Introduction about Copulas.

Duc T. Nguyen

Ulm University

November 29, 2020

The talk is based on two following references:

1. **An Introduction to Copulas.**
   (2006) Roger B. Nelsen, Springer.
2. **Elements of Copula Modeling with R.**
   (2018) M. Hofert, I.Kojadinovic, M. Machler, J. Yan, Springer.

Outline

- First part: Motivating Example.
- Second part: Analysis Approach to Copula.
  - Subcopula and Copula
  - Sklar's Theorem.
- Third part: Copulas and Random variables.
  - Sklar's Theorem.
  - Fréchet-Hoeffding Bounds
- Forth part: Survival Copula and Symmetric Copula.

PART 1: Motivating Example.

The data set.

- Let $(X_1, X_2)^{(n)} = (X_{i_1}, X_{i_2})_{i=1}^n$, $(Y_1, Y_2)^{(n)} = (Y_{i_1}, Y_{i_2})_{i=1}^n$ be two given data set.

Goal: Comparing the dependence between between X_1 and X_2 with Y_1 and Y_2.

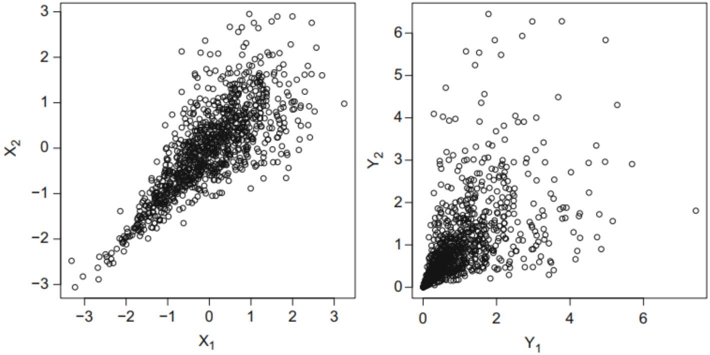

Figure: $n = 1000$ independent observations of (X_1, X_2) and (Y_1, Y_2).

Kernel Estimation.

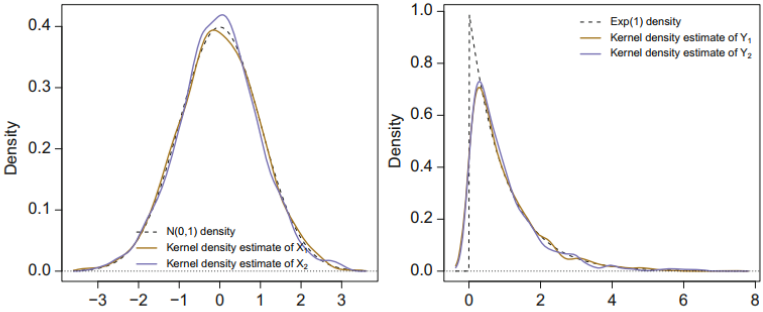

Figure: Kernel density estimates of the densities of (X_1, X_2) and (Y_1, Y_2).

- X_1 and X_2 are likely to follow **standard normal distribution**.
- Y_1 and Y_2 are likely to follow **standard exponential distribution**.

Transform to the uniform distributions.

Lemma: Let $\boldsymbol{X}$ be continuous random variable with distribution function $\boldsymbol{F}$, then $\boldsymbol{F(X)}$ is a standard uniform random variable, i.e., $\boldsymbol{F(X) \sim U(0,1)}$

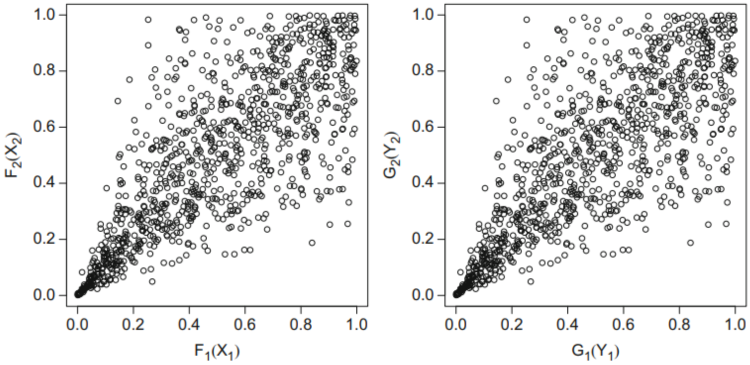

Figure: Scatter plots of $\boldsymbol{(F(X_1), F(X_2))}$ and $\boldsymbol{(G(Y_1), G(Y_2))}$

Comments.

- The distribution of $(F(X_1), F(X_2))$ and $(G(Y_1), G(Y_2))$ seem to be identical.
- Two given data sets is indistinguishable in terms of **dependence** and only differ in terms of the underlying **marginal distribution function**.

Question: Is there any tool to measure the **dependence** between random variables?

Structure of the talk

Analysis approach to Copula.

$\Downarrow$

Sklar's Theorem.

$+$

Fréchet-Hoeffding Bounds.

$\Downarrow$

Copula and Random Variable.

PART 2: Analysis approach to Copula.

- A rectangle K in $\overline{\mathbb{R}}^2$ is denoted by:

$$K = [x_1, x_2] \times [y_1, y_2].$$

- For any function $H : \overline{\mathbb{R}}^2 \to \mathbb{R}$, $DomH$ is its domain and $RanH$ is its range. Assume that $DomH = S_1 \times S_2$.
- Denote a_1, a_2 be the least elements of S_1, S_2 respectively and b_1, b_2 be the greastest elements of S_1, S_2 respectively
- $H - volume$ of K:

  $$V_H(K) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

  - H deduces a measure on $\overline{\mathbb{R}}^2$.

2 - increasing function.

- A 2 - place real-valued function H is **2 - increasing** function if $V_H(K) \geq 0$ for all rectangles within $\mathit{Dom}H$.
- Let a_1, a_2 be the least elements of S_1, S_2 respectively. The function $H : S_1 \times S_2 \to \mathbb{R}$ is **grounded** if

$$H(x, a_2) = 0 = H(a_1, y) \text{ for all } (x, y) \in S_1 \times S_2.$$

Remark:

$$2 - \textbf{increasing} \nrightarrow \textbf{grounded} : H(x, y) = (2x - 1)(2y - 1)$$
$$\textbf{grounded} \nrightarrow 2 - \textbf{increasing} : H(x, y) = \max(x, y)$$

Lemma: If H be a **grounded, 2 - increasing** function H is nondecreasing in each argument.

Subcopula.

- A two - dimensional real-valued subcopula C' is a function which has the following properties:
  - ***Dom*** C' is an subset of square unit $\mathbb{I}^2$ containing the points 0 and 1.
  - C' is **grounded** and **2 - increasing**.
  - For every (u, v) in $S_1 \times S_2$, $C'(u, 1) = u$, $C'(1, v) = v$.
- A two - dimensional real-valued **copula** is a **subcopula** whose domain is $\mathbb{I}^2$.

Example of copula.

- Considering the **Frank** Copula

$$C^F(u, v) = \frac{1}{9} \log\left(1 + \frac{(e^{9u} - 1)(e^{9v} - 1)}{e^9 - 1}\right), (u, v) \in [0, 1]^2$$

One can check it satisfies three conditions to be a copula.

Important Lemmas

Lemma: Let C' be a subcopula. Then for every $(u_1, v_1), (u_2, v_2)$ in $DomC'$,

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Hence C is **uniformly continuous** on its domain.

Lemma: Let C' be a subcopula. Then for any (u, v) in $DomC'$,

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v).$$

Lemma: Let C be a copula. For any v in $\mathbb{I}$, the partial derivative $\partial C(u, v)/\partial u$ exist for almost all u, and for such v and u,

$$0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1.$$

Distribution Function.

- A distribution function is a function F with domain $\overline{\mathbb{R}}$ such that
  - F is nondecreasing
  - $F(-\infty) = 0$ and $F(\infty) = 1$

- A joint distribution function is a function H whose domain $\overline{\mathbb{R}}^2$ such that
  - H is 2 - increasing
  - $H(x, -\infty) = 0 = H(-\infty, y)$ and $H(\infty, \infty) = 1$.

- The margins F and G of H are given by and .
  - $F(x) = H(x, \infty)$
  - $G(y) = H(\infty, y)$

Sklar's Theorem

Let H be a joint distribution function with margins F and G. Then there exists a **copula** C such that for all x, y in $\overline{\mathbb{R}}$,

$$H(x,y) = C(F(x), G(y)). \tag{1}$$

If F and G are continuous, then C is **unique**; otherwise, C is **uniquely determined** on $RanF \times RanG$.

Conversely, if C is a copula and F, G are distributions function, then the function H defined by (1) is a joint distribution function with margins F and G.

Procedure of the proof.

There exists a unique subcopula $C'; H(x, y) = C'(F(x), G(y))$.

$+$

There exists a copula C, it is coincide with C' over the $DomC'$.

$+$

If F and G are continuous, then $RanF = RanG = \mathbb{I} \to C \equiv C'$.

$\Downarrow$

Theorem is proved.

Lemma 1

Let H be the joint distribution function with margins F and G. Then there exists a unique subcopula C' such that

- $DomC' = RanF \times RanF$
- For all x, y in $\overline{\mathbb{R}}$, $H(x, y) = C'(F(x), G(y))$.

Proof: For any $(x_1, y_1), (x_2, y_2) \in \overline{\mathbb{R}}^2$, one has

$$|H(x_2, x_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

$H(x_1, y_1) = H(x_2, y_2)$ when $(F(x_1), G(y_1)) = (F(x_2), G(y_2))$.
Define the following mapping:

$$C' : RanF \times RanG \to [0, 1]$$

where $C'(F(x), G(y)) := H(x, y)$.
Hence, it is unique on $DomC' = RanF \times RanF$.

Proof.

Lemma 2

Let C' be a subcopula. Then there exists a copula C such that it is coincide with C' over the $DomC'$ and this extension is non-unique.

Proof: Because of the continuity of the subcopula C', it will be a subcopula C'' over $\overline{S_1} \times \overline{S_2}$.

For any point $(a, b) \in \mathbb{I}^2$, denote by

- a_1, b_1 be the greatest element in S_1, S_2 respectively such that

$$a_1 \leq a, b_1 \leq b$$

- a_2, b_2 be the least element in S_1, S_2 respectively such that

$$a_1 \geq a, b_1 \geq b$$

Proof.

$$\lambda_1 = \begin{cases} (a - a_1)/(a_2 - a_1), & \text{if } a_1 < a_2 \\ 1, & \text{if } a_1 = a_2 \end{cases}$$

$$\mu_1 = \begin{cases} (b - b_1)/(b_2 - b_1), & \text{if } b_1 < b_2 \\ 1, & \text{if } b_1 = b_2 \end{cases}$$

$$C(a,b) = (1-\lambda_1)(1-\mu_1)C''(a_1,b_1) + (1-\lambda_1)\mu_1 C''(a_1,b_2) + \lambda_1(1-\mu_1)C''(a_2,b_1) + \lambda_1\mu_1 C''(a_2,b_3)\,.$$

One can check $C(a,b)$ is **grounded** and **2 - increasing**.

Creating a rectangle B in this unit square requires one more point $(c,d) \in \mathbb{I}^2$ with $c \geq a, d \geq b$.

Let $c_1, d_1, c_2, d_2, \lambda_2, \mu_2$ in the same way with a and b.

Proof.

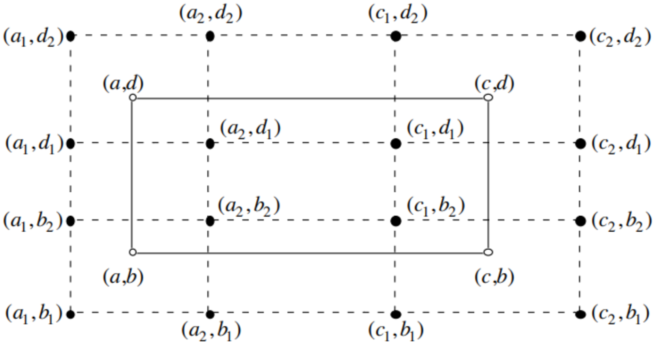

Figure: Scheme of points.

The C - volume of B can be decomposed into the summation of nine C - volumes of nine sub-rectangles, which are all non-negative.

PART III. Copulas and Random variables.

Lemma: Let X be continuous random variable with distribution function F, then $F(X)$ is a standard uniform random variable, i.e., $F(X) \sim U(0,1)$.

- For any random variable X with distribution function F, the *quantile function* $F^{\leftarrow}$ is defined by

$$F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}, \qquad y \in [0,1],$$

with the convention that $\inf \emptyset = \infty$.

Lemma: Let $X \sim U(0,1)$ and F be any distribution function. Then $F^{\leftarrow}(X)$ is a random variable with distribution function F.

Sklar's Theorem

Let X and Y be random variables with distribution functions F and G, respectively, and joint distribution function H. Then there exists a copula C satisfying. Then there exists a **copula** C such that for all x, y in $\overline{\mathbb{R}}$,

$$H(x, y) = C(F(x), G(y)). \tag{2}$$

If F and G are continuous, then C is **unique**; otherwise, C is **uniquely determined** on $RanF \times RanG$.

- The copula in this theorem will be called the **copula of X and Y**, denoted C_{XY}

Remark.

- Copula is a multivariate distribution function with standard uniform univariate margins, that is, $U(0,1)$ margins.

- For any $(u,v) \in RanF \times RanG$, (2) can be rewritten as follows
$$H(F^{\leftarrow}(u), G^{\leftarrow}(v)) = C(u,v).$$

- A random vector has a continuous joint distribution function if and only if it has continuous margins.

- If one can estimate the copula and the margin from a given data, then the multivariate is provided.

Example.

Let X, Y be random variables with joint distribution H

$$H(x,y) = \begin{cases} \dfrac{(x+1)(e^y-1)}{x+2e^y-1}, & (x,y) \in [-1,1] \times [0,\infty] \\ 1-e^{-y}, & (x,y) \in (1,\infty] \times [0,\infty] \\ 0, & \text{else where.} \end{cases}$$

with margins F, G given by

$$F(x) = \begin{cases} 0, & x < -1 \\ (x+1)/2, & x \in [-1,1] \\ 1, & x > 1 \end{cases}, \quad G(y) = \begin{cases} 0, & y < 0 \\ 1-e^{-y}, & y \geq 0 \end{cases}$$

Example.

Obviously F, G are continuous functions, $RanF = RanG = \mathbb{I}$ and the quantile functions of F, G are

$$F^{\leftarrow}(x) = 2x - 1, \qquad G^{\leftarrow}(y) = -\ln(1 - v).$$

Because $RanF = RanG = \mathbb{I}$, one can find the copula

$$C(u, v) = \frac{uv}{u + v - uv}.$$

Hence,

$$H(x, y) = \frac{F(x)G(y)}{F(x) + G(y) - F(x)G(y)}$$

Independent Copula

Theorem

Let X and Y be continuous random variable. Then X and Y are independent if and only if $C_{XY} = \prod$

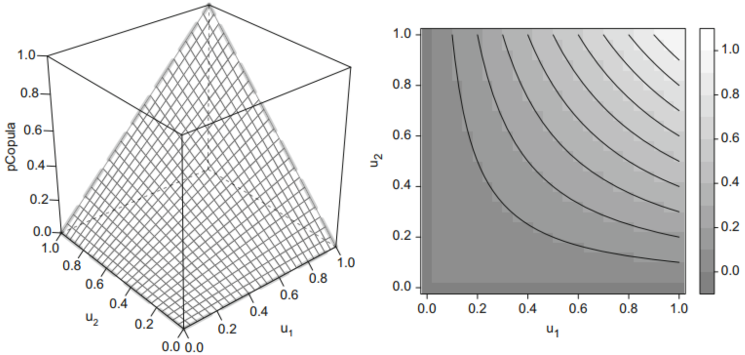

Figure: Graph of product copula and its contour graph.

Invariance principle.

Theorem.

Let X and Y are continuous random variables with distribution functions F and G respectively and a copula C_{XY}. If α, β are strictly increasing on $RanX \times RanY$, respectively, then

$$C_{XY} = C_{\alpha(X)\beta(Y)}.$$

- If α is strictly increasing and β is strictly decreasing, then

$$C_{\alpha(X),\beta(Y)}(u,v) = u - C_{XY}(u, 1-v)$$

- If α is strictly decreasing and β is strictly increasing, then

$$C_{\alpha(X),\beta(Y)}(u,v) = v - C_{XY}(1-u, v)$$

- If α and β are both strictly decreasing, then

$$C_{\alpha(X),\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v)$$

Theorem.

If X and Y are two random variables with F, G are their distribution function respectively. Then for any x, y in $\overline{\mathbb{R}}$, and any u, v in $[0, 1]$, one has

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$$

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y))$$

- Call $W(u, v) = \max(u + v - 1, 0)$, $M(u, v) = \min(u, v)$. Then $W(u, v)$ and $M(u, v)$ are copulas

The Fréchet-Hoeffding Bounds

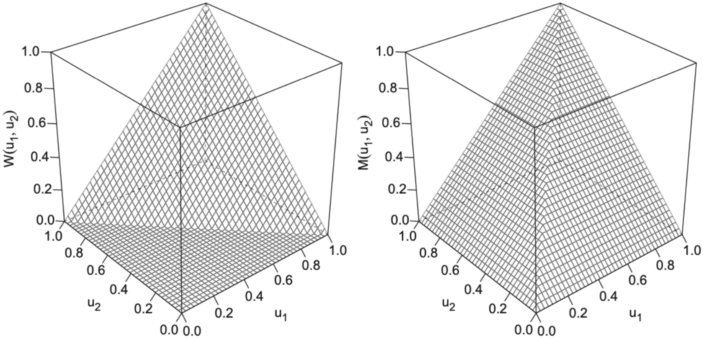

Figure: The Fréchet-Hoeffding Bounds Lower bound and Upper bound.

Example

- Considering the **Frank** Copula

$$C^F(u,v) = \frac{1}{9}\log\left(1 + \frac{(e^{9u}-1)(e^{9v}-1)}{e^9 - 1}\right), (u,v) \in [0,1]^2$$

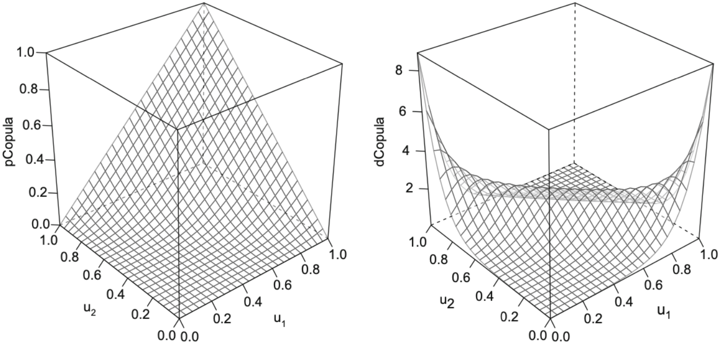

Figure: Frank copula and its density.

Example

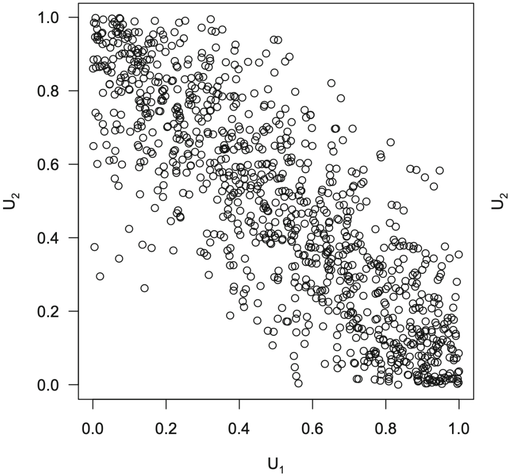

Figure: $n = 1000$ independent observations $(U_1, U_2) \sim C^F$.

Question

Natural Question: What happens when the equality occurs?

- A subset S of $\overline{\mathbb{R}}^2$ is nondecreasing if for any $(x, y), (u, v)$ in S, $x < u$ implies $y \leq v$.

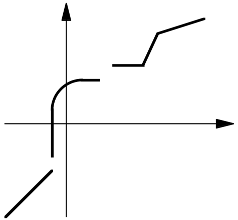

Figure: An example of nondecreasing set.

- A subset S of $\overline{\mathbb{R}}^2$ is nonincreasing if for any $(x, y), (u, v)$ in S, $x < u$ implies $y \geq v$.

The equalities.

- The **support** of a copula C is the complement of the union of all open subset of $\mathbb{I}^2$ with C - measure zero.

Theorem.

Let X and Y be random variables with joint distribution function H. Then H is equal to its **Fréchet-Hoeffding upper bound** if and only if the support of H is a **nondecreasing subset** of $\overline{\mathbb{R}}^2$.

Analogously, H is equal to its **Fréchet-Hoeffding lower bound** if and only if the support of H is a **nonincreasing subset** of $\overline{\mathbb{R}}^2$.

Example: The Fréchet-Hoeffding upper bound is a copula and its support set is the main diagonal.