Journal of Multivariate Analysis 137 (2015) 179–186

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Maximal coupling of empirical copulas for discrete vectors

Olivier P. Faugeras

Toulouse School of Economics - Université Toulouse 1 Capitole - GREMAQ, Manufacture des Tabacs, Bureau MF319, 21 Allée de Brienne, 31000 Toulouse, France

ARTICLE INFO

Article history:
Received 30 June 2014
Available online 24 February 2015

AMS 2010 subject classifications:
60E05
62H05
62H12
62E20

Keywords:
Discrete vector
Maximal coupling
a.s. constructions
Empirical copula
Ergodicity

ABSTRACT

For a vector **X** with a purely discrete multivariate distribution, we give simple short proofs of uniform a.s. convergence on their whole domain of two versions of genuine empirical copula functions, obtained either via probabilistic continuation, i.e. kernel smoothing, or via the distributional transform. These results give a positive answer to some delicate issues related to the convergence of copula functions in the discrete case. They are obtained under the very weak hypothesis of ergodicity of the sample, a framework which encompasses most types of serial dependence encountered in practice. Moreover, they allow to derive, as simple corollaries, almost sure consistency results for some recent extensions of concordance measures attached to discrete vectors. The proofs are based on a maximal coupling construction of the empirical cdf, a result of independent interest.

1. Introduction

Let **X** be a d-variate real-valued random vector with cdf F, and corresponding vector of marginal cdfs $\mathbf{G} = (G_1, \dots, G_d)$, namely $G_i(x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$. We denote vectors by bold letters, and interpret operations between vectors componentwise. Let $\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$ be the l_1 norm on $\mathbb{R}^d$, and $\|\cdot\|_\infty$ the supremum norm on $\mathbb{R}$.

1.1. Copula functions

Recall that a d-dimensional copula function $C : [0, 1]^d \mapsto [0, 1]$ is defined analytically as a grounded, d-increasing function, with uniform marginals whose domain is $[0, 1]^d$ (see Nelsen [14]). Alternatively, it can be defined probabilistically as the restriction to $[0, 1]^d$ of the multivariate cdf of a random vector **U**, called a *copula representer*, whose marginals are uniformly distributed on [0, 1] (see Rüschendorf [18,19]). Their interest stems from Sklar's theorem (see [21,22]), which asserts that, for every random vector $\mathbf{X} \sim F$, there exists a copula function *connecting, or associated with* **X**, in the sense that:

Theorem 1.1. *For every multivariate cdf F, with marginal cdfs* **G**, *there exists some copula function C such that*

$$F(\mathbf{x}) = C(\mathbf{G}(\mathbf{x})), \quad \forall \mathbf{x} \in \mathbb{R}^d, \tag{1}$$

where $\mathbf{G}(\mathbf{x}) = (G_1(x_1), \dots, G_d(x_d))$. Conversely, if C is a copula function and **G** *a set of marginal distribution functions, then the function F defined by (1) is a joint distribution function with marginals* **G**.

E-mail address: olivier.faugeras@tse-fr.eu.

http://dx.doi.org/10.1016/j.jmva.2015.02.013

When $\mathbf{G}$ is continuous, the copula C associated with $\mathbf{X}$ in relation (1) is unique and can be defined from F either analytically by $C = F \circ \mathbf{G}^{-1}$, where $\mathbf{G}^{-1} = (G_1^{-1}, \dots, G_d^{-1})$ is the vector of marginal quantile functions, or probabilistically as the cdf of the multivariate Probability Integral Transform, namely $C(\mathbf{u}) = P(\mathbf{G}(\mathbf{X}) \leq \mathbf{u}), \mathbf{u} \in [0, 1]^d$. Whenever discontinuity is present, C is no longer unique: in other words C, as a functional parameter, is not identifiable from F alone. In such a case, probabilistic constructions of a copula representer $\mathbf{U}$ associated with $\mathbf{X}$ can be based on:

(i) the d-variate distributional transform $\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$ where $G_j(x_j, \lambda) = P(X_j < x_j) + \lambda P(X_j = x_j), j = 1, \dots, d, \lambda \in [0, 1]$, and $\mathbf{V}$ is a vector of uniform [0, 1] marginals, independent of $\mathbf{X}$ (see Moore and Spruill [13], Rüschendorf [17–19], Neslehova [15]);
(ii) probabilistic continuation, i.e. by taking the limit of $\mathbf{U}_h = \hat{\mathbf{G}}_h(\mathbf{X}_h)$ in distribution along a subsequence, where $\hat{\mathbf{G}}_h$ is the vector of marginal cdf of the continued $\mathbf{X}_h = \mathbf{X} + h\mathbf{Z}$, where $\mathbf{Z}$ is continuous and $h \downarrow 0$ (see Faugeras [6]).

1.2. Empirical copulas for continuous distributions

If F is unknown, but one has instead a sample $\mathbf{X}_1, \mathbf{X}_2, \dots$ of copies distributed according to F on a probability space $(\Omega, \mathcal{A}, P)$, one can define the ecdf F_n,

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\mathbf{X_i} \leq \mathbf{x}},$$

and the corresponding vector of marginal ecdfs $\mathbf{G}_n$. Sklar's theorem therefore entails that there exists some copula function C_n associated with F_n. As the ecdf is discrete, C_n is no longer unique and can no longer be defined, in parallel with the continuous case, as $C_n^* := F_n \circ \mathbf{G}_n^{-1}$, or as $C_n^{**}(\mathbf{u}) := P^*(\mathbf{G}_n(\mathbf{X}_n^*) \leq \mathbf{u})$, with $\mathbf{X}_n^* \sim F_n$, conditionally on the sample, and where P^* is the corresponding probability (more on this below). Indeed, C_n^* and C_n^{**} do not have uniform marginals and hence are not genuine copula functions associated with F_n. C_n^* and C_n^{**} are versions of the improperly called empirical "copula" functions, introduced by Rüschendorf [16] under the name of multivariate rank order function and Deheuvels [2,3] under the name of empirical dependence function.

When F is continuous, the disadvantage of estimating $C = F \circ \mathbf{G}^{-1}$ by estimators which are not proper, in the sense that they do not belong to the same class of the parameter to be estimated, is mitigated by the fact that these estimators coincide, with any copula function associated with F_n on the grid of points $\mathbf{u}_k = (k_1/n, \dots, k_d/n)$ for $k_1, \dots, k_d = 0, \dots, n$; see Deheuvels [3]. Moreover, any version of the corresponding empirical "copula" process weakly converges, see e.g. Fermanian et al. [7], or Rüschendorf [16]. Hence, in the continuous case, the choice of which "empirical copula" function to use is often of little relevance for statistical purposes.

1.3. Empirical copulas for discrete distributions

To the contrary, when F has a discrete component, the indeterminacy of C and hence of C_n is more acute: Deheuvels [1] then Sempi [20] and Lindner and Szimayer [9] show that if C_n and C are copula functions associated with F_n, F, with respective marginals $\mathbf{G}_n, \mathbf{G}$, then, $F_n \xrightarrow{d} F$ if and only if

(i) the margins weakly converge: $F_{n,j} \xrightarrow{d} F_j, j = 1, \dots, d$;
(ii) and C_n converges uniformly to C on $\overline{Ran\mathbf{G}}$.

Quoting Lindner and Szimayer, "since in [the discrete] case, the copula of $\mathbf{X}$ does not need to be unique, convergence of the copulas on $[0, 1]^d$ cannot be expected" and they present a counter-example. See also Neslehova [15,11] for a similar discussion.

The question thus arises as to what can be expected when one applies the previous schemes mentioned in Section 1.1 in order to obtain some specific genuine empirical copula functions associated with the sample. More precisely, conditionally on the sample, on an extra probability space $(\Omega^*, \mathcal{A}^*, P^*)$,

(i) let $\mathbf{X}_n^* \sim F_n$. Set $\mathbf{U}_n := \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$ the multivariate distributional transforms for the ecdf F_n, with randomization $\mathbf{V}$. Denote as C_n^1 the cdf of $\mathbf{U}_n$, i.e. the copula function associated with F_n;
(ii) let $\mathbf{X}_n^* \sim F_n$ as before and $\mathbf{Z} \sim K$ a continuous vector, independent of $(\mathbf{X}_1^*, \mathbf{X}_2^*, \dots)$. Set

$$\mathbf{Y}_n := \mathbf{X}_n^* + h_n\mathbf{Z}, \quad h_n \downarrow 0,$$

the smoothing of $\mathbf{X}_n^*$. Denote as $\hat{F}_n, \hat{\mathbf{G}}_n$ the corresponding (continuous) joint and marginal cdfs. Note that $\hat{F}_n$ is the Parzen–Rosenblatt kernel-smoothed empirical cdf,

$$P^*(\mathbf{Y}_n \leq \mathbf{x}) = \int P^*\left(\mathbf{Z} \leq \frac{\mathbf{x} - \mathbf{y}}{h_n}\right) dF_n(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right).$$

Define $\hat{\mathbf{U}}_n := \hat{\mathbf{G}}_n(\mathbf{Y}_n)$, the copula representer associated with $\hat{F}_n$, obtained via this probability integral transform. Denote as C_n^2 the cdf of $\hat{\mathbf{U}}_n$, i.e. the copula function associated with $\hat{F}_n$.

Faugeras' [6] Theorems 3.1 and 4.1 explore the latter scheme: he shows that the sequence of copula representers $\hat{\mathbf{U}}_{\mathbf{n}}$, built from the empirical cdf by probabilistic continuation, converges to *some* corresponding copula representer $\hat{\mathbf{U}}$ associated with the original $\mathbf{X}$. When $\mathbf{X}$ has a discrete component, Theorem 4.1 of [6] gives, via a double asymptotic argument, the uniform a.s. convergence of the (kernel smoothed) empirical copula function C_n^2 on the whole $[0, 1]^d$, as a simple corollary.

At first sight, a reader might be tempted to think that such a uniform convergence result of copula functions, in the discrete case, on the whole $[0, 1]^d$ and not solely on $\overline{Ran\mathbf{G}}$, is in contradiction with the assertions of Lindner, Szimayer and Deheuvels. However, this is not so, and the paradox comes from the fact that the copula functions in Deheuvels and Lindner and Szimayer's result are left unspecified, which leaves some space for counterexamples. To the contrary, the indeterminacy on the copula functions is raised in Theorem 4.1 of Faugeras [6], by the specific construction of $\hat{\mathbf{U}}_{\mathbf{n}}$, $\hat{\mathbf{U}}$ involved there.

1.4. Outline

The present article greatly improves and completes these results, by investigating the asymptotic behavior of the aforementioned empirical copulas C_n^1 and C_n^2, when the original distribution F of $\mathbf{X}$ is purely discrete. More precisely, we present in Section 2 a maximal coupling construction, inspired by Thorisson [24], which is applied to the empirical cdf of a purely discrete distribution. This preliminary key result, which is of independent interest, is obtained under the very weak assumption of ergodicity of the sample $\mathbf{X}_1, \mathbf{X}_2, \ldots$. This framework encompasses most types of serial dependence encountered in practice for which statistical estimation remains a sensible question (e.g. when the sample is an irreducible aperiodic positive recurrent Markov chain which possesses an invariant probability distribution, when it is a stationary ergodic dynamical system, and, of course, the i.i.d. case).

Regarding empirical copulas, the interest of this construction is twofold: for the empirical copula function C_n^2 derived by probabilistic continuation, it allows to obtain a simple proof of uniform a.s. consistency on $[0, 1]^d$ of C_n^2, which bypasses the double asymptotic framework involved in Theorem 4.1 of [6]. More importantly, for the empirical copula function C_n^1 obtained via the distributional transform, we also obtain the new result of a.s. convergence of the copula representers $\mathbf{U}_n \stackrel{a.s.}{\to} \mathbf{U}$, and the corresponding uniform a.s. convergence on $[0, 1]^d$ of C_n^1. These results are contrasted with those of Genest et al. [8] and Rüschendorf [18]. Eventually, we show, in Section 4, how these results allow to obtain as simple corollaries consistency of some estimators of extensions of concordance measures for discrete vectors, such as Neslehova's [15] and Mesfioui and Quessy's [12] extensions of Kendall's tau.

2. Maximal coupling of the discrete empirical cdf

2.1. Coupling of discrete random vectors

Coupling is a powerful probabilistic method which allows to turn distributional properties into their corresponding counterparts in terms of random variables. See e.g. Lindvall [10] and Thorisson [24] for some excellent treatises on the subject. As in Thorisson [24, chapter 1], recall that if $\{\mathbf{X}_i, i \in \mathbb{I}\}$ be a collection of random vectors on $\mathbb{R}^d$, where $\mathbb{I}$ is an index set, then a *coupling* of $\{\mathbf{X}_i, i \in \mathbb{I}\}$ is a family of random vectors $(\hat{\mathbf{X}}_i, i \in \mathbb{I})$ defined on the *same* probability space, such that $\mathbf{X}_i \stackrel{d}{=} \hat{\mathbf{X}}_i$, for all $i \in \mathbb{I}$. $\hat{\mathbf{X}}_i$ is called a *representer* or a *copy* of $\mathbf{X}_i$. Recall also that a *coupling event* A is an event such that if A occurs, then all $\hat{\mathbf{X}}_i$ coincide. Theorem 4.1 in chapter 1 of [24] gives an upper bound on the probability of a coupling event (see also inequalities (2.6) and (2.8) in chapter 1 of [10]). Theorem 4.2 in chapter 1 of [24] or Theorem 5.2 in chapter 1 of [10] shows that the upper bound can be attained for a special construction, called a *maximal coupling* in the terminology of Thorisson, or a γ-coupling in the terminology of Lindvall [10].

As asserted above, it is well known that the distributional property of weak convergence of measures can be treated in a probabilistic manner by coupling methods: Skorokhod–Dudley–Wichura's theorem (see Skorokhod [23], Dudley [5]), asserts that, if $\mathbf{X}_n \sim F_n, \mathbf{X} \sim F, n \in \mathbb{N}$, weak convergence of F_n towards F, $F_n \stackrel{d}{\to} F$, is equivalent to the existence of a coupling $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \ldots, \hat{\mathbf{X}})$ of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}$, such that $\hat{\mathbf{X}}_n \stackrel{a.s.}{\to} \hat{\mathbf{X}}$.

In the case the random vectors involved are all purely discrete and take their values on a common countable set E, Thorisson [24] somehow strengthens this result: pointwise convergence of probability mass functions of random vectors can be turned into pointwise convergence of representers which eventually hit the limit, via a maximal coupling construction.

Theorem 2.1 (*Thorisson Theorem 1.6.1*). *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\infty$ be a sequence of discrete random vectors with values on a common countable set E. Then,*

$$\lim_{n\to\infty} P(\mathbf{X_n} = \mathbf{x}) = P(\mathbf{X}_\infty = \mathbf{x}), \quad \textit{for all } \mathbf{x} \in E,$$

holds if and only if there exists a coupling $(\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_\infty)$ of $\mathbf{X}_1, \ldots, \mathbf{X}_\infty$ and a finite random integer $\mathcal{N}$ s.t.

$$\hat{\mathbf{X}}_n = \hat{\mathbf{X}}_\infty, \quad n \geq \mathcal{N}.$$

Proof. See Appendix, where Thorisson's proof is reproduced for convenience. He proved the result for real valued random-variables but it extends effortlessly to random vectors. □

Remark 1. This result can be thought of as a stochastic analog of the fact that if (u_n) is a deterministic sequence with values in, say, $\mathbb{Z}$, then (u_n) converges iff (u_n) is eventually stationary in the sense that there exists $p \in \mathbb{N}$, s.t. $\forall n \geq p$, $u_{n+1} = u_n$.

Remark 2. Note that pointwise convergence of probability mass functions is obviously stronger than weak convergence: for $X_n = 1/n$, $n \in \mathbb{N}$ and $X_\infty = 0$, clearly $X_n \xrightarrow{d} X_\infty$, whereas $P(X_n = 0) = 0$, for all $n \in \mathbb{N}$, and $P(X_\infty = 0) = 1$. Therefore, Theorem 2.1 is not, strictly speaking, a strengthening of Skorokhod's theorem. See Thorisson, [24, chapter 1] for a discussion.

2.2. Maximal coupling of the empirical cdf for discrete vectors

Let μ be a multivariate atomic probability measure, i.e. charging a countable set $E = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ of values, with cdf F. Denote its probability mass function by $p(\mathbf{x})$: $p(\mathbf{x}) > 0$, $\forall \mathbf{x} \in E$ and $\sum_{\mathbf{x}\in E} p(\mathbf{x}) = 1$. Assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ an ergodic sample of μ, in the sense that, $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ take their values in E, and, for all $\mathbf{x} \in E$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\mathbf{X_i}=\mathbf{x}} \xrightarrow{a.s.} p(\mathbf{x}), \quad \text{as } n \to \infty. \tag{2}$$

Denote F_n the empirical cdf and p_n its corresponding empirical probability mass function. One then has the following key proposition:

Proposition 2.2. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ be an ergodic sample of purely discrete random vectors distributed as F, s.t. (2) is satisfied. Then, for almost every realization $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ there exists some probability space s.t. one can define jointly on it*

(i) *a sequence $\hat{\mathbf{X}}_n^* \sim F_n$,*
(ii) *$\hat{\mathbf{X}}^* \sim F$,*
(iii) *a finite random integer $\mathcal{N}$,*

s.t. $\hat{\mathbf{X}}_n^* = \hat{\mathbf{X}}^*$, for $n \geq \mathcal{N}$.

Proof. Assume the sample $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ is defined on some probability space $(\Omega, \mathcal{A}, P)$. For example, take $(\Omega, \mathcal{A})$ the infinite countable product $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}) = (S \times S \times \cdots, \mathcal{S} \times \mathcal{S} \times \cdots)$ where $S = \mathbb{R}^d$ and $\mathcal{S} = \mathcal{B}(S)$ the corresponding Borel σ-field and let $\mathbf{X}_i(\omega) : (S^{\mathbb{N}}, \mathbb{S}^{\mathbb{N}}) \mapsto (S, \mathbb{S})$, $i = 1, 2, \ldots,$ be the coordinate-projection mapping, defined for $\omega = (\omega_1, \omega_2, \ldots) \in S^{\mathbb{N}}$, by $\mathbf{X}_i(\omega) = \omega_i$.

Call $P_n^\omega(\cdot) = \frac{1}{n}\sum \delta_{\mathbf{X}_i(\omega)}(\cdot)$ the empirical measure based on the sample. Denote $p_n(\mathbf{x}) = P_n^\omega(\{\mathbf{x}\})$ the empirical probability mass function, (where we suppressed the dependence on ω). Assumption (2) means that there exists some set $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ s.t. for all $\omega \in \Omega_0$, for all $\mathbf{x} \in E$, $p_n(\mathbf{x}) \to p(\mathbf{x})$.

Fix $\omega \in \Omega_0$. On some extra probability space $(\Omega^*, \mathcal{A}^*, P^*)$, define a sequence $\mathbf{X}_n^* \overset{P^*}{\sim} F_n$ of random vectors distributed according to F_n and $\mathbf{X}^* \overset{P^*}{\sim} F$ distributed according to F. Note that the set of values of $\mathbf{X}_n^* \sim F_n$ is finite and included in E, but possibly not equal to E. However, it is not difficult to see that Theorem 2.1 remains valid in this setting: since the empirical measures (P_n^ω) are absolutely continuous w.r.t. to a single dominating measure μ, one can use Thorisson [24, Theorem 7.1 in chapter 1] (see also his theorem 9.2 in chapter 3) instead of his Theorem 1.6.1, with Radon–Nikodym derivatives instead of probability mass functions.

One can therefore apply the construction of Theorem 2.1: there exists a coupling $(\hat{\mathbf{X}}_1^*, \hat{\mathbf{X}}_2^*, \ldots, \hat{\mathbf{X}}^*)$ of $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}^*$ and a finite random integer $\mathcal{N}$ s.t.

$$\hat{\mathbf{X}}_n^* = \hat{\mathbf{X}}^*, \quad n \geq \mathcal{N}.$$

The construction is valid for every $\omega \in \Omega_0$, with $P(\Omega_0) = 1$, hence the assertion. $\square$

3. Uniform a.s. convergence of some genuine empirical copula functions for purely discrete data

Let $\mathbf{X} \in \mathbb{R}^d$ be a purely discrete-valued random vector with cdf F, marginal cdfs $\mathbf{G}$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an ergodic sample of $\mathbf{X}$ in the sense of (2) and $F_n, \mathbf{G}_n$ the empirical cdfs as before. The preliminary construction of Section 2 allows to obtain simple proofs of a.s. convergence of empirical copula associated with a sample from a discrete distribution.

3.1. Convergence of distributionally transformed empirical copula

The a.s. convergence of the distributionally transformed empirical copula representer is presented in the next theorem.

Theorem 3.1. *If F is purely discrete, then for almost every realization $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ there exists a probability space which carries copula representers $\hat{\mathbf{U}}_\mathbf{n} \sim C_n^1$, $\hat{\mathbf{U}} \sim C$, such that:*

(i) *C_n^1, C are copula functions associated with F_n, F, respectively;*
(ii) *$\hat{\mathbf{U}}_\mathbf{n} \xrightarrow{a.s.} \hat{\mathbf{U}}$.*

Proof. Let $\mathbf{X}^* \sim F$, $(\mathbf{X}_n^* \sim F_n)_{n\in\mathbb{N}}$, $\mathbf{V}$ be defined on the same product probability space with $\mathbf{V}$ a vector with uniform marginals, independent of $(\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}^*)$. Set $\mathbf{U} := \mathbf{G}(\mathbf{X}^*, \mathbf{V})$, $\mathbf{U}_n := \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$, the multivariate distributional transforms for the cdf F and the ecdf F_n, with the same randomization $\mathbf{V}$. By Rüschendorf's [18], $\mathbf{U}$, $\mathbf{U_n}$ are vectors with uniform marginals whose cdf C, C_n^1 are copula functions and satisfy Sklar's theorem for F, F_n, respectively.

By Proposition 2.2, there exists a coupling $(\hat{\mathbf{X}}_1^*, \hat{\mathbf{X}}_2^*, \ldots, \hat{\mathbf{X}}^*)$ of $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}^*$ and a finite random integer $\mathcal{N}$ such that

(i) $\hat{\mathbf{X}}_n^* \stackrel{d}{=} \mathbf{X}_n^*$, $\hat{\mathbf{X}}^* \stackrel{d}{=} \mathbf{X}$, $n \geq 1$,
(ii) $\hat{\mathbf{X}}_n^* = \hat{\mathbf{X}}^*$, $n \geq \mathcal{N}$.

Enlarge the corresponding probability space if necessary to carry some vector $\hat{\mathbf{V}} \stackrel{d}{=} \mathbf{V}$, independent of $(\hat{\mathbf{X}}_1^*, \hat{\mathbf{X}}_2^*, \ldots, \hat{\mathbf{X}}^*, \mathcal{N})$. Thus, $\mathbf{U}_n = \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) \stackrel{d}{=} \hat{\mathbf{U}}_n := \mathbf{G}_n(\hat{\mathbf{X}}_n^*, \hat{\mathbf{V}}) = \mathbf{G}_n(\hat{\mathbf{X}}^*, \hat{\mathbf{V}})$ for $n \geq \mathcal{N}$. $p_n(\mathbf{x}) \to p(\mathbf{x})$ entails $\mathbf{G}_n(\mathbf{x}, \boldsymbol{\lambda}) \to \mathbf{G}(\mathbf{x}, \boldsymbol{\lambda})$ pointwise. In turn, as $n \to \infty$, since $\mathcal{N}$ is finite,

$$\hat{\mathbf{U}}_n \stackrel{a.s.}{\to} \hat{\mathbf{U}} := \mathbf{G}(\hat{\mathbf{X}}^*, \hat{\mathbf{V}}) \stackrel{d}{=} \mathbf{G}(\mathbf{X}^*, \mathbf{V}) = \mathbf{U}. \quad \square$$

Remark 3. The fact that $\mathbf{G}$ is not continuous prevents one to mimic the proof by Faugeras [6, Theorem 4.1], i.e. to use Skorokhod's representation theorem and the continuous mapping theorem directly to $\mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$.

Corollary 3.2. *With P-probability one,*

$$\sup_{\mathbf{u}\in[0,1]^d} |C_n^1(\mathbf{u}) - C(\mathbf{u})| \to 0.$$

Proof. Theorem 3.1 implies $\hat{\mathbf{U}}_n \stackrel{d}{\to} \hat{\mathbf{U}}$. Since C is continuous, Polya's lemma entails the desired uniform consistency of cdfs. Since the result is valid for all $\omega \in \Omega_0$ with $P(\Omega_0) = 1$, the assertion occurs. $\square$

A referee has drawn our attention to the recent results obtained by Genest et al. [8], who consider weak convergence of the empirical copula process for count data. In our notation, they study weak convergence of the process $\sqrt{n}(C_n^1(\cdot) - C(\cdot)) := \hat{\mathbb{C}}_n^{\maltese}$, with a randomizer $\mathbf{V}$ made of independent uniform components. Their main result, Theorem 3.1, gives weak convergence of the empirical copula process on a compact subset K included in $[0, 1]^d$. As the limit process $\hat{\mathbb{C}}^{\maltese}$ of $\hat{\mathbb{C}}_n^{\maltese}$ depends on the partial derivatives of C, which may not be continuous in general at the (F_{k+}, F_{+l}) values corresponding to the support values of the marginals, weak convergence in the space of continuous functions of $[0, 1]^d$ equipped with the uniform norm cannot be expected in general. Hence, weak convergence is studied in the subspace $\mathcal{C}(K)$, for any compact subset $K \subset \mathcal{O}$, where $\mathcal{O}$ is a dense open subset of $[0, 1]^d$, whose complement is the Cartesian product of the ranges of the marginal distribution functions.

The main differences with the results obtained in the present article are as follows: on the one hand, Corollary 3.2 only deals with consistency of C_n^1 towards C, i.e. it is a first order statement, whereas Theorem 3.1 by [8] is a second order statement, i.e. it gives the limiting process $\hat{\mathbb{C}}^{\maltese}$ of $\hat{\mathbb{C}}_n^{\maltese}$. On the other hand, Corollary 3.2 is an almost sure statement, a stronger form of convergence than weak convergence, as in Theorem 3.1 by [8], or convergence in probability, as in their Corollary 4.1. Moreover, the key point is to realize that our goal is to derive uniform a.s. consistency results under the weakest possible assumption on the serial dependence in the sample, i.e. ergodicity, and not solely under the i.i.d. case, as in [8]. In such a general setting, it is impossible to obtain second-order asymptotics without adding further assumptions and structure on the data generating process. Also, the interest of the present manuscript lies also in the novel mathematical proof via the powerful coupling method. For all these reasons, the results presented here are not directly comparable and should be regarded as complementary as those of these authors.

Note also that a result similar to the main Theorem 3.1 by [8] already appears in Theorem 4.1 by Rüschendorf [18]. Indeed, thanks to Rademacher's theorem, the Lipschitz character of the copula function ensures the existence λ^d-almost everywhere of its partial derivatives. Moreover, direct evaluation of these partial derivatives via the distributional transform shows they are continuous λ^d-almost everywhere. Therefore, the limit Gaussian process L_0 in [18] or $\hat{\mathbb{C}}^{\maltese}$ in [8] is well defined and continuous on a λ^d-a.e. subset of $[0, 1]^d$, where λ stands for the univariate Lebesgue measure. Rüschendorf then argues that one can easily extend weak convergence from the continuous to the discrete case, hence the formulation of weak convergence on a subset K in [8]. Notice that [18] considers the more general sequential version of the empirical copula process, but that weak convergence is studied via an older technology, i.e. in the Skorokhod space of cadlag functions.

3.2. Convergence of empirical copulas regularized by convolution

The a.s. convergence of the kernel-smoothed empirical copula representer is presented in the next theorem.

Theorem 3.3. *If F purely discrete, then, for almost every realization $\mathbf{X}_1, \mathbf{X}_2, \ldots$, there exists a probability space which carries copula representers $\hat{\mathbf{U}}_n \sim C_n^2$, $\hat{\mathbf{U}} \sim C$ respectively, such that:*

(i) *C^2_n, C are copula functions, C is associated with F, C^2_n is associated with the kernel-smoothed empirical cdf;*
(ii) *$\hat{\mathbf{U}}_{n'} \xrightarrow{d} \hat{\mathbf{U}}$, for a subsequence $n' \to \infty$.*

Proof. Denote as before $(\hat{\mathbf{X}}^*_1, \hat{\mathbf{X}}^*_2, \dots, \hat{\mathbf{X}}^*)$ Thorisson's coupling of $\mathbf{X}^*_1, \mathbf{X}^*_2, \dots, \mathbf{X}$ and $\mathcal{N}$ the corresponding random finite integer such that $\hat{\mathbf{X}}^*_n = \hat{\mathbf{X}}^*$ for $n \geq \mathcal{N}$, with $\hat{\mathbf{X}}^*_n \sim F_n$, $\hat{\mathbf{X}}^* \sim F$. Enlarge the corresponding probability space if necessary to carry a continuous vector $\mathbf{Z} \sim K$, independent of $(\hat{\mathbf{X}}^*_1, \hat{\mathbf{X}}^*_2, \dots, \hat{\mathbf{X}}^*)$. Set

$$\mathbf{Y}_n := \hat{\mathbf{X}}^*_n + h_n\mathbf{Z}, \quad h_n \downarrow 0,$$

the smoothing of $\hat{\mathbf{X}}^*_n$; denote as $\hat{F}_n$, $\hat{\mathbf{G}}_n$ the corresponding joint and marginal cdfs and set

$$\hat{\mathbf{U}}_n := \hat{\mathbf{G}}_n(\mathbf{Y}_n)$$

the copula representer. $\hat{F}_n$ is the kernel-smoothed empirical cdf. Set

$$\bar{\mathbf{X}}_n = \hat{\mathbf{X}}^* + h_n\mathbf{Z},$$

denote as $\bar{F}_n$, $\bar{\mathbf{G}}_n$ the joint and vector of marginal cdfs of $\bar{\mathbf{X}}_n$, and let

$$\mathbf{U}_n := \bar{\mathbf{G}}_n(\bar{\mathbf{X}}_n).$$

For $n \geq \mathcal{N}$, $\mathbf{Y}_n = \hat{\mathbf{X}}^* + h_n\mathbf{Z} = \bar{\mathbf{X}}_n$. Thus,

$$\hat{\mathbf{U}}_n - \mathbf{U}_n = \hat{\mathbf{G}}_n(\bar{\mathbf{X}}_n) - \bar{\mathbf{G}}_n(\bar{\mathbf{X}}_n), \quad \text{for } n \geq \mathcal{N}.$$

But $\|\hat{\mathbf{G}}_n - \bar{\mathbf{G}}_n\|_\infty \leq \|\mathbf{G}_n - \mathbf{G}\|_\infty \to 0$, as $n \to \infty$, since pointwise convergence of probability mass function is equivalent to total variation convergence for discrete distributions (see Thorisson [24, Section 1.6.1]). Since $\mathcal{N}$ is finite, letting $n \to \infty$ therefore entails

$$\|\hat{\mathbf{U}}_n - \mathbf{U}_n\|_1 \leq \|\,\|\mathbf{G}_n - \mathbf{G}\,\|_\infty\,\|_1 \xrightarrow{a.s.} 0,$$

where we denoted $\|\,\|\mathbf{G}\,\|_\infty\,\|_1 = \sum_{i=1}^d \sup_{x_i \in \mathbb{R}} |G_i(x_i)|$.

Faugeras [6, Theorem 1.1], entails $\mathbf{U}_{n'} \xrightarrow{d} \mathbf{U}$, for some subsequence $n' \to \infty$ and some $\mathbf{U}$, whose cdf is a copula and satisfies Sklar's theorem associated with F. Slutsky's theorem yields the desired result. □

Corollary 3.4. *With P-probability one,*

$$\sup_{\mathbf{u} \in [0,1]^d} |C^2_{n'}(\mathbf{u}) - C(\mathbf{u})| \to 0.$$

Proof. As in Corollary 3.2. □

4. Consistency of concordance measure for discrete vectors

As a consequence of Theorem 3.1, one automatically obtains strong consistency of any continuous functional of $\mathbf{U}$ by the continuous mapping theorem. For example, one can obtain strong consistency of concordance measures for discrete vectors, as is shown below. We focus mainly on Kendall's τ, but similar results can be obtained for Spearman's ρ.

4.1. Concordance measure for bivariate continuous vectors

Let $\mathbf{X} = (X, Y)$ be a bivariate vector with cdf F. The idea of concordance is that an increase (resp. a decrease) in X is associated with an increase (resp. a decrease) in Y. More precisely, let $\mathbf{X}' = (X', Y')$ be an independent copy of $\mathbf{X}$. The probability of concordance is defined as

$$Q(\mathbf{X}, \mathbf{X}') := P((X - X')(Y - Y') > 0),$$

and the probability of discordance as

$$\bar{Q}(\mathbf{X}, \mathbf{X}') := P((X - X')(Y - Y') < 0).$$

Kendall's τ coefficient is defined as $\tau(F) := \mathcal{Q}(\mathbf{X}, \mathbf{X}')$, where

$$\mathcal{Q}(\mathbf{X}, \mathbf{X}') := Q(\mathbf{X}, \mathbf{X}') - \bar{Q}(\mathbf{X}, \mathbf{X}')$$

is called the concordance function, see Nelsen [14].

If F is continuous, $P(tie) := P((X - X')(Y - Y') = 0) = 0$, thus

$$\tau(F) = 2Q(\mathbf{X}, \mathbf{X}') - 1 = 4\int C(\mathbf{u})dC(\mathbf{u}) - 1 = 2Q(\mathbf{U}, \mathbf{U}') - 1$$

where $\mathbf{U} := \mathbf{G}(\mathbf{X}) \sim C := F \circ \mathbf{G}^{-1}$ is the unique copula representer associated to $\mathbf{X}$, and similarly for $\mathbf{U}' = \mathbf{G}(\mathbf{X}') \sim C' = C$, see Nelsen [14]. Hence, in the continuous case, $\tau(F) = \tau(C)$, i.e the coefficient is the same whether one computes it at the observational or at the copula level. Thus, it only depends on the copula function and as such remains invariant w.r.t. any strictly increasing transformations of the variables. Hence, it is scale invariant in the utmost manner, which makes it an attractive measure of dependence.

4.2. Concordance measures for bivariate discrete vectors

In the discrete case, $P(tie) \neq 0$ and there are many copulas associated to $\mathbf{X}$, as explained in Section 1.1. Therefore, expressing a population version of Kendall's τ in terms of a copula function associated to $\mathbf{X}$ is no longer straightforward. Denuit and Lambert [4], Neslehova [15] and Mesfioui and Quessy [12] propose generalizations of concordance coefficient for discrete vectors, based on the distributional transform.

Let $\mathbf{V} := (V_X, V_Y) \sim R$ be a bivariate vector with uniform $[0, 1]$ marginals. Let $\mathbf{V}'$ be an independent copy of $\mathbf{V}$, with $(\mathbf{V}, \mathbf{V}')$ also independent of $(\mathbf{X}, \mathbf{X}')$. Set $\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$ (resp. $\mathbf{U}' = \mathbf{G}(\mathbf{X}', \mathbf{V}')$) the copula representer associated to $\mathbf{X}$ (resp. $\mathbf{X}'$) with randomizer $\mathbf{V}$ (resp $\mathbf{V}'$). [12,4] propose to take a randomizer with independent marginals, i.e. $R = \Pi$ the independence (copula) distribution, and to define Kendall's τ in terms of the probability of concordance of the corresponding copula representer, i.e.

$$\tau(F) := \mathcal{Q}(\mathbf{U}, \mathbf{U}') = 2Q(\mathbf{U}, \mathbf{U}') - 1,$$

see Mesfioui and Quessy [12, Section 4]. (Some modified versions based on normalizations are also proposed in Mesfioui and Quessy [12] and Neslehova [15]). This approach is based on the fact that the distributional transform does not perturb the concordance function, see Neslehova [15, Theorem 5],

$$\mathcal{Q}(\mathbf{X}, \mathbf{X}') = \mathcal{Q}(\mathbf{U}, \mathbf{U}') = 2Q(\mathbf{U}, \mathbf{U}') - 1,$$

so that, if one sets C as the cdf of $\mathbf{U}$, the concordance function writes as a function of the copula C,

$$\mathcal{Q}(\mathbf{X}, \mathbf{X}') = 4\int C(\mathbf{u})dC(\mathbf{u}) - 1, \tag{3}$$

in parallel with the continuous case.

4.3. A.s. consistency of concordance measures for purely discrete vectors

Let $\mathbf{X} \in \mathbb{R}^d$ be a purely discrete-valued random vector with cdf F, $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ be an ergodic sample of $\mathbf{X}$ and F_n, $\mathbf{G}_n$ the empirical cdfs. Construct, as in Section 3, conditionally on the sample, a sequence $\hat{\mathbf{X}}_n^* \sim F_n$, $\mathbf{X}^* \sim F$, the copula representers $\mathbf{U}_n^* = \mathbf{G_n}(\mathbf{X_n^*}, \mathbf{V}) \sim C_n$, $\mathbf{U}^* = \mathbf{G}(\mathbf{X}^*, \mathbf{V}) \sim C$ associated with $\mathbf{X}_n^*$, $\mathbf{X}^*$, with $\mathbf{V}$ of independent marginals, as in the previous section.

Note that Eq. (3) writes, $\tau(F) = 4E^*[C(\mathbf{U}^*)] - 1$, so that it becomes straightforward to define the empirical counterpart of Kendall's coefficient: denote

$$\tau(F_n) := 4E^*[C_n^1(\mathbf{U}_n^*)] - 1,$$

where expectation E^* is conditional on the sample. The almost sure consistency w.r.t. the original probability measure is proved, as a simple corollary of Theorem 3.1,

Corollary 4.1. *With P-probability one,*

$$\tau(F_n) \to \tau(F),$$

as $n \to \infty$.

Proof. As in the previous proofs, one works conditionally on the sample, i.e. for a fixed $\omega \in \Omega_0$ with $P(\omega_0) = 1$. Then,

$$|C_n^1(\mathbf{U}_n^*) - C(\mathbf{U}^*)| \leq \|C_n^1 - C\|_\infty + |C(\mathbf{U}_n^*) - C(\mathbf{U}^*)|.$$

Both terms on the right-hand side of the inequality go to zero a.s., the first one by Corollary 3.2 and the second one is identically zero for $n \geq \mathcal{N}$, by Theorem 3.1. Uniform boundedness of these terms yield the desired result by dominated convergence. This reasoning is valid for every fixed $\omega \in \Omega_0$ with $P(\omega_0) = 1$, hence the assertion. □

Appendix

Proof of Theorem 2.1. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\infty$ be discrete random vectors with values in a finite or countable set E, s.t.

$$P(\mathbf{X}_n = \mathbf{x}) \to P(\mathbf{X}_\infty = \mathbf{x}), \quad \text{as } n \to, \infty, \ \forall \mathbf{x} \in E. \tag{4}$$

Let $q_0 := 0$ and $q_n(\mathbf{x}) := \inf_{n \le k < \infty} P(\mathbf{X}_k = \mathbf{x})$. Then, (4) entails that, for all $\mathbf{x} \in E$,

$$q_n(\mathbf{x}) \uparrow P(\mathbf{X}_\infty = \mathbf{x}), \quad \text{as } n \to \infty. \tag{5}$$

Let $\mathcal{N}, \mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{W}_1, \mathbf{W}_2, \ldots$ be independent random elements such that for $1 \le n < \infty$ and $\mathbf{x} \in E$,

$$P(\mathcal{N} = n) = \sum_{\mathbf{x} \in E} q_n(\mathbf{x}) - \sum_{\mathbf{x} \in E} q_{n-1}(\mathbf{x})$$

$$P(\mathbf{V}_n = \mathbf{x}) = \begin{cases} \dfrac{q_n(\mathbf{x}) - q_{n-1}(\mathbf{x})}{P(\mathcal{N} = n)} & \text{if } P(\mathcal{N} = n) > 0 \\ \text{arbitrary} & \text{if } P(\mathcal{N} = n) = 0 \end{cases}$$

$$P(\mathbf{W}_n = \mathbf{x}) = \begin{cases} \dfrac{P(\mathbf{X}_n = \mathbf{x}) - q_n(\mathbf{x})}{P(\mathcal{N} > n)} & \text{if } P(\mathcal{N} > n) > 0 \\ \text{arbitrary} & \text{if } P(\mathcal{N} > n) = 0. \end{cases}$$

By dominated convergence, (5) entails

$$P(\mathcal{N} \le n) = \sum_{\mathbf{x} \in E} q_n(\mathbf{x}) \uparrow \sum_{\mathbf{x} \in E} P(\mathbf{X}_\infty = \mathbf{x}) = 1,$$

as $n \to \infty$, i.e. $\mathcal{N}$ is finite. Define, for $1 \le n \le \infty$,

$$\hat{\mathbf{X}}_n = \begin{cases} \mathbf{V}_{\mathcal{N}} & \text{if } n \ge \mathcal{N}, \\ \mathbf{W}_n & \text{if } n < \mathcal{N}, \end{cases}.$$

Then $\hat{\mathbf{X}}_\infty = \mathbf{V}_{\mathcal{N}}$ and simple computations show that $P(\hat{\mathbf{X}}_n = \mathbf{x}) = P(\mathbf{X}_n = \mathbf{x})$ and $P(\hat{\mathbf{X}}_\infty = \mathbf{x}) = P(\mathbf{X}_\infty = \mathbf{x})$ for each $\mathbf{x} \in E$. $\square$

References

[1] Paul Deheuvels, Caractérisation complète des lois extrêmes mutivariées et de la convergence des types extrêmes, Publ. Inst. Stat. Univ. Paris 23 (3-4) (1978) 1–36.
[2] Paul Deheuvels, La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance, Bull. Acad. R. Belg. 65 (f.6) (1979) 274–292.
[3] Paul Deheuvels, A multivariate Bahadur–Kiefer representation for the empirical copula process, J. Math. Sci. 163 (2009) 382–398.
[4] Michel Denuit, Philippe Lambert, Constraints on concordance measures in bivariate discrete data, J. Multivariate Anal. 93 (1) (2005) 40–57.
[5] R.M. Dudley, Real Analysis and Probability, second ed., Cambridge University Press, 2002.
[6] Olivier P. Faugeras, Sklar's theorem derived using probabilistic continuation and two consistency results, J. Multivariate Anal. 122 (2013) 271–277.
[7] Jean-David Fermanian, Dragan Radulovic, Marten Wegkamp, Weak convergence of empirical copula processes, Bernoulli 10 (5) (2004) 847–860.
[8] Christian Genest, Johanna G. Neslehova, Bruno Rémillard, On the empirical multilinear copula process for count data, Bernoulli 20 (3) (2014) 1344–1371.
[9] Alexander M. Lindner, Alexander Szimayer, A limit theorem for copulas. Preprint, Discussion Paper 433, Sonderforschungsbereich 386, 2005. Available at epub.ub.uni-muenchen.de.
[10] Torgny Lindvall, Lectures on the Coupling Method, J. Wiley and Sons, 1992.
[11] Albert W. Marshall, Copulas, Marginals, and Joint Distributions, in: Lecture Notes–Monograph Series, vol. 28, Institute of Mathematical Statistics, Hayward, CA, 1996, pp. 213–222.
[12] M. Mesfioui, J.-F. Quessy, Concordance measures for multivariate non-continuous random vectors, J. Multivariate Anal. 101 (2010) 2398–2410.
[13] David S. Moore, M.C. Spruill, Unified large-sample theory of general chi-squared statistics for tests of fit, Ann. Statist. 3 (3) (1975) 599–616.
[14] Roger B. Nelsen, An Introduction to Copulas, Springer, 2006.
[15] J. Neslehova, On rank correlation measures for non-continuous random variables, J. Multivariate Anal. 98 (2007) 544–567.
[16] Ludger Rüschendorf, Asymptotic distributions of multivariate rank order statistics, Ann. Statist. 4 (5) (1976) 912–923.
[17] Ludger Rüschendorf, Stochastically ordered distributions and monotonicity of the oc-function of sequential probability ratio tests, Math. Oper.forsch. Stat. Ser. Stat. 12 (3) (1981) 327–338.
[18] Ludger Rüschendorf, On the distributional transform, Sklar's theorem, and the empirical copula process, J. Statist. Plann. Inference 139 (11) (2009) 3921–3927.
[19] Ludger Rüschendorf, Mathematical Risk Analysis — Dependence, Risk Bounds, Optimal Allocations and Portfolios, Springer, 2013.
[20] Carlo Sempi, Convergence of copulas: critical remarks, Rad. Mat. 12 (2004) 241–249.
[21] Abe Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Stat. Univ. Paris 8 (1959) 229–231.
[22] Abe Sklar, Random variables, joint distribution functions, and copulas, Kybernetika 9 (6) (1973) 449–460.
[23] A.V. Skorokhod, Limit theorems for stochastic processes, Theory Probab. Appl. 1 (1956) 261–290.
[24] Hermann Thorisson, Coupling, Stationarity, and Regeneration, Springer, 2000.