

# The Philosophy of Copula Modeling: A Conversation with ChatGPT

MARIUS HOFERT<sup>1,\*</sup>

<sup>1</sup>*The University of Hong Kong, Department of Statistics and Actuarial Science, Hong Kong*

## Abstract

In the form of a scholarly exchange with ChatGPT, we cover fundamentals of modeling stochastic dependence with copulas. The conversation is aimed at a broad audience and provides a light introduction to the topic of copula modeling, a field of potential relevance in all areas where more than one random variable appears in the modeling process. Topics covered include the definition, Sklar’s theorem, the invariance principle, pseudo-observations, tail dependence and stochastic representations. The conversation also shows to what degree it can be useful (or not) to learn about such concepts by interacting with the current version of a chatbot.

**Keywords** *introduction; invariance principle; pseudo-observations; Sklar’s theorem; stochastic representation; tail dependence*

## 1 Introduction

We engage in a scholarly discussion with artificial intelligence chatbot ChatGPT, version 4, about the notion of copulas with the goal of learning how to quantify and model stochastic dependence. Articles assessing ChatGPT’s capabilities in different academic contexts recently appeared, for example Joshi et al. (2023) on undergraduate computer science knowledge or Wood et al. (2023) on exam questions in accounting. There are a large amount of applications of ChatGPT known to date, see for example Lopez-Lira and Tang (2023). For using ChatGPT to find the right machine learning models and approaches for data science purposes (so as a “personal data scientist”), see Hassan et al. (2023). For ChatGPT and its use in data science education, see Ellis and Slade (2023) and Zheng (2023).

The following conversation with ChatGPT, version 4, was led with hyperparameter ‘Max response’ (maximal number of tokens per response; an integer between 1 and 4000) set to 800, ‘Temperature’ (controlling randomness of the output; a real number between 0 and 1) set to 0.7, ‘Top  $p$ ’ (probability  $p$  between 0 and 1 defining the smallest set of next tokens to choose from whose cumulative probability exceeds  $p$ ) set to 0.95, and with the ‘Past messages included’ parameter set to 10 messages (providing the memory or length of context of the conversation); all parameter choices are the corresponding defaults. We comment on ChatGPT’s answers (those are the parts starting with “MH to reader”) and, as quality assessment, rate each response according to a five-point Likert scale (5-star rating; 0 stars = poor answer; 5 stars = excellent answer; half-stars allowed). For readability, questions and answers involving lists or mathematical notation have been converted to  $\text{\LaTeX}$ .

---

\* Email: [mhofert@hku.hk](mailto:mhofert@hku.hk).

## 2 Conversation

**MH to ChatGPT:** *Dear ChatGPT, I would like to ask you some fundamentals of stochastic dependence modeling with copulas in multivariate data. To give you context, your answers will be read by a broad audience interested in learning about copulas.*

**ChatGPT's response:**

Dear Reader,

I am glad to provide an overview of the fundamentals of stochastic dependence modeling with copulas in multivariate data. Copulas are a powerful statistical tool used to study the dependence between random variables in multivariate data. Here, we will discuss the basics of copulas, their properties, and their applications in modeling dependence.

1. What are copulas?

Copulas are mathematical functions that join multivariate distribution functions to their one-dimensional marginal distribution functions. In other words, they describe the dependence structure between random variables, irrespective of their individual marginal distributions. This feature makes copulas particularly useful when we want to analyze the dependence between variables with different types of distributions.

2. Sklar's Theorem

The foundation of copula theory is Sklar's theorem, which states that for any multivariate cumulative distribution function (CDF)  $F$  with marginals  $F_1, F_2, \dots, F_n$ , there exists a copula function  $C$  such that:

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

Sklar's theorem also states that if the marginals are continuous, the copula  $C$  is unique.

3. Properties of copulas

Some essential properties of copulas include:

- Copulas are defined on the unit hypercube, i.e., their range and domain are  $[0, 1]^n$  for an  $n$ -dimensional copula.
- Copulas are grounded, which means  $C(0, \dots, 0) = 0$  and  $C(1, \dots, 1) = 1$ .
- Copulas are increasing functions and satisfy the 2-increasing property.

4. Types of copulas

There are several families of copulas, each with unique properties that characterize different types of dependence structures. Some common copula families include:

- Gaussian copula: Derived from the multivariate normal distribution, Gaussian copulas can capture linear dependence structures.
- Clayton copula: Suitable for modeling lower tail dependence, it works well for situations where extreme low values of one variable are associated with extreme low values of another variable.
- Gumbel copula: This copula is suitable for modeling upper tail dependence, which is useful when extreme high values of one variable are associated with extreme high values of another variable.

- Frank copula: Useful for modeling non-monotonic dependence structures, Frank copula can capture both positive and negative dependence.

#### 5. Applications of copulas

Copulas have numerous applications in various fields, including:

- Finance: Modeling portfolio risk, pricing financial derivatives, and estimating value-at-risk.
- Actuarial science: Analyzing the dependence between risks and estimating joint probabilities of extreme events.
- Hydrology: Studying the dependence between different sources of water, such as rainfall and river flow.
- Environmental science: Investigating the relationship between air pollution and public health.

In conclusion, copulas provide a flexible and robust framework to model the dependence between random variables in multivariate data. By capturing the dependence structure separately from the marginal distributions, copulas allow for a more accurate analysis of multivariate relationships, which is essential in numerous fields and applications.

★★★★☆

**MH to reader:** A simple “Sure, how can I help you?” would have been preferred, but while we are at it, let’s have a closer look at ChatGPT’s response to our context, even though this already requires some knowledge about copulas, knowledge we come back to in more detail later.

1. The answer provided under the first numbered section is in general fine, although no proper definition of copulas is given. We also see this behavior of ChatGPT later, it first provides a general (hit-and-miss) answer and then provides a more technical answer.
2. The version of Sklar’s theorem provided is correct, but is missing an important second statement; see also later. For now we can add to the correct first statement that the copula  $C$  is given in terms of the marginal quantile functions by  $C(u_1, u_2, \dots, u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_n^{-1}(u_n))$ , uniquely defined on the product of the ranges of the margins  $F_1, F_2, \dots, F_n$ ; we will use this result later.
3. Here we see the first real inaccuracies or flaws:
  - As a distribution function, the range of an  $n$ -dimensional copula is always  $[0, 1]$  (not:  $[0, 1]^n$ ).
  - The groundedness property of copulas says that  $C(\mathbf{u}) = 0$  if  $u_j = 0$  for *at least one*  $j = 1, \dots, n$ , so the statement can be replaced by a more general one. Also,  $C(1, \dots, 1) = 1$  does not follow from groundedness but rather from the fact that a copula has  $U(0, 1)$  univariate margins, so  $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$  for all  $j = 1, \dots, n$ .
  - An  $n$ -dimensional copula  $C$  must be  $n$ -increasing (not only 2-increasing), which means that it assigns non-negative probability mass to all possible non-empty  $n$ -dimensional hyperrectangles. More digestible is to say that if  $C$  has a density, this density must indeed be non-negative.
4. The “unique” in the first sentence is too vague to always be correct, the properties of the copulas listed thereafter do not uniquely characterize them. And that “Gaussian copulas can capture linear dependence structures” (a statement more widely found in the literature) is another vague or even void sentence. The multivariate normal distribution captures linear dependence through the notion of Pearson’s correlation coefficient  $\rho$ . However, at the level of the Gaussian copula,  $\rho$  merely acts as a parameter of the copula, there is no linear relationship

between the underlying random variables (copulas are highly non-linear on their compact support, the unit hypercube). In particular, if  $\rho = 1$  for a bivariate normal distribution, the corresponding random variables with finite second moments are linearly dependent almost surely. But  $\rho = 1$  for a bivariate Gaussian copula means that it is the comonotone copula  $M(u_1, u_2) = \min\{u_1, u_2\}$  (with stochastic representation  $(U, U) \sim M$  for (the same)  $U \sim U(0, 1)$ ), so *any* two random variables  $X_1 \sim F_1$  and  $X_2 \sim F_2$  combined with this Gaussian copula with  $\rho = 1$  are comonotone in the sense that  $(X_1, X_2)$  is in distribution equal to  $(F_1^{-1}(U), F_2^{-1}(U))$  for a single  $U \sim U(0, 1)$ , irrespectively of whether  $F_1, F_2$  have any finite moments. And, clearly,  $F_1^{-1}(U)$  and  $F_2^{-1}(U)$  are not necessarily linearly dependent. Finally, in order for the “non-monotonic” adjective given for Frank copulas to make sense, one should replace it by “non-comonotonic” or by “discordant”; also “copula” should be “copulas” for this entry.

We already see that ChatGPT can be wrong or at least a bit off sometimes, an effect known as *hallucinations*; see Crouse (2023), Rodger (2023) or Hofert (2023a). Even if statements are often correct, we cannot trust the answers; to make things worse, sometimes ChatGPT sounds highly convincing. ChatGPT rather gives us interesting output to enjoy, analyze and validate our understanding of the topic under consideration against; see also Hofert (2023a) and Hofert (2023b) in this regard. One should also keep in mind that, although correct in principle, the notation used by ChatGPT is sometimes pedagogically questionable at best, which is especially relevant when first encountering a mathematical concept to wrap one’s head around; for example,  $n$  is typically reserved for the sample size (see also later, where ChatGPT also uses  $n$  as sample size) but is used by ChatGPT here for the dimension of the copula (the latter is typically referred to as  $d$ , with the frequently used  $p$  in more classical areas of statistics being reserved for the dimension of the parameter space when considering parametric copulas).

**MH to ChatGPT:** *Okay, those are a lot of things already. Let’s start with the definition of copulas. Can you provide a mathematical definition of the stochastic notion of a copula?*

**ChatGPT’s response:**

Sure! A copula is a mathematical function used to model the dependence structure between multiple random variables. It essentially describes the joint distribution of these random variables while separating the marginal distributions.

Formally, a copula is a multivariate distribution function  $C(u_1, u_2, \dots, u_n)$  with uniform marginals on the interval  $[0, 1]$ , where  $u_i = F_i(x_i)$  for  $i = 1, 2, \dots, n$ . Here,  $F_i(x_i)$  is the cumulative distribution function (CDF) of the  $i$ -th random variable  $X_i$ .

The copula  $C(u_1, u_2, \dots, u_n)$  has the following properties:

1.  $C(u_1, u_2, \dots, u_n)$  is a multivariate distribution function with domain  $[0, 1]^n$ .
2. The marginal distribution functions of  $C$  are uniform on the interval  $[0, 1]$ , i.e., the one-dimensional marginals of  $C$  are given by  $C_i(u_i) = u_i$ , where  $i = 1, 2, \dots, n$ .
3.  $C(u_1, u_2, \dots, u_n)$ , is invariant under any strictly increasing transformation of its arguments, meaning that the dependence structure remains unchanged even if the marginal distributions are transformed.

The main idea behind copulas is to separate the dependence structure from the marginal distributions, so one can model them independently and then combine them using the copula function. This provides flexibility in modeling multivariate distributions with different types of dependence structures and marginals.

★★★★☆

**MH to reader:** ChatGPT’s answer “A copula is a mathematical function used to model the dependence structure between multiple random variables.” is not a proper definition. A proper definition would be that an  $n$ -dimensional copula is an  $n$ -dimensional distribution function with standard uniform univariate margins.

The sentence “Formally, a copula is a multivariate distribution function  $C(u_1, u_2, \dots, u_n)$  with uniform marginals on the interval  $[0, 1]$ , where  $u_i = F_i(x_i)$  for  $i = 1, 2, \dots, n$ .” is almost correct, but the second part (“[...] where  $u_i = F_i(x_i)$  for  $i = 1, 2, \dots, n$ ”) should be omitted. There should appear no connection to arbitrary marginal distribution functions  $F_1, \dots, F_n$  when defining copulas; and if they are stated as  $U(0, 1)$  distributions, then we have  $F_i(x_i) = x_i$ ,  $x_i \in [0, 1]$ , anyways, so we can omit them. Also here, note again the rather unusual notation with the index  $i$  running across dimensions,  $i (= 1, \dots, n)$  is typically used to index observations from a copula and  $j (= 1, \dots, d)$  is used as an index across dimensions. Even though we did not ask ChatGPT about copula properties, let’s have a quick look at the list of properties it provided:

1. The statement that copulas are distribution functions with domain being the unit hypercube is correct.
2. The sentence “The marginal distribution functions of  $C$  are uniform on the interval  $[0, 1]$ ” is fine (even though it would be more precise to say “The *univariate* marginal distribution functions of  $C$  are uniform on the interval  $[0, 1]$ ”). The left-hand side of “[...]  $C_i(u_i) = u_i$  [...]” in the second part should be interpreted as  $C_i(u_i) := C_i(1, \dots, 1, u_i, 1, \dots, 1)$  to be correct, where the entry  $u_i$  appears as  $i$ -th component of  $C$  and all other components are 1.
3. The statement about the invariance of copulas under strictly increasing transformations of its arguments is correct but needs to be interpreted in the context of the so-called invariance principle (typically best conveyed in terms of the underlying random variables); see later.

The last part is fine, but again not part of what we asked for.

**MH to ChatGPT:** *You already mentioned Sklar’s theorem as a central result in copula modeling, and you provided the statement of one of its parts. Can you state the second part of the theorem, too, and explain its meaning?*

**ChatGPT’s response:**

Certainly! Sklar’s theorem consists of two parts. The first part, which we discussed earlier, states that any multivariate joint distribution function can be represented as a copula function applied to the marginal distribution functions. Now, let’s move to the second part of the theorem:

The second part of Sklar’s theorem states that if the marginal distribution functions are continuous, then the copula  $C$  is unique; otherwise,  $C$  is uniquely determined only on the range of the marginal distribution functions.

In simpler terms, this means that if the marginal distributions are continuous (i.e., they don’t have any jumps or breaks), then there exists a unique copula that connects them to form the joint distribution. This uniqueness is important because it ensures that the copula function will consistently model the dependence structure between the variables.

However, if the marginal distributions are not continuous (e.g., they have jumps or some kind of discontinuity), then the copula is only unique on the specific range

where the marginal distributions are defined. In such cases, the copula might not provide a consistent representation of the dependence structure across the entire range of the variables.

To summarize, the second part of Sklar's theorem emphasizes the importance of continuous marginal distributions for the uniqueness of the copula function in modeling the dependence structure between variables in a multivariate dataset.

★★★★☆☆

**MH to reader:** This answer is not what we were looking for, but nonetheless of remarkable quality, especially in comparison to the struggles ChatGPT, version 3.5, had with such type of questions; see, for example, Hofert (2023a) and Hofert (2023b). The second part of Sklar's Theorem states that for any  $d$ -dimensional copula  $C$  and any univariate distribution functions  $F_1, \dots, F_d$ , the function  $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ ,  $\mathbf{x} \in \mathbb{R}$ , is a  $d$ -dimensional distribution function with margins  $F_1, \dots, F_d$ ; note that this result holds irrespectively of whether the margins are continuous. Let's see if we can guide ChatGPT there (we hereby also use the more common notation for the copula dimension as mentioned before).

**MH to ChatGPT:** *This was not what I meant. The first part of Sklar's Theorem you described can be viewed as a decomposition, namely any multivariate distribution function  $F$  with continuous margins  $F_1, \dots, F_d$  can be uniquely decomposed into  $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ , so into a copula and the marginal distribution functions corresponding to  $F$ . The second part of Sklar's theorem can be viewed as a composition, it addresses what happens if we combine any copula with any marginal distribution functions  $F_1, \dots, F_d$  in the form of  $C(F_1(x_1), \dots, F_d(x_d))$ . Do you remember the correct statement now?*

**ChatGPT's response:**

Apologies for the confusion earlier. Yes, I understand now. The second part of Sklar's theorem states that if you combine any copula  $C$  with any marginal distribution functions  $F_1, \dots, F_d$  in the form  $C(F_1(x_1), \dots, F_d(x_d))$ , then the resulting function is a valid multivariate distribution function with continuous margins  $F_1, \dots, F_d$ .

In other words, Sklar's theorem not only guarantees the decomposition of a multivariate distribution function with continuous margins into a copula and its marginal distributions, but also ensures that any valid copula combined with continuous marginal distributions forms a valid multivariate distribution function. This property highlights the flexibility and usefulness of copulas in constructing and modeling multivariate distributions.

★★★★★☆☆

**MH to reader:** The answer is excellent apart from the fact that both paragraphs contain an unnecessary limitation highly relevant for model building with copulas. Did you spot it? The point is, as mentioned before, that the second part of Sklar's theorem (the composition) does *not* require the margins to be continuous.

Why is that? A simple argument can be given here by a comparison with the univariate case: The two parts of "Sklar's theorem for  $d = 1$ " consider the well-known probability integral transformation and its inverse, the quantile transformation. Concerning the probability integral transformation, if  $X \sim F$  and  $F$  is continuous, then  $F(X) \sim U(0, 1)$ ; here continuity of  $F$  is crucial as otherwise not all values in  $(0, 1)$  are reached and therefore  $F(X)$  would not be continuously distributed. Concerning the quantile transformation, if  $F$  is *any* univariate distribution function and  $U \sim U(0, 1)$ , then  $F^{-1}(U) \sim F$ . This holds for all univariate distribution functions and is

well known as *inversion method* for sampling. In fact, normal distributions in R are sampled with the inversion method based on accurate and fast approximations of the  $N(0, 1)$  quantile function  $\Phi^{-1}$ .

The first part of Sklar's theorem (the decomposition) is mostly relevant in statistical applications. For example, if a multivariate  $F$  consists of  $d = 100$  continuous margins  $F_1, \dots, F_d$ , each with only two parameters and a copula  $C$  with one parameter, then estimating all 201 parameters of  $F$  simultaneously may not be tractable or numerically well-behaved. However, the copula decomposition allows this problem to be viewed and handled as 101 estimation problems with at most two parameters each, which can all be done in parallel; see, for example, Hofert et al. (2018, Chapter 4) for the details, and below for the relevant pseudo-observations and corresponding 'price' one has to pay for this convenience.

The second part of Sklar's theorem (the composition) is especially popular in applications such as finance, insurance or risk management, where tailor-made distributions are of interest. For example, in a portfolio of  $d$  insurance contracts, one may have enough individual information to know or estimate  $F_1, \dots, F_d$ , but too little or no information about the dependence. Combining  $F_1, \dots, F_d$  with different copulas allows one to assess how relevant the dependence structure is for the overall risk of the portfolio or which copula properties translate to extreme portfolio risks, etc.; this is often done by Monte Carlo simulation. In essence, the more sensitive the portfolio's risk is to changes in the copula, the more careful we need to be in correctly modeling the dependence. Turning this around, we can implement *stress testing* this way, by combining the known margins with copulas reflecting stress scenarios, and then assess the portfolio's risk across these stress scenarios. Like playing Lego, we can combine marginal bricks (distribution functions) and dependence bricks (copulas) to create flexible objects (multivariate distribution functions) with the second part of Sklar's theorem.

**MH to ChatGPT:** *Yes, that was what I was looking for. Can you give an example how this second part of Sklar's theorem can be used to construct a flexible multivariate distribution function  $F$ ?*

**ChatGPT's response:**

Sure! The second part of Sklar's theorem allows you to create a flexible multivariate distribution function by combining any copula with any marginal distribution functions. Here's a simple example:

Suppose you have two random variables  $X$  and  $Y$ , and you want to model their joint distribution  $F(x, y)$ . You know the marginal distribution functions for  $X$  and  $Y$ , which are  $F_X(x)$  and  $F_Y(y)$ , respectively. Let's assume that  $X$  follows a normal distribution with mean  $\mu_X$  and standard deviation  $\sigma_X$ , and  $Y$  follows an exponential distribution with rate parameter  $\lambda$ :

$$F_X(x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right),$$

$$F_Y(y) = 1 - \exp(-\lambda y),$$

(where  $\Phi$  is the standard normal cumulative distribution function).

Now, you need to choose a copula  $C(u, v)$  to model the dependence structure between  $X$  and  $Y$ . For example, you can use the Gaussian copula, which is defined by the correlation parameter  $\rho$ :

$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$  (where  $\Phi_\rho$  is the bivariate standard normal cumulative distribution function with correlation  $\rho$ )

Now, using the second part of Sklar's theorem, you can construct the joint distribution function  $F(x, y)$  by combining the chosen copula with the marginal distribution functions:

$$\begin{aligned} F(x, y) &= C_\rho(F_X(x), F_Y(y)) \\ &= \Phi_\rho\left(\Phi\left(\frac{x - \mu_X}{\sigma_X}\right), 1 - \exp(-\lambda y)\right). \end{aligned}$$

This joint distribution function  $F(x, y)$  represents a flexible multivariate distribution that captures the dependence between the random variables  $X$  and  $Y$  using the Gaussian copula, while maintaining their specified marginal distributions. By changing the copula or the marginal distributions, you can create various multivariate distribution functions to suit your specific modeling needs.

★★★★★

**MH to reader:** Interestingly, here ChatGPT correctly mentioned that any marginal distribution functions can be used in the composition part of Sklar's theorem. It also now refers to the correct second part of Sklar's theorem. The answer overall is almost excellent, with one flaw, namely  $F$  should be

$$F(x, y) = \Phi_\rho\left(\frac{x - \mu_X}{\sigma_X}, \Phi^{-1}(1 - \exp(-\lambda y))\right)$$

instead of the one given. These mistakes are no wonder as ChatGPT does not have built-in logic, it is merely a predictive model. What is rather surprising, though, is the coherence of the idea of the model construction across this longer answer compared with the partly rather weak technical answers of version 3.5 in Hofert (2023a) and Hofert (2023b).

Of course, there are pedagogically more meaningful examples, one is the following. Suppose you want to construct a model for  $(X_1, \dots, X_d)$ . Except for the too limited multivariate normal you recall the multivariate Student's  $t$  distribution as a more flexible choice. However, this (also) implies that all margins have to be of the same *type*, so linear (or: location-scale) transforms of each other. If there are two margins that violate this property, the multivariate  $t$  is not the right model. That means, if two margins come from  $t$  distributions with *different* degrees-of-freedom parameters, we do not have an adequate model anymore. However, by the second part of Sklar's theorem, we can simply use the copula implied by the multivariate  $t_\nu$  distribution – the  $t_\nu$  copula – and combine it with univariate  $t_{\nu_j}$ ,  $j = 1, \dots, d$ , margins to obtain an already much more flexible model with possibly  $d + 1$  *different* degrees-of-freedom parameters. Of course, one can also rephrase this example in terms of a single margin not being  $t$  distributed at all, in which case we could simply model that margin with a more suitable univariate distribution function and then combine it with the other univariate  $t$  margins and the  $t$  copula to construct an adequate joint model.

**MH to ChatGPT:** *The modern proof of the two parts of Sklar's theorem (utilizing stochastic representations) is straightforward by applying marginal probability integral transformations and by applying marginal quantile transformations. Essentially, if  $\mathbf{X} = (X_1, \dots, X_d)$  is a  $d$ -dimensional random vector following a distribution function  $H$  with continuous margins  $F_1, \dots, F_d$ , then  $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$  (marginal probability integral transformations) follows the*



copula  $C$  such that  $H(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ . And if  $\mathbf{U} = (U_1, \dots, U_d)$  is a  $d$ -dimensional random vector following any copula  $C$ , then  $\mathbf{X} = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$  (marginal quantile transformations) is a  $d$ -dimensional random vector following the distribution function  $H(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$  with margins  $F_1, \dots, F_d$ . More generally, if  $\mathbf{Y} = (T_1(X_1), \dots, T_d(X_d))$  is the  $d$ -dimensional random vector obtained by marginally applying strictly increasing transformations to the components of  $\mathbf{X} = (X_1, \dots, X_d)$ , why is the copula of  $\mathbf{Y}$  still the one of  $\mathbf{X}$ ? In other words, why does the copula not change by applying such transformations?

#### ChatGPT's response:

The copula of a random vector remains unchanged when applying strictly increasing transformations to the components because copulas capture the dependence structure between variables, not the specific marginal distributions. When you apply a strictly increasing transformation to a component, you do not alter the underlying dependence between the components.

To understand this better, let's consider the random vector  $\mathbf{X} = (X_1, X_2)$  with a copula  $C$ . Now let's apply strictly increasing transformations  $T_1$  and  $T_2$  to the components  $X_1$  and  $X_2$ , resulting in a new random vector  $\mathbf{Y} = (Y_1, Y_2) = (T_1(X_1), T_2(X_2))$ .

The joint distribution of  $\mathbf{Y}$  can be expressed as  $H(y_1, y_2) = C(F_1(y_1), F_2(y_2))$ , where  $F_1$  and  $F_2$  are the marginal distribution functions of  $Y_1$  and  $Y_2$ . Since  $T_1$  and  $T_2$  are strictly increasing, their inverse functions  $T_1^{-1}$  and  $T_2^{-1}$  exist. We can express the joint distribution of  $\mathbf{X}$  as  $F(x_1, x_2) = C(G_1(x_1), G_2(x_2))$ , where  $G_1$  and  $G_2$  are the marginal distribution functions of  $X_1$  and  $X_2$ .

Now, let's substitute  $x_1 = T_1^{-1}(y_1)$  and  $x_2 = T_2^{-1}(y_2)$  in the expression for  $F(x_1, x_2)$ . We get:

$$F(T_1^{-1}(y_1), T_2^{-1}(y_2)) = C(G_1(T_1^{-1}(y_1)), G_2(T_2^{-1}(y_2))).$$

Since  $T_1$  and  $T_2$  are strictly increasing,  $G_1(T_1^{-1}(y_1)) = F_1(y_1)$  and  $G_2(T_2^{-1}(y_2)) = F_2(y_2)$ . Therefore, we obtain:

$$H(y_1, y_2) = C(F_1(y_1), F_2(y_2)).$$

This shows that the copula  $C$  remains unchanged when applying strictly increasing transformations to the components of  $\mathbf{X}$ , as it still represents the dependence structure between the transformed variables  $Y_1$  and  $Y_2$ .

★★★★☆

**MH to reader:** As mentioned earlier, the first sentence tries to provide a convincing argument, but it is not a mathematical one. The second sentence merely provides the statement again. Then, ChatGPT provides a mathematical argument (in  $d = 2$  instead of for general  $d \geq 2$ , but this is sufficient here as the general case follows the exact same steps), albeit in a less pedagogically meaningful way. More rigorously, like ChatGPT did, let us assume that  $\mathbf{X} = (X_1, X_2) \sim F$  with  $F(x_1, x_2) = C(G_1(x_1), G_2(x_2))$  and that the transformed  $\mathbf{Y} = (Y_1, Y_2) = (T_1(X_1), T_2(X_2))$  has distribution function  $H$  with margins  $F_1, F_2$ ; note that we cannot yet assume the corresponding copula to be the same as the copula  $C$  of  $(X_1, X_2)$ , this is what we have to show! Then the correct argument is as follows (for more details about working with quantile functions, see Embrechts and Hofert (2013)). We first derive the form of  $H$ , which is

$$\begin{aligned} H(y_1, y_2) &= \mathbb{P}(T_1(X_1) \leq y_1, T_2(X_2) \leq y_2) = \mathbb{P}(X_1 \leq T_1^{-1}(y_1), X_2 \leq T_2^{-1}(y_2)) \\ &= F(T_1^{-1}(y_1), T_2^{-1}(y_2)) = C(G_1(T_1^{-1}(y_1)), G_2(T_2^{-1}(y_2))), \end{aligned}$$

and from there the margins of  $H$ , which are

$$F_j(y_j) = G_j(T_j^{-1}(y_j)), \quad j = 1, 2.$$

The corresponding generalized inverses are  $F_j^{-1}(u_j) = T_j(G_j^{-1}(u_j))$ ,  $j = 1, 2$ . By the first part of Sklar's Theorem, the copula of  $\mathbf{Y}$  is given by

$$\begin{aligned} H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) &= C(G_1(T_1^{-1}(T_1(G_1^{-1}(u_1))))), G_2(T_2^{-1}(T_2(G_2^{-1}(u_2)))))) \\ &= C(u_1, \dots, u_d), \end{aligned}$$

so, indeed, the same copula  $C$  as that of  $F$  (or  $\mathbf{X}$ ).

**MH to ChatGPT:** *One application of the invariance principle is obtained by taking as componentwise strictly increasing transformations the distribution functions  $F_1, \dots, F_d$  of  $X_1, \dots, X_d$ , respectively, as in the proof of the first part of Sklar's theorem. This idea is also underlying the construction of data from  $C$  based on  $n$  observations from  $(X_1, \dots, X_d)$ , known as pseudo-observations. Can you explain mathematically how the pseudo-observations are constructed?*

**ChatGPT's response:**

Certainly. Pseudo-observations are constructed by applying the empirical distribution function to the original data points, essentially transforming them to the unit hypercube, i.e., the space of the copula.

Given a dataset of  $n$  observations from a  $d$ -dimensional random vector  $(X_1, \dots, X_d)$ , we first compute the empirical distribution function for each variable. For the  $i$ -th variable, the empirical distribution function is given by:

$$F_i(x) = \frac{1}{n} \sum_{j=1}^n I(X_{ij} \leq x),$$

where  $X_{ij}$  represents the  $j$ -th observation of the  $i$ -th variable and  $I(\cdot)$  is the indicator function, which equals 1 if the condition inside the parentheses is true and 0 otherwise.

Next, we apply the empirical distribution function to each observation of the original dataset to construct the pseudo-observations. The pseudo-observation corresponding to the  $j$ -th observation is given by:

$$U_j = (F_1(X_{1j}), \dots, F_d(X_{dj})).$$

This procedure transforms the original data  $(X_1, \dots, X_d)$  to the copula space  $(U_1, \dots, U_d)$ , which lie in the unit hypercube  $[0, 1]^d$ . Pseudo-observations can then be used to study the dependence structure among variables, estimate the copula parameters, and perform goodness-of-fit tests.

★★★★☆

**MH to reader:** The first paragraph describes the right idea, but is slightly inaccurate as one conceptually applies the empirical distribution function of the  $j$ -th margin to the  $j$ -th component sample (which is a more accurate description than ChatGPT's). The second paragraph is correct, but, as mentioned before, the interchange of the index  $i$  (normally for the  $i$ -th observed vector)

and  $j$  (normally for the  $j$ -th variable in the observed vector) is confusing, especially when first learning about these concepts with ChatGPT and then consulting other references or using software. Also, there is an even easier way of computing the pseudo-observations. Let's see if ChatGPT knows about it.

**MH to ChatGPT:** *To compute the pseudo-observations, one can avoid having to first compute the empirical distribution functions of the component samples. Do you know how?*

**ChatGPT's response:**

Yes, there's an alternative way to compute pseudo-observations without explicitly computing the empirical distribution functions. You can use the ranks of the data points instead. Here's the step-by-step procedure:

1. For each component  $X_i$  ( $i = 1, \dots, d$ ), obtain its sample of  $n$  observations  $\{x_{i1}, \dots, x_{in}\}$ .
2. For each component  $X_i$ , compute the ranks  $R_{i1}, \dots, R_{in}$  of the observations  $x_{i1}, \dots, x_{in}$ . The rank of an observation is its position in the sorted sample. In case of ties, you can use the average rank.
3. Convert the ranks  $R_{ij}$  into pseudo-observations  $U_{ij}$  by using the formula  $U_{ij} = (R_{ij} - 0.5)/n$ , where  $j = 1, \dots, n$ .

This method provides an approximation of the empirical distribution function without explicitly computing it. The resulting pseudo-observations  $U_{ij}$  can then be used for further analysis with copulas.

★★★★★

**MH to reader:** This is a great answer and somewhat unexpected given the lack of performance of ChatGPT, version 3.5, on follow-up questions of this sort; see Hofert (2023a) and Hofert (2023b). There are minor details that could be improved or mentioned, such as the aforementioned notation of the indices and the fact that pseudo-observations are typically computed via the slightly different scaling  $U_{ij} = R_{ij}/(n + 1)$ ,  $j = 1, \dots, n$  (where we used ChatGPT's notation).

**MH to ChatGPT:** *The pseudo-observations also show how copula modeling differs from general multivariate modeling based on observations from independent and identically distributed random vectors. Due to the componentwise scaled ranks, the pseudo-observations are not independent anymore. Do you know a situation in which this particularly matters?*

**ChatGPT's response:**

Yes, one situation in which the dependence between pseudo-observations particularly matters is when performing statistical inference, such as parameter estimation and hypothesis testing for copula models. The dependence between the pseudo-observations can lead to biased estimates and incorrect conclusions when using classical methods that assume the data are independent and identically distributed (i.i.d.).

For example, when fitting a copula model to the data and estimating its parameters, the dependence between pseudo-observations can cause underestimation of the standard errors, leading to overconfidence in the estimated parameters. Similarly, when performing hypothesis tests, such as goodness-of-fit tests, the dependence between pseudo-observations can cause an increased probability of type I errors (rejecting a true null hypothesis) or type II errors (failing to reject a false null hypothesis).

To account for the dependence between pseudo-observations, modified statistical methods have been developed specifically for copula modeling, such as the parametric bootstrap, the two-step estimation procedure, and the block-maxima and peaks-over-threshold methods for extreme value copulas. These methods take into account the dependence structure in the data and provide more accurate and reliable results in the context of copula modeling.

★★★★★

**MH to reader:** Again, this is overall a great answer, at least without the part after “[...] such as the parametric bootstrap [...]” in the last paragraph, which mentions procedures that are not relevant for the point being made; for example, the two-stage (or two-step) estimation procedure actually leads to the problem of non-i.i.d. data, so is rather the cause than the solution of the problem of dependent ranks.

**MH to ChatGPT:** *Copula modeling is widely applied in finance, insurance and risk management, partly because of the importance of the notion of tail dependence. Can you provide the definitions of the lower and upper tail dependence coefficients and quickly describe an application where tail dependence matters?*

**ChatGPT’s response:**

Lower and upper tail dependence coefficients are measures that quantify the dependence between the lower and upper tails of the marginal distributions in a multivariate setting. They help assess the potential for extreme events occurring simultaneously.

The lower tail dependence coefficient between two random variables  $X$  and  $Y$  is defined as:

$$\lambda_L = \lim_{u \rightarrow 0^+} \mathbb{P}(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u)).$$

The upper tail dependence coefficient between two random variables  $X$  and  $Y$  is defined as:

$$\lambda_U = \lim_{u \rightarrow 1^-} \mathbb{P}(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)).$$

In both cases,  $F_X$  and  $F_Y$  are the marginal distribution functions of  $X$  and  $Y$ , respectively, and  $F_X^{-1}$  and  $F_Y^{-1}$  are their inverse functions (quantile functions).

Tail dependence is particularly important in finance, insurance, and risk management because these fields often deal with the occurrence of rare and extreme events, such as large financial losses or natural disasters. For example, in financial risk management, tail dependence is used to assess the likelihood of simultaneous extreme losses in different assets within a portfolio. If the tail dependence is high, it means that the assets are more likely to experience large losses together, which can be detrimental to the portfolio. In this context, understanding tail dependence helps risk managers to diversify their portfolios and mitigate the potential impact of extreme events.

★★★★★

**MH to reader:** Also here, a great answer. Although the first paragraph could be misinterpreted (in the sense of tail dependence coefficients quantifying the dependence between the lower and upper tail of the same univariate distribution – they do not), the correct definitions thereafter

leave no room for misinterpretation. The last paragraph is also excellent, it is even formulated in such a way that it remains correct, irrespective of whether losses are modeled as negative realizations (more popular in finance where the focus is on gains) or positive realizations (more popular in risk management where the focus is on losses) of loss random variables.

**MH to ChatGPT:** *Can you provide definitions of the lower and upper tail dependence coefficients that only depend on the underlying copula under consideration?*

**ChatGPT's response:**

Yes, the lower and upper tail dependence coefficients can be defined in terms of the underlying copula  $C$ . For a bivariate copula  $C(u, v)$  with  $u, v \in [0, 1]$ , the lower and upper tail dependence coefficients are defined as follows:

Lower tail dependence coefficient:

$$\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(V \leq F_v^{-1}(u) \mid U \leq F_u^{-1}(u)) = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}.$$

Upper tail dependence coefficient:

$$\lambda_U = \lim_{u \rightarrow 1} \mathbb{P}(V > F_v^{-1}(u) \mid U > F_u^{-1}(u)) = \lim_{u \rightarrow 1} \frac{1 - C(u, u)}{1 - u}.$$

These definitions only depend on the copula  $C$  and capture the dependence structure between the two variables  $U$  and  $V$  in the lower and upper tails of their joint distribution. In applications, tail dependence coefficients are important for understanding and managing the risk of extreme events, such as financial crashes or natural disasters, where the dependence between variables in the tail regions plays a crucial role.

★★☆☆☆

**MH to reader:** If the limits exist,

$$\lambda_L = \lim_{u \rightarrow 0+} \frac{C(u, u)}{u} \quad \text{and} \quad \lambda_U = \lim_{u \rightarrow 1-} \frac{1 - 2u + C(u, u)}{1 - u}$$

are the lower and upper tail dependence coefficients of  $(U, V) \sim C$ , respectively. Comparing with ChatGPT's answer, the following comes up:

- We are missing that  $(U, V) \sim C$ .
- Also, the limits are not formulated as one-sided limits anymore.
- Moreover,  $F_u, F_v$  should better be denoted by  $F_U, F_V$ ; the expression  $F_u^{-1}(u)$  is especially a bad choice of notation.
- A larger mistake concerns the form of  $\lambda_U$ , it must be  $\lambda_U = \lim_{u \rightarrow 1-} (1 - 2u + C(u, u))/(1 - u)$  (or forms derived thereof). In the form as stated by ChatGPT, the independence copula  $C(u, v) = uv$  leads to  $\lambda_U = 2$  (instead of  $\lambda_U = 0$ ), which is not even in  $[0, 1]$ .

It again shows that one cannot trust ChatGPT's answers much. After several close to excellent answers to seemingly much harder questions, it messes up a rather simple definition.

**MH to ChatGPT:** *You already mentioned the Gaussian copula before, an example of the class of elliptical copulas which also contains Student's  $t$  copula. What are the lower tail dependence coefficients of these two copulas? And why are the lower and upper tail dependence coefficients equal for each of these copula families?*

**ChatGPT's response:**

For the Gaussian copula with correlation matrix  $\Sigma$ , the tail dependence coefficients are given by:

$$\text{Lower tail dependence coefficient } \lambda_L = 2\Phi(\Phi^{-1}(1/2)\rho) - 1,$$

$$\text{Upper tail dependence coefficient } \lambda_U = 2\Phi(\Phi^{-1}(1/2)\rho) - 1,$$

where  $\Phi$  is the standard normal cumulative distribution function,  $\Phi^{-1}$  is its inverse, and  $\rho$  is the correlation coefficient between the variables.

For the Student's  $t$  copula with correlation matrix  $\Sigma$  and degrees of freedom  $\nu$ , the tail dependence coefficients are given by:

$$\text{Lower tail dependence coefficient } \lambda_L = 2t_\nu(t_\nu^{-1}(1/2, \nu)\rho) - 1,$$

$$\text{Upper tail dependence coefficient } \lambda_U = 2t_\nu(t_\nu^{-1}(1/2, \nu)\rho) - 1,$$

where  $t_\nu$  is the cumulative distribution function of the  $t$ -distribution with  $\nu$  degrees of freedom,  $t_\nu^{-1}$  is its inverse, and  $\rho$  is the correlation coefficient between the variables.

For both the Gaussian and Student's  $t$  copulas, the lower and upper tail dependence coefficients are equal because these copulas have symmetric dependence structures. This means that the dependence between the variables in the lower tail of the distribution is the same as the dependence in the upper tail. This property is typical for elliptical copulas, as they exhibit a symmetric dependence structure.

★ ★ ☆ ☆ ☆

**MH to reader:** All four provided formulas are wrong. For Gaussian copulas,

$$\lambda_L = \lambda_U = \mathbb{1}_{\{\rho=1\}},$$

so 0 unless  $\rho = 1$ , a fact known since Sibuya (1959) and painfully rediscovered in the aftermath of the financial crisis of 2007–2008. And for  $t$  copulas,

$$\lambda_L = \lambda_U = 2F_{t_{\nu+1}}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right),$$

so positive unless  $\rho = -1$  or unless  $\nu \rightarrow \infty$  and  $\rho \in [-1, 1)$ ; here  $F_{t_{\nu+1}}$  denotes the  $t_{\nu+1}$  distribution function. The reason ChatGPT provides for why  $\lambda_L = \lambda_U$  is also rather vague, the correct answer being that, as elliptical copulas, Gaussian and  $t$  copulas are radially symmetric. A random vector  $\mathbf{U}$  following a radially symmetric copula satisfies that  $\mathbf{U}$  is in distribution equal to  $\mathbf{1} - \mathbf{U}$ , thus lower and upper parts of a radially symmetric copula are equal and so are the tail dependence coefficients. But let's go back to the provided formulas again.

**MH to ChatGPT:** *Are the formulas for the tail dependence coefficients of Gaussian and  $t$  copulas you provided correct?*

**ChatGPT's response:**

Yes, the formulas I provided for the tail dependence coefficients of Gaussian and  $t$  copulas are correct. To recap:

For the Gaussian copula with correlation parameter  $\rho$ , the lower (and upper) tail dependence coefficient is given by:

$$\lambda_L = \lambda_U = \Phi\left(\frac{\rho}{\sqrt{1-\rho^2}}\right),$$

where  $\Phi$  denotes the standard normal cumulative distribution function.

For the  $t$  copula with correlation parameter  $\rho$  and degrees of freedom  $\nu$ , the lower (and upper) tail dependence coefficient is given by:

$$\lambda_L = \lambda_U = T_{\nu,\rho}\left(\rho\sqrt{\frac{\nu}{\nu-2}}\right),$$

where  $T_{\nu,\rho}$  denotes the bivariate  $t$  distribution with  $\nu$  degrees of freedom and correlation  $\rho$ .

The lower and upper tail dependence coefficients are equal for each of these copula families because they are both symmetric copulas, meaning they exhibit the same dependence structure in the lower and upper tails. This property is a result of the underlying elliptical distribution, which has symmetric contours.

★☆☆☆☆

**MH to reader:** The reason for  $\lambda_L = \lambda_U$  due to radial symmetry is now slightly better explained than before. But overall, we clearly see one of the major problems when working with ChatGPT: Although at first convinced that the aforementioned wrong tail dependence coefficients were correct, ChatGPT now provides two new versions – that are also wrong. Learning precise mathematical concepts from ChatGPT is currently not advisable due to the lack of built-in mathematical logic. Learning with ChatGPT about such concepts to test your knowledge can be beneficial, though. For example, at this point, one could start to derive the correct tail dependence coefficients for Gaussian and  $t$  copulas.

**MH to ChatGPT:** *All copulas used in practice have stochastic representations (representations in terms of random variables that are well-known and easy to simulate from), which often allows one to understand how the dependence between the components of a random vector arises under these models. Can you provide some example copulas with their corresponding stochastic representations?*

**ChatGPT's response:**

Sure, here are some examples of copulas with their corresponding stochastic representations:

1. Gaussian Copula:

The Gaussian copula is defined by the correlation matrix  $R$  and can be represented as:  $C(u_1, \dots, u_d) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$ , where  $\Phi$  is the cumulative distribution function (CDF) of the standard Gaussian distribution,  $\Phi^{-1}$  is its inverse, and  $\Phi_R$  is the joint CDF of a multivariate Gaussian distribution with mean vector  $\mathbf{0}$  and correlation matrix  $R$ .

Stochastic representation:

If  $\mathbf{Z} = (Z_1, \dots, Z_d)$  is a multivariate Gaussian random vector with mean vector  $\mathbf{0}$  and correlation matrix  $R$ , then the Gaussian copula can be represented as:  $U_i = \Phi(Z_i)$ , for  $i = 1, \dots, d$ .

2. Student's  $t$  Copula:

The Student's  $t$  copula is defined by the correlation matrix  $R$  and the degrees of freedom parameter  $\nu$ . It can be represented as:  $C(u_1, \dots, u_d) = T_{R,\nu}(T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_d))$ , where  $T_\nu$  is the CDF of the univariate Student's  $t$  distribution with  $\nu$  degrees of freedom,  $T_\nu^{-1}$  is its inverse, and  $T_{R,\nu}$  is the joint CDF of a multivariate Student's  $t$  distribution with mean vector  $\mathbf{0}$ , correlation matrix  $R$ , and  $\nu$  degrees of freedom.

Stochastic representation:

If  $\mathbf{Z} = (Z_1, \dots, Z_d)$  is a multivariate Student's  $t$  random vector with mean vector  $\mathbf{0}$ , correlation matrix  $R$ , and  $\nu$  degrees of freedom, then the Student's  $t$  copula can be represented as:  $U_i = T_\nu(Z_i)$ , for  $i = 1, \dots, d$ .

## 3. Clayton Copula:

The Clayton copula is defined by the parameter  $\alpha > 0$  and can be represented as:  $C(u_1, \dots, u_d) = (u_1^{-\alpha} + \dots + u_d^{-\alpha} - d + 1)^{-1/\alpha}$ .

Stochastic representation:

If  $W_1, \dots, W_d$  are independent and identically distributed (i.i.d.) random variables following an exponential distribution with mean  $1/\alpha$ , and  $V$  is an independent random variable following a gamma distribution with shape parameter  $d$  and scale parameter 1, then the Clayton copula can be represented as:  $U_i = W_i/(W_i + V)$ , for  $i = 1, \dots, d$ .

## 4. Gumbel Copula:

The Gumbel copula is defined by the parameter  $\theta \geq 1$  and can be represented as:  $C(u_1, \dots, u_d) = \exp(-((-\log(u_1))^\theta + \dots + (-\log(u_d))^\theta)^{1/\theta})$ .

Stochastic representation:

If  $W_1, \dots, W_d$  are i.i.d. random variables following a standard exponential distribution, and  $V$  is an independent random variable following a gamma distribution with shape parameter  $\theta$  and scale parameter 1, then the Gumbel copula can be represented as:  $U_i = \exp(-W_i/V^{1/\theta})$ , for  $i = 1, \dots, d$ .

★★★☆☆

**MH to reader:** The stochastic representations of Gaussian and  $t$  copulas are correct, but those of Clayton and Gumbel copulas are wrong – they do not even lead to  $U(0, 1)$  margins. It would have also been more meaningful to group these four examples into two categories, the first two examples are copulas of normal variance mixtures, the last two examples are Archimedean copulas. For both of these well-known copula classes, one knows stochastic representations; see also Hofert et al. (2018, Chapter 3) and references therein. A quick overview can be given as follows.

*Normal variance mixtures* are distributions with stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{\mathbf{W}}\mathbf{A}\mathbf{Z}, \quad (1)$$

where (typically, there are other options)  $\boldsymbol{\mu} \in \mathbb{R}^d$ ,  $\mathbf{A}$  is the Cholesky factor of a covariance matrix  $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$  (being lower triangular and satisfying  $\mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ ), and  $\mathbf{W}$  is an almost surely positive random variable independent of  $\mathbf{Z} = (Z_1, \dots, Z_d)$  for  $Z_j \stackrel{\text{ind.}}{\sim} \mathbf{N}(0, 1)$ ,  $j = 1, \dots, d$ .

*Normal variance mixture copulas*, such as the Gaussian and  $t$ , are the copulas of  $\mathbf{X}$  arising from the first part of Sklar's Theorem. So if  $F_1, \dots, F_d$  are the margins of  $\mathbf{X}$ , then  $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$  follows the respective normal variance mixture copula. Similarly for the larger class of elliptical copulas. The popularity of the Gaussian (or: normal) copula and the  $t$



copula comes from the fact that, for these two cases,  $F_1, \dots, F_d$  are known numerically to the degree – and availability in software – that they can be considered known. In case of the Gaussian copula, as ChatGPT correctly identified,  $F_1 = \dots = F_d = \Phi$ , and in case of the  $t$  copula,  $F_1 = \dots = F_d = F_{t_\nu}$  (the distribution function of the standard Student's  $t_\nu$  distribution with  $\nu$  degrees-of-freedom). However, we need to be careful here,  $\Phi$  and  $F_{t_\nu}$  only lead to the correct  $\mathbf{U}$  if we start the construction with  $\mathbf{X}$  in (1) having  $\boldsymbol{\mu} = \mathbf{0}$  and with  $\boldsymbol{\Sigma}$  being a correlation matrix. Why can we make these two assumptions without loss of generality you may ask? Because of the invariance principle:  $\mathbf{Y} = (Y_1, \dots, Y_d)$  with  $Y_j = (X_j - \mu_j)/\sigma_j$ ,  $j = 1, \dots, d$ , and  $\mathbf{X}$  have the same copula!

Clayton and Gumbel copulas are examples of *Archimedean copulas*, that is copulas of the form

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \quad \mathbf{u} \in [0, 1]^d,$$

where  $\psi : [0, \infty) \rightarrow [0, 1]$  is the (*Archimedean*) *generator*. In case of Clayton,  $\psi(t) = (1+t)^{-1/\theta}$ ,  $t \geq 0$ , for  $\theta > 0$ , and in case of Gumbel,  $\psi(t) = \exp(-t^{1/\theta})$ ,  $t \geq 0$ , for  $\theta \geq 1$ . According to Marshall and Olkin (1988), Archimedean copulas with generators  $\psi$  that are Laplace–Stieltjes transforms  $\mathcal{L}\mathcal{S}[F_V]$  of distributions  $F_V$  on the positive real line allow for the stochastic representation

$$\mathbf{U} = \left( \psi\left(\frac{E_1}{V}\right), \dots, \psi\left(\frac{E_d}{V}\right) \right),$$

where  $V \sim F_V$  is independent of  $E_1, \dots, E_d \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ . In case of Clayton,  $F_V$  is the distribution function of the  $\Gamma(1/\theta, 1)$  distribution (the gamma distribution with density  $f_V(v) = \frac{1}{\Gamma(1/\theta)} v^{1/\theta-1} \exp(-v)$ ,  $v > 0$ ), and for Gumbel,  $F_V$  is the distribution function of the  $S(1/\theta, 1, \cos^\theta(\pi/(2\theta)), \mathbb{1}_{\{\theta=1\}}; 1)$  distribution (a stable distribution); see, for example, Hofert et al. (2018, Table 3.1) for more details. This construction principle is widely used for Archimedean copulas. We also learn from it that Archimedean copulas are *conditional independence* copulas, so the components of  $\mathbf{U}$  are conditionally independent given the random variable  $V$ . This often leads to analytically tractable results for Archimedean copulas.

**MH to ChatGPT:** *Thank you for this conversation.*

**ChatGPT's response:**

You're welcome! If you have any more questions or need further clarification on any topic, feel free to ask. Have a great day!

★★★★★

**MH to reader:** For the great conversational abilities of ChatGPT to help us learn more about copulas, we give a 5-star rating here.

### 3 Conclusion

We provided fundamental aspects of stochastic dependence modeling with copulas through the lens of a scholarly exchange with chatbot ChatGPT. Covered topics included the definition of copulas, Sklar's theorem, the invariance principle, pseudo-observations, tail dependence and stochastic representations. In total, ChatGPT reached 3.4 out of 5 stars in 15 questions. Our conclusions are:

- Version 4 of ChatGPT showed partly a remarkable knowledge on general aspects of copula modeling, especially in comparison to version 3.5.
- However, ChatGPT is merely predictive and does not possess built-in logic to get mathematical arguments right. Sometimes it could not even produce well-known facts it should have come across during training (this problem can in principle be solved by an even larger training sample).
- Even if ChatGPT sometimes obtained correct results, they may not be presented in a pedagogically meaningful way, be it because of unusual notation, or, as in most cases, because ChatGPT tries to provide arguments in terms of analytical functions (such as distribution functions), whereas a more pedagogically meaningful approach to learn about dependence is to consider stochastic representations (random variables).
- Without providing references, it is difficult to verify how ChatGPT came up with certain wrong statements. Due to its generative nature (to a great deal non-reproducible), references would probably also not be particularly helpful (after all, ChatGPT is a chatbot, not an online encyclopedia).
- One feels the urge to correct ChatGPT if it gets facts wrong, which may partly come from the feeling one chats with a real person. However, ChatGPT can (currently) not learn this way, its knowledge is derived from past data only.
- Where ChatGPT can be useful is to test one's (already otherwise acquired) knowledge about a subject, for example as exam preparation or study buddy. This provides an interesting way to interact with the topics covered, allows to make connections between the topics, compare them, and thus to learn about them.

## References

- Crouse M (2023). ChatGPT cheat sheet: Complete guide for 2023. TechRepublic, URL: [techrepublic.com/article/chatgpt-cheat-sheet](https://techrepublic.com/article/chatgpt-cheat-sheet).
- Ellis AR, Slade E (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2): 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Embrechts P, Hofert M (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3): 423–432. <https://doi.org/10.1007/s00186-013-0436-7>
- Hassan MM, Knipper A, Santu SKK (2023). ChatGPT as your personal data scientist. arXiv preprint: <https://arxiv.org/abs/2305.13657>.
- Hofert M (2023a). Assessing ChatGPT's proficiency in quantitative risk management. *Risks*, 11(9):166. <https://doi.org/10.3390/risks11090166>
- Hofert M (2023b). Correlation pitfalls with ChatGPT: Would you fall for them? *Risks*, 11(7):115. <https://doi.org/10.3390/risks11070115>
- Hofert M, Kojadinovic I, Mächler M, Yan J (2018). *Elements of Copula Modeling with R*. Use R! Series, Springer. <https://doi.org/10.1007/978-3-319-89635-9>. ISBN: 978-3-319-89635-9.
- Joshi I, Budhiraja R, Dev H, Kadia J, Ataulloh MO, Mitra S, et al. (2023). ChatGPT and the future of undergraduate computer science: Challenges, opportunities and recommendations. arXiv preprint: <https://arxiv.org/abs/2304.14993v2>.
- Lopez-Lira A, Tang Y (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. arXiv preprint: <https://arxiv.org/pdf/2304.07619.pdf>.

- Marshall AW, Olkin I (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403): 834–841. <https://doi.org/10.1080/01621459.1988.10478671>
- Rodger J (2023). Google issues urgent warning to anybody using ChatGPT. Birmingham Live. URL: [birminghammail.co.uk/news/midlands-news/google-issues-urgent-warning-anybody-26227871](http://birminghammail.co.uk/news/midlands-news/google-issues-urgent-warning-anybody-26227871).
- Sibuya M (1959). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics*, 11(2): 195–210.
- Wood DA, Achhpilia MP, Adams MT, Aghazadeh S, Akinyele K, Akpan M, et al. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? *Issues in Accounting Education*. <https://doi.org/10.2308/ISSUES-2023-013>
- Zheng Y (2023). ChatGPT for teaching and learning: An experience from data science education. arXiv preprint: <https://arxiv.org/abs/2307.16650>.

## Afterword

We would like to thank the author for his submission, “The philosophy of copula modeling: a conversation with ChatGPT”. Whilst highly optimistic about the potential of large language models (LLMs), we believe Marius has achieved something that is currently outside the reach of these technologies: almost 20 straight pages of insightful discussion of a statistical method, delivered engagingly and with humor.

This article is a welcomed contribution to the “Philosophy of Data Science” section, where our goal is to apply critical thinking to the use and application to data science’s methods and tools. With tools like ChatGPT reducing the dependency on websites like Stack Overflow, Google, and Wikipedia, we need to think critically about the use of these new tools to instruct on complex methodologies and assimilate new concepts.

The conversation with ChatGPT highlights both the promise and pitfalls of this trend. On one hand, ChatGPT’s ability to immediately summarize a modelling method showcases how AI tools can serve as valuable on-demand learning aids. On the other, its inaccuracies and omissions emphasize the need for discernment and critical evaluation. When should students accept an answer from one of these technologies? With what level of certainty? More importantly, how should student learn to prod and scrutinize the initial answers to ensure that the LLM has not erred or omitted crucial information? To find out, we’ll need to have more conversations... with AI and experts alike.

We look forward to more of these conversations in the future. We are extremely interested in seeing how LLMs would attempt to instruct on other statistics or data science methods. We would like to hear human experts (and LLMs themselves) provide strategies for how non-experts can get accurate answers from these tools.

This is just the beginning of conversations we’d like between humans and machines!

---

Glen Wright Colopy

Section Editor: The Philosophy of Data Science

Jun Yan

Editor: Journal of Data Science