Identifiability and the Simplifying Assumption

Roger Cooke[1*] and Tim Bedford[2]

[1*]Department Mahematics, Delft University of Technology and Resources for the Future, Delft, Washington DC , The Netherlands, USA.

[2]Department of Management Science, University of Strathclyde, Glasgow, Scotland.

*Corresponding author(s). E-mail(s): cooke@rff.org;
Contributing authors: Tim.Bedford@Strath.ac.uk;

Abstract

A parameter in a statistical model is identifiable if it has a consistent estimator. We consider a conditional copula as a parameter in a regular vine model. As a baseline, no assumptions are made regarding bivariate margins. A non-simplified representation of a conditional copula expresses this copula as a convex combination of other copulas indexed by a measurable function of the conditioning variables. From this baseline three possibilities arise: (a) a non-simplified representation does not exist, (b) a non-simplified representation exists and is identifiable and (c) non-simplified representations exist but are not identifiable. The data generation process may either (1) sample the same conditional copula for each value of the conditioning variables, or (2) sample different conditional copulas for different values of the conditioning variables. The combinations (a.1), (b.1), (b.2), (c.1) and (c.2) are all possible. Moreover, different generation processes may produce the same joint distribution. In cases of (c) we may have strong evidence against the simplifying assumption, but need additional assumptions to render the non-simplified copula identifiable and define a consistent estimator. This article is an introductory exploration of these issues. Distinct continuous convex decompositions of the independent copula are found with different correlation functions. We find new extreme copulas on the unit square using bijective and non-bijective measure preserving maps, and conjecture that these constitute the set of extreme copula.

Keywords: Regular vine, copula, identifiability, simplifying assumption

1 Introduction

The history of vines or vine copulas is described in (Cooke, Joe, and Aas, 2010). The paper introducing vines [1], see also [2] associated the nodes of a regular vine with sets of (conditional) bivariate copulas and remarked that for applications it is convenient to choose copulas indexed by a single parameter such as rank correlation and to apply minimal information constraints - which sum along the nodes of the vine - relative to the product of one-dimensional margins. This implies that conditional copulas depend on the values of the conditioning variables only through the conditional CDF's of the copulated variables and not on the conditioning values directly. For example:

$$C_{F_{X|y},F_{Z|y}}(F_{X|y}(x),F_{Z|y}(z)|y)=C_{F_{X|y},F_{F_Z|y}}(F_{X|y}(x),F_{Z|y}(z)). \quad (1)$$

The right hand side, integrated over $dF(y)$, is called a *partial* or *simplified copula.* These two notions diverge in higher dimensions depending on the fitting over lower order conditional copulas, [3]. We skirt this issue here by restricting to 3-vines with one conditional copula and refer only to "partial copula", in analogy with partial versus conditional correlation. The constraint in (1) has become known as the *"simplifying assumption"*. The term "assumption" implies a proof burden; with equal justice one could speak of a "simplifying decision" to replace a complex random quantity with its expectation. In any event, a distinction is made between the graphical structure (regular vine) and the assignment of copulas to nodes (regular vine specification). Every multivariate density can be uniquely represented on every regular vine [4] by allowing the conditional copula to depend on the values of the conditioning variable. This result immediately shows that a regular vine without the simplifying assumption is not identifiable from the data. There is no "true" regular vine just as there is no "true" coordinate system.

Restricting the choice of copulas to a single parameter per copula restricts the modeling of joint dependence of n variables to a choice of $\binom{n}{2}$ scalars. This is of course a strong restriction. While preserving the polynomial complexity of the modeling, additional flexibility is obtained by allowing k parameter families of bivariate copula.

A different issue altogether is allowing the conditional copula to depend on the value of the conditioning variables. Regular vines with the simplifying assumption are known as "simplified vines". This blurs the distinction between the graphical structure (regular vine) and the assignment of copulas to nodes, but draws needed attention to this issue. Unlike general regular vines, simplified vines are not all created equal, some simplified vines may fit the data better than others. [5] cataloged large differences in fidelity among all simplified vines with up to 8 variables. The question of identifiability is also recast: given a simplified vine, is there a non-simplified representation which is identifiable from the data?

The literature to date on the simplifying assumption is recently summarized in [6]: [3] show that the subclass of simplified vines is dense in the class of all distributions for the supremum norm but not with regard to stricter norms. "In fact, we prove that the family of simplified copulas is even *nowhere dense* with respect to either of these four topologies[1] (emphasis in original)...". To illustrate the difference between measure theoretic and topological notions of size, the real line can be decomposed as a disjunct union of a Lebesgue null set and a countable union of nowhere dense sets and the set of differentiable functions on $[0,1]$ is a countable union of nowhere dense subsets of the set of continuous functions on $[0,1]$([7] theorem 1.6, Chap. 11).

The simplifying assumption can be tested [8] and [9]. Fitting non-simplified vines using the approaches of [10] and [11] requires large sample sizes to obtain stable estimates. [3] state "...since the (conditional) univariate marginal distribution functions $F_{1|3}(.|t)$ and $F_{2|3}(.|t)$ may fail to be continuous the (conditional) bivariate copulas $C^t_{1,2;3}$ are not unique in general. To the best of the authors' knowledge, [this] observation has not yet been addressed in the literature ...". The most recent review of the current state of play, [12], makes the important observation that the simplifying assumption constrains only the $\binom{n}{2} - n$ conditional copulas in the regular vine, which is a small fraction of all conditional bivariate copulas. [13] show that if $\{X_1, X_2\}$ are independent and $\{X_2, X_3\}$ are independent and if the conditional copula function $c_{1,3;2}$ does not depend on the value of X_2, then X_2 is independent of $\{X_1, X_3\}$. This strengthens an analogous result for partial correlations: $\rho_{1,2} = \rho_{2,3} = 0$ implies $\rho_{1,3;2} = \rho_{1,3}$.

This article studies the identifiability of non-simplified vines given the graphical structure and restricts attention to the simple D-vine of three uniform random variables (X, Y, Z). Section 2 defines identifiability of non-simplified conditional copulas in this context. Section 3 presents examples to illustrate the identifiability issue with respect to partial copulas on the $N \times N$ grid. Section 4 distinguishes identifying non-simplified representations of the partial copula and discovering the data generating process. Section 5 considers the identifiability of arbitrary copula on the $N \times N$ grid. Section 6 addresses the general case of bivariate copulas on $[0,1]^2$. Section 7 concludes.

2 Identifiability

By the Glivenko-Cantelli theorem, the unconditional copulas $C(X,Y), C(Y,Z)$ as well as the partial copula $\int C_{F_{X|y},F_{Z|y}}(F_{X|y}(x), F_{Z|y}(z))dF_Y(y)$ can be obtained as a.s. limits of functions of data (x_i, y_i, z_i), $i = 1 \dots N$ as $N \to \infty$. Indeed, the full joint distribution is identifiable, albeit under the curse of dimensionality, but because of [4], the $R-$vine representations are not. In the spirit of [14] we introduce [2]:

Definition: In *D-vine* $X - Y - Z$ with partial copula $\int C_{F_{X|y},F_{Z|y}}(F_{X|y}(x), F_{Z|y}(z))dF_Y(y)$, a copula valued function $C(y) = C_{F_{X|y},F_{F_Z|y}}(F_{X|y}(x), F_{Z|y}(z)\,|\,y)$ is an identifiable non-simplified representation of the partial copula if:

1. $E(C(Y)(x,z)) = \int C_{F_{X|y},F_{Z|y}}(F_{X|y}(x), F_{Z|y}(z))dF(y)$
2. $P\{\,y\,|C(y)(x,y) \neq E(C(Y)(x,z))\} > 0$
3. If $\tilde{C}(Y)$ is another copula valued function satisfying (1) and (2) then $\tilde{C}(Y) = C(Y)$ a.s.

[1]D1, weak conditional convergence, total variation and the Kullback-Leibler divergence
[2]For the definition of higher order partial copulas see [8].

Conditions (1), (2) say that $C(Y)$ is a non-simplified representation of the partial copula: the partial copula can be written as a non-trivial convex combination of the copulas $C(y)$. (3) says that this representation is a.s. unique. If the partial copula is extreme in the closed convex set of copulas on $[0,1]^2$ then condition (2) cannot be satisfied and the non-simplified copula does not exist. If the partial copula can be written as a convex combination of extreme copula in different $y-$measurable ways, then condition (3) fails. In this case non-simplified copulas exist but are not identifiable. "Identifiability is a natural and even a necessary condition: If the parameter is not identifiable, then consistent estimators cannot exist." [14], p. 62.

A non-simplified copula may be non-identifiable but nonetheless may be discoverable with exogenous information about the data generation process. An analogy with right censored life variables may be instructive. Suppose we observe the smaller of life variables X, Y and observe which it is. The conditional subsurvival functions $P(X > x|X < Y)$ and $P(Y > y|Y < X)$ are identifiable from the data but without further assumptions or exogenous information the unconditional survival functions are not. [15] showed that the survival functions are identifiable, whatever the conditional subsurvivor functions, *if* X and Y are assumed independent (see also [14] p.49). Examination of the data generation process may reveal that X is time to component failure and Y is preventive maintenance triggered by observed degradation during periodic inspection. This exogenous information constrains the shapes of the conditional subsurvival functions and may or may not be consistent with observed data. In case of consistency, the exogenous information enables identification of the survival functions which are then dependent [16]. In this case we can adopt either the independent model or the dependent model, whereby the dependent model is credited with passing an empirical test whereas the independent model is not.

3 Examples

This, and the following two sections, are restricted to copulas with densities measurable on the $N \times N$ grid of the unit square: Their densities are constant on the open squares, (i, j) with upper right corner at coordinates $(i/N, j/N), i = 1 \dots N, j = 1...N$. These are instances of checkerboard copulas studied in [17], which can uniformly approximate any d-dimensional copula (Theorem 4.1.5). We denote $(x, y, z)_{1,3} = (x, z)$.

We consider a simple D-vine on uniform variables X, Y, Z. For convenience we let (X, Y) and (Y, Z) be independent. The CDF of $X|\,y$ is $F_{X|Y=y}(x) = x$ and similarly for Z. The conditional copula $C_{F_{X|y}(x), F_{Z|y}(z)}\,|\,y$ can be written simply as $C(x, z|\,y)$. If $C(x, z|\,y)$ does not depend on y it is simply the (X, Z) copula, $C(X, Z)$, and can be chosen independently of y. We choose $C(X, Z)$ as the copula pictured in Figure 1 left pane with mass distributed uniformly on the shaded squares. $C(X, Z)$ can be written as a $(\frac{1}{2}, \frac{1}{2})$ convex combination of the co- and counter-monotonic copulas on the $N \times N$ grid. It can also be written as an a $(\frac{1}{2}, \frac{1}{2})$ convex combination of the green and red copulas in Figure 1, right pane. In fact, for each of the $2^5 - 2 = 30$ non-trivial subsets of $\{1, 2, 3, 4, 5\}$ we can form distinct $(\frac{1}{2}, \frac{1}{2})$ convex combinations which decompose the partial copula in Figure 1 left pane, into distinct pairs of extreme copulas.

We now define G as a copula valued function of y:

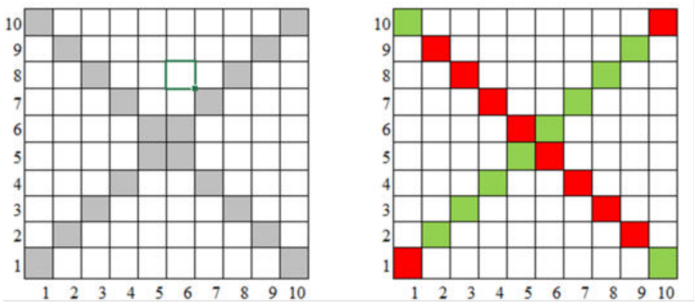

Fig. 1 Left: $C(x, z|\, y)$, (can be expresssed as a $(\frac{1}{2}, \frac{1}{2})$ convex combination of co- and counter monotonic copulas); Right: $C(x, z|\, y)$ expressed as a $(\frac{1}{2}, \frac{1}{2})$ convex combination of green and red copulas.

$$\begin{aligned} G(y) &= \text{co-monotonic copula if } (x, y, z)_{1,3} \text{ is on diagonal and} \\ &= \text{counter-monotonic copula if } (x, y, z)_{1,3} \text{ is on anti-diagonal} \end{aligned} . \tag{2}$$

Note that $G(y)$ is well defined except on the null set $(x, y, z) \cap (x', y, z'), x \neq x$' and/or $z \neq z$'. We can also define a different copula valued function of y:

$$\begin{aligned} H(y) &= \text{green copula if} (x, y, z)_{1,3} \in \text{supp green copula in Figure 1 Right and} \\ &= \text{red copula if} (x, y, z)_{1,3} \in \text{supp red copula in Figure 1 Right.} \end{aligned}$$

This gives a different non-simplified representation of the partial copula $C(X, Z)$. No amount of data from (X, Y, Z) will enable us to identify whether G or H is "correct". A non-simplified representation of the partial copula in Figure 1 is possible but not identifiable. Note that the co- and counter-monotonic copulas on $[0, 1]^2$ are extreme in the set of bivariate copulas. For such copulas non-simplified representations do not exist. More examples are given in Section 6.

Whether a non-simplified representation of the partial copula is identifiable depends on the partial copula. In Figure 2, a partial copula with uniform mass 1/20 on the colored squares can be given a non-simplified representation in only one way in terms of 2 copulas taking values 1/10 via the copula valued function

$$\begin{aligned} J(x, y, z) &= \text{green copula if } (x, y, z)_{1,3} \text{ is in a green square of Figure 2} \\ &= \text{red copula if } (x, y, z)_{1,3} \text{ is in a red square of Figure 2} \end{aligned} . \tag{3}$$

To appreciate the difference between Figures 1 and 2, define a graph whose nodes are the colored cells and with an edge between two cells if they share a common row or common column. This is a bipartite graph with chromatic index 2. A cycle is a non-empty path in which only the first and last elements are identical. Figure 2 has only one cycle, whereas Figure 1 right has five.

Figures 3 and 4 show examples of non-unique convex decompositions. Figure 3 shows a partial copula consisting of 3 cycles, each of which can be uniquely decomposed

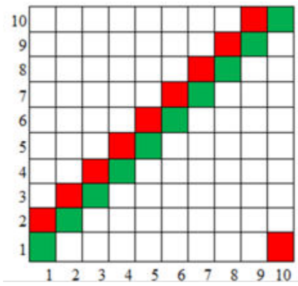

Fig. 2 Partial copula taking value 1/20 on shaded squares is uniquely decomposed into convex sum of two copulas.

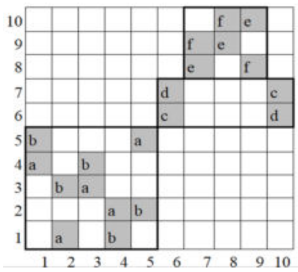

Fig. 3 Graph with 3 cycles.

into 2 disjunct intra-cycle matrices having one shaded square in each row and column as indicated by the letters. There are 2^3 ways of joining these to form pairs of copulas on the whole grid. The partial copula can be given a $(\frac{1}{2}, \frac{1}{2})$ convex decomposition into copulas taking value 1/10 on shaded squares in 2^3 ways. Figure 4 shows that uniform convex combinations of overlapping copulas result in non-uniform partial copulas. These possibilities are explored further in Section 5.

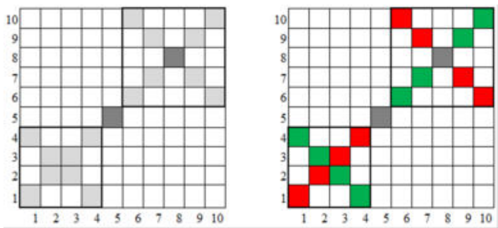

Fig. 4 Left: Partial copula with mass 1/20 and 1/10 on light and dark squares respectively; Right: Convex decomposition into green and red copula, intersecting on dark squares.

4 Identifying versus Discovering

This Section addresses the distinction between identifying a non-simplified conditional copula and discovering the data generation process. Consider generating processes:

1. For the copula in Figure 2: for each y a fair coin is flipped; if heads the green copula is chosen, otherwise the red copula is chosen. A square in the chosen copula is selected by lottery with probability 1/10 and (x, z) is sampled uniformly from that square
2. For the copulas in Figure 1: for some $A \subset [0, 1]$ of Lebesgue measure 1/2, for each y, if $y \in A$, the co-monotonic copula in Figure 1 left is chosen, otherwise the counter-monotonic copula in Figure 1 left is chosen. A square in the chosen copula is selected by lottery with probability 1/10 and (x, z) are sampled uniformly from that square
3. For the copulas in Figure 1: for some $B \subset [0, 1]$ of Lebesgue measure 1/2, for each y, if $y \in B$, the green copula in Figure 1 right is chosen, otherwise the red copula in Figure 1 right is chosen. A square in the chosen copula is selected by lottery with probability 1/10 and (x, z) are sampled uniformly from that square

Unlike the functions G and H in equations (2) and (3), these generating processes appeal to random mechanisms, they are not functions of the data. Process (1) corresponds to the simplified partial copula: even though different copulas are used on each sample and even though the non-simplified copula is identifiable, the choice of conditional copula is independent of y. For a given choice of $A \subset [0, 1]$ process (2) could be tested and rejected if data point (x, y, z) is found with $y \in A$ and $(x, y, z)_{1,3}$ on the anti-diagonal. Of course it is possible that (2) and (3) are both non-rejected for some A, B. Put $A = \{y|(x, y, z)_{1,3} \in \textit{supp comonotonic copula}\}$ and $B = \{y|(x, y, z)_{1,3} \in \textit{supp green copula}\}$.

Contrast this with the situation in Figure 2 where there is only one way to express the partial copula as a convex sum of two distinct copulas. Various random mechanisms could be proposed and tested for choosing between the red and green copulas in Figure 2 as functions of y. If we find a such mechanism that passes all tests then we may say that the non-simplified copula is identi*fied* but in any case the non-simplified copula is identi*fiable* since there is only one pair of copulas from which to choose.

Suppose process (1) is changed to:

(1'): In Figure 2: A colored square is chosen randomly and (x, z) is sampled uniformly from that square.

Processes (1) and (1') are physically distinct but yield the same distribution. In general the data generating process cannot be inferred from the data alone, but hypotheses regarding that process can be proposed and tested for consistency with the data.

5 Convex Decompositions of Copulas into Extreme Copulas

The Krein Milman theorem [18] states that a compact convex set is the closed convex hull of its extreme points. The set of all bivariate copulas on the closed unit square is convex and compact with the uniform metric; however, the set of extreme copulas has not been fully characterized [17]. For this reason we focus in this Section on copula densities measurable on an $N \times N$ grid of the unit square.

A doubly stochastic, or bistochastic matrix is a non-negative matrix whose row and column sums are 1. A permutation matrix is a bistochastic matrix in which each row and each column has exactly one non zero entry whose value is therefore 1. The permutation matrices are the extreme points in the convex compact set of bistochastic matrices. The Birkoff von Neumann theorem [19] , [20] , [21] states that a bistochastic matrix can be written as a (generally non-unique) convex sum of permutation matrices. The problem of finding a convex decomposition into a minimal set of permutation matrices is NP hard [22] [23]. The problem of finding all convex decompositions of a given bistochastic matrix into permutation matrices is therefore also NP hard. The volume of the set of bistochastic matrices is known up to dimension 10 [24]. [25] gives a method of generating bistochastic matrices which is an instance of iterative proportional fitting [26], [27] whose properties are well studied. A simple method [22] uses the circulant matrix of a probability N-vector to generate bistochastic matrices by rotating the vector's components as illustrated by a probability 4-vector in Figure 5. Note that the left-most column and the top row are the same, idem second row and second column, etc. The circulant of the N vector with one 1 and $(N-1)\,0$'s is a permutation matrix.

0.4	0.1	0.3	0.2
0.1	0.3	0.2	0.4
0.3	0.2	0.4	0.1
0.2	0.4	0.1	0.3

Fig. 5 Circulant of a probability 4-vector.

The set of copulas on the $N \times N$ grid is convex compact. Dividing entries of an $N \times N$ bistochastic matrix by N yields a copula. Assigning uniform mass of $1/N$ to the non-zero cells of an N-permutation matrix results in a *N-permutation copula*. N-permutation copulas are the extreme copulas in the compact convex set of copulas measurable on the $N \times N$ grid.

We focus on the case $N = 4$ for reasons that will appear shortly. A permutation matrix can be represented as a row 4-vector where each entry gives the row in which a "1" is located. For example, the identity permutation is represented by the vector $(4, 3, 2, 1)$, where 1 appears in the 4th row of the first column, in the 3rd row of the 2nd column etc. Circulants provide a handy way of organizing permutation matrices. The circulant of $(4, 3, 2, 1)$ is:

4	3	2	1
3	2	1	4
2	1	4	3
1	4	3	2

Fig. 6 Circulant of $(4, 3, 2, 1)$.

Figure 6 represents 4 permutations. Dividing the entries of the permutation matrices by 4, the circulant of $(4, 3, 2, 1)$ gives the four permutation copulas (a, b, c, d) in Figure 7.

copula a				copula b				copula c				copula d			
4	3	2	1	3	2	1	4	2	1	4	3	1	4	3	2
0.25							0.25			0.25			0.25		
	0.25			0.25							0.25			0.25	
		0.25			0.25			0.25							0.25
			0.25			0.25			0.25			0.25			

Fig. 7 Copulas represented by the circulant in Figure 6.

These 4 copulas together constitute a tiling of the $N \times N$ grid. They also constitute a convex decomposition of the uniform copula into 4 disjunct extreme copulas. Therefore, they also constitute a non-simplified representation of the uniform copula via a copula-valued function of sample data:

$$\begin{aligned} G(x, y, z) &= \text{copula a if } (x, y, z)_{1,3} \text{ is in an orange square} \\ &= \text{copula b if } (x, y, z)_{1,3} \text{ is in a blue square} \\ &= \text{copula c if } (x, y, z)_{1,3} \text{ is in a green square} \\ &= \text{copula d if } (x, y, z)_{1,3} \text{ is in a gray square} \end{aligned}$$

There are 6 permutations of $(3, 2, 1)$, each producing a distinct set of 4 permutations yielding 6 distinct tilings of the uniform distribution on the 4×4 grid with no common permutations. The $4! = 24$ permutations of $\{1, 2, 3, 4\}$ are broken down into $(4 - 1)! = 6$ disjunct convex decompositions of the uniform copula into extreme copula, each consisting of 4 disjunct extreme copula. 4 is the smallest integer N for which $(N - 1)! > N$.

Figure 8 with 16 uncolored rows and 24 columns shows the 24 extreme copulas corresponding to permutations, grouped by circulant. Tiling boundaries are bolded. For each tiling we have a distinct convex decomposition of the uniform copula into extreme copulas. The non-identifiability of convex decompositions of the uniform copula on N×N in terms of extreme copula grows with $(N - 1)!$. Note that because of the circulation property each column of Figure 8 encodes one extreme copula in its tiling.

Thus, the first column concatenates the columns of the first 4×4 matrix in the first tiling, the second column concatenates the second 4×4 matrix in the first tiling, etc.

tiling generators for extreme copula																							
4	**3**	**2**	**1**	**4**	**3**	**1**	**2**	**4**	**2**	**1**	**3**	**4**	**2**	**3**	**1**	**4**	**1**	**2**	**3**	**4**	**1**	**3**	**2**
0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0
0	0.25	0	0	0	0.25	0	0	0	0	0	0.25	0	0	0.25	0	0	0	0	0.25	0	0	0.25	0
0	0	0.25	0	0	0	0	0.25	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0	0	0	0	0.25
0	0	0	0.25	0	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0	0.25	0	0	0	0.25	0	0
0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25
0.25	0	0	0	0.25	0	0	0	0	0	0.25	0	0	0.25	0	0	0	0	0.25	0	0	0.25	0	0
0	0.25	0	0	0	0	0.25	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0	0	0	0	0.25	0
0	0	0.25	0	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0	0.25	0	0	0	0.25	0	0	0
0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0
0	0	0	0.25	0	0	0	0.25	0	0.25	0	0	0.25	0	0	0	0	0.25	0	0	0.25	0	0	0
0.25	0	0	0	0	0.25	0	0	0	0	0	0.25	0	0	0	0.25	0.25	0	0	0	0	0.25	0	0
0	0.25	0	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0	0	0	0	0	0.25	0	0	0	0.25
0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0	0	0.25	0	0
0	0	0.25	0	0	0	0.25	0	0.25	0	0	0	0	0	0	0.25	0.25	0	0	0	0	0	0	0.25
0	0	0	0.25	0.25	0	0	0	0	0	0.25	0	0	0	0.25	0	0	0	0	0.25	0.25	0	0	0
0.25	0	0	0	0	0	0	0.25	0	0	0	0.25	0.25	0	0	0	0	0	0.25	0	0	0	0.25	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

Fig. 8 24 extreme 4×4 copulas organized by circulant.

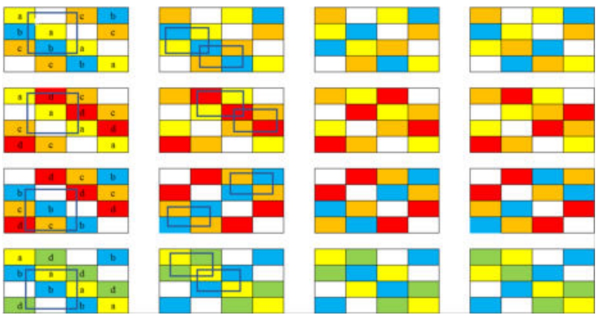

Fig. 9 For the leftmost tiling in Figure 8, the permutation copulas are named a, b, c, d , the 4 supports obtained by individually removing each successively are shown with 4 non-simplified representations for each. These representations are formed by switching row entries within a cycle, the cycles are indicated by the corners of the embedded rectangles.

Tilings give some insight into non-identifiability of equal convex combinations of permutations in a tiling. The union of the supports of three permutations in a tiling is the compliment of the support of the excluded permutation. Since this excluded permutation occurs only in one tiling, the union of the supports of the other three is also unique to this tiling. Label the permutation copulas in the leftmost tiling of Figure 8 as a, b, c, d. The rows in Figure 9 are non-simplified versions of the uniform copulas on the supports after excluding one of the permutation copula. These are generated by switching row and column entries in a cycle in the graphs of 2-copulas.

The cycles are shown as the corners of the embedded rectangles. The uniform copula on the support of 3-permutation copulas in a tiling can be represented as a uniform mixture of permutation copulas in 4 different ways.

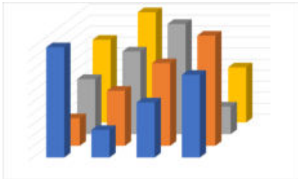

Fig. 10 Target copula.

Starting distribution assigns weight 1 to copula 1...24 (with RMS error)																							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1.26E-06	1.79E-06	3.96E-07	4.58E-07	1.01E-06	3.43E-07	3.84E-06	1.32E-06	1.18E-06	6.01E-06	1.02E-06	2.05E-06	2.48E-05	1.89E-05	2.32E-06	7.09E-07	1.22E-06	1.12E-07	4.52E-06	1.27E-06	1.35E-07	1.21E-07	7.18E-07	9.55E-07
0.3411	0.1286	0.1391	0.1552	0.1129	0.1338	0.1455	0.1267	0.1323	0.1498	0.1252	0.1100	0.1131	0.1171	0.1463	0.1191	0.0855	0.1419	0.0902	0.1219	0.1276	0.1208	0.1298	0.1511
0.0000	0.0114	0.0000	0.0000	0.0030	0.0000	0.0043	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0039	0.0000	0.0001	0.0000	0.0000	0.0027	0.0000	0.0000	0.0001	0.0000	0.0002
0.0457	0.0184	0.1125	0.0520	0.0273	0.0020	0.0063	0.0000	0.0000	0.0086	0.0001	0.0272	0.0211	0.0037	0.0070	0.0034	0.0005	0.0181	0.0008	0.0148	0.0006	0.0313	0.0055	0.0149
0.0955	0.0719	0.0884	0.2229	0.0675	0.0749	0.0758	0.0198	0.0787	0.0734	0.0174	0.0675	0.0694	0.0763	0.0752	0.0112	0.1136	0.0496	0.1161	0.0531	0.0183	0.0831	0.0843	0.0674
0.0065	0.0590	0.0856	0.0457	0.0950	0.0666	0.0781	0.0809	0.0235	0.0827	0.0299	0.0687	0.0553	0.0231	0.0567	0.0845	0.0528	0.0749	0.0537	0.0322	0.0268	0.0631	0.0750	0.0173
0.0000	0.0080	0.0000	0.0228	0.0043	0.0759	0.0068	0.0145	0.0036	0.0002	0.0154	0.0001	0.0113	0.0000	0.0000	0.0087	0.0001	0.0004	0.0118	0.0000	0.0000	0.0000	0.0000	0.0136
0.0462	0.0102	0.0018	0.0006	0.0196	0.0158	0.0741	0.0379	0.0001	0.0290	0.0000	0.0209	0.0003	0.0122	0.0031	0.0019	0.0001	0.0010	0.0004	0.0158	0.0377	0.0112	0.0031	0.0000
0.0554	0.0689	0.0560	0.0104	0.0802	0.0811	0.0939	0.1999	0.0838	0.0376	0.1049	0.0564	0.0521	0.0627	0.0931	0.0523	0.0975	0.0770	0.0006	0.0756	0.0516	0.0853	0.0875	0.0490
0.0023	0.0004	0.0000	0.0252	0.0000	0.0040	0.0063	0.0182	0.0747	0.0001	0.0140	0.0052	0.0012	0.0005	0.0086	0.0000	0.0033	0.0008	0.0127	0.0000	0.0157	0.0146	0.0000	0.0000
0.0494	0.0112	0.0001	0.0003	0.0196	0.0017	0.0241	0.0000	0.0236	0.0900	0.0389	0.0199	0.0168	0.0000	0.0009	0.0393	0.0001	0.0001	0.0009	0.0141	0.0000	0.0056	0.0190	0.0035
0.0503	0.0754	0.0550	0.0139	0.0579	0.0889	0.0448	0.1074	0.0793	0.0942	0.1998	0.0808	0.0842	0.1025	0.0442	0.0571	0.1028	0.0812	0.0040	0.0657	0.0553	0.0567	0.0512	0.1111
0.0006	0.0569	0.0864	0.0518	0.0683	0.0165	0.0786	0.0302	0.0665	0.0781	0.0846	0.0993	0.0607	0.0754	0.0186	0.0433	0.0613	0.0804	0.0576	0.0239	0.0838	0.0548	0.0228	0.0623
0.0007	0.0672	0.0880	0.0491	0.0625	0.0799	0.0073	0.0305	0.0250	0.0668	0.0853	0.0633	0.0959	0.0817	0.0775	0.0828	0.0595	0.0191	0.0573	0.0758	0.0305	0.0735	0.0198	0.0868
0.0001	0.0003	0.0000	0.0260	0.0010	0.0013	0.0067	0.0000	0.0000	0.0001	0.0163	0.0065	0.0070	0.0691	0.0091	0.0141	0.0194	0.0002	0.0040	0.0007	0.0157	0.0000	0.0025	0.0005
0.0557	0.0177	0.0002	0.0001	0.0191	0.0046	0.0001	0.0419	0.0284	0.0005	0.0001	0.0001	0.0165	0.0185	0.0812	0.0351	0.0004	0.0180	0.0004	0.0009	0.0000	0.0351	0.0001	0.0232
0.0575	0.0734	0.0549	0.0118	0.0830	0.0993	0.0471	0.0496	0.0693	0.0937	0.0533	0.0573	0.0814	0.0821	0.0875	0.1919	0.0031	0.0690	0.0988	0.0803	0.1045	0.0547	0.0829	0.0394
0.0001	0.0691	0.0299	0.1081	0.0700	0.0536	0.0656	0.0941	0.0556	0.0685	0.0928	0.0699	0.0687	0.0986	0.0682	0.0014	0.1920	0.0871	0.0851	0.0891	0.0001	0.0754	0.0860	0.0645
0.0499	0.0052	0.0102	0.0006	0.0106	0.0002	0.0017	0.0003	0.0001	0.0000	0.0007	0.0100	0.0000	0.0001	0.0107	0.0023	0.0224	0.0780	0.0241	0.0003	0.0003	0.0000	0.0000	0.0130
0.0013	0.0728	0.0306	0.1090	0.0717	0.0886	0.0685	0.0000	0.0923	0.0635	0.0001	0.0705	0.0757	0.0615	0.0643	0.0896	0.0815	0.0806	0.1915	0.0947	0.0906	0.0650	0.0630	0.0508
0.0429	0.0003	0.0134	0.0008	0.0009	0.0028	0.0040	0.0004	0.0014	0.0170	0.0000	0.0000	0.0091	0.0005	0.0001	0.0000	0.0235	0.0001	0.0247	0.0671	0.0004	0.0094	0.0000	0.0007
0.0491	0.0757	0.0574	0.0166	0.0597	0.0622	0.0971	0.0495	0.0888	0.0322	0.0528	0.0829	0.0656	0.0792	0.0426	0.1121	0.0069	0.0762	0.1010	0.0811	0.1993	0.0525	0.0894	0.0803
0.0007	0.0639	0.0897	0.0502	0.0616	0.0209	0.0564	0.0849	0.0726	0.0097	0.0288	0.0635	0.0649	0.0269	0.0793	0.0356	0.0580	0.0212	0.0565	0.0848	0.0839	0.0943	0.0785	0.0692
0.0008	0.0056	0.0000	0.0245	0.0045	0.0001	0.0062	0.0127	0.0000	0.0042	0.0000	0.0006	0.0024	0.0016	0.0001	0.0128	0.0147	0.0011	0.0028	0.0081	0.0158	0.0006	0.0770	0.0000
0.0481	0.0285	0.0009	0.0023	0.0001	0.0253	0.0005	0.0002	0.0003	0.0002	0.0398	0.0196	0.0274	0.0030	0.0255	0.0013	0.0010	0.0240	0.0022	0.0000	0.0414	0.0128	0.0225	0.0812

Fig. 11 Starting with weight 1 on extreme copula $1 \ldots 24$, solve for weights over the 24 extreme copula which yield the target distribution Figure 10, up to numerical error.

To illustrate the non-identifiability of a non-constant copula, the copula of Figure 5, is chosen as the target and illustrated in Figure 10. In general the problem of finding a mixture of extreme copula recovering the target copula has no unique solution; a solver's result will depend on the starting distribution of weights over extreme copulas. As a numerical exercise, a non-linear solver is started at 24 starting weight vectors, each giving weight 1 to one extreme copula. The solver converges numerically in each case to a distinct solution vector shown in Figure 11. The $RMSE$ is 6.5×10^{-6}. The mean correlation between the 24 solution vectors is 0.631. Interestingly, the rank of the matrix in Figure 8 is 10 and $10/16 = 0.625$.

Each weight vector in Figure 11 represents a decomposition of the target copula in Figure 10 into a distinct mixture of the 24 extreme copulas. Let A be the matrix in Figure 8 without the colored rows, let $B_i, i = 1, \ldots 24$ be the i^{th} column vector in the matrix in Figure 11 and let $\mathcal{I} = I_1, \ldots I_{24}$ be a measurable partition of the unit interval into measurable disjunct non-null subsets. For each such $\mathcal{I}$ the following copula valued function defines a distinct non simplified representation of the partial copula in Figure 9:

$G_{\mathcal{I}}(y) = A \times B_i$, if $y \in I_i, i = 1, \ldots 24$.

6 Continuous Convex Decompositions of the Independent Copula

The assumptions regarding the D-vine (X, Y, Z) in Section 3 and the notation $C_{x,z|y}$ remain in force. Continuous convex decompositions of the independent copula with different correlation functions are constructed using box copulas (Section 6.1) and shuffles of min (Section 6.2). As noted, the set of copulas on $[0,1]^2$ is convex and compact but the complete characterization of extreme copulas is still unknown [17], p.30. Proposition 6.1 describes extreme copulas based on N-permutation copulas, which are special cases of the "Dirac copulas" discussed in Section 6.3 where a more general result is proved.

Proposition 1 *For any integer N, replacing the uniform density of the non-empty cells of an $N-$ permutation copula with uniform mass $1/N$ on the cells' diagonal or anti diagonal is an extreme copula in the set of copulas on* $[0,1]^2$.

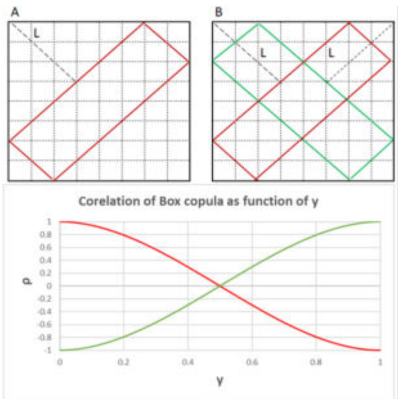

Fig. 12 (A): Family of box copulas parametrized by $L = 0 \dots \frac{\sqrt{2}}{2}$; (B): The copula in A together with its rotation by 90^o. Correlation functions for red and green copulas are shown below.

6.1 Box Copulas

Figure 12 (A) depicts a family of "box copulas" parametrized by the length of the dotted line $L = 0 \dots \frac{\sqrt{2}}{2}$. For each L the length of the boundary of the red box is $2\sqrt{2}$. Distributing unit mass uniformly on the red line, the density at each point on the boundary is $\frac{1}{2\sqrt{2}}$. Figure 12 (B) shows a "double box copula", the $(\frac{1}{2}, \frac{1}{2})$ convex combination in Figure 12 (A) (red) together with the result of rotating by 90^o (green).

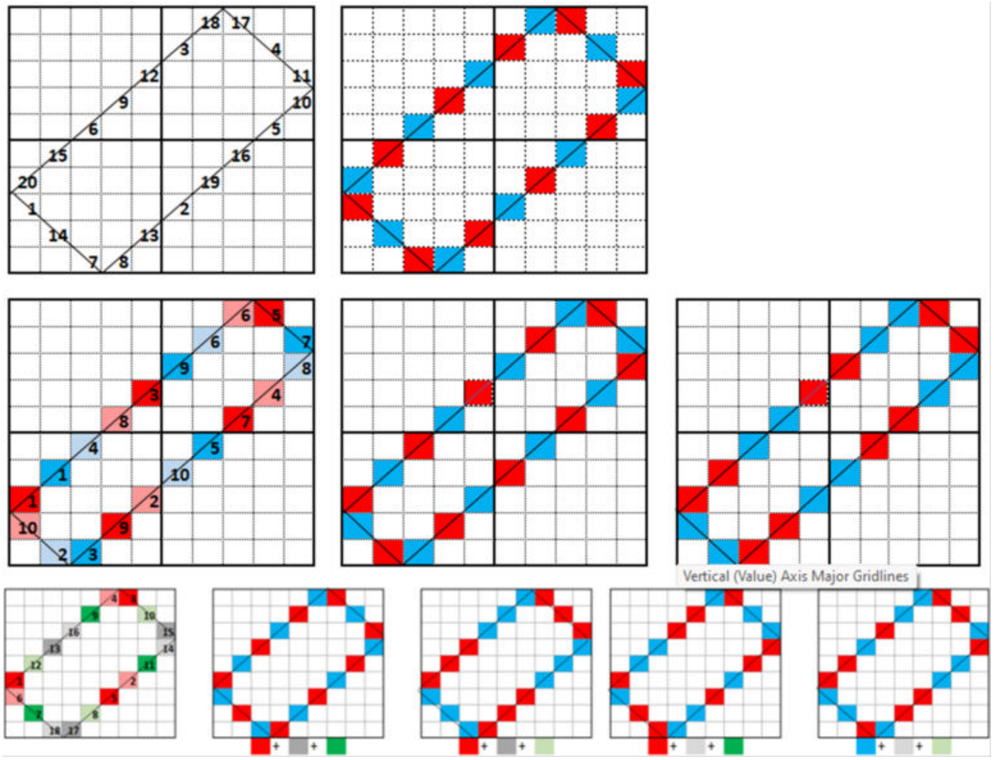

Fig. 13 Top row: a box copulas on the 10×10 grid with one cycle (identifiable). Middle row a box copula on the 10×10 grid with 2 cycles (non-identifiable). Bottom row: A box copula on a 9×9 grid with 3 cycles (non-identifiable).

Integrating either family (A) or (B) over L returns the independent copula[3]. Integrating over $L = 0, \ldots \alpha$, $\alpha < \frac{\sqrt{2}}{2}$ returns the 'diagonal band' copula used in risk analysis [28].

Setting $y = 1 - \frac{2L}{\sqrt{2}}$ a red box copula is associated with each value of y. For $y = 0$ this is the comonotonic copula, for $y = 1$, the counter monotonic copula. More generally, the red box copula $R(y)$'s correlation can be computed as (see apendix)

$$\rho(R(y)) = \frac{\frac{1}{3}y^3 - \frac{1}{2}y^2 + \frac{1}{3} - \frac{1}{4}}{\frac{1}{12}}; \; y \in [0,\, 1]. \tag{4}$$

One verifies that $\rho(R(y)) = -\rho(R(1-y))$. Therefore, the correlation $\rho(G(y))$ of the green copula in Figure 12 (B), is $-\rho(R(y))$. If the generating process samples a box

[3]As $L = 0 \ldots \frac{\sqrt{2}}{2}$, each point gets 'painted' twice with density $\frac{1}{2\sqrt{2}}$. Normalizing the integration interval, the mass at each point is $2 \times \frac{1}{2\sqrt{2}} \times \frac{2}{\sqrt{2}} = 1$

copula $R(y)$ from Figure 12 (A) with parameter $y = 1 - \frac{2L}{\sqrt{2}}$ conditional on y then :

$$\rho(C_{F_{X|y},F_{Z|y}}(F_{X|y}(x), F_{Z|y}(z)|y)) = \rho(R(y)). \tag{5}$$

The conditional correlations range from 1 to -1.

The box copulas are not extreme and the problem of resolving these copulas into extreme copulas is illustratively hard. Copulas on an $N \times N$ grid are dense in the set of copulas on $[0,1]^2$ and provide a line of attack. A box copula is "*inscribed on the $N \times N$ grid* if it can be represented with a copula on the $N \times N$ grid by replacing the uniform mass of the non-empty gird cells with equivalent mass on the diagonal or anti-diagonal of these cells. Connecting cells which share a common row or column yields a bipartite graph with chromatic index 2. Figure 13 gives an illustration. In the top row, starting with cell number 1, the cells are numbered successively by following the connections. Color the even numbered cells blue and the odd numbered cells red. The red and blue cells constitute a $(\frac{1}{2}, \frac{1}{2})$ convex decomposition of the box copula into two extreme copulas. In this case there is only one cycle in the graph so that the convex decomposition is unique and a non-simplified version of the box copula is identifiable. The middle row shows another possibility; the odd cells are colored dark red or dark blue while the even cells are colored light red or light blue. The red and blue cells constitute 2 cycles, and members of these cycles may be combined to form extreme copulas in two ways. A non-simplified version of the box copula is not identifiable. The bottom row shows a box copula with 3 cycles and 4 non simplified representations.

6.2 Shuffles of Min

Shuffles of min are extreme copulas formed by permuting rows and columns of the co-monotonic copula on a $N \times N$ grid, not necessarily equally spaced and possibly flipping the mass from the diagonal to the anti diagonal on any grid cell. [29] used them to prove that the independent copula can be approximated by shuffles of the min. [30] showed they are dense in the set of copulas with respect to the supremum norm. Since extreme copula are simplified, [3] concluded that simplified copulas are dense in the supremum norm.

Figure 14 shows 6 simple examples, A, B ("parallel copulas") and C, D, E, F ("hammer copulas") as function of parameter $h \in [0,1]$. $A(h)...F(h)$ denote the copulas $A \ldots F$ for parameter value h. Beneath each copula the correlation function $\rho(W(h)), W \in \{A, B, C, D, E, F\}$ is given and exhibited in the graph. The bottom row shows results of integrating the copulas over h. The calculation is sketched in the appendix. From the graphs we see that $\rho(C(h))...\rho(F(h))$ are invertible whereas $\rho(A(h)), \rho(B(h))$ are not.

The families $A(h)$, $B(h)$, $h \in [0,1]$ constitute distinct continuous convex decompositions of the independent copula into extreme copulas. Copulas $C(h) \ldots$ $F(h)$ decompose curious half triangular copulas, shown in Figure 14 bottom. The fact that these families realize all correlation values means that their mixtures can realize any measurable correlation function $\rho(h), h \in [0, 1]$.

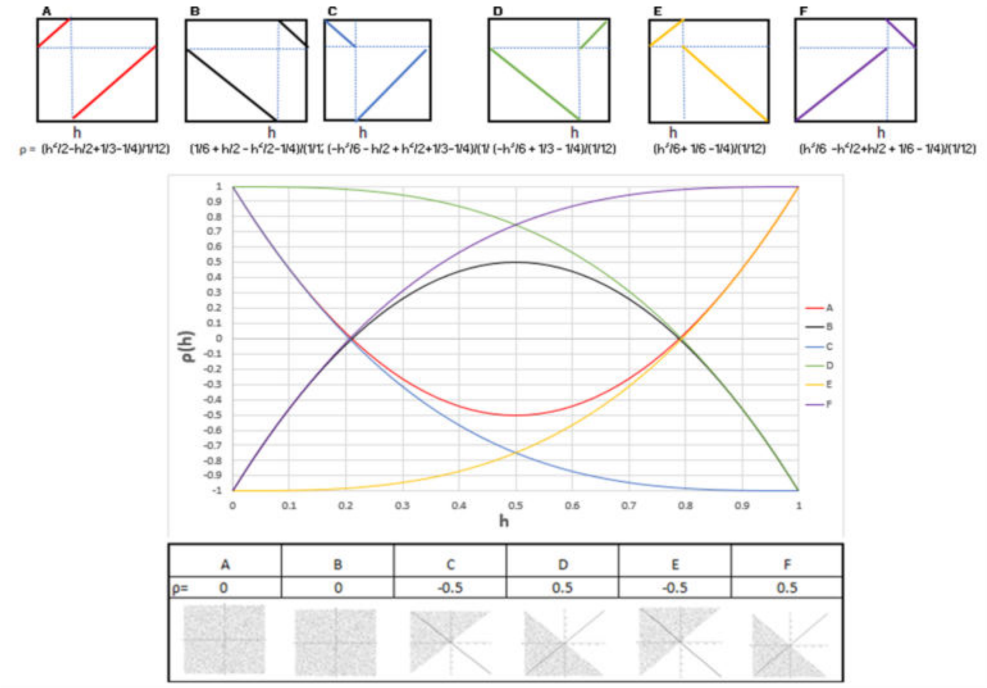

Fig. 14 Six shuffles of min, A, B are parallel copulas, C...F are hammer copulas. Correlation functions and partial copulas found by integrating over h are shown below.

Define $h_W : [-1, 1] \to [0, 1]$ as:

$$h_W(r) = \rho_W^{-1}(r), W \in \{C, D, E, F\}.$$

For each $W \in \{C, D, E, F\}$, $W(h_W(\rho(C_{X,Z|y}))$, defines a family of copulas satisfying

$$\rho(\int_{y=0}^{1} W(h_W(\rho(C_{X,Z|y}))dy) = \rho(C_{X,Z|y}); W \in \{C, D, E, F\}.$$

The results in this Section do not *approximate* but *represent* the independent copula as mixtures of parallel copulas (extreme) or box copulas (not extreme).

6.3 Dirac copulas

This Section introduces Dirac copulas which generalize those of the previous two subsections. Let λ denote Lebesgue measure, $\mathcal{F}$ the σ-field of Lebesgue measurable subsets of $\| =\ [0, 1]$, ϕ a measure preserving map $\| \to \|$ such that $\forall B \in \mathcal{F}, \lambda(\phi^{-1}(B)) = \lambda(B)$. If μ is a measure, $supp(\mu)$ denotes the support of μ (the intersection of all closed sets having full measure). Unless stated otherwise, all integrals are over $\|$.

The Dirac measure for $A \in \mathcal{F}$, $\delta_{xo}(A) = 1$ if $xo \in A$ and $= 0$ otherwise. This can also be written as

$$\int 1_A(x)\delta_{xo}(dx) = 1_A(xo).$$

$\delta_{xo}(A)$ is not absolutely continuous with respect to λ, indeed $\delta_{xo}(\{xo\}) = 1$. The heuristic shorthand $\delta_{xo}(dx) \simeq \delta_{xo}(x)dx$ suggests that $\delta_{xo}(x)$ is a Radon Nikodym derivative, which does not exist. For any continuous function f on $\|$, the portmanteau theorem says $\int f(s)\delta_{xo}(dx) = f(xo)$ [31].

The copulas in Figure 14 are Dirac copulas based on bijective measure preserving transformations. We generalize this to simply measure preserving and illustrate with measures which are measure-proportional on elements of a partition. Examples of this construction are given in [32] p.60. Let $0 = a_0 < a_1 < \ldots < a_n = 1$, Let $A_i = (a_{i-1}, a_i]$ and let $\phi_i : A_i \to [0, 1]$:

$$\phi_i(x) = \frac{x - a_i}{a_{i+1} - ai} \text{ or } = \frac{a_{i+1} - x}{a_{i+1} - a_i}. \tag{6}$$

The slope of ϕ_i may be positive or negative. It is easy to check that for $B \in \mathcal{F}, \lambda(\phi_i^{-1}(B)) = \lambda(A_i)\lambda(B)$. We now define

$$\phi(x) = \sum_{i=1}^{n} 1_{A_i}(x)\phi_i(x).$$

One verifies for $B \in \mathcal{F}$:

$$\lambda(\phi^{-1}(B)) = \lambda(\bigcup_i \phi_i^{-1}(B)) = \sum_i \lambda(B)\lambda(A_i) = \lambda(B).$$

We define a Dirac copula as the measure $c_\phi(x, y) = \delta_{\phi(x)}(dy)dx$; $x, y \in \|$. Nota bene, we cannot reverse the order of integration as the Fubini theorem doesn't hold and $\delta_{\phi(x)}(dy)$ depends on x, so we must integrate $\delta_{\phi(x)}(dy)$ first.

Proposition 2 *c_ϕ is an extreme copula.*

Proof We have

$$c_\phi(\|, \|) = \int\int 1_{\{\|,\|\}}(x, y)\delta_{\phi(x)}(dy)dx = \int 1_{\{\|,\|\}}(x, \phi(x))dx = 1.$$

To show that c_ϕ has uniform margins, choose $B \in \mathcal{F}$.

$$c_\phi(B, \|) = \int\int 1_{\{B,\|\}}(x, y)\delta_{\phi(x)}(dy)dx = \int 1_{\{B,\|\}}(x, \phi(x))dx = \int 1_B dx = \lambda(B).$$

$$c_\phi(\|, B) = \int\int 1_{\{\|,B\}}(x, y)\delta_{\phi(x)}(dy)dx = \int 1_{\{\|,B\}}(x, \phi(x))dx =$$
$$\int 1_B(\phi(x))dx = \int 1_{\{\phi^{-1}(B)\}}dx = \lambda(\phi^{-1}(B)) = \lambda(B)$$

since ϕ is measure preserving. This shows that c_ϕ is a copula. To show it is also extreme, suppose that $c_\phi = \alpha c_1 + (1-\alpha)c_2$, $1 > \alpha > 0$. $supp(c_\phi) = \{(x, \phi(x)) \mid x \in \|\} \in \mathcal{F}$. Let $A = supp(c_\phi) \setminus supp(c_1)$ and suppose $\lambda\{x|(x,\phi(x)) \in A\} > 0$. We have $c_\phi(A) = \alpha \times 0 + (1-\alpha)c_2$, contradicting $\alpha > 0$. Therefore, $supp(c_\phi) = supp(c_1) = supp(c_2)$. Let $B = \{x \mid c_\phi(B, \|) < c_1(B, \|)\}$ and suppose $\lambda(B) > 0$. $c_\phi(B, \|) = \lambda(B) < c_1(B, \|)$ so that c_1 does not have uniform margins. □

Remark 1 The above proof also goes through, albeit with a profusion of notation, if we replace equation (6) by broken linear functions as in the parallel or hammer copulas of Figure 14. In this case ϕ_i is still measure proportional and ϕ is still measure preserving.

We have shown that the set of extreme copulas on $[0, 1]^2$ is larger than the set of Dirac copulas with bijective measure preserving maps and includes more general measure preserving maps. We conjecture that this set is in fact the set of extreme copula on $[0, 1]^2$, though the proof has not yet been given.

We now further extend the construction given above, noting that the restriction that the partition of $[0, 1]$ into a finite partition $A_1, A_2, \ldots A_n$ can be weakened to a countably infinite non-intersecting closed intervals of $[0, 1]$ with Lebesgue measure 1. On each partition element A_i we define $\phi_i : A_i \to [0, 1]$ to be one of the two linear maps that map A_i bijectively to $[0, 1]$.

For convenience we are going to express real numbers in their base 3 representation $x = 0.x_1x_2x_3\ldots.$, where $x \in [0, 1]$ and each x_i is an integer taking the value 0, 1 or 2. The real number $x = \sum_{i=1}^{\infty} x_i/3^i$.

Following the construction given above we construct the Cantor copula as a probability measure on the graph of a piecewise linear function on $[0, 1]$ (Note that $1 = 0.2222222\ldots.$, where the sequence of 2's continuous indefinitely, when expressed to base 3).

Given $x \in [0, 1]$, $x = 0.x_1x_2x_3\ldots$, set $k(x)$ to be the smallest integer such that $x_{k(x)} = 1$. Clearly given a particular $x_1 \ldots x_{n-1}$ of 0's and 2's there is an interval of points x with base 3 representation $0.x_1 \ldots x_{n-1}1$ and for all such points x, $k(x) = n$.

For $x = 0.x_1x_2x_3\ldots.$ we define a function $\phi_{x_1 \ldots x_{k(x)}}(x) = (x - 0.x_1 \ldots x_{k(x)})x3^{k(x)}$, or $= 0$ if $k(x) = \infty$.

The set of points where $k(x)$ is finite has Lebesgue measure 1, as is easily seen since, for each $n \geq 1$ there are 2^{n-1} possible intervals of points for which $k(x) = n$, each of length 3^{-n}. On each of these intervals $\phi_{x_1 \ldots x_{k(x)}}(x)$ is locally a linear bijection from a closed set $A_{x_1 \ldots x_{k(x)}}$ onto $[0, 1]$. The total length of these intervals is $\sum_{n=1}^{\infty} 2^{n-1}/3^{-n} = 1$.

The construction of the Dirac copula given above goes through in this case as well, as is easy to see that the mass on any rectangle $A \times B$, where A is a subinterval of an $A_{x_1 \ldots x_{k(x)}}$ and B is a subinterval of $[0, 1]$ is equal to

$$\lambda(A \cap \phi^{-1}_{x_1 \ldots x_{k(x)}}(B).$$

Figure 15 displays the Cantor copula.

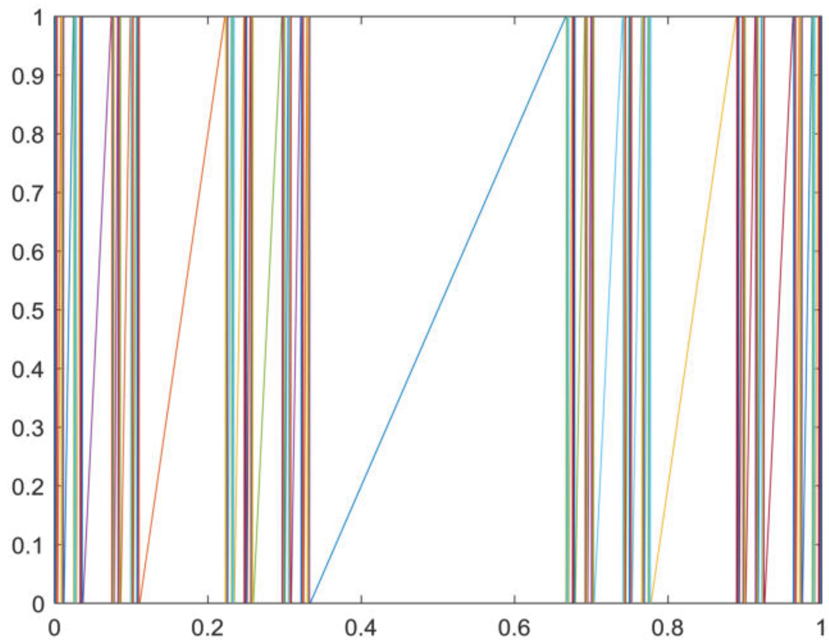

Fig. 15 The Cantor copula.

Note that this construction is based on the Middle Third Cantor Set $C \subset [0,1]$, which is well known to be "self similar" in the sense that it satisfies the property that $C = \psi_1(C)\psi_2(C)$ where $\psi_0(x) = x/3$ and $\psi_2(x) = 2/3 + x/3$, and is the unique closed set that satisfies that property [33] For $x = 0.x_1x_2x_3...$ we have $\psi_0(x) = 0.0x_1x_2x_3...$ and $\psi_2(x) = 0.2x_1x_2x_3...$, and so we see that the functions $(x, y) \mapsto (\psi_0(x), y)$ and $(x, y) \mapsto (\psi_2(x), y)$ maps the graph of the function onto itself. This explains the recurrence of structure with ever increasing slopes in the graph. A key feature of the Cantor copula that goes beyond the finite piecewise linear Dirac copulas discussed above, is that the gradients are unbounded. If we take the interval of $x = 0.x_1x_2x_3..x_n\ldots$ with no $x_i = 1$ for $i = 1, \ldots, n$, the slope is 3^n and the conditional distribution for Y given that X is in this interval is uniform.

7 Conclusions

The "simplifying assumption" might also be called the "simplifying decision" to represent a complex random quantity by its expectation, in casu replacing the conditional copula by the partial copula. The partial copula is identifiable, that is it can be estimated from the data without further assumptions. If the conditional copulas were not in fact constant then their use would evidently provide a more faithful representation of the data. This paper illustrates three possibilities: (a) a non-simplified representation of the partial copula does not exist, (b) a non simplified representation of the partial copula exists and is identifiable and (c) non-simplified representations of the partial copula exist but are not identifiable without further assumptions or exogenous information. The independent partial copula falls in case (c), as does a non-simplified copula chosen on the basis of its correlation function. Approaches such as [8] and [9] involving partitioning the space of conditioning variables and estimating "conditional partial copulas" holds promise if we can judiciously select which conditional partial copulas to estimate. [5] show that there is much to be gained with such an approach.

Appendix A Box Copula Correlation

Consider a box copula intersecting the x, y axes at $(h, 0), (1, 1-h), (1-h, 1), (0, h)$. The computation of its correlation can be written in detail using Dirac copulas, using $\delta_{xo}(dx/a) = a\delta_{xo}(dx), a \in \|$ For a sketch of the computation, first compute $\int_0^1 x \times y(x)dx$:

1. $\int_0^h (x(-x+h) + x(x+h))dx = h^3$
2. $\int_h^{1-h} (x(x+h) + x(x-h))dx = \frac{2}{3} - 2h + 2h^2 - \frac{4}{3}h^3$
3. $\int_{1-h}^1 (x(x-h) + x(-x+2-h))dx = 2h - 3h^2 + h^3$.

Adding these contributions,

$$\int_0^1 x \times y(x)dx = \frac{2}{3}h^3 - h^2 + \frac{2}{3}.$$

This integral double counts the mass. Dividing by 2, subtracting the mean squared and dividing by the variance we arrive at equation 6.

Appendix B Correlation Functions

The calculation of the correlation function for copula family $A(h)$ is sketched, the others are similar. For $x \in [0, h), y(x) = x + 1 - h$; for $x \in [h, 1],\ y(x) = x - h$.

$$\int_0^1 xy(x)dx = \int_0^h x(x+1-h)dx + \int_h^1 x(x-h)dx = -\frac{h^3}{6} + \frac{h^2}{2} + \frac{h^3}{6} - \frac{h}{2} + \frac{1}{3}.$$

Subtracting the product of the means of x, y and dividing by the product of their standard deviations:

$$\rho(h) = \frac{\frac{h^2}{2} - \frac{h}{2} + \frac{1}{3} - \frac{1}{4}}{\frac{1}{12}}.$$

References

[1] Cooke, R.M.: Markov and entropy properties of tree and vine-dependent variables. In: Proceedings of the ASA Section on Bayesian Statistical Science, (1997)

[2] Bedford, T., Cooke, R.M.: Vines: A new graphical model for dependent random variables. Annals of Statistics, 1031–1068 (2002)

[3] Mroz, T., Fuchs, S., Trutschnig, W.: How simplifying and flexible is the simplifying assumption in pair-copula constructions – analytic answers in dimension three and a glimpse beyond. Electronic Journal of Statistics **15**(1), 1951–1992 (2021) https://doi.org/10.1214/21-EJS1832

[4] Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial intelligence **32**(1-4), 245–268 (2001)

[5] Morales-Nápoles, O., Rajabi-Bahaabadi, M., Torres-Alves, G.A., Hart, C.M.P.: Chimera: An atlas of regular vines on up to 8 nodes. Scientific Data **10**(1), 337 (2023)

[6] Czado, C.: Vine copula based structural equation models. Computational Statistics & Data Analysis, 108076 (2024)

[7] Oxtoby, J.: Measure and Category. Springer, ??? (1980)

[8] Kurz, M.S., Spanhel, F.: Testing the simplifying assumption in high-dimensional vine copulas. Electronic Journal of Statistics **16**(2), 5226–5276 (2022)

[9] Derumigny, A., Fermanian, J.-D.: About tests of the “simplifying” assumption for conditional copulas. Dependence Modeling **5**(1), 154–197 (2017)

[10] Acar, E.F., Genest, C., Nešlehová, J.: Beyond simplified pair-copula constructions. Journal of Multivariate Analysis **110**, 74–90 (2012)

[11] Vatter, T., Nagler, T.: Generalized additive models for pair-copula constructions. Journal of Computational and Graphical Statistics **27**(4), 715–727 (2018)

[12] Nagler, T.: Simplified vine copula models: state of science and affairs. arXiv preprint arXiv:2410.16806 (2025)

[13] Cooke, R.M., Kurowicka, D., Wilson, K.: Sampling, conditionalizing, counting, merging, searching regular vines. Journal of Multivariate Analysis **138**, 4–18

(2015)

[14] Vaart, A.W.: Asymptotic Statistics vol. 3. Cambridge university press, ??? (2000)

[15] Tsiatis, A.: A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences **72**(1), 20–22 (1975)

[16] Cooke, R., Bedford, T.: Reliability databases in perspective. IEEE Transactions on Reliability **51**(3), 294–310 (2002)

[17] Durante, F., Sempi, C., *et al.*: Principles of Copula Theory vol. 474. CRC press Boca Raton, FL, ??? (2016)

[18] Krein, M., Milman, D.: On extreme points of regular convex sets. Studia Mathematica **9**, 133–138 (1940)

[19] Birkhoff, G.: Three observations on linear algebra. Univ. Nac. Tacuman, Rev. Ser. A **5**, 147–151 (1946)

[20] Von Neumann, J.: A certain zero-sum two-person game equivalent to the optimal assignment problem. Contributions to the Theory of Games **2**(0), 5–12 (1953)

[21] Konig, D.: Theorie der Endlichen und Unendlichen Graphen vol. 72. American Mathematical Soc., ??? (2001)

[22] Dufossé, F., Uçar, B.: Notes on birkhoff–von neumann decomposition of doubly stochastic matrices. Linear Algebra and its Applications **497**, 108–115 (2016)

[23] Dufossé, F., Kaya, K., Panagiotas, I., Uçar, B.: Further notes on birkhoff–von neumann decomposition of doubly stochastic matrices. Linear Algebra and its Applications **554**, 68–78 (2018)

[24] Cappellini, V., Sommers, H.-J., Bruzda, W., Życzkowski, K.: Random bistochastic matrices. Journal of Physics A: Mathematical and Theoretical **42**(36), 365209 (2009)

[25] Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. The annals of mathematical statistics **35**(2), 876–879 (1964)

[26] Csiszár, I.: I-divergence geometry of probability distributions and minimization problems. The annals of probability, 146–158 (1975)

[27] Csiszár, I.: Information geometry and alternating minimization procedures. Statistics and Decisions, Dedewicz **1**, 205–237 (1984)

[28] Cooke, R.M., Waij, R.: Monte carlo sampling for generalized knowledge dependence with application to human reliability. Risk Analysis **6**(3), 335–343 (1986)

[29] Kimeldorf, G., Sampson, A.R.: Monotone dependence. The Annals of Statistics, 895–903 (1978)

[30] Mikusinski, P., Sherwood, H., Taylor, M.D.: Shuffles of min. Stochastica **13**(1), 61–74 (1992)

[31] Billingsley, P.: Convergence of Probability Measures. John Wiley & Sons, ??? (2013)

[32] Nelsen, R.B.: An Introduction to Copulas. Springer, ??? (2006)

[33] Hutchinson, J.E.: Fractals and self similarity. Indiana University Mathematics Journal **30**(5), 713–747 (1981)