Mean square error for histograms when estimating Radon-Nikodym derivatives

Paulo Eduardo Oliveira*
Univ. Coimbra, Dep Matemática
Apartado 3008, 3000 Coimbra, Portugal

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 thus obtained recovers the $n^{-2/3}$ 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 completement results by Jacob, Oliveira [8, 10], where histograms were considered to estimate Radon-Nikodym derivatives of point processes or, to be more accurate, compound point processes. The general idea is to define two integrable point processes 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 [5] for a particular choice of ξ and η, in order to address density estimation, Jacob, Mendes Lopes [6], where instead of point processes the authors considered absolutely continuous random measures, thus setting the problem in terms of the random densities associated, Jacob, Oliveira [8, 9, 10] wit the same framework as stated here. All the references quoted above suppose independent sampling and, with the exception of [8] and [10] 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] and Roussas [11, 12, 13] considered associated sampling. All these extensions deal with kernel estimates.

Besides the convergence in several modes of the estimators defined, some references address also the mean squared convergence rate. 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. The complements of the results of Jacob, Oliveira [8, 10] 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].

*This author was supported by grant PRAXIS/2/2.1/MAT/19/94 from JNICT and Instituto de Telecomunicações - Pólo de Coimbra

2 Auxiliary results

Although in [8] and [10] the authors considered point processes 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 $\mathbb{I}_A$ the indicator function of the set A.

- (Ellis [5]) Density estimation: take $\xi = \nu$ a.s., $\eta = \delta_X$, where X is a random variable. 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 $\alpha_n = \mathbb{I}_B(T_n)$, 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 is characterized by a markovian kernel of distributions $(\pi_x,\, x \in \mathbb{R}^p)$ with means $(\mu_x,\, x \in \mathbb{R}^p)$ 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 $B \in \mathcal{B}$ fixed, we have $\frac{d\mu}{d\nu}(x) = \mu_x(B)$ 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 $\mu_{x_1} \otimes \cdots \otimes \mu_{x_n}$ (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 $\frac{d\mu}{d\nu}(x) = \mu_x(B) = \mathrm{P}(Y \in B|X = x)$.

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 $k \in \mathbb{N}$, $\Pi_k \subset \mathcal{B}$;

(P2) for each $k \in \mathbb{N}$, Π_k is finite;

(P3) $\sup\{\text{diam}(I) : I \in \Pi_k\} \longrightarrow 0$;

(P4) for each $k \in \mathbb{N}$ and $I \in \Pi_k$, $\nu(I) > 0$;

(P5) for each $k \in \mathbb{N}$, $h_k = \lambda(I)$, $I \in \Pi_k$ and $\lim_{k \to +\infty} h_k = 0$.

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 precised to obtain the convergence.

As we are not working with embedded partitions we need the following assumptions, as in Jacob, Oliveira [10]. A measure m on $\mathbb{R}^p \times \mathbb{R}^p$ verifies 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, such that $\nu^*(A^*) = \nu(A)$ 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 [10] 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 serve as a tool for establishing the convergence rates.

The following is essential for the analysis of the estimator.

Theorem 2.1 (Jacob, Oliveira [10]) *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 [10] if

$$nh_n \longrightarrow +\infty \tag{1}$$

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

$$\mathrm{E}\xi^k(I)\leq R^{k-2}k!\,\mathrm{E}\xi^2(I)$$

$$\mathrm{E}\eta^k(I)\leq R^{k-2}k!\,\mathrm{E}\eta^2(I)$$

the convergence is uniformly almost complete.

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 $\mathrm{E}\xi\otimes\xi$, $\mathrm{E}\eta\otimes\eta$ verify **(M)** and (1) is satisfied,

$$f_n(s)=\frac{1}{nh_n}\sum_{i=1}^{n}\xi_i(I_n(s))\longrightarrow f(s)$$

$$g_n(s)=\frac{1}{nh_n}\sum_{i=1}^{n}\eta_i(I_n(s))\longrightarrow g(s)$$

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

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 [7]) *Let X and Y be non-negative integrable random variables then, for $\varepsilon>0$ small enough,*

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

3 The convergence rates

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 *Let f and g be versions of $\frac{d\nu}{d\lambda}$ and $\frac{d\mu}{d\lambda}$, respectively, continuously differentiable on the compact set B. Suppose the sequence of partitions Π_k, $k\in\mathbb{N}$, verifies* **(P1)**-**(P5)***, the moment measures $\mathrm{E}\xi\otimes\xi$, $\mathrm{E}\eta\otimes\eta$ both verify* **(M)** *and (1) holds. Then*

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

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

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

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

$$\mathrm{E}f_n(s) = \frac{1}{h_n}\nu(I_n(s)) = \frac{1}{h_n}\int_{I_n} f(t)\ \lambda(dt)$$

(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 $t=(t_1,\ldots,t_p)$,

$$f(t) = f(s) + \sum_{k=1}^{p} \frac{\partial f}{\partial x_k}(s)(t_k-s_k) + O(\| t-s \|^2),$$

thus

$$\begin{aligned}
\mathrm{E}f_n(s) - f(s) &= \frac{1}{h_n}\int_{I_n} \sum_{k=1}^{p} \frac{\partial f}{\partial x_k}(s)(t_k-s_k)\ \lambda(dt) + \frac{1}{h_n}\int_{I_n} O(\| t-s \|^2)\ \lambda(dt) \\
&= O(h_n)\sum_{k=1}^{p} \frac{\partial f}{\partial x_k}(s) + O(h_n^2)
\end{aligned}$$

as $\| t-s \| \le h_n$ and $\lambda(I_n)=h_n$. On the other hand

$$\mathrm{Var}(f_n(s)) = \frac{1}{nh_n^2}\mathrm{E}\xi^2(I_n) - \frac{1}{nh_n^2}\mathrm{E}^2\xi(I_n).$$

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

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

and

$$\begin{aligned}
\frac{m_2^{\xi,\xi}(I_n^*)}{h_n} &= \frac{1}{h_n}\int_{I_n^*} \gamma_2^{\xi,\xi}(t,t)\ \lambda^*(dt) \\
&= \frac{1}{h_n}\int_{I_n^*} \gamma_2^{\xi,\xi}(s,s)\ \lambda^*(dt) + \frac{1}{h_n}\int_{I_n^*} \sum_{k=1}^{p} \frac{\partial\gamma_2^{\xi,\xi}}{\partial x_k}(s+\theta_k(t_k-s_k))(t_k-s_k)\ \lambda^*(dt) \\
&= \gamma_2^{\xi,\xi}(s,s) + \sum_{k=1}^{p} \frac{\partial\gamma_2^{\xi,\xi}}{\partial x_k}(s)O(h_n) + +o(\frac{1}{nh_n}),
\end{aligned}$$

so, according to (1),

$$\frac{m_2^{\xi,x\xi}(I_n^*)}{nh_n^2} = \frac{\gamma_2^{\xi,\xi}(s,s)}{nh_n} + o(\frac{1}{nh_n}).$$

The remaining term $\frac{1}{nh_n^2}\mathrm{E}^2\xi(I_n) = \frac{1}{n}(\frac{\nu(I_n)}{h_n})^2$ is clearly an $O(\frac{1}{n})$, so, gathering all these approximations, we get the result announced. 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 development of $\mathrm{E}f_n(s) - f(s)$ we would find the integral

$$\frac{1}{h_{n,1}\cdots h_{n,p}}\int_{a_{n,1}}^{a_{n,1}+h_{n,1}}\cdots\int_{a_{n,p}}^{a_{n,p}+h_{n,p}}\sum_{k=1}^{p}\frac{\partial f}{\partial x_k}(s)(t_k-s_k)\,dt_1\cdots dt_p = \sum_{k=1}^{p}\frac{h_{n,k}^2-2h_{n,k}(s_k-a_{n,k})}{2h_{n,k}}.$$

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

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

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

Lemma 3.2 *Let f be a version of $\frac{d\mu}{d\nu}$ continuously differentiable on B. Suppose the sequence of partitions Π_k, $k\in\mathbb{N}$, verifies* **(P1)**-**(P5)**, *the moment measure $\mathrm{E}\xi\otimes\xi$ verify* **(M)**, *that there exists $R>0$ such that, for every $I\subset B$ and $k=3,4$,*

$$\mathrm{E}\xi^k(I) \le R\mathrm{E}\xi^2(I), \qquad \mathrm{E}\eta^k(I)\le R\mathrm{E}\eta^2(I). \tag{3}$$

Finally, if (1) holds,

$$\mathrm{E}(f_n(s)-f(s))^4 = O\left(h_n^2 + \frac{h_n}{n} + \frac{1}{n^2h_n^2}\right).$$

Proof : We develop

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

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

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

and

$$\mathrm{E}(f_n(s)-f(s))^2(\mathrm{E}f_n(s)-f(s))^2 = O(\frac{h_n^2}{nh_n}).$$

Developing now the third order moment, we find

$$\mathrm{E}(f_n(s)-\mathrm{E}f_n(s))^3 = \frac{1}{n^2h_n^3}\mathrm{E}\xi^2(I_n) - \frac{3}{n^2h_n^3}\mathrm{E}\xi^2(I_n)\nu(I_n) + \frac{2}{n^2h_n^3}\nu^3(I_n).$$

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

$$\frac{\mathrm{E}\xi^3(I_n)}{n^2h_n^3} \leq R\frac{\mathrm{E}\xi^2(I_n)}{n^2h_n^3} = \frac{R}{n^2h_n^2}\left(\frac{m_1^{\xi,\xi}(I_n\times I_n)}{h_n} + \frac{m_2^{\xi,\xi}(I_n^*)}{h_n}\right) = O(\frac{1}{n^2h_n^2})$$

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

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

$$\mathrm{E}(f_n(s)-\mathrm{E}f_n(s))^4 =$$

$$= \frac{1}{n^3h_n^4}\mathrm{E}\xi^4(I_n) - \frac{4}{n^3h_n^4}\mathrm{E}\xi^3(I_n)\nu(I_n) + \frac{6}{n^nh_n^4}\mathrm{E}\xi^2(I_n)\nu^2(I_n) - \frac{4}{n^3h_n^4}\nu^4(I_n) +$$

$$+\frac{3(n-1)}{n^3h_n^4}\mathrm{E}^{\ 2}(\xi(I_n)-\nu(I_n))^2.$$

Applying again (3) and reproducing the same arguments as above, it is easily checked that the sum of the first four terms is an $O(\frac{1}{n^2h_n^3})$. The last term, again after development and using **(M)** is easily found to be an $O(\frac{1}{n^2h_n^2})$. So as (1) holds we finally get $\mathrm{E}(f_n(s)-\mathrm{E}f_n(s))^4 = O(\frac{1}{n^2h_n^2})$, 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 *Let f and g be versions of $\frac{d\nu}{d\lambda}$ and $\frac{d\mu}{d\lambda}$, respectively, continuously differentiable on B. Suppose the sequence of partitions Π_k, $k\in\mathbb{N}$, verifies* **(P1)-(P5)**, *the moment measures $\mathrm{E}\xi\otimes\xi, \mathrm{E}\eta\otimes\eta$ both verify* **(M)** *and (3) hold. Further, suppose that there exist real numbers $\beta>\alpha>0$ such that*

$$nh_n^{4\alpha+2\beta+1} \longrightarrow +\infty \tag{4}$$

and that

$$\mathrm{E}(\varphi_n^4(s)\mathrm{I\!I}_{\{\varphi_n(s)>h_n^{-\alpha}\}}) \longrightarrow 0, \tag{5}$$

then

$$\mathrm{E}(\varphi_n(s)-\varphi(s))^2 = \frac{O(h_n^2)}{f^2(s)}\left(\varphi(s)\sum_{k=1}^p\frac{\partial f}{\partial x_k}(s) - \sum_{k=1}^p\frac{\partial g}{\partial x_k}(s)\right)^2 +$$

$$+\frac{1}{nh_nf^2(s)}\left(\varphi^2(s)\gamma_2^{\xi,\xi}(s,s) - 2\varphi(s)\gamma_2\xi,\eta(s,s) + \gamma_2\eta,\eta(s,s)\right) +$$

$$+o(h_n^2+\frac{h_n^{1/2}}{n^{1/2}}+\frac{1}{nh_n}+\frac{h_n^{5/4}}{n^{1/4}}+\frac{1}{n^{3/4}h_n^{1/4}})$$

Proof : We will go through each term in (2) to derive the convenient rates to 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 $o(h_n^2+\frac{h_n^{1/2}}{n^{1/2}}+\frac{1}{nh_n}+\frac{h_n^{5/4}}{n^{1/4}}+\frac{1}{n^{3/4}h_n^{1/4}})$. For this put $\widetilde{\varphi}_n(s)=\frac{\mathrm{E}g_n(s)}{\mathrm{E}f_n(s)}$ and write

$$
\begin{aligned}
&\mathrm{E}[(\varphi_n^2(s)-\varphi^2(s))(f_n(s)-f(s))^2]= \\
&\qquad =\mathrm{E}[(\varphi_n^2(s)-\widetilde{\varphi}_n^2(s))(f_n(s)-f(s))^2]+(\widetilde{\varphi}_n^2(s)-\varphi^2(s))\mathrm{E}(f_n(s)-f(s))^2.
\end{aligned}
$$

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

$$(\widetilde{\varphi}_n^2(s)-\varphi^2(s))\mathrm{E}(f_n(s)-f(s))^2=o(h_n^2+\frac{1}{nh_n}).$$

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

$$
\begin{aligned}
&\mathrm{E}[(\varphi_n^2(s)-\widetilde{\varphi}_n^2(s))(f_n(s)-f(s))^2]\leq \\
&\quad\leq(\alpha_n+\widetilde{\varphi}_n(s))\mathrm{E}\left[|\varphi_n(s)-\widetilde{\varphi}_n(s)|\,(f_n(s)-f(s))^2\mathbb{I}_{\{\varphi_n(s)\leq\alpha_n\}}\mathbb{I}_{\{|\varphi_n(s)-\widetilde{\varphi}_n(s)|\leq\varepsilon\}}\right]+ \\
&\quad+(\alpha_n+\widetilde{\varphi}_n(s))\mathrm{E}\left[|\varphi_n(s)-\widetilde{\varphi}_n(s)|\,(f_n(s)-f(s))^2\mathbb{I}_{\{\varphi_n(s)\leq\alpha_n\}}\mathbb{I}_{\{|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon\}}\right]+ \qquad (6) \\
&\quad+\mathrm{E}[(\varphi_n^2(s)-\widetilde{\varphi}_n^2(s))(f_n(s)-f(s))^2\mathbb{I}_{\{\varphi_n(s)>\alpha_n\}}].
\end{aligned}
$$

The first term of this expansion is bounded above by

$$(\alpha_n+\widetilde{\varphi}_n(s))\varepsilon_n\mathrm{E}(f_n(s)-f(s))^2=o(h_n^2+\frac{1}{nh_n}),$$

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

The second term in (6) is bounded above by

$$
\begin{aligned}
&(\alpha_n+\widetilde{\varphi}_n(s))^2\mathrm{E}\left[(f_n(s)-f(s))^2\mathbb{I}_{\{|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon\}}\right]\leq \\
&\quad\leq(\alpha_n+\widetilde{\varphi}_n(s))^2\mathrm{E}^{1/2}(f_n(s)-f(s))^4\,\mathrm{P}^{1/2}\left(|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon\right). \qquad (7)
\end{aligned}
$$

According to lemma 2.2, we have

$$\mathrm{P}\left(|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon\right)\leq\mathrm{P}\left(|g_n(s)-\mathrm{E}g_n(s)|>\frac{\varepsilon_n}{4}\mathrm{E}f_n(s)\right)+\mathrm{P}\left(|f_n(s)-\mathrm{E}f_n(s)|>\frac{\varepsilon_n}{4}\frac{\mathrm{E}^2f_n(s)}{\mathrm{E}g_n(s)}\right).$$

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

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

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

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

The choice of the sequences α_n and ε_n acerts that

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

so the second term in (6) is an $o(h_n^2 + \frac{h_n^{1/2}}{n^{1/2}} + \frac{1}{nh_n})$.

We look now at the third term in (6). Applying Hölder's inequality, this term is bounded above by

$$\mathrm{E}^{1/2}(f_n(s) - f(s))^4 \mathrm{E}^{1/2} \left(\varphi_n^4(s) \mathrm{I\!I}_{\{\varphi_n(s) > \alpha_n\}} \right) + \mathrm{E}^{1/2}(f_n(s) - f(s))^4 \widetilde{\varphi}_n^2(s) \, \mathrm{P}^{1/2} \left(\varphi_n(s) > \alpha_n \right)$$

and this is an $o(h_n^2 + \frac{h_n^{1/2}}{n^{1/2}} + \frac{1}{nh_n})$ according to (5).

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

$$\mathrm{E}[(\varphi_n(s) - \varphi(s))(f_n(s) - f(s))(g_n(s) - g(s))] \leq$$

$$\leq \mathrm{E}^{1/2} \left[(\varphi_n(s) - \varphi(s))^2 (f_n(s) - f(s))^2 \right] \mathrm{E}^{1/2}(g_n(s) - g(s))^2.$$

The first factor is further bounded by

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

the analysis of which proceeds as the one made for the second term from (2), showing a convergence rate of $o^{1/2}(h_n^2 + \frac{h_n^{1/2}}{n^{1/2}} + \frac{1}{nh_n})$. The factor $\mathrm{E}(g_n(s) - g(s))^2 = O(\frac{1}{nh_n} + h_n^2)$, so we finally have a convergence rate $o(h_n^2 + \frac{h_n^{1/2}}{n^{1/2}} + \frac{1}{nh_n} + \frac{h_n^{5/4}}{n^{1/4}} + \frac{1}{n^{3/4} h_n^{1/4}})$, which concludes the proof. ■

The optimization of h_n indicates one should choose $h_n = cn^{-1/3}$, whether one bases this optimization on the convergence rate given in lemma 3.2 or in theorem 3.3. In this case, we get

$$\mathrm{E}(f_n(s) - f(s))^2 = O(n^{-2/3})$$

$$\mathrm{E}(g_n(s) - g(s))^2 = O(n^{-2/3})$$

$$\mathrm{E}[(f_n(s) - f(s))(g_n(s) - g(s))] = O(n^{-2/3})$$

$$\mathrm{E}(f_n(s) - f(s))^4 = O(n^{-4/3})$$

$$\mathrm{E}(\varphi_n(s) - \varphi(s))^2 = O(n^{-2/3}),$$

thus finding the $n^{-2/3}$ convergence rate which is well known for density or regression estimation, for example, although (5) means some restrictions in each case. Also, theorem 3.3 is applicable choosing, for instance, $\beta = 2\alpha < 1/2$.

References

[1] Bensaïd, N., Fabre, J., Estimation par noyau d'une derivée de Radon-Nikodym sous des conditions de mélange, Canadian J. Statist., to appear in 1998

[2] Bensaïd, N., Jacob, P., Erreur quadratique asymptotique optimale pour les estimateurs à noyau de dérivées de Radon-Nikodym générales, Preprint

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

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

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

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

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

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

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

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

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