Simulation Results of Finite Dimensional Distributions of the Empirical Process

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

CMAT - Centro de Matemática

Universidade do Minho

4710-057 Braga, Portugal

cecilia@math.uminho.pt

Departamento de Matemática

Universidade de Coimbra

3000 Coimbra, Portugal

paulo@mat.uc.pt

Key Words: association; kernel estimator; optimal bandwidth; asymptotic normality.

Abstract

We consider kernel estimation of the p-dimensional marginal distribution function of a stationary associated sequence. We restate this setting available results concerning the asymptotic behaviour of the kernel estimator on the literature for somewhat reduced framework. We present some simulation results concerning the empirical process constructed form the estimator, illustrating the asymptotic normality recalled here.

1. Introduction and Assumptions

Estimation of distribution functions has been one of the classical problems in statistics. We will address this estimation using a kernel estimator and based on a stationary sequence satisfying a positive association assumption, and we will be interested in the p-dimensional marginal distribution function. The dependence structure of associated sequences is completely characterized by the covariances between the variables, as described by Theorem 10

[1]Author partially supported by CMAT and FCT under the program POCI 2010.

[2]Author partially supported by CMUC and FCT.

in Newman (1984). It is thus natural to seek sufficient conditions for the convergence of the estimators imposing some convenient decrease rate on the covariances. Estimation of distribution functions based on associated samples seems to have attracted attention first: Roussas (1991) studied the one-dimensional case, Cai and Roussas (1998) considered the same one-dimensional case but under positive quadrant dependence, a positive dependence related to association. 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 in Azevedo and Oliveira (2000) using the kernel estimator and in Henriques and Oliveira (2003) using the histogram estimator. This note, after rewriting the results in Azevedo and Oliveira (2000) for the p-dimensional setting, presents some simulation results concerning the empirical process. These simulations illustrate the convergence of the finite dimensional distributions of the empirical process, 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 being still dependent. The influence of the parameter is also illustrated. It is clear that, for sequences that are close to independence the asymptotic normality happens with quite fast convergence. We will include only sketches of proofs, as the adaptation of the two-dimensional results to the p-dimensional is almost straightforward, referring the reader to Azevedo and Oliveira (2000) for details.

Definition 1 (Esary et al. (1967)) *The random variables $X_1, X_2, \dots$ are said to be associated, if for every $k \in \mathbb{N}$ and any real-valued coordinatewise increasing functions $G,\, H\colon \mathbb{R}^k \longrightarrow \mathbb{R}$,*

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

whenever this covariance exists.

A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to be associated if, for every $k \in \mathbb{N}$, the random variables $X_1,\dots,X_k$ are associated.

Given the sequence $\{X_n\}_{n\in\mathbb{N}}$ denote by $\mathbf{F}_p$ its p-dimensional marginal distribution function, $\mathbf{X}_{i,p} = (X_{i+1},\dots,X_{i+p})$ and $\mathbf{x} = (x_1,\dots,x_p)$. Consider $\mathbf{U}$ a given p-variate distribution function, and $\{h_n\}_{n\in\mathbb{N}}$, a sequence of positive real numbers such that $h_n \longrightarrow 0$. Assuming the stationarity, the estimator for $\mathbf{F}_p$ is defined as

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

We will refer to $\mathbf{U}$ as the kernel function. This is the natural extension of the kernel estimator for distribution functions.

For independent samples, Jin and Shao (1999) proved the almost sure consistency and described the mean square error of $\widehat{F}_{n,p}$, 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 the 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 optimal bandwidth rate is shown to be depend on the decay rate of the covariances. In particular, in Cai and Roussas (1998), strengthening the assumptions on the covariances, the authors recover the optimal bandwidth rate $n^{-1/3}$, as for independent sequences. In Azevedo and Oliveira (2000), the two-dimensional estimation of the distribution function of (X_1, X_{k+1}) was considered with results similar to those by Cai and Roussas (1998).

For future reference we list the assumptions that will be used in the sequel. This set of conditions is basically the same as in Cai and Roussas (1998) and in Jin and Shao (1999).

Assumptions

(A_1) $\{X_n\}_{n\in\mathbb{N}}$ is a strictly stationary sequence of associated random variables with bounded density function f;

(A_2) The distribution function $\mathbf{F}_p$ of the random vector $\mathbf{X} = (X_1, \ldots, X_p)$ has bounded and continuous partial derivatives of first and second orders;

(A_3) For each $j \in \mathbb{N}$, the distribution function $\mathbf{F}_{p,j}$ of the $2p$-dimensional random vector $(\mathbf{X}_{1,p}, \mathbf{X}_{j,p})$ has bounded and continuous partial derivatives of first and second orders;

(A_4) 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

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

(A_5) $nh_n^2 \longrightarrow 0$;

(A_5^*) $nh_n^4 \longrightarrow 0$;

(A_6) $\displaystyle\sum_{n=1}^{\infty} n \operatorname{Cov}^{1/3}(X_1, X_n) < \infty$;

(A_6^*) There exists $\tau \in (0,1)$ such that $\displaystyle\sum_{j=1}^{\infty} \Big(\operatorname{Cov}(X_1, X_{j+1}) \Big)^{\frac{1-\tau}{3}} < \infty$;

(A_7) The function $\mathbf{V} = \frac{\partial^p \mathbf{U}^2}{\partial x_1 \ldots \partial x_p}$, satisfies $\displaystyle\int_{\mathbb{R}^p} \mathbf{x}\, \mathbf{x}^T\, \mathbf{V}(\mathbf{x})\, d\mathbf{x} < \infty$.

Remark 2 *Assumptions (A_1) and (A_6) have already been used in Cai and Roussas (1998) for the treatment of the univariate case. They state the regularity of the one-dimensional distribution function and a convenient decrease rate on the covariances. This assumption enables the control of pairs of random variables.*

Remark 3 *The strengthened assumptions (A_5^*) and (A_6^*) have also been used in Cai and Roussas (1998) in the one-dimensional case to obtain an optimal bandwidth characterization with the same rate as for independent sequences.*

Finally, denoting $\mathbf{x} = (x_1, \ldots, x_p)$ and $\mathbf{t} = (t_1, \ldots, t_p)$, we define the auxiliar real valued functions $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ and $\mathbf{V}_4$ on $\mathbb{R}^p$:

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

- $\displaystyle \mathbf{V}_2(\mathbf{x}) = \sum_{i=1}^{p} \frac{\partial \mathbf{F}_p}{\partial x_i}(\mathbf{x}) \int_{\mathbb{R}^p} t_i \mathbf{V}(\mathbf{t})\, d\mathbf{t}$;

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

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

2. Consistency of the Estimator

In this section we look at the almost sure consistency of the estimator $\widehat{F}_{n,p}$. This is be accomplished by applying a strong law of large numbers to the sequence of random variables $\mathbf{U}\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)$, $i = 1, \ldots, n-p$, that appear in the definition of $\widehat{F}_{n,p}$. To achieve this step we shall to characterize the asymptotic behavior of each entry of the covariance matrix of the random vector whose entries are the preceding variables. The almost sure consistency will then follow from the asymptotic unbiasedness of $\widehat{F}_{n,p}$.

Theorem 4 *Suppose $\{X_n\}_{n\in\mathbb{N}}$ satisfies (A_1). Then, for each $\mathbf{x} \in \mathbb{R}^p$,*

(i) $\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right) \longrightarrow \mathbf{F}_p(\mathbf{x})$;

(ii) if further (A_2) and (A_4) are satisfied, $\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right) = \mathbf{F}_p(\mathbf{x}) + \frac{V_1(\mathbf{x})}{2} h_n^2 + o(h_n^2)$.

Proof: (i) follows from an application of the Dominated Convergence Theorem. As for (ii), rewrite $\widehat{F}_{n,p}$ as

$$\widehat{F}_{n,p}(\mathbf{x}) = \int_{\mathbb{R}^p} \mathbf{U}\left(\frac{\mathbf{x}-\mathbf{s}}{h_n}\right) d\widehat{\phi_n}(\mathbf{s}), \tag{2}$$

where $\widehat{\phi_n}(\mathbf{x}) = \frac{1}{n-p}\sum_{i=1}^{n-p} \mathbb{1}_{(-\infty,x_1]\times\cdots\times(-\infty,x_p]}(\mathbf{X}_{i,p})$, and $\mathbb{1}_A$ is the characteristic function of the set A. It is easily verified that $\mathbb{E}\left(\widehat{\phi_n}(\mathbf{x})\right) = \mathbf{F}_p(\mathbf{x})$, so the result follows from (2), applying Fubini's Theorem, making a standard change of variable and using a Taylor expansion taking into account (A_2) and (A_4). ■

In order to prove the strong law of large numbers that we seek we need the following auxiliary result.

Lemma 5 (Lebowitz (1972)) *Let $A, B \subset \{1, \ldots, n\}$ and, for each $i \in A \cup B$, let $x_i \in \mathbb{R}$. Define $H_{A,B} = P(X_i > x_i, i \in A\cup B) - P(X_j > x_j, j \in A)P(X_k > x_k, k \in B)$. If the random variables $X_1, \ldots, X_n$ are associated then, $0 \le H_{A,B} \le \sum_{i\in A,\, j\in B} H_{\{i\},\{j\}}$.*

As mentioned before, the assumptions that we need to verify for the strong law of large numbers require a convenient control on the covariances of the terms that are summed in (1). For this purpose, define

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

and

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

Lemma 6 *Suppose that the sequence* $\{X_n\}_{n\in\mathbb{N}}$ *satisfies* (A_1), (A_2), (A_3) *and* (A_4). *Then, for each* $\mathbf{x} \in \mathbb{R}^p$,

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

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

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

Proof: To prove assertion (i) write

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

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 (A_3) to make the linear terms equal to zero and (A_4) to control the terms that multiply h_n^2, to find the term $\mathbf{F}_{pj}(\mathbf{x},\mathbf{x}) + O(h_n^2)$. This together with the characterization of the behavior of $\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right)$, as given in Theorem 4, completes the proof of (i).

To prove (ii), first use Lemma 5 to find

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

From (A_1) and Lemma 2.6 of Roussas (1995), there exits a constant $M > 0$ such that,

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

Inserting this into (3) and taking into account the stationarity of the random variables the conclusion follows. ■

Remark 7 *Note that if we assume, as was done in Cai and Roussas (1998), that the covariance sequence* $\{\mathrm{Cov}\,(X_1, X_{j+1})\}_{j\in\mathbb{N}}$ *is decreasing, then, under the same assumptions as in the previous lemma, the following upper bound holds* $\mathbf{I}_j(\mathbf{x}) \le p^2 \mathrm{Cov}^{1/3}(X_1, X_{j+1})$.

We may finally conclude the almost sure consistency of the estimator $\widehat{F}_{n,p}$.

Theorem 8 *Suppose the variables* $\{X_n\}_{n\in\mathbb{N}}$ *satisfy* (A_1), (A_2), (A_3), (A_4) *and* (A_6). *Then, for each* $\mathbf{x} \in \mathbb{R}^p$, $\widehat{F}_{n,p}(\mathbf{x}) \longrightarrow \mathbf{F}_p(\mathbf{x})$ *almost surely.*

Proof: As proved in Theorem 4, $\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right) \longrightarrow \mathbf{F}_p(\mathbf{x})$, so it is enough to verify that the variables $\mathbf{U}\left(\frac{\mathbf{x}-\mathbf{X}_{m,p}}{h_n}\right)$, $m \ge 1$, satisfy a strong law of large numbers. The stationarity of the variables is obvious. Further, as $\mathbf{U}$ is a distribution function, it is coordinatewise increasing, so these variables are also 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 strong law of large numbers. Now, it follows from Lemma 6 that

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

so (5) is a consequence of (A_6) and the association of the variables. ■

3. The Behavior of the Mean Square Error

In this section we study the asymptotics and convergence rate of the mean square error (MSE) of $\widehat{F}_{n,p}$. From the characterization obtained the optimal rate of the bandwidth is of order n^{-1}, thus a different convergence rate than the one for the independent case, as was already noticed for the one-dimensional case in Cai and Roussas (1998). The optimal rate for the bandwidth, when dealing with independent variables, is, for every dimension, of order $n^{-1/3}$, 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 $n^{-1/3}$ rate for the bandwidth.

As usual write MSE $\left(\widehat{F}_{n,p}(\mathbf{x})\right) = \text{Var}\left(\widehat{F}_{n,p}(\mathbf{x})\right) + \left(\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right) - \mathbf{F}_p(\mathbf{x})\right)^2$. The behavior of $\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right)$ being known (cf. Theorem 4), we need to describe the asymptotics and convergence rate for the variance term.

Lemma 9 *Suppose the sequence* $\{X_n\}_{n\in\mathbb{N}}$ *satisfies* (A_1), (A_2), (A_3), (A_4) *and* (A_7). *Then, for each* $\mathbf{x} \in \mathbb{R}^p$,

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

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

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

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

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 (A_4) to find the $o(h_n^2)$ term.

In order to verify (ii) decompose the variance in the standard way and apply (i) together with Theorem 4 to conclude the proof. ■

For a more convenient characterization of the mean square error of the estimator, let us introduce the following notation:

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

Theorem 10 *Suppose the sequence* $\{X_n\}_{n\in\mathbb{N}}$ *satisfies* (A_1), (A_2), (A_3), (A_4), (A_5), (A_6) *and* (A_7). *Then, for each* $\mathbf{x}\in\mathbb{R}^p$,

$$(n-p)\text{Var}\left(\widehat{F}_{n,p}(\mathbf{x})\right) = \sigma^2(\mathbf{x}) - h_n\mathbf{V}_2(\mathbf{x}) + (n-p-1)h_n^2\left(\mathbf{V}_4(\mathbf{x}) - \mathbf{F}_p(\mathbf{x})\mathbf{V}_1(\mathbf{x})\right) + O(h_n^2) - \mathbf{c}_n(\mathbf{x}).$$

Proof: Use the stationarity of the random variables and Lemmas 6 and 9 to write

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

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 to describe the behavior of the mean square error.

Theorem 11 *Suppose the sequence* $\{X_n\}_{n\in\mathbb{N}}$ *satisfies* (A_1), (A_2), (A_3), (A_4), (A_5), (A_6) *and* (A_7). *Then, for each* $x \in \mathbb{R}^p$,

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

Note that, with assumptions made, $c_n \longrightarrow 0$ and is independent of the bandwidth choice. So, to find the optimal bandwidth rate it is enough to minimize the $o(\cdot)$ term, which is achieved by choosing $h_n = O(n^{-1})$, for each dimension p. Tracing the coefficients of the above approximations, 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 instead (A_5^*), (A_6^*), and show that this reflects on the optimal bandwidth rate, recovering the same rate as in the independent case.

Theorem 12 *Suppose the covariances* $\mathrm{Cov}\,(X_1, X_{j+1})$ *decrease as* j *increases and the sequence* $\{X_n\}_{n\in\mathbb{N}}$ *satisfies* (A_1), (A_2), (A_3), (A_4), (A_5^*), (A_6^*) *and* (A_7). *Then, for each* $\mathbf{x} \in \mathbb{R}^p$,

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

Proof: Recall that, as shown in Lemma 6, $\mathbf{I}_{nj}(\mathbf{x})-\mathbf{I}_j(\mathbf{x}) = O(h_n^2)$, and, as noted in Remark 7, when the covariances are decreasing, we have $\mathbf{I}_j(\mathbf{x}) \leq p^2 \mathrm{Cov}^{1/3}(X_1, X_{j+1})$, and the same inequality holds for $\mathbf{I}_{nj}(\mathbf{x})$. It follows then that, there exists a constant $c > 0$ such that,

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

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

$$\begin{aligned}
&(n-p)\mathrm{Var}\left(\widehat{F}_{n,p}(\mathbf{x})\right) \\
&\quad = \mathrm{Var}\left(\mathbf{U}\left(\frac{\mathbf{x}-\mathbf{X}_{1,p}}{h_n}\right)\right) + \frac{2}{n-p}\sum_{j=2}^{n-p}(n-p-j+1)\left(\mathbf{I}_{nj}(\mathbf{x})-\mathbf{I}_j(\mathbf{x})\right) + \sum_{j=2}^{n-p}(n-p-j+1)\mathbf{I}_j(\mathbf{x}).
\end{aligned}$$

Using (A_6^*), we have that,

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

The result now follows readily repeating the arguments as in the proof of Theorem 10. ■

An optimization of this mean square error 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.

4. Finite Dimensional Distributions

We now look at the asymptotic behavior of the finite dimensional distributions of the estimator $\widehat{F}_{n,p}$. The method of proof is based on a decomposition of the sum that defines the estimator 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 to these independent copies. The distance between the original blocks and the coupling variables is controlled via Newman's inequality, cf. Newman (1984).

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

$$\alpha_n(\mathbf{x}) = \sqrt{n-p}\left(\widehat{F}_{n,p}(\mathbf{x}) - \mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right)\right),$$

$$\widetilde{\alpha}_n(\mathbf{x}) = \sqrt{n-p}\left(\widehat{F}_{n,p}(\mathbf{x}) - F_p(\mathbf{x})\right) = \alpha_n(\mathbf{x}) + \sqrt{n-p}\left(\mathbb{E}\left(\widehat{F}_{n,p}(\mathbf{x})\right) - F_p(\mathbf{x})\right).$$

The last term converges to zero using Theorem 4, so we need to concentrate our attention on $\alpha_n(\mathbf{x})$. Define further

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

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

Theorem 13 *Suppose $\{X_n\}_{n\in\mathbb{N}}$ satisfies (A_1), (A_2), (A_3), (A_4), (A_5), (A_6) and (A_7). Then, given $s\in\mathbb{N}$, $\mathbf{x_1},\ldots,\mathbf{x_s}\in\mathbb{R}^p$, 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*

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

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

Lemma 14 *Suppose the assumptions of Theorem 13 are satisfied. Then, for every $\mathbf{x},\mathbf{y}\in\mathbb{R}^p$,* $\mathrm{Cov}\,(\alpha_n(\mathbf{x}),\alpha_n(\mathbf{y})) \longrightarrow \varsigma^2(\mathbf{x},\mathbf{y})$.

Proof: Use the stationarity and argue as in Lemma 6. ■

Define now the block decomposition variables. Given an integer $r\le n-p$, let m be the largest integer less or equal to $(n-p)/r$. Denote

$$T_{n,i}(\mathbf{x}) = U\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right) - \mathbb{E}\left(U\left(\frac{\mathbf{x}-\mathbf{X}_{i,p}}{h_n}\right)\right),$$

$$Y_j^r(\mathbf{x}) = \frac{1}{\sqrt{r}}\sum_{i=(j-1)r+1}^{jr} T_{n,i}(\mathbf{x}), \qquad W_j^r = \sum_{q=1}^{s} c_q Y_j^r(\mathbf{x}_q),$$

and

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

The random variable Z_n is the linear combination of the coordinates of $(\alpha_n(\mathbf{x}_1),\ldots,\alpha_n(\mathbf{x}_s))$ required for the application of the Cramer-Wold Theorem. Define further

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

which replaces the sum up to $n-p$ by a sum with a multiple of r numbers of terms. Note also that, as follows from Lemma 14,

$$\mathrm{Var}\ (Z^*_{mr}) \longrightarrow \varsigma^2 := \sum_{q=1}^{s} c_q^2 \varsigma^2(\mathbf{x}_q,\mathbf{x}_q) + 2\sum_{q=1}^{s-1}\sum_{l=q+1}^{s} c_q c_l \varsigma^2(\mathbf{x}_q,\mathbf{x}_l). \tag{6}$$

Further, for each r fixed, it follows from Lemma 6 (i) that

$$\mathrm{Var}\ (Y_1^r(\mathbf{x})) = \frac{1}{r}\sum_{i,i'=1}^{r}\left(\mathbf{F}_{p,|i'-i+1|}(\mathbf{x},\mathbf{x}) - \mathbf{F}_p^2(\mathbf{x})\right) + O(rh_n^2), \tag{7}$$

and

$$\mathrm{Var}\ \left(W_j^r\right) = \sum_{q,q'=1}^{s} c_q c_{q'} \frac{1}{r}\sum_{i,i'=1}^{r}\left(\mathbf{F}_{p,|i'-i+1|}(\mathbf{x}_q,\mathbf{x}_{q'}) - \mathbf{F}_p(\mathbf{x}_q)\mathbf{F}_p(\mathbf{x}_{q'})\right) + O(rh_n^2). \tag{8}$$

We now proceed directly to the proof of Theorem 13. 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 15 *Suppose the assumptions of Theorem 13 are satisfied and let r be fixed. Then*

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

Proof: Use Hölder's inequality, the boundedness of $Z_{n,i}$, (6) and $mr/(n-p) \longrightarrow 1$. ■

We may now replace the sum Z_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_j^r, $j \geq 1$. Each of these variables is a linear combination of the Y_j^r which are decreasing functions of the original variables X_n, $n \geq 1$. So the Y_j^r are associated and we may apply a convenient variation of Newman's inequality, Newman (1984), to the variables W_j^r, $j \geq 1$ as proved in Lemma 4.1 from Jacob and Oliveira (1999), when coupling these variables with independent ones with the same distribution as each of the W_j^r.

Lemma 16 *Suppose the assumptions of Theorem 13 are satisfied and let r be fixed. Then*

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

The next step is a Central Limit Theorem for the coupling variables. In order to keep the notation simpler, we will denote these variables also by W_j^r. To describe the variances appearing on the next lemma let us define

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

Lemma 17 *Suppose the assumptions of Theorem 13 are satisfied and let r be fixed. Then*

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

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

Proof: **(of Theorem 13)** Let us define $a=\sum_{q,q'=1}^{s}c_q c_{q'}\varsigma^2(\mathbf{x}_q,\mathbf{x}_{q'})$. We have

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

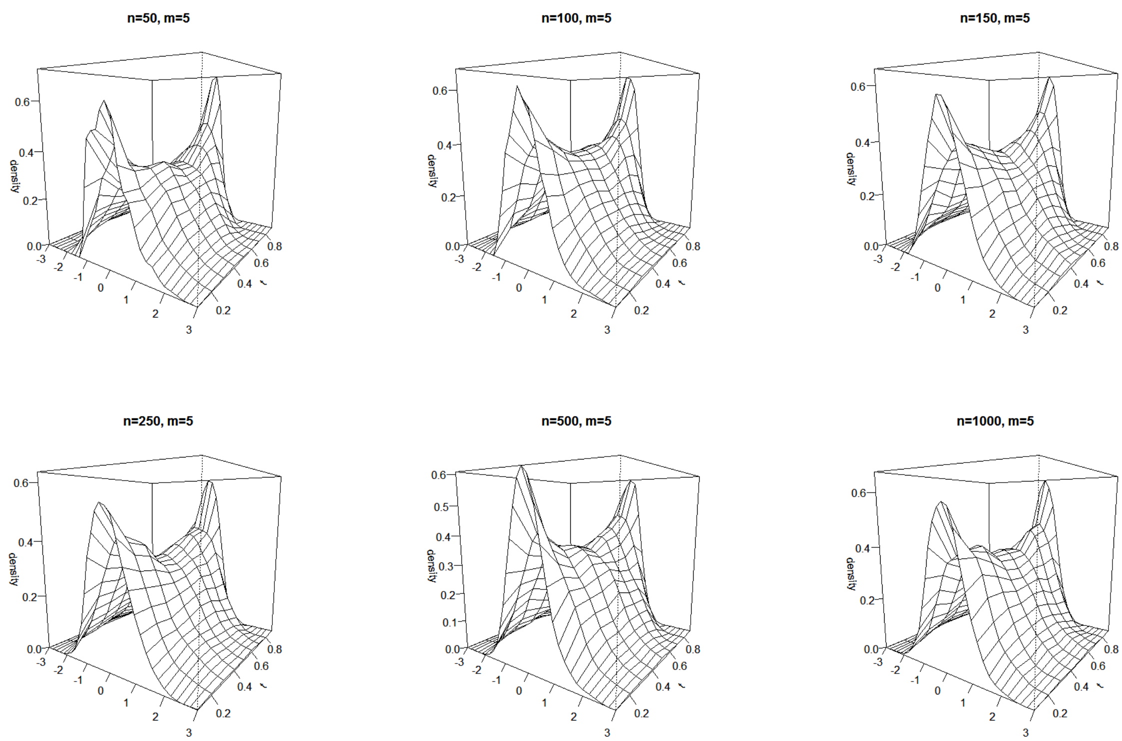
Supposing, for the moment, that r is fixed, it follows from the previous lemmas that

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

Now, if we let $r\longrightarrow+\infty$, it follows that this upper bound converges to zero on account of (7) and the stationarity of the variables X_n, $n\geq 1$, thus proving the theorem. ■

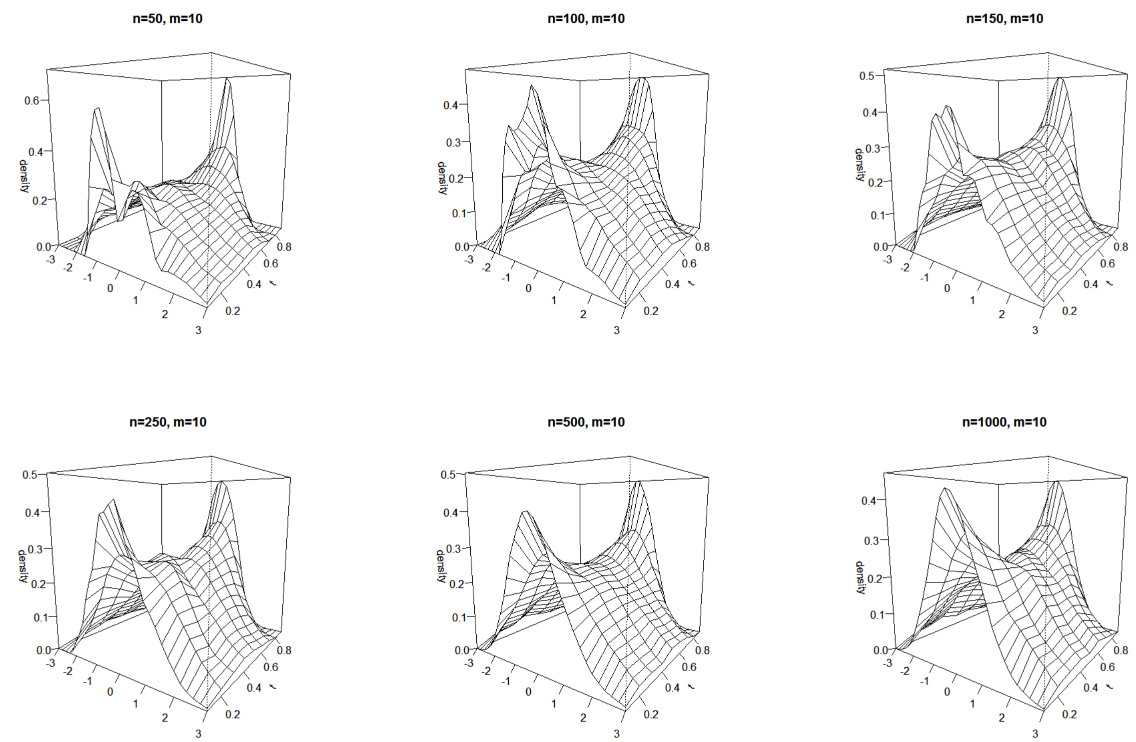
5. Simulation Results

We now 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 $m \in \mathbb{N}$ and 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 Y variables is easily chosen so that the X_i's are uniformly distributed in $[0, 1]$. Note that m may be interpreted as a measure of how far the variables are dependent. For each n and each m we simulated 1 000 paths for the empirical process and approximate, based on these paths, the density of $\widetilde{\alpha}_n(t_i)$ for a fixed set of points $t_1, \ldots, t_L \in [0, 1]$, using a kernel estimator. Note that Theorem 13 only proves the convergence of the finite dimensional distributions and not the functional convergence of the empirical process itself. Below we graph the approximations obtained for $n = 50$, 100, 150, 250, 500 and 1 000, and for $m = 5$:

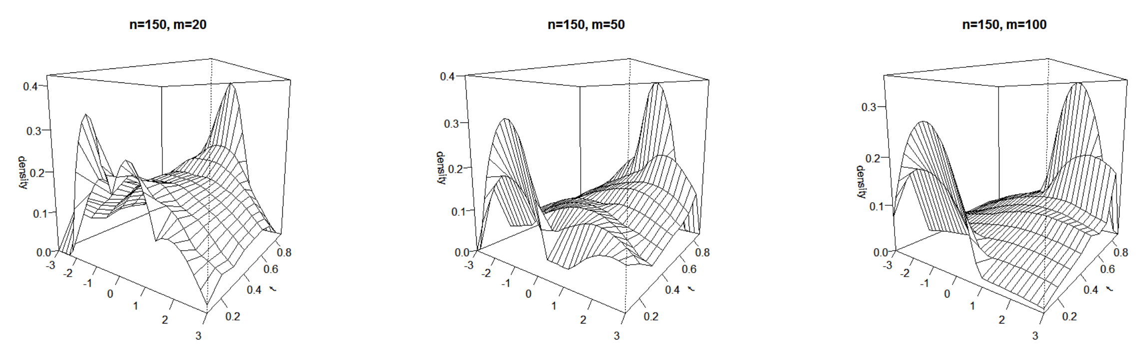


These graphs show a nice behaviour, close to gaussianity, but the base variables are "almost independent", so this is not very susprising.

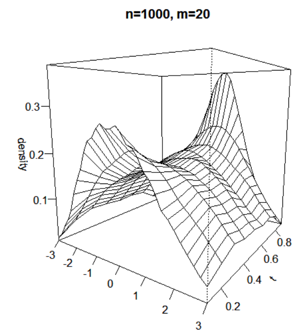
We now show the graphs, for the same values of n, increasing the degree of dependence between the variables, by considering $m = 10$. The convergence to a gaussian distribution is slower, as expected, but the approximations seem quite good for larger values of n:

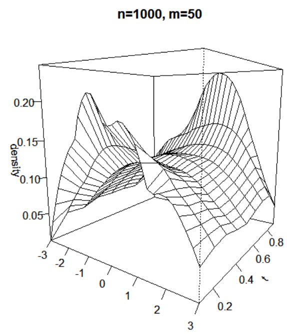


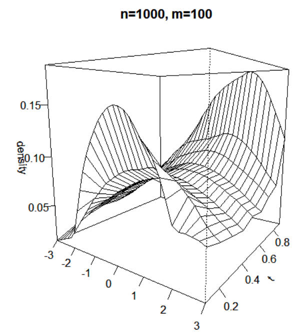
For a fixed number of values the effect of the degree of dependence is dramatic. We illustrate this for $n = 150$ and allowing m to take the values, 5, 10, shown above, and also 20, 50, 100.



These three graphs do not show much similarity with a gaussian distribution. Finally, we include the graphs for the same values of m as before, but with $n = 1\ 000$.







These confirm the impression from the previous case, although with a better behaviour for the smaller values of m. It is clear that, even with a very large number of points, the influence of m, measuring the degree of dependence, is by no means negligible.

Bibliography

Azevedo, C. and Oliveira, P. E. (2000). Kernel-type estimation of bivariate distribution function for associated random variables. In *New Trends in Probability and Statistics*, 17–25.

Cai Z. and Roussas, G. (1998). Efficient estimation of a distribution function under quadrant dependence. *Scand. J. Statist.*, **25**, 211–224.

Masry, E. (2002). Multivariate probability density estimation for associated processes: strong consistency and rates. *Statist. Probab. Lett.*, **58**, 205–219.

Henriques, C. and Oliveira, P. E. (2003). Estimation of a two-dimensional distribution function under association. *J. Statist. Plann. Inference*, **113**, 137–150.

Henriques, C. and Oliveira, P. E. (2005). Exponential rates for kernel density estimation under association. *Statist. Neerlandica* **59**, 448–466

Jacob, P. and Oliveira, P. E. (1999). Histograms and associated point processes. *Statist. Inf. Stoch. Proc.*, **2**, 227–251.

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

Lebowitz, J. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.*, **28**, 313–321.

Newman, C. M. (1980). Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.*, **74**, 119–128.

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

Oliveira P. E. and Suquet, C. (1996). An $L^2[0,1]$ invariance principle for LPQD random variables. *Portugal. Mathematica*, **53**, 367–379.

Roussas, G. (1991) Kernel estimates under association: strong uniform consistency. *Statist. Probab. Lett.*, **12**,393–403.

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

Roussas, G. (2000). Asymptotic normality of the kernel estimate of a probability density function under association. *Statist. Probab. Lett.*, **50**, 1–12.

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

Utev, S. A. (1990). On the central limit theorem for ϕ mixing arrays of random variables. *Theory Probab. Appl.*, **35**, 131–139.