

## Accepted Manuscript

Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions

Georg Mainik

PII: S0047-259X(15)00168-2

DOI: <http://dx.doi.org/10.1016/j.jmva.2015.07.008>

Reference: YJMVA 3967

To appear in: *Journal of Multivariate Analysis*

Received date: 4 March 2012



Please cite this article as: G. Mainik, Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions, *Journal of Multivariate Analysis* (2015), <http://dx.doi.org/10.1016/j.jmva.2015.07.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions

Georg Mainik<sup>1</sup>

*RiskLab, Department of Mathematics, ETH Zurich  
Raemistrasse 101, 8092 Zurich, Switzerland*

---

## Abstract

This paper studies convergence properties of multivariate distributions constructed by endowing empirical margins with a copula. This setting includes Latin Hypercube Sampling with dependence, also known as the Iman–Conover method. The primary question addressed here is the convergence of the component sum, which is relevant to risk aggregation in insurance and finance. This paper shows that a CLT for the aggregated risk distribution is not available, so that the underlying mathematical problem goes beyond classic functional CLTs for empirical copulas. This issue is relevant to Monte-Carlo based risk aggregation in all multivariate models generated by plugging empirical margins into a copula. Instead of a functional CLT, this paper establishes strong uniform consistency of the estimated sum distribution function and provides a sufficient criterion for the convergence rate  $O(n^{-1/2})$  in probability. These convergence results hold for all copulas with bounded densities. Examples with unbounded densities include bivariate Clayton and Gauss copulas. The convergence results are not specific to the component sum and hold also for any other component-wise non-decreasing aggregation function. On the other hand, convergence of estimates for the joint distribution is much easier to prove, including CLTs. Beyond Iman–Conover estimates, the results of this paper apply to multivariate distributions obtained by plugging empirical margins into an exact copula or by plugging exact margins into an empirical copula.

*Keywords:* Risk aggregation, sum distribution, empirical margins, empirical copula, functional CLT, Iman–Conover, Latin hypercube sampling

---

## 1. Introduction

In various real-world applications, multivariate stochastic models are constructed upon empirical marginal data and an assumption on the dependence

---

<sup>1</sup>research@georgmainik.com

structure between the margins. This dependence assumption is often formulated in terms of copulas. The major reason for this set-up is the lack of multivariate data sets, as it is often the case in insurance and finance. This approach may appear artificial from the statistical point of view, but it arises naturally in the context of stress testing. In addition to finance and insurance, relevant application areas include engineering and environmental studies. Sometimes the marginal data is not even based on observations, but is generated by a univariate model that is considered reliable. Many of these models are so complex that the resulting distributions cannot be expressed analytically. In such cases exact marginal distributions are replaced by empirical distributions of simulated univariate samples. These empirical margins are endowed with some dependence structure to obtain a multivariate distribution. The computation of aggregated risk or other characteristics of this multivariate model is typically based on Monte-Carlo techniques.

*Iman–Conover: dependence “injection” by sample reordering*

Related methods include generation of synthetic multivariate samples from univariate data sets. Whilst the margins of such a synthetic sample accord with the univariate data, its dependence structure is modified to fit the application’s needs. The most basic example is the classic Latin Hypercube Sampling method, which mimics independent margins. It is a popular tool for removing spurious correlations from multivariate data sets. This method is also applied to variance reduction in the simulation of independent random variables (cf. McKay et al., 1979; Stein, 1987; Owen, 1992; Iman, 2008). Similar applications to dependent random variables include variance reduction in Monte-Carlo methods (Packham and Schmidt, 2010) and in copula estimation (Genest and Segers, 2010).

An extension of Latin Hypercube Sampling that brings dependence into the samples was proposed by Iman and Conover (1982). The original description of the Iman–Conover method uses random reordering of marginal samples, and the intention there was to control the rank correlations in the synthetic multivariate sample. The reordering is performed according to the vectors of marginal ranks in an i.i.d. sample of some multivariate distribution, say,  $H$ , with continuous margins. Thus rank correlations of  $H$  are “injected” into the synthetic sample. This procedure is equivalent to plugging empirical margins (obtained from asynchronous observations) into the rank based empirical copula of a sample of  $H$  (Arbenz et al., 2012). Moreover, it turned out that the Iman–Conover method allows to introduce not only the rank correlations of  $H$  into the synthetic samples, but the entire copula of  $H$  (cf. Arbenz et al., 2012; Mildenhall, 2005). In somewhat weaker sense, these results are related to the approximation of stochastic dependence by deterministic functions and to the pioneering result by Kimeldorf and Sampson (1978). Further developments in that area include measure preserving transformations (Vitale, 1990) and shuffles of min (Durante et al., 2009). In statistical optimization, reordering techniques were also used by Rüschemdorf (1983). A very recent, related application in quantitative risk management is a rearrangement algorithm that computes worst-case bounds for

the aggregated loss quantiles in a portfolio with given marginal distributions (cf. Embrechts et al., 2013, and references therein).

Using explicit reorderings of univariate marginal samples, the Iman–Conover method has a unique algorithmic tractability. It is implemented in various software packages, and it serves as a standard tool in dependence modelling and uncertainty analysis. The reordering algorithm allows even to construct synthetic samples with hierarchical dependence structures that meet the needs of risk aggregation in insurance and reinsurance companies (Arbenz et al., 2012). The distribution of the aggregated risk is estimated by the empirical distribution of the component sums  $\tilde{X}_1^{(k)} + \dots + \tilde{X}_d^{(k)}$  of the synthetic samples  $\tilde{X}^{(k)} = (\tilde{X}_1^{(k)}, \dots, \tilde{X}_d^{(k)})$  for  $k = 1, \dots, n$ . This Monte-Carlo approach has computational advantages. The resulting convergence rate of  $n^{-1/2}$  (or even faster with Quasi-Monte-Carlo using special sequences) allows to outperform explicit calculation of sum distributions already for moderate dimensions  $d \geq 4$  (cf. Arbenz et al., 2011).

*Challenge and contribution: convergence proofs*

Despite its popularity, some applications of the Iman–Conover method have been justified by simulations rather than by mathematical proofs. The original publication (Iman and Conover, 1982) derives its conclusions from promising simulation results for the distribution of the following function of a 4-dimensional random vector:  $f(X_1, \dots, X_4) = X_1 + X_2(X_3 - \log |X_1|) + \exp(X_4/4)$ . Yet a rigorous proof is still missing. The present paper provides a convergence proof for Iman–Conover estimates of the component sum distribution. It also includes a proof sketch for the much simpler case of the estimated joint distribution. Both problems have been open until now.

The solutions given in this paper are derived from the empirical process theory as presented in van der Vaart and Wellner (1996). Under appropriate regularity assumptions, Iman–Conover estimates of the sum distribution are strongly uniformly consistent with convergence rate  $O_P(n^{-1/2})$  (see Theorems 4.1 and 4.2). The convergence of Iman–Conover estimates for the joint distribution is discussed in cf. Remark 4.8. All these findings are not specific to the component sum and extend immediately to all componentwise non-decreasing functions (see Corollary 4.10). Moreover, Theorems 4.1 and 4.2 also cover the convergence of aggregated risk distributions obtained by Monte-Carlo sampling of a multivariate model constructed by plugging empirical margins into a copula (see Remark 4.9). In fact, both sampling methods (reordering by Iman–Conover and classic top-down sampling with empirical margins instead of the exact ones) lead to the same mathematical problem. This is discussed in Remark 3.2(d).

The regularity assumptions used here to establish the  $O_P(n^{-1/2})$  convergence rate for Iman–Conover estimates of sum distributions are satisfied for all copulas with bounded densities. This case includes the independence copula in arbitrary dimension  $d \geq 2$ . The assumptions are also satisfied for all bivariate Clayton copulas and for bivariate Gauss copulas with correlation parameter  $\rho \geq 0$ . The

convergence rate for  $\rho < 0$  is, if at all, only slightly weaker. The best bound that is currently available for  $\rho < 0$  is  $O_{\mathbb{P}}(n^{-1/2}\sqrt{\log n})$ .

The regularity assumptions for the marginal distributions involved in the Iman–Conover method are absolutely natural, and they are always satisfied by empirical distribution of i.i.d. samples: Strong uniform consistency of Iman–Conover estimates needs strong uniform consistency of consistency of empirical margins, whereas the uniform  $O_{\mathbb{P}}(n^{-1/2})$  convergence rate of Iman–Conover requires the same uniform convergence rate of  $O_{\mathbb{P}}(n^{-1/2})$  in the margins.

#### *Why a precise CLT remains elusive*

The convergence results obtained here are related to standard convergence results for empirical copulas (cf. Rüschendorf, 1976; Deheuvels, 1979; Fermanian et al., 2004; Segers, 2012). However, the mathematical problem for the sum distribution goes beyond the standard setting, where empirical measures are evaluated on rectangular sets. In the case of sum distributions, the usage of empirical margins in the construction of the multivariate model significantly extends the class of sets on which the empirical process of the copula sample should converge. As shown in Section 3, the canonical way to prove asymptotic normality for the Iman–Conover estimator of the sum distribution would need a uniform CLT for the copula sample on the collection of so-called lower layers in  $[0, 1]^d$ . However, this class is too complex for a uniform CLT (cf. Dudley, 1999, Theorems 8.3.2, 12.4.1, and 12.4.2). For this reason the proofs of consistency and convergence rate presented here sacrifice the precise asymptotic variance and use approximations that allow to simplify the problem. This technical difficulty is not specific to the Iman–Conover method. It also arises in any other application where multivariate samples are generated from a simulated copula sample and empirical marginal distributions. As mentioned above, this approach is very popular in practice, especially for computational reasons. Similar problems also arise in applications that combine exact marginal distributions with empirical copulas.

#### *Structure of the paper*

The paper is organized as follows. Section 2 introduces the reordering method and highlights the relations between sample reordering and empirical copulas. The complexity issues are discussed in Section 3. The convergence results are established in Section 4. The underlying regularity assumptions are discussed in Section 5, including examples of copula families that satisfy them. Conclusions are stated in Section 6.

## **2. Empirical copulas and sample reordering**

Let  $X = (X_1, \dots, X_d)$  be a random vector in  $\mathbb{R}^d$  with joint distribution function  $F$ , marginal distribution functions  $F_1, \dots, F_d$ , and copula  $C$ . That is,

$$F(x) = C(F_1(x_1), \dots, F_d(x_d)), \quad (1)$$

where  $C$  is a probability distribution function on  $[0, 1]^d$  with uniform margins. By Sklar's Theorem, any multivariate distribution function  $F$  admits this representation.

We assume throughout the following that  $F_i$  are unknown and that we have some uniform approximations  $F_{i,n}$ ,  $i = 1, \dots, d$ . The true margins  $F_i$  need not be continuous. In this case the representation (1) is not unique, but it is not an issue in our application, which is rather computational than statistical. As sketched in the Introduction, we consider the case where only univariate, asynchronous observations of the components  $X_i$  are available, and the copula  $C$  is set by expert judgement to compute the resulting distribution of the component sum. In practice, the choice of the copula  $C$  aims at dependence characteristics that are known or assumed for the random vector  $X$ . This choice also depends on the ability to sample  $C$  or any other multivariate distribution with continuous margins and copula  $C$ .

To keep the presentation simple, we assume that  $F_{i,n}$  are empirical distribution functions of some univariate samples  $X_i^{(1)}, \dots, X_i^{(n)}$  for  $i = 1, \dots, d$ :

$$F_{i,n}(t) := \frac{1}{n} \sum_{k=1}^n 1 \{X_i^{(k)} \leq t\}, \quad t \in \mathbb{R}. \quad (2)$$

These samples need not be i.i.d. We will only assume that  $\|F_{i,n} - F_i\|_\infty \rightarrow 0$ , either P-a.s. or in probability. Extensions to the general case will be given later on.

Let  $X_i^{(1:n)} \leq \dots \leq X_i^{(n:n)}$  denote the order statistics of the  $i$ -th component  $X_i$  for  $i = 1, \dots, d$ , and let  $P_H$  denote the probability measure with distribution function  $H$ . The Iman-Conover method approximates  $P_F$  by the empirical measure of the following synthetic multivariate sample:

$$\tilde{X}^{(j)} := \left( X_1^{(R_1^{(j):n})}, \dots, X_d^{(R_d^{(j):n})} \right), \quad j = 1, \dots, n, \quad (3)$$

where  $R_i^{(1)}, \dots, R_i^{(n)}$  for  $i = 1, \dots, d$  are the marginal ranks of a simulated i.i.d. sample  $U^{(1)}, \dots, U^{(n)} \sim C$ :

$$R_i^{(j)} = \sum_{k=1}^n 1 \{U_i^{(j)} \geq U_i^{(k)}\}, \quad i = 1, \dots, d, \quad j = 1, \dots, n.$$

It is easy to verify (cf. Arbenz et al., 2012, Theorem 3.2) that the empirical distribution function of the synthetic sample (3) is equal to

$$F_n^*(x) := C_n^*(F_{1,n}(x_1), \dots, F_{d,n}(x_d)), \quad (4)$$

where  $C_n^*$  is the rank based *empirical copula* of  $U^{(1)}, \dots, U^{(n)}$ :

$$C_n^*(u) := \frac{1}{n} \sum_{k=1}^n 1 \left\{ \frac{1}{n} R_1^{(k)} \leq u_1, \dots, \frac{1}{n} R_d^{(k)} \leq u_d \right\}. \quad (5)$$

This links the convergence of the Iman–Conover method to the convergence of  $P_{F_n^*}$  to  $P_F$ , and hence to the convergence of  $P_{C_n^*}$  to  $P_C$ .

*Remark 2.1.* (a) The central application of the Iman–Conover method discussed in the present paper is the computation of the aggregated risk distribution. The most common risk aggregation function is the sum. In this case one must compute the probability distribution of the random variable  $\sum_{i=1}^d X_i$  for  $(X_1, \dots, X_d) \sim F$  with  $F$  defined in (1). Iman–Conover involves two approximations: replacing the unknown margins  $F_i$  by their empirical versions  $F_{i,n}$ , and replacing the known (or treated as known) copula  $C$  by its empirical version  $C_n^*$ .

- (b) Using  $C_n^*$  may appear unnecessary because one can also compute the sum distribution for a random vector with margins  $F_{i,n}$  and exact copula  $C$ . However, computation of sum distributions from margins and copulas is quite difficult in practice. It involves numeric integration on non-rectangular sets, which cannot be reduced to taking the value of  $C(F_{1,n}(x_1), \dots, F_{d,n}(x_d))$  for a few points  $x = (x_1, \dots, x_d)$ . Implementations of this kind are exposed to the curse of dimensions. Monte-Carlo methods, which Iman–Conover belongs to, have the convergence rate of  $1/\sqrt{n}$ , and Quasi-Monte-Carlo methods using special sequences may even allow to achieve the rate  $1/n$ . According to Arbenz et al. (2011), explicit computation of sum distributions is outperformed by Monte-Carlo already for  $d = 4$ .
- (c) Another motivation of the Iman–Conover method is its flexibility and algorithmic tractability. It only includes reordering of samples and works in the same way for any dimension. Moreover, sample reordering is compatible with hierarchical dependence structures that can be described as trees with univariate distributions in leaves and copulas in branching nodes (cf. Arbenz et al., 2012). In each branching node, the marginal distributions are aggregated according to the node’s copula and the resulting aggregated (typically, sum) distribution is propagated to the next aggregation level. As shown in Arbenz et al. (2012), sample reordering can be implemented for a whole tree. The setting with one copula and margins discussed in the present paper is the basic element of such aggregation trees. The results presented here allow to prove the convergence of the aggregated (say, sum) distribution in every tree node, including the total sum.

Now let us return to the technical details of the Iman–Conover estimator for the aggregated sum distribution. As the random variables  $U_i^{(k)}$  are continuously distributed, they have no ties P-a.s. Thus  $P_{C_n^*}$  consists P-a.s. of  $n$  atoms of size  $1/n$ . Moreover, these  $n$  atoms build a Latin hypercube on the  $d$ -variate grid  $\{\frac{1}{n}, \frac{2}{n}, \dots, 1\}^d$ , i.e., each section  $\{x \in \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}^d : x_i = \frac{j}{n}\}$  for  $i = 1, \dots, d$  and  $j = 1, \dots, n$  contains precisely one atom. Therefore the Iman–Conover method is also called Latin Hypercube Sampling with dependence.

For  $X \sim F$ , let  $G$  denote the distribution function of the component sum:  $G(t) := P(X_1 + \dots + X_d \leq t)$ ,  $t \in \mathbb{R}$ . The relation between  $G$  and  $P_C$  can be expressed as follows.

**Lemma 2.2.**

$$\forall t \in \mathbb{R} \quad G(t) = P_C(\langle T(A_t) \rangle) \quad (6)$$

where  $A_t := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i \leq t\}$ ,  $T(x) := (F_1(x_1), \dots, F_d(x_d))$ , and

$$\langle B \rangle := \cup_{u \in B} [0, u] \quad \text{for } B \subset [0, 1]^d.$$

The notation  $[0, u]$  refers to the closed  $d$ -dimensional interval between 0 and  $u$ :  $[0, u] := [0, u_1] \times \dots \times [0, u_d]$ .

*Proof.* Let  $U \sim C$  and denote

$$T^{\leftarrow}(u) := (F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))$$

for  $u \in [0, 1]^d$ , where  $F_i^{\leftarrow}(y) := \inf\{t \in \mathbb{R} : F_i(t) \geq y\}$  is the quantile function of  $F_i$ . It is well known that  $T^{\leftarrow}(U) \sim F$ . Hence

$$G(t) = P_F(A_t) = P(T^{\leftarrow}(U) \in A_t) = P_C(\{u \in [0, 1]^d : T^{\leftarrow}(u) \in A_t\}),$$

and it suffices to show that  $v \in \langle T(A_t) \rangle$  is equivalent to  $T^{\leftarrow}(v) \in A_t$ .

If  $T^{\leftarrow}(v) \in A_t$ , then  $T \circ T^{\leftarrow}(v) \in \langle T(A_t) \rangle$ . Due to  $F_i \circ F_i^{\leftarrow}(v_i) \geq v_i$  for all  $i$  this implies that  $v \in [0, T \circ T^{\leftarrow}(v)] \subset \langle T(A_t) \rangle$ .

If  $v \in \langle T(A_t) \rangle$ , then  $v \leq T(x)$  (componentwise) for some  $x \in A_t$ . Since  $F_i^{\leftarrow} \circ F(x_i) \leq x_i$  for all  $i$ , this yields  $T^{\leftarrow}(v) \leq x$ . As the function  $x \mapsto \sum_{i=1}^d x_i$  is componentwise non-decreasing, we obtain that  $T^{\leftarrow}(v) \in A_t$ .  $\square$

*Remark 2.3.* (a) The measurability of  $\langle T(A_t) \rangle$  follows from the equivalence of  $v \in \langle T(A_t) \rangle$  and  $T^{\leftarrow}(v) \in A_t$ .

(b) The purpose of the operator  $\langle \cdot \rangle$  is to guarantee that for  $u \in \langle B \rangle$  and  $v \in [0, 1]^d$  the componentwise ordering  $v \leq u$  implies  $v \in \langle B \rangle$ . This immediately yields  $\langle \langle B \rangle \rangle = \langle B \rangle$ . Consistently with Dudley (1999), we will call  $\langle B \rangle$  the *lower layer* of  $B$ . This set class is also mentioned in the context of nonparametric regression (cf. Wright, 1981, and references therein).

(c) The sets  $\langle T(A_t) \rangle$  are closed if the marginal distributions  $F_i$  have bounded domains, but not necessarily in the general case. If, for instance,  $F_1 = F_2$  are standard normal distributions, then  $\langle T(A_0) \rangle = \{u \in [0, 1]^2 : u_1 + u_2 \leq 1\} \setminus \{(0, 1), (1, 0)\}$ . This example with punctured corners is quite prototypical. It is easy to show that if a sequence  $u^{(n)}$  in  $\langle T(A_t) \rangle$  converges to  $u \notin \langle T(A_t) \rangle$ , then  $u$  is on the boundary of  $[0, 1]^d$ . Indeed, if  $u \in (0, 1)^d$ , then  $u^{(n)} \in (0, 1)^d$  for sufficiently large  $n$ . This allows to construct a sequence  $v^{(n)} \rightarrow u$  such that  $v^{(n)} \leq u^{(n)}$  and  $v^{(n)} \leq u$  for all  $n$ . For any  $u^{(n)} \in \langle T(A_t) \rangle$  there exists  $x^{(n)} \in A_t$  such that  $T(x^{(n)}) \geq u^{(n)}$ . As  $T^{\leftarrow}$  is non-decreasing and  $F_i^{\leftarrow} \circ F(x_i) \leq x_i$  for all  $i$ , we have  $x^{(n)} \geq T^{\leftarrow}(u^{(n)}) \geq T^{\leftarrow}(v^{(n)})$  and hence  $T^{\leftarrow}(v^{(n)}) \in A_t$  for all  $n$ . Since all  $F_i^{\leftarrow}$  are left continuous on  $(0, 1)$ , we obtain  $T^{\leftarrow}(v^{(n)}) \rightarrow T^{\leftarrow}(u)$  and  $T^{\leftarrow}(u) \in A_t$ . As  $F_i \circ F_i^{\leftarrow}(u_i) \geq u_i$  for all  $i$ , we obtain  $u \in \langle T(A_t) \rangle$ . Thus  $u \notin \langle T(A_t) \rangle$  is only possible for  $u \in \partial[0, 1]^d$ .

By construction,  $\langle T(A_t) \rangle$  includes all points  $u \in \partial[0, 1]^d$  such that  $u + \varepsilon e_i \in \langle T(A_t) \rangle$  for some unit vector  $e_i$ ,  $i = 1, \dots, d$ , and  $\varepsilon > 0$ . Thus the area where the set  $\langle T(A_t) \rangle$  does not include its boundary points is very small.

Let us now return to the estimation of the sum distribution  $G(t) = P_F(A_t)$ . The empirical distribution of the component sum in the synthetic sample (3) is nothing else than the empirical multivariate distribution of this sample evaluated at the sets  $A_t$ :

$$\begin{aligned} G_n^*(t) &:= \frac{1}{n} \sum_{k=1}^n \mathbb{1} \left\{ X_1^{(R_1^{(k):n})} + \dots + X_d^{(R_d^{(k):n})} \leq t \right\} \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{1} \left\{ \left( X_1^{(R_1^{(k):n})}, \dots, X_d^{(R_d^{(k):n})} \right) \in A_t \right\}. \end{aligned} \quad (7)$$

Analogously to (6),  $G_n^*(t) = P_{F_n^*}(A_t)$  can be written in terms of the empirical copula  $C_n^*$  defined in (5) and, as next step, in terms of the empirical distribution  $C_n$  of the i.i.d. copula sample  $U^{(1)}, \dots, U^{(n)}$ :

$$C_n(u) := \frac{1}{n} \sum_{k=1}^n \mathbb{1} \{ U^{(k)} \in [0, u] \}, \quad u \in [0, 1]^d.$$

Let  $C_{i,n}$  denote the margins of  $C_n$ , and let  $C_{i,n}^{\leftarrow}$  denote the corresponding quantile functions. To avoid technicalities, we consider  $C_{i,n}^{\leftarrow}$  as mappings from  $[0, 1]$  to  $[0, 1]$ :

$$C_{i,n}^{\leftarrow}(u) := \inf \{ v \in [0, 1] : C_{i,n}(v) \geq u \}, \quad u \in [0, 1].$$

Denote  $\tau_n(x) := (F_{1,n}(x_1), \dots, F_{d,n}(x_d))$  and  $T_n := \rho_n^{\leftarrow} \circ \tau_n$ , where  $\rho_n^{\leftarrow} := (C_{1,n}^{\leftarrow}(x_1), \dots, C_{d,n}^{\leftarrow}(x_d))$ . Then we can state the following result.

**Corollary 2.4.**

$$\forall t \in \mathbb{R} \quad G_n^*(t) = P_{C_n^*}(\langle \tau_n(A_t) \rangle) \quad (8)$$

and, with probability 1,

$$\forall t \in \mathbb{R} \quad G_n^*(t) = P_{C_n}(\langle T_n(A_t) \rangle). \quad (9)$$

*Proof.* It is easy to see that the synthetic sample (3) can be written as

$$\tilde{X}^{(k)} = \tau_n^{\leftarrow} \circ \rho_n \left( U^{(k)} \right), \quad k = 1, \dots, n,$$

where  $\tau_n^{\leftarrow}(x) := (F_{1,n}^{\leftarrow}(x_1), \dots, F_{d,n}^{\leftarrow}(x_d))$  and  $\rho_n(x) := (C_{1,n}(x_1), \dots, C_{d,n}(x_d))$ . This yields

$$G_n^*(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1} \left\{ \tau_n^{\leftarrow} \circ \rho_n \left( U^{(k)} \right) \in A_t \right\}. \quad (10)$$

According to the proof of Lemma 2.2,  $\tau_n^{\leftarrow}(x) \in A_t$  is equivalent to  $x \in \langle \tau_n(A_t) \rangle$ . Hence (10) implies

$$G_n^*(t) = \frac{1}{n} \sum_{k=1}^n \mathbf{1} \left\{ \rho_n \left( U^{(k)} \right) \in \langle \tau_n A_t \rangle \right\},$$

which is the same as (8) because  $C_n^*$  is the empirical distribution function of  $\rho_n(U^{(1)}), \dots, \rho_n(U^{(n)})$ .

Being continuously distributed,  $U_i^{(1)}, \dots, U_i^{(n)}$  have different values P-a.s. for each  $i$ . Hence the mapping  $\rho_n^{\leftarrow}$  is componentwise P-a.s. strictly increasing on  $\{\frac{1}{n}, \dots, 1\}^d$  with probability 1, and therefore

$$\mathbf{1} \left\{ \rho_n \left( U^{(k)} \right) \in \langle \tau_n(A_t) \rangle \right\} = \mathbf{1} \left\{ \rho_n^{\leftarrow} \circ \rho_n \left( U^{(k)} \right) \in \langle \rho_n^{\leftarrow}(\langle \tau_n(A_t) \rangle) \rangle \right\} \quad \text{P-a.s.}$$

Thus (9) follows from  $\rho_n^{\leftarrow} \circ \rho_n(U^{(k)}) = U^{(k)}$  and  $\langle \rho_n^{\leftarrow}(\langle \tau_n(A_t) \rangle) \rangle = \langle \rho_n^{\leftarrow} \circ \tau_n(A_t) \rangle$ .  $\square$

*Remark 2.5.* Uniform consistency of  $F_{i,n}$  and  $C_{i,n}^{\leftarrow}$  implies  $T_n \rightarrow T$  in  $l^\infty(\mathbb{R}^d)$ .

### 3. Complexity of the problem

The representation (9) translates the asymptotic normality of  $G_n^*$  into a CLT for  $C_n$  uniformly on the random set sequence  $(\langle T_n(A_t) \rangle : n \in \mathbb{N})$ . The canonical way to prove results of this kind is to establish a uniform CLT on the set class  $\mathcal{T}_d$  of all possible  $\langle T_n(A_t) \rangle$  and  $\langle T(A_t) \rangle$ . This is the natural set class to work with if  $F_i$  are unknown and estimated empirically. The index  $d$  in the notation  $\mathcal{T}_d$  highlights the dimension. Since  $F_i$  need not be continuous, the set of all  $T$  includes all  $T_n$ , so that  $\mathcal{T}_d$  is simply the collection of all possible  $\langle T(A_t) \rangle$ . Furthermore, if each unknown margin  $F_i$  has a positive density on entire  $\mathbb{R}$ , then the resulting empirical distributions  $F_{i,n}$  can take any value in the class of all possible stair functions on  $\mathbb{R}$  with steps of size  $1/n$  going from 0 to 1. In this case the class of all possible  $\langle T_n(A_t) \rangle$  is dense (w.r.t. Hausdorff metric) in the class of all possible  $T(A_t)$ . Thus, even though there is only one limit transformation  $T$  that really matters to us, considering the class of all possible  $\langle T(A_t) \rangle$  does not add more complexity to the problem.

It is also easy to see that even pointwise asymptotic normality of  $G_n^*(t)$  in some  $t = t_0$  would require a uniform CLT on  $\mathcal{T}_d$ . Shifting the unknown margins  $F_i$ , one can easily generate all possible sets  $T(A_t)$  from a single  $A_{t_0}$ . Thus the complexity of the problem is the same for the uniform and for the pointwise asymptotic normality of  $G_n^*$ .

There are various functional CLTs for empirical copulas (cf. Rüschendorf, 1976; Deheuvels, 1979; Fermanian et al., 2004; Segers, 2012, and references therein). However, empirical copula estimates are empirical measures evaluated on the set class  $\mathcal{R}_d$  of rectangle cells  $(-\infty, a] := (-\infty, a_1] \times \dots \times (-\infty, a_d]$  for  $a \in \mathbb{R}^d$ . This set class is simple enough to be *universally Donsker*. A set

class  $\mathcal{C}$  is called *P-Donsker* if the empirical measure  $\mathbb{P}_n(B) := \frac{1}{n} \sum_{k=1}^n 1_B(Y^{(k)})$  of an i.i.d. sample  $Y^{(1)}, \dots, Y^{(n)} \sim P$  satisfies

$$\sqrt{n}(\mathbb{P}_n(B) - P(B)) \xrightarrow{w} \mathbb{G}_P(B) \quad (11)$$

as a mapping in  $l^\infty(\mathcal{C})$ , where  $\mathbb{G}_P$  is the so-called Brownian bridge “with time”  $P$ . That is,  $\mathbb{G}_P$  is a centred Gaussian process with index  $B \in \mathcal{C}$  and covariance structure

$$\text{cov}(\mathbb{G}_P(A), \mathbb{G}_P(B)) = P(A \cap B) - P(A)P(B).$$

The Donsker property of  $\mathcal{C}$  is called *universal* if it holds for any probability measure  $P$  on the sample space.

The symbol  $\xrightarrow{w}$  in (11) refers to the extended notion of weak convergence for non-measurable mappings in  $l^\infty(\mathcal{C})$  as used in van der Vaart and Wellner (1996). See Remark 4.5 for further details.

Sufficient conditions for a set class to be Donsker can be obtained from the *entropy* of this set class. Entropy conditions can be formulated in terms of *covering numbers* or *bracketing numbers* (cf. van der Vaart and Wellner, 1996, Sections 2.1 and 2.2). Entropy bounds that do not depend on the underlying probability measure are called *uniform*. The most common sufficient criterion for uniform entropy bounds guaranteeing that a set class is universally Donsker is the *Vapnik-Červonenkis (VC)* property. A set class  $\mathcal{C}$  is VC if it does not *shatter* any  $n$ -point set  $\{x^{(1)}, \dots, x^{(n)}\}$  for sufficiently large  $n$ . The set  $\{x^{(1)}, \dots, x^{(n)}\}$  is shattered by  $\mathcal{C}$  if every subset  $A \subset \{x^{(1)}, \dots, x^{(n)}\}$  can be obtained as  $A = B \cap \{x^{(1)}, \dots, x^{(n)}\}$  with some  $B \in \mathcal{C}$ . The smallest  $n$  such that no  $n$ -point set is shattered by  $\mathcal{C}$  is called *VC-index* of  $\mathcal{C}$ .

It is well known that the set class  $\mathcal{R}_d$  is VC with index  $d + 1$  (cf. van der Vaart and Wellner, 1996, Example 2.6.1). This yields asymptotic normality of  $C_n(u)$  uniformly in  $u \in [0, 1]^d$ . The asymptotic normality of empirical copulas follows then by functional Delta method. These functional CLTs allow to prove asymptotic normality for the estimators of the multivariate distribution function  $F(x)$  that are derived from  $C_n$  or  $C_n^*$ .

Unfortunately, the problem for  $G_n^*$  is much more difficult. As shown above, a functional CLT for  $G_n^*$  is closely related to a uniform CLT for  $C_n$  on the set class  $\mathcal{T}_d$ . The complexity of  $\mathcal{T}_d$  is much higher than that of  $\mathcal{R}_d$ . Lemma 3.1 stated below implies that  $\mathcal{T}_d$  is not VC, and Remark 3.2(c) shows that the complexity of this set class is even so high that a uniform CLT on  $\mathcal{T}_d$  does not hold.

Let  $\mathcal{H}_d$  denote the collection of all lower layers in  $[0, 1]^d$ :

$$\mathcal{H}_d := \{B \subset [0, 1]^d : \langle B \rangle = B\},$$

and denote  $\overline{\mathcal{H}}_d := \{\overline{B} : B \in \mathcal{H}_d\}$ . Analogously, denote  $\overline{\mathcal{T}}_d := \{\overline{B} : B \in \mathcal{T}_d\}$ . According to Remark 2.3(c),  $B \in \mathcal{T}_d$  implies that  $\overline{B} \setminus B \subset \partial[0, 1]^d$ , which is a  $P_C$ -null set for any copula  $C$ . Thus, for empirical processes constructed from copula samples, uniform convergence on  $\mathcal{T}_d$  is equivalent to uniform convergence on  $\overline{\mathcal{T}}_d$ .

It is obvious that  $\overline{\mathcal{T}}_d \subset \overline{\mathcal{H}}_d$ . The following result shows that for  $d = 2$  these set classes are almost identical.

**Lemma 3.1.** *If  $B \in \overline{\mathcal{H}_2}$ , then  $B \cup \Lambda_2 \in \overline{\mathcal{T}_2}$ , where  $\Lambda_2 := \{u \in [0, 1]^2 : u_1 u_2 = 0\}$  is the union of lower faces of  $[0, 1]^2$ .*

*Proof.* Denote  $B' := B \cup \Lambda_2$ . It suffices to find probability distribution functions  $F_1, F_2$  such that  $\langle T(A_0) \rangle = B'$  with  $T(x) = (F_1(x_1), F_2(x_2))$ . Denote

$$F_1(t) := t1_{[0,1]}(t) + 1_{[1,\infty)}(t)$$

and

$$F_2(t) := \begin{cases} 0 & \text{if } t < -1 \\ \sup\{s \in [0, 1] : (-t, s) \in B'\} & \text{if } t \in [-1, 0) \\ 1 & \text{if } t \geq 0 \end{cases}$$

Since  $B' \in \overline{\mathcal{H}_2}$ , the function  $F_2$  is non-decreasing in  $t$ . Indeed, if  $(-t, s) \in B'$ , then  $(-t - \delta, s) \in B'$  for any  $\delta \in (0, t]$ , and hence  $F_2(t + \delta) \geq F_2(t)$ . The maximal value of  $F_2$  is 1, and  $F_2$  is right continuous because  $B'$  is closed. Thus  $F_2$  is a probability distribution function.

As  $F_1$  and  $F_2$  are non-decreasing, we have

$$\langle T(A_0) \rangle = \langle T(\partial A_0) \rangle$$

where  $\partial A_0 := \{(-t, t) : t \in \mathbb{R}\}$  is the boundary of  $A_0$ . Now observe that

$$T(\partial A_0) = \{(F_1(t), F_2(-t)) : t \in \mathbb{R}\} = \{(t, F_2(-t)) : t \in [0, 1]\}.$$

Hence  $\langle T(A_0) \rangle$  is the area enclosed between the zero line and the graph of  $F_2(-t)$  for  $t \in [0, 1]$ . This is precisely  $B'$ .  $\square$

*Remark 3.2.* (a) Since  $\Lambda_2$  is a  $P_C$ -null set for any copula  $C$ , the modification of  $B$  into  $B \cup \Lambda_2$  in Lemma 3.1 has no influence on the uniform convergence of empirical processes obtained from copula samples.

(b) The set classes  $\overline{\mathcal{H}_d}$  and  $\mathcal{H}_d$  are not VC. For instance, they shatter all sets  $\{u \in \{0, \frac{1}{n}, \dots, 1\}^d : u_1 + \dots + u_d = 1\}$  for  $n \in \mathbb{N}$ . Any subset  $B$  of this hyperplane in  $\{0, \frac{1}{n}, \dots, 1\}^d$  can be picked out by  $\langle B \rangle \in \overline{\mathcal{H}_d} \subset \mathcal{H}_d$ . Similar arguments apply to the modified set class  $\{B \cup \Lambda_d : B \in \overline{\mathcal{H}_d}\}$ . Hence Lemma 3.1 implies that  $\overline{\mathcal{T}_2}$  is not VC. This rules out the canonical usage of VC criteria in convergence proofs for  $G_n^*$ .

(c) The problem is even more difficult, and also more remarkable. In fact, the set classes  $\mathcal{H}_d$  and  $\overline{\mathcal{H}_d}$  for  $d \geq 2$  are not Donsker with respect to the Lebesgue measure on  $[0, 1]^d$  (cf. Dudley, 1999, Theorems 8.3.2, 12.4.1, and 12.4.2). Thus Lemma 3.1 implies that a uniform CLT on  $\overline{\mathcal{T}_2}$  does not hold in the most basic case, when  $C$  is the independence copula. Therefore one cannot prove asymptotic normality of  $G_n^*$  via uniform CLT on  $\mathcal{T}_2$ , and precise asymptotic variance of  $G_n^*$  also seems out of reach.

- (d) The complexity issues are not specific to the estimator  $G_n^*$  obtained by plugging empirical margins  $F_{i,n}$  into the rank based empirical copula  $C_n^*$ . They also affect models generated by plugging  $F_{i,n}$  directly into the “exact” copula  $C$ . Top-down simulation of such models means marginal transformation of copula samples  $U^{(1)}, \dots, U^{(n)} \sim C$  by  $F_{i,n}^{\leftarrow}$ . The resulting estimate of the component sum distribution  $G(t)$  can be written as  $P_{C_n}(\langle \tau_n(A_t) \rangle)$  where  $\tau_n(x) := (F_{1,n}(x_1), \dots, F_{d,n}(x_1))$ . The sets  $\langle \tau_n(A_t) \rangle$  feature the same stair shape as the sets  $\langle T_n(A_t) \rangle$ . Asymptotic normality of  $P_{C_n}(\langle \tau_n(A_t) \rangle)$  leads us again to a uniform CLT on  $\overline{\mathcal{T}}_d$ .
- (e) It is not yet clear whether Lemma 3.1 can be extended to  $B \in \overline{\mathcal{H}}_d$  for  $d > 2$ . However, the complexity of  $\overline{\mathcal{H}}_d$  can only increase for greater  $d$ . It is easy to embed  $\overline{\mathcal{H}}_2$  in  $\overline{\mathcal{H}}_d$  for  $d > 2$  by identifying  $\overline{\mathcal{H}}_2$  with the following subclass:

$$\overline{\mathcal{H}}_{d|1,2} \{B \in \overline{\mathcal{H}}_d : B = B' \times [0, 1]^{d-2}, B' \in \overline{\mathcal{H}}_2\}.$$

Setting  $F_i(t) := 1_{[0,\infty)}(t)$  for  $i = 3, \dots, d$  in the proof of Lemma 3.1, the result obtained there can be extended to  $B \cup \Lambda_d \in \overline{\mathcal{T}}_d$  for  $B \in \overline{\mathcal{H}}_{d|1,2}$ . This allows to extend the conclusions in (b,c,d) to all dimensions  $d > 2$ .

*Remark 3.3.* (a) The results of this section can be summarized as follows: The true target set class  $\mathcal{H}_T := \{ \langle T(A_t) \rangle : t \in \mathbb{R} \}$  is simple (it will be shown in Remark 4.12 that  $\mathcal{H}_T$  is VC with index 2), but unknown. Replacing these unknown margins by the empirical ones, we obtain random elements of the set class  $\overline{\mathcal{T}}_d$ , which is too complex for a uniform CLT. This is the reason why the convergence proofs presented below sacrifice precise asymptotic variance. The resulting loss of precision can be considered as the price one is forced to be pay for using empirical margins  $F_{i,n}$  instead of the true ones.

- (b) Similar issues can also arise when the margins are known, but the copula is not. The implicit use of transformations  $\rho_n$  in  $C_n^*$  (cf. proof of Corollary 2.4) entails deformations of target sets  $B \in \mathcal{H}_T$  that are very similar to the ones caused by  $T_n$  or  $\tau_n$ . Thus, depending on the application, loss of precision may also be caused by the use of an empirical copula. In particular, a uniform CLT for  $C_n^*$  on the set class  $\mathcal{H}_T$  is still an open problem, and the foregoing results suggest a plausible explanation why this problem is so hard.
- (c) A deeper reason behind these complexity issues is the typical shape of the target sets  $B \in \mathcal{H}_T$ . If the margins or the copula are estimated empirically, corresponding random transformations of  $B$  ( $\tau_n$  and  $\rho_n^{\leftarrow}$  in Corollary 2.4) significantly increase the complexity of the problem. Depending on the application, the resulting loss of precision can be attributed to empirical margins, empirical copulas, or both.

#### 4. Convergence results

The major problem studied in this paper is the uniform convergence of the Iman–Conover estimator  $G_n^*$  introduced in (7). Strong consistency in  $l^\infty(\mathbb{R})$  is

established in Theorem 4.1. A sufficient condition for the convergence rate to be  $n^{-1/2}$  is given in Theorem 4.2. These results are stated below and followed by some corollaries, remarks, and auxiliary results needed in the proofs. The technical proofs of the auxiliary results are provided in Section 4.1.

**Theorem 4.1.** *Let  $G(t) := P(X_1 + \dots + X_d \leq t)$  for  $X \sim F$  as defined in (1), and let  $G_n^*$  be the Iman–Conover estimator of  $G$  introduced in (7). Assume that the functions  $F_{i,n}$  defined in (2) satisfy*

$$\|F_{i,n} - F_i\|_\infty \rightarrow 0 \text{ P-a.s.}, \quad i = 1, \dots, d. \quad (12)$$

*If the copula  $C$  of  $F$  is Lebesgue absolutely continuous, then  $\|G_n^* - G\|_\infty \rightarrow 0$  P-a.s.*

As discussed in Remark 3.2(c), a CLT for  $G_n^*$  seems out of reach, so that the convergence rate is established as an  $O_P(n^{-1/2})$  bound. This notation is related to tightness:  $Z_n = O_P(1)$  means that  $Z_n$  is tight, and  $Z_n = O_P(a_n)$  is equivalent to  $a_n^{-1}Z_n = O_P(1)$ . In particular, if  $|Z_n| \leq |Y_n|$  and  $Y_n \xrightarrow{w} Y$ , then  $Z_n = O_P(1)$ .

The regularity assumptions also need some additional notation. In the following, let  $B_t$  denote the “upper” boundary of  $\langle T(A_t) \rangle$ :

$$B_t := \left\{ x \in \overline{\langle T(A_t) \rangle} : \forall \varepsilon > 0 \quad x + (\varepsilon, \dots, \varepsilon) \notin \overline{\langle T(A_t) \rangle} \right\} \quad t \in \mathbb{R},$$

and let  $U_\delta(B_t)$  denote the closed  $\delta$ -neighbourhood of  $B_t$  in Euclidean distance:

$$U_\delta(B_t) := \{u \in [0, 1]^d : |u - v| \leq \delta \text{ for some } v \in B_t\}. \quad (13)$$

One of the regularity assumptions in Theorem 4.2 specifies the probability mass that the copula  $C$  assigns to  $U_\delta(B_t)$ . The other one involves the Lebesgue density  $c$  of  $C$ . For  $\varepsilon \in (0, 1/2)$ , we denote

$$K(\varepsilon) := \text{ess sup} \{c(u) : u \in [\varepsilon, 1 - \varepsilon]^d\}.$$

The growth of  $K(\varepsilon)$  for  $\varepsilon \rightarrow 0$  specifies the behaviour of  $c$  near the boundary of  $[0, 1]^d$ .

**Theorem 4.2.** *Let  $G(t) := P(X_1 + \dots + X_d \leq t)$  for  $X \sim F$  as defined in (1), and let  $G_n^*$  be the Iman–Conover estimator of  $G$  introduced in (7). Assume that the functions  $F_{i,n}$  defined in (2) satisfy*

$$\|F_{i,n} - F_i\|_\infty = O_P(n^{-1/2}), \quad i = 1, \dots, d, \quad (14)$$

*and that the copula  $C$  of  $F$  is absolutely continuous and satisfies*

$$\sup_{t \in \mathbb{R}} P_C(U_\delta(B_t)) = O(\delta) \quad (15)$$

*for  $\delta \rightarrow 0$  and  $U_\delta(B_t)$  defined in (13). Further, assume that*

$$\int_0^{1/2} \sqrt{\log K(\varepsilon^2)} d\varepsilon < \infty. \quad (16)$$

*Then  $\|G_n^* - G\|_\infty = O_P(n^{-1/2})$ .*

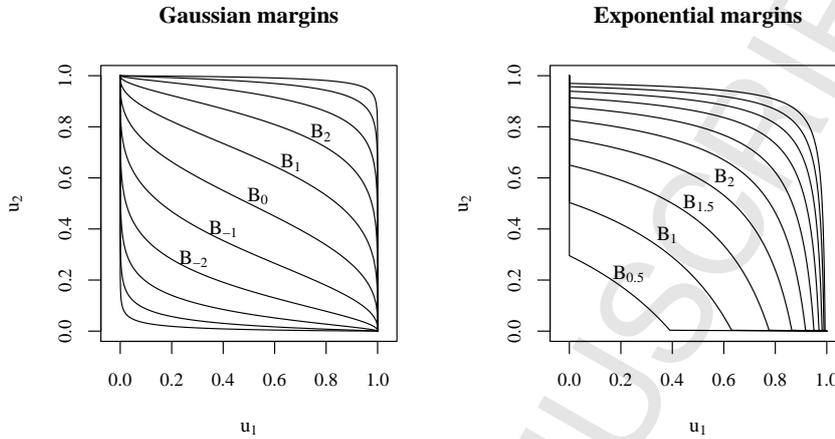


Figure 1: Set families  $\{B_t : t \in \mathbb{R}\}$ . Left hand side:  $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2 \sim \mathcal{N}(0, 1/4)$ . Right hand side:  $X_1 \sim \text{Exp}(1)$  and  $X_2 \sim \text{Exp}(0.7)$ .

*Remark 4.3.* The shapes of the sets  $B_t$  strongly depend on the marginal distributions  $F_i$ . Two examples of set families  $\{B_t : t \in \mathbb{R}\}$  are given in Figure 1. The verification of (15) for some copula families is discussed in Section 5. The examples presented there suggest that this condition is non-trivial, and that it depends on the interplay between the copula  $C$  and the true, unknown margins  $F_i$ .

We proceed with an auxiliary result that gives us an upper bound for the volume of  $U_\delta(B_t)$ . It follows from the componentwise monotonicity of the transformation  $T$ . The general idea behind this result is that the “surface area” of  $B_t$  is bounded by the sum of its  $d$  projections on the lower faces of the unit square  $[0, 1]^d$ , and the “thickness” of  $U_\delta(B_t)$  is roughly  $2\delta$ . The proof is given in Section 4.1.

**Lemma 4.4.** *Let  $\lambda$  denote the Lebesgue measure on  $[0, 1]^d$ . Then  $\lambda(U_\delta(B_t)) \leq 2d\delta$ .*

The next lemma provides the Glivenko–Cantelli and Donsker properties for two set classes involved in the proofs of Theorems 4.1 and 4.2. The Donsker property is defined in (11). A set class  $\mathcal{C}$  is called *P-Glivenko–Cantelli* if the empirical measure  $\mathbb{P}_n(B) := \frac{1}{n} \sum_{k=1}^n 1_B(Y^{(k)})$  of an i.i.d. sample  $Y^{(1)}, \dots, Y^{(n)} \sim \mathbb{P}$  satisfies

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{C}} := \sup_{B \in \mathcal{C}} |\mathbb{P}_n(B) - \mathbb{P}(B)| \xrightarrow{\text{a.s.}} 0 \quad (17)$$

This notation emphasizes that the convergence also depends on the true distribution that is sampled to construct  $\mathbb{P}_n$ .

*Remark 4.5.* One technical aspect of (17) and (11) needs an additional comment. These statements regard  $\mathbb{P}_n$  and  $\mathbb{P}$  as mappings from the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  to  $l^\infty(\mathcal{C})$ . However,  $\mathbb{P}_n$  need not be measurable with respect to the Borel  $\sigma$ -field on  $l^\infty(\mathcal{C})$  (cf. Billingsley, 1968, Chapter 18). This issue can be solved by extended versions of almost sure and weak convergence as presented in van der Vaart and Wellner (1996). In the following,  $\xrightarrow{\text{a.s.}}$  and  $\xrightarrow{\text{w}}$  are understood according to that monograph. In case of measurability these extended notions coincide with the standard ones.

**Lemma 4.6.** (a) *The set class*

$$\mathcal{H}_T := \{\langle T(A_t) \rangle : t \in \mathbb{R}\}$$

for a fixed  $T(x) = (F_1(x_1), \dots, F_d(x_d))$  is universally Glivenko–Cantelli and Donsker.

(b) *If  $C$  is Lebesgue absolutely continuous, then the set class*

$$\mathcal{D}_{\delta_0} := \{U_\delta(B_t) : \delta \in [0, \delta_0], t \in \mathbb{R}\}$$

is  $\mathbb{P}_C$ -Glivenko–Cantelli for any  $\delta_0 > 0$ .

(c) *If  $C$  is Lebesgue absolutely continuous and satisfies (16), then  $\mathcal{D}_{\delta_0}$  is  $\mathbb{P}_C$ -Donsker for any  $\delta_0 > 0$ .*

The proof of this auxiliary result is given in Section 4.1. Now we proceed with the proofs of Theorems 4.1 and 4.2.

*Proof of Theorem 4.1.* For the sake of simplicity, we will write  $\mu\langle B \rangle$  instead of  $\mu(\langle B \rangle)$  for any measure  $\mu$  on  $[0, 1]^d$ . According to (9), we have to show that  $\mathbb{P}_{C_n}\langle T_n(A_t) \rangle \rightarrow \mathbb{P}_C\langle T(A_t) \rangle$  uniformly in  $t \in \mathbb{R}$ . It is easy to see that

$$\begin{aligned} & |\mathbb{P}_{C_n}\langle T_n(A_t) \rangle - \mathbb{P}_C\langle T(A_t) \rangle| \\ & \leq \mathbb{P}_{C_n}(\langle T_n(A_t) \rangle \triangle \langle T(A_t) \rangle) + |\mathbb{P}_{C_n}\langle T(A_t) \rangle - \mathbb{P}_C\langle T(A_t) \rangle|, \end{aligned} \quad (18)$$

where  $\triangle$  denotes the symmetric difference:  $A \triangle B := (A \setminus B) \cup (B \setminus A)$ . According to Lemma 4.6(a), the set class  $\mathcal{H}_T$  is  $\mathbb{P}_C$ -Glivenko–Cantelli. Hence the second term in (18) converges to 0  $\mathbb{P}$ -a.s. uniformly in  $t \in \mathbb{R}$ .

Now consider the first term in (18) and denote

$$Y_n := \|T_n(x) - T(x)\|_\infty. \quad (19)$$

As the transformations  $T_n$  and  $T$  are componentwise non-decreasing,  $Y_n$  is a measurable random variable. Furthermore, symmetry arguments give us  $\|C_{i,n}^{\leftarrow} - \text{id}_{[0,1]}\|_\infty = \|C_{i,n} - \text{id}_{[0,1]}\|_\infty$ , where  $\text{id}_{[0,1]}(u) := u$  for  $u \in [0, 1]$ . Hence (12) and the classic Glivenko–Cantelli theorem for  $C_{i,n}$  yield  $T_n \xrightarrow{\text{a.s.}} T$  in  $l^\infty(\mathbb{R}^d)$ . This implies that  $Y_n \xrightarrow{\text{a.s.}} 0$ .

It is also easy to see that

$$\langle T_n(A_t) \rangle \triangle \langle T(A_t) \rangle \subset U_{Y_n}(B_t), \quad (20)$$

where  $U_\delta(B_t)$  is the set introduced in (13). Moreover, for any  $\delta > 0$  we have

$$P_{C_n}(U_\delta(B_t)) \leq P_C(U_\delta(B_t)) + |P_{C_n}(U_\delta(B_t)) - P_C(U_\delta(B_t))|. \quad (21)$$

As  $Y_n \xrightarrow{\text{a.s.}} 0$ , it suffices to show that for  $\delta \rightarrow 0$  both terms on the right hand side of (21) vanish with probability 1 uniformly in  $t$ . In particular, for the second term it suffices to show that for some  $\delta_0 > 0$

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}, \delta \in [0, \delta_0]} |P_{C_n}(U_\delta(B_t)) - P_C(U_\delta(B_t))| = 0 \quad \text{P-a.s.}$$

This follows from Lemma 4.6(b).

The first term on the right hand side of (21) vanishes due to the absolute continuity of the copula  $C$ . Indeed, let  $\varepsilon > 0$ . Since the density  $c$  of  $C$  is non-negative and  $\int c(u) d\lambda(u) = 1$ , there exists  $M > 0$  such that  $\int_{\{c > M\}} c(u) d\lambda(u) < \varepsilon/2$ . Then, for  $\delta \leq \varepsilon/(4dM)$ , Lemma 4.4 yields

$$\begin{aligned} P_C(U_\delta(B_t)) &\leq P_{C_n}(U_\delta(B_t) \cap \{c \leq M\}) + \frac{\varepsilon}{2} \\ &\leq M\lambda(U_\delta(B_t)) + \frac{\varepsilon}{2} \\ &\leq M2d\delta + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

That is,  $P_C(U_\delta(B_t)) \rightarrow 0$  for  $\delta \rightarrow 0$  □

*Proof of Theorem 4.2.* According to (18) and (20), we have that

$$\begin{aligned} \sqrt{n} |G_n^*(t) - G(t)| \\ \leq \sqrt{n} P_{C_n}(U_{Y_n}(B_t)) + \sqrt{n} |P_{C_n}(T(A_t)) - P_C(T(A_t))|. \end{aligned} \quad (22)$$

The second term in (22) is  $O_P(1)$  uniformly in  $t \in \mathbb{R}$  due to Lemma 4.6(a).

Now consider the first term in (22) and observe that

$$\begin{aligned} \sqrt{n} P_{C_n}(U_{Y_n}(B_t)) \\ = \sqrt{n} (P_{C_n}(U_{Y_n}(B_t)) - P_C(U_{Y_n}(B_t))) + \sqrt{n} P_C(U_{Y_n}(B_t)). \end{aligned} \quad (23)$$

Applying the classic Donsker Theorem (cf. van der Vaart and Wellner, 1996, Theorem 2.5.7) to  $C_{i,n}$ , we obtain that  $\|C_{i,n}^- - \text{id}_{[0,1]}\|_\infty = \|C_{i,n} - \text{id}_{[0,1]}\|_\infty$  is  $O_P(n^{-1/2})$ . Hence assumption (14) yields  $Y_n = O_P(n^{-1/2})$ , and assumption (15) implies that the second term in (23) is  $O_P(1)$ .

Let  $Z_n$  denote the the first term in (23). As  $Y_n = o_P(1)$ , we have

$$Z_n = 1\{Y_n \leq \delta_0\} \sqrt{n} (P_{C_n}(U_{Y_n}(B_t)) - P_C(U_{Y_n}(B_t))) + o_P(1).$$

for any  $\delta_0 > 0$ . Now assumption (16) and Lemma 4.6(c) imply that  $Z_n$  is weakly convergent, and hence  $O_P(1)$ . □

The following corollary allows to replace the empirical marginal distributions  $F_{i,n}$  in Theorems 4.1 and 4.2 by any other consistent approximations of the true, unknown margins  $F_i$ .

**Corollary 4.7.** *Let  $F_{i,n}$ ,  $i = 1, \dots, d$ ,  $n \in \mathbb{N}$ , be arbitrary distribution functions on  $\mathbb{R}$ , and let  $G_n^*(t) := \mathbb{P}_{F_n^*}(A_t)$  with  $F_n^*(x) := C_n^*(F_{1,n}(x_1), \dots, F_{d,n}(x_d))$ .*

- (a) *If  $F_{i,n}$  satisfy (12) and  $C$  is absolutely continuous, then  $\|G_n^* - G\|_\infty \rightarrow 0$  P-a.s.*
- (b) *If  $F_{i,n}$  satisfy (14) and  $C$  satisfies (15) and (16), then  $\|G_n^* - G\|_\infty = O_P(n^{-1/2})$ .*

*Proof.* Part (a). For each  $F_{i,n}$  there is an approximation  $\tilde{F}_{i,n} : \mathbb{R} \rightarrow \{0, \frac{1}{n}, \dots, 1\}$  that minimizes  $\|F_{i,n} - \tilde{F}_{i,n}\|_\infty$ . It is obvious that  $\|F_{i,n} - \tilde{F}_{i,n}\|_\infty \leq 1/n$ . Hence the estimator  $\tilde{G}_n(t) := C_n^*(\langle \tilde{T}_n(A_t) \rangle)$  with  $\tilde{T}_n(A) := (C_{1,n}^{\leftarrow} \circ \tilde{F}_{1,n}(x_1), \dots, C_{d,n}^{\leftarrow} \circ \tilde{F}_{d,n}(x_d))$  satisfies the assumptions of Theorem 4.1, and therefore  $\|\tilde{G}_n - G\|_\infty \xrightarrow{\text{a.s.}} 0$ . Moreover,

$$\begin{aligned} |G_n^*(t) - \tilde{G}_n(t)| &\leq \mathbb{P}_{C_n}(\langle \tilde{T}_n(A_t) \rangle \triangle \langle T_n(A_t) \rangle) \\ &\leq \mathbb{P}_{C_n}(U_{\|\tilde{T}_n - T_n\|_\infty + \|T_n - T\|_\infty}(B_t)). \end{aligned} \quad (24)$$

As  $\|\tilde{T}_n - T_n\|_\infty \xrightarrow{\text{a.s.}} 0$  and  $\|T_n - T\|_\infty \xrightarrow{\text{a.s.}} 0$ , the term (24) vanishes with probability 1 uniformly in  $t \in \mathbb{R}$  analogously to the first term in (21). This yields  $\|G_n^* - G\|_\infty \xrightarrow{\text{a.s.}} 0$ .

Part (b). If  $F_{i,n}$  satisfy (14), then so do  $\tilde{F}_{i,n}$ . Hence Theorem 4.2 yields  $\|\tilde{G}_n - G\|_\infty = O_P(n^{-1/2})$ . Furthermore, (24) implies that

$$|\tilde{G}_n(t) - G_n^*(t)| \leq |\mathbb{P}_{C_n}(U_{\tilde{Y}_n}(B_t)) - \mathbb{P}_C(U_{\tilde{Y}_n}(B_t))| + \mathbb{P}_C(U_{\tilde{Y}_n}(B_t)) \quad (25)$$

for  $\tilde{Y}_n := \|\tilde{T}_n - T\|_\infty + \|T_n - T\|_\infty$ . As  $\tilde{Y}_n = O_P(n^{-1/2})$ , assumption (15) implies that  $\sup_{t \in \mathbb{R}} \mathbb{P}_C(U_{\tilde{Y}_n}(B_t)) = O_P(n^{-1/2})$ . The first term on the right hand side of (25) is  $O_P(n^{-1/2})$  uniformly in  $t \in \mathbb{R}$  due to Lemma 4.6(c).  $\square$

*Remark 4.8.* (a) Compared to  $G_n^*$ , the multivariate distribution function  $F_n^*(x)$  obtained by plugging  $F_{i,n}$  into  $C_n^*$  is much easier to handle. The deeper reason here is that  $F_n^*(x)$  can be written as the empirical measure  $\mathbb{P}_{C_n}$  indexed with random elements of the rectangle set class  $\mathcal{R}_d$ . In particular, if  $F_{i,n}$  are defined according to (2), then, analogously to (9), we have

$$\begin{aligned} F_n^*(x) &= C_n^*(F_{1,n}(x_1), \dots, F_{d,n}(x_d)) \\ &= C_n(C_{1,n}^{\leftarrow} \circ F_{1,n}(x_1), \dots, C_{d,n}^{\leftarrow} \circ F_{d,n}(x_d)) = \mathbb{P}_{C_n}(\langle T_n(x) \rangle). \end{aligned} \quad (26)$$

As mentioned above,  $\mathcal{R}_d$  is VC, and hence universally Donsker and Glivenko–Cantelli. Thus, due to  $\langle T_n(x) \rangle \in \mathcal{R}_d$ , we can apply standard results to  $F_n^*$ .

- (b) To prove strong consistency of  $F_n^*$ , recall that any copula is a Lipschitz function with Lipschitz constant 1 (cf. Nelsen, 2006, Theorem 2.2.4). Therefore (26) yields

$$\|F_n^* - F\|_\infty \leq \|C_n - C\|_\infty + Y_n. \quad (27)$$

As mentioned below (19), assumption (12) implies that  $Y_n \xrightarrow{\text{a.s.}} 0$ . Hence  $\|F_n^* - F\|_\infty \rightarrow 0$  P-a.s. due to the classic Glivenko–Cantelli Theorem for empirical distribution functions. The extension to general  $F_{i,n}$  is analogous to Corollary 4.7(a).

- (c) In the proof of Theorem 4.2 it is shown that assumption (14) entails  $Y_n = O_P(n^{-1/2})$ . Hence the  $O_P(n^{-1/2})$  convergence rate for  $F_n^*$  follows from (27) and the classic Donsker Theorem for empirical distribution functions. The extension to general  $F_{i,n}$  is analogous to Corollary 4.7(b).
- (d) If  $F_{i,n}$  satisfy a functional CLT, then the functional Delta method yields a functional CLT for  $F_n^*$ , with precise asymptotic variance – see van der Vaart and Wellner (1996, Lemma 3.9.28) and Segers (2012) for further details. Unfortunately, this does not imply a functional CLT for  $G_n^*$ , as  $G_n^*$  is obtained by indexing  $P_{C_n}$  with a totally different set class.

*Remark 4.9.* Theorems 4.1 and 4.2, along with all their extensions and corollaries, also apply to multivariate models generated by plugging empirical margins  $F_{i,n}$  directly into the copula  $C$ . According to Remark 3.2(d), the resulting estimator of  $G(t)$  can be written as  $P_{C_n}(\langle \tau_n(A_t) \rangle)$  with  $\tau_n(x) = (F_{1,n}(x_1), \dots, F_{d,n}(x_d))$ . Since  $G_n^* = P_{C_n}(\langle T_n(A_t) \rangle)$ , extension of convergence results to  $P_{C_n}(\langle \tau_n(A_t) \rangle)$  is straightforward. A closer look at the proof of Corollary 2.4 suggests that same is true for the convergence of  $P_{C_n}(\langle \rho_n^\leftarrow \circ T(A_t) \rangle)$  uniformly in  $t \in \mathbb{R}$ , where  $T(x) = (F_1(x_1), \dots, F_d(x_d))$  and  $\rho_n(x) = (C_{1,n}(x_1), \dots, C_{d,n}(x_d))$ . This setting corresponds to the combination of exact margins with the empirical copula  $C_n^*$ .

The final result in this section generalizes all foregoing results to a broader class of aggregation functions. Revising the proofs above, it is easy to see that the only property of the component sum used there is that it is componentwise non-decreasing. This immediately yields the following extension.

**Corollary 4.10.** *Let a function  $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfy*

$$\Psi(x) \leq \Psi(y) \quad \text{if } x_i \leq y_i \text{ for } i = 1, \dots, d.$$

*Then all results stated above for the sum distribution  $G$  also hold for the distribution function  $G_\Psi$  of the aggregated random variable  $\Psi(X)$ . In particular, the estimator  $G_{\Psi,n}^*(t) := P_{F_n^*}(\{x \in \mathbb{R}^d : \Psi(x) \leq t\})$  converges P-a.s. to  $G_\Psi$  in  $l^\infty(\mathbb{R})$  under the assumptions of Theorem 4.1 and has convergence rate  $O_P(n^{-1/2})$  under the assumptions of Theorem 4.2.*

*Remark 4.11.* (a) It depends on the aggregation function  $\Psi$  whether the generalization stated above is advantageous. In some special cases even stronger

results are possible. If, for instance,  $\Psi(x) = \max\{x_1, \dots, x_d\}$ , then the convergence of  $G_{\Psi, n}^*$  in  $l^\infty$  is related to the uniform convergence of the empirical measure  $P_{C_n}$  on the rectangle set class  $\mathcal{R}_d \cap [0, 1]^d$ . As the latter set class is VC, one can derive a Donsker Theorem for  $G_{\Psi, n}^*$  with a precise asymptotic variance.

- (b) Another remarkable example is the *Kendall process*, which is obtained by taking the joint distribution function  $F$  as aggregating function  $\Psi$ . The resulting aggregated distribution function is  $H(t) := P(F(X) \leq t)$ . Using the notation from above, this means  $H := G_\Psi$  for  $\Psi = F$ . The aggregated distribution function  $H(t)$  can be estimated by the empirical distribution  $H_n := n^{-1} \sum_{j=1}^n 1\{F_n^*(\tilde{X}^{(j)}) \leq t\}$ , where  $\tilde{X}^{(j)}$  are the Iman–Conover synthetic variables defined in (3) and  $F_n^*$  is their empirical distribution function (cf. (4)). If the margins  $F_i$  are continuous, then  $F(X)$  has the same distribution as  $C(U)$  for  $U \sim C$ . Moreover,  $F_n^*(\tilde{X})$  can always be written as  $C_n^*(U)$  for  $U \sim C$ . Thus the distribution of the process  $\sqrt{n}(H_n(t) - H(t))$  does not depend on the margins  $F_i$ . In this case asymptotic normality is also available (cf. van der Vaart and Wellner, 2007; Ghoudi and Rémillard, 1998; Barbe et al., 1996).
- (c) In the general case, however, extensions of Theorems 4.1 and 4.2 indeed go beyond available convergence results for empirical multivariate distribution functions.

#### 4.1. Proofs of auxiliary results

*Proof of Lemma 4.4.* Denote

$$W_{\delta, t}^{(0)} := \langle (T(A_t)) + \delta(1, \dots, 1) \rangle \cap [0, 1]^d$$

and, subsequently,

$$W_{\delta, t}^{(i)} := \left( W_{\delta, t}^{(i-1)} - 2\delta e_i \right) \cap [0, 1]^d, \quad i = 1, \dots, d.$$

The notation  $A + x$  for  $A \subset \mathbb{R}^d$  and  $x \in \mathbb{R}^d$  represents a shift of the set  $A$ , i.e.,  $A + x := \{a + x : a \in A\}$ . Further, denote

$$V_{\delta, t}^{(i)} := \overline{W_{\delta, t}^{(i-1)}} \setminus W_{\delta, t}^{(i)}, \quad i = 1, \dots, d.$$

The boundaries of  $V_{\delta, t}^{(i)}$  are Lebesgue null sets, because any  $\langle A \rangle$  for  $A \subset [0, 1]^d$  is Lebesgue-boundary-less. Indeed, the construction of  $\langle A \rangle$  guarantees that if  $u \in \langle A \rangle$  and  $v \in [0, 1]^d$ , then  $v \leq u$  (componentwise) implies  $v \in \langle A \rangle$ . Analogously, if  $u \in [0, 1]^d \setminus \langle A \rangle$  and  $v \in [0, 1]^d$  with  $v \geq u$ , then  $v \in [0, 1]^d \setminus \langle A \rangle$ . This monotonicity property allows to cover the boundary  $\partial \langle A \rangle$  by  $O(\varepsilon^{1-d})$   $d$ -dimensional cubes with edge length  $\varepsilon$  for any  $\varepsilon > 0$ . The total volume of this coverage is  $O(\varepsilon)$ , so that sending  $\varepsilon \rightarrow 0$  we obtain  $\lambda(\partial \langle A \rangle) = 0$ . This implies that all sets  $W_{\delta, t}^{(j)}$  and  $V_{\delta, t}^{(i)}$  are Lebesgue-boundary-less.

It is obvious that  $U_\delta(B_t) \subset \overline{W_{\delta,t}^{(0)} \setminus W_{\delta,t}^{(d)}}$ . Moreover, the construction of  $V_{\delta,t}^{(i)}$  entails that

$$\overline{W_{\delta,t}^{(0)} \setminus W_{\delta,t}^{(d)}} = \bigcup_{i=1}^d V_{\delta,t}^{(i)}$$

and  $\lambda(V_{\delta,t}^{(i)}) \leq 2\delta$  for all  $i$ . This yields  $\lambda U_\delta(B_t) \leq 2d\delta$ .  $\square$

*Proof of Lemma 4.6.* According to van der Vaart and Wellner (1996, Theorem 2.4.1), a set class  $\mathcal{C}$  is P-Glivenko–Cantelli if the *bracketing number*  $N_{[\cdot]}(\varepsilon, \mathcal{C}, L_1(\mathbb{P}))$  is finite for any  $\varepsilon > 0$ . The number  $N_{[\cdot]}(\varepsilon, \mathcal{H}_T, L_1(\mathbb{P}))$  is the minimal amount of so-called  $\varepsilon$ -brackets  $[V, W]$  needed to cover  $\mathcal{C}$ . An  $\varepsilon$ -bracket  $[V, W]$  with respect to  $L_1(\mathbb{P})$  is a pair of sets satisfying  $V \subset W$  and  $\mathbb{P}(W \setminus V) \leq \varepsilon$ . A set class  $\mathcal{C}$  is covered by brackets  $[V_i, W_i]$ ,  $i = 1, \dots, N$ , if each  $A \in \mathcal{C}$  satisfies  $V_i \subset A \subset W_i$  for some  $i$ . The criterion cited above is stated in terms of function classes, but it easily applies to set classes by identifying sets with their indicator functions.

A sufficient condition for  $\mathcal{C}$  to be P-Donsker is

$$\int_0^\infty \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{C}, L_2(\mathbb{P}))} d\varepsilon < \infty \quad (28)$$

(cf. van der Vaart and Wellner, 1996, Section 2.5.2). The distance of two sets  $A$  and  $B$  in  $L_2(\mathbb{P})$  is related to their distance in  $L_1(\mathbb{P})$  via

$$d_{L_2(\mathbb{P})}(A, B) = \|1_A - 1_B\|_{L_2(\mathbb{P})} = d_{L_1(\mathbb{P})}^{1/2}(A, B).$$

Hence the  $L_2(\mathbb{P})$  bracketing entropy condition (28) is equivalent to

$$\int_0^\infty \sqrt{\log N_{[\cdot]}(\varepsilon^2, \mathcal{C}, L_1(\mathbb{P}))} d\varepsilon < \infty. \quad (29)$$

Part (a). The set class  $\mathcal{H}_T$  is the collection of all  $\langle T(A_t) \rangle$  for  $t \in \mathbb{R}$  with a fixed  $T$ . Since the sets  $A_t$  are increasing in  $t$ , and  $T$  is componentwise non-decreasing, we have  $\langle T(A_t) \rangle \subset \langle T(A_s) \rangle$  for  $t \leq s$ . Consequently,  $\mathcal{H}_T$  can be covered by  $O(1/\varepsilon)$  brackets of size  $\varepsilon$  with respect to  $L_1(\mathbb{P}')$  for any probability measure  $\mathbb{P}'$  on  $[0, 1]^d$ . The brackets  $[V_i, W_i]$  can be chosen as  $V_i = \langle T(A_{t_i}) \rangle$ ,  $W_i = \cup_{t < t_{i+1}} \langle T(A_t) \rangle$  with an appropriate finite sequence  $t_1 < \dots < t_N$ . If  $\mathbb{P}'\langle T(A_t) \rangle$  has jumps, then it may be difficult to choose  $t_i$  such that  $\mathbb{P}'(W_i \setminus V_i) = \varepsilon$  for all  $i$ . In this case we may have  $\mathbb{P}'(W_i \setminus V_i) < \varepsilon$  for some  $i$ , but the total number of brackets is still  $O(1/\varepsilon)$ . Thus we have

$$N_{[\cdot]}(\varepsilon, \mathcal{H}_T, L_1(\mathbb{P}')) = O(1/\varepsilon),$$

and  $\mathcal{H}_T$  is universally Glivenko–Cantelli. If the maximal bracket size is  $\varepsilon^2$ , then one needs  $O(1/\varepsilon^2)$  brackets to cover  $\mathcal{H}_T$ . This is sufficient for (29), and hence  $\mathcal{H}_T$  is universally Donsker.

*Remark 4.12.* It is also easy to show that the set class  $\mathcal{H}_T$  is VC with index 2. As the sets  $\langle T(A_t) \rangle$  are increasing in  $t$ , they cannot shatter any two-point set.

Let  $u^{(1)}, u^{(2)} \in [0, 1]^d$ , and let  $t_1, t_2 \in \mathbb{R}$  be such that for  $k = 1, 2$   $x^{(k)} \in \langle T(A_t) \rangle$  is equivalent to  $t \geq t_k$ . Without loss of generality let  $t_1 \leq t_2$ . Then  $\mathcal{H}_T$  cannot pick out  $\{u^{(2)}\}$ , and hence  $\mathcal{H}_T$  is VC. From here, universal Glivenko–Cantelli and Donsker properties follow if we verify  $P'$ -measurability of  $\mathcal{H}_T$  for any probability measure  $P'$  on  $[0, 1]^d$  (cf. van der Vaart and Wellner, 1996, Definition 2.3.3). This can also be done.

However, since the VC property of  $\mathcal{D}_{\delta_0}$  remains elusive, the proofs of Parts (b) and (c) are based on bracketing entropy. This is the reason why the proof of Part (a) presented above is also based on bracketing. It gives a short preview of the ideas used below.

Part (c). Fix  $\varepsilon > 0$ . As  $[\emptyset, [0, 1]^d]$  is a bracket of size 1 in  $L_1(P_C)$  covering any subset of  $[0, 1]^d$ , we can assume that  $\varepsilon < 1$ . For this  $\varepsilon$  we define

$$\gamma := \frac{\varepsilon}{8dK(\varepsilon/8d)}$$

and

$$\delta_1 := (\delta_0 - \gamma)_+.$$

Given an arbitrary  $t_0 \in \mathbb{R}$  and  $t_1 \geq t_0$ , consider the following sets:

$$\begin{aligned} W_{t_0, t_1} &:= \bigcup_{t \in [t_0, t_1]} U_{\delta_0}(B_t), \\ W'_{t_0, t_1} &:= \bigcup_{t \in [t_0, t_1]} U_{\delta_1}(B_t), \\ V_{t_0, t_1} &:= \bigcap_{t \in [t_0, t_1]} U_{\delta_1}(B_t). \end{aligned}$$

The bracket  $[V_{t_0, t_1}, W_{t_0, t_1}]$  covers all  $U_{\delta}(B_t)$  for  $\delta \in [\delta_0, \delta_1]$  and  $t \in [t_0, t_1]$ . The size of this bracket in  $L_1(P_C)$  equals

$$P_C(W_{t_0, t_1} \setminus W'_{t_0, t_1}) + P_C(W'_{t_0, t_1} \setminus V_{t_0, t_1}).$$

Denote  $J_{\eta} := [\eta, 1 - \eta]^d$  for  $\eta \in [0, 1/2)$ , and let  $\lambda$  be the Lebesgue measure on  $J_0 = [0, 1]^d$ . Then

$$P_C(W_{t_0, t_1} \setminus W'_{t_0, t_1}) \leq P_C(J_0 \setminus J_{\varepsilon/8d}) + \lambda(W_{t_0, t_1} \setminus W'_{t_0, t_1})K(\varepsilon/8d). \quad (30)$$

As  $C$  is a copula and has uniform marginal distributions, the first term on the right hand side satisfies

$$P_C(J_0 \setminus J_{\varepsilon/8d}) \leq \sum_{i=1}^d P_C(\{u \in [0, 1]^d : u_i \notin [\varepsilon/8d, 1 - \varepsilon/8d]\}) = \varepsilon/4.$$

To obtain an upper bound for  $\lambda(W_{t_0, t_1} \setminus W'_{t_0, t_1})$ , observe that

$$W_{t_0, t_1} \setminus W'_{t_0, t_1} \subset (\langle U_{\delta_0}(B_{t_1}) \rangle \setminus \langle U_{\delta_1}(B_{t_1}) \rangle) \cup (\langle U_{\delta_0}(B_{t_0}) \rangle_{\text{up}} \setminus \langle U_{\delta_1}(B_{t_0}) \rangle_{\text{up}})$$

where  $\langle B \rangle_{\text{up}} := \cup_{u \in B} [u, (1, \dots, 1)]$  is the upper layer of  $B$  in  $[0, 1]^d$ . Moreover, analogously to Lemma 4.4, one can obtain that

$$\lambda(\langle U_{\delta_0}(B_{t_1}) \rangle \setminus \langle U_{\delta_1}(B_{t_1}) \rangle) \leq d(\delta_0 - \delta_1) \leq d\gamma = \frac{\varepsilon}{8K(\varepsilon/8d)}.$$

Same bound holds for  $\lambda(\langle U_{\delta_0}(B_{t_0}) \rangle_{\text{up}} \setminus \langle U_{\delta_1}(B_{t_0}) \rangle_{\text{up}})$ . Thus (30) yields

$$\mathbb{P}_C(W_{t_0, t_1} \setminus W'_{t_0, t_1}) \leq \frac{\varepsilon}{2}. \quad (31)$$

Consequently, as  $\mathbb{P}_C$  is absolutely continuous, we can choose  $t_1 > t_0$  such that either  $t_1 < \infty$  and  $\mathbb{P}_C(W_{t_0, t_1} \setminus V_{t_0, t_1}) = \varepsilon$  or  $\mathbb{P}_C(W_{t_0, t_1} \setminus V_{t_0, t_1}) < \varepsilon$  and  $t_1 = \infty$ . If  $t_1 < \infty$ , then (31) implies that

$$\mathbb{P}_C(W'_{t_0, t_1} \setminus V_{t_0, t_1}) \geq \frac{\varepsilon}{2}. \quad (32)$$

Proceeding in the same way as above, we obtain an increasing sequence  $t_0 < t_1 < t_2 < \dots$  that eventually terminates at  $\infty$ . Technical difficulties related to possible jumps of  $\mathbb{P}_C(W'_{t_i, t} \setminus V_{t_i, t})$  for  $t > t_i$  can be handled analogously to the proof of Part (a). A similar construction yields a decreasing sequence  $t_0 > t_{-1} > \dots$  that eventually terminates at  $-\infty$ .

We still have to show that the sequence  $t_k$  is always finite, i.e., that  $t_k$  indeed assumes  $\pm\infty$  for some  $k$ . Consider the sets  $S_k := W'_{t_k, t_{k+1}} \setminus U_{\delta_1}(B_{t_{k+1}})$  and  $S'_k := W'_{t_k, t_{k+1}} \setminus U_{\delta_1}(B_{t_k})$ . As  $S_k$  are disjoint for different  $k$ , we have  $\sum_k \mathbb{P}_C(S_k) \leq 1$  and, analogously,  $\sum_k \mathbb{P}_C(S'_k) \leq 1$ . It is obvious that

$$S_k \cup S'_k = W'_{t_k, t_{k+1}} \setminus (U_{\delta_1}(B_{t_k}) \cap U_{\delta_1}(B_{t_{k+1}})).$$

Furthermore, monotonicity of  $T$  implies that

$$U_{\delta_1}(B_{t_k}) \cap U_{\delta_1}(B_{t_{k+1}}) = \bigcap_{t \in [t_k, t_{k+1}]} U_{\delta_1}(B_t) \subset V_{t_k, t_{k+1}}.$$

This immediately yields  $W'_{t_k, t_{k+1}} \setminus V_{t_k, t_{k+1}} \subset S_k \cup S'_k$ . Applying (32), we obtain that  $\mathbb{P}_C(S_k) + \mathbb{P}_C(S'_k) \geq \varepsilon/2$  if  $t_k$  and  $t_{k+1}$  are finite. As the sum  $\sum_k (\mathbb{P}_C(S_k) + \mathbb{P}_C(S'_k))$  is bounded by 2, the length of the sequence  $(t_k)$  is bounded by  $4\lceil 1/\varepsilon \rceil + 4$ . Possible discontinuities of  $\mathbb{P}_C(W'_{t_k, t} \setminus V_{t_k, t})$  may increase this number at most  $\lceil 2/\varepsilon \rceil$  additional steps (cf. proof of Part (a)).

Thus we have shown that the set class  $\{U_{\delta}(B_t) : t \in \mathbb{R}, \delta \in [\delta_1, \delta_0]\}$  can be covered by  $6\lceil 1/\varepsilon \rceil + 4$  brackets of size  $\varepsilon$  in  $L_1(\mathbb{P}_C)$ . Defining  $\delta_k := (\delta_{k-1} - \gamma)_+$  for  $k = 2, 3, \dots$ , we reach 0 after  $\lceil \delta_0/\gamma \rceil = \lceil 8\delta_0 dK(\varepsilon/8d)/\varepsilon \rceil$  steps. As the arguments above apply to any interval  $[\delta_k, \delta_{k-1}]$ , we obtain a coverage for the set class  $\mathcal{D}_{\delta_0}$  and hence an upper bound for the bracketing number:

$$N_{[]}(\varepsilon, \mathcal{D}_{\delta_0}, L_1(\mathbb{P}_C)) = O(\varepsilon^{-2} K(\varepsilon/8d)).$$

According to (29), we need to verify that

$$\int_0^1 \sqrt{\log(K(\varepsilon^2/4d)\varepsilon^{-4})} d\varepsilon < \infty. \quad (33)$$

Changing the upper integral bound from  $\infty$  to 1 is justified by the fact that any set class can be covered by a single bracket of size 1.

As  $\int_0^1 \sqrt{\log(1/\varepsilon)} d\varepsilon < \infty$  and

$$\sqrt{\log(K(\varepsilon^2/4d))\varepsilon^{-4}} \leq \sqrt{\log K(\varepsilon^2/4d)} + 2\sqrt{\log(1/\varepsilon)},$$

the integrability condition (33) follows from the assumption (16).

Part (b). According to proof of part (c),  $N_{[]}(\varepsilon, \mathcal{D}_{\delta_0}, L_1(\mathbf{P}_C)) < \infty$  for any  $\varepsilon$ . Assumption (16) is needed only to verify (29).  $\square$

## 5. Examples

In this section we discuss the regularity assumptions of Theorems 4.1 and 4.2. The main results are stated in Propositions 5.2 and 5.3, verifying all regularity assumptions for bivariate Clayton copulas and bivariate Gauss copulas with correlation parameter  $\rho \geq 0$ . The case  $\rho < 0$  is treated in Proposition 5.3(c), which guarantees the convergence rate  $O_{\mathbf{P}}(n^{-1/2}\sqrt{\log n})$ . This is almost as good as  $O_{\mathbf{P}}(n^{-1/2})$ .

We start the discussion with a remark covering the mild integrability condition (16) and copulas with bounded densities.

*Remark 5.1.* (a) It is easy to see that  $K(\varepsilon) = O(\exp(\varepsilon^{-1+\eta}))$  for  $\varepsilon \rightarrow 0$  with some  $\eta > 0$  implies (16). In particular, any polynomial bound  $K(\varepsilon) = O(\varepsilon^{-k})$  for  $k > 0$  is sufficient.

(b) An immediate consequence of Lemma 4.4 is that all copulas with bounded densities satisfy all regularity conditions of Theorems 4.1 and 4.2. A particularly important copula example with a bounded density is the *independence copula*  $C(u) = \prod_{i=1}^d u_i$ . The Iman–Conover method with independence copula is a standard tool in applications with empirically margins based on real data. It is applied to generate multivariate samples with margins that are close to independent or to remove spurious correlations from multivariate data sets.

Unfortunately, many popular copulas, such as Gauss, Clayton, Gumbel, or  $t$  copulas, have unbounded densities. In particular, a bounded copula density implies that all tail dependence coefficients are zero. Thus applications related to dependence of rare events demand a deeper study of copulas with unbounded densities. The present paper provides two bivariate examples: the Clayton and the Gauss copula.

The bivariate Clayton copula with parameter  $\theta \in (0, \infty)$  is defined as

$$C_{\theta}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

The density  $c_{\theta}$  can be obtained by differentiation:

$$\begin{aligned} c_{\theta}(u_1, u_2) &= \partial_{u_2} \partial_{u_1} C_{\theta}(u_1, u_2) \\ &= (u_1^{-\theta} + u_2^{-\theta} - 1)^{-2-1/\theta} (\theta + 1)(u_1 u_2)^{-\theta-1}. \end{aligned} \quad (34)$$

The next result states that this copula family satisfies all regularity assumptions of Theorem 4.2.

**Proposition 5.2.** *Any bivariate Clayton copula  $C_\theta$  with  $\theta \in (0, \infty)$  satisfies (15) and (16).*

*Proof.* To verify (16), it suffices to show that  $K(\varepsilon)$  is polynomial (cf. Remark 5.1(a)). The density is given in (34). It is easy to see that  $(u_1^{-\theta} + u_2^{-\theta} - 1)^{-2-1/\theta} \leq 1$  for  $u \in (0, 1)^2$ . Hence the order of magnitude of  $K(\varepsilon)$  is determined by  $\sup_{u \in (\varepsilon, 1-\varepsilon)^2} (u_1 u_2)^{-\theta-1}$ , which is clearly polynomial.

To verify (15), recall that the proof of Lemma 4.4 used the following coverage of the set  $U_\delta(B_t)$ :

$$U_\delta(B_t) \subset \bigcup_{i=1}^d V_{\delta,t}^{(i)}.$$

Hence, for  $d = 2$ , we have

$$\mathbb{P}_{C_\theta}(U_\delta(B_t)) \leq \mathbb{P}_{C_\theta}(V_{\delta,t}^{(1)}) + \mathbb{P}_{C_\theta}(V_{\delta,t}^{(2)}). \quad (35)$$

The arguments that yield  $O(\delta)$  bounds for  $\mathbb{P}_{C_\theta}(V_{\delta,t}^{(i)})$  are symmetric in  $i = 1, 2$ , so that it suffices to consider  $V_{\delta,t}^{(1)}$ . For  $u_2 \in (0, 1)$ , denote

$$\bar{u}_1 = \bar{u}_1(u_2) := \sup \left\{ u_1 : (u_1, u_2) \in V_{\delta,t}^{(1)} \right\}$$

and

$$\underline{u}_1 = \underline{u}_1(u_2) := \inf \left\{ u_1 : (u_1, u_2) \in V_{\delta,t}^{(1)} \right\}.$$

It is easy to see that

$$\mathbb{P}_{C_\theta}(V_{\delta,t}^{(1)}) = \int_0^1 \int_{\underline{u}_1}^{\bar{u}_1} c_\theta(u_1, u_2) du_1 du_2.$$

Moreover, the construction of  $V_{\delta,t}^{(1)}$  implies that

$$\forall u_2 \in (0, 1) \quad \bar{u}_1(u_2) - \underline{u}_1(u_2) \leq 2\delta \quad \lambda\text{-a.s.}$$

Partial differentiation of  $\log c_\theta$  yields

$$\partial_{u_i} \log c_\theta(u_1, u_2) = \frac{(2\theta + 1)u_i^{-\theta-1}}{u_1^{-\theta} + u_2^{-\theta} - 1} - \frac{\theta + 1}{u_i}, \quad i = 1, 2.$$

Hence  $\partial_{u_1} \log c_\theta(u_1, u_2) = 0$  is equivalent to

$$\frac{(2\theta + 1)u_1^{-\theta}}{u_1^{-\theta} + u_2^{-\theta} - 1} = \theta + 1,$$

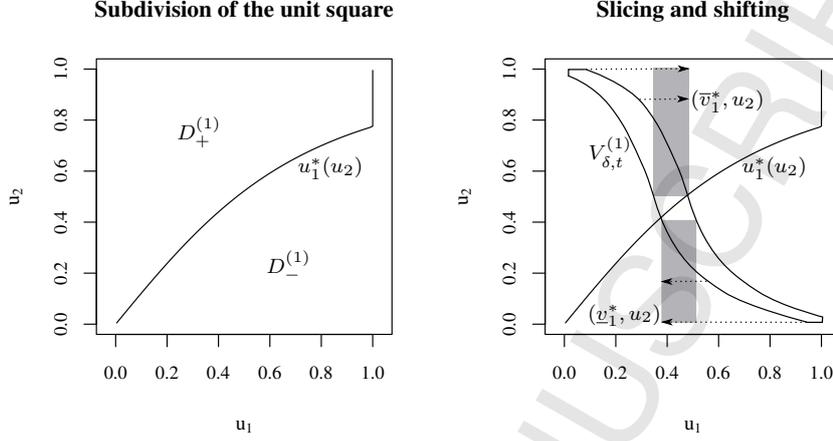


Figure 2: Bivariate Clayton copula: the curve  $u_2 \mapsto (u_1^*(u_2), u_2)$  subdividing the unit square (left) and the resulting slicing and shifting argument (right).

and for fixed  $u_2 \in (0, 1)$  the copula density  $c_\theta(u_1, u_2)$  attains its maximum at

$$u_1^* = u_1^*(u_2) := \min \left\{ \left( \frac{\theta + 1}{\theta} (u_2^{-\theta} - 1) \right)^{-1/\theta}, 1 \right\}.$$

Furthermore,  $c_\theta(u_1, u_2)$  is increasing in  $u_1$  for  $u_1 < u_1^*$  and decreasing in  $u_1$  for  $u_1 > u_1^*$ . Let  $D_+^{(1)}$  and  $D_-^{(1)}$  denote the corresponding sub-domains of  $(0, 1)^2$ :

$$D_+^{(1)} := \{u \in (0, 1)^2 : u_1 < u_1^*\}, \quad D_-^{(1)} := \{u \in (0, 1)^2 : u_1 > u_1^*\}.$$

An exemplary plot of the function  $u_1^*$  with resulting sets  $D_+^{(1)}$ ,  $D_-^{(1)}$  is given in Figure 2. Note that the function  $u_2 \mapsto u_1^*(u_2)$  is non-decreasing for any  $\theta > 0$ .

We will show that  $P_{C_\theta}(V_{\delta,t}^{(1)} \cap D_+^{(1)})$  and  $P_{C_\theta}(V_{\delta,t}^{(1)} \cap D_-^{(1)})$  are bounded by  $2\delta$ . Denote

$$I_+^{(1)} := \pi_2(V_{\delta,t}^{(1)} \cap D_+^{(1)}) \quad \text{and} \quad I_-^{(1)} := \pi_2(V_{\delta,t}^{(1)} \cap D_-^{(1)}),$$

where  $\pi_2$  is the projection on the second coordinate:  $\pi_2((u_1, u_2)) := u_2$ . Further denote  $S_*^{(1)} := \{u \in V_{\delta,t}^{(1)} : u_1 = u_1^*(u_2)\}$  and  $I_* := \pi_2(S_*^{(1)})$ . It is easy to see that  $I_*^{(1)} = \overline{I_+^{(1)}} \cap \overline{I_-^{(1)}}$ . Hence we can write

$$P_{C_\theta}(V_{\delta,t}^{(1)} \cap D_+^{(1)}) = \int_{I_+^{(1)} \setminus I_*^{(1)}} \int_{\underline{u}_1}^{\overline{u}_1} c_\theta(u) du_1 du_2 + \int_{I_*^{(1)}} \int_{\underline{u}_1}^{u_1^*} c_\theta(u) du_1 du_2 \quad (36)$$

Denote  $\bar{v}_1^* := \sup\{u_1 : (u_1, u_2) \in V_{\delta,t}^{(1)} \cap D_+^{(1)}\}$ . This definition implies that  $\bar{u}_1(u_2) \leq \bar{v}_1^*$  for all  $u_2 \in I_+^{(1)} \setminus I_*^{(1)}$ . Moreover, it is easy to see that  $\bar{v}_1^* = \sup\{u_1^*(u_2) : u_2 \in I_*^{(1)}\}$ . As  $u_1^*(u_2)$  is non-decreasing, this yields  $\bar{v}_1^* \leq u_1^*(u_2)$  for  $u_2 \in I_+^{(1)} \setminus I_*^{(1)}$ . This gives us

$$\forall u_2 \in I_+^{(1)} \setminus I_*^{(1)} \quad \bar{u}_1(u_2) \leq \bar{v}_1^* \leq u_1^*(u_2)$$

and, as a consequence,  $V_{\delta,t}^{(1)} \cap D_+^{(1)} \subset [0, \bar{v}_1^*] \times (0, 1)$ . If  $\bar{v}_1^* \leq 2\delta$ , then the uniform margins of the copula  $C_\theta$  immediately yield  $\mathbb{P}_{C_\theta}(V_{\delta,t}^{(1)} \cap D_+^{(1)}) \leq 2\delta$ . Hence, without loss of generality, we assume that  $\bar{v}_1^* > 2\delta$ .

As  $c_\theta$  is increasing in  $u_1$  on  $D_+^{(1)}$  and  $\bar{u}_1 - u_1 \leq 2\delta$   $\lambda$ -a.s., we obtain that

$$\forall u_2 \in I_+^{(1)} \setminus I_*^{(1)} \quad \int_{\underline{u}_1}^{\bar{u}_1} c_\theta(u) du_1 \leq \int_{\bar{v}_1^* - 2\delta}^{\bar{v}_1^*} c_\theta(u) du_1 \quad \lambda\text{-a.s.} \quad (37)$$

Moreover, it is easy to see that if  $u_2 \in I_*^{(1)}$  and  $u_1 \in [\underline{u}_1, u_1^*]$ , then  $u_1 \geq \bar{v}_1^* - 2\delta$ . This yields

$$\forall u_2 \in I_*^{(1)} \quad \int_{\underline{u}_1}^{u_1^*} c_\theta(u) du_1 \leq \int_{\bar{v}_1^* - 2\delta}^{\bar{v}_1^*} c_\theta(u) du_1. \quad (38)$$

Combining (37), (38), and (36), we obtain that

$$\mathbb{P}_{C_\theta}(V_{\delta,t}^{(1)} \cap D_+^{(1)}) \leq \mathbb{P}_{C_\theta}([\bar{v}_1^* - 2\delta, \bar{v}_1^*] \times [0, 1]) = 2\delta. \quad (39)$$

The latter equality is due to the uniform margins of the copula  $C$ .

The proof of (37) formalizes the idea of slicing the set  $V_{\delta,t}^{(1)}$  along  $u_1$  for every  $u_2 \in I_+^{(1)} \setminus I_*^{(1)}$  and shifting each slice  $[\underline{u}_1, \bar{u}_1] \times \{u_2\}$  upwards along  $u_1$  until this slice touches the point  $(\bar{v}_1^*, u_2)$  as illustrated in Figure 2. Since  $c_\theta$  is increasing in  $u_1$  on  $D_+^{(1)}$ , the transformed set has a larger probability under  $\mathbb{P}_{C_\theta}$ .

Using the fact that  $c_\theta$  is decreasing in  $u_1$  on  $D_-^{(1)}$ , one easily obtains the following analogue to (37):

$$\forall u_2 \in I_-^{(1)} \setminus I_*^{(1)} \quad \int_{\underline{u}_1}^{\bar{u}_1} c_\theta(u) du_1 \leq \int_{\underline{v}_1^*}^{\underline{v}_1^* + 2\delta} c_\theta(u) du_1 \quad \lambda\text{-a.s.},$$

where  $\underline{v}_1^* := \inf\{u_1 : (u_1, u_2) \in V_{\delta,t}^{(1)} \cap D_-^{(1)}\}$ . This result is obtained by slicing  $V_{\delta,t}^{(1)}$  along  $u_1$  for all  $u_2 \in I_-^{(1)} \setminus I_*^{(1)}$  and shifting each slice  $[\underline{u}_1, \bar{u}_1] \times \{u_2\}$  downwards along  $u_1$  until it touches the point  $(\underline{v}_1^*, u_2)$ , cf. Figure 2.

Similarly to (38), we have that

$$\forall u_2 \in I_*^{(1)} \quad \int_{\underline{u}_1^*}^{\bar{u}_1} c_\theta(u) du_1 \leq \int_{\underline{v}_1^*}^{\underline{v}_1^* + 2\delta} c_\theta(u) du_1.$$

Hence, analogously to (39), we obtain that

$$P_{C_\theta} \left( V_{\delta,t}^{(1)} \cap D_-^{(1)} \right) \leq 2\delta,$$

and, consequently,  $P_{C_\theta}(V_{\delta,t}^{(1)}) \leq 4\delta$ .

As the Clayton copula is symmetric in  $u_1$  and  $u_2$ , we also have  $P_{C_\theta}(V_{\delta,t}^{(2)}) \leq 4\delta$ . Thus (35) yields  $P_{C_\theta}(U_\delta(B_t)) \leq 8\delta$ , and condition (15) is satisfied.  $\square$

The next example is the bivariate Gauss copula  $C_\rho$ , defined as the copula of a bivariate normal distribution with correlation parameter  $\rho \in (-1, 1)$ . The parameter value  $\rho = 0$  yields the independence copula, which is the uniform distribution on the unit square  $(0, 1)^2$ . In this case all regularity conditions are satisfied (cf. Remark 5.1(b)).

Let  $\Phi$  denote the distribution function of the univariate standard normal distribution, and for  $u_1, u_2 \in (0, 1)$  let  $q_i = q_i(u_i) := \Phi^{-1}(u_i)$  denote the corresponding standard normal quantiles. Further, let  $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$  denote the  $2 \times 2$  correlation matrix corresponding to  $\rho$ . Then the bivariate Gauss copula  $C_\rho$  for  $\rho \neq 0$  can be written as

$$C_\rho(u_1, u_2) = \int_{-\infty}^{q_1} \int_{-\infty}^{q_2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}x^\top \Sigma^{-1}x\right) dx_1 dx_2,$$

where  $x$  is considered as a bivariate column vector and  $x^\top$  is the transposed of  $x$ . The copula density  $c_\rho$  can be obtained by differentiation:

$$\begin{aligned} c_\rho(u_1, u_2) &= \partial_{u_2} \partial_{u_1} C_\rho(u_1, u_2) \\ &= \det(\Sigma)^{-1/2} \exp\left(\frac{1}{2}q^\top (I - \Sigma^{-1})q\right), \end{aligned}$$

where  $q = (q_1, q_2)^\top$  and  $I$  is the identity matrix.

**Proposition 5.3.** (a) *The bivariate Gauss copula  $C_\rho$  always satisfies (16).*

(b) *If  $\rho \geq 0$ , then  $C_\rho$  satisfies (15).*

(c) *If  $\rho < 0$ , then the estimate  $G_n^*$  of the sum distribution  $G$  satisfies*

$$\|G_n^* - G\|_\infty = O_P\left(n^{-1/2}\sqrt{\log n}\right). \quad (40)$$

*Proof.* As mentioned above, the case  $\rho = 0$  is trivial. Hence we assume  $\rho \neq 0$ .

Part (a) It is easy to verify that  $\frac{1}{2}q^\top (I - \Sigma^{-1})q \leq Mq^\top q$  for any fixed  $\rho \in (-1, 0) \cup (0, 1)$  with some constant  $M = M(\rho) > 0$ . Hence, for  $\varepsilon \rightarrow 0$ ,

$$K(\varepsilon) = O\left(\exp\left(M|\Phi^{-1}(\varepsilon)|^2\right)\right).$$

Let  $\phi$  denote the standard normal density:  $\phi(t) := (2\pi)^{-1/2} \exp(-\frac{1}{2}t^2)$ . It is obvious that  $\Phi(t) = \int_{-\infty}^t \phi(s) ds < \int_{-\infty}^t -s\phi(s) ds = \phi(t)$  for  $t < -1$ . This yields

$$|\Phi^{-1}(\varepsilon)| < |\phi^{-1}(\varepsilon)| = \sqrt{-2 \log(\sqrt{2\pi\varepsilon})} \quad (41)$$

for sufficiently small  $\varepsilon > 0$ . Hence we obtain that

$$K(\varepsilon) = O\left(\exp\left(-2M \log\left(\sqrt{2\pi\varepsilon}\right)\right)\right) = O\left(\varepsilon^{-2M}\right), \quad \varepsilon \rightarrow 0.$$

Thus condition (16) follows from Remark 5.1(a).

Part (b). The verification of (15) for  $\rho > 0$  is analogous to Proposition 5.2. Due to

$$I - \Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} -\rho^2 & \rho \\ \rho & -\rho^2 \end{pmatrix}, \quad (42)$$

partial differentiation of  $\log c_\rho$  in  $u_1$  yields

$$\partial_{u_1} \log c_\rho(u_1, u_2) = \frac{1}{(1 - \rho^2)\phi(q_1)} (-\rho^2 q_1 + \rho q_2).$$

Hence  $c_\rho(u) = 0$  is equivalent to

$$u_1 = u_1^*(u_2) := \Phi(\Phi^{-1}(u_2)/\rho). \quad (43)$$

Moreover, it is easy to see that  $c_\rho(u)$  with  $\rho > 0$  is increasing in  $u_1$  if  $u_1 < u_1^*$  and decreasing in  $u_1$  if  $u_1 > u_1^*$ . This is illustrated in Figure 3. As  $u_1^*(u_2)$  is increasing, the slicing and shifting argument used in the proof of Proposition 5.2 applies here, and we obtain that  $P_{C_\rho}(V_{\delta,t}^{(1)}) \leq 4\delta$ . Due to the symmetry of  $c_\rho(u)$  in  $u_1$  and  $u_2$ , partial differentiation in  $u_2$  and the slicing and shifting method yield  $P_{C_\rho}(V_{\delta,t}^{(2)}) \leq 4\delta$ . Hence (35) gives us  $P_{C_\rho}(U_\delta(B_t)) \leq 8\delta$ .

Part (c). The situation for  $\rho < 0$  is different. The copula density  $c_\rho(u)$  is still increasing in  $u_1$  for  $u_1 \leq u_1^*$  and decreasing in  $u_1$  for  $u_1 > u_1^*$ , but the function  $u_1^*(u_2)$  is decreasing (cf. Figure 3). Thus (37) does not hold here. Instead of shifting each slice  $[\underline{u}_1, \bar{u}_1] \times \{u_2\}$  as in the proof of Proposition 5.2, one can shift it until it touches the point  $(u_1^*, u_2)$ . If  $u_1^* \in [\underline{u}_1, \bar{u}_1]$ , no shift is needed. That is, we replace the interval  $[\underline{u}_1, \bar{u}_1]$  by the interval  $[\underline{u}_1, \bar{u}_1] + \Delta$ , where

$$\Delta = \Delta(u_2) := \begin{cases} u_1^* - \bar{u}_1 & \text{if } \bar{u}_1 < u_1^*, \\ u_1^* - \underline{u}_1 & \text{if } \underline{u}_1 > u_1^*, \\ 0 & \text{else.} \end{cases}$$

It is easy to see that

$$\forall u_2 \in (0, 1) \quad \int_{\underline{u}_1}^{\bar{u}_1} c_\rho(u) du_1 \leq \int_{\underline{u}_1 + \Delta}^{\bar{u}_1 + \Delta} c_\rho(u) du_1.$$

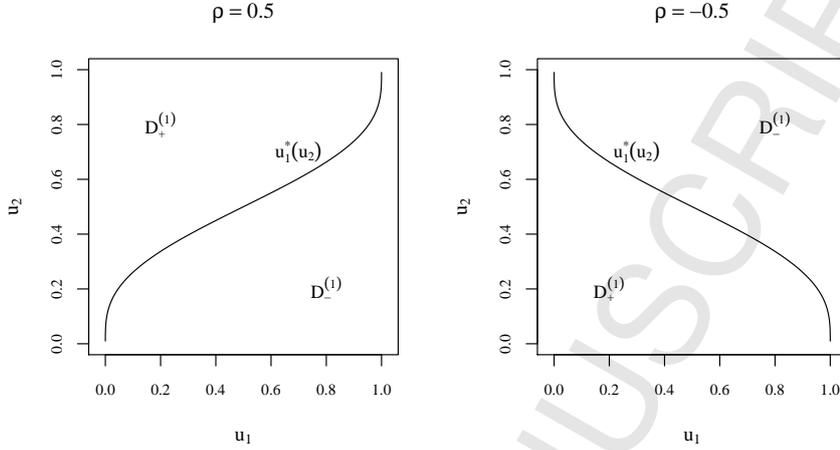


Figure 3: The curves  $u_2 \mapsto (u_1^*(u_2), u_2)$  for the bivariate Gauss copula with correlation  $\rho = \pm 0.5$ .

Integrating over  $u_2$ , we obtain that

$$P_{C_\rho} \left( V_{\delta,t}^{(1)} \right) \leq P_{C_\rho} (U^*),$$

where  $U^* := \{u \in (0, 1)^2 : |u_1 - u_1^*(u_2)| < 2\delta\}$ . It is easy to see that

$$\{u \in U^* : u_1^*(u_2) \notin (2\delta, 1 - 2\delta)\} \subset ([0, 4\delta] \cup [1 - 4\delta, 1]) \times [0, 1].$$

Hence, as  $C_\rho$  is a copula and has uniform margins, we obtain that

$$P_{C_\rho} (\{u \in U^* : u_1^*(u_2) \notin (2\delta, 1 - 2\delta)\}) \leq 8\delta.$$

Denote the remaining part of  $U^*$  by  $U_0^*$ :

$$U_0^* := \{u \in U^* : u_1^*(u_2) \in (2\delta, 1 - 2\delta)\}.$$

As  $u_1^*(u_2)$  maximizes  $c_\rho$  for fixed  $u_2$ , we have that

$$P_{C_\rho} (U_0^*) \leq 4\delta \int_{u_1^{*-1}(1-2\delta)}^{u_1^{*-1}(2\delta)} c_\rho(u_1^*(u_2), u_2) du_2.$$

Applying (42) and (43) we obtain that

$$c_\rho(u_1^*(u_2), u_2) = \sqrt{1 - \rho^2} \exp\left(\frac{1}{2} (\Phi^{-1}(u_2))^2\right).$$

Thus we need an upper bound for the integral

$$I(\delta) := \int_{u_1^{*-1}(1-2\delta)}^{u_1^{*-1}(2\delta)} \exp\left(\frac{1}{2} (\Phi^{-1}(u_2))^2\right) du_2.$$

The substitution  $u_2 = \Phi(t)$  yields

$$I(\delta) = \int_{\frac{1}{\rho}\Phi^{-1}(1-2\delta)}^{\frac{1}{\rho}\Phi^{-1}(2\delta)} \frac{1}{\sqrt{2\pi}} dt = \frac{\sqrt{2}}{\rho\sqrt{\pi}} |\Phi^{-1}(2\delta)|.$$

Finally, applying (41), we obtain that

$$I(\delta) \leq \frac{\sqrt{2}}{\rho\sqrt{\pi}} \sqrt{-2\log(2\delta\sqrt{2\pi})} = O\left(\sqrt{|\log \delta|}\right)$$

for  $\delta \rightarrow 0$ . This implies that

$$P_{C_\rho}\left(V_{\delta,t}^{(1)}\right) = O\left(\delta\sqrt{|\log \delta|}\right). \quad (44)$$

Symmetry arguments yield the same order of magnitude for  $V_{\delta,t}^{(2)}$  and, as a consequence, for  $P_{C_\rho}(U_\delta(B_t))$ .

Obviously, (44) is slightly weaker than  $O(\delta)$  in assumption (15). Revising the proof of Theorem 4.2, we see that (15) is used to guarantee that the second term in (23) is  $O_P(1)$ . Hence, replacing (15) by (44) in Theorem 4.2, one obtains (40).  $\square$

*Remark 5.4.* (a) It is currently an open question whether the weaker result of Proposition 5.3(c) reflects the reality or simply arises from the approximations used in the proof. However, it should be noted that the case  $\rho < 0$  is indeed more difficult than  $\rho > 0$ . For  $\rho < 0$  the curve  $(u_1^*(u_2), u_2)$  may be much closer to the set  $B_t$ , and hence  $P_{C_\rho}(U_\delta(B_t))$  may be substantially larger than for  $\rho > 0$ . In particular, if  $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2 \sim \mathcal{N}(0, 1/\rho^2)$ , then

$$\begin{aligned} \{(u_1^*(u_2), u_2) : u_2 \in (0, 1)\} &= T(\{x \in \mathbb{R}^2 : x_1 + x_2 = 0\}) \\ &= B_0 \cap (0, 1)^2. \end{aligned}$$

That is, for  $\rho < 0$  one can be confronted with the worst case when some  $B_t$  entirely falls into the area where the copula density is at its largest. A graphic example to this issue is given on the left hand side of Figure 1. The set  $B_0$  in that plot coincides with the set  $\{(u_1^*(u_2), u_2) : u_2 \in (0, 1)\}$  from the right hand side of Figure 3 ( $\rho = -0.5$ ). Such coincidence is not possible for  $\rho > 0$  or for the Clayton copula.

- (b) Intuitively speaking, this problem originates from the negative dependence for  $\rho < 0$ , where large values of  $X_1$  tend to be associated with small values of  $X_2$  and vice versa. As a consequence, the probability  $P(X_1 + X_2 \leq 0)$  is influenced by tail events. This effect is much weaker for Gauss copulas with  $\rho > 0$  and, analogously, for many other copulas with positive dependence.
- (c) The margins  $F_i$  may also influence the convergence of the estimated sum distribution function  $G_n^*$ . In insurance and reinsurance applications the

components  $X_i \sim F_i$  are often non-negative. In this case  $X_1 + X_2 \leq t$  implies that  $X_i \leq t$  for  $i = 1, 2$ , so that the tail events have no influence on  $G(t) = P(X_1 + X_2 \leq t)$  for moderately large  $t$ . The resulting sets  $B_t$  for  $t < \infty$  do not contain any internal points of the unit square that are close to the upper left or to the lower right vertex. That is,  $u \in B_t$  and  $u_1 > 1 - \varepsilon$  with a small  $\varepsilon$  implies  $u_2 = 0$ , and  $u_2 > 1 - \varepsilon$  implies  $u_1 = 0$ . These  $B_t$  avoid the areas where the density of the Gauss copula with  $\rho < 0$  is at its highest. Thus non-negative margins simplify the estimation of the function  $G$  for the Gauss copula with  $\rho < 0$  and, analogously, for many other copulas with negative dependence. An illustration to the different types of sets  $B_t$  is given in Figure 1.

## 6. Conclusions

This paper proves that Iman–Conover based estimates for the distribution function of the component sum are strongly uniformly consistent, and it provides sufficient conditions for the convergence rate  $O_P(n^{-1/2})$ . Besides the component sum, these results hold for any other componentwise non-decreasing function. The underlying mathematical problem goes beyond the classic uniform convergence results for empirical copulas. In the context of the Iman–Conover method, the primary cause for this technical difficulty is the implicit usage of empirical marginal distributions. Similar issues also arise in all multivariate models generated by plugging empirical margins into an exact copula or by plugging exact marginal distributions into an empirical copula. The marginal transformations involved in these applications complicate the resulting estimation problem in a way that does not allow to establish asymptotic normality of the estimated sum distribution. Therefore the weaker  $O_P(n^{-1/2})$  statement is quite the best result one can achieve.

The results proved here for the Iman–Conover method extend to all models generated by plugging empirical margins into an exact copula or by plugging exact margins into an empirical copula. The regularity conditions needed for the convergence rate  $O_P(n^{-1/2})$  are satisfied for all copulas with bounded densities, all bivariate Clayton copulas, and bivariate Gauss copulas with correlation parameter  $\rho > 0$ . The best convergence rate that could be established for the bivariate Gauss copulas with  $\rho < 0$  is  $O_P(n^{-1/2}\sqrt{\log n})$ . This result suggests that negative dependence may slow down the convergence of estimates. On the other hand, non-negative components  $X_i$ , as typical in insurance applications, may simplify the problem. The proof technique used for the bivariate Clayton copula applies to bivariate Gauss copulas with  $\rho > 0$  and may also work for other bivariate copulas with positive dependence. A straightforward generalization of this method to higher dimensions is, however, not feasible. Thus the question for the convergence rate in the  $d$ -variate case with  $d > 2$  and an unbounded copula density is currently open.

### Acknowledgements

The author would like to thank RiskLab, ETH Zurich, for financial support. Further thanks for helpful discussions and comments are due to Philipp Arbenz, Sara van de Geer, Jan Beran, Fabrizio Durante, and Paul Embrechts. Last but not least, the author thanks the anonymous referees, whose valuable comments and suggestions helped to improve this paper.

### References

- Arbenz, P., Embrechts, P., Puccetti, G., 2011. The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. *Bernoulli* 17, 562–591. doi:10.3150/10-BEJ284.
- Arbenz, P., Hummel, C., Mainik, G., 2012. Copula based risk aggregation through sample reordering. *Insurance: Mathematics and Economics* 51, 122 – 133. doi:10.1016/j.insmatheco.2012.03.009.
- Barbe, P., Genest, C., Ghoudi, K., Rémillard, B., 1996. On Kendall's process. *Journal of Multivariate Analysis* 58, 197 – 229. doi:10.1006/jmva.1996.0048.
- Billingsley, P., 1968. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York.
- 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, 274–292.
- Dudley, R.M., 1999. *Uniform Central Limit Theorems*. Cambridge University Press.
- Durante, F., Sarkoci, P., Sempì, C., 2009. Shuffles of copulas. *Journal of Mathematical Analysis and Applications* 352, 914 – 921. doi:10.1016/j.jmaa.2008.11.064.
- Embrechts, P., Puccetti, G., Rüschendorf, L., 2013. Model uncertainty and var aggregation. *Journal of Banking & Finance* 37, 2750 – 2764. doi:10.1016/j.jbankfin.2013.03.014.
- Fermanian, J.D., Radulović, D., Wegkamp, M., 2004. Weak convergence of empirical copula processes. *Bernoulli* 10, 847–860. doi:10.3150/bj/1099579158.
- Genest, C., Segers, J., 2010. On the covariance of the asymptotic empirical copula process. *Journal of Multivariate Analysis* 101, 1837 – 1845. doi:10.1016/j.jmva.2010.03.018.
- Ghoudi, K., Rémillard, B., 1998. Empirical processes based on pseudo-observations, in: *Asymptotic Methods in Probability and Statistics*. North-Holland, Amsterdam, pp. 171 – 197. doi:10.1016/B978-044450083-0/50012-5.
- Iman, R., Conover, W., 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Communications in Statistics - Simulation and Computation* 11, 311–334.
- Iman, R.L., 2008. *Latin Hypercube Sampling*. John Wiley & Sons, Ltd. doi:10.1002/9780470061596.risk0299.

- Kimeldorf, G., Sampson, A.R., 1978. Monotone dependence. *The Annals of Statistics* 6. URL: <http://www.jstor.org/stable/2958865>.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics, Chemical and Engineering Sciences* 21, 239–245.
- Mildenhall, S., 2005. Correlation and aggregate loss distributions with an emphasis on the Iman-Conover Method, in: *The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources*. Casualty Actuarial Society Forum, pp. 103–204. URL: <http://myn1.com/wp/ic.pdf>.
- Nelsen, R.B., 2006. *An introduction to copulas*. Second ed., Springer, New York.
- Owen, A.B., 1992. A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B* 54, 541–551.
- Packham, N., Schmidt, W.M., 2010. Latin hypercube sampling with dependence and applications in finance. *Journal of Computational Finance* 13.
- Rüschendorf, L., 1976. Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* 4, 912–923.
- Rüschendorf, L., 1983. Solution of a statistical optimization problem by rearrangement methods. *Metrika* 30, 55–61. doi:10.1007/BF02056901.
- Segers, J., 2012. Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* 18, 764–782. doi:10.3150/11-BEJ387.
- Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. *Technometrics* 29, 143–151. doi:10.2307/1269769.
- van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes*. Springer, New York. Corrected 2nd printing 2000.
- van der Vaart, A.W., Wellner, J.A., 2007. Empirical processes indexed by estimated functions, in: *Asymptotics: Particles, Processes and Inverse Problems*. Institute of Mathematical Statistics, Beachwood, Ohio. IMS Lecture Notes – Monograph Series, pp. 234–252.
- Vitale, R., 1990. On stochastic dependence and a class of degenerate distributions, in: Block, H., Sampson, A., Savits, T. (Eds.), *Topics in statistical dependence*, IMS Lecture Notes Monograph Series volume 16. Institute of Mathematical Statistics, Hayward, pp. 459–469.
- Wright, F.T., 1981. The empirical discrepancy over lower layers and a related law of large numbers. *The Annals of Probability* .

## Highlights

- Asynchronously sampled data can be endowed with any copula by a reordering technique.
- Popular since 1982, this method gets a rigorous convergence proof in the present paper.
- Related estimates of sum distribution functions converge uniformly with rate  $O_P(1/\sqrt{n})$ .
- The underlying problem is not covered by classic empirical process results.
- CLT fails in this case. This issue affects many real-world applications.