A NOTE ON MINIMUM DISTANCE ESTIMATION OF COPULA DENSITIES

GÉRARD BIAU† AND MARTEN WEGKAMP‡

Abstract. This paper introduces a minimum L_1 distance estimate for parametric copula densities. It is shown that the expected L_1 error of the estimate is within a given constant multiple of the best possible error plus an additive remainder term which is small under mild assumptions. The proof is based on an oracle inequality and a maximal inequality for the empirical copula process indexed by sets.

1. Introduction

Any multivariate, cumulative distribution function H on $\mathbb{R}^p$ with *continuous* marginal distribution functions $F_1, \ldots, F_p$ can be written as

$$H(x_1, \ldots, x_p) = C\big(F_1(x_1), \ldots, F_p(x_p)\big), \quad -\infty \le x_1, \ldots, x_p \le +\infty\,, \tag{1}$$

for some function $C : [0,1]^p \to [0,1]$, which is called the *copula* or *dependence function* associated to H. This representation is unique on the range of $(F_1, \ldots, F_p)$ [cf. Sklar (1959)] and C is a distribution function in itself on $[0,1]^p$ with uniform marginals.

Copulas capture the dependence structure among the components X_j of the random vector $(X_1, \ldots, X_p)$ irrespectively of their marginal distributions F_j. In other words, they allow us to model the dependence structure and marginal distributions *separately.* Thus, by coupling different marginal distributions with different copula functions, copula-based models are able to model distributions that exhibit a wide variety of marginal and dependence

Date: December 7, 2004.
†Corresponding author. Email: biau@math.univ-montp2.fr . Institut de Mathématiques et de Modélisation de Montpellier – UMR CNRS 5149, Equipe de Probabilités et Statistique, Université Montpellier II, CC 051 – Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.
‡Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, United States of America.

properties such as skewness, fat tails, asymmetric dependence, and positive tail dependence, to name just a few. Because of this ability, the literature on copulas is growing at a fast pace, especially in the finance literature. We limit ourselves here to the recent surveys by Joe (1997) and Nelsen (1999).

In many applications, the copula function is described by one or more parameters, and a crucial step is the selection and the calibration of the model from the data at hand. For a comprehensive review of the classical methods to estimate the parameters of a determined copula (such as maximum likelihood or inference functions for marginals), we refer the reader to the comparative study of Romano (2002).

Recently, using ideas from Yatracos (1985), Devroye and Lugosi (2001) explore a new paradigm, called *minimum distance estimation*, for the data-based or automatic selection of the free parameters of density estimates in general so that the expected L_1 error is within a given constant multiple of the best possible error plus an additive remainder term which is small under mild assumptions. The purpose of the present paper is to extend the existing statistical theory on minimum distance estimation to the setting of parametric copula function selection. The score function of many copula densities is unbounded near the boundaries of $[0,1]^p$ [cf. Hu (1998), or Chen and Fan (2001)], what renders parameter estimation difficult. In addition, the maximum likelihood estimate does not minimize the L_1 distance or, equivalently, the total variation distance as noted by e.g. Devroye and Lugosi (2001, page 52). The minimum distance estimate, however, is specifically designed for achieving this.

The paper is organized as follows. In Section 2, we introduce the necessary notation and define the minimum distance copula density estimate. Section 3 proves a new maximal inequality for the empirical copula process indexed by sets. Finally, we state the main result in Section 4, which we illustrate with a few examples.

2. Minimum distance copula density estimation

In this section we introduce a new copula density estimate. Formally, a copula is a multivariate distribution function whose marginals are uniform on the interval $[0,1]$. For sake of simplicity, we will limit our exposition to the two-dimensional case. All results, however, can be easily extended to the multivariate case ($p \geq 2$).

Let (X, Y) be a bivariate random vector with distribution function $H(x, y)$ and *continuous* marginal distribution functions $F(x)$ and $G(y)$. Its associated copula function is defined, for any real numbers x and y, by [cf. (1) with $p = 2$]

$$H(x, y) = C\big(F(x), G(y)\big)\,.$$

Since F and G are continuous, the copula C defined above exists and is unique. This result is known as Sklar's theorem [Sklar (1959)], and we may write

$$C(u, v) = H\big(F^{-}(u), G^{-}(v)\big)\,, \quad 0 \leq u, v \leq 1\,,$$

where F^{-} is the *generalized quantile function* of F, that is,

$$F^{-}(u) = \inf\{x \in \mathbb{R} \ : \ F(x) \geq u\}\,, \quad 0 \leq u \leq 1\,,$$

and G^{-} is the generalized quantile function of G.

Based on independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), we construct the empirical distribution function

$$\mathbb{H}_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x, Y_i \leq y\}}\,, \quad -\infty \leq x, y \leq +\infty\,,$$

and let $\mathbb{F}_n(x)$ and $\mathbb{G}_n(y)$ be its associated marginal distributions, that is,

$$\mathbb{F}_n(x) = \mathbb{H}_n(x, +\infty) \text{ and } \mathbb{G}_n(y) = \mathbb{H}_n(+\infty, y)\,, \quad -\infty \leq x, y \leq +\infty\,.$$

The *empirical copula function* $\mathbb{C}_n$ is defined by

$$\mathbb{C}_n(u, v) = \mathbb{H}_n\big(\mathbb{F}_n^{-}(u), \mathbb{G}_n^{-}(v)\big)\,, \quad 0 \leq u, v \leq 1\,.$$

Ruymgaart (1973, pages 6–13) briefly discussed $\mathbb{C}_n$ in the introduction of his doctoral thesis. Deheuvels (1979) investigated the consistency of $\mathbb{C}_n$ as an estimate of C, and Deheuvels (1981a, b) proved a functional central limit theorem for the *empirical copula process*

$$\mathbb{Z}_n(u,v) = \sqrt{n}(\mathbb{C}_n - C)(u,v), \quad 0 \le u, v \le 1\,,$$

when X and Y are independent. Rüschendorf (1976, Theorem 3.3) and Gänssler and Stute (1987) proved weak convergence of the empirical copula process $\mathbb{Z}_n$ in the Skorokhod space $D\big([0,1]^2\big)$. Fermanian, Radulović, and Wegkamp (2003) established weak convergence in the space $\ell^\infty\big([0,1]^2\big)$ using the functional delta-method and they showed that the assumption that C has continuous partial derivatives is necessary and sufficient.

From now on, we assume that the copula function $C(u,v)$, which is a distribution in itself, has a density $c(u,v)$ with respect to the Lebesgue measure on the unit square $[0,1]^2$. The purpose of this study is to estimate c given a class $\mathcal{C}$ of candidate copula densities on $[0,1]^2$. We will use the minimum distance paradigm for the data-based selection of the free parameters of density estimates in general presented in Devroye and Lugosi (2001). For other references to this theory, see Hengartner and Wegkamp (2001), Biau and Devroye (2004a, 2004b), and Bunea and Wegkamp (2004).

We assume that the candidate copula densities in the class $\mathcal{C}$ are indexed by a parameter $\alpha \in \Xi$. We define a density estimate $\widehat{c}_n$ in $\mathcal{C}$ as follows. First introduce the class of sets

$$\mathcal{A} = \Big\{ \big\{ (u,v) \in [0,1]^2 \;:\; c_\alpha(u,v) > c_{\alpha'}(u,v) \big\} \;:\; (\alpha, \alpha') \in \Xi \times \Xi \Big\}$$

[$\mathcal{A}$ is the so-called *Yatracos class* associated with $\mathcal{C}$, following the terminology introduced by Devroye and Lugosi (2001)] and the goodness of fit criterion for a density $c_\alpha \in \mathcal{C}$:

$$\Delta_\alpha = \sup_{A \in \mathcal{A}} \left| \int_A c_\alpha(u,v)\,\mathrm{d}u\mathrm{d}v - \int_A \mathrm{d}\mathbb{C}_n(u,v) \right|.$$

Then the *minimum distance estimate* $\widehat{c}_n$ is defined as any density estimate selected among those densities $c_\alpha \in \mathcal{C}$ with

$$\Delta_\alpha < \inf_{\alpha^* \in \Xi} \Delta_{\alpha^*} + \frac{1}{n}\,. \tag{2}$$

The $1/n$ term here is added to ensure the existence of such a density estimate.

The interpretation of this estimate can be found in Devroye and Lugosi (2001). If $\mathbb{C}_n$ is replaced by C and $\mathcal{A}$ by all Borel sets in $[0,1]^2$, we would minimize the total variation distance between c and elements in $\mathcal{C}$. Since $\mathbb{C}_n$ is not absolutely continuous, we need to use a smaller collection of sets, which still gives a meaningful minimization. Using the same arguments as in Devroye and Lugosi (2001, Theorem 6.4), we easily find the following oracle inequality:

Lemma 2.1. *Let $\widehat{c}_n$ be the minimum distance estimate as defined above. Then, for all n, we have*

$$\int_{[0,1]^2} |\widehat{c}_n - c| \le 3 \inf_{\alpha \in \Xi} \int_{[0,1]^2} |c_\alpha - c| + 4 \sup_{A \in \mathcal{A}} \left| \int_A \mathrm{d}(\mathbb{C}_n - C) \right| + \frac{3}{n}\,. \tag{3}$$

The proof of Lemma 2.1 is left to the reader, since it follows the lines of the proof of Theorem 6.4 in Devroye and Lugosi (2001), with μ_n and μ replaced by $\mathbb{C}_n$ and C, respectively. The first term on the right of display (3) can be thought of as an approximation term, which decreases as the collection $\mathcal{C}$ expands, and we see that the minimum distance estimate tries to find the density in $\mathcal{C}$ which is closest to c in terms of the total variation distance (up to a factor 3). The second term can be regarded as an estimation term, which increases as $\mathcal{C}$ expands. Lemma 2.1 makes a crucial connection between the error of the selected density estimate and the supremum norm of the empirical copula process uniformly over a certain class of sets. We need to establish a Vapnik-Chervonenkis type of (maximal) inequality for the empirical copula process

$$\int_A \mathrm{d}\mathbb{Z}_n = \int_A \mathrm{d}\sqrt{n}(\mathbb{C}_n - C)\,,$$

indexed by *sets* A. As far as we know, this has not been addressed in the literature.

3. A maximal inequality for the empirical copula process indexed by sets

The purpose of this section is to prove a maximal inequality for the empirical copula process $\mathbb{Z}_n$ indexed by sets. Let us first recall that the *Vapnik-Chervonenkis (VC)-dimension* $V_{\mathcal{A}}$ of a given class of sets $\mathcal{A}$ [Vapnik and Chervonenkis (1971), Vapnik (1988)] is defined as the largest integer p such that

$$\mathcal{S}_{\mathcal{A}}(p) = 2^p,$$

where $\mathcal{S}_{\mathcal{A}}(p)$ is the *Vapnik-Chervonenkis shatter coefficient*, defined by

$$\mathcal{S}_{\mathcal{A}}(p) = \max_{x_1,\ldots,x_p \in \mathbb{R}^d} \mathrm{Card}\big\{\{x_1,\ldots,x_p\} \cap A \ : \ A \in \mathcal{A}\big\}.$$

If $\mathcal{S}_{\mathcal{A}}(p) = 2^p$ for all p, then we say that $V_{\mathcal{A}} = \infty$. Otherwise we say that $\mathcal{A}$ is a VC-class. Finally, we will denote by d the Euclidean distance on $[0,1]^2$, and, for each $A \in \mathcal{A}$, $\delta > 0$, we will use the notation

$$A^i_\delta = \{z \in A \ : \ d(z, A^c) \le \delta\} \ \text{and} \ A^e_\delta = \{z \in A^c \ : \ d(z, A) \le \delta\}\ ,$$

and

$$A_\delta = A^i_\delta \cup A^e_\delta\,.$$

Our basic assumptions are the following:

- (C.1) The distribution functions $F(x)$ and $G(y)$ are *continuous* and the copula function $C(u,v)$ admits a *bounded density* $c(u,v)$ with respect to the Lebesgue measure λ on the unit square.
- (C.2) The Yatracos class $\mathcal{A}$ is a VC-class of finite VC-dimension $V_{\mathcal{A}}$.
- (C.3) Let $\delta_n = \sqrt{\log(2n)/n}$. For each $A \in \mathcal{A}$, there exists a set $\overline{A}_{\delta_n}$ which contains the set A_{δ_n}. The collection of these majorizing sets, $\overline{\mathcal{A}}$, is a VC-class of finite VC-dimension $V_{\overline{\mathcal{A}}}$, independent of n.
- (C.4) Let $\delta_n = \sqrt{\log(2n)/n}$. There exists a sequence $\phi_n \downarrow 0$ such that

$$\sup_{\overline{A}_{\delta_n} \in \overline{\mathcal{A}}} \lambda(\overline{A}_{\delta_n}) \le \phi_n\,. \tag{4}$$

Let us briefly discuss these conditions. (C.1) requires little regularity from the distribution of (X, Y) and ensures that the copula function exists and is unique [cf. Sklar (1959)]. The last three conditions concern the Yatracos class $\mathcal{A}$ only. (C.2) states that the class $\mathcal{A}$ cannot be too large in a combinatorial sense. Roughly, (C.3) and (C.4) require that $\mathcal{A}$ is not too ill-behaved, in that the complexity of the collection of border tubes around sets in $\mathcal{A}$ remains under control in a combinatorial sense. Examples will be discussed at the end of Section 4. Our maximal inequality is as follows:

Theorem 3.1. *Let $\phi_n \downarrow 0$ be as defined in (4) of (C.4). Under assumptions (C.1)–(C.4), we have, for every n,*

$$\mathbb{E}\left\{ \sup_{A\in\mathcal{A}} \left| \int_A \mathrm{d}(\mathbb{C}_n - C) \right| \right\} \le K\left(\frac{\sqrt{V_{\mathcal{A}}} + \sqrt{V_{\overline{\mathcal{A}}}}}{\sqrt{n}} \right) + \|c\|_\infty \, \phi_n \,,$$

where K is a positive universal constant.

Proof. The proof is based on a series of lemmata. We first introduce some more notation. Define the pseudo-variables

$$(X^*, Y^*) = \big(F(X), G(Y)\big)\,,$$

with distribution function

$$H^*(u, v) = \mathbb{P}\{X^* \le u, Y^* \le v\} = H\big(F^-(u), G^-(v)\big) = C(u, v)\,.$$

Notice that the marginal distributions $F^*(u) = H^*(u, 1)$ and $G^*(v) = H^*(1, v)$ are uniform on $[0, 1]$. Finally, let $\mathbb{H}_n^*(u, v)$ be the empirical distribution function based on $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$ with marginal distributions $\mathbb{F}_n^*(u) = \mathbb{H}_n^*(u, 1)$ and $\mathbb{G}_n^*(v) = \mathbb{H}_n^*(1, v)$.

Lemma 3.2. *For every n, we have*

$$\begin{aligned}\sup_{A\in\mathcal{A}} \left| \int_A \mathrm{d}(\mathbb{C}_n - C) \right| &\le \sup_{A\in\mathcal{A}} \left| \int_A \mathrm{d}(\mathbb{H}_n^* - H^*) \right| \\ &\quad + \sup_{A\in\mathcal{A}} \left| \int_{\mathbb{R}^2} \left(\mathbf{1}_{\{(\mathbb{F}_n(x), \mathbb{G}_n(y)) \in A\}} - \mathbf{1}_{\{(F(x), G(y)) \in A\}} \right) \mathrm{d}\mathbb{H}_n(x, y) \right|.\end{aligned}$$

Proof. The claim follows directly from the definition of the copula function and adding and subtracting appropriate terms. Namely, observe that for every $A \in \mathcal{A}$, we have

$$\begin{aligned}
&\int_A \mathrm{d}(\mathbb{C}_n - C)(u,v) \\
&\quad = \int_A \mathrm{d}\Big\{\mathbb{H}_n\big(\mathbb{F}_n^-(u), \mathbb{G}_n^-(v)\big) - H\big(F^-(u), G^-(v)\big)\Big\} \\
&\quad = \int_A \mathrm{d}\Big\{\mathbb{H}_n\big(F^-(u), G^-(v)\big) - H\big(F^-(u), G^-(v)\big)\Big\} \\
&\qquad + \int_A \mathrm{d}\Big\{\mathbb{H}_n\big(\mathbb{F}_n^-(u), \mathbb{G}_n^-(v)\big) - \mathbb{H}_n\big(F^-(u), G^-(v)\big)\Big\} \\
&\quad = \int_A \mathrm{d}(\mathbb{H}_n^* - H^*)(u,v) \\
&\qquad + \int_{\mathbb{R}^2} \Big(\mathbf{1}_{\{(\mathbb{F}_n(x), \mathbb{G}_n(y)) \in A\}} - \mathbf{1}_{\{(F(x), G(y)) \in A\}}\Big)\, \mathrm{d}\mathbb{H}_n(x,y)\,,
\end{aligned}$$

which proves the lemma. □

Lemma 3.3. *For every n, we have*

$$\mathbb{E}\Big\{\sup_{A \in \mathcal{A}} \Big| \int_A \mathrm{d}(\mathbb{H}_n^* - H^*) \Big| \Big\} \le K \sqrt{\frac{V_{\mathcal{A}}}{n}}\,,$$

where K is a positive universal constant.

Proof. This well-known result is due to Dudley (1978). □

Define now the sequence $\delta_n = \sqrt{\log(2n)/n}$, and the sets

$$\Omega_n = \Big\{\sqrt{\|\mathbb{F}_n - F\|_\infty^2 + \|\mathbb{G}_n - G\|_\infty^2} \le \delta_n\Big\}\,.$$

Lemma 3.4. *For every n, we have*

$$\mathbb{P}\{\Omega_n\} \ge 1 - \frac{2}{n}\,.$$

Proof. The lemma follows from Massart (1990), who proved the inequality

$$\mathbb{P}\big\{\|\mathbb{F}_n - F\|_\infty > t\big\} \le 2e^{-2nt^2}$$

for any $t > 0$. □

Recall that d denotes the Euclidean distance on $[0,1]^2$ and, for each $A \in \mathcal{A}$, $\delta > 0$, the sets

$$A_\delta^i = \{z \in A \;:\; d(z, A^c) \le \delta\}\,, \quad A_\delta^e = \{z \in A^c \;:\; d(z, A) \le \delta\}$$

and $A_\delta = A_\delta^i \cup A_\delta^e$.

Lemma 3.5. *Let $z, \overline{z} \in [0,1]^2$ with $d(z,\overline{z}) \le \delta$. Then*

$$\left|\mathbf{1}_{\{z\in A\}} - \mathbf{1}_{\{\overline{z}\in A\}}\right| \le \mathbf{1}_{\{z\in A_\delta\}}.$$

Proof. We consider two cases:

- *Case i*: $z \in A, \overline{z} \notin A$. Suppose $z \notin A_\delta$. Then $d(z, A^c) > \delta$, and since $\overline{z} \in A^c$, we find that $d(z,\overline{z}) \ge d(z, A^c) > \delta$, which is a contradiction.
- *Case ii*: $z \notin A, \overline{z} \in A$. Suppose $z \notin A_\delta$. Then $d(z, A) > \delta$, and since $\overline{z} \in A$, we find that $d(z,\overline{z}) \ge d(z, A) > \delta$, which is a contradiction.

The proof of the lemma is complete. □

We are now in position to conclude the proof of Theorem 3.1. Observe that

$$\begin{aligned}
&\mathbb{E}\left\{\sup_{A\in\mathcal{A}}\left|\int_{\mathbb{R}^2}\left(\mathbf{1}_{\{(\mathbb{F}_n(x),\mathbb{G}_n(y))\in A\}} - \mathbf{1}_{\{(F(x),G(y))\in A\}}\right)\mathrm{d}\mathbb{H}_n(x,y)\right|\right\} \\
&\quad\le \mathbb{P}\{\Omega_n^c\} \\
&\qquad + \mathbb{E}\left\{\mathbf{1}_{\Omega_n}\sup_{A\in\mathcal{A}}\left|\int_{\mathbb{R}^2}\left(1_{\{(\mathbb{F}_n(x),\mathbb{G}_n(y))\in A\}} - \mathbf{1}_{\{(F(x),G(y))\in A\}}\right)\mathrm{d}\mathbb{H}_n(x,y)\right|\right\}
\end{aligned}$$

(since the integrand is bounded by 1).

The last term on the right can be bounded by

$$\begin{aligned}
&\mathbb{E}\left\{\mathbf{1}_{\Omega_n}\sup_{A\in\mathcal{A}}\left|\int_{\mathbb{R}^2}\left(\mathbf{1}_{\{(\mathbb{F}_n(x),\mathbb{G}_n(y))\in A\}} - \mathbf{1}_{\{(F(x),G(y))\in A\}}\right)\mathrm{d}\mathbb{H}_n(x,y)\right|\right\} \\
&\quad\le \mathbb{E}\left\{\mathbf{1}_{\Omega_n}\sup_{A\in\mathcal{A}}\int_{\mathbb{R}^2}\mathbf{1}_{\{(F(x),G(y))\in A_{\delta_n}\}}\,\mathrm{d}\mathbb{H}_n(x,y)\right\} \\
&\qquad\text{(by Lemma 3.5)} \\
&\quad\le \mathbb{E}\left\{\mathbf{1}_{\Omega_n}\sup_{\overline{A}_{\delta_n}\in\overline{\mathcal{A}}}\int_{\mathbb{R}^2}\mathbf{1}_{\{(F(x),G(y))\in\overline{A}_{\delta_n}\}}\,\mathrm{d}\mathbb{H}_n(x,y)\right\} \\
&\quad\le \mathbb{E}\left\{\sup_{\overline{A}_{\delta_n}\in\overline{\mathcal{A}}}\left|\int_{\mathbb{R}^2}\mathbf{1}_{\{(F(x),G(y))\in\overline{A}_{\delta_n}\}}\,\mathrm{d}(\mathbb{H}_n - H)(x,y)\right|\right\} \\
&\qquad + \mathbb{E}\left\{\sup_{\overline{A}_{\delta_n}\in\overline{\mathcal{A}}}\int_{\mathbb{R}^2}\mathbf{1}_{\{(F(x),G(y))\in\overline{A}_{\delta_n}\}}\,\mathrm{d}H(x,y)\right\} \\
&\quad= \mathbb{E}\left\{\sup_{\overline{A}_{\delta_n}\in\overline{\mathcal{A}}}\left|\int_{\overline{A}_{\delta_n}}\mathrm{d}(\mathbb{H}_n^* - H^*)(u,v)\right|\right\} + \mathbb{E}\left\{\sup_{\overline{A}_{\delta_n}\in\overline{\mathcal{A}}}\int_{\overline{A}_{\delta_n}}\mathrm{d}C(u,v)\right\} \\
&\quad\le K\sqrt{\frac{V_{\overline{\mathcal{A}}}}{n}} + \|c\|_\infty\,\phi_n
\end{aligned}$$

(by Lemma 3.3 and the very definition of ϕ_n).

This inequality combined with Lemma 3.2, 3.3, and 3.4 completes the proof of Theorem 3.1. □

4. Main result

Our main result is a direct corollary of the oracle inequality (Lemma 2.1) and the maximal inequality for the empirical copula process (Theorem 3.1).

Theorem 4.1. *Let $\phi_n \downarrow 0$ be as defined in (4) of (C.4) and let $\widehat{c}_n$ be the minimum distance estimate as defined in (2). Assume that (C.1)–(C.4) hold. Then, for all n, we have*

$$\begin{aligned}&\mathbb{E}\Big\{\int_{[0,1]^2} |\widehat{c}_n - c|\Big\}\\ &\quad \le 3\inf_{\alpha\in\Xi}\int_{[0,1]^2}|c_\alpha - c| + K_1\left(\frac{\sqrt{V_{\mathcal{A}}}+\sqrt{V_{\overline{\mathcal{A}}}}}{\sqrt{n}}\right) + \|c\|_\infty\,\phi_n\,,\end{aligned}$$

where K_1 is a positive universal constant.

Remark. The bound can be extended for situations with a general dimension $p \ge 2$. Note that the VC-dimensions $V_{\mathcal{A}}$ and $V_{\overline{\mathcal{A}}}$ might depend on p. Moreover, by taking $\delta_n = \sqrt{(p/2n)\log(np)}$, Lemma 3.4 has $1 - 2/n$ as a lower bound for the probability, but ϕ_n, through its dependence on δ_n, depends in a non-trivial way on p.

We illustrate the theory with examples of two standard copula families.

Example 1. Consider the classical bivariate *Gaussian copula* given by

$$C_\alpha(u,v) = \Phi_\alpha\big(\Phi^{-1}(u), \Phi^{-1}(v)\big), \quad 0 \le u, v \le 1\,, \tag{5}$$

where $\Phi(.)$ is the distribution function of a univariate normal random variable with mean 0 and variance 1, and $\Phi_\alpha(.,.)$ is the distribution function of a bivariate normal with means 0, variances 1, and correlation coefficient $\alpha \in (0,1)$. Denote by $\varphi(.)$ and $\varphi_\alpha(.,.)$ the densities of $\Phi(.)$ and $\Phi_\alpha(.,.)$, respectively. Then the density c_α of the Gaussian copula function C_α reads, for $(u,v) \in [0,1]^2$,

$$
\begin{aligned}
&c_\alpha(u,v) \\
&= \frac{\varphi_\alpha\big(\Phi^{-1}(u), \Phi^{-1}(v)\big)}{\varphi\big(\Phi^{-1}(u)\big)\,\varphi\big(\Phi^{-1}(v)\big)} \\
&= \frac{1}{\sqrt{1-\alpha^2}} \exp\Big(\frac{\alpha}{1-\alpha^2}\,\Phi^{-1}(u)\Phi^{-1}(v) - \frac{\alpha^2}{2(1-\alpha^2)}\big([\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2\big)\Big).
\end{aligned}
$$

A little calculation shows that every set in the Yatracos class may be written in the general form

$$\big\{(u,v) \in [0,1]^2 \ : \ [\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2 + a\Phi^{-1}(u)\Phi^{-1}(v) > b\big\}$$

or

$$\big\{(u,v) \in [0,1]^2 \ : \ [\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2 + a\Phi^{-1}(u)\Phi^{-1}(v) < b\big\},$$

where a is a *non-zero* real number and b is a *positive* constant. This class has finite VC-dimension, see, e.g., van der Vaart and Wellner (1996, Chapter 2). Consequently, assumption (C.2) holds. The validation of assumptions (C.3) and (C.4) needs a bit more care. To this aim, consider a general set in the Yatracos class, say A, and assume, without loss of generality, that $A = \big\{(u,v) \in [0,1]^2 \ : \ [\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2 + a\Phi^{-1}(u)\Phi^{-1}(v) > b\big\}$. In view of the property $\Phi^{-1}(u) = -\Phi^{-1}(1-u)$, the set A is symmetric with respect to the center $0 = (1/2, 1/2)$ of the unit square. As we increase the parameter a, the set A takes the form of the complement in $[0,1]^2$ of a connected ellipse-shaped manifold (Figure 1, left) whose opposite sides separate as a crosses the value 2 (Figure 1, right).

Clearly, the boundary of A is a smooth curve (symmetric with respect to the center of the square) with arc length certainly less than 4 (the perimeter of the square, see Figure 1 that depicts a typical example). Hence, for each $\delta > 0$, the set A_δ is contained in the union of two tubes of diameter 2δ, with Lebesgue measure not exceeding 4δ. Finally, the family of tubes has finite VC-dimension independent of δ and not exceeding two times the VC-dimension of $\mathcal{A}$. Assumptions (C.3) and (C.4) follow easily by taking the majorizing sequence ϕ_n in (4) as $\phi_n = 8\sqrt{\log(2n)/n}$.

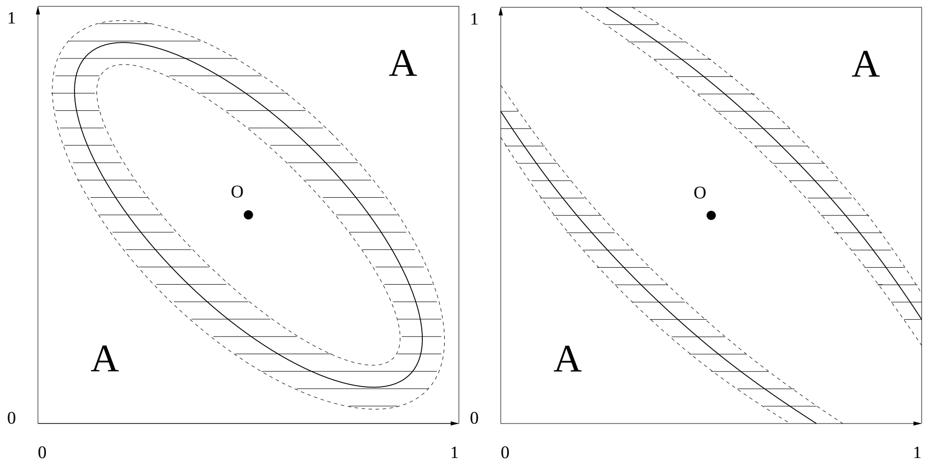

FIGURE 1. Example of Yatracos set A with $a = 1.5$ and $b = 1$ (left) and $a = 2.5$ and $b = 1$ (right). The set A is symmetric with respect to the center $0 = (1/2, 1/2)$ of the unit square. A typical tube-set is dashed.

Clearly, definition (5) of Gaussian copula functions readily extends to the multivariate setting ($p \geq 2$). As observed by a referee, a challenging question is then to determine how the VC-dimension of corresponding tubes depends on p. We also suspect that $\phi_n \sim V_{p-1}\delta_n$, where V_{p-1} is the volume of the unit sphere $\mathcal{S}_{p-1}$ (for the supremum norm) in $\mathbb{R}^p$. Here it is not clear how the curse of dimensionality kicks in.

Example 2. As a second example, consider the Farlie-Gumbel-Morgenstern copula

$$C_\alpha(u,v) = uv\big[1 + \alpha(1-u)(1-v)\big], \quad 0 \leq u, v \leq 1\,,$$

with density

$$c_\alpha(u,v) = 1 + \alpha(1-2u)(1-2v), \quad |\alpha| \leq 1.$$

Here $\mathcal{A}$ consists of two sets only:

$$\big\{(u,v)\ :\ 0 \leq u < 1/2,\ 0 \leq v < 1/2\big\} \cup \big\{(u,v)\ :\ 1/2 < u \leq 1,\ 1/2 < v \leq 1\big\}$$

and its complement in $[0,1]^2$. Assumptions (C.1)-(C.4) hold straightforwardly with $\phi_n = 6\sqrt{\log(2n)/n}$.

Acknowledgments. The authors wish to thank an anonymous referee for valuable comments.

References

[1] Biau, G. and Devroye, L. (2004a). A note on density model size testing, *IEEE Transactions on Information Theory*, **50**, 576-581.

[2] Biau, G. and Devroye, L. (2004b). Density estimation by the penalized combinatorial method, *Journal of Multivariate Analysis*, in press.

[3] Bunea, F. and Wegkamp, M. (2003). A note on penalized minimum distance estimation in nonparametric regression, *The Canadian Journal of Statistics*, **31**, 267-274.

[4] Chen, X. and Fan, Y. (2002). Estimation of copula-based semiparametric time series models, *Technical Report*, Vanderbilt University.

[5] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance, *Bulletin de l'Académie Royale de Belgique, Classe des Sciences*, **65**, 274-292.

[6] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*, Springer–Verlag, New York.

[7] Dudley, R. M. (1978). Central limit theorems for empirical measures, *The Annals of Probability*, **6**, 899-929.

[8] Fermanian, J. D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes, *Bernoulli*, **10**, 847-860.

[9] Gänssler, P. and Stute, W. (1987). *Seminar on Empirical Processes*, DMV Seminar, Band 9, Birkhäuser.

[10] Hengartner, N. and Wegkamp, M. (2001). Estimation and selection procedures in regression: an L_1 approach, *The Canadian Journal of Statistics*, **29**, 621-632.

[11] Hu, L. (1998). *Large Sample Theory for Pseudo-Maximum Likelihood Estimates in Semiparametric Models*, PhD Thesis, University of Washington.

[12] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

[13] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, *The Annals of Probability*, **18**, 1269-1283.

[14] Nelsen, R. B. (1999). *An Introduction to Copulas*, Lecture Notes in Statistics, **139**, Springer–Verlag, New York.

[15] Romano, C. (2002). Calibrating and simulating copula functions: An application to the Italian Stock Market, *Working paper n. 12/2002*, CIDEM, Università degli Studi di Roma "La Sapienza".

[16] Rüschendorf, L. (1976). Asymptotic normality of multivariate rank order statistics, *The Annals of Statistics*, **4**, 912-923.

[17] Ruymgaart, F. H. (1973). *Asymptotic Theory for Rank Tests for Independence*, MC-tract 43, Mathematisch Instituut, Amsterdam.

[18] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229-231.

[19] Yatracos, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov's entropy, *The Annals of Statistics*, **13**, 768-774.

[20] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes – With Applications to Statistics*, Springer–Verlag, New York.

[21] Vapnik, V. N. and Chervonenkis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, **16**, 264-280.

[22] Vapnik, V. N. (1998). *Statistical Learning Theory*, John Wiley & Sons, New York.