Bivariate distribution function estimation for associated variables

Cecília Azevedo[*]
Dep. Matemática, Univ. Minho
Campus de Gualtar, 4710-057 Braga, Portugal

Paulo Eduardo Oliveira[†]
Dep. Matemática, Univ. Coimbra
Apartado 3008, 3001-454 Coimbra, Portugal

Abstract

The estimation of distribution functions of pairs of associated variables is addressed based on a kernel estimator. This problem is motivated by the need to approximate covariance functions appearing as the limiting covariances of the empirical process sequence. Results characterizing the asymptotics and convergence rates of the estimator are obtained. From these we derive the optimal bandwidth convergence rate, which is of order n^{-1}. Finally, we give conditions for the asymptotic normality of the finite dimensional distributions, characterizing their limit covariance matrix. Besides some usual conditions on the kernel function, the conditions typically impose a convenient decrease rate on the covariances $\mathrm{Cov}(X_1, X_n)$.

1 Introduction

The interest on approximating distribution functions of random pairs arises from the characterizations of the limiting distribution of empirical processes, which has been a subject of interest for many statisticians. In general, given random variables X_n, $n \geq 1$, with common distribution function G, the empirical process is defined

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbb{I}_{[-\infty,\, t]}(X_i) - G(t) \Big), \qquad t \in \mathbb{R},$$

where $\mathbb{I}_A$ represents the characteristic function of the set A. Many testing procedures are based on the asymptotic properties of the sequence Z_n or on some convenient transformation of Z_n. Some examples include the "goodness of fit" tests proposed by Watson [23] or by Anderson, Darling [1], where integral transforms of Z_n are used as testing statistic. There exist, in the literature, many more examples of integral transformations of Z_n being used of which we mention the Cramer-von Mises test. In some other applications, it is required to compute the supremum of $Z_n(t)$. These suggested the study of the asymptotic distribution of Z_n in the Skhorohod space. Note that, for the characterization of the convergence of the empirical process, considering the quantile function corresponding to G, it is enough to consider the variables X_n, $n \geq 1$, to be uniformly distributed on $[0, 1]$. The first results concerning the asymptotic distribution of the sequence date back to Donsker [5], for independent underlying variables X_n, $n \geq 1$, where the limit process was found to be the Brownian bridge. The extension of this characterization to nonindependent variables was eventually studied. Supposing the sequence X_n, $n \geq 1$, to be stationary, the asymptotic

[*]This author was partially supported by Centro de Matemática da Universidade do Minho
[†]This author was supported by grant PRAXIS/2/2.1/MAT/19/94 from FCT and Centro de Matemática da Universidade de Coimbra

distributions were studied by Billingsley [2] and Sen [18] for ϕ-mixing sequences, later replaced by strong mixing sequences. The best rate of convergence of the strong mixing coefficients α_n for which the convergence has been proved, was obtained by Shao [19] requiring $\alpha_n = O(n^{-a})$, with $a > 2$. In all cases, the limiting process is Gaussian centered with covariance function

$$
\begin{aligned}
\Gamma(x,y) &= G(x \wedge y) - G(x)G(y) + \\
&+ \sum_{k=1}^{\infty} \Big[(\mathrm{P}(X_1 \leq x, X_{k+1} \leq y) - G(x)G(y)) + (\mathrm{P}(X_1 \leq y, X_{k+1} \leq x) - G(x)G(y)) \Big],
\end{aligned}
\tag{1}
$$

where $x \wedge y = \min(x,y)$.

Another way of controlling dependence is association, introduced by Esary, Proschan, Walkup [6]. The random variables X_n, $n \geq 1$, are associated if, given $n \in \mathbb{N}$ and $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ coordinatewise increasing,

$$
\mathrm{Cov}\Big(f(X_1, \ldots, X_n),\, g(X_1, \ldots, X_n) \Big) \geq 0,
$$

whenever the covariance exists. For associated variables, it follows from Theorem 10 in Newman [12] that the covariances $\mathrm{Cov}(X_i,\, X_j)$ completely determine the dependence structure. So, for associated stationary variables it is natural to impose conditions on the decrease rate of $\mathrm{Cov}(X_1,\, X_n)$. The convergence in distribution of the empirical process for associated underlying variables was studied by Yu [22] and later by Shao, Yu [20] who obtained convergence for uniform $[0,1]$ variables such that $\mathrm{Cov}(X_1,\, X_n) = O(n^{-a})$ with $a > (3+\sqrt{33})/2 \approx 4.373$.

As mentioned before, in many applications, we consider integral transforms of the empirical process, so it is natural to restate the problem seeking convergence in the weaker space L^2. This case was studied in Oliveira, Suquet [13] who proved that, for uniform $[0,1]$ variables, it suffices that

$$
\sum_{n=2}^{\infty} \mathrm{Cov}^{1/3}(X_1,\, X_n) < \infty. \tag{2}
$$

In [13] were also considered strong mixing coefficients. As, in this article, we will consider only the associated case we do not quote the strong mixing characterization here. Extensions of these results to $L^p[0,1]$, $p \geq 2$, were considered in Oliveira, Suquet [14].

Note the appearance of the exponent $1/3$. This is due to the inequality,

$$
\mathrm{Cov}\left(\mathbb{I}_{(-\infty,\, s]}(Y_1),\, \mathbb{I}_{(-\infty,\, t]}(Y_2) \right) \leq B\ \mathrm{Cov}^{1/3}(Y_1,\, Y_2), \tag{3}
$$

for some constant $B > 0$, where Y_1, Y_2 are associated random variables with common distribution function with a bounded density (see Sadikova [17]).

The characterizations described above are of theoretical nature, giving a limiting covariance function Γ defined by (1). The purpose of the present article is to study approximations of each term $\mathrm{P}(X_1 \leq x, X_{k+1} \leq y)$ in (1).

2 Definitions and assumptions

As described above, we are interested in the estimation of $F(x,y) = \mathrm{P}(X_1 \leq x,\, X_{k+1} \leq y)$ with k fixed, a bivariate distribution function. A natural estimator would be the empirical distribution function

$$
\widehat{\varphi}_n(x,y) = \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(-\infty,\, x] \times (-\infty,\, y]}(X_i, X_{k+i}).
$$

The behaviour of this estimator was studied in Henriques, Oliveira [7]. Here we will be interested in the kernel estimator of F, defined by

$$\widehat{F}_n(x,y)=\frac{1}{n-k}\sum_{i=1}^{n-k}\mathcal{U}\left(\frac{x-X_i}{h_n},\,\frac{y-X_{k+i}}{h_n}\right)=\int_{\mathbb{R}^2}\mathcal{U}\left(\frac{x-s}{h_n},\,\frac{y-t}{h_n}\right)\,d\widehat{\varphi}_n(ds,dt), \tag{4}$$

where $\mathcal{U}$ is a given distribution function and h_n, $n\geq 1$, is a sequence of positive numbers converging to zero.

Analogous estimation problems have been addressed to by Cai, Roussas [3] and Roussas [15] for univariate distribution functions based on associated samples, and by Jin, Shao [10] for multivariate distributions functions but based on independent samples.

We now introduce the set of assumptions that will be referenced to throughout the text.

(A1) X_n, $n\geq 1$, is a strictly stationary sequence of associated random variables with bounded density function g;

(A2) k is a fixed integer and F the distribution function of (X_1,X_{k+1}). F has bounded and continuous partial derivatives of first and second orders;

(A3) For each positive integer j, F_j is the distribution function of $(X_1,X_{k+1},X_j,X_{k+j})$. F_j has bounded and continuous partial derivatives of first and second orders;

(A4) $\mathcal{U}$ is twice differentiable. If $u=\frac{\partial^2\mathcal{U}}{\partial x\partial y}$ it satisfies

$$\int_{\mathbb{R}^2} x\,u(x,y)\,dxdy=\int_{\mathbb{R}^2} y\,u(x,y)\,dxdy=0$$

$$\int_{\mathbb{R}^2} x^2\,u(x,y)\,dxdy<\infty,\qquad \int_{\mathbb{R}^2} y^2\,u(x,y)\,dxdy<\infty;$$

(A5) The sequence of bandwidths is such that $n\,h_n^2\longrightarrow 0$;

(A6) $\sum_{n=1}^{\infty} n\,\mathrm{Cov}^{1/3}(X_1,\,X_n)<\infty$;

(A7) $V=\frac{\partial^2 U^2}{\partial x\partial y}$ is such that

$$\int_{\mathbb{R}^2} x^2\,V(x,y)\,dxdy<\infty\qquad\text{and}\qquad\int_{\mathbb{R}^2} y^2\,V(x,y)\,dxdy<\infty.$$

Note that conditions **(A1)**, **(A2)**, **(A4)** and **(A6)** have already been used in Cai, Roussas [3] for the treatment of the univariate case. Note further that **(A6)** implies (2) which, as mentioned previously, implies the $L^2[0,1]$ weak convergence of the empirical process, as proved in Oliveira, Suquet [13].

In sections 3 and 4 we study the convergence and mean square error of $\widehat{F}_n$, with a treatment that follows the same lines as in Cai, Roussas [3]. Section 5 considers the asymptotic distribution of the finite dimensional distributions of $\widehat{F}_n$. Similar results were obtained by Roussas [16] for the estimation of density functions but only for unidimensional marginal distributions of the estimator, and by Cai, Roussas [4] for the estimation of distribution functions considering multivariate dimensional margins but for negatively associated variables. Here we consider multidimensional marginal distributions of $\widehat{F}_n$, the estimator of the bivariate distribution function we are interested in. Note that, changing our focus from estimating density functions, as in Roussas, to estimating distributions functions, as in this article, and from supposing the variables to be negatively associated, as in Cai, Roussas [4], to supposing the variables to be (positively) associated enables a relaxation on the conditions imposed, mainly on the kernel function. In fact, we will need only some second order integrability assumptions on the density associated to our kernel, as described by conditions **(A4)** and **(A7)**, whereas Roussas [16] supposed the kernel to be of bounded variation and decreasing fast enough to zero at infinity. The main reason for this difference relies on the fact that, being interested on estimating a distribution function, we transform the variables by a conveniently chosen distribution function, thus keeping the association property after the transformation. Besides, we do not need any assumption linking the bandwidth and the covariances $\mathrm{Cov}(X_1, X_n)$ as was needed in Roussas [16] or Cai, Roussas [4]

3 Consistency of the estimator

We first show that $\widehat{F}_n$ is asymptotically unbiased, characterizing also the convergence rate of $\mathrm{E}\left[\widehat{F}_n(x,y)\right]$. Then, to derive the asymptotic consistency of $\widehat{F}_n$, we apply a strong law of large numbers to the random variables $\mathcal{U}\left(\frac{x-X_i}{h_n}, \frac{y-X_{k+i}}{h_n}\right)$, $i = 1, \ldots, n-k$. To achieve this last step we shall need to characterize the behaviour of each entry of the covariance matrix of the random vector whose entries are the preceding variables, establishing their limits and convergence rates.

Theorem 3.1 *Suppose the variables X_n, $n \geq 1$, satisfy* **(A1), (A2)** *and* **(A4)***. Then, for each $x, y \in \mathbb{R}$,*

$$\mathrm{E}\left[\widehat{F}_n(x,y)\right] = F(x,y) + \frac{h_n^2}{2}\left[\frac{\partial^2 F}{\partial x^2}(x,y)\int s^2\, u(s,t)\, dsdt + \right.$$
$$\left. + \frac{\partial^2 F}{\partial x \partial y}(x,y)\int st\, u(s,t)\, dsdt + \frac{\partial^2 F}{\partial y^2}(x,y)\int t^2\, u(s,t)\, dsdt\right] + o(h_n^2).$$

Proof : As $\mathrm{E}\left[\widehat{\varphi}_n(x,y)\right] = F(x,y)$ it follows from (4) that

$$\mathrm{E}\left[\widehat{F}_n(x,y)\right] = \int_{\mathbb{R}^2} \mathcal{U}\left(\frac{x-s}{h_n}, \frac{y-t}{h_n}\right)\, dF(s,t) = \int_{\mathbb{R}^2} u(w,v)\, F(x - w\, h_n,\, y - v\, h_n)\, dwdv.$$

Using a Taylor expansion of order 2 of F and taking account of **(A2)** and **(A4)**, and of the continuity of the second order partial derivatives of F (assumption **(A2)**) the theorem follows. ■

Note that assumptions **(A2)** and **(A4)** are only required in order to establish a convergence rate. In fact, $\mathrm{E}\left[\widehat{F}_n(x,y)\right] \longrightarrow F(x,y)$ follows from an application of the Dominated Convergence Theorem.

In order to establish the almost sure convergence of $\widehat{F}_n$ we will apply a strong law of large numbers proved by Newman [11]. In course of proof we will need to control some covariances that are described in the following lemma.

Lemma 3.2 *Suppose the variables X_n, $n \geq 1$, satisfy* **(A1), (A2), (A3)** *and* **(A4)***. Then, for each $j \in \mathbb{N}$ and $x, y \in \mathbb{R}$,*

$$\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), \mathcal{U}\left(\frac{x-X_j}{h_n}, \frac{y-X_{k+j}}{h_n}\right)\right] = F_j(x,y,x,y) - F^2(x,y) + O(h_n^2).$$

Proof : Rewrite the covariance as

$$\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right),\mathcal{U}\left(\frac{x-X_j}{h_n},\frac{y-X_{k+j}}{h_n}\right)\right]=$$

$$=\int_{\mathbb{R}^4}\mathcal{U}\left(\frac{x-s}{h_n},\frac{y-t}{h_n}\right)\mathcal{U}\left(\frac{x-w}{h_n},\frac{y-v}{h_n}\right)\,dF_j(s,t,w,v)-\left(\int_{\mathbb{R}^2}\mathcal{U}\left(\frac{x-s}{h_n},\frac{y-t}{h_n}\right)\,dF(s,t)\right)^2.$$

The second term on the right is just $\mathrm{E}^2\left[\widehat{F}_n(x,y)\right]$, so its behaviour has been described in Theorem 3.1. As for the first term, writing the function $\mathcal{U}$ as an integral and using Fubini's Theorem,

$$\int_{\mathbb{R}^4}\mathcal{U}\left(\frac{x-s}{h_n},\frac{y-t}{h_n}\right)\mathcal{U}\left(\frac{x-w}{h_n},\frac{y-v}{h_n}\right)\,dF_j(s,t,w,v)=$$

$$=\int_{\mathbb{R}^4}u(a,b)\,u(c,d)F_j(x-ah_n,y-bh_n,x-ch_n,y-dh_n)\,da\,db\,dc\,dd.$$

Expanding F_j to the second order and using **(A3)** and **(A4)**, this integral is equal to $F_j(x,y,x,y)+O(h_n^2)$, which, together with the mentioned behaviour of $\mathrm{E}\left[\widehat{F}_n(x,y)\right]$, completes the proof of the lemma. ∎

We may now prove the almost sure convergence of the estimator $\widehat{F}_n$.

Theorem 3.3 *Suppose the variables X_n, $n\geq 1$, satisfy* **(A1), (A2), (A3), (A4), (A7)** *and (2). Then, for every $x,y\in\mathbb{R}$, $\widehat{F}_n(x,y)\longrightarrow F(x,y)$ almost surely.*

Proof : As proved in Theorem 3.1, $\mathrm{E}\left[\widehat{F}_n(x,y)\right]\longrightarrow F(x,y)$, so it is enough to prove that the variables $\mathcal{U}\left(\frac{x-X_m}{h_n},\frac{y-X_{k+m}}{h_n}\right)$, $m\geq 1$, satisfy a strong law of large numbers. These variables are stationary and associated, as $\mathcal{U}$ is coordinatewise nondecreasing. Then, according to Newman [11], they satisfy a strong law of large numbers if

$$\lim_{n\to\infty}\frac{1}{n-k}\sum_{j=1}^{n-k}\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right),\mathcal{U}\left(\frac{x-X_j}{h_n},\frac{y-X_{k+j}}{h_n}\right)\right]=0. \qquad (5)$$

From Lemma 3.2 and using (3), it follows

$$\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right),\mathcal{U}\left(\frac{x-X_j}{h_n},\frac{y-X_{k+j}}{h_n}\right)\right]=F_j(x,y,x,y)-F^2(x,y)+O(h_n^2)\leq$$

$$\leq 4\,B\,\mathrm{Cov}^{1/3}(X_1,\,X_j)+O(h_n^2).$$

Now, condition (5) is a consequence of (2) and association, so the theorem follows. ∎

Requiring an exponential decrease rate on the covariances, instead of (2), we may give a rate for the preceding convergence.

Theorem 3.4 *Suppose the variables X_n, $n\geq 1$, are strictly stationary and that there exists $a>1$ such that*

$$\mathrm{Cov}(X_1,\,X_n)=O(a^{-n}). \qquad (6)$$

Choose $\alpha_n\longrightarrow+\infty$ such that $\frac{\alpha_n\log n}{n^2}\longrightarrow 0$ and $\psi_n=c_1\frac{n^{1/3}}{\alpha_n^{2/3}\log^{2/3}n}$, for some constant $c_1>0$. Then

$$\psi_n\left(\widehat{F}_n(x,y)-\mathrm{E}\left[\widehat{F}_n(x,y)\right]\right)\longrightarrow 0\qquad a.s..$$

Proof : Using (3) it follows from (6) that

$$C(x,y,p) := \mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), \mathcal{U}\left(\frac{x-X_p}{h_n}, \frac{y-X_{k+p}}{h_n}\right)\right] = O\left(a^{-p/3}\right).$$

Now, according to Ioannides, Roussas [8], for each $\varepsilon > 0$, there exist positive constants c_0 and c such that,

$$\mathrm{P}\left[\psi_n\left(\widehat{F}_n(x,y) - \mathrm{E}\left[\widehat{F}_n(x,y)\right]\right) > \varepsilon\right] \le c_0\, e^{-c\frac{r_n\varepsilon^2}{\psi^2}} \tag{7}$$

where r_n is the largest integer less or equal than $\frac{n}{2p_n}$ and $p_n \longrightarrow +\infty$, provided that

$$C(x,y,p_n) \le \exp\left(-\frac{r_n\,\varepsilon}{\psi_n}\right).$$

It is also required that $r_n \longrightarrow +\infty$. We now check that it is possible to find such sequences. This last inequality follows from

$$p_n > c'\,\frac{r_n\,\varepsilon}{\psi_n} = c'\,\frac{r_n\,\varepsilon\,\alpha^{2/3}\log^{2/3} n}{n^{1/3}},$$

where c' stands for some positive constants, not necessarily the same. As $r_n \sim \frac{n}{2p_n}$, this is equivalent to

$$\frac{n}{r_n} > c'\,\frac{r_n\,\varepsilon\,\alpha^{2/3}\log^{2/3}}{n^{1/3}} \Leftrightarrow \frac{c'}{\varepsilon\, r_n^2} > \frac{\alpha_n^{2/3}\log^{2/3} n}{n^{4/3}},$$

so that the choice of r_n is compatible with the choice made for the sequence α_n. We should then choose r_n as large as possible fulfilling this last inequality, that is, we choose

$$r_n = c'\,\frac{n^{2/3}}{\alpha_n^{1/3}\log^{1/3} n}.$$

It follows then that

$$p_n \sim n^{1/3}\alpha_n^{1/3}\log^{1/3} n \quad \text{and} \quad \frac{r_n\,\varepsilon^2}{\psi^2} = c'\alpha_n \log n,$$

so that the probability in (7) defines a convergent series, thus the almost sure convergence follows from the Borel-Cantelli Lemma. ■

We may, in fact, prove the uniform consistency of the estimator under the same set of conditions as in Theorem 3.3.

Theorem 3.5 *Suppose the variables X_n, $n \ge 1$, satisfy* **(A1), (A2), (A3), (A4), (A7)** *and (2). Then*

$$\sup_{x,y\in\mathbb{R}} \left|\widehat{F}_n(x,y) - F(x,y)\right| \longrightarrow 0 \qquad a.s.$$

Proof : The proof follows the usual steps after a decomposition of $\mathbb{R}^2$ on a fixed set of suitably chosen points and establishing convenient inequalities. Let $M > 1$ be fixed and Q the quantile function corresponding to G (recall that G is the marginal distribution of both the coordinates of (X_1, X_{k+1})). Define the points $x_{M,i} = Q(i/M)$, $i = 1, \ldots, M$. Then, from Theorem 3.3, it follows that

$$\Delta_{M,n} = \max_{0\le i,j\le M}\left|\widehat{F}_n\left(\frac{i}{M}, \frac{j}{M}\right) - F\left(\frac{i}{M}, \frac{j}{M}\right)\right| \longrightarrow 0 \qquad \text{a.s..}$$

Now, as $\widehat{F}_n$ is nondecreasing, it follows easily that, for all $x, y \in \mathbb{R}$,

$$\left|\widehat{F}_n(x,y) - F(x,y)\right| \le \Delta_{M,n} + \frac{2}{M},$$

from which the theorem follows, as M is arbitrary. ■

4 The behaviour of the mean square error

In this section we study the asymptotics and convergence rate of the mean square error. This characterization will then be used to derive the optimal bandwidth convergence rate. This convergence rate for the bandwidth is, as it will be explained later, of order n^{-1}, thus a different convergence rate than the one in the independent case. This confirms a modification on the behaviour of h_n already noticed in Cai, Roussas [3].

As usual we write

$$\mathrm{MSE}\left[\widehat{F}_n(x,y)\right] = \mathrm{Var}\left[\widehat{F}_n(x,y)\right] + \left(\mathrm{E}\left[\widehat{F}_n(x,y)\right] - F(x,y)\right)^2 .$$

The behaviour of $\mathrm{E}\left[\widehat{F}_n(x,y)\right]$ being known (cf. Theorem 3.1), we need to describe the asymptotics and convergence rate for the variance term. For this purpose write

$$\begin{aligned}\mathrm{Var}\left[\widehat{F}_n(x,y)\right] &= \frac{1}{n-k}\,\mathrm{Var}\left[\mathcal{U}\left(\frac{x-X_1}{h_n},\,\frac{y-X_{k+1}}{h_n}\right)\right] + \\ &+\frac{2}{(n-k)^2}\sum_{j=2}^{n-k}(n-k-j+1)\,\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n},\,\frac{y-X_{k+1}}{h_n}\right),\mathcal{U}\left(\frac{x-X_j}{h_n},\,\frac{y-X_{k+j}}{h_n}\right)\right].\end{aligned} \tag{8}$$

The asymptotic behaviour of all these terms has been described in Lemma 3.2. Just notice that the variance term, which corresponds to the choice $j=1$ in Lemma 3.2, gives as limit $F_1(x,y,x,y) - F^2(x,y) = F(x,y) - F^2(x,y)$. The convergence rate for $\mathrm{MSE}\left[\widehat{F}_n(x,y)\right]$ now follows readily. We now state the result that summarizes the procedure.

Theorem 4.1 *Suppose the variables X_n, $n \geq 1$, satisfy* **(A1), (A2), (A3), (A4), (A5), (A6)** *and* **(A7)**. *Then, for all $x, y \in \mathbb{R}$,*

$$(n-k)\,\mathrm{MSE}\left[\widehat{F}_n(x,y)\right] = F(x,y) - F^2(x,y) + 2\sum_{j=2}^{\infty}\left(F_j(x,y,x,y) - F^2(x,y)\right) + O(h_n + n\,h_n^2) + a_n$$

where

$$a_n = \frac{1}{n-k}\sum_{j=2}^{\infty}(j-1)\left(F_j(x,y,x,y) - F^2(x,y)\right) - 2\sum_{j=n-k-1}^{\infty}\left(F_j(x,y,x,y) - F^2(x,y)\right).$$

Note that $a_n \longrightarrow 0$, according to the assumptions made, and that a_n is independent of the bandwidth choice.

It is now evident that an optimization of the convergence rate of the MSE is achieved by choosing $h_n = c\,n^{-1}$.

5 Finite dimensional distributions

We now study the asymptotic behaviour of the finite dimensional distributions of the estimator. The method of proof is based on a decomposition of the sum (4) into the sum of several blocks. These blocks will afterwards be coupled with independent variables with the same distributions as the original blocks followed by an application of the Lindeberg Central Limit Theorem. This coupling is controlled via Newman's inequality [12]. As the proof is somewhat long and quite technical we will divide it into several lemmas.

In order to state our result in a more tractable way let us define, for every $x, y, w, v \in \mathbb{R}$,

$$\alpha_n(x,y) = \sqrt{n-k}\, \left(\widehat{F}_n(x,y) - \mathrm{E}\left[\widehat{F}_n(x,y)\right]\right)$$

$$\sigma^2(x,y,w,v) = F(x\wedge w, y\wedge v) - F(x,y)F(w,v) + 2\sum_{j=2}^{\infty}\Big(F_j(x,y,w,v) - F(x,y)F(w,v)\Big).$$

Theorem 5.1 *Suppose that the random variables X_n, $n \geq 1$, satisfy **(A1), (A2), (A3), (A4), (A5), (A6)** and **(A7)**. Then, given $s \in \mathbb{N}$, and $x_1, \ldots, x_s, y_1, \ldots, y_s \in \mathbb{R}$, the random vector $\big(\alpha_n(x_1,y_1), \ldots, \alpha_n(x_s,y_s)\big)$ converges in distribution to a Gaussian centered random vector with covariance matrix*

$$\Sigma = \begin{bmatrix} \sigma^2(x_1,y_1,x_1,y_1) & \sigma^2(x_1,y_1,x_2,y_2) & \cdots & \sigma^2(x_1,y_1,x_q,y_q) \\ \sigma^2(x_2,y_2,x_1,y_1) & \sigma^2(x_2,y_2,x_2,y_2) & \cdots & \sigma^2(x_2,y_2,x_q,y_q) \\ \cdots & \cdots & \cdots & \cdots \\ \sigma^2(x_q,y_q,x_1,y_1) & \sigma^2(x_q,y_q,x_2,y_2) & \cdots & \sigma^2(x_q,y_q,x_q,y_q) \end{bmatrix}.$$

We start by describing the asymptotics of the covariances depending on the α_n at different points.

Lemma 5.2 *Suppose that the random variables X_n, $n \geq 1$, satisfy **(A1), (A2), (A3), (A4), (A5), (A6)** and **(A7)**. Then, for every $x, y, w, v \in \mathbb{R}$,*

$$\mathrm{Cov}\left[\alpha_n(x,y),\, \alpha_n(w,v)\right] \longrightarrow \sigma^2(x,y,w,v).$$

Proof : Using the stationarity of the variables we may write

$$\mathrm{Cov}\left(\alpha_n(x,y),\, \alpha_n(w,v)\right) =$$

$$= \mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), \mathcal{U}\left(\frac{w-X_1}{h_n}, \frac{v-X_{k+1}}{h_n}\right)\right] + \tag{9}$$

$$+\frac{2}{n-k}\sum_{j=2}^{n-k}(n-k-j+1)\,\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), \mathcal{U}\left(\frac{w-X_j}{h_n}, \frac{v-X_{k+j}}{h_n}\right)\right].$$

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

$$\mathrm{Cov}\left[\mathcal{U}\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), \mathcal{U}\left(\frac{w-X_j}{h_n}, \frac{v-X_{k+j}}{h_n}\right)\right] = F_j(x,y,w,v) - F(x,y)F(w,v) + O(h_n^2).$$

Inserting these characterizations in (9) we find that the sum is equal to

$$\sum_{j=2}^{n-k}\Big(F_j(x,y,w,v) - F(x,y)F(w,v)\Big) - \frac{1}{n-k}\sum_{j=2}^{n-k}(j-1)\Big(F_j(x,y,w,v) - F(x,y)F(w,v)\Big) + O(n\,h_n^2).$$

Now, using (3), it follows

$$\frac{1}{n-k}\sum_{j=2}^{n-k}(j-1)\Big(F_j(x,y,w,v)-F(x,y)F(w,v)\Big)\leq$$

$$\leq\frac{1}{n-k}\sum_{j=2}^{n-k} j\,\mathrm{Cov}\Big(\mathbb{I}_{(-\infty,x]\times(-\infty,y]}(X_1,X_{k+1}),\ \mathbb{I}_{(-\infty,w]\times(-\infty,v]}(X_j,X_{k+j})\Big)\leq$$

$$\leq\frac{4\,B}{n-k}\sum_{j=2}^{n-k} j\,\mathrm{Cov}^{1/3}(X_1,X_j)\longrightarrow 0.$$

according to **(A6)**. ■

For the lemmas concerning directly the proof of the asymptotic normality we need some further notation. Denote $\widetilde{n}=n-k$ and, given an integer $r\leq\widetilde{n}$, let m be the largest integer less or equal to $\widetilde{n}/r$. Define

$$T_{\widetilde{n},i}(x,y)=\mathcal{U}\left(\frac{x-X_i}{h_n},\frac{y-X_{k+i}}{h_n}\right)-\mathrm{E}\left[\mathcal{U}\left(\frac{x-X_i}{h_n},\frac{y-X_{k+i}}{h_n}\right)\right]$$

$$Y_j^r(x,y)=\frac{1}{\sqrt{r}}\sum_{i=(j-1)r+1}^{jr}T_{\widetilde{n},i}(x,y),\qquad W_j^r=\sum_{q=1}^{s}c_qY_j^r(x_q,y_q),$$

and

$$Z_{\widetilde{n},i}=\sum_{q=1}^{s}c_qT_{\widetilde{n},i}(x_q,y_q),\qquad Z_{\widetilde{n}}=\frac{1}{\sqrt{\widetilde{n}}}\sum_{i=1}^{\widetilde{n}}Z_{\widetilde{n},i}=\frac{1}{\sqrt{\widetilde{n}}}\sum_{q=1}^{s}c_q\sum_{i=1}^{\widetilde{n}}T_{\widetilde{n},i}(x_q,y_q).$$

The random variable $Z_{\widetilde{n}}$ is the linear combination of the coordinates of $\Big(\alpha_n(x_1,y_1),\ldots,\alpha_n(x_s,y_s)\Big)$ required for the application of the Cramer-Wold Theorem. Define further

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

which replaces the sum up to $\widetilde{n}$ by a sum with a multiple of r numbers of terms. Note also that, as follows from Lemma 5.2,

$$\mathrm{Var}(Z_{\widetilde{n}})\longrightarrow\sigma^2:=\sum_{q=1}^{s}c_q^2\sigma^2(x_q,y_q,x_q,y_q)+2\sum_{q=1}^{s-1}\sum_{l=q+1}^{s}c_qc_l\sigma^2(x_q,y_q,x_l,y_l).\tag{10}$$

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

$$\begin{aligned}\mathrm{Var}\left[Y_1^r(x,y)\right]&=\mathrm{Cov}\left[\frac{1}{\sqrt{r}}\sum_{i=1}^{r}T_{\widetilde{n},i}(x,y),\ \frac{1}{\sqrt{r}}\sum_{i'=1}^{r}T_{\widetilde{n},i'}(x,y)\right]=\\&=\frac{1}{r}\sum_{i,i'=1}^{r}\Big(F_{|i'-i+1|}(x,y,x,y)-F^2(x,y)\Big)+O(r\,h_n^2)\end{aligned}\tag{11}$$

and

$$\mathrm{Var}(W_j^r) = \sum_{q,q'=1}^{s} c_q c_{q'} \,\mathrm{Cov}\left[Y_j^r(x_q,y_q),\, Y_j^r(x_{q'},y_{q'})\right] =$$

$$= \sum_{q,q'=1}^{s} c_q c_{q'} \,\frac{1}{r} \sum_{i,i'=1}^{r} \left(F_{|i'-i+1|}(x_q,y_q,x_{q'},y_{q'}) - F(x_q,y_q)F(x_{q'},y_{q'})\right) + O(r\,h_n^2).$$

We now proceed directly to the proof of Theorem 5.1. First replace the sum of $\widetilde{n}$ terms defined by $Z_{\widetilde{n}}$ by the sum Z_{mr}^* to get only a sum of the blocks W_j^r.

Lemma 5.3 *Suppose the assumptions of Theorem 5.1 are satisfied and let r be fixed. Then*

$$\left|\mathrm{E}e^{itZ_{\widetilde{n}}} - \mathrm{E}e^{itZ_{mr}^*}\right| \longrightarrow 0.$$

Proof : Using Hölder's inequality, we find

$$\begin{aligned}\left|\mathrm{E}e^{itZ_{\widetilde{n}}} - \mathrm{E}e^{itZ_{mr}^*}\right| &\le 2\,|t|\,\mathrm{E}\left|Z_{\widetilde{n}} - Z_{mr}^*\right| \le 2\,|t|\,\mathrm{Var}^{1/2}(Z_{\widetilde{n}} - Z_{mr}^*) \\ &\le 2\sqrt{2}\,|t| \left[\left(\frac{1}{\sqrt{mr}} - \frac{1}{\sqrt{\widetilde{n}}}\right)^2 \mathrm{E}\left(\sum_{i=1}^{mr} Z_{\widetilde{n},i}\right)^2 + \frac{1}{\widetilde{n}}\,\mathrm{E}\left(\sum_{i=mr+1}^{\widetilde{n}} Z_{\widetilde{n},i}\right)^2\right]^{1/2}.\end{aligned} \tag{12}$$

Now, as $\left|Z_{\widetilde{n},i}\right| \le 2\,\sum_{q=1}^{s}\,|c_q|$, it follows

$$\left|\mathrm{E}e^{itZ_{\widetilde{n}}} - \mathrm{E}e^{itZ_{mr}}\right| \le 2\sqrt{2}\,|t| \left[\left(\frac{1}{\sqrt{mr}} - \frac{1}{\sqrt{\widetilde{n}}}\right)^2 \mathrm{Var}(Z_{mr}^*) + 2\,\frac{\widetilde{n}-mr}{\widetilde{n}} \sum_{q=1}^{s} |c_q|\right]^{1/2} \longrightarrow 0,$$

according to (10). ■

We may now replace the sum $Z_{\widetilde{n}}$ by the sum Z_{mr}^* as what convergence in distribution is regarded. The variable Z_{mr}^* is a sum of m blocks, so we are trying to prove a Central Limit Theorem for the sum of the dependent variables W_1^r, $W_2^r, \ldots$. Each of these variables is a linear combination of the Y_j^r which are decreasing functions of the original variables X_n, $n \ge 1$. It follows then that the Y_j^r are associated and we may apply a convenient variation of Newman's inequality [12] to the variables W_1^r, $W_2^r, \ldots$ as proved in Lemma 4.1 from Jacob, Oliveira [9] when coupling these variables with independent ones with the same distribution as each of the W_j^r.

Lemma 5.4 *Suppose the assumptions of Theorem 5.1 are satisfied and let r be fixed. Then*

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

Proof : According to Lemma 4.1 in [9] we have

$$\left|\mathrm{E}e^{itZ_{mr}^*} - \prod_{j=1}^{m} \mathrm{E}e^{\frac{it}{\sqrt{m}}W_j^r}\right| \le 2\,\frac{t^2}{m} \sum_{\substack{i,j=1\\ i\ne j}}^{m} \mathrm{Cov}(W_i^r,\, W_j^r) =$$

$$= 2\,\frac{t^2}{m} \sum_{\substack{i,j=1\\ i\neq j}}^{m} \sum_{q,q'=1}^{s} c_q c_{q'} \operatorname{Cov}\left[Y_i^r(x_q,y_q),\, Y_j^r(x_{q'},y_{q'})\right] =$$

$$= 2\,t^2 \sum_{q,q'=1}^{s} c_q c_{q'} \sum_{\substack{i,j=1\\ i\neq j}}^{m} \operatorname{Cov}\left[\frac{1}{\sqrt{m}} Y_i^r(x_q,y_q),\, \frac{1}{\sqrt{m}} Y_j^r(x_{q'},y_{q'})\right] =$$

$$= 2\,t^2 \sum_{q,q'=1}^{s} c_q c_{q'} \left[\operatorname{Var}\left(\frac{1}{\sqrt{mr}} \sum_{j=1}^{mr} T_{\widetilde{n},i}\right) - \sum_{j=1}^{m} \operatorname{Var}\left(\frac{1}{\sqrt{m}}\, Y_j^r\right)\right] =$$

$$= 2\,t^2 \sum_{q,q'=1}^{s} c_q c_{q'} \Big(\operatorname{Var}\left(\frac{1}{\sqrt{mr}} \sum_{j=1}^{mr} T_{\widetilde{n},i}\right) - \operatorname{Var}(Y_1^r)\Big),$$

due to the stationarity of the variables. ■

The next step is the proof of a Central Limit Theorem for the coupling of the variables W_j^r. In order to keep the notation as simple as it seems possible, we will denote these variables also by W_j^r. Of course, during the next lemma, and on this lemma only, we will suppose that the variables are independent. To describe the variances appearing on the next lemma let us define

$$\sigma_r^2 = \sum_{q,q'=1}^{s} c_q c_{q'}\, \frac{1}{r} \sum_{i,i'=1}^{r} \Big(F_{|i'-i+1|}(x_q,y_q,x_{q'},y_{q'}) - F(x_q,y_q)F(x_{q'},y_{q'})\Big).$$

Lemma 5.5 *Suppose the assumptions of Theorem 5.1 are satisfied and let r be fixed. Then*

$$\left|\prod_{j=1}^{m} \mathrm{E} e^{\frac{it}{\sqrt{m}} W_j^r} - e^{-\frac{t^2\sigma_r^2}{2}}\right| \longrightarrow 0.$$

Proof : We will apply the Lindeberg condition to the variables $m^{-1/2}W_j^r$, $j = 1,\ldots,m$. Remembering that $Z_{mr}^* = m^{-1/2}\sum_{j=1}^{m} W_j^r$ and (10), it follows that the verification of the Lindeberg condition reduces to checking that

$$\sum_{j=1}^{m} \int_{\left\{|m^{-1/2}W_j^r| \geq \varepsilon\sigma_r^2\right\}} \frac{1}{m} (W_j^r)^2 \, d\mathrm{P} \longrightarrow 0. \tag{13}$$

Using now Lemma 4 in Utev [21] the integral in (13) is bounded above by

$$\sum_{j=1}^{m} \frac{1}{m} \sum_{i=(j-1)r+1}^{jr} \int_{\left\{\left|T_{\widetilde{n},i}\right| \geq \varepsilon \frac{\sigma_r^2\sqrt{mr}}{r}\right\}} T_{\widetilde{n},i}^2 \, d\mathrm{P}\,.$$

Now, as for each $i = 1,\ldots,mr$ and $\widetilde{n} \in \mathbb{N}$, $T_{\widetilde{n},i} \leq 2\sum_{q=1}^{s} c_q$ the integration set is, for $\widetilde{n}$ large enough, empty, so each integral in this last sum is equal to zero. ■

Proof (of Theorem 5.1) : Let us define $a = \sum_{q,q'=1}^{s} c_q c_{q'} \sigma^2(x_q,y_q,x_{q'},y_{q'})$. The proof of the theorem reduces to verifying that

$$\left|\mathrm{E} e^{itZ_{\widetilde{n}}} - e^{-\frac{t^2 a}{2}}\right| \longrightarrow 0.$$

We have

$$\left|\mathrm{E}e^{itZ_{\widetilde{n}}} - e^{-\frac{t^2 a}{2}}\right| \leq$$

$$\leq \left|\mathrm{E}e^{itZ_{\widetilde{n}}} - \mathrm{E}e^{itZ^*_{mr}}\right| + \left|\mathrm{E}e^{itZ^*_{mr}} - \prod_{j=1}^{m} \mathrm{E}e^{\frac{it}{\sqrt{m}}W_j^r}\right| + \left|\prod_{j=1}^{m} \mathrm{E}e^{\frac{it}{\sqrt{m}}W_j^r} - e^{-\frac{t^2\sigma_r^2}{2}}\right| + \left|e^{\frac{t^2\sigma_r^2}{2}} - e^{-\frac{t^2 a}{2}}\right|$$

Supposing, for the moment, that r is fixed, it follows from the previous lemmas that

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

Letting now $r \longrightarrow +\infty$ it follows that this upper bound converges to zero on account of (11) and the stationarity of the variables X_n, $n \geq 1$, thus proving the theorem. ■

References

[1] Anderson, T. W., Darling, D. A., Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, Ann. Math. Statist. 23 (1952), 193–212.

[2] Billingsley, P., Convergence of Probability Measures. Wiley, New York, 1968.

[3] Cai, Z., Roussas, G., Efficient Estimation of a Distribution Function under Quadrant Dependence, Scand. J. Statist. 25 (1998), 211–224.

[4] Cai, Z., Roussas, G., Weak convergence for smooth estimator of a distribution function under negative association, Stochastic Anal. Appl. 17 (1999), 145–168.

[5] Donsker, M. D., An invariance principle for certain probability limit theorems. Mem. Amer. Math. Soc. 6 (1951).

[6] Esary, J. D., Proschan, F., Walkup, D. W., Association of random variables, with applications. Ann. Math. Statist. 38 (1967), 1466-1474.

[7] Henriques, C., Oliveira, P.E., Covariance estimation for associated random variables, Preprint, Dep. Matemática Univ. Coimbra 99-03 (1999).

[8] Ioannides, D. A., Roussas, G. G., Exponential inequality for associated random variables, Statist. Probab. Lett. 42 (1999), 423–431.

[9] Jacob, P., Oliveira, P. E., Histograms and associated point processes, to appear in Statist Inf. Stoch. Proc.

[10] Jin, Z., Shao, Y., On kernel estimation of a multivariate distribution function. Statist. Probab. Lett. 41 (1999), 163–168.

[11] Newman, C.M. (1980) Normal fluctuations and the FKG inequalities, Comm. Math. Phys., 75 (1980), 119–128.

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

[13] Oliveira, P. E., Suquet, C., $L^2[0,1]$ weak convergence of the empirical process for dependent variables. Actes des XV Rencontres Franco-Belges de Statisticiens (Ondelettes et Statistique), Lecture Notes in Statistics 103 (1995), Wavelets and Statistics, Ed. A. Antoniadis, G. Oppenheim, 331-344.

[14] Oliveira, P. E., Suquet, C., Weak convergence in $L^p[0,1]$ of the uniform empirical process under dependence, Statist. Probab. Letters 39 (1998), 363-370.

[15] Roussas, G., Asymptotic normality of a smooth estimate of a random field distribution function under association, Statist. Probab. Lett., 24 (1995) , 77–90.

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

[17] Sadikova, S. M., Two dimensional analogues of an inequality of Esséen with applications to the central limi theorem, Theory Probab. Appl. 11 (1966), 325–335.

[18] Sen, P. K., A note on weak convergence of empirical processes for sequences of ϕ-mixing random variables. Ann. Math. Statist. 42 (1971), 2131-2133.

[19] Shao, Q. M., Weak convergence of multidimensional empirical processes for strong mixing sequences, Chinese Ann. Math. Ser. A 7 (1986), 547-552.

[20] Shao, Q. M., Yu, H., Weak convergence for weighted empirical process of dependent sequences, Ann. Probab. 24 (1996), 2098-2127.

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

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

[23] Watson, G. S.,Goodness-of-fit tests on a circle, Biometrika 48 (1961), 109–114.