Strong approximation of empirical copula processes by Gaussian processes

Salim Bouzebda, Tarek Zari

Une approximation forte du processus empirique de la copule par une suite de processus gaussiens

Salim BOUZEBDA and Tarek ZARI

L.S.T.A., Université Pierre et Marie Curie-Paris 6. 175, rue du Chevaleret, 8ème étage, bâtiment A,75013 PARIS FRANCE.

Abstract. This paper investigates the problem of strong approximation of the empirical copula processes for arbitrary dimension, with continuous unknown margins. The idea of the proof is based on the results obtained in Deheuvels and al.(2006), and the theorem of strong approximation for an arbitrary distribution function proved in Csorgo and 1988. Using these results, we derive the normality for smoothed empirical copula processes and the L.I.L. for empirical copula processes.
Keywords : Copulas ; strong approximation ; Kiefer process.
Résumé. Nous établissons un résultat d'approximation forte pour le processus empirique multivarié de la copule, sous réserve que les marges soient continues et inconnues. Nous ferons usage d'un résultat de Csörgő et Horváth (1988), ainsi que de quelques techniques développées par Deheuvels et al., (2006). Grâce à ce résultat, nous obtenons une loi de logarithme iterée de ce processus, ainsi qu'une approximation forte du processus empirique de la copule lissée.
Mots clés : Copules ; principe d'invariance ; L.I.L ; le processus de Kiefer.

1 Notations and Definitions

Let $\mathbf{X}_i = (X_{1i}, \ldots, X_{di}), i = 1, 2, \ldots$, be an independent replications of a d-dimensional random vector $\mathbf{X} \in \mathbb{R}^d$ with distribution function [df] $\mathbf{F}(x_1, \ldots, x_d) := \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$. Throughout this paper, we assume that the corresponding marginal df's $F_i(x_i) := \mathbb{P}(X_i \leq x_i)$, for $i = 1, \ldots, d$ of $\mathbf{X}_i$ are continuous. The quantile functions pertaining to $F_i(\cdot)$ for $i = 1, \ldots, d$ are denoted, respectively, by $F_i^-(u_i) = \inf\{x_i : F_i(x_i) \geq u_i\}$, for $0 < u_i < 1$ and $i = 1, \ldots, d$. The copula function (see, e.g., Sklar (1959) and Deheuvels (1979, 1980, 1981b) among others) of $\mathbf{F}(\cdot)$ is defined as the distribution function

$$\mathbf{C}(u_1, \ldots, u_d) = P(F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d), \quad \text{for every} \quad \mathbf{u} \in [0, 1]^d$$

of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. For some historic notes, we refer to Schweizer (1991) and the recent survey conducted by Joe (1997) and Nelsen (2006). It's known that the copula associated with $\mathbf{F}(\cdot)$ is uniquely determined and is given by

$$\mathbf{C}(u_1, \ldots, u_d) := \mathbf{F}(F_1^-(u_1), \ldots, F_d^-(u_d)), \quad \text{for every} \quad \mathbf{u} \in [0, 1]^d. \tag{1}$$

Let's $\mathbb{1}_A(\cdot)$ be the indicator function of the set A. We define, for each $n \geq 1$, the empirical counterparts of $\mathbf{F}(\cdot)$, $F_1(\cdot), \ldots, F_d(\cdot)$ and $F_1^-(\cdot), \ldots, F_d^-(\cdot)$, respectively, by setting, for $j = 1, \ldots, d$,

$$\mathbf{F}_n(x_1, \ldots, x_d) := \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbb{1}\{X_{ji} \leq x_j\}, \qquad \text{for} \quad \mathbf{x} \in \mathbb{R}^d,$$

$$F_{jn}(x_j) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_{ij} \le x_j\} = \mathbf{F}_n(1,\ldots,x_j,\ldots,1), \quad \text{for } x_j \in \mathbb{R},$$

and

$$F_{in}^{-}(u_i) = \inf\{t_i \in \mathbb{R} : F_{in}(t_i) \ge u_i\}, \quad \text{for } u_i \in [0,1].$$

In order to characterize the copula function given in (1), we define the empirical copula function of $\mathbf{F}_n(\cdot)$ (or, equivalently, associated with $\mathbf{X}_1,\ldots,\mathbf{X}_n$), as any copula $\mathbf{C}_n(\cdot)$, through the following identity

$$\mathbf{C}_n(u_1,\ldots,u_d) := \mathbf{F}_n(F_{1n}^{-}(u_1),\ldots,F_{dn}^{-}(u_d)), \quad \text{for } \mathbf{u} \in [0,1]^d.$$

Set $U_{ji} := F_j(X_{ji})$ for $i = 1,\ldots,n,\ j = 1,\ldots,d$ and $\mathbf{U}_i := (U_{1i},\ldots,U_{di})$. Thus, the random vectors $\mathbf{U}_i$ are distributed according to the copula $\mathbf{C}(\cdot)$. For each $n \ge 1$, $0 \le u_j \le 1$ and $1 \le j \le d$, set

$$\widetilde{\mathbf{C}}_n(u_1,\ldots,u_d) := \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} \mathbb{1}\{U_{ji} \le u_j\}, \quad \text{and} \quad G_{jn}(u_j) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{U_{ij} \le u_j\}.$$

Consider the empirical processes defined, respectively, for $n \ge 1$, $\mathbf{u} \in [0,1]^d$ and $0 \le u_j \le 1$, for $j = 1,\ldots,d$, by

$$\begin{aligned}
\alpha_n(\mathbf{u}) &:= n^{1/2}(\widetilde{\mathbf{C}}_n(\mathbf{u}) - \mathbf{C}(\mathbf{u})), &(2)\\
\alpha_n^{j}(u_j) &:= n^{1/2}\{G_{jn}(u_j) - u_j\}, &(3)\\
\beta_{jn}(u_j) &:= n^{1/2}\{G_{jn}^{-}(u_j) - u_j\}. &(4)
\end{aligned}$$

Then, with $X \stackrel{d}{=} Y$ which means that X and Y have the same distribution, we have

$$\mathbf{F}_n(x_1,\ldots,x_d) \stackrel{d}{=} \widetilde{\mathbf{C}}_n(F_1(x_1),\ldots,F_d(x_d))$$

and

$$(F_{1n}(x_1),\ldots,F_{dn}(x_d)) \stackrel{d}{=} (G_{1n}(F_1(x_1)),\ldots,G_{dn}(F_d(x_d))).$$

From the definition of $F_{in}^{-}(\cdot)$, it follows that

$$G_{in}^{-}(u_i) \stackrel{d}{=} F_i\{F_{in}^{-}(u_i)\},$$

and we also have

$$\mathbf{C}_n(u_1,\ldots,u_d) \stackrel{d}{=} \widetilde{\mathbf{C}}_n(G_{1n}^{-}(u_1),\ldots,G_{dn}^{-}(u_d)).$$

Since the copula function associated with $\mathbf{F}_n(\cdot)$ is not unique, it is convenient to investigate the *modified empirical copula process*, which is, in turn, uniquely defined by

$$\gamma_n(\mathbf{u}) = n^{1/2}(\widetilde{\mathbf{C}}_n(G_{1n}^{-}(u_1),\ldots,G_{dn}^{-}(u_d)) - \mathbf{C}(u_1,\ldots,u_d)), \text{ for } \mathbf{u} \in [0,1]^d. \qquad (5)$$

If the margins are known, Neuhaus (1971) and Bickel and Wichura (1971) showed

$$\gamma_n(\mathbf{u}) \stackrel{d}{\rightarrow} \mathbb{B}_{\mathbf{C}}(\mathbf{u}) \qquad (6)$$

for $J_1^{(d)}$-topology on D_d as $n \to \infty$, where $B_{\mathbf{C}}$ is a continuous Gaussian process with

$$\mathbb{E}(B_{\mathbf{C}}(\mathbf{u})) = 0, \quad \mathbb{E}(B_{\mathbf{C}}(\mathbf{u})B_{\mathbf{C}}(\mathbf{v})) = \mathbf{C}(\mathbf{u} \wedge \mathbf{v}) - \mathbf{C}(\mathbf{u})\mathbf{C}(\mathbf{v})$$

where $\mathbf{u} \wedge \mathbf{v} = (\min(u_1, v_1), \ldots, \min(u_d, v_d))$. For the precise definition of the space D_d and the $J_1^{(d)}$-topology, and for more multivariate empirical processes, we refer to Einmahl (1987). Fermanian, Radulovic and Wegkamp (2004) show that the weak convergence of $\gamma_n(\cdot)$ to $\mathbb{B}^*_{\mathbf{C}}(\cdot)$, in the bivariate case, holds on $[0,1]^2$ when $\mathbf{C}(\cdot)$ has continuous partial derivatives on $[0,1]^2$, where

$$\mathbb{B}^*_{\mathbf{C}}(\mathbf{u}) = B_{\mathbf{C}}(\mathbf{u}) - B_{\mathbf{C}}(u_1, 1)\frac{\partial \mathbf{C}(\mathbf{u})}{\partial u_1} - B_{\mathbf{C}}(1, u_2)\frac{\partial \mathbf{C}(\mathbf{u})}{\partial u_2}, \quad \mathbf{u} \in [0,1]^2.$$

Deheuvels (1979, 1980, 1981a) described the limiting behavior $\gamma_n(\cdot)$ in the case of independence of margins. The empirical copula process has been studied in full generality in Gaenssler and Stute (1987). Recently, Deheuvels (2008) derive optimal rates for strong approximation of empirical copula processes by sequences of Gaussian processes. In the last reference, a full characterization of empirical copulas is provided. We can refer to Deheuvels (2008) and the references therein concerning the invariance principles for copulas.

In the present paper, we are concerned with strong approximations of the empirical copula processes $\{\gamma_n(\mathbf{u}) : \mathbf{u} \in [0,1]^d\}$, based upon $\mathbf{X}_1, \ldots, \mathbf{X}_n$, by a single Gaussian process. Denote by $\{\mathbb{K}_{\mathbf{C}}(\mathbf{u}, t) : \mathbf{u} \in [0,1]^d, t \geq 0\}$ a multivariate Gaussian process. This process has continuous sample paths and fulfills

$$\mathbb{E}\left(\mathbb{K}_{\mathbf{C}}(\mathbf{u}, s)\right) = 0 \quad \text{and} \quad \mathbb{E}\left(\mathbb{K}_{\mathbf{C}}(\mathbf{u}, s)\mathbb{K}_{\mathbf{C}}(\mathbf{v}, t)\right) = (s \wedge t)\left\{\mathbf{C}(\mathbf{u} \wedge \mathbf{v}) - \mathbf{C}(\mathbf{u})\mathbf{C}(\mathbf{v})\right\},$$

for $\mathbf{u}, \mathbf{v} \in [0,1]^d$ and $s,\ t \geq 0$. Note that the process $\{\mathbb{K}_{\mathbf{C}}(\mathbf{u}, t) : \mathbf{u} \in [0,1]^d, t \geq 0\}$ is known in the literature under the name of Kiefer process. For each $n > 0$, $u_j \in [0,1]$ and $j = 1, \ldots, d$, the copula Gaussian process is defined by

$$\begin{aligned}
\mathbb{K}^*_{\mathbf{C}}(\mathbf{u}, n) &= \mathbb{K}_{\mathbf{C}}(\mathbf{u}, n) - \sum_{j=1}^{d} \mathbb{K}_{\mathbf{C}}(1, \ldots, 1, u_j, 1, \ldots, 1, n)\frac{\partial \mathbf{C}(\mathbf{u})}{\partial u_j} \\
&=: \mathbb{K}_{\mathbf{C}}(\mathbf{u}, n) - \sum_{j=1}^{d} \mathbb{K}^{(j)}_{\mathbf{C}}(\mathbf{1}, u_j, \mathbf{1}, n)\frac{\partial \mathbf{C}(\mathbf{u})}{\partial u_j}, \quad \mathbf{u} \in [0,1]^d. \qquad (7)
\end{aligned}$$

The remaining of our paper is organized as follows. In the next section we will give our main result concerning the strong approximation of empirical copula processes by a Gaussian process, which is stated in Theorem 2.1 below. In section 3 we will give some applications of Theorem 2.1, more precisely we will give the limit law of smoothed empirical copula and the law of iterated logarithm for the empirical copula processes.

2 Results

We may now state our strong approximation theorem as follows.

Theorem 2.1 *Assume that $\mathbf{C}(\cdot)$, associated with $\mathbf{F}(\cdot)$, is twice continuously differentiable on $(0,1)^d$, and the second derivative be continuous on $[0,1]^d$. On a suitable probability space, we may define the empirical copula processes $\{\gamma_n(\mathbf{u}) : \mathbf{u} \in [0,1]^d; n > 0\}$ in combination with a Gaussian process $\{\mathbb{K}^*_{\mathbf{C}}(\mathbf{u}, t) : \mathbf{u} \in [0,1]^d; t \geq 0\}$, in such a way that, almost surely as $n \to \infty$*

$$\sup_{\mathbf{u} \in [0,1]^d} |\sqrt{n}\gamma_n(\mathbf{u}) - \mathbb{K}^*_{\mathbf{C}}(\mathbf{u}, n)| = O\left(n^{1/2 - 1/(4d)}(\log n)^{3/2}\right), \qquad (8)$$

where $\mathbb{K}^*_{\mathbf{C}}(\mathbf{u}, t)$ is defined in (7).

Remark 1 In the particular case of independence, i.e., $\mathbf{C}(\mathbf{u}) = \prod_{i=1}^d u_i$ *the process* $\{\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},n) : \mathbf{u} \in [0,1]^d; n \geq 0\}$ *is equal to*

$$\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},n) =: \mathbb{K}_{\mathbf{C}}(\mathbf{u},n) - \sum_{j=1}^d \mathbb{K}^{(j)}_{\mathbf{C}}(\mathbf{1},u_j,\mathbf{1},n)\prod_{i\neq j}^d u_i, \qquad \mathbf{u} \in [0,1]^d,$$

with mean zero and covariance functions

$$\mathbb{E}\left(\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},s)\mathbb{K}^*_{\mathbf{C}}(\mathbf{v},t)\right) = (s\wedge t)\left\{\prod_{i=1}^d (u_i\wedge v_i) + (d-1)\prod_{i=1}^d u_iv_i - \sum_{i=1}^d (u_i\wedge v_i)\prod_{i\neq j} u_iv_j\right\}$$

where $\mathbf{u},\mathbf{v} \in [0,1]^d$ *and* $s,\ t \geq 0$*. For more details the reader may refer to Csörgő (1979).*

3 Applications

3.1 Smoothed empirical copula processes

The smoothed empirical copula function $\widehat{\mathbf{C}}_n(\cdot)$ is defined by

$$\widehat{\mathbf{C}}_n(\mathbf{u}) = \frac{1}{h}\int_{[0,1]^d} k\left(\frac{\mathbf{u}-\mathbf{v}}{h^{1/d}}\right)\mathbf{C}_n(\mathbf{v})d\mathbf{v} \quad \text{for} \quad \mathbf{u} \in [0,1]^d, \tag{9}$$

where $k(\cdot)$ is a kernel function and $h = h(n)$ is the smoothing parameter. For notational convenience, we have chosen the same bandwidth sequence for each margins. This assumption can be dropped easily. Similarly to the previous section, we define the smoothed empirical copulas process by

$$\widehat{\gamma}_n(\mathbf{u}) := \sqrt{n}(\widehat{\mathbf{C}}_n(\mathbf{u}) - \mathbf{C}(\mathbf{u})) \quad \text{for} \quad \mathbf{u} \in [0,1]^d. \tag{10}$$

We will describe the asymptotic properties of the smoothed empirical copulas process $\widehat{\gamma}_n(\cdot)$ under the following conditions.

(F.1). Assume that the copula function $\mathbf{C}(\cdot)$ has a bounded sth derivative.

Suppose that $\{h(n)\}_{n\geq 1}$ is a sequence of positive constants satisfies the condition.

(C.1). $h = h(n) \to 0$, $nh \to \infty$ and $\sqrt{n}h^{s/d} \to 0$ as $n \to \infty$;

and the kernel function $K(\cdot)$ fulfills the following conditions.

(C.2). $k(\cdot)$ is a continuous density function and compactly supported;

(C.3). $k(\cdot)$ is of order s.

The following corollary establishes the limiting behaviour of the smoothed empirical copulas process $\widehat{\gamma}_n(\cdot)$.

Corollary 3.1 *Assume that (F.1) and (C.1)-(C.3) hold. Then, on a suitable probability space, we may define the empirical copula processes* $\{\widehat{\gamma}_n(\mathbf{u}) : \mathbf{u} \in [0,1]^d; n > 0\}$ *in combination with a Gaussian process* $\{\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},t) : \mathbf{u} \in [0,1]^d; t \geq 0\}$*, in such a way that, as* $n \to \infty$

$$\sup_{\mathbf{u}\in[0,1]^d} |\widehat{\gamma}_n(\mathbf{u}) - \frac{1}{\sqrt{n}}\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},n)| = o_P(1). \tag{11}$$

Remark 2 1. *Corollary 3.1 remains valid when replacing the condition that the kernel function $k(\cdot)$ having compact support in (C.2) by*

(C.4). There exists a sequence of positive real numbers a_n such that $a_n h$ tends to zero when n tends to infinity, and

$$\sqrt{n}\int_{\{\|\mathbf{v}\|>a_n\}} |k(\mathbf{v})|d\mathbf{v} \to 0.$$

2. *Note that the conditions of Corollary 3.1 are grouped to control the deviations between the* modified empirical copula process $\gamma_n(\cdot)$ *and the smoothed empirical copula process* $\widehat{\gamma}_n(\cdot)$.

3.2 The law of iterated logarithm for empirical copula processes

From Theorem 2.1, we have almost surely

$$\limsup_{n\to\infty}\left\{\left(\frac{n}{2\log\log n}\right)^{1/2}\sup_{\mathbf{u}\in[0,1]^d}|\mathbf{C}_n(\mathbf{u})-\mathbf{C}(\mathbf{u})|\right\} = \limsup_{n\to\infty}\frac{\sup_{\mathbf{u}\in[0,1]^d}|\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},n)|}{(2n\log\log n)^{1/2}}. \tag{12}$$

According to Wichura (1973), we have almost surely, as $n\to\infty$,

$$\limsup_{n\to\infty}\frac{\sup_{\mathbf{u}\in[0,1]^d}|\mathbb{K}^*_{\mathbf{C}}(\mathbf{u},n)|}{(2n\log\log n)^{1/2}} \leq d+1. \tag{13}$$

Combining (12) and (13), we can state the following Corollary.

Corollary 3.2 *Under the same conditions of the Theorem 2.1, we have*

$$\limsup_{n\to\infty}\left\{\left(\frac{n}{2\log\log n}\right)^{1/2}\sup_{\mathbf{u}\in[0,1]^d}|\mathbf{C}_n(\mathbf{u})-\mathbf{C}(\mathbf{u})|\right\} \leq d+1. \tag{14}$$

Remark 3 The result of Corollary 3.2 was obtained by Deheuvels (1979) (refer to Theorem 3.1) using a different method.

Références

Bickel, P. and Wichura, M. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Stat.*, **42**, 1656–1670.

Csörgő, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process. *J. Multivariate Anal.*, **9**(1), 84–100.

Csörgő, M. and Horváth, L. (1988). A note on strong approximations of multivariate empirical processes. *Stochastic Process. Appl.*, **28**(1), 101–109.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, **65**(6), 274–292.

Deheuvels, P. (1980). Nonparametric test of independence. In *Nonparametric asymptotic statistics (Proc. Conf., Rouen, 1979) (French)*, volume 821 of *Lecture Notes in Math.*, pages 95–107. Springer, Berlin.

Deheuvels, P. (1981a). A Kolmogorov-Smirnov type test for independence and multivariate samples. *Rev. Roumaine Math. Pures Appl.*, **26**(2), 213–226.

Deheuvels, P. (1981b). Multivariate tests of independence. In *Analytical methods in probability theory (Oberwolfach, 1980)*, volume 861 of *Lecture Notes in Math.*, pages 42–50. Springer, Berlin.

Deheuvels, P. (2008). A Multivariate Bahadur-Kiefer Representation for the Empirical Copula Process. *Preprint*.

Deheuvels, P., Peccati, G., and Yor, M. (2006). On quadratic functionals of the Brownian sheet and related processes. *Stochastic Process. Appl.*, **116**(3), 493–538.

Einmahl, J. H. J. (1987). *Multivariate empirical processes*, volume 32 of *CWI Tract*. Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica, Amsterdam.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**(5), 847–860.

Gaenssler, P. and Stute, W. (1987). *Seminar on empirical processes. (Notes based on lectures given at the Seminar on Empirical Processes held at Schloß Mickeln, Düsseldorf, FRG, September 8-13, 1985).* DMV Seminar, Bd. 9. Basel - Boston : Birkhäuser Verlag. VI, 110 p.; DM 38.00 .

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Stat.*, **42**, 1285–1295.

Schweizer, B. (1991). Thirty years of copulas. In *Advances in probability distributions with given marginals (Rome, 1990)*, volume 67 of *Math. Appl.*, pages 13–50. Kluwer Acad. Publ., Dordrecht.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

Wichura, M. J. (1973). Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments. *Ann. Probability*, **1**, 272–296.