Histograms and associated point processes*

Pierre Jacob
Université de Montpellier II
34095 Montpellier Cedex 05, France

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

Abstract

Non parametric inference for point processes is approached using histograms, which provide a nice tool for the analysis of on-line data. The construction of histograms depend on a sequence of partitions, which we take to be non embedded. This is quite natural in what regards applications, but presents some theoretical problems. On another direction, we drop the usual independence assumption on the sample, replacing it by an association hypothesis. Under this setting, we study the convergence of the histogram, in probability and almost surely, finding conditions on the covariance structure, which is well known to be the determinant factor under association, to ensure the convergence. On the final section we look at the similar question regarding the finite dimensional distributions, proving a convergence in distribution to a gaussian centered vector with a covariance we can describe. The main tool of analysis will be a decomposition of second order moment measures.

AMS Classification: 62G05, 62G20

1 Introduction

Non parametric inference for point processes has developed using methods based on those used in classical functional estimation, where the estimators belong essentially to two big families, histograms on one side and kernel on the other. Although kernel type estimators have became increasingly popular as they produce smoother estimators, the use of histograms still proves efficient in many situations. Besides, some recent variations on the classical histogram enables an improvement of the convergence rates for this type of estimator, see Beirlant, Berlinet, Györfi [2]. Histograms have been used in estimation of several models depending on point processes. Some examples include regression, as in Bensaïd [3], Palm distributions, as in Karr [22, 23, 24] or Niéré [28], mean local distributions of composed random measures, as in Mendes Lopes [25] or Saleh [35, 36], or density estimation, as in Ellis [11]. This list of references by no means pretends to give an account of the existing literature, but only mentions some examples illustrating each problem. For a more complete list of publications on these subjects we refer the reader to one of the following monographs: Bosq, [6], Bosq, Lecoutre [8], Bosq, Nguyen [7] or Karr [22]. All the above mentioned problems produce results which look similar, in many cases reducing to translations to a different setting of what is known in some previous one. This similarity is due to the fact that these problems may be addressed to in an unified way by defining a convenient general setting, reducing the estimation of the functions in each case to the estimation of a Radon-Nikodym derivative between the

*Preprint

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

means of two given random measures. Some examples on how this framework may include some of the referred problems will be given later. This general framework has been used in Bensaïd, Fabre [4], Ellis [11], Ferrieux [13, 14], Jacob, Mendes Lopes [16], Jacob, Oliveira [18, 19, 20] and Roussas [32, 33, 34], where the articles [18] and [20] are concerned with histograms, while the others study kernel type estimators, although [11] and [32, 33, 34] used a somewhat reduced framework by imposing some special properties on the random measures, namely, supposing one of them, the one with respect to which we need the absolute continuity to be verified, typically a Lebesgue measure, to be almost surely fixed, and [16] deals with absolutely continuous random measures, thus passing the problem to an analysis of the random densities involved. Half of the articles referenced study estimation based on an independent sampling of the point processes. Some results for dependent sampling have been obtained by Bensaïd, Fabre [4] where the kernel estimator is studied based on samples which are strong mixing. On another direction on suppressing the independence assumption, Roussas [32] and, more recently Ferrieux [13, 14], considered associated samples, both studying kernel estimators. Roussas [33, 34] also studied kernel estimates for associated random fields.

Here we will be concerned with histograms for associated compound point processes. These models provide the more interesting examples to illustrate our conditions. The use of histograms relies on the choice of a sequence of partitions of the base space, which in some cases, is constructed by splitting some of the sets of a partition to construct the next one. This procedure produces nice embedded partitions which are very convenient as they allow the use of martingale results as a tool for proving the required convergences. This was used by the authors in [18]. This procedure is quite unnatural from an applications point of view, where it is more common to require that the sets in each partition are of same size, with respect to some reference measure. This requirement, together with the embedding procedure, produces sets which decrease quite fast, and this may mean that the results thus obtained are of limited interest, as the number of new observations needed to change to the next partition would be very large. Some work-arounds to this feature have been made by, for example, Abou-Jaoudé [1], Grenander [15] or Karr [22]. The conditions used typically linked the number of sets in each partition to the moments of the unknown distribution, as it is done in Karr [22]. The authors gave another solution to this problem, using the same general framework as here, but for independent samples in [20], where the conditions imposed depend only on the distribution or only on the sizes of the sets. As this seems a more natural procedure to apply, the results in [20] will be the base for the extension to the associated sampling setting.

2 Preliminaries

In order to define more precisely our framework 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 non negative 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 topology of vague convergence (we refer the reader to Daley, Vere-Jones [10], Kallenberg [21] or Karr [22] for basic properties on random measures). In what follows ξ and η are random measures that are supposed integrable, that is, such that the set functions $\mu(B) = \mathrm{E}\eta(B)$ and $\nu(B) = \mathrm{E}\xi(B)$ define elements of $\mathcal{M}$, and that these mean measures verify the absolute continuity relation $\mu \ll \nu$. As it will be evident, we will be interested in estimating a version of the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$. We will denote by $\mathbb{I}_A$ the indicator function of the set A.

We now exhibit how some of the mentioned estimation problems may be included in the present framework. We will be interested on the interpretations, in each setting, of the Radon Nikodym derivative $\frac{d\mu}{d\nu}$.

- (Ellis [11]) 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 absolutly 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 non negative 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 $\alpha_n = \mathbb{1}_B(T_n)$, 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 $\mathbf{S}\times\mathbf{S}$ 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 $\frac{d\mu}{d\nu}(x) = \mu_x(B) = \mathrm{P}(Y \in B|X = x)$.

So, as illustrated by the examples, we will be concerned with the estimation of $\frac{d\mu}{d\nu}$ based on a sample $((\xi_1, \eta_1), \ldots, (\xi_n, \eta_n))$ of the random pair (ξ, η). As mentioned before, we suppose the pairs (ξ_i, η_i), $i = 1, \ldots, n$, to be associated, that is, given $n \in \mathbb{N}$ and any two coordinatewise non decreasing functions f, g defined on $\mathcal{M}^{2n}$, for which the covariance below exists, we have

$$\mathrm{Cov}(f(\xi_1, \eta_1, \ldots, \xi_n, \eta_n),\, g(\xi_1, \eta_1, \ldots, \xi_n, \eta_n)) \geq 0.$$

(Given $\zeta_1, \zeta_2 \in \mathcal{M}$, we say that $\zeta_1 \leq \zeta_2$ if $\zeta_2 - \zeta_1 \in \mathcal{M}$). For basic results on association we refer the reader to Newman [26], and for association of random measures to Burton, Waymire [9] or Evans [12]. An account of the relevant results for our purposes may be found in Ferrieux [13, 14].

We note that the density estimation and the regression cases mentioned above do not make sense for the associated sampling. In fact, it is easily checked that, whenever a point process has a fixed number of independent points, it can not be associated with itself, so it is impossible to construct a sequence of associated point processes with that same distribution. To check the above suppose $\xi = \delta_X$. Then it is easily seen that $\mathrm{Cov}(\xi,\xi) = \mathrm{E}\delta_{(X,X)} - \mathrm{P}_X \otimes \mathrm{P}_X$. More generally, if $\xi = \sum_{i=1}^n \delta_{X_i}$, for some independent random elements X_i with distributions P_{X_i}, not necessarily equal, then

$$\mathrm{Cov}(\xi,\xi) = \sum_{i=1}^n \left(\mathrm{E}\delta_{(X_i,X_i)} - \mathrm{P}_{X_i} \otimes \mathrm{P}_{X_i} \right).$$

As $\mathrm{E}\delta_{(X_i,X_i)}$ is a measure on $\mathbf{S}\times\mathbf{S}$ with support included in the diagonal and $\mathrm{P}_{X_i} \otimes \mathrm{P}_{X_i}$ is not supported by the diagonal (except in degenerate cases), we really find a signed measure.

It should also be noted that it is not clear if there is any connexion between the $X_1,\ldots,X_n$ being associated and the $\delta_{X_1},\ldots,\delta_{X_n}$ being associated, thus meaning that there is probably no overlap with the work of Ellis [11] or Roussas [32, 33, 34].

In order to define our histograms we need a sequence of partitions. For reasons that will be explained later we will take Π_k, $k \in \mathbb{N}$, to be a sequence of partitions of a fixed compact set $B \subset \mathbf{S}$, instead of partitions of the whole space. We consider the following assumptions

(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 the assumptions **(P4)** or **(P5)** we may specify the measure to which these conditions are to be verified, meaning that ν is to be replaced by some other measure. When not stating anything about this, it means that we really keep the measure ν. The correct indication of this measure is of importance when coupled with conditions **(M1)** and **(M2)**, to be introduced later, where there exists a measure playing a role of reference. We will need these two reference measures to be identical.

Before proceeding introducing further assumptions, we may introduce an approximator of (a suitable version of) $\frac{d\mu}{d\nu}$. Given $s \in B$ we denote by $I_k(s)$ the unique set of Π_k containing the point s, and define, for each $k \in \mathbb{N}$, the function

$$g_k(s) = \sum_{I\in\Pi_k} \frac{\mu(I)}{\nu(I)} \mathbb{1}_I(s) = \frac{\mu(I_k(s))}{\nu(I_k(s))}.$$

In the case of embedded partitions the convergence of g_k to some version of $\frac{d\mu}{d\nu}$ is just a martingale result, which is no longer available in our setting. As it is well known, if there exists a continuous version of the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$, and if the sequence of partitions Π_k, $k \in \mathbb{N}$, verifies **(P1)-(P4)**, the convergence

$$\sup_{s\in B} |f(s) - g_k(s)| \longrightarrow 0$$

holds. The fact that everything is happening within a compact set is crucial to the proof of this uniform convergence. This is why we only consider partitions of a fixed compact set B.

Based on the sample $((\xi_1,\eta_1),\ldots,(\xi_n,\eta_n))$, define $\overline{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$ and $\overline{\eta}_n = \frac{1}{n}\sum_{i=1}^n \eta_i$. The histogram estimator is then

$$f_n(s) = \sum_{I\in\Pi_k} \frac{\overline{\eta}_n(I)}{\overline{\xi}_n(I)} \mathbb{1}_I(s) = \frac{\overline{\eta}_n(I_k(s))}{\overline{\xi}_n(I_k(s))}$$

(we define $f_n(s)$ as zero whenever the denominator vanishes, as usual), where the dependence of k on n is to be precised later. The convergence of f_n to some version of $\frac{d\mu}{d\nu}$ follows from the convergence of $f_n - g_k$ to zero. The treatment of this later convergence was made, in the independent case, via a martingale result concerning product measures of type $\mathrm{E}\zeta_1 \otimes \zeta_2$, where $\zeta_1, \zeta_2 \in \{\xi, \eta\}$ (see Lemma 3.1 in [18]). Again, this was a consequence of the embedding of the partitions, no longer available in our present framework. To get around the difficulty we consider an assumption concerning a decomposition of measures on the product space $\mathbf{S}\times\mathbf{S}$, as considered in [20]. We will say that a measure m on $\mathbf{S}\times\mathbf{S}$ verifies condition **(M)** with respect to the measure ν on $\mathbf{S}$ if $m = m_1 + m_2$ where m_2 is a measure on Δ, the diagonal of $\mathbf{S}\times\mathbf{S}$ and m_1 is a measure on $\mathbf{S}\times\mathbf{S} \setminus \Delta$, 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 continuous version γ_2 of the Radon-Nikodym derivative $\frac{dm_2}{d\nu^*}$.

Then the following result, which will play the role of the above mentioned martingale Lemma in the independent case, holds.

Theorem 2.1 ([20]) *Suppose m is a measure on $\mathbf{S}\times\mathbf{S}$ 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.

Proof : Using the decomposition included in **(M)** we have two terms to look at, corresponding to m_1 and m_2. As for the first

$$\sum_{I\in\Pi_k} \frac{m_1(I\times I)}{\nu(I)} \mathbb{I}_I(s) = \sum_{I\in\Pi_k} \frac{1}{\nu(I)} \int_{I\times I} \gamma_1 \, d\nu\otimes\nu \, \mathbb{I}_I(s) \leq$$

$$\leq \sup_{s,t\in B} |\gamma_1(s,t)| \sum_{I\in\Pi_k} \nu(I) \mathbb{I}_I(s) \leq \sup_{s,t\in B} |\gamma_1(s,t)| \max_{I\in\Pi_k} \nu(I) \longrightarrow 0.$$

As for the second term

$$\sum_{I\in\Pi_k} \frac{m_2(I\times I)}{\nu(I)} \mathbb{I}_I(s) = \sum_{I\in\Pi_k} \frac{m_2(I^*)}{\nu^*(I^*)} \mathbb{I}_I(s) = \sum_{I\in\Pi_k} \left(\frac{1}{\nu^*(I^*)} \int_{I^*} \gamma_2 \, d\nu^* \right) \mathbb{I}_I(s)$$

and the uniform convergence of this expression to $\gamma_2(s,s)$ is just another version of the result giving the already mentioned convergence of the sequence g_k, $k \in \mathbb{N}$. ■

Note that **(M)** must be defined with respect to some measure. If we do not mention any such measure, it will be understood as being ν. As stated after introducing conditions **(P1)-(P5)** what will be important is that the reference measure is the same in both cases. Then, the convergence stated in Theorem 2.1 still holds with the obvious modification on the definition of γ_2, becoming the Radon-Nikodym derivative of m_2 with respect to the lifting of the used reference measure.

We finish this section quoting an useful result, enabling the separation of the variables in the expression f_n.

Lemma 2.2 ([17]) *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,

$$\begin{aligned} \{|f_n(s) - g_k(s)| > \varepsilon\} &= \left\{\left|\frac{\overline{\eta}_n(I_k(s))}{\overline{\xi}_n(I_k(s))} - \frac{\mu(I_k(s))}{\nu(I_k(s)}\right| > \varepsilon\right\} \\ &\subset \left\{|\overline{\eta}_n(I_k(s)) - \mu(I_k(s))| > \frac{\varepsilon}{4}\nu(I_k(s))\right\} \cup \left\{\left|\overline{\xi}_n(I_k(s)) - \nu(I_k(s))\right| > \frac{\varepsilon}{4}\frac{\nu^2(I_k(s))}{\mu(I_k(s))}\right\}. \end{aligned} \tag{1}$$

3 Convergence of the estimator

Having introduced all the definitions and preliminary results needed, we may now look at the convergence of the estimator f_n. We begin by the convergence in probability, for which we state two versions, the second one extendable to an almost complete result which we will not state here for reasons that will be explained later. In order to be more explicit about the dependence between the different indexes used, we will denote the set involved by $I_{k(n)}$, to stress on the dependence of k, identifying which partition is to be considered, on n, the size of the sample.

Theorem 3.1 *Let $B \in \mathcal{B}$ be compact and f a version of $\frac{d\mu}{d\nu}$ continuous on B. Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, verifies* **(P1)-(P5)***, that there exist measures $m^{\xi,\xi}$ and $m^{\eta,\eta}$ such that, for every $n \in \mathbb{N}$,*

$$\frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(\xi_i, \xi_j) \leq m^{\xi,\xi} \qquad \text{and} \qquad \frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(\eta_i, \eta_j) \leq m^{\eta,\eta}$$

with $m^{\xi,\xi}$ and $m^{\eta,\eta}$ both verifying **(M)** *with respect to ν and*

$$n \min_{I \in \Pi_{k(n)}} \nu(I) \longrightarrow +\infty. \tag{2}$$

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

Proof : After separation of variables using (1), we apply Chebyshev's inequality. The term corresponding to η leads to

$$\begin{aligned} &\mathrm{P}\left(\left|\overline{\eta}_n(I_{k(n)}(s)) - \mu(I_{k(n)}(S))\right| > \frac{\varepsilon\nu(I_{k(n)}(s))}{4}\right) \leq \\ &\leq \frac{16}{\varepsilon^2 n\nu^2(I_{k(n)}(s))}\frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(\eta_i(I_{k(n)}(s)), \eta_j(I_{k(n)}(s))) \leq \\ &\leq \frac{16}{\varepsilon^2 n\nu(I_{k(n)}(s))}\,\frac{m_1^{\eta,\eta}(I_{k(n)}(s) \times I_{k(n)}(s)) + m_2^{\eta,\eta}(I_{k(n)}(s) \times I_{k(n)}(s))}{\nu(I_{k(n)}(s))}, \end{aligned}$$

and this converges to zero according to (2) and Theorem 2.1. The other term after separation of variables is treated analogously. ■

Note that in the preceding result, association implies that the covariance measures introduced are really measures and not just signed measures. It is possible to relax a little the requirements imposed on the bounds for the covariances. If we just impose that those covariances are uniformly bounded by some constant, without supposing there is any kind of additivity on the constants for each set, we may still find the convergence in probability at the cost of a slower decrease rate of measures of the sets.

Corollary 3.2 *Let $B \in \mathcal{B}$ and f a version of $\frac{d\mu}{d\nu}$ continuous on B. Suppose there exist constants $c_1, c_2 > 0$ such that*

$$\begin{aligned} \frac{1}{n} \sum_{i,j=1}^{n} \operatorname{Cov}\left(\eta_i(B), \eta_j(B)\right) &\leq c_1 \\ \frac{1}{n} \sum_{i,j=1}^{n} \operatorname{Cov}\left(\xi_i(B), \xi_j(B)\right) &\leq c_2. \end{aligned} \tag{3}$$

If

$$n^{1/2} \min_{I \in \Pi_{k(n)}} \nu(I) \longrightarrow +\infty, \tag{4}$$

then $f_n(s)$ converges in probability to $f(s)$ ν-almost everywhere in B.

The proof follows the same steps as the proof of the theorem, using association to replace each set by B and Markov's inequality leading to r^{th} order moments, that are, afterwards, controlled using inequalities proved by Birkel [5].

Note that conditions (3), for the case $\eta = \delta_X$, rewrites as

$$\frac{1}{n} \sum_{i,j=1}^{n} \left[\mathrm{P}\left(X_i \in I_{k(n)}(s), X_j \in I_{k(n)}(s)\right) - \mathrm{P}\left(X_i \in I_{k(n)}(s)\right) \mathrm{P}\left(X_j \in I_{k(n)}(s)\right)\right] \leq c_1.$$

This kind of sum appears in other situations when studying association. In fact, a general condition for thightness of empirical processes in $L^2[0,1]$ is the uniform convergence of these expressions, as proved in Oliveira, Suquet [29, 30]. The same problem, but in the space $D[0,1]$, also depends on a convenient treatment of these expressions, as in Yu [39] and Shao, Yu [37].

A result about almost complete convergence for associated sampling seems to be out of reach, unless we impose a significantly slower decrease rate on the sets of each partition. This is due to the fact that there is no available version of the Bernstein inequality valid in this setting. The method used for proving Corollary 3.2, based on moment inequalities for sums of associated variables by Birkel [5], may be used to derive an almost complete result but leads to conditions on the partitions requiring a quite slow convergence rate of the sets used at each step and, further, this convergence rate should be well tuned with the decrease rate of the covariance structure of the sequences $\xi_n(B)$, $\eta_n(B)$, $n \in \mathbb{N}$. Thus, we would find conditions with the same drawbacks as those already mentioned linking the size of the sets to the moments of the unknown distribution, that we are trying to avoid here. So, we must look for another type of convergence. Instead of using separation of variables based on Lemma 2.2, the crucial step towards an almost sure theorem is to observe that we do not change the partition each time a new observation is added to our sample, that is, we go on using the same sets until the number of observations increases enough to justify the use of the next partition. This is what is implicitly included in conditions such us (2) or (4). Besides, we will not look at the difference $f_n - g_k$, but rewrite

$$f_n(s) = \frac{\mu(I_{k(n)}(s))}{\nu(I_{k(n)}(s))} \frac{\dfrac{\overline{\eta}_n(I_{k(n)}(s))}{\mu(I_{k(n)}(s))}}{\dfrac{\overline{\xi}_n(I_{k(n)}(s))}{\nu(I_{k(n)}(s))}},$$

so, to prove the almost sure convergence, it is enough to prove that both $\frac{\overline{\eta}_n(I_{k(n)}(s))}{\mu(I_{k(n)}(s))}$ and $\frac{\overline{\xi}_n(I_{k(n)}(s))}{\nu(I_{k(n)}(s))}$ converge almost surely to 1. We will suppress the mention to the point s in all cases where confusion does not arise. For the almost sure convergence we need to identify where we really change from one partition to the next one. Define t_k as the size of the sample where we use, for the first time, sets belonging to partition Π_k. These numbers will report how much more information will be needed to change partition, and their increase must be well balanced in order to obtain the almost sure convergence, as it is proved in the following result.

Theorem 3.3 *Let $B \in \mathcal{B}$ be compact and f a version of $\frac{d\mu}{d\nu}$ continuous and bounded away from zero on B. Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, verifies* **(P1)-(P5)**, *that there exist measures $m^{\xi,\xi}$ and $m^{\eta,\eta}$, such that, for every $n \in \mathbb{N}$,*

$$\frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(\xi_i, \xi_j) \leq m^{\xi,\xi} \qquad \text{and} \qquad \frac{1}{n}\sum_{i,j=1}^{n} \mathrm{Cov}(\eta_i, \eta_j) \leq m^{\eta,\eta}$$

with $m^{\xi,\xi}$ and $m^{\eta,\eta}$ both verifying **(M)**, *and*

$$\frac{t_{k+1}}{t_k} \text{ is bounded} \tag{5}$$

$$\sum_{k=1}^{\infty} \frac{1}{t_k \min_{I\in\Pi_k} \nu(I)} < \infty. \tag{6}$$

Then $f_n(s)$ converges almost surely to $f(s)$, for every $s \in B$.

Proof : We shall check that, under the hypothesis of the theorem $\frac{\overline{\eta}_n(I_{k(n)})}{\mu(I_{k(n)})}$ converges to 1 a.s., the term corresponding to ξ being treated analogously. The proof will follow the classical approach: first we verify the convergence along the subsequence defined by the indexes t_k, $k \in \mathbb{N}$, and then control the difference between these and the remaining terms of the sequence. The first step reduces to an application of Chebyshev's inequality, as follows

$$\mathrm{P}\left(\left|\frac{\overline{\eta}_{t_k}(I_k)}{\nu(I_k)} - 1\right| > \varepsilon\right) = \mathrm{P}\left(\left|\sum_{i=1}^{t_k} (\eta_i(I_k) - \mu(I_k))\right| > \varepsilon t_k \mu(I_k)\right) \leq$$

$$\leq \frac{1}{\varepsilon^2 t_k^2 \mu(I_k)} \sum_{i,j=1}^{t_k} \mathrm{Cov}(\eta_i(I_k), \eta_j(I_k)) \leq \frac{1}{\varepsilon^2} \frac{1}{t_k \nu(I_k)} \frac{\nu^2(I_k)}{\mu^2(I_k)} \frac{m_1^{\eta,\eta}(I_k \times I_k) + m_2^{\eta,\eta}(I_k \times I_k)}{\nu(I_k)}$$

and this defines a convergent series, according to (6) and Theorem 2.1.

Suppose now that $n \in [t_k, t_{k+1})$. According to the definition of t_k, it follows that $I_{k(n)} = I_k$, so

$$\frac{\overline{\eta}_n(I_{k(n)})}{\mu(I_{k(n)})} - \frac{\overline{\eta}_{t_k}(I_k)}{\mu(I_k)} = \sum_{i=1}^{t_k} \left(\frac{1}{n} - \frac{1}{t_k}\right) \frac{\eta_i(I_k) - \mu(I_k)}{\mu(I_k)} + \frac{1}{n} \sum_{i=t_k+1}^{n} \frac{\eta_i(I_k) - \mu(I_k)}{\mu(I_k)}. \tag{7}$$

The first term equals $\left(\frac{t_k}{n} - 1\right)\left(\frac{\overline{\eta}_{t_k}(I_k)}{\mu(I_k)} - 1\right)$. As $t_k \leq n$, the first factor is bounded, and the other factor in this last expression converges almost surely to 0, as proved in the first step. As for the second term in (7), we have, using the generalization of Kolmogorov's inequality for associated variables proved by Newman, Wright [27],

$$\begin{aligned}
\mathrm{P}\left(\max_{t_k \leq n < t_{k+1}} \frac{1}{n}\left|\sum_{i=t_k+1}^{n} \frac{\eta_i(I_k) - \mu(I_k)}{\mu(I_k)}\right| > \varepsilon\right) &\leq \\
\leq \mathrm{P}\left(\max_{t_k \leq n < t_{k+1}} \left|\sum_{i=t_k+1}^{n} [\eta_i(I_k) - \mu(I_k)]\right| > \varepsilon t_k \mu(I_k)\right) &\leq \\
\leq \frac{2}{\varepsilon^2 t_k^2 \mu^2(_k)} \sum_{i,j=t_k}^{t_{k+1}} \mathrm{Cov}(\eta_i(I_k), \eta_j(I_k)) &\leq \\
\leq \frac{2}{\varepsilon^2} \frac{t_{k+1}}{t_k} \frac{1}{t_k \nu(I_k)} \frac{\nu^2(I_k)}{\mu^2(I_k)} \frac{m_1^{\eta,\eta}(I_k \times I_k) + m_2^{\eta,\eta}(I_k \times I_k)}{\nu(I_k)}&,
\end{aligned}$$

which defines a convergent series according to (5), (6) and Theorem 2.1, so the second term in (7) also converges almost surely to zero, and this concludes the proof. ■

4 Finite dimensional distributions

We now look at the finite dimensional asymptotics of $f_n - g_k$, conveniently normalized. As in Jacob, Oliveira [20], in this section we will suppose that ν is absolutely continuous with respect to some fixed non atomic measure λ on $\mathbf{S}$, with Radon-Nikodym derivative f_ν continuous on the compact set B, and that the sets in each partition have equal λ measure. Denote by h_n the λ measure of each set in $\Pi_{k(n)}$. Obviously, μ will also be absolutely continuous with respect to λ and we will denote by f_μ a version of the Radon-Nikodym derivative $\frac{d\mu}{d\lambda}$ that we will suppose also continuous on B. Further, we will suppose that both f_ν and f_μ are bounded away from zero on B. Let us fix $s_1, \ldots, s_r \in B$ and denote by $I_{n,1}, \ldots, I_{n,r}$ the sets in partition $\Pi_{k(n)}$ containing each one of the given points. To prove the convergence in distribution of the finite dimensional distributions we will need some weak form of weak stationarity on the sample, expressed on the conditions we will impose on the decomposition of the covariance measures (8). The proof is based on the method used in the proof of Theorem 9 in Oliveira, Suquet [31], consisting in approximating the sums involved by the sums of conveniently defined blocks and showing that we may reason as if these blocks were independent. For this later part, the main tool is the inequality proved in Theorem 16 in Newman [26], relating the characteristic functions of associated random vectors with what should appear if the coordinates were independent, setting up an upper bound using the covariances of the variables. Before proceeding to the result on the finite dimensional distributions of the estimator, we state a lemma giving a conveniently adapted version of this inequality.

Lemma 4.1 *Let Y_n, $n \in \mathbb{N}$, be associated random variables, $r \in \mathbb{N}$ and $\alpha_0, \ldots, \alpha_r \in \mathbb{R}$. For each $n \in \mathbb{N}$, define*

$$X_n = \sum_{k=0}^{r} \alpha_k Y_{k+n} \qquad \text{and} \qquad \overline{X}_n = \sum_{k=0}^{r} \alpha_k Y_{k+n}.$$

Then, for every $u_1, \ldots, u_r \in \mathbb{R}$,

$$\left| \mathrm{E} e^{i \sum_{j=1}^{m} u_j X_j} - \prod_{j=1}^{m} \mathrm{E} e^{i u_j X_j} \right| \leq 2 \sum_{k \neq l} \left| u_k u_l \operatorname{Cov}(\overline{X}_k, \overline{X}_l) \right|.$$

Proof : For each $n \in \mathbb{N}$ define the functions $f_n(y_1, y_2, \ldots) = \sum_{k=0}^{r} \alpha_k y_{k+n}$ and $\overline{f}_n(y_1, y_2, \ldots) = \sum_{k=0}^{r} |\alpha_k| \, y_{k+n}$. Then $f_n(y_1, y_2, \ldots) + \overline{f}_n(y_1, y_2, \ldots) = \sum_{k=0}^{r} (\alpha_k + |\alpha_k|) y_{k+n}$ and $\overline{f}_n(y_1, y_2, \ldots) - f_n(y_1, y_2, \ldots) = \sum_{k=0}^{r} (|\alpha_k| - \alpha_k) y_{k+n}$, both are coordinatewise increasing, as the coefficients of these linear combinations are non negative. Thus we may apply Theorem 16 from Newman [26], which gives the conclusion of this lemma. ■

Let us introduce the measures, for each $j, k \in \mathbb{N}$,

$$\theta_{j,k} = \frac{1}{k} \sum_{l,l'=(j-1)k+1}^{jk} \operatorname{Cov}(\zeta_{1,l}, \zeta_{2,l'}), \tag{8}$$

where $\zeta_{1,l} = \xi_l$ or $\zeta_{1,l} = \eta_l$ for every $l \in \mathbb{N}$, and analogously for $\zeta_{2,l}$. Decomposition **(M)** defines measures that we will denote $m_{1,j,k}^{\zeta_1,\zeta_2}$ and $m_{2,j,k}^{\zeta_1,\zeta_2}$, where $\zeta_1, \zeta_2 \in \{\xi, \eta\}$, and analogously for the corresponding Radon-Nikodym derivatives.

In the course of proof of the next theorem we need to suppose that the sequence t_k, $k \in \mathbb{N}$, is such that the differences $t_{k+1} - t_k$ are strictly increasing. As this is true, at least for some subsequence, we will assume on the sequel this property verified.

Theorem 4.2 *Suppose the sequence of partitions Π_k, $k \in \mathbb{N}$, verify* **(P1)**-**(P5)** *with respect to λ and*

$$n h_n \longrightarrow +\infty \tag{9}$$

$$\frac{h_{n+1}}{h_n} \longrightarrow 1. \tag{10}$$

Given $k \in \mathbb{N}$ denote m the largest integer less or equal than n/k. Suppose that the measures $\theta_{j,k}$ verify condition **(M)** *with respect to λ and the Radon-Nikodym derivatives defined there verify, for each choice of $\zeta_1, \zeta_2 \in \{\xi, \eta\}$,*

$$\sup_{j,k,n \in \mathbb{N}, jk \leq n} \ \sup_{x \in B} \left| \gamma_{1,j,k}^{\zeta_1,\zeta_2}(x) \right| \leq c_0 < \infty \tag{11}$$

$$\lim_{m \to +\infty} \frac{1}{m} \sum_{j=1}^{m} \gamma_{2,j,k}^{\zeta_1,\zeta_2} = g_{2,k}^{\zeta_1,\zeta_2} \qquad \text{uniformly on } B \tag{12}$$

$$\lim_{k \to +\infty} g_{2,k}^{\zeta_1,\zeta_2} = g_2^{\zeta_1,\zeta_2} \qquad \text{uniformly on } B \tag{13}$$

for some functions $g_{2,k}^{\zeta_1,\zeta_2}$ and $g_2^{\zeta_1,\zeta_2}$ continuous on B. Suppose further that for every sequence $I_n \in \cup_{k=1}^{\infty} \Pi_k$ decreasing to a discrete set and every constant $C > 0$,

$$\int_{\{\zeta_2^2(I_n) > C n h_n\}} \frac{1}{h_n} \zeta_1^2(I_n) \, d\mathrm{P} \longrightarrow 0 \tag{14}$$

for every choice $\zeta_1, \zeta_2 \in \{\xi, \eta\}$. Then, the random vector

$$n^{1/2} h_n^{-1/2} \left(\overline{\eta}_n(I_{n,1}) - \mu(I_{n,1}), \ldots, \overline{\eta}_n(I_{n,r}) - \mu(I_{n,r}), \overline{\xi}_n(I_{n,1}) - \nu(I_{n,1}), \ldots, \overline{\xi}_n(I_{n,r}) - \nu(I_{n,r}) \right) \tag{15}$$

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

$$\Gamma = \begin{bmatrix} g_2^{\eta,\eta}(s_1,s_1) & 0 & \cdots & 0 & g_2^{\xi,\eta}(s_1,s_1) & 0 & \cdots & 0 \\ 0 & g_2^{\eta,\eta}(s_2,s_2) & \cdots & 0 & 0 & g_2^{\xi,\eta}(s_2,s_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_2^{\eta,\eta}(s_r,s_r) & 0 & 0 & \cdots & g_2^{\xi,\eta}(s_r,s_r) \\ g_2^{\xi,\eta}(s_1,s_1) & 0 & \cdots & 0 & g_2^{\xi,\xi}(s_1,s_1) & 0 & \cdots & 0 \\ 0 & g_2^{\xi,\eta}(s_2,s_2) & \cdots & 0 & 0 & g_2^{\xi,\xi}(s_2,s_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_2^{\xi,\eta}(s_r,s_r) & 0 & 0 & \cdots & g_2^{\xi,\xi}(s_r,s_r) \end{bmatrix}$$

Proof : Let $c_1, \ldots, c_r, d_1, \ldots, d_r \in \mathbb{R}$ be fixed and define, for each $n \in \mathbb{N}$, $i = 1, \ldots, n$, $q = 1, \ldots, r$ the random variables

$$T_{n,i}^q = \frac{1}{\sqrt{h_n}} \left[c_q(\xi_i(I_{n,q}) - \nu(I_{n,q})) + d_q(\eta_i(I_{n,q}) - \mu(I_{n,q})) \right]$$

and

$$T_n^q = \frac{1}{\sqrt{n}} \sum_{i=1}^n T_{n,i}^q, \qquad Z_{n,i} = \sum_{q=1}^r T_{n,i}^q, \qquad Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{n,i} = \sum_{q=1}^r T_n^q.$$

For each $j = 1, \ldots, m$ and $q = 1, \ldots, r$ define

$$Y_{n,j}^q = \frac{1}{\sqrt{k}} \sum_{l=(j-1)k+1}^{jk} T_{n,l}^q.$$

Then

$$T_{mk}^q = \frac{1}{\sqrt{mk}} \sum_{i=1}^{mk} T_{mk,i}^q = \frac{1}{\sqrt{m}} \sum_{j=1}^m Y_{mk,j}^q.$$

The variable Z_n is the linear combination of the coordinates of (15), needed to use the Cramer-Wold Theorem, while the variables $Y_{n,j}^q$ correspond to the blocks in which we will decompose our sums. The proof will be accomplished in five steps.

Step 1: We first approximate the characteristic function of Z_n by the characteristic function of Z indexed by a convenient multiple of k, which will be mk or $(m+1)k$ as explained in the following, with $k \in \mathbb{N}$ fixed. As k is fixed and for n large enough, there is at most one change of partition between the sample sizes mk and $(m+1)k$. Suppose for the moment there are no changes of partitions, or, if there is one corresponding to the sample size $t_l \in [mk, (m+1)k)$, that $mk \leq n < t_l$. In this case we approximate Z_n by Z_{mk}.

$$\left| \mathrm{E} e^{iuZ_n} - \mathrm{E} e^{iuZ_{mk}} \right| \leq |u| \operatorname{Var}^{1/2}(Z_n - Z_{mk}) \leq$$

$$\leq |u| \left[\frac{1}{\sqrt{n}} \operatorname{Var}^{1/2} \left(\sum_{i=1}^{mk} (Z_{n,i} - Z_{mk,i}) \right) + \left(\frac{1}{\sqrt{mk}} - \frac{1}{\sqrt{n}} \right) \operatorname{Var}^{1/2} \left(\sum_{i=1}^{mk} Z_{mk,i} \right) + \right. \tag{16}$$

$$+\frac{1}{\sqrt{n}}\mathrm{Var}^{1/2}\left(\sum_{i=mk+1}^{n} Z_{n,i}\right)\Bigg].$$

We now prove that this sum converges to zero. The square of the first term is

$$\begin{aligned}\frac{1}{n}\mathrm{Var}\left(\sum_{i=1}^{mk}(Z_{n,i}-Z_{mk,i})\right) &= \\ &= \frac{1}{n}\sum_{i,j=1}^{mk}\left[\mathrm{Cov}(Z_{n,i},Z_{n,j})-\mathrm{Cov}(Z_{n,i},Z_{mk,j})-\mathrm{Cov}(Z_{mk,i},Z_{n,j})+\mathrm{Cov}(Z_{mk,i},Z_{mk,j})\right].\end{aligned} \tag{17}$$

Developing the first of these terms we find

$$\begin{aligned}\frac{1}{nh_n}\sum_{q,q'=1}^{r}\sum_{i,j=1}^{mk}&\left[c_qc_{q'}\,\mathrm{Cov}(\xi_i(I_{n,q}),\xi_j(I_{n,q'}))+c_qd_{q'}\,\mathrm{Cov}(\xi_i(I_{n,q}),\eta_j(I_{n,q'}))+\right. \\ &\left.+d_qc_{q'}\,\mathrm{Cov}(\eta_i(I_{n,q}),\xi_j(I_{n,q'}))+d_qd_{q'}\,\mathrm{Cov}(\eta_i(I_{n,q}),\eta_j(I_{n,q'}))\right].\end{aligned} \tag{18}$$

The first term of this last expansion equals, using the decomposition **(M)**,

$$\frac{mkh_{mk}}{nh_n}\,\frac{m_{1,1,mk}^{\xi,\xi}(I_{n,q}\times I_{n,q})}{h_{mk}}+\frac{mk}{n}\,\frac{m_{2,1,mk}^{\xi,\xi}(I_{n,q}\times I_{n,q})}{h_{mk}}.$$

Now,

$$\frac{m_{1,1,mk}^{\xi,\xi}(I_{n,q}\times I_{n,q})}{h_{mk}}\leq c_0\frac{\lambda(I_{n,q})\lambda(I_{n,q'})}{h_n}\longrightarrow 0$$

according to the assumptions on the partitions. The second term on the decomposition equals 0 if $q\neq q'$, as in this case the set $I_{n,q}\times I_{n,q'}$ does not intersect the diagonal of the product space. When $q=q'$ we find, according to (13),

$$\frac{m_{2,1,mk}^{\xi,\xi}(I_{n,q}\times I_{n,q})}{h_{mk}}=\frac{m_{2,1,mk}^{\xi,\xi}(I_{n,q}^{*})}{\lambda^{*}(I_{n,q}^{*})}\longrightarrow g_2^{\xi,\xi}(s_q,s_q).$$

The remaining terms in (18) are treated analogously, thus we get, remembering that $\frac{mk}{n}\longrightarrow 1$,

$$\frac{1}{n}\sum_{i,j=1}^{mk}\mathrm{Cov}(Z_{n,i},Z_{n,j})\longrightarrow\sum_{q=1}^{r}\left(c_q^2g_2^{\xi,\xi}(s_q,s_q)+2c_qd_qg_2^{\xi,\eta}(s_q,s_q)+d_q^2g_2^{\eta,\eta}(s_q,s_q)\right)$$

(note that we should consider two terms corresponding to $g_2^{\xi,\eta}$ and to $g_2^{\eta,\xi}$, but as we only need their values on the diagonal, these coincide). The fourth term in (17) is analogous to the one just treated, but the second and third are slightly different, requiring the use of the sequence t_l, $l\in\mathbb{N}$.

In fact,

$$\frac{1}{n}\sum_{i,j=1}^{mk}\operatorname{Cov}(Z_{n,i},Z_{mk,j})=$$

$$=\frac{1}{n\sqrt{h_nh_{mk}}}\sum_{q,q'=1}^{r}\sum_{i,j=1}^{mk}\left[c_qc_{q'}\operatorname{Cov}(\xi_i(I_{n,q}),\xi_j(I_{mk,q'}))+c_qd_{q'}\operatorname{Cov}(\xi_i(I_{n,q}),\eta_j(I_{mk,q'}))+\right.$$

$$\left.+d_qc_{q'}\operatorname{Cov}(\eta_i(I_{n,q}),\xi_j(I_{mk,q'}))+d_qd_{q'}\operatorname{Cov}(\eta_i(I_{n,q}),\eta_j(I_{mk,q'}))\right].$$

As we supposed that there was no change of partition between mk and $(m+1)k$ or that $mk\leq n<t_l<(m+1)k$, in either case, it follows that $I_{mk,q'}=I_{n,q'}$, so the convergence of this expression to $\sum_{q=1}^{r}\left(c_q^2g_2^{\xi,\xi}(s_q,s_q)+2c_qd_qg_2^{\xi,\eta}(s_q,s_q)+d_q^2g_2^{\eta,\eta}(s_q,s_q)\right)$ follows as in the analysis of the first term in (17).

So, adding up these terms, we finally get that

$$\frac{1}{\sqrt{n}}\operatorname{Var}\left(\sum_{i=1}^{mk}(Z_{n,i}-Z_{mk,i})\right)\longrightarrow 0.$$

We proceed now to the second term in (16). Again developing its square we find

$$\left(\frac{1}{\sqrt{mk}}-\frac{1}{\sqrt{n}}\right)^2\frac{1}{h_{mk}}\sum_{i,j=1}^{mk}\operatorname{Cov}(Z_{mk,i},Z_{mk,j})=$$

$$=\left(1-\frac{\sqrt{mk}}{\sqrt{n}}\right)^2\frac{1}{mkh_{mk}}\sum_{q,q'=1}^{r}\sum_{i,j=1}^{mk}\left[c_qc_{q'}\operatorname{Cov}(\xi_i(I_{mk,q}),\xi_j(I_{mk,q'}))+\right.$$

$$+c_qd_{q'}\operatorname{Cov}(\xi_i(I_{mk,q}),\eta_j(I_{mk,q'}))+d_qc_{q'}\operatorname{Cov}(\eta_i(I_{mk,q}),\xi_j(I_{mk,q'}))+$$

$$\left.+d_qd_{q'}\operatorname{Cov}(\eta_i(I_{mk,q}),\eta_j(I_{mk,q'}))\right].$$

All the terms have now the same form as those of (18), so this converges to zero, as $\left(1-\frac{\sqrt{mk}}{\sqrt{n}}\right)^2\longrightarrow 0$.

Finally, we look at the third term in (16). Developing its square we find

$$\frac{1}{n}\sum_{i,j=mk+1}^{n}\operatorname{Cov}(Z_{n,i},Z_{n,j})=$$

$$=\frac{1}{nh_n}\sum_{q,q'=1}^{r}\sum_{i,j=mk+1}^{n}\left[c_qc_{q'}\operatorname{Cov}(\xi_i(I_{n,q}),\xi_j(I_{n,q'}))+c_qd_{q'}\operatorname{Cov}(\xi_i(I_{n,q}),\eta_j(I_{n,q'}))+\right.$$

$$\left.+d_qc_{q'}\operatorname{Cov}(\eta_i(I_{n,q}),\xi_j(I_{n,q'}))+d_qd_{q'}\operatorname{Cov}(\eta_i(I_{n,q}),\eta_j(I_{n,q'}))\right],$$

and these terms all converge to zero because of (10) and the non negativity of the covariances due to association of the variables.

So, we have finally proved that

$$\left|\mathrm{E}e^{iuZ_n} - \mathrm{E}e^{iuZ_{mk}}\right| \longrightarrow 0.$$

It remains to check the case $mk \le t_l \le n < (m+1)k$. In this case we approximate the characteristic function of Z_n by the one of $Z_{(m+1)k}$. The majorizations carried before modify as follows. In the first two terms of (16) just replace m by $m+1$. This does not affect the arguments used in the subsequent analysis. In fact, this change reflects on (17) and (18), thus the remarks made before hold, as now $I_{n,q} = I_{(m+1)k,q}$, thus we still have the convergence to zero of these two first terms in (16). The third term in (16) is replaced by

$$\frac{1}{\sqrt{n}}\mathrm{Var}^{1/2}\left(\sum_{i=n+1}^{(m+1)k} Z_{n,i}\right),$$

and this converges to zero as the corresponding term in the previous case. So, also in this case, we have finally that

$$\left|\mathrm{E}e^{iuZ_n} - \mathrm{E}e^{iuZ_{(m+1)k}}\right| \longrightarrow 0,$$

thus we mat proceed the proof looking only at those values of n that are multiples of k.

Step 2: We now control the difference between $Z_{mk} = \sum_q T^q_{mk}$ and what we would find if the variables $T^1_{mk}, \dots, T^r_{mk}$ were independent. For each $n \in \mathbb{N}$, $i = 1, \dots, n$, $q = 1, \dots, r$ define the random variables

$$\overline{T}^q_{n,i} = \frac{1}{\sqrt{h_n}}\left[|c_q|\left(\xi_i(I_{n,q}) - \nu(I_{n,q})\right) + |d_q|\left(\eta_i(I_{n,q}) - \mu(I_{n,q})\right)\right] \quad \text{and} \quad \overline{T}^q_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \overline{T}^q_{n,i}.$$

From Lemma 4.1,

$$\left|\mathrm{E}e^{iuZ_{mk}} - \prod_{q=1}^{r} \mathrm{E}e^{iuT^q_{mk}}\right| = \left|\mathrm{E}e^{iu\sum_q T^q_{mk}} - \prod_{q=1}^{r} \mathrm{E}e^{iuT^q_{mk}}\right| \le 2u^2 \sum_{q\ne q'} \mathrm{Cov}(\overline{T}^q_{mk}, \overline{T}^{q'}_{mk}).$$

This converges to zero, as it is easily seen developing one of the covariance terms,

$$\mathrm{Cov}(\overline{T}^q_{mk}, \overline{T}^{q'}_{mk}) = \frac{1}{mk}\sum_{i,j=1}^{mk} \mathrm{Cov}(\overline{T}^q_{mk,i}, \overline{T}^{q'}_{mk,j}) =$$

$$= \frac{1}{mkh_{mk}}\sum_{i,j=1}^{mk}\left[|c_qc_{q'}|\,\mathrm{Cov}(\xi_i(I_{mk,q}), \xi_j(I_{mk,q'})) + |c_qd_{q'}|\,\mathrm{Cov}(\xi_i(I_{mk,q}), \eta_j(I_{mk,q'})) + \right.$$

$$\left. + |d_qc_{q'}|\,\mathrm{Cov}(\eta_i(I_{mk,q}), \xi_j(I_{mk,q'})) + |d_qd_{q'}|\,\mathrm{Cov}(\eta_i(I_{mk,q}), \eta_j(I_{mk,q'}))\right].$$

Using decomposition **(M)**, the first term of this sum equals

$$\frac{m^{\xi,\xi}_{1,1,mk}(I_{mk,q}\times I_{mk,q'})}{h_{mk}} + \frac{m^{\xi,\xi}_{2,1,mk}(I_{mk,q}\times I_{mk,q'})}{h_{mk}},$$

and this converges to zero, as before, taking account that $q \ne q'$, so for n large enough $I_{mk,q}\cap I_{mk,q'} = \emptyset$.

Step 3: The sums $T^q_{mk} = \frac{1}{\sqrt{m}} \sum_{j=1}^m Y^q_{mk,j}$ are now approximated by what we would find if the summands were independent. We may reason with q fixed as

$$\left| \prod_{q=1}^r \mathrm{E} e^{iuT^q_{mk}} - \prod_{q=1}^r \prod_{j=1}^m \mathrm{E} e^{i\frac{u}{\sqrt{m}} Y^q_{mk,j}} \right| \leq \sum_{q=1}^r \left| \mathrm{E} e^{iuT^q_{mk}} - \prod_{j=1}^m \mathrm{E} e^{i\frac{u}{\sqrt{m}} Y^q_{mk,j}} \right|.$$

For each $j = 1, \ldots, m$ and $q = 1, \ldots, r$ define

$$Y^q_{n,j} = \frac{1}{\sqrt{k}} \sum_{l=(j-1)k+1}^{jk} \overline{T}^q_{n,l}.$$

Then, another application of Lemma 4.1 yields

$$\left| \mathrm{E} e^{iuT^q_{mk}} - \prod_{j=1}^m \mathrm{E} e^{i\frac{u}{\sqrt{m}} Y^q_{mk,j}} \right| = \left| \mathrm{E} e^{i\frac{u}{\sqrt{m}} \sum_{j=1}^m Y^q_{mk,j}} - \prod_{j=1}^m \mathrm{E} e^{i\frac{u}{\sqrt{m}} Y^q_{mk,j}} \right| \leq$$

$$\leq 2u^2 \sum_{j \neq j'} \frac{1}{m} \operatorname{Cov}(\overline{Y}^q_{mk,j}, \overline{Y}^q_{mk,j'}) = 2u^2 \left[\frac{1}{mk} \sum_{i,j=1}^{mk} \operatorname{Cov}(\overline{T}^q_{n,i}, \overline{T}^q_{n,j}) - \frac{1}{m} \sum_{j=1}^m \operatorname{Cov}(\overline{Y}^q_{mk,j}, \overline{Y}^q_{mk,j}) \right]. \tag{19}$$

The sum

$$\frac{1}{mk} \sum_{i,j=1}^{mk} \operatorname{Cov}(\overline{T}^q_{n,i}, \overline{T}^q_{n,j}) =$$

$$= \frac{1}{mkh_{mk}} \sum_{i,j=1}^{mk} [c_q^2 \operatorname{Cov}(\xi_i(I_{mk,q}), \xi_j(I_{mk,q})) + |c_q d_q| \operatorname{Cov}(\xi_i(I_{mk,q}), \eta_j(I_{mk,q})) +$$

$$+ |d_q c_q| \operatorname{Cov}(\eta_i(I_{mk,q}), \xi_j(I_{mk,q})) + d_q^2 \operatorname{Cov}(\eta_i(I_{mk,q}), \eta_j(I_{mk,q}))]$$

which, as seen before, converges to

$$\overline{a}^q := c_q^2 g_2^{\xi,\xi}(s_q, s_q) + 2\, |c_q d_q|\, g_2^{\xi,\eta}(s_q, s_q) + d_q^2 g_2^{\eta,\eta}(s_q, s_q).$$

The remaining term in (19) may also be treated using decomposition **(M)**, as

$$\operatorname{Cov}(\overline{Y}^q_{mk,j}, \overline{Y}_{mk,j}) =$$

$$= \frac{1}{kh_{mk}} \sum_{l,l'=(j-1)k+1}^{jk} [c_q^2 \operatorname{Cov}(\xi_l(I_{mk,q}), \xi_{l'}(I_{mk,q})) + |c_q d_q| \operatorname{Cov}(\xi_l(I_{mk,q}), \eta_{l'}(I_{mk,q})) +$$

$$+ |d_q c_q| \operatorname{Cov}(\eta_l(I_{mk,q}), \xi_{l'}(I_{mk,q})) + d_q^2 \operatorname{Cov}(\eta_l(I_{mk,q}), \eta_{l'}(I_{mk,q}))].$$

Reproducing the same arguments as before, this converges to

$$c_q^2 \gamma^{\xi,\xi}_{2,j,k}(s_q, s_q) + 2\, |c_q d_q|\, \gamma^{\xi,\eta}_{2,j,k}(s_q, s_q) + d_q^2 \gamma^{\eta,\eta}_{2,j,k}(s_q, s_q),$$

so

$$\frac{1}{m} \sum_{j=1}^m \operatorname{Cov}(\overline{Y}^q_{mk,j}, \overline{Y}^q_{mk,j}) \longrightarrow \overline{a}^q_k := c_q^2 g^{\xi,\xi}_{2,k}(s_q, s_q) + 2\, |c_q d_q|\, g^{\xi,\eta}_{2,k}(s_q, s_q) + d_q^2 g^{\eta,\eta}_{2,k}(s_q, s_q),$$

using (12). Going back to (19), we have thus that

$$\limsup_{m\to+\infty}\left|\mathrm{E}e^{iuT^q_{mk}}-\prod_{j=1}^m \mathrm{E}e^{i\frac{u}{\sqrt{m}}Y^q_{mk,j}}\right|\leq u^2(\overline{a}^q-\overline{a}^q_k).$$

Step 4: To prove now the Central Limit Theorem we may proceed as if the variables $Y^q_{mk,j}$, $j=1,\ldots,m$, were independent. In this case we may verify the Lindeberg condition which reduces to

$$\sum_{j=1}^m\int_{\left\{\left|Y^q_{mk,j}\right|>\varepsilon a^q_k\sqrt{m}\right\}}\frac{1}{m}(Y^q_{mk,j})^2\,d\mathrm{P}\longrightarrow 0,$$

as the variance of the sum of these variables is easily shown to converge to $a^q_k:=c_q^2g_{2,k}^{\xi,\xi}(s_q,s_q)+2c_qd_qg_{2,k}^{\xi,\eta}(s_q,s_q)+d_q^2g_{2,k}^{\eta,\eta}(s_q,s_q)$ reproducing the preceding arguments. Applying Lemma 4 of Utev [38] to this last integral we find the upper bound

$$\begin{aligned}
&\sum_{j=1}^m\int_{\left\{\left|\sum_{l=(j-1)k+1}^{jk}T^q_{mk,l}\right|>\varepsilon a^q_k\sqrt{mk}\right\}}\frac{1}{mk}\left(\sum_{l=(j-1)k+1}^{jk}T^q_{mk,l}\right)^2 d\mathrm{P}\leq\\
&\quad\leq\frac{2}{m}\sum_{j=1}^m\sum_{j=(j-1)k+1}^{jk}\int_{\left\{\left|T^q_{mk,l}\right|>\frac{\varepsilon a^q_k}{2}\sqrt{\frac{m}{k}}\right\}}(T^q_{mk,l})^2\,d\mathrm{P}=\\
&\quad=\frac{2}{m}\sum_{j=1}^{mk}\int_{\left\{\left|T^q_{mk,j}\right|>\frac{\varepsilon a^q_k}{2k}\sqrt{mk}\right\}}(T^q_{mk,j})^2\,d\mathrm{P}\,.
\end{aligned}$$

As k is fixed, this sum has the same form as the one treated in the proof of Theorem 4.1 in Jacob, Oliveira [20] (see also, Theorem 6.1 in [18]), where it was proved to converge to zero from (14). Thus we have the convergence in distribution of the vector $(Y^q_{mk,1},\ldots,Y^q_{mk,m})$ to a centered gaussian vector with covariance matrix of the same form as Γ but with the $g_2^{\zeta_1,\zeta_2}$ replaced by $g_{2,k}^{\zeta_1,\zeta_2}$.

Step 5: To finish now the proof we write, with $a^q=c_q^2g_2^{\xi,\xi}(s_q,s_q)+2c_qd_qg_2^{\xi,\eta}(s_q,s_q)+d_q^2g_2^{\eta,\eta}(s_q,s_q)$,

$$\begin{aligned}
&\left|\mathrm{E}e^{iuZ_{mk}}-e^{-\frac{u^2}{2}\sum_{q=1}^r a^q}\right|\leq\\
&\quad\leq\left|\mathrm{E}e^{iuZ_{mk}}-\prod_{q=1}^r\mathrm{E}e^{iuT^q_{mk}}\right|+\sum_{q=1}^r\left|\mathrm{E}e^{iuT^q_{mk}}-\prod_{j=1}^m\mathrm{E}e^{\frac{iu}{\sqrt{m}}Y^q_{mk,j}}\right|+\\
&\quad+\sum_{q=1}^r\left|\prod_{j=1}^m\mathrm{E}e^{\frac{iu}{\sqrt{m}}Y^q_{mk,j}}-e^{-\frac{u^2}{2}a^q_k}\right|+\sum_{q=1}^r\left|e^{-\frac{u^2}{2}a^q_k}-e^{-\frac{u^2}{2}a^q}\right|
\end{aligned}$$

Consider k fixed for the moment. The first term in this upper bound converges to zero according to *Step 2*. The third term converges to zero according to *Step 4*. So, taking account of *Step 3*, we have, for each $k\in\mathbb{N}$,

$$\limsup_{m\to+\infty}\left|\mathrm{E}e^{iuZ_{mk}}-e^{-\frac{u^2}{2}\sum_{q=1}^r a^q}\right|\leq 2u^2\sum_{q=1}^r(\overline{a}^q-\overline{a}^q_k)+\sum_{q=1}^r\left|e^{-\frac{u^2}{2}a^q_k}-e^{-\frac{u^2}{2}a^q}\right|.$$

Finally, letting $k \longrightarrow +\infty$, this converges to zero, according to (13), thus proving the convergence in distribution of (15). ■

A discussion concerning conditions (14) has been carried by the authors in [18], showing that it is a reasonable condition which is verified for Poisson processes and some other point processes constructed from Poisson processes.

An application of the δ-method yields the convergence of the finite dimensional distributions of the estimator f_n itself.

Theorem 4.3 *Suppose the conditions of theorem 4.2 are verified. Then*

$$n^{1/2}h_n^{1/2}\left(\frac{\overline{\eta}_n(I_{n,1})}{\overline{\xi}_n(I_{n,1})}-\frac{\mu(I_{n,1})}{\nu(I_{n,1})},\ldots,\frac{\overline{\eta}_n(I_{n,r})}{\overline{\xi}_n(I_{n,r})}-\frac{\mu(I_{n,r})}{\nu(I_{n,r})}\right) \tag{20}$$

converges in distribution to a centered gaussian random vector with diagonal covariance matrix Γ^*, with $\gamma^*_{q,q}=\frac{g_2^{\xi,\xi}(s_q,s_q)}{f_\mu^2(s_q)}-\frac{2g_2^{\xi,\eta}(s_q,s_q)f_\mu(s_q)}{f_\nu^3(s_q)}+\frac{f_\mu^2(s_q)g_2^{\eta,\eta}(s_q,s_q)}{f_\nu^4(s_q)}$, $q=1,\ldots,r$.

Proof : Define the random vector $U_n=\left(\overline{\eta}_n(I_{n,1}),\ldots,\overline{\eta}_n(I_{n,r}),\overline{\xi}_n(I_{n,1}),\ldots,\overline{\xi}_n(I_{n,r})\right)$ and the real valued function φ on $\mathbb{R}^{2r}$ by $\varphi(y)=\sum_{q=1}^r b_q\frac{y_q}{y_{r+q}}$, where $b_1,\ldots,b_q$ are real numbers, so that $\sqrt{nh_n}\,(\varphi(U_n)-\varphi(\mathrm{E}U_n))$ is a linear combination of the coordinates of (20). Using the Taylor expansion, we find

$$\sqrt{nh_n}\,(\varphi(U_n)-\varphi(\mathrm{E}U_n)) =$$

$$=\sum_{q=1}^{2r} h_n\frac{\partial\varphi}{\partial y_q}(\mathrm{E}U_n)\sqrt{\frac{n}{h_n}}\,(U_{n,q}-\mathrm{E}U_{n,q})+h_n\sqrt{\frac{n}{h_n}}\parallel U_n-\mathrm{E}U_n\parallel\varepsilon(\parallel U_n-\mathrm{E}U_n\parallel)$$

where ε is continuous and $\lim_{y\to 0}\varepsilon(y)=0$. As $U_n\longrightarrow\mathrm{E}U_n$ in probability, according to Theorem 3.1, and $n^{1/2}h_n^{-1/2}\parallel U_n-\mathrm{E}U_n\parallel$ converges in distribution, the last term converges in probability to zero. Consider now the vector $u=(f_\mu(s_1),\ldots,f_\mu(s_r),f_\nu(s_1),\ldots,f_\nu(s_r))$ and rewrite the first term of the Taylor expansion as

$$\sum_{q=1}^{2r}\frac{\partial\varphi}{\partial y_q}(u)\sqrt{\frac{n}{h_n}}\,(U_{n,q}-\mathrm{E}U_{n,q})+\sum_{q=1}^{2r}\left(h_n\frac{\partial\varphi}{\partial y_q}(\mathrm{E}U_n)-\frac{\partial\varphi}{\partial y_q}(u)\right)\sqrt{\frac{n}{h_n}}\,(U_{n,q}-\mathrm{E}U_{n,q}).$$

Computing the derivatives, it is easily checked that $h_n\frac{\partial\varphi}{\partial y_q}(\mathrm{E}U_n)\longrightarrow\frac{\partial\varphi}{\partial y_q}(u)$, so the limiting distribution of $\sqrt{nh_n}\,(\varphi(U_n)-\varphi(\mathrm{E}U_n))$ is characterized by the limit of $\sum_q^{2r}\frac{\partial\varphi}{\partial y_q}(u)\sqrt{\frac{n}{h_n}}\,(U_{n,q}-\mathrm{E}U_{n,q})$, which has been shown to be gaussian in the previous theorem. Its variance is easily shown to be

$$\sum_{q=1}^{2r} b_q^2\left(\frac{g_2^{\xi,\xi}(s_q,s_q)}{f_\mu^2(s_q)}-\frac{2g_2^{\xi,\eta}(s_q,s_q)f_\mu(s_q)}{f_\nu^3(s_q)}+\frac{f_\mu^2(s_q)g_2^{\eta,\eta}(s_q,s_q)}{f_\nu^4(s_q,s_q)}\right),$$

replacing, in the computation of the variance in the previous theorem, c_q by $\frac{b_q}{f_\mu(s_q)}$ and d_q by $\frac{b_qf_\mu(s_q)}{f_\nu^2(s_q)}$, according to the derivatives of φ. ■

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] Beirlant, J., Berlinet, A., Györfi, L., On Piecewise Linear Density Estimators, Statist. Neerlandica, to appear in 1998

[3] Bensaïd, N., *Contribution à l'estimation et à la prévision non paramétrique d'un processus ponctuel multidimensionnel*, PhD Dissertation, Univ. Paris VI, 1992

[4] 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.

[5] Birkel, T., Moment Bounds for associated sequences, Ann. Probab., 16 (1988), 1184-1193

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

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

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

[9] Burton, R. M., Waymire, E., The central limit problem for infinitely divisible random measures, Dependence in Probability and Statistics, Eds E. Eberlein, M. S. Taqqu, Birkhäuser, 1986, 383-395

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

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

[12] Evans, S., Association and random measures, Probab. Th. Related Fields, 86 (1990), 1-19

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

[14] 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

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

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

[17] 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

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

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

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

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

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

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

[24] 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

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

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

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

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

[29] Oliveira, P. E., Suquet, C., $L^2[0,1]$ weak invariance principle of the empirical process for dependent variables, Actes des XVèmes Rencontres Franco-Belges de Statisticiens (Ondelettes et Statistique), Lecture Notes in Statistics 103, Wavelets and Statistics, Ed. A. Antoniadis, G. Oppenheim, 1995

[30] Oliveira, P. E., Suquet, C., Processus empirique sous dépendence dans L^2, Compte Rendus Acad. Sciences de Paris 321(1995), Série I, 939-944

[31] Oliveira, P. E., Suquet, C., An $L^2[0,1]$ invariance principle for LPQD random variables, Port. Math. 53 (1996), 367-379

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

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

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

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

[36] 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

[37] Shao, Q. M., Yu, H., Weak convergence for weighted empirical processes of dependent sequences, Ann. Probab. 24 (1996), 2052-2078

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

[39] Yu, H., A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences, Probab. Th. Related Fields, 95 (1993), 357-370