Chilean Journal of Statistics
2011, iFirst article

Nonparametric Statistics
Research Paper

On the kernel estimation of a multivariate distribution function under positive dependence

Cecília Azevedo[1,*] and Paulo E. Oliveira[2]

[1]Centro de Matemática – CMat, Universidade do Minho, Braga, Portugal
[2]CMUC, Departamento de Matemática, Universidade de Coimbra, Portugal
(Received: 10 September 2009 · Accepted in final form: 17 December 2010)
(Received 00 Month 200x · Accepted in final form: 00 Month 200x)

Abstract

In this paper, we consider the kernel estimator of the p-dimensional marginal distribution function of a stationary, positively associated sequence of random variables. For this setting, we state results concerning the asymptotic behaviour of this estimator extending some characterizations available in the literature. In addition, we present a simulation study about the empirical process constructed from such a estimator illustrating its asymptotic normality.

Keywords: Asymptotic normality · Empirical process · Nonparametric estimation · Optimal bandwidth · Positive association.

Mathematics Subject Classification: Primary 62G05 · Secondary 62G07, 62G20.

1. Introduction

Estimation of distribution functions has been, in parallel to estimation of density functions, one of the classical problems in statistics. For a stationary sequence of random variables, the kernel estimator of its p-dimensional marginal distribution function can be considered, assuming that the available sample satisfies some kind of positive dependence. Lehmann (1966) and Esary et al. (1967) introduced a notion of positive dependence. After of these works, various other types of dependence have also been proposed. These dependence structures have received some attention in the statistical literature, especially since the early 1990's. In what regards the asymptotic behaviour with respect to convergence in distribution, the dependence structure introduced by Esary et al. (1967), it is completely characterized by the covariance structure between the variables, as described in Newman (1984, Theorem 10). Thus, it is natural and convenient to seek sufficient conditions for the convergence of kernel estimators imposing some adequate decrease rate on the covariances.

For an one-dimensional marginal distribution function, its kernel estimator has been studied by Roussas (1991, 2000). For this problem, Cai and Roussas (1998) considered quadrant positive dependence, a type of dependence that is a weaker form of positive

*Corresponding author: Cecília Azevedo, Centro de Matemática – CMat, Universidade do Minho, Largo do Paço, 4710 – 057 Braga, Portugal. Email: cecilia@math.uminho.pt

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

dependence. Motivated by the need to approximate covariance functions appearing in the study of empirical processes, the two-dimensional case based on associated samples was addressed by Azevedo and Oliveira (2000) using the kernel estimator, whereas Henriques and Oliveira (2003) employed the histogram estimator.

The aims of this work are mainly two. First, to extend to higher dimensions some asymptotic results of the kernel estimator of a multivariate marginal distribution function under the notion positive dependence introduced by Esary et al. (1967) for a sequence of random variables. Second, to conduct a simulation study concerning the empirical process of this estimator. Specifically, we extend to the p-dimensional case the asymptotic normality characterizations obtained previously by Azevedo and Oliveira (2000) and the almost sure consistency of the estimator studied by Azevedo and Oliveira (2005). By means of the simulation study, we illustrate the convergence of the finite dimensional distributions of the empirical process induced by the estimator, giving some information about the finite sample behaviour. The simulation model depends on a parameter that may be interpreted as a measure of how far away the variables can be while remaining dependent. The influence of this parameter is also illustrated. It is clear that, for sequences that are close to independence, i.e., for small values of the above mentioned parameter, the asymptotic normality happens with a quite fast convergence rate.

The paper is organized as follows. Section 2 presents some preliminary aspects for this study. Section 3 discusses about consistency of the kernel estimator of the distribution function. Section 4 characterizes the asymptotic behaviour and the convergence rate of the mean square error (MSE) of this estimator. Section 5 looks at the asymptotic normality of the finite dimensional distributions of the empirical process constructed from the above mentioned estimator. Section 6 carries out the simulation study that allows us to describe the behaviour of the empirical process. Finally, Section 7 sketches some conclusions.

2. Preliminaries

In this section, we present the framework and assumptions for kernel estimation of a distribution function under positive dependence for the p-dimensional setting.

We recall the definition of association introduced in Esary et al. (1967).

Definition 2.1 The random variables $X_1, X_2, \ldots$ are said to be positively associated if, for every $k \geq 1$ and any real-valued coordinatewise increasing functions G, H: $\mathbb{R}^k \longrightarrow \mathbb{R}$,

$$\mathrm{Cov}\,(G(X_1, \ldots, X_k), H(X_1, \ldots, X_k)) \geq 0,$$

whenever this covariance exists. A sequence of random variables $\{X_n, n \geq 1\}$ is said to be associated if, for every $k \in \mathbb{N}$, the random variables $X_1, \ldots, X_k$ are associated.

Given a stationary sequence of random variables $\{X_n, n \geq 1\}$, denote by $\mathbf{F}_p$ its p-dimensional marginal distribution function. Let $\mathbf{U}$ be a p-variate distribution function, and h_n, for $n \geq 1$, a sequence of positive real numbers known as bandwidth such that $h_n \to 0$. Taking into account the stationarity, an estimator for $\mathbf{F}_p$ is defined by

$$\widehat{(\mathbf{F}_p)}_n(\mathbf{x}) = \frac{1}{(n-p)} \sum_{i=1}^{n-p} \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right), \tag{1}$$

where $\mathbf{X}_{i,p} = (X_i, \ldots, X_{i+p-1})$ and $\mathbf{x} = (x_1, \ldots, x_p)$. We refer to $\mathbf{U}$ as the kernel function. This is the natural extension of the kernel estimator for density functions to distribution functions.

For independent samples, Jin and Shao (1999) proved the almost sure consistency of the MSE of $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$, deriving that, for every dimension p, the optimal bandwidth rate is of order $n^{-1/3}$. For associated samples, it follows from Cai and Roussas (1998) that, under assumptions on the covariance structure that imply the consistency of the estimator, the optimal bandwidth rate for the one-dimensional case is of order n^{-1}. This characterization of the optimal bandwidth rate depends on the decay rate of the covariances as shown in Cai and Roussas (1998), where strengthening of the assumptions on the covariances the optimal bandwidth rate of order $n^{-1/3}$, as for independent sequences, is recovered. Azevedo and Oliveira (2000) considered the two-dimensional kernel estimation of the distribution function of (X_1, X_{k+1}), characterizing the optimal bandwidth rate, with results similar to those given in Cai and Roussas (1998). Next, we list the assumptions that are used in the sequel. This set of conditions is basically the same as in Cai and Roussas (1998) and Jin and Shao (1999).

Assumptions

- (A1) Let $\{X_n, n \geq 1\}$ be a strictly stationary sequence of associated random variables with bounded density function f;
- (A2) The distribution function $\mathbf{F}_p$ of the random vector $\mathbf{X} = (X_1, \ldots, X_p)$ has bounded, continuous partial derivatives of first and second orders;
- (A3) For each $j \geq 1$, the distribution function $\mathbf{F}_{p,j}$ of the $2p$-dimensional random vector $(\mathbf{X}_{1,p}, \mathbf{X}_{j,p})$ has bounded, continuous partial derivatives of first and second orders;
- (A4) The kernel function $\mathbf{U}$ is p times differentiable and $\mathbf{u} = \frac{\partial^p \mathbf{U}}{\partial x_1 \ldots \partial x_p}$ satisfies

$$\text{(i)} \int_{\mathbb{R}^p} \mathbf{u}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1; \qquad \text{(ii)} \int_{\mathbb{R}^p} \mathbf{x}\,\mathbf{u}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0; \qquad \text{(iii)} \int_{\mathbb{R}^p} \mathbf{x}\,\mathbf{x}^\top\,\mathbf{u}(\mathbf{x})\,\mathrm{d}\mathbf{x} < \infty;$$

- (A5) $n\,h_n^2 \to 0$;
- (A5*) $n\,h_n^4 \to 0$;
- (A6) $\displaystyle\sum_{n=1}^{\infty} n\,(\mathrm{Cov}\,(X_1, X_n))^{1/3} < \infty$;
- (A6*) There exists $\tau \in (0,1)$ such that $\displaystyle\sum_{j=1}^{\infty} \big(\mathrm{Cov}\,(X_1, X_{j+1})\big)^{(1-\tau)/3} < \infty$;
- (A7) The function $\mathbf{V} = \frac{\partial^p \mathbf{U}^2}{\partial x_1 \cdots \partial x_p}$ satisfies

$$\int_{\mathbb{R}^p} \mathbf{x}\,\mathbf{x}^\top\,\mathbf{V}(\mathbf{x})\,\mathrm{d}\mathbf{x} < \infty.$$

Remark 2.1 (A1) and (A6) have been used by Cai and Roussas (1998) for the treatment of the univariate case. These assumptions state the regularity of the one-dimensional distribution function and a convenient decrease rate on the covariances. The later assumption enables the control of pairs of random variables.

Remark 2.2 The strengthened assumptions (A5*) and (A6*) have also been used in the one-dimensional case by Cai and Roussas (1998). The authors obtained optimal bandwidth characterization with the same rate as for independent sequences of random variables.

Remark 2.3 It is obvious that (A6) implies (A6*). This later assumption has been shown to imply the $L^2[0,1]$ weak convergence of the one-dimensional empirical process based on associated random variables; see Oliveira and Suquet (1996).

Finally, let us define some auxiliary real valued functions $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ and $\mathbf{V}_4$ on $\mathbb{R}^p$ as:

- $\mathbf{V}_1(\mathbf{x}) = \sum_{i=1}^{p} \frac{\partial^2 \mathbf{F}_p}{\partial x_i^2}(\mathbf{x}) \int_{\mathbb{R}^p} t_i^2 \mathbf{u}(\mathbf{t})\,\mathrm{d}\mathbf{t} + 2\sum_{j=1}^{p-1} \sum_{i=j+1}^{p} \frac{\partial^2 \mathbf{F}_p}{\partial x_j \partial x_i}(\mathbf{x}) \int_{\mathbb{R}^p} t_i\, t_j\, \mathbf{u}(\mathbf{t})\,\mathrm{d}\mathbf{t}$;
- $\mathbf{V}_2(\mathbf{x}) = \sum_{i=1}^{p} \frac{\partial \mathbf{F}_p}{\partial x_i}(\mathbf{x}) \int_{\mathbb{R}^p} t_i\, \mathbf{V}(\mathbf{t})\,\mathrm{d}\mathbf{t}$;
- $\mathbf{V}_3(\mathbf{x}) = \sum_{i=1}^{p} \frac{\partial^2 \mathbf{F}_p}{\partial x_i^2}(\mathbf{x}) \int_{\mathbb{R}^p} t_i^2\, \mathbf{V}(\mathbf{t})\,\mathrm{d}\mathbf{t} + 2\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \frac{\partial^2 \mathbf{F}_p}{\partial x_j \partial x_i}(\mathbf{x}) \int_{\mathbb{R}^p} t_i\, t_j\, \mathbf{V}(\mathbf{t})\,\mathrm{d}\mathbf{t}$;
- $\mathbf{V}_4(\mathbf{x}) = \sum_{i=1}^{2p} \frac{\partial^2 \mathbf{F}_{p,j}}{\partial x_i^2}(\mathbf{x},\mathbf{x}) \int_{\mathbb{R}^p} t_i^2\, \mathbf{u}(\mathbf{t})\,\mathrm{d}\mathbf{t} + 2\sum_{i=1}^{2p-1} \sum_{j=i+1}^{2p} \frac{\partial^2 \mathbf{F}_{p,j}}{\partial x_j \partial x_i}(\mathbf{x},\mathbf{x}) \int_{\mathbb{R}^p} t_i\, t_j\, \mathbf{u}(\mathbf{t})\,\mathrm{d}\mathbf{t}$,

where $\mathbf{t} = (t_1, \ldots, t_p)$.

3. Consistency

In this section, we look at the almost sure consistency of $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$. This results follows from an application of a strong law of large numbers (SLLN) to the sequence of random variables $\mathbf{U}(\{\mathbf{x} - \mathbf{X}_{i,p}\}/h_n)$, for $i = 1, \ldots, n-p$, that appears in the definition of $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$. Notice that, as $\mathbf{U}$ is a distribution function, thus nondecreasing, these variables are associated. To prove this SLLN, we characterize the asymptotic behaviour of the covariances between the corresponding variables. The almost sure consistency then follows from the asymptotic unbiasedness of $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$, which is proved in the following theorem.

Theorem 3.1 Assume the sequence of random variables $\{X_n, n \geq 1\}$ satisfy (A1). Then, for each $\mathbf{x} \in \mathbb{R}^p$,

(i) $\mathbb{E}\big[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\big] \to \mathbf{F}_p(\mathbf{x})$;

(ii) If (A2) and (A4) are also satisfied, $\mathbb{E}\big[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\big] = \mathbf{F}_p(\mathbf{x}) + \mathbf{V}_1(\mathbf{x})\, h_n^2/2 + o(h_n^2)$.

Proof Part (i) follows from an application of the dominated convergence theorem. For part (ii), rewrite $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$ as

$$\widehat{(\mathbf{F}_p)}_n(\mathbf{x}) = \int_{\mathbb{R}^p} \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{s}}{h_n}\right) \mathrm{d}\widehat{\phi}_n(\mathbf{s}), \tag{2}$$

where $\widehat{\phi}_n(\mathbf{x}) = \frac{1}{(n-p)} \sum_{i=1}^{n-p} \mathbb{I}_{(-\infty,x_1]\times\cdots\times(-\infty,x_p]}(\mathbf{X}_{i,p})$. It is easily verified that $\mathbb{E}(\widehat{\phi}_n(\mathbf{x})) = \mathbf{F}_p(\mathbf{x})$, so the result follows from Equation (2), applying Fubini's Theorem, making a standard change of variable and using a Taylor expansion taking into account (A2) and (A4). ∎

As mentioned, the SLLN that gives the almost sure consistency of $\widehat{(\mathbf{F}_p)}_n$ follows from a convenient control on the covariances between the terms summed in the definition of the estimator given in Equation (1).

LEMMA 3.2 Assume the sequence of random variables $\{X_n, n \geq 1\}$ satisfy (A1)-(A4). Then, for each $\mathbf{x} \in \mathbb{R}^p$,

(i) $\mathbf{I}_{n,j}(\mathbf{x}) = \mathbf{I}_j(\mathbf{x}) + O(h_n^2) = \mathbf{F}_{p,j}(\mathbf{x},\mathbf{x}) - \mathbf{F}_p^2(\mathbf{x}) + O(h_n^2)$, for each $j \in \mathbb{N}$;

(ii) There exists a constant $M > 0$, independent from $\mathbf{x}$, such that, for each $j > p-1$,

$$\mathbf{I}_j(\mathbf{x}) \leq M \sum_{k=1}^{p} (p-k+1)(\mathrm{Cov}\,(X_1, X_{j+k}))^{1/3} + M \sum_{k=1}^{p-1} (p-k)(\mathrm{Cov}\,(X_1, X_{j-k+1}))^{1/3},$$

where

$$\mathbf{I}_{n,j}(\mathbf{x}) = \mathrm{Cov}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right), \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{j,p}}{h_n}\right)\right),$$

$$\mathbf{I}_j(\mathbf{x}) = \mathrm{Cov}\left(\mathbb{I}_{(-\infty,\mathbf{x}]}(\mathbf{X}_{1,p}), \mathbb{I}_{(-\infty,\mathbf{x}]}(\mathbf{X}_{j,p})\right)$$

and $\mathbb{I}_A$ is the indicator function of the set A.

We first quote a result by Lebowitz (1972) needed to prove Lemma 3.2.

LEMMA 3.3 Let $A, B \subset \{1, \ldots, n\}$ and, for each $i \in A \cup B$, let $x_i \in \mathbb{R}$. Define $H_{A,B} = \mathrm{P}(X_i > x_i, i \in A \cup B) - \mathrm{P}(X_j > x_j, j \in A)\,\mathrm{P}(X_k > x_k, k \in B)$. If the random variables $X_1, \ldots, X_n$ are associated then, $0 \leq H_{A,B} \leq \sum_{i \in A,\, j \in B} H_{\{i\},\{j\}}$.

PROOF [LEMMA 3.2] To prove part (i), write

$$\mathbf{I}_{n,j} = \int_{\mathbb{R}^{2p}} \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{s}}{h_n}\right) \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{t}}{h_n}\right) \mathrm{d}\mathbf{F}_{p,j}(\mathbf{s},\mathbf{t}) - \left(\int_{\mathbb{R}^p} \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{s}}{h_n}\right) \mathrm{d}\mathbf{F}_p(\mathbf{s})\right)^2.$$

We only need to take care of the first integral. As $\mathbf{U}$ is an integral, we may use Fubini's Theorem followed by a standard change of variable, as before. Next, expand $\mathbf{F}_{p,j}$ to the second order, use (A3) to make the linear terms equal to zero and (A4) to control the coefficients of h_n^2, finding $\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x}) + O(h_n^2)$. This together with the characterization of the behaviour of $\mathbb{E}[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})]$, as given in Theorem 3.1, completes the proof of part (i). To prove part (ii), first use Lemma 3.3 to find

$$\mathrm{Cov}\left(\mathbb{I}_{(-\infty,\mathbf{x}]}(\mathbf{X}_{1,p}), \mathbb{I}_{(-\infty,\mathbf{x}]}(\mathbf{X}_{j,p})\right) \leq \sum_{k=1}^{p} \sum_{i=1}^{p} \mathrm{Cov}\left(\mathbb{I}_{(-\infty,x_k]}(X_k), \mathbb{I}_{(-\infty,x_{j+i}]}(X_{j+i})\right). \quad (3)$$

From (A1) and Roussas (1995, Lemma 2.6), there exits a constant $M > 0$ such that

$$\mathrm{Cov}\left(\mathbb{I}_{(-\infty,x_k]}(X_k), \mathbb{I}_{(-\infty,x_{j+i}]}(X_{j+i})\right) \leq M\,(\mathrm{Cov}\,(X_k, X_{j+i}))^{1/3}. \quad (4)$$

Inserting Equation (4) in Equation (3) and taking into account the stationarity of the random variables, the proofs follows. ■

REMARK 3.1 Notice that such as in Cai and Roussas (1998), if we assume that the covariance sequence $\mathrm{Cov}\,(X_1, X_{j+1})$, for $j \geq 1$, is decreasing, then, under the same assumptions as in Lemma 3.2, the upper bound $\mathbf{I}_j(\mathbf{x}) \leq p^2(\mathrm{Cov}\,(X_1, X_{j+1}))^{1/3}$ holds.

We may finally conclude the almost sure consistency of the estimator $\widehat{(\mathbf{F}_p)}_n$ in the following theorem.

THEOREM 3.4 Assume the sequence of random variables $\{X_n, n \geq 1\}$ satisfy (A1)-(A4) and (A6). Then, for each $\mathbf{x} \in \mathbb{R}^p$, $\widehat{(\mathbf{F}_p)}_n(\mathbf{x}) \to \mathbf{F}_p(\mathbf{x})$ almost surely.

PROOF As proved in Theorem 3.1, $\mathbb{E}[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})] \to \mathbf{F}_p(\mathbf{x})$. Thus, it is enough to verify that $\mathbf{U}(\{\mathbf{x}-\mathbf{X}_{m,p}\}/h_n)$, for $m \geq 1$, satisfy a SLLN. The stationarity of $\mathbf{U}$ is obvious. In addition, as $\mathbf{U}$ is a distribution function, it is coordinatewise increasing and so these variables are also statistically associated. Then, according to Newman (1980), the condition

$$\lim_{n\to\infty} \frac{1}{(n-p)} \sum_{j=1}^{n-p} I_{n,j}(\mathbf{x}) = 0 \tag{5}$$

implies the SLLN. Now, it follows from Lemma 3.2 that

$$\mathbf{I}_{n,j}(\mathbf{x}) \leq M \sum_{k=1}^{p} (p-k+1)(\mathrm{Cov}\,(X_1, X_{j+k}))^{1/3} + M \sum_{k=1}^{p-1} (p-k)(\mathrm{Cov}\,(X_1, X_{j-k+1}))^{1/3} + O(h_n^2)$$

and so Equation (5) is a consequence of (A6) as well as the association of the variables. ■

4. Mean Square Error

In this section, we characterize the asymptotic behaviour and convergence rate of the MSE of $\widehat{(\mathbf{F}_p)}_n$. From the results obtained below, it follows immediately the optimal bandwidth convergence rate of order n^{-1}. Thus, we have a different convergence rate than that for the independent case, such as was already noted in Cai and Roussas (1998) for the one-dimensional case. The optimal rate for the bandwidth, when dealing with independent variables, is of order $n^{-1/3}$ for every dimension, such as shown in Jin and Shao (1999). Again, strengthening the assumptions on the decay rate of the covariances as done in Cai and Roussas (1998), we find a different description of the MSE, which gives, for associated variables,the optimal bandwidth rate of order $n^{-1/3}$.

As usual, let us to write

$$\mathrm{MSE}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right] = \mathrm{Var}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right] + \left(\mathbb{E}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right] - \mathbf{F}_p(\mathbf{x})\right)^2. \tag{6}$$

Since the behaviour of $\mathbb{E}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right]$ in Equation (6) is known (cf. Theorem 3.1), we need to describe the asymptotic behaviour and convergence rate for the variance term given in this equation, which is shown in the following lemma.

LEMMA 4.1 Assume the sequence of random variables $\{X_n, n \geq 1\}$ satisfy (A1)-(A4) and (A7). Then, for each $\mathbf{x} \in \mathbb{R}^p$,

(i) $\mathbb{E}\left(\mathbf{U}^2\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)\right) = \mathbf{F}_p(\mathbf{x}) - h_n \mathbf{V}_2(\mathbf{x}) + \frac{h_n^2}{2}\mathbf{V}_3(\mathbf{x}) + o(h_n^2)$;

(ii) $\left|\mathrm{Var}\left(\mathbf{U}\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)\right) - \mathbf{F}_p(\mathbf{x})(1-\mathbf{F}_p(\mathbf{x})) + h_n \mathbf{V}_2(\mathbf{x})\right|$
$= h_n^2\left(\mathbf{V}_3(\mathbf{x}) - \mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right) + o(h_n^2).$

Proof In what concerns part (i), we have, recalling the definition of $\mathbf{V}$,

$$\mathbb{E}\left(\left[\mathbf{U}\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)\right]^2\right)=\int_{\mathbb{R}^p}\left(\int_{(-\infty,\frac{\mathbf{x}-\mathbf{s}}{h_n}]}\mathbf{V}(\mathbf{a})\,\mathrm{d}\mathbf{a}\right)d\mathbf{F}_p(\mathbf{s})=\int_{\mathbb{R}^p}\mathbf{V}(\mathbf{a})\mathbf{F}_p(\mathbf{x}-\mathbf{a}h_n)\,\mathrm{d}\mathbf{a},$$

using Fubini's Theorem. Expand now $\mathbf{F}_p(\mathbf{x}-\mathbf{a}h_n)$ to the second order, recall the definitions of the auxiliary functions $\mathbf{V}_2$ and $\mathbf{V}_3$, and take into account (A4) to find the $o(h_n^2)$ term. In order to verify part (ii), decompose the variance in the standard way and apply part (i) together with Theorem 3.1 to conclude the proof. ■

Theorem 4.2 Assume the sequence of random variables $\{X_n, n\geq 1\}$ satisfy (A1), (A2), (A3), (A4), (A5), (A6) and (A7). Then, for each $\mathbf{x}\in\mathbb{R}^p$,

$$\begin{aligned}(n-p)\mathrm{Var}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right]&=\sigma^2(\mathbf{x})-h_n\mathbf{V}_2(\mathbf{x})+(n-p-1)h_n^2\left(\mathbf{V}_4(\mathbf{x})-\mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right)\\&\quad+O(h_n^2)-\mathbf{c}_n(\mathbf{x}),\end{aligned}$$

where $\sigma^2(\mathbf{x})=\mathbf{F}_p(\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})+2\sum_{j=2}^{\infty}\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right)$, for $\mathbf{x}\in\mathbb{R}^p$, and

$$\mathbf{c}_n(\mathbf{x})=2\sum_{j=n-p+1}^{\infty}\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right)+\frac{2}{(n-p)}\sum_{j=2}^{n-p}(j-1)\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right),\ \mathbf{x}\in\mathbb{R}^p.$$

Proof Use the stationarity of the random variables and Lemmas 3.2 and 4.1 to write

$$\begin{aligned}(n-p)\mathrm{Var}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right]&=\mathbf{F}_p(\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})-\mathbf{V}_2(\mathbf{x})h_n+\left(\mathbf{V}_3(\mathbf{x})-\mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right)h_n^2\\&\quad+2\sum_{j=2}^{n-p}\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right)+(n-p-1)h_n^2\left(\mathbf{V}_4(\mathbf{x})-\mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right)\\&\quad-\frac{2}{(n-p)}\sum_{j=2}^{n-p}(j-1)\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right)+O(h_n^2).\end{aligned}$$

Summing and subtracting terms of the form $\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{x})-\mathbf{F}_p^2(\mathbf{x})\right)$, the result follows. ■

We may now summarize the above results to describe the behaviour of the mean square error.

Theorem 4.3 Assume the sequence of random variables $\{X_n, n\geq 1\}$ satisfy (A1), (A2), (A3), (A4), (A5), (A6) and (A7). Then, for each $x\in\mathbb{R}^p$,

$$(n-p)\mathrm{MSE}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right]=\sigma^2(\mathbf{x})-h_n\,\mathbf{V}_2(\mathbf{x})+O(n\,h_n^2)+o(h_n+n\,h_n^2)-\mathbf{c}_n(\mathbf{x}).$$

Notice that, with the assumptions made, $c_n\to 0$ and it is independent of the bandwidth choice. So, to find the optimal bandwidth rate, it is enough to minimize the term $o(\cdot)$, which is achieved by choosing $h_n=O(n^{-1})$, for each dimension p. It is possible to give a more explicit expression to the optimal bandwidth taking into account the characterizations in Lemma 4.1 and Theorems 4.2 and 4.3. Tracing the coefficients given in the expressions of these results, it is possible to check that we should choose

$$h_n(\mathbf{x}) = \frac{\mathbf{V}_2(\mathbf{x})}{2(n-p-1)\left(\mathbf{V}_4(\mathbf{x}) - \mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right)}.$$

We now strengthen the assumptions on the covariance decrease rate assuming the stronger conditions (A5*) and (A6*) and showing that this reflects on the optimal bandwidth rate, recovering the same rate as in the independent case.

THEOREM 4.4 Assume $\mathrm{Cov}\,(X_1, X_{j+1})$ decreases as j increases and the variables $\{X_n, n \geq 1\}$ satisfy (A1), (A2), (A3), (A4), (A5*), (A6*) and (A7). Then, for each $\mathbf{x} \in \mathbb{R}^p$,

$$(n-p)\mathrm{MSE}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right] = \sigma^2(\mathbf{x}) - h_n\,\mathbf{V}_2(\mathbf{x}) + O(n\,h_n^4) + o(h_n + n\,h_n^4) - \mathbf{c}_n(\mathbf{x}).$$

PROOF Recall that, as shown in Lemma 3.2, $\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x}) = O(h_n^2)$, and, as noted in Remark 3.1, when the covariances are decreasing, we have $\mathbf{I}_j(\mathbf{x}) \leq p^2(\mathrm{Cov}\,(X_1, X_{j+1}))^{1/3}$, and the same inequality holds for $\mathbf{I}_{n,j}(\mathbf{x})$. Then, it follows that, for a constant $c > 0$,

$$|\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})| = |\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})|^{\tau}\,|\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})|^{1-\tau} \leq \tilde{c}\,h_n^{2\tau}\left|(\mathrm{Cov}\,(X_1, X_{j+1}))^{(1-\tau)/3}\right|,$$

where $\tilde{c} = c^{\tau}\,p^{2(1-\tau)}$. Let us now write the variance of the estimator $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$ as

$$(n-p)\mathrm{Var}\left[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\right] = \mathrm{Var}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right)\right) + \frac{2}{(n-p)}\sum_{j=2}^{n-p}(n-p-j+1)\left(\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})\right)$$
$$+\sum_{j=2}^{n-p}(n-p-j+1)\mathbf{I}_j(\mathbf{x}).$$

Using (A6*), we have that

$$\frac{1}{(n-p)}\sum_{j=2}^{n-p}(n-p-j+1)\,|\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})| \leq \sum_{j=2}^{n-p}|\mathbf{I}_{n,j}(\mathbf{x}) - \mathbf{I}_j(\mathbf{x})|$$
$$\leq \tilde{c}\,h_n^{2\tau}\sum_{j=2}^{\infty}\left(\mathrm{Cov}\,(X_1, X_{j+1})\right)^{(1-\tau)/3} = O(h_n^{2\tau}),$$

The result now follows readily repeating the arguments as in the proof of Theorem 4.2. ■

An optimization of the MSE of $\widehat{(\mathbf{F}_p)}_n(\mathbf{x})$ leads now to the choice of a bandwidth of order $n^{-1/3}$, the optimal rate for the estimator when dealing with an independent sequence of random variables.

5. Finite Dimensional Distributions

In this section, we look at the asymptotic behaviour of the finite dimensional distributions of the empirical process induced by estimator $\widehat{(\mathbf{F}_p)}_n$.

The method is based on a decomposition of the sum defining the estimator into several blocks. These blocks are afterwards replaced by independent variables with the same distributions as the original blocks, followed by an application of the Lindeberg central limit theorem (CLT) to these independent copies. The distance between the original blocks and the replacing variables is controlled via Newman's characteristic functions inequality; see Newman (1984).

In order to state our result in a more tractable way, let us define, for each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\alpha_n(\mathbf{x}) = \sqrt{n-p}\left(\widehat{(\mathbf{F}_p)}_n(\mathbf{x}) - \mathbb{E}\big[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\big]\right) \tag{7}$$

and

$$\widetilde{\alpha}_n(\mathbf{x}) = \sqrt{n-p}\big(\widehat{(\mathbf{F}_p)}_n(\mathbf{x}) - F_p(\mathbf{x})\big) = \alpha_n(\mathbf{x}) + \sqrt{n-p}\left(\mathbb{E}\big[\widehat{(\mathbf{F}_p)}_n(\mathbf{x})\big] - F_p(\mathbf{x})\right). \tag{8}$$

The last term on the right hand of Equation (8) converges to zero using Theorem 3.1. Thus, we need to concentrate our attention on $\alpha_n(\mathbf{x})$ given in Equation (7). Define further

$$\varsigma^2(\mathbf{x}, \mathbf{y}) = \mathbf{F}_p(\mathbf{x} \wedge \mathbf{y}) - \mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y}) + \sum_{j=2}^{\infty} \left(\mathbf{F}_{p,j}(\mathbf{x}, \mathbf{y}) + \mathbf{F}_{p,j}(\mathbf{y}, \mathbf{x}) - 2\mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y})\right), \tag{9}$$

where $\mathbf{x} \wedge \mathbf{y}$ denotes the vector $(\min\{x_1, y_1\}, \ldots, \min\{x_p, y_p\})$. Notice that $\varsigma^2(\mathbf{x}, \mathbf{y})$ given in Equation (9) is symmetric in $\mathbf{x}$ and $\mathbf{y}$.

Theorem 5.1 Assume the variables $\{X_n, n \geq 1\}$ satisfy (A1), (A2), (A3), (A4), (A5), (A6) and (A7). Then, for $\mathbf{x_1}, \ldots, \mathbf{x_s} \in \mathbb{R}^p$, with $s \geq 1$, the random vector $(\widetilde{\alpha}_n(\mathbf{x_1}), \ldots, \widetilde{\alpha}_n(\mathbf{x_s}))$ converges in distribution to a Gaussian centered random vector with covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \varsigma^2(\mathbf{x_1}, \mathbf{x_1}) & \varsigma^2(\mathbf{x_1}, \mathbf{x_2}) & \cdots & \varsigma^2(\mathbf{x_1}, \mathbf{x_s}) \\ \varsigma^2(\mathbf{x_2}, \mathbf{x_1}) & \varsigma^2(\mathbf{x_2}, \mathbf{x_2}) & \cdots & \varsigma^2(\mathbf{x_2}, \mathbf{x_s}) \\ \vdots & \vdots & \ddots & \vdots \\ \varsigma^2(\mathbf{x_s}, \mathbf{x_1}) & \varsigma^2(\mathbf{x_s}, \mathbf{x_2}) & \cdots & \varsigma^2(\mathbf{x_s}, \mathbf{x_s}) \end{pmatrix}$$

The proof of Theorem 5.1 is divided into several lemmas. We start by describing the asymptotic behaviour of the covariances of the α_n at different points. These are needed to characterize the variance of some auxiliary variables for the proof of Theorem 5.1.

Lemma 5.2 Under the assumptions of Theorem 5.1, it holds that, for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, $\text{Cov}\,(\alpha_n(\mathbf{x}), \alpha_n(\mathbf{y})) \to \varsigma^2(\mathbf{x}, \mathbf{y})$.

Proof Using the stationarity of the variables, we may write

$$\begin{aligned}\text{Cov}\,(\alpha_n(\mathbf{x}), \alpha_n(\mathbf{y})) = {} & \text{Cov}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right), \mathbf{U}\left(\tfrac{\mathbf{y}-\mathbf{X}_{1,p}}{h_n}\right)\right) \\ & + \frac{1}{(n-p)} \sum_{j=2}^{n-p} (n-p-j+1)\Bigg(\text{Cov}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right), \mathbf{U}\left(\tfrac{\mathbf{y}-\mathbf{X}_{j,p}}{h_n}\right)\right) \\ & + \text{Cov}\left(\mathbf{U}\left(\tfrac{\mathbf{y}-\mathbf{X}_{1,p}}{h_n}\right), \mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{j,p}}{h_n}\right)\right)\Bigg).\end{aligned}$$

Repeating the arguments for the proof of Lemma 3.2, it follows that, for $j=1,\ldots,n-p$,

$$\mathrm{Cov}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right),\mathbf{U}\left(\tfrac{\mathbf{y}-\mathbf{X}_{1,p}}{h_n}\right)\right)=\mathbf{F}_{p,j}(\mathbf{x},\mathbf{y})-\mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y})+O(h_n^2).$$

Inserting this characterization in Equation (10), we find that the sum in this expression is equal to

$$\sum_{j=2}^{n-p}\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{y})-\mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y})\right)-\frac{1}{(n-p)}\sum_{j=2}^{n-p}(j-1)\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{y})-\mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y})\right)+O(nh_n^2).$$

Now, using Equation (4), it follows that

$$\begin{aligned}\sum_{j=2}^{n-p}\frac{(j-1)}{(n-p)}\left(\mathbf{F}_{p,j}(\mathbf{x},\mathbf{y})-\mathbf{F}_p(\mathbf{x})\mathbf{F}_p(\mathbf{y})\right)&\le\frac{1}{(n-p)}\sum_{j=2}^{n-p}j\mathrm{Cov}\left(\mathbb{I}_{(-\infty,\mathbf{x}]}(\mathbf{X}_{1,p}),\mathbb{I}_{(-\infty,\mathbf{y}]}(\mathbf{X}_{j,p})\right)\\&\le\frac{M}{(n-p)}\sum_{j=2}^{n-p}j\left(\sum_{k=1}^{p}(p-k+1)(\mathrm{Cov}\,(X_1,X_{j+k}))^{1/3}\right.\\&\qquad\left.+\sum_{k=1}^{p-1}(p-k)(\mathrm{Cov}\,(X_1,X_{j-k+1}))^{1/3}\right).\end{aligned}$$

Taking into account assumption (A6), this converges to zero. ■

Define now the decomposition of the sum into several blocks. Given an integer $r\le n-p$, let m be the largest integer less or equal than $(n-p)/r$. Denote

$$T_{n,i}(\mathbf{x})=\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)-\mathbb{E}\left(\mathbf{U}\left(\tfrac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)\right),\ \ Y_j^r(\mathbf{x})=\frac{1}{\sqrt{r}}\sum_{i=(j-1)r+1}^{jr}T_{n,i}(\mathbf{x}),$$

$$W_j^r=\sum_{q=1}^{s}c_qY_j^r(\mathbf{x}_q),\ \ Z_{n,i}=\sum_{q=1}^{s}c_qT_{n,i}(\mathbf{x}_q),\ \ Z_n=\frac{1}{\sqrt{n-p}}\sum_{q=1}^{s}c_q\sum_{i=1}^{n-p}T_{n,i}(\mathbf{x}_q),$$

where $c_1,\ldots,c_q\in\mathbb{R}$. The variable Z_n is the linear combination of the coordinates of $(\alpha_n(\mathbf{x}_1),\ldots,\alpha_n(\mathbf{x}_s))$ needed for the application of the Cramer-Wold Theorem. Define further

$$Z_{mr}^*=\frac{1}{\sqrt{m}}\sum_{q=1}^{s}c_q\sum_{j=1}^{r}Y_j^r(\mathbf{x}_q)=\frac{1}{\sqrt{m}}\sum_{j=1}^{m}W_j^r=\frac{1}{\sqrt{mr}}\sum_{j=1}^{mr}Z_{n,i},$$

which replaces the sum up to $n-p$ by a sum with a multiple of r terms. Notice also that, as follows from Lemma 5.2,

$$\mathrm{Var}\,(Z_{mr}^*)\to\varsigma^2=\sum_{q=1}^{s}c_q^2\varsigma^2(\mathbf{x}_q,\mathbf{x}_q)+2\sum_{q=1}^{s-1}\sum_{\ell=q+1}^{s}c_qc_\ell\varsigma^2(\mathbf{x}_q,\mathbf{x}_\ell).\tag{10}$$

Further, for each r fixed, it follows from Lemma 3.2 (i) that

$$\text{Var}\,(Y_1^r(\mathbf{x})) = \frac{1}{r}\sum_{i,i'=1}^{r}\left(\mathbf{F}_{p,\,|i'-i+1|}(\mathbf{x},\mathbf{x}) - \mathbf{F}_p^2(\mathbf{x})\right) + O(rh_n^2), \tag{11}$$

$$\text{Var}\,\left(W_j^r\right) = \sum_{q,q'=1}^{s} c_q c_{q'} \frac{1}{r}\sum_{i,i'=1}^{r}\left(\mathbf{F}_{p,\,|i'-i+1|}(\mathbf{x}_q,\mathbf{x}_{q'}) - \mathbf{F}_p(\mathbf{x}_q)\mathbf{F}_p(\mathbf{x}_{q'})\right) + O(rh_n^2). \tag{12}$$

We now proceed with the steps for the proof of Theorem 5.1. First, we show that we may replace the sum of $n-p$ terms defined by Z_n by the sum Z_{mr}^* to get only a sum of the blocks W_j^r.

Lemma 5.3 Under the assumptions of Theorem 5.1, for each r fixed, we have that

$$\left|\mathbb{E}(\mathrm{e}^{itZ_n}) - \mathbb{E}(\mathrm{e}^{itZ_{mr}^*})\right| \to 0.$$

Proof Using Hölder's inequality, we find

$$\begin{aligned}\left|\mathbb{E}\left(\mathrm{e}^{itZ_n}\right) - \mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}\right)\right| &\leq 2\,|t|\,\mathbb{E}\,|Z_n - Z_{mr}^*| \leq 2\,|t|\,(\text{Var}\,(Z_n - Z_{mr}^*))^{1/2} \qquad (13)\\ &\leq 2\sqrt{2}\,|t|\left(\left(\frac{1}{\sqrt{mr}} - \frac{1}{\sqrt{n-p}}\right)^2 \mathbb{E}\left(\sum_{i=1}^{mr} Z_{n,i}\right)^2 + \right.\\ &\qquad \left.\frac{1}{(n-p)}\,\mathbb{E}\left(\sum_{i=mr+1}^{n-p} Z_{\mathbf{n},i}\right)^2\right)^{1/2}.\end{aligned}$$

As $|Z_{n,i}| \leq 2\sum_{q=1}^{s}|c_q|$, it follows that

$$\left|\mathbb{E}\left(\mathrm{e}^{itZ_n}\right) - \mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}\right)\right| \leq \sqrt{8}\,|t|\left(\left(\tfrac{1}{\sqrt{mr}} - \tfrac{1}{\sqrt{n-p}}\right)^2 \text{Var}\,(Z_{mr}^*) + \tfrac{2(n-p-mr)}{(n-p)}\sum_{q=1}^{s}|c_q|\right)^{1/2},$$

which approaches zero, using Equation (10) and $mr/(n-p)$ approaches one, as follows from the choice of the integers m. ■

Thus, as what convergence in distribution is regarded, we may now replace Z_n by Z_{mr}^*. This last variable is a sum of m blocks. Thus, we are trying to prove a CLT for the sum of the dependent variables W_j^r, for $j \geq 1$. Each of these variables is a linear combination of the Y_j^r, which are decreasing functions of the original variables X_n, for $n \geq 1$. So, the Y_j^r's are statistically associated and we may apply a convenient variation of Newman's inequality, to the variables W_j^r, for $j \geq 1$, as proved in Jacob and Oliveira (1999, Lemma 4.1), when coupling these variables with the independent ones following the same distribution as that of each W_j^r.

Lemma 5.4 Under the assumptions of Theorem 5.1, for each r fixed, we have that

$$\left|\mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*} - \prod_{j=1}^{m}\mathbb{E}\left(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\right)\right)\right| \leq 2t^2\left(\text{Var}\left(\frac{1}{\sqrt{mr}}\sum_{j=1}^{mr}T_{n,i}\right) - \text{Var}\,(Y_1^r)\right)\sum_{q,q'=1}^{s}c_q c_{q'}.$$

Proof Taking into account Jacob and Oliveira (1999, Lemma 4.1), we have

$$\begin{aligned}\left|\mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}-\prod_{j=1}^m \mathbb{E}\left(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\right)\right)\right| &\leq \frac{2t^2}{m}\sum_{\substack{i,j=1\\ i\neq j}}^m \mathrm{Cov}\left(W_i^r,W_j^r\right)\\ &= \frac{2t^2}{m}\sum_{\substack{i,j=1\\ i\neq j}}^m \sum_{q,q'=1}^s c_q c_{q'}\mathrm{Cov}\left(Y_i^r(\mathbf{x}_q),Y_j^r(\mathbf{x}_{q'})\right)\\ &= 2t^2\sum_{q,q'=1}^s c_q c_{q'}\sum_{\substack{i,j=1\\ i\neq j}}^m \mathrm{Cov}\left(\frac{1}{\sqrt{m}}Y_i^r(\mathbf{x}_q),\frac{1}{\sqrt{m}}Y_j^r(\mathbf{x}_{q'})\right).\end{aligned}$$

Using now the stationarity of the variables, it follows that

$$\begin{aligned}\left|\mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}-\prod_{j=1}^m \mathbb{E}\left(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\right)\right)\right| &\leq 2t^2\sum_{q,q'=1}^s c_q c_{q'}\left(\mathrm{Var}\left(\frac{1}{\sqrt{mr}}\sum_{j=1}^{mr}T_{n,i}\right)-\sum_{j=1}^m \mathrm{Var}\left(\frac{1}{\sqrt{m}}Y_j^r\right)\right)\\ &= 2t^2\sum_{q,q'=1}^s c_q c_{q'}\left(\mathrm{Var}\left(\frac{1}{\sqrt{mr}}\sum_{j=1}^{mr}T_{n,i}\right)-\mathrm{Var}\left(Y_1^r\right)\right).\end{aligned}$$

■

The next step is a CLT for the coupling variables. In order to keep the notation simpler, we denote these variables also by W_j^r. On the next lemma, and on this lemma only, we assume the variables are independent (in order to formally correct, we should introduce a new family of random variables with the same distribution as that of W_j^r).

Lemma 5.5 Under the assumptions of Theorem 5.1, for each r fixed, we have that

$$\left|\prod_{j=1}^m \mathbb{E}\big(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\big)-\mathrm{e}^{-\frac{t^2\varsigma_r^2}{2}}\right|\to 0,$$

where $\varsigma_r^2=\sum_{q,q'=1}^s c_q c_{q'}\frac{1}{r}\sum_{i,i'=1}^r\left(\mathbf{F}_{p,\,|i'-i+1|}(\mathbf{x}_q,\mathbf{x}_{q'})-\mathbf{F}_p(\mathbf{x}_q)\mathbf{F}_p(\mathbf{x}_{q'})\right)$.

Proof Apply the Lindeberg condition to the variables $m^{-1/2}W_j^r$, for $j=1,\ldots,m$. As these variables are sums, use Lema 4 in Utev (1990) to separate the variables and the unbounded of the $T_{n,i}$'s to conclude the proof. ■

Now, the proof of Theorem 5.1 follows by putting together all these partial results.

Proof [Theorem 5.1] We have

$$\begin{aligned}\left|\mathbb{E}\left(\mathrm{e}^{itZ_{\mathbf{n}}}\right)-\mathrm{e}^{-\frac{t^2a}{2}}\right| &\leq \left|\mathbb{E}\left(\mathrm{e}^{itZ_n}\right)-\mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}\right)\right|+\left|\mathbb{E}\left(\mathrm{e}^{itZ_{mr}^*}\right)-\prod_{j=1}^m \mathbb{E}\left(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\right)\right|\\ &\quad+\left|\mathbb{E}\left(\mathrm{e}^{\frac{it}{\sqrt{m}}W_j^r}\right)-\mathrm{e}^{-\frac{t^2\varsigma_r^2}{2}}\right|+\left|\mathrm{e}^{-\frac{t^2\varsigma_r^2}{2}}-\mathrm{e}^{-\frac{t^2a}{2}}\right|,\end{aligned}$$

where $a=\sum_{q,q'=1}^s c_q c_{q'}\varsigma^2(\mathbf{x}_q,\mathbf{x}_{q'})$. Assuming for the moment that r is fixed, it follows from

Lemmas 5.2-5.5 that

$$\limsup_{m\to+\infty}\left|\mathbb{E}\left(\mathrm{e}^{itZ_n}\right)-\mathrm{e}^{-\frac{t^2a}{2}}\right|\leq 2t^2\left[\mathrm{Var}\left(\sum_{j=1}^{mr}\frac{T_{n,i}}{\sqrt{mr}}\right)-\mathrm{Var}\left(Y_1^r\right)\right]\sum_{q,q'=1}^{s}c_qc_{q'}+\left|\mathrm{e}^{-\frac{t^2\varsigma_r^2}{2}}-\mathrm{e}^{-\frac{t^2a}{2}}\right|.$$

Now, if we let $r\to+\infty$, it follows that this upper bound converges to zero taking into account Equation (11) and the stationarity of the sequence of variables $\{X_n,\ n\geq 1\}$, thus proving the theorem. ■

6. Simulation Study

In this section, we show some simulation results describing the behaviour of the empirical process for finite values of n.

In order to obtain associated variables, we fix an integer $m\in\mathbb{N}$, simulate $Y_1,\ldots,Y_{n+m}$ with a suitable distribution and construct $X_i=\min(Y_i,\ldots,Y_{n+m-1})$. The distribution of the variables Y_i's is easily chosen so that the X_i's are uniformly distributed in $[0,1]$. Notice that m may be interpreted as a measure of until how far the variables are dependent. For each n and each m, we simulate 5000 paths for the empirical process. The first set of results correspond to the one dimensional case, for which we can produce graphical representations of the underlying densities. Thus, Figures 1-3 show the approximations, based on the simulated paths of the density of $\widetilde{\alpha}_n(t_i)$ for a fixed set of points $t_1,\ldots,t_L\in[0,1]$. The simulations were performed using the statistical software `R`, which can be freely downloaded at `http://www.r-project.org`; see R Development Core Team (2009). Notice that Theorem 5.1 only proves the convergence of the finite dimensional distributions and not the functional convergence of the empirical process itself. In Figure 1, we graph the approximations obtained for $n=50,250,1000$, and for $m=5$. These graphs show a nice behaviour, close to normality, but the basis variables are "almost independent", so this is not very surprising. In Figure 2, we show the graphs, for the same values of n, by increasing the degree of dependence between the variables and considering $m=10$. The convergence to a Gaussian distribution is slower, as expected, but the approximations seem quite good for larger values of n. For a fixed size of the sample, the effect of the degree of dependence is dramatic. We illustrate this in Figure 3 for $n=250$, allowing m to take the values 20, 50, 100. These graphs do not show much similarity with a Gaussian distribution. It is clear that the influence of m, measuring the degree of dependence, is by no means negligible.

In order to illustrate the behaviour for higher dimensional cases, we compute the theoretical and simulated MSE's for the two-dimensional and three-dimensional cases for sample size $n=50,250,1000$ with $m=5,10$, as above. Table 1 reports the largest absolute deviations between these two functions. As expected, the differences increase with the dependence parameter. There is a curious effect when going from one-dimensional to two-dimensional simulations, as the differences consistently decrease. This was unexpected and seems to be due to border effects on the estimation procedure.

Table 1. Largest deviations between theoretical and simulated MSE's.

	$n=50$			$n=250$			$n=1000$		
m	1D	2D	3D	1D	2D	3D	1D	2D	3D
5	0.78727	0.40643	0.52861	0.73905	0.35171	0.48225	0.67744	0.27498	0.41971
10	1.16461	0.73072	0.93648	1.00028	0.61839	0.82205	0.93897	0.50870	0.71263

The `R` code to perform simulations, create graphics and compute the theoretical and simulated MSE's is available from the authors upon request.

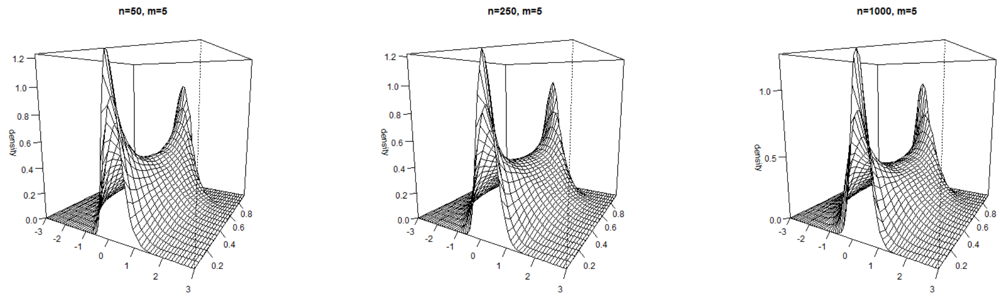

Figure 1. Simulated empirical processes for $n = 50, 250, 1000$ and $m = 5$.

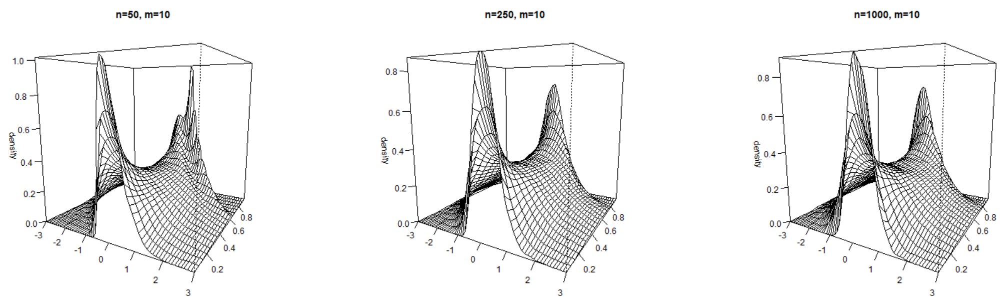

Figure 2. Simulated empirical processes for $n = 50, 250, 1000$ and $m = 10$.

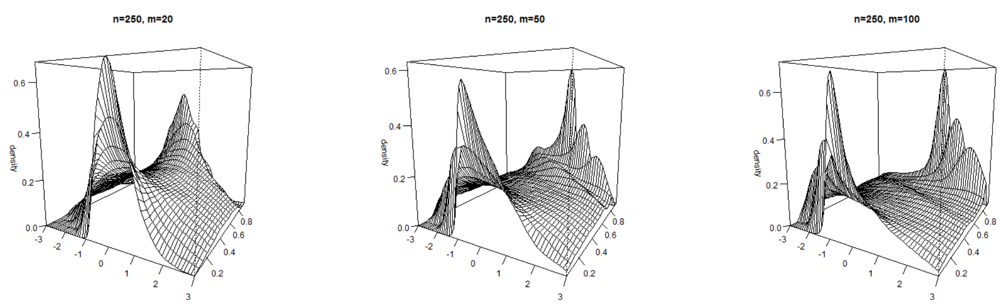

Figure 3. Simulated empirical processes for $n = 250$ and $m = 20, 50, 100$.

7. Conclusions

In this paper, we have considered the kernel estimator of the p-dimensional marginal distribution function of a stationary, positively associated sequence of random variables. For this setting, we have stated theoretical results about the asymptotical characterizations of this estimator. A simulation study has provided information about the behaviour of the finite dimensional distributions of the empirical process induced by the estimator. The convergence rates seem to be reasonable and mainly affected the parameter of the simulation model describing the degree of dependence, since this parameter may be interpreted as a measure of how far away the variables can be while remaining dependent. This should reflect the fact that the estimation process requires then considering more terms on the definition of the asymptotic covariance, thus needing a larger amount of information to find the same quality of approximation. From the point of view of applying the asymptotic characterizations, the results have indicated that reasonable sample sizes are required to make a reliable use of the asymptotic normality, unless one can find information regarding the degree of dependence parameter. Thus, we have established that, for sequences that are close to independence, i.e., for small values of the mentioned parameter, the asymptotic normality happens with a quite fast convergence rate.

Acknowledgements

The authors wish to thank the Executive Editor Víctor Leiva and two anonymous referees for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This study was partially supported by the CMat through the FCT Pluriannual Funding Program, Portugal.

References

Azevedo, C., Oliveira, P.E., 2000. Kernel-type estimation of bivariate distribution function for associated random variables. In New Trends in Probability and Statistics, Vol. 5 (Tartu, 1999), 1725, VSP, Utrecht, 2000.

Azevedo, C., Oliveira, P.E., 2005. On the multivariate kernel distribution estimator for distribution functions under association. In Proceedings of the 11th Symposium on ASMDA, May, 17- 20, Brest, France (Jacques Janssen and Philippe Lenca (eds.)), 1475 – 1482.

Cai, Z., Roussas, G., 1998. Efficient estimation of a distribution function under quadrant dependence. Scandinavian Journal of Statistics, 25, 211-224.

Esary, F., Proshan, F., Walkup, D.W., 1967. Association of random variables, with applications. The Annals of Mathematical Statistics, 36, 1466-1474.

Henriques, C., Oliveira, P.E., 2003. Estimation of a two-dimensional distribution function under association. Journal of Statistical Planning and Inference, 113, 137-150.

Jacob, P., Oliveira, P.E., 1999. Histograms and associated point processes. Statistical Inference for Stochastic Processes, 2, 227-251.

Jin, Z., Shao, Y., 1999. On kernel estimation of a multivariate distribution function. Statistics and Probability Letters, 41, 163-168.

Lehmann, E., 1966. Some concepts of dependence. The Annals of Mathematical Statistics, 37, 1137-1153.

Lebowitz, J., 1972. Bounds on the correlations and analyticity properties of ferromagnetic ising spin systems. Communications in Mathematical Physics, 28, 313-321.

Newman, C.M., 1980. Normal fluctuations and the FKG inequalities. Communications in Mathematical Physics, 74, 119-128.

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

Oliveira, P.E., Suquet, C., 1996. An $L^2[0,1]$ invariance principle for LPQD random variables. Portugaliae Mathematica, 53, 367-379.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Roussas, G., 1991. Kernel estimates under association: strong uniform consistency. Statistics and Probability Letters, 12, 393-403.

Roussas, G., 1995. Asymptotic normality of a smooth estimate of a random field distribution function under association. Statistics and Probability Letters, 24, 77-90.

Roussas, G., 2000. Asymptotic normality of the kernel estimate of a probability density function under association. Statistics and Probability Letters, 50, 1-12.

Utev, S.A., 1990. On the central limit theorem for ϕ mixing arrays of random variables. SIAM: Theory of Probability and its Applications, 35, 131-139.