STAT 542: Multivariate Analysis **Spring 2021**

Lecture 12: Copula

Instructor: Yen-Chi Chen

References.

- Chapters 1, 3, and 4 of

  Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. (2010). *Copula theory and its applications* (Vol. 198). New York: Springer.

12.1 Introduction

Copula is a powerful way to model the dependence of a random vector. One key insight is due to the famous Sklar theorem: the distribution of any continuous random vector can be expressed using copula and the marginal distribution. It is easy to estimate the marginals of a random vector, so all we need is to estimate the copula function and this would lead to an estimator of the joint distribution.

12.2 Sklar's theorem and copulas

Let $X \in \mathbb{R}^d$ be a random vector. Let $F(x)$ be the CDF of X, i.e., $F(x) = P(X_1 \leq x_1, \cdots, X_d \leq x_d)$. Further, we denote $F_1, \cdots, F_d$ to be the marginal CDF of $X_1, \cdots, X_n$.

A *copula* is a function $C : [0,1]^d \mapsto [0,1]$ with the following properties:

(C1) *Marginal.* For any $j = 1, \cdots, d$, $C(u_j, u_{-j} = 1) = u_j$, where $u_{-j} = 1$ means that all augments except j-th augment is 1.

(C2) *Isotonic.* $C(u) \leq C(v)$ if $u \leq v$, where $u \leq v$ means that $u_j \leq v_j$ for all $j = 1, \cdots, d$.

(C3) *d-increasing.* C is d-increasing, i.e., for any box $[a, b] \subset [0,1]^d$ with non-empty volume, $C([a, b]) > 0$.

Note that when there are d variables, C is often called the d-copula. A copula can be viewed as *a CDF of d-dimensional random vector U* such that $U_j \sim \mathsf{Unif}[0,1]$.

While the copula may seem to be an abstract object, the following theorem shows its importance to a multivariate CDF.

Theorem 12.1 (Sklar's theorem) *For a random vector X with CDF F and univariate marginal CDFs $F_1, \cdots, F_d$. There exists a copula C such that*

$$F(x_1, \cdots, x_d) = C(F_1(x_1), \cdots, F_d(x_d)).$$

If X is continuous, then such a copula C is unique.

Theorem **??** shows that we can write the joint CDF in terms of marginal CDFs and the copula. Conversely, if we know the joint CDF F and the marginals $F_1, \cdots, F_d$, we can find the copula via

$$C(u_1, \cdots, u_d) = F(F_1^{-1}(u_1), \cdots, F_d^{-1}(u_d)), \tag{12.1}$$

where $F_j^{-1}(t) = \inf\{s : F_j(s) \geq t\}$.

12.2.1 Copula density

Given that a copula can be viewed as a CDF of a uniform random vector, we can then think about the corresponding PDF. Such density function is known as a copula density.

Interestingly, the Sklar theorem (Theorem **??**) implies that the PDF of X can be written as

$$p(x_1, \cdots, x_d) = c(F_1(x_1), \cdots, F_d(x_2))p_1(x_1) \cdots p_d(x_d), \tag{12.2}$$

where $p_j(x_j)$ is the PDF of X_j and c is the copula density. You can think of c as the corresponding PDF of the uniform random vector with joint CDF C.

12.3 Examples of copulas

Theorem **??** shows the power of copulas. Here we introduce some basic examples of copulas.

- *Independence copula.* $C_{\mathsf{ind}}(u_1, \cdots, u_d) = u_1 \times u_2 \times \cdots \times u_d$. This copula corresponds to the case where all random variables are independent.
- *Comonotonicity copula.* $C_{\mathsf{co}}(u_1, \cdots, u_d) = \min\{u_1, \cdots, u_d\}$. If we view the copula as a CDF of a uniform random vector U, this is the case where $U_1 = U_2 = \cdots = U_d$ a.s.
- *Counter-monotonicity copula.* $W_{\mathsf{counter}}(u_1, \cdots, u_d) = \max\{u_1 + u_2 + \cdots + u_d - d + 1, 0\}$. Note that W_{counter} is not a copula except for $d = 2$. In the case of $d = 2$, this corresponds to the case where $U_2 = 1 - U_1$.

These basic example provides a lower and an upper bound of any copula, as stated in the following theorem.

Theorem 12.2 (FréchetHoeffding bounds) *For any d-copula C, we have*

$$W_{\mathsf{counter}}(u) \leq C(u) \leq C_{\mathsf{co}}(u).$$

Moreover, the bound is pointwisely sharp, i.e., for each fixed u,

$$\inf_{C \in \mathcal{C}} C(u) = W_{\mathsf{counter}}(u), \quad \sup_{C \in \mathcal{C}} C(u) = C_{\mathsf{co}}(u).$$

The above examples shows the connection of copulas and uniform random vector U. Based on the Sklar's theorem (Theorem **??**), we know that the copula would provide information on the corresponding random vector X. The following proposition provides a concrete statement about this:

Proposition 12.3 *Suppose the random vector X has a copula C. Let $T_1, \cdots, T_d$ be strictly increasing functions. Then the copula of $(T_1(X_1), \cdots, T_d(X_d))$ is also C.*

The power of Proposition **??** is that any coordinate-wise monotone transformation will not change the underlying copula! As a results, we have the following properties:

- X has a copula $C_{\mathsf{ind}} \Leftrightarrow X_1, \cdots, X_n$ are independent.
- X has a copula $C_{\mathsf{co}} \Leftrightarrow$ there exists a random variable Z and increasing functions $T_1, \cdots, T_d$ such that $X_j = T_j(Z)$.
- In the case of $d = 2$, X has a copula $W_{\mathsf{counter}} \Leftrightarrow$ there exists a strictly decreasing function T such that $X_2 = T(X_1)$.

12.3.1 Common families of copulas

Here we briefly introduce some popular family of copulas.

- *Bivariate Gaussian copula.* For the case of $d = 2$, a popular copula is the bivariate Gaussian copula, which has a parameter θ such that

  $$C(u_1, u_2; \theta) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{\sqrt{2\pi(1-\theta^2)}} \exp\left(-\frac{1}{2(1-\theta^2)}(s^2 - 2\theta st + t^2)\right) dsdt.$$

  Note that the Gaussian copula belongs to a larger family known as the *elliptical copula.*
- *Archimedean copula.* The Archimedean copula is a class of copulas that can be written as

  $$C(u) = \psi(\psi^{-1}(u_1) + \cdots, \psi^{-1}(u_d))$$

  for some suitable function ψ. The *Gumbel-Hougaard copula* is an Archimedean copula with

  $$C_{GH}(u;\theta) = \exp\left(-\left(\sum_{j=1}^{d}(-\log u_j)^{\theta}\right)^{1/\theta}\right).$$

  When $\theta = 1$, we recover the independence copula and when $\theta \to \infty$, we recover the comonotonicity copula. The *Mardia-Takahasi-Clayton copula* is also an Archimedean copula with

  $$C_{MTC}(u;\theta) = \max\left\{\left(\sum_{j=1}^{d} u_j^{-\theta} - (d-1)\right)^{-1/\theta}, 0\right\},$$

  where $\theta \geq \frac{-1}{d-1}$. Note that when $\theta = 0$, this reduces to independence copula.
- *EFGM copula.* The EFGM (Eyraud-Farlie-Gumbel-Morgenstern) copula is a family of copula generated by a product rule and a set of parameter. In the case of $d = 2$,

  $$C_{EFGM}(u_1, u_2; \theta_{12}) = u_1 u_2 (1 + \theta_{12}(1-u_1)(1-u_2)).$$

  In the case of $d = 3$,

  $$\begin{aligned} C_{EFGM}(u_1, u_2, u_3; \theta) = u_1 u_2 u_3 (1 &+ \theta_{12}(1-u_1)(1-u_2) + \theta_{13}(1-u_1)(1-u_3) \\ &+ \theta_{23}(1-u_2)(1-u_3) + \theta_{123}(1-u_1)(1-u_2)(1-u_3)). \end{aligned}$$

  This copula can be generalized to any d by the same reasoning.

Remark. You may be wondering if we can place a parametric model on the copula density. Yes we could do that, but we have to be very careful. The major reason is that the copula density is the PDF of marginally uniform random variables. So there are constraints on how the PDF will be like in the marginal cases. Not any PDF over $[0,1]^d$ will satisfy this constraint.

12.4 Estimation of copula

With a slightly abuse of notations, we let our random sample be random vectors $X_1, \cdots, X_n \in \mathbb{R}^d \sim F$ and let C be the corresponding unknown copula and $F_1, \cdots, F_d$ be the corresponding unknown marginal CDF.

In practice, we need to estimate C and $F_1, \cdots, F_d$ to obtain an estimator of the joint CDF F. There are three popular approaches to this ends: parametric, semi-parametric, and nonparametric approaches.

12.4.1 Parametric approach

A simple parametric approach is that we place a parametric model of each marginal CDF $F_j(\cdot;\alpha_j)$ and let $p_j(\cdot;\alpha_j)$ be the corresponding PDF. Similarly, we also place a parametric model of the copula density $c(u_1, \cdots, u_d; \theta)$.

By the Sklar's theorem and the copula density form in equation (**??**), the joint PDF is

$$p(x;\alpha,\theta) = c(F_1(x_1;\alpha_1), \cdots, F_d(x_d;\alpha_d);\theta) p_1(x_1;\alpha_1)\cdots p_d(x_d;\alpha_d).$$

Thus, the log-likelihood function is

$$\ell(\alpha,\theta|x) = \ell_c(\alpha,\theta|x) + \sum_{j=1}^d \log p_j(\alpha_j|x_j),$$

where $\ell_c(\alpha,\theta|x) = \log c(F_1(x_1;\alpha_1), \cdots, F_d(x_d;\alpha_d);\theta)$.

Ideally, we want to estimate the parameters α, θ by maximizing the above log-likelihood function. But this could be challenging since the parameter α also appears in the copula part.

A common approach to reduce the complexity is that we first estimate α by marginal distribution, i.e.,

$$\widehat{\alpha}_{j,\mathsf{IFM}} = \mathsf{argmax}_{\alpha_j} \sum_{i=1}^n \log p_j(\alpha_j|X_{i,j})$$

and then plug-in this to $\ell_c(\widehat{\alpha}_{\mathsf{IFM}}, \theta|X)$ and solve for θ, i.e.,

$$\widehat{\theta}_{\mathsf{IFM}} = \mathsf{argmax}_\theta \sum_{i=1}^n \ell_c(\widehat{\alpha}_{\mathsf{IFM}}, \theta|X_i).$$

This idea is called *the method of inference functions for margins (IFM).*

12.4.2 Semi-parametric approach

Since the marginal CDF can be estimated easily via the marginal empirical distribution function (EDF), we may only place a parametric model on the copula part and use the marginal EDF to handle the marginal

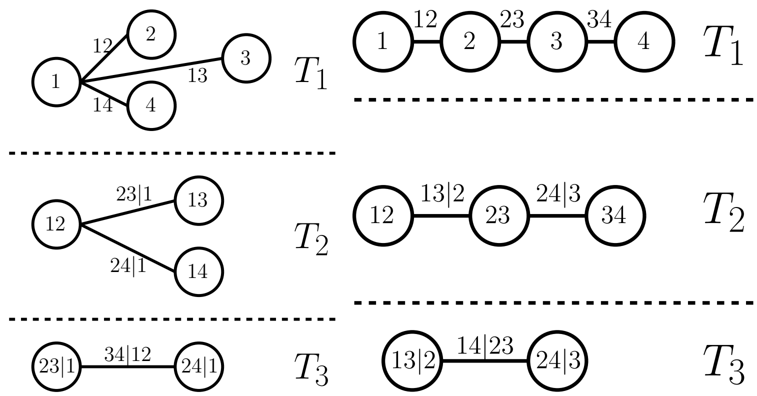

Figure 12.1: Two vine graphs of $p(x_1, \cdots, x_4)$.

CDF. Namely, we only use a parametric model of $c(F_1(x_1), \cdots, F_d(x_d); \theta)$. Specifically, we use

$$\widehat{F}_j(x_j) = \frac{1}{n}\sum_{i=1}^{n} I(X_{i,j} \leq x_j)$$

as an estimator of F_j. Then we compute the likelihood function of θ

$$\ell_c(\theta|x) = \log c(\widehat{F}_1(x_1), \cdots, \widehat{F}_d(x_d); \theta).$$

Note that $\widehat{F}_j^{-1}(X_{i,j})$ is the normalized *rank* of $X_{i,j}$ among $X_{1,j}, \cdots, X_{n,j}$. The normalized rank is the rank divided by n. With this, one can use the MLE to find $\widehat{\theta}$.

12.4.3 Nonparametric approach

Given that the copula $C(u) = F(F_1^{-1}(u_1), \cdots F_d^{-1}(u_d))$, a simple nonparametric estimator is

$$\widehat{C}(u) = \widehat{F}(\widehat{F}_1^{-1}(u_1), \cdots \widehat{F}_d^{-1}(u_d)),$$

where $\widehat{F}$ is the joint EDF of $X_1, \cdots, X_d$ and $\widehat{F}_j$ is the marginal EDF.

While this estimator has nice theoretical properties, it does not give a smooth copula and we cannot use it to estimate the copula density.

If the goal is to obtain a smooth copula or a copula density estimator, we may apply a density estimator using $\widehat{U}_1, \cdots, \widehat{U}_n$,, where

$$\widehat{U}_{i,j} = \widehat{F}_j(X_{i,j}).$$

Namely, we first transform $X_1, \cdots, X_n$ based on the marginal ranks so they become rank vectors $\widehat{U}_1, \cdots, \widehat{U}_n$. Then we apply a nonparametric density estimator to $\widehat{U}_1, \cdots, \widehat{U}_n$. One can use the KDE, histogram, basis or any estimator that will be suitable in this case.

12.5 Pair-copula and vine construction

The vine copula is a representation of the joint distribution using a set of bivariate copulas. To give a concrete example, consider $d = 4$ cases. We can factorize the joint PDF as

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_1, x_2, x_3). \tag{12.3}$$

Here is an interesting insight from the bivariate copula density. Any joint PDF of two variables, say X_1, X_2, can be factorized as

$$p(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2))p(x_1)p(x_2). \tag{12.4}$$

The above factorization can be applied to conditional PDF as well:

$$p(x_1, x_2|x_j) = c_{12|j}(F_{1|j}(x_1|x_j), F_{2|j}(x_2|x_j))p(x_1|x_j)p(x_2|x_j) \tag{12.5}$$

for $j = 3, 4$. Equation (**??**) further implies that

$$p(x_1|x_2, x_j) = \frac{p(x_1, x_2|x_j)}{p(x_2|x_j)} = c_{12|j}(F_{1|j}(x_1|x_j), F_{2|j}(x_2|x_j))p(x_1|x_j).$$

Thus,

$$\begin{aligned}
p(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2))p(x_2) \\
&= c_{12} \cdot p_2, \\
p(x_3|x_1, x_2) &= c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1))p(x_3|x_1) \\
&= c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1))c_{13}(F_1(x_1), F_3(x_3))p(x_3) \\
&= c_{23|1} \cdot c_{13} \cdot p_3 \\
p(x_4|x_1, x_2, x_3) &= c_{34|12}(F_{3|12}(x_3|x_1, x_2), F_{4|12}(x_4|x_1, x_2))p(x_4|x_1, x_2) \\
&= c_{34|12} \cdot c_{24|1} \cdot c_{14} \cdot p_4.
\end{aligned}$$

As a result, equation (**??**) can be written as

$$p(x_1, x_2, x_3, x_4) = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{34|12}. \tag{12.6}$$

Note that each copula here is a bivariate copula (the conditional component only affects the augment's distribution, not the copula function). So we only need specify/estimate each bivariate copula and the corresponding conditional CDF to obtain an estimator of the joint PDF. The factorization in equation (**??**) is known as Canonical vine representation (C-vine). It is called *vine* because there is an elegant graphical representation of how each component is constructed; see the left panel of Figure **??**.

An important note is that the factorization in equation (**??**) is *NOT unique*. The same joint distribution can admit several different factorizations, which leads to different vine representations. The C-vine is one of the most popular representation. Another popular factorization is the D-vine (D: drawable):

$$p(x_1, x_2, x_3, x_4) = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{14|23}. \tag{12.7}$$

Equation (**??**) is based on the graph in the right panel of Figure **??**.

This idea can be applied to any arbitrary d. So we can factorize the joint distribution using a vine copula representation. Note that while different vine graphs lead to the same distribution, when we place parametric models on copulas, they may lead to different joint distributions. So some works have been proposed to choose the vine graph that is best fitted to the observed data. Formally, the 'best fitted vine graph' is under the specific models of bivariate copulas. If we estimate all copulas (and conditional CDFs) nonparametrically, any vine graph representation will eventually lead to the same distribution.

Here is a paper that provides a gentle introduction on vine copulas:

> Czado, C. (2010). Pair-copula constructions of multivariate copulas. In Copula theory and its applications (pp. 93-109). Springer, Berlin, Heidelberg. https://mediatum.ub.tum.de/doc/1079253/file.pdf

12.6 Compatibility and Fréchet classes

The compatibility is an important issue when we are trying to use copula to analyze the data. To give a concrete example, consider three uniform random variables U_1, U_2, U_3. When we analyze three variables together, we use a 3-copula $C(u_1, u_2, u_3)$. If now we ignore one of the variable and focus on the other two, we will be using 2-copula. There will be three possible 2-copulas C_{12}, C_{13}, C_{23}, depending on which two variables we are using. If we start with 3-copula and derive the implied 2-copula, we are generally fine. But if we start with the three 2-copulas C_{12}, C_{13}, C_{23}, we may not be able to find a 3-copula whose marginals agree with these 2-copulas. Namely, *any three arbitrary 2-copulas C_{12}, C_{13}, C_{23} may NOT lead to a 3-copula.* Given C_{12}, C_{13}, C_{23}, if there exists a 3-copula whose marginals are these three, then we say that C_{12}, C_{13}, C_{23} are compatible. If C_{12}, C_{13}, C_{23} are compatible, the set of 3-copulas with these three 2-copulas as the marginals is called the Fréchet class.

There has been some results on the compatibility of 2-copula, see, e.g.,

> Durante, F., Klement, E. P., & Quesada-Molina, J. J. (2007). Copulas: compatibility and Fréchet classes. arXiv preprint arXiv:0711.2409.

But this is still an open question for general d-copula cases.