Histogram estimation of Radon-Nikodym derivatives for strong mixing data*

Nadia Bensaïd
Lab. Probab. Stat.,Univ. Montpellier II
34095 Montpellier Cedex 05, France

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

Abstract

Nonparametric inference for point processes is discussed by way histograms, which provide a nice tool for the analysis of on-line data. The construction of histograms depends on a sequence of partitions, which we take to be nonembedded to allow partitions with sets of equal measure. This presents some theoretical problems, which are addressed with an assumption on the decomposition of second order moments. In another direction, we drop the usual independence assumption on the sample, replacing it by a strong mixing assumption. Under this setting, we study the convergence of the histogram in probability, which depends on approximation conditions between the distributions of random pairs and the product of their marginal distributions, and almost completely, which is based on the decomposition of the second order moments. This last convergence is stated on two versions according to the assumption of Laplace transforms or the Cramer moment conditions. These are somewhat stronger, bet enable us to recover the usual condition on the decrease rate of sets on each partition. In the final section we prove that the finite dimensional distributions converge in distribution to a gaussian centered vector with a specified covariance.

AMS Classification: 62G05, 62G20

1 Introduction

The use of point processes or random measures to address estimation problems, besides the interesting mathematical questions it raises by its own nature, provides a general setting that is able to include several classical functional estimation problems, such as density or regression estimation. These particular models have attracted the attention of a large number of statisticians producing a wide literature. An account of this literature and available results may be found in Bosq, Lecoutre [6] for a more classical approach to these questions, or more recently in Bosq [4], Bosq, Nguyen [5] or Karr [24] for a point process or stochastic process point of view. Other estimation problems, more directly concerned with point processes, include estimation of Palm distributions, Karr [24, 25, 26], Niéré [28] or Crétois [9], mean local distributions of composed random measures, Mendes Lopes [27] or Saleh [36, 37], or thinning functions, Bensaïd [2], to give some references where a framework similar to ours was used, that is, two integrable point processes were considered such that their mean measures satisfy an absolute continuity relation and we are interested in estimating the therefore existing Radon-Nikodym derivative. Later in this section, we will briefly show some examples

*Preprint

†This author was supported by grant PRAXIS/2/2.1/MAT/400/94 from FCT and Centro de Matemática da Universidade de Coimbra.

of how the derivative may be conveniently interpreted. The method for construction estimators divides essentially between two types: histograms, used by Crétois [9], Karr [25, 26], Jacob, Oliveira [19, 21, 22] or Oliveira [29], or kernel used by Bensaïd, Fabre [3], Ellis [12], Ferrieux [14, 15], Jacob, Mendes Lopes [17], Jacob, Oliveira [20] or Roussas [31, 32, 33]. In some of these references some particular properties on the point processes or random measures were supposed, namely, supposing one of the random measures to be deterministic, as in [12, 31, 32, 33], or supposing that both the random measures are absolutely continuous with respect to some reference measure as in [17]. The kernel approach has become quite popular as it produces smoother approximations and provides better convergence rates. Histograms are, nevertheless, more suitable to real time data although the convergence rates are slower. In this article we will study the above mentioned approximations based on histograms, following the work in Jacob, Oliveira [19, 20, 22]. This family of estimators requires the introduction of a suitable sequence of partitions of the base space or, at least, of some convenient fixed subset. For the theoretical treatment the simplest is to require that the this sequence of partitions is embedded, that is, to pass from one partition to the next one we divide some of the sets. This requirement was adopted, for example, in Jacob, Oliveira [19] and gives access to martingale characterizations which are very convenient for the asymptotic treatment. On another hand, for a more practical point of view, it is quite natural to impose that the sets in each partition are of same size (with respect to some fixed measure). These two requirements together produce partitions with sets that decrease to fast, preventing the known convergence conditions to be accomplished in some cases. Some solutions to this problem link the size of the sets to moments of the unknown distribution as is done in Karr [24], for example. Other approaches were proposed, among others, by Abou-Jaoudé [1] or Grenander [16]. In Jacob, Oliveira [21, 22] another approach to this question was proposed. To our view the last approach seems better suited for point processes and keeps conditions on the sets and conditions on the moments of the unknown distributions quite well separated, which we believe is advantageous.

The behaviour of the estimator under dependence assumptions on the sample is another matter of interest. For independent sampling this study of the convergence was carried out in Jacob, Oliveira [19] with the convergence rates studied in Oliveira [29] for histograms and in Ellis [12], Jacob, Mendes Lopes [17] or Jacob, Oliveira [20] for kernels. The treatment of dependent samples was addressed to in Bensaïd [2], Bensaïd, Fabre [3] , Cai, Roussas [8] for strong mixing samples and by Ferrieux [14, 15] and Roussas [31, 32, 33] for associated samples with kernels. Histograms were studied under associated sampling by Jacob, Oliveira [22]. Here we will be interested in investigating the behaviour of histograms under strong mixing sampling, establishing conditions for the convergence in probability and almost completely of the estimator and also for the convergence in distribution of its finite dimensional distributions.

We now define our framework more precisely. Let $\mathbf{S}$ be a complete separable and locally compact metric space, $\mathcal{B}$ the ring of relatively compact Borel subsets of $\mathbf{S}$, and $\mathcal{M}$ the space of nonnegative Radon measures on $\mathbf{S}$. A random measure is any function defined on some probability space with values in $\mathcal{M}$ measurable with respect to the σ-algebra induced by the vague topology (see Daley, Vere-Jones [10], Kallenberg [23] or Karr [24] for basic properties on random measures and characterizations of vague convergence). A point process is, as usual, a random measure that, almost surely, takes values in the subspace of $\mathcal{M}$ of the integer valued measures. In what follows ξ and η are random measures or, in most cases, point processes, that are supposed integrable, that is, $\nu(B) = \mathrm{E}[\xi(B)]$ and $\mu(B) = \mathrm{E}[\eta(B)]$ are nonnegative Radon measures on $\mathbf{S}$, and such that their mean measures $\mu = \mathrm{E}(\eta)$, $\nu = \mathrm{E}(\xi)$ satisfy the absolute continuity assumption $\mu \ll \nu$. We will be interested in estimating a version of the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$. Throughout this article we will denote by $\mathbb{I}_A$ the indicator function of the set A.

For sake of completeness we mention some examples of the interpretation of the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$.

- Density estimation: let ν be a measure on $\mathbf{S}$ and take $\xi = \nu$ a.s., $\eta = \delta_X$, where X is a random variable with absolutely continuous distribution 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 nonnegative real random variable and X a random variable on $\mathbf{S}$. 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 $\mathbf{S}$, 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 $p = \frac{d\mu}{d\nu}$ 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 $\mathbf{S} \times \mathrm{T}$ 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 $\mathbf{S}\times\mathbf{S}$ 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 $(\pi_x,\ x \in \mathbf{S})$ with means $(\mu_x,\ x \in \mathbf{S})$ 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 $(Y_1, \ldots, Y_n)$ 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 $\frac{d\mu}{d\nu}(x) = \mu_x(B) = \mathrm{P}(Y \in B|X = x)$.

As suggested by the examples above, we are interested in estimating a version of $\frac{d\mu}{d\nu}$ based on a sample $\Big((\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)\Big)$ of the random pair (ξ, η).

2 Assumptions and auxiliary results

In order to define the histograms we need a sequence of partitions Π_k, $k \in \mathbb{N}$. As we do not use embedded partitions it will not possible to construct approximations on the whole space $\mathbf{S}$ but only on some fixed compact set $B \subset \mathbf{S}$, as it will be justified below. We make the following assumptions on the partitions:

(P1) for each $k \in \mathbb{N}$, $\Pi_k \subset \mathcal{B}$;

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

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

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

(P5) $\max_{I\in\Pi_k} \nu(I) \longrightarrow 0$.

When including in our conditions **(P4)** or **(P5)** we may specify the measure to which these conditions are to be satisfied, meaning that ν is to be replaced by some other measure. When we do not state any such reference we mean **(P4)** and **(P5)** with the current statement.

Given $s \in B$ we denote by $I_k(s)$, or only I_k if there is no confusion about the point s, the unique set of Π_k containing the point s, and define, for each $k \in \mathbb{N}$, the function

$$\widetilde{f}_k(s) = \sum_{I\in\Pi_k} \frac{\mu(I)}{\nu(I)} \mathbb{I}_I(s) = \frac{\mu(I_k(s))}{\nu(I_k(s))}.$$

It is well known that, if there exists a continuous version f of the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$, and if the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies **(P1)-(P4)**, then

$$\lim_{k\to+\infty} \sup_{s\in B} \left| f(s) - \widetilde{f}_k(s) \right| = 0. \tag{1}$$

The fact that we take the supremum on a compact set is essential for this convergence and justifies our choice on partitioning only a fixed compact set. Also due to this property, we suppose on the sequel that there exists a continuous version f of $\frac{d\mu}{d\nu}$.

We now set our first assumptions on the point processes. Throughout this article we will consider the sequence (ξ_n, η_n), $n \in \mathbb{N}$, of point processes to be strictly stationary and strong mixing with coefficients denoted, as usual, by α_n (see Doukhan [11] for details on strong mixing and other types of mixing). We will be interested in constructing an estimator of f based on the elements $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ of the above mentioned sequence. For this purpose, define the random measures $\overline{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$ and $\overline{\eta}_n = \frac{1}{n}\sum_{i=1}^n \eta_i$. The histogram estimator of f is then defined by

$$f_n(s) = \sum_{I\in\Pi_{k(n)}} \frac{\overline{\eta}_n(I)}{\overline{\xi}_n(I)} \mathbb{I}_I(s) = \frac{\overline{\eta}_n(I_{k(n)}(s))}{\overline{\xi}_n(I_{k(n)}(s))}, \tag{2}$$

defining, as usual, $f_n(s)$ as zero whenever the denominator vanishes. The convergence of f_n to f follows from the convergence to zero of $f_n - \widetilde{f}_{k(n)}$ if the sequence of partitions satisfies **(P1)-(P4)**. When considering embedded partitions this convergence reduces to a martingale argument which is no longer available in the present framework. This argument is replaced by convenient assumptions on second order moment measures which we describe next.

We will say that a measure m on $\mathbf{S}\times\mathbf{S}$ satisfies condition **(M)** with respect to ν if

(M)

- $m = m_1 + m_2$ where m_2 is a measure on Δ, the diagonal of $\mathbf{S}\times\mathbf{S}$, and m_1 a measure on $\mathbf{S}\times\mathbf{S} \setminus \Delta$;
- $m_1 \ll \nu\otimes\nu$ and there exists a bounded version of $\frac{dm_1}{d\nu\otimes\nu}$;
- $m_2 \ll \nu^*$, where ν^* is the measure on Δ defined by lifting ν, that is, given $A^* \subset \Delta$ we have $A^* = \{(s,s) : s \in A\}$ and $\nu^*(A^*) = \nu(A)$ and there exits a bounded and continuous version of $\frac{dm_2}{d\nu^*}$.

Note that the absolute continuity is with respect to measures that depend on ν. As it will be apparent later it is important that this is the same measure appearing in **(P4)** and **(P5)**.

This condition **(M)** was first used in Jacob, Oliveira [20] for the treatment of the kernel estimator for independent sampling and later in Oliveira [29] for the histogram under independence, Ferrieux [14, 15] for the kernel estimator, Jacob, Oliveira [22] for the histogram for associated samples for these three last references. A 3-dimensional variation of this type of conditions was considered in Bensaïd, Fabre [3] for the study of the kernel estimator under strong mixing sampling.

In order to control $\mathrm{Var}(\overline{\xi}_n(I_{k(n)}))$ and $\mathrm{Var}(\overline{\eta}_n(I_{k(n)}))$ we introduce a further assumption on the sequence (ξ_n, η_n), $n \in \mathbb{N}$ (recall that we have supposed this sequence to be strictly stationary).

(M1)

Let $\zeta_{1,1} \in \{\xi_1, \eta_1\}$ and $\zeta_{2,k} = \xi_k$ for every $k \in \mathbb{N}$ or $\zeta_{2,k} = \eta_k$ for every $k \in \mathbb{N}$;

Each measure $\mathrm{E}(\zeta_{1,1} \otimes \zeta_{2,k})$ satisfies **(M)** with respect to ν with Radon-Nikodym derivatives $g^1_{\zeta_{1,1},\zeta_{2,k}}$ and $g^2_{\zeta_{1,1},\zeta_{2,k}}$ for the nondiagonal and diagonal components, respectively;

$\sum_{k=2}^{\infty} \left| g^1_{\zeta_{1,1},\zeta_{2,k}} - f_{\zeta_{1,1}} \otimes f_{\zeta_{2,k}} \right|$ converges uniformly on $B \times B$ to a bounded function $G^1_{\zeta_1,\zeta_2}$, where $f_{\zeta_{i,k}}$ is the function identically equal to 1 if $\zeta_{i,k} \in \{\xi_1, \xi_2, \ldots\}$ or is the function f if $\zeta_{i,k} \in \{\eta_1, \eta_2, \ldots\}$.

$\sum_{k=2}^{\infty} g^2_{\zeta_{1,1},\zeta_{2,k}}$ converges uniformly on $B \times B$ to a continuous function $G^2_{\zeta_1,\zeta_2}$.

This condition has been used in Bensaïd, Fabre [3] for the study of the kernel estimator under present framework. Conditions similar to **(M1)** appear also in Bosq [4] or in Roussas [33, 35]. In Jacob, Oliveira [22] the treatment of the histogram estimator for associated sampling is carried under assumption **(M)** only. The reason we need the extra condition **(M1)** is due to the fact that, under our present framework, the measures $\mathrm{Cov}(\zeta_1, \zeta_2)$ are, in general, signed measures. This makes the decomposition **(M)** not the most natural to control these covariance measures.

The assumptions being presented we now proceed with some auxiliary results. To separate the variables in (2) we use the following lemma.

Lemma 2.1 ([18]) *Let X and Y be non-negative integrable random variables then, for $\varepsilon > 0$ small enough,*

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

Using this Lemma, it follows that, for $\varepsilon > 0$ small enough,

$$\begin{aligned} \left\{ \left| f_n(s) - \widetilde{f}_{k(n)}(s) \right| > \varepsilon \right\} &= \left\{ \left| \frac{\overline{\eta}_n(I_{k(n)}(s))}{\overline{\xi}_n(I_{k(n)}(s))} - \frac{\mu(I_{k(n)}(s))}{\nu(I_{k(n)}(s)} \right| > \varepsilon \right\} \\ &\subset \left\{ \left| \overline{\eta}_n(I_{k(n)}(s)) - \mu(I_{k(n)}(s)) \right| > \frac{\varepsilon}{4} \nu(I_{k(n)}(s)) \right\} \cup \left\{ \left| \overline{\xi}_n(I_{k(n)}(s)) - \nu(I_{k(n)}(s)) \right| > \frac{\varepsilon}{4} \frac{\nu^2(I_{k(n)}(s))}{\mu(I_{k(n)}(s))} \right\}. \end{aligned} \tag{3}$$

There are several versions of the Bernstein inequality. We will use in this article the following version which, up to a constant change, is contained in Theorem 2.1.5 of Fabre [13].

Lemma 2.2 *Let X_k, $k \in \mathbb{Z}$, a strong mixing sequence of real random variables and, for each $n \in \mathbb{N}$, $S_n = \sum_{i=1}^n [X_i - \mathrm{E}(X_i)]$. Suppose that the sequence of Laplace transforms of $X_k, k \in \mathbb{Z}$, is uniformly bounded in some interval $[-\delta, \delta]$. Then, for every $n, \gamma \geq 2$, $\varepsilon > 0$ and $p \in [1, n/2]$, we have*

$$\mathrm{P}\left(|S_n| \geq n\varepsilon\right) \leq 6 \exp\left(\frac{-nt\varepsilon}{30p}\right) + 6\frac{n}{p}\left(\frac{10M_\gamma}{\varepsilon} + 1\right)^{\frac{\gamma}{2\gamma+1}} \alpha_p^{\frac{2\gamma}{2\gamma+1}}, \tag{4}$$

where $M_\gamma = \sup_{k \in \mathbb{Z}} \| X_k \|_\gamma$, $t = \min(\frac{\delta}{2}, \frac{\varepsilon}{3c})$ and $c = 4 \sup_{k \in \mathbb{Z}} \sum_{i=2}^\infty \frac{\delta^{i-2}}{i!} \mathrm{E}(|X_k|^i)$.

Next, we quote the following lemma, which is a modification, by Rhomari [30] (see Lemma 1.2 in [4]), of the well known coupling lemma due to Bradley [7].

Lemma 2.3 ([7, 30]) *Let (X, Y) be a $\mathbb{R}^p \times \mathbb{R}$ valued random vector such that $Y \in \mathcal{L}^q$ for some $q \in [1, +\infty]$. Let $c \in \mathbb{R}$ be such that $\| Y + c \|_q > 0$ and $u \in (0, \| Y + c \|_q]$. Then there exists a random variable Y^* such that*

a) Y^* has the same distribution as Y and is independent of X;

b) $\mathrm{P}\left(|Y^* - Y| > u\right) \leq 11 \left(\frac{\| Y + c \|_q}{u}\right)^{\frac{q}{2q+1}} \left[\alpha(\sigma(X), \sigma(Y))\right]^{\frac{2q}{2q+1}}$

where $\alpha(\sigma(X), \sigma(Y)) = \sup\left\{|\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)|,\ A \in \sigma(X),\ B \in \sigma(Y)\right\}$.

The next two results concern the use of our conditions on the decomposition of moment measures **(M)** and **(M1)**. For the use of **(M)** we have the following convergence.

Lemma 2.4 ([21, 22]) *Suppose m is a measure on* $\mathbf{S}\times\mathbf{S}$ *that satisfies condition* **(M)** *with respect to ν and the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies* **(P1)**-**(P5)**. *Then, denoting by g^2 a bounded and continuous version of $\frac{dm_2}{d\nu^*}$,*

$$\sum_{I \in \Pi_k} \frac{m(I \times I)}{\nu(I)}\, \mathbb{I}_I(s) \longrightarrow g^2(s, s)$$

uniformly on B.

From **(M1)** we may characterize the convergence of convenient sums of covariances.

Lemma 2.5 (Bensaïd, Fabre [3]) *Suppose that* **(M1)** *are satisfied with respect to ν and the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies* **(P1)**-**(P5)**. *Then,*

$$\frac{1}{n\,\nu(I_{k(n)})} \sum_{i=1}^n \sum_{j=1}^i \mathrm{Cov}\Big(\zeta_{1,1}(I_{k(n)}),\, \zeta_{2,j}(I_{k(n)})\Big) \longrightarrow \sum_{j=1}^\infty g^2_{\zeta_{1,1},\zeta_{2,j}}.$$

3 Convergence of the estimator

As (1) states the uniform convergence on B of $\widetilde{f}_k$ to f, it suffices, in order to prove the convergence of the estimator, to check that $f_n - \widetilde{f}_{k(n)}$ converges to zero. Now, taking account of (3), this will follow from the convergence to zero of both $\overline{\xi}_n(I_{k(n)}) - \nu(I_{k(n)})$ and $\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})$. With this in mind, we first look at the convergence in probability of the histogram.

Theorem 3.1 *Let (ξ_n, η_n), $n \in \mathbb{N}$, be a strictly stationary strong mixing sequence of integrable random measures satisfying* **(M1)**. *Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies* **(P1)**-**(P5)** *and*

$$n \min_{I \in \Pi_{k(n)}} \nu(I) \longrightarrow +\infty. \tag{5}$$

Then $f_n(s)$ converges in probability to $f(s)$ for every $s \in B$.

Proof : We will only look at the term corresponding to ξ appearing from (3), the other being treated analogously. Applying Chebyshev's inequality we find the upper bound

$$\mathrm{P}\left(\left|\overline{\xi}_n(I_{k(n)}) - \nu(I_{k(n)})\right| > \frac{\varepsilon}{4}\,\frac{\nu^2(I_{k(n)})}{\mu(I_{k(n)})}\right) \leq \frac{16}{\varepsilon^2}\,\frac{\mu^2(I_{k(n)})}{\nu^2(I_{k(n)})}\,\frac{1}{n\,\nu(I_{k(n)})}\,\frac{n\,\mathrm{Var}\Big(\overline{\xi}_n(I_{k(n)})\Big)}{\nu(I_{k(n)})}.$$

As $\frac{\mu^2(I_{k(n)})}{\nu^2(I_{k(n)})}$ converges according to (1) it suffices to check that the last quotient is convergent. Now

$$\mathrm{Var}\Big(\overline{\xi}_n(I_{k(n)})\Big) = \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{i}\mathrm{Cov}\Big(\xi_1(I_{k(n)}),\, \xi_j(I_{k(n)})\Big) - \frac{1}{n}\,\mathrm{Var}\Big(\xi(I_{k(n)})\Big),$$

so it follows from Lemmas 2.4 and 2.5 that

$$\frac{n}{\nu(I_{k(n)})}\,\mathrm{Var}\Big(\overline{\xi}_n(I_{k(n)})\Big) \longrightarrow g^2_{\xi_1,\xi_1}(s,s) + 2\sum_{j=2}^{\infty} g^2_{\xi_1,\xi_j}(s,s).$$

Finally, taking account of (5) the theorem is proved. ■

We now state two versions of the almost complete convergence, one supposing the existence of the Laplace transforms and another supposing a variation of the Cramer moment conditions. These, of course, imply the existence of the Laplace transforms, but enable the proof under the usual convergence rate on the sizes of the sets in each partition, which seems out of reach under the simple existence of Laplace transforms.

Theorem 3.2 *Suppose the sequence (ξ_n, η_n), $n \in \mathbb{N}$, is strictly stationary strong mixing with coefficients $\alpha_n = a\,\rho^n$, for some $a > 0$ and $\rho \in (0,1)$, and such that the Laplace transforms of $\xi(B)$ and $\eta(B)$ are finite on some interval $[-\delta, \delta]$. Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies* **(P1)**-**(P5)** *and that*

$$h_n = \min_{I \in \Pi_{k(n)}} \nu(I) = \frac{\log n}{\sqrt{n}\,\varepsilon_n} \tag{6}$$

where $\varepsilon_n \longrightarrow 0$. Then $f_n(s)$ converges almost completely to $f(s)$ for every $s \in B$.

Proof : Let us prove that

$$\sum_{n=1}^{\infty} \mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})\right| > \frac{\varepsilon}{4}\,\nu(I_{k(n)})\right) < \infty. \tag{7}$$

The corresponding term relative to ξ, according to (3), is afterwards treated analogously. Applying Lemma 2.2 to each term in (7) we have

$$\mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})\right| > \frac{\varepsilon}{4}\,\nu(I_{k(n)})\right) \leq$$

$$\leq 6\,\exp\left(\frac{-n\,t_n\varepsilon\,\nu(I_{k(n)})}{120\,p_n}\right) + 6\,\frac{n}{p_n}\left(\frac{40\,\|\,\eta(B)\,\|_\gamma}{\varepsilon\,\nu(I_{k(n)})} + 1\right)^{\frac{\gamma}{2\gamma+1}} a^{\frac{2\gamma}{2\gamma+1}}\,\rho^{\frac{2\gamma}{2\gamma+1}p_n},$$

where $p_n \in [1, n/2]$, $t_n = \min(\delta/2,\, \varepsilon\, \nu(I_{k(n)})/(12\, c_\eta))$ and $c_\eta = 4 \sum_{k=2}^{\infty} \frac{\delta^{k-2}}{k!}\, \mathrm{E}(\eta^k(B))$, so that, for n large enough, $t_n = \frac{\varepsilon\, \nu(I_{k(n)})}{12\, c_\eta}$. Choosing now $p_n = \sqrt{n}\, h_n$, it follows

$$\mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})\right| > \frac{\varepsilon}{4}\, \nu(I_{k(n)})\right) \leq$$

$$\leq 6\, \exp\left(\frac{-\sqrt{n}\, \varepsilon^2 \nu(I_{k(n)})}{1440\, c_\eta}\right) + 6\, \frac{\sqrt{n}}{\nu(I_{k(n)})} \left(\frac{40 \parallel \eta(B) \parallel_\gamma}{\varepsilon\, \nu(I_{k(n)})} + 1\right)^{\frac{\gamma}{2\gamma+1}} a^{\frac{2\gamma}{2\gamma+1}} \rho^{\frac{2\gamma}{2\gamma+1} \sqrt{n}\, \nu(I_{k(n)})} \leq$$

$$\leq 6\, \exp\left(\frac{-\varepsilon^2 \sqrt{n}\, h_n}{1440\, c_\eta}\right) + 6\, \frac{\sqrt{n}}{h_n} \left(\frac{40 \parallel \eta(B) \parallel_\gamma}{\varepsilon h_n} + 1\right)^{\frac{\gamma}{2\gamma+1}} a^{\frac{2\gamma}{2\gamma+1}} \rho^{\frac{2\gamma}{2\gamma+1} \sqrt{n}\, h_n}.$$

Now (6) ensures this last upper bound defines a convergent series, thus implying (7). ■

Notice that the usual convergence rate require that $\frac{n\, h_n}{\log n} \longrightarrow +\infty$ which is weaker than (6). We may prove a result under this convergence rate strengthening a little some conditions.

Corollary 3.3 *Suppose the sequence (ξ_n, η_n), $n \in \mathbb{N}$, is strictly stationary strong mixing with coefficients $\alpha_n = a\, \rho^n$, for some $a > 0$ and $\rho \in (0,1)$, and there exists a constant $M > 0$ such that, for every $I \subset B$, $k \geq 2$ and $\zeta \in \{\xi, \eta\}$,*

$$\mathrm{E}\left[\zeta^k(I)\right] \leq M^{k-2} k! \mathrm{E}\left[\zeta^2(I)\right]. \tag{8}$$

Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfies **(P1)**-**(P5)**, *the moment measures $\mathrm{E}(\xi \otimes \xi)$ and $\mathrm{E}(\eta \otimes \eta)$ both satisfy* ***(M)*** *and that*

$$h_n = \min_{I \in \Pi_{k(n)}} \nu(I) = \frac{\log^2 n}{n\, \varepsilon_n} \tag{9}$$

where $\varepsilon_n \longrightarrow 0$. Then $f_n(s)$ converges almost completely to $f(s)$ for every $s \in B$.

Proof : From (8) it follows that the Laplace transform of $\xi(B)$ and $\eta(B)$ are finite in some interval $[-\delta, \delta]$ where we may choose δ such that $M\delta < 1$. Thus we apply (4) to the random variables $\eta_i(I_{k(n)})$, $i = 1, \ldots, n$, to find

$$\mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})\right| > \frac{\varepsilon}{4}\, \nu(I_{k(n)})\right) \leq$$

$$\leq 6\, \exp\left(\frac{-n\, t_n\, \varepsilon\, \nu(I_{k(n)})}{120\, p_n}\right) + 6\, \frac{n}{p_n} \left(\frac{40 \parallel \eta(B) \parallel_\gamma}{\varepsilon\, \nu(I_{k(n)})} + 1\right)^{\frac{\gamma}{2\gamma+1}} a^{\frac{2\gamma}{2\gamma+1}} \rho^{\frac{2\gamma}{2\gamma+1} p_n}.$$

where $t_n = \min\left(\frac{\delta}{2}, \frac{\varepsilon\, \nu(I_{k(n)})}{12\, c_{n,\eta}}\right)$ and $c_{n,\eta} = \sum_{j=2}^{\infty} \frac{\delta^{j-2}}{j!}\, \mathrm{E}[\eta^j(I_{k(n)})]$. Using (8), it follows easily that $c_{n,\eta} \leq \frac{\mathrm{E}[\eta^2(I_{k(n)})]}{1 - M\delta} < \infty$, so

$$\frac{\varepsilon\, \nu(I_{k(n)})}{12\, c_{n,\eta}} \geq \frac{\varepsilon(1 - M\delta)\nu(I_{k(n)})}{12\, \mathrm{E}[\eta^2(I_{k(n)})]}.$$

Now $\frac{\mathrm{E}[\eta^2(I_{k(n)})]}{\nu(I_{k(n)})} = \frac{\mathrm{E}(\eta \otimes \eta)(I_{k(n)} \times I_{k(n)})}{\nu(I_{k(n)})}$ has a finite limit according to Lemma 2.4, thus

$$\omega_\eta = \liminf_{n \to +\infty} \frac{\varepsilon\, \nu(I_{k(n)})}{12\, c_{n,\eta}} > 0. \tag{10}$$

Choosing $p_n = n^{1/2} h_n^{1/2}$ and $\delta_\eta = \min\left(\frac{\delta}{2}, \frac{\omega_\eta}{2}\right)$, we obtain, for n large enough,

$$\mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}) - \mu(I_{k(n)})\right| > \frac{\varepsilon}{4}\,\nu(I_{k(n)})\right) \leq$$

$$\leq 6\,\exp\left(\frac{-\varepsilon\,\delta_\eta\, n^{1/2}\, h_n^{1/2}}{120}\right) + 6\,\frac{n^{1/2}}{h_n^{1/2}}\left(\frac{40\parallel \eta(B)\parallel_\gamma}{\varepsilon\, h_n} + 1\right)^{\frac{\gamma}{2\gamma+1}} a^{\frac{2\gamma}{2\gamma+1}} \rho^{\frac{2\gamma}{2\gamma+1}\, n^{1/2} h_n^{1/2}},$$

and we may conclude the proof as in the preceding case taking account of (9). ■

4 Finite dimensional distributions

On this section we suppose that the measure ν is absolutely continuous with respect to some fixed nonatomic measure λ on $\mathbf{S}$, with Radon-Nikodym derivative f_ν which is supposed continuous on the compact set B. Obviously, it follows that $\mu \ll \lambda$ and we denote by f_μ a continuous (on B) version of $\frac{d\mu}{d\lambda}$. We also suppose that f_ν and f_μ are bounded away from zero on B. Let $s_1, \ldots, s_m \in B$ be fixed throughout this section and denote $I_{n,1}, \ldots, I_{n,m}$ the sets in $\Pi_{k(n)}$ containing each one of the given points. Further suppose that the sets in each partition are of same measure with respect to λ and denote by h_n the λ measure of each set in $\Pi_{k(n)}$.

We will first prove a general convergence theorem concerning a random vector with coordinates of the form $\zeta(I_{n,j}) - \mathrm{E}(\zeta(I_{n,j}))$, $\zeta \in \{\xi, \eta\}$. Finally we prove a version for $\mathbf{S} = \mathbb{R}^d$ for the vector with coordinates of the form $\zeta(I_{n,j}) - f_{\mathrm{E}(\zeta)}(s_j)$, $\zeta \in \{\xi, \eta\}$.

Theorem 4.1 *Let (ξ_n, η_n), $n \in \mathbb{N}$, be a strictly stationary strong mixing sequence satisfying* **(M1)** *with respect to λ. Suppose there exist a constant $R > 0$ such that, for every $I \subset B$ and $\zeta \in \{\xi, \eta\}$,*

$$\mathrm{E}\left[\zeta^3(I)\right] \leq R\,\mathrm{E}\left[\zeta^2(I)\right]. \tag{11}$$

Suppose further that the sequence of partitions Π_k, $k \in \mathbb{N}$ satisfy **(P1)**-**(P5)** *with respect to λ, there exists $a > 1$ such that*

$$n\, h_n^a \longrightarrow +\infty \tag{12}$$

and the strong mixing coefficients of the sequence (ξ_n, η_n), $n \in \mathbb{N}$, are such that $\alpha_n = O(n^{-\beta})$ for some

$$\beta > \frac{3a}{a-1}\,. \tag{13}$$

Then the random vector

$$\begin{aligned} n^{1/2} h_n^{-1/2} \Big(&\overline{\eta}_n(I_{n,1}) - \mu(I_{n,1}), \ldots, \overline{\eta}_n(I_{n,m}) - \mu(I_{n,m}),\\ &\overline{\xi}_n(I_{n,1}) - \nu(I_{n,1}), \ldots, \overline{\xi}_n(I_{n,m}) - \nu(I_{n,m})\Big) \end{aligned} \tag{14}$$

converges in distribution to a centered gaussian random vector with covariance matrix

$$\Gamma = \begin{bmatrix} \gamma_{\eta,\eta}(s_1) & \cdots & 0 & \gamma_{\xi,\eta}(s_1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_{\eta,\eta}(s_m) & 0 & \cdots & \gamma_{\xi,\eta}(s_m) \\ \gamma_{\xi,\eta}(s_1) & \cdots & 0 & \gamma_{\xi,\xi}(s_1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_{\xi,\eta}(s_m) & 0 & \cdots & \gamma_{\xi,\xi}(s_m) \end{bmatrix} \tag{15}$$

where $\gamma_{\zeta_1,\zeta_2}(s) = g^2_{\zeta_{1,1},\zeta_{2,1}}(s,s) + 2\sum_{k=2}^{\infty} g^2_{\zeta_{1,1},\zeta_{2,k}}(s,s)$.

Proof : Let $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathbb{R}$ be fixed and define, for each $n \in \mathbb{N}$, $k = 1, \ldots, n$, the random variables

$$X_{k,n} = h_n^{-1/2} \sum_{j=1}^{m} \left[a_j(\xi_k(I_{n,j}) - \nu(I_{n,j})) + b_j(\eta_k(I_{n,j}) - \mu(I_{n,j}))\right]$$

and

$$S_n = \sum_{k=1}^{n} X_{k,n},$$

the linear combination of the coordinates of (14) needed to use the Cramer-Wold Theorem. The method of proof consists on defining large and small blocks, of sizes p_n and q_n, respectively, of variables and couple them using Lemma 2.3. On what follows we will drop the subscript n on p_n and q_n to avoid a heavy notation. Let $r \in \mathbb{N}$ be such that $r(p+q) \leq n < r(p+q+1)$ (again we should denote r_n, which we do nor for the same reason as before) and define

$$\begin{aligned}
V_{1,n} &= X_{1,n} + \cdots + X_{p,n}, & V'_{1,n} &= X_{p+1,n} + \cdots + X_{p+q,n} \\
&\vdots & &\vdots \\
V_{r,n} &= X_{(r-1)(p+q)+1,n} + \cdots + X_{rp+(r-1)q,n}, & V'_{r,n} &= X_{rp+(r-1)q+1,n} + \cdots + X_{r(p+q),n}.
\end{aligned}$$

Suppose that $r = O(n^c)$, $p = O(n^{1-c})$, $q = O(n^d)$, with $c, d \in (0,1)$ suitably choosen, as it will be explained in course of proof.

Using Lemma 2.3, there exist independent variables $W_{1,n}, \ldots, W_{r,n}$ with distributions $\mathrm{P}_{W_{j,n}} = \mathrm{P}_{V_{j,n}}$ and such that

$$\mathrm{P}\left(|V_{j,n} - W_{j,n}| > u_n\right) \leq 11 \left(\frac{\| V_{j,n} + c_n \|_2}{u_n}\right)^{2/5} \alpha_q^{4/5} \tag{16}$$

where $u_n \in (0, \| V_{j,n} + c_n \|_2]$. Now

$$\begin{aligned}
\| X_{1,n} \|_2^2 = \frac{1}{h_n} \sum_{j,j'=1}^{m} & \left[a_j a_{j'} \operatorname{Cov}(\xi_1(I_{n,j}), \xi_1(I_{n,j'})) + a_j b_{j'} \operatorname{Cov}(\xi_1(I_{n,j}), \eta_1(I_{n,j'}))\right. \\
& \left. + b_j a_{j'} \operatorname{Cov}(\eta_1(I_{n,j}), \xi_1(I_{n,j'})) + b_j b_{j'} \operatorname{Cov}(\eta_1(I_{n,j}), \eta_1(I_{n,j'}))\right],
\end{aligned}$$

so that, recalling that $\lambda(I_{n,j}) = \lambda(I_{n,j'}) = h_n$ and using Lemma 2.4, converges to

$$A = \sum_{j=1}^{m} \left[a_j^2 \, g_{\xi,\xi}^2(s_j, s_j) + 2a_j \, b_j \, g_{\xi,\eta}^2(s_j, s_j) + b_j^2 \, g_{\eta,\eta}^2(s_j, s_j)\right].$$

Choosing in (16) $c_n = 3p\, A^{1/2}$ it follows that, for n large enough,

$$p\, A^{1/2} \leq \| V_{j,n} + c_n \|_2 \leq 5p\, A^{1/2},$$

so that, for n large enough,

$$\mathrm{P}\left(|V_{j,n} - W_{j,n}| > u_n\right) = O\left(\frac{p^{2/5}\, \alpha_q^{4/5}}{u_n^{2/5}}\right). \tag{17}$$

If we consider now

$$\Delta_n = \frac{\sum_{j=1}^{r} W_{j,n}}{\sqrt{rp}} - \frac{\sum_{j=1}^{r} V_{j,n}}{\sqrt{rp}}$$

it follows from (17)

$$\mathrm{P}\left(|\Delta_n| > \varepsilon\right) = \mathrm{P}\left(\left|\sum_{j=1}^{r} (V_{j,n} - W_{j,n})\right| > \varepsilon\sqrt{rp}\right) \leq$$

$$\leq \sum_{j=1}^{r} \mathrm{P}\left(|V_{j,n} - W_{j,n}| > \varepsilon\, p^{1/2}\, r^{-1/2}\right) = O\left(r^{6/5}\, p^{1/5}\, \alpha_q^{4/5}\right).$$

Taking account of the convergence rate supposed for the strong mixing coefficients, it follows that

$$\mathrm{P}\left(|\Delta_n| > \varepsilon\right) = O\left(n^{(5c+1-4\beta d)/5}\right)$$

which converges to zero provided that

$$\beta > \frac{5c+1}{4d}\,. \tag{18}$$

We prove now the asymptotic normality of $(rp)^{-1/2}\sum_{j=1}^{r} W_{j,n}$ verifying the Lyapounov condition. For this we will verify that

$$z_n = \frac{\sum_{j=1}^{r} \mathrm{E}(|W_{j,n}|^{\rho})}{(r\,\mathrm{Var}(W_{1,n}))^{\rho/2}} \longrightarrow 0\,. \tag{19}$$

Using Lemma 2.5, it is easily verified that

$$\mathrm{Var}(W_{1,n}) = \mathrm{Var}(V_{1,n}) \sim p \sum_{j=1}^{m} \left[a_j^2 \gamma_{\xi,\xi}(s_j) + 2a_j b_j \gamma_{\xi,\eta}(s_j) + b_j^2 \gamma_{\eta,\eta}(s_j)\right].$$

We shall control the numerator in (19) by using Yokoyama's inequality [38]. First we need to control the third moment of the $X_{k,n}$ conveniently normalized.

$$\begin{aligned}
h_n^{1/2}\,\mathrm{E}(|X_{k,n}|^3) \leq{}& \frac{1}{h_n}\sum_{j=1}^{m}\left|a_j^3\,\mathrm{E}\left[(\xi_k(I_{n,j}) - \nu(I_{j,n}))^3\right]\right| + \\
&+\frac{1}{h_n}\sum_{j,j'=1}^{m}\left|a_j^2 b_{j'}\mathrm{E}\left[(\xi_k(I_{n,j}) - \nu(I_{j,n}))^2(\eta_k(I_{n,j}) - \mu(I_{j,n}))\right]\right| + \\
&+\frac{1}{h_n}\sum_{j,j'=1}^{m}\left|a_j b_{j'}^2\mathrm{E}\left[(\xi_k(I_{n,j}) - \nu(I_{j,n}))(\eta_k(I_{n,j}) - \mu(I_{j,n}))^2\right]\right| + \\
&+\frac{1}{h_n}\sum_{j=1}^{m}\left|b_j^3\,\mathrm{E}\left[(\eta_k(I_{n,j}) - \mu(I_{j,n}))^3\right]\right|.
\end{aligned} \tag{20}$$

The first term on the right of (20) is a sum of terms of the form

$$\frac{1}{h_n}\left|\mathrm{E}\left[(\xi_k(I_{n,j}) - \nu(I_{j,n}))^3\right]\right| = \frac{1}{h_n}\,\mathrm{E}(\xi_k^3(I_{n,j})) - \frac{3}{h_n}\,\mathrm{E}(\xi_k^2(I_{n,j}))\nu(I_{n,j}) + \frac{4}{h_n}\,\nu^3(I_{n,j}). \tag{21}$$

Now, using (11) the first term is bounded above by a convergent term, according to **(M)**, so it is bounded. The remaining terms in (21) converge to zero as a consequence of **(M)**. That is, the first term in (20) is bounded. Evidently, so is the last term in the right side of (20). The remaining two terms in (20) are also bounded as, using Hölder's inequality, we may reduce their control to the third order moments of each factor. Collecting this information, we have verified that

$$h_n^{1/2}\,\mathrm{E}(|X_{k,n}|^3) < \infty, \qquad k = 1, \ldots, n.$$

In order to apply Yokoyama's inequality we verify that there exists $\rho \in (2,3)$ such that $\sum_n (n+1)^{\rho/2-1}\alpha_n^{(3-\rho)/3} < \infty$. Given the assumptions made on the mixing coefficients, the convergence of the series follows from

$$\beta > \frac{3\,\rho}{2(3-\rho)}\,. \tag{22}$$

Now, we may apply Yokoyama's inequality to derive $\mathrm{E}\Big((h_n^{\rho/6}\;|W_{j,n}|^\rho\Big) = O(p^{\rho/2})$, so that

$$z_n = O\left(\frac{r\,p^{\rho/2}}{r^{\rho/2}\,p^{\rho/2}\,h_n^{\rho/6}}\right) = O\left(\frac{n^{c-c\rho/2}}{h_n^{\rho/6}}\right). \tag{23}$$

This converges to zero, taking account of (12), provided that $c \geq \frac{\rho}{3\,a(\rho-2)}$. As we want to choose $c \in (0,1)$, we impose that $\rho > \frac{6\,a}{3\,a-1}$. Using this on (22) we need to impose (13). This proves the asymptotic normality of $(rp)^{-1/2}\sum_{j=1}^r W_{j,n}$ from which follows the asymptotic normality of $(rp)^{-1/2}\sum_{j=1}^r V_{j,n}$.

To finish our proof write

$$S_n = \sum_{j=1}^r V_{j,n} + \sum_{j=1}^r V'_{j,n} + R_n.$$

It is easy to check that

$$\mathrm{Var}\left(\frac{\sum_{j=1}^r V'_{j,n}}{\sqrt{rp}}\right) = O(n^{d-1+c})$$

and

$$\mathrm{Var}\left(\frac{R_n}{\sqrt{rp}}\right) = O(n^{c-1})$$

Thus $(rp)^{-1/2}R_n$ converges in probability to zero. The term $(rp)^{-1/2}\sum_{j=1}^r V'_{j,n}$ also converges to zero if $d < 1-c$. Using this on (18), we derive $\beta > (5+a)/(4(a-1))$ which follows from (13), so the theorem is proved. ■

Remark that in many situations h_n is chosen to behave like some power of n, that is $h_n = O(n^{-a})$ with $a < 1$. In this case (13) is replaced by

$$\beta > \frac{3}{1-a}$$

as follows from the proof of the theorem.

Finally we prove a more precise version of the convergence in distribution adapted to case $\mathbf{S} = \mathbb{R}^d$ for some $d \in \mathbb{N}$, with the reference measure λ being the Lebesgue measure on $\mathbb{R}^d$. For this case we put, to be coherent with the traditional notations, $h_n^d = \lambda(I)$, $I \in \Pi_{k(n)}$. It is known that, for independent sampling, the optimal choice of h_n is $h_n = O(n^{-1/(d+2)})$ (see Oliveira [29] for a proof in the present framework) so we will state a result based on this choice of h_n.

Theorem 4.2 *Let (ξ_n, η_n), $n \in \mathbb{N}$, be a strictly stationary strong mixing sequence satisfying* **(M1)** *with respect to λ the sequence of partitions Π_k, $k \in \mathbb{N}$, satisfy* **(P1)-(P5)** *with respect to λ and (11) holds. Suppose there exist versions f_μ and f_ν of the Radon-Nikodym derivatives $\frac{d\mu}{d\lambda}$ and $\frac{d\nu}{d\lambda}$, respectively, that, besides being bounded away from zero, are continuously differentiable on B. If*

$$h_n = c n^{-1/(d+2)} \varepsilon_n, \tag{24}$$

with $c > 0$, $\varepsilon_n \longrightarrow 0$ and $\alpha_n = O(n^{-\beta})$ with

$$\beta > \frac{3d+6}{2},$$

then

$$\begin{aligned} n^{1/2} h_n^{-d/2} \Bigg(& \frac{\overline{\eta}_n(I_{n,1})}{h_n^d} - f_\mu(s_1), \ldots, \frac{\overline{\eta}_n(I_{n,m})}{h_n^d} - f_\mu(s_m), \\ & \frac{\overline{\xi}_n(I_{n,1})}{h_n^d} - f_\nu(s_1), \ldots, \frac{\overline{\xi}_n(I_{n,m})}{h_n^d} - f_\nu(s_m) \Bigg) \end{aligned} \tag{25}$$

converges in distribution to a centered gaussian random vector with covariance matrix (15).

Proof : The convergence to the required gaussian vector of

$$\begin{aligned} n^{1/2} h_n^{-d/2} \big(& \overline{\eta}_n(I_{n,1}) - \mu(I_{n,1}), \ldots, \overline{\eta}_n(I_{n,m}) - \mu(I_{n,m}), \\ & \overline{\xi}_n(I_{n,1}) - \nu(I_{n,1}), \ldots, \overline{\xi}_n(I_{n,m}) - \nu(I_{n,m}) \big) \end{aligned}$$

follows the same arguments as in the proof of Theorem 4.1.

Now, from the differentiability assumptions, it follows that (see [29])

$$n^{1/2} h_n^{d/2} \Big(\frac{\mu(I_{n,j})}{h_n^d} - f_\mu(s_i) \Big) = n^{1/2} h_n^{d/2}\, O(h_n) = O\left(\varepsilon_n^{(d+2)/2} \right)$$

which converges to zero. The terms corresponding to ν are treated analogously, so the theorem is proved. ■

References

[1] Abou-Jaoudé, S., *La convergence L^1 et L^∞ de certains estimateurs d'une densité de probabilité*, Thèse d'Etat, Univ. Pierre et Marie Curie, Paris, 1977

[2] Bensaïd, N., Nonparametric inference for thinned point process, Statist. Probab. Lett. 33 (1997), 253-258

[3] Bensaïd, N., Fabre, J. P., Convergence de l'estimateur à noyau de derivées de Radon-Nikodym générales dans le cas mélangeant, Canadian J. Statist. 26 (1998), 267-282

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

[5] Bosq, D., Nguyen, H. T., *A course in stochastic processes. Stochastic models and statistical inference*, Kluwers Academic Publisher, 1996

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

[7] Bradley, R. C., Approximation theorems for strongly mixing random variables, Michigan Math. J. 30 (1983) 303-306

[8] Cai, Z. W., Roussas, G. G., Uniform strong estimation under α-mixing, with rates, Statist. Probab. Lett. 15 (1992), 47-55

[9] Crétois, E., Estimation of reduced Palm distributions by random methods for Cox processes with unknown probability law, Appl. Math. (Warsaw) 23 (1995), 247-259

[10] Daley, D. J., Vere-Jones, D., *An introduction to the theory of point processes*, Springer Series in Statistics, 1988

[11] Doukhan, P., *Mixing. Properties and examples*, Lecture Notes in Statistics, 85, Springer-Verlag, New York, 1994

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

[13] Fabre, J. P., *Suites Mélangeantes de mesures aléatoires: estimation fonctionnelle et inégalités de grande deviation*, PhD Dissertation, Univ. Montpellier II, 1998

[14] Ferrieux, D., *Estimation de densités de mesures moyennes de processus ponctuels associés*, PhD dissertation, Univ. Montpellier II, 1996

[15] 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

[16] Grenander, U., *Abstract inference*, Wiley, 1981

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

[18] 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

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

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

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

[22] Jacob, P., Oliveira, P. E., Histograms and associated point processes, Preprint, Pré-publicações Dep. Matemática, Univ. Coimbra, 98-10, 1998

[23] Kallenberg, O., *Random measures*, Academic Press, 1983

[24] Karr, A., *Point processes and their statistical inference*, Marcel Dekker, New York, 1986

[25] Karr, A., Estimation of Palm measures of stationary point processes, Probab. Th. Related Fields, 74 (1987), 55-69

[26] Karr, A., Palm distributions of point processes and their applications to statistical inference. Statistical inference from stochastic processes, Contemp. Math., 87 (1988), Amer. Math. Soc., Providence

[27] Mendes Lopes, N., Uniform consistency of an estimator class of mean local distribution of a random composite measure, Pub. IRMA Lille, 31 (1988), I

[28] Niéré, L., *Estimation des distributions de Palm*, Thèse de 3^{eme} cycle, Univ, Lille, 1987

[29] Oliveira, P. E., Mean square error for histograms when estimating Radon-Nikodym derivatives, to appear in Port. Math.

[30] Rhomari, N., *Filtrage non paramètrique pour les processus non Markoviens. Applications*, PhD Dissertation, Univ. Pierre et Marie Curie, Paris, 1994.

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

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

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

[34] Roussas, G. G., Efficient estimation of a smooth distribution function under $alpha$-mixing, Research developments in probability and statistics 205-217, VSP, Utrecht, 1996

[35] Roussas, G. G., Asymptotic Normality of the Kernel Estimate of a Probability Density Function under Association, Univ. Californa, Davis, Tech. Report # 334, 1998

[36] Saleh, S., Mesure composite aléatoire, Pub. Inst. Stat. Univ. Paris 29 (1984), 59-67

[37] Saleh, S., Estimation de la distribution locale moyenne d'une measure composite aléatoire, Comptes Rendus de l'Acad. Sciences Paris 301, Série 1, 545-548

[38] Yokoyama, R., Moment bounds for stationary absolutely regular sequences, Z. Wahrsch. Verw. Gebiete 52 (1980), 45-57