M5L11

Introduction to Copulas

1. Introduction

In this lecture, the theory of copula is introduced. The definition of copula function with explanation, its properties, basic terminologies, Fréchet-Hoeffding Bounds, Sklar theorem and measures of dependence are discussed.

2. Introduction to Copula

Origin of the word *copula* is the Latin word *copulare*, which means 'to join together'. In many cases of statistical modeling, it is essential to obtain the joint pdf between two or more random variables. Even though the marginal distributions of each of the random variables are known, their joint distributions may not be easy to derive from these marginal distributions. If the information on scale-free measures of dependence between random variables is available, *copula* can be used to obtain their joint distribution.

3. Definition of Copula

3.1. Classical Definition

A copula, C, is a function that joins or couples multiple distribution functions to their one-dimensional marginal distribution functions (*Nelsen*, 2006). Application of copula to probability and statistics is achieved through Sklar Theorem (*Sklar*, 1959), which states that, if $H_{X,Y}(x,y)$ is a joint distribution function, then there exists a copula $C(u,v)$ such that for all x, y in $\overline{R} \in (-\infty,\infty)$,

$$H_{X,Y}(x,y) = C(F_X(x),\ F_Y(y)) \tag{1}$$

where, $F_X(x)$ and $F_Y(y)$ are the marginal distributions of X and Y respectively.

The conceptual definition of copula can be expressed in another way also. Let X and Y be a pair of random variables with cumulative distribution functions (CDF) of $F_X(x)$ and $F_Y(y)$ respectively. Also let their joint CDF be $H_{X,Y}(x,y)$. Each pair, (x, y), of real numbers leads to a point, $(F_X(x), F_Y(y))$, in the unit square $[0,1]\times[0,1]$ and this ordered pair, in turn,

corresponds to a number, $H_{X,Y}(x, y)$, in [0,1]. This correspondence is indeed a function, which is known as Copula (*Nelsen*, 2006).

It is worthwhile to note here that such correspondence is independent of the marginal distributions of the random variables. In other words, any form of marginal distribution can be coupled to get their joint distribution, which is the reason for the popularity of copula theory in many areas of research.

3.2. Mathematical Definition

A 2-dimensional copula (2-copula) is a function C' having following properties:

i. $DomC' = S_1 \times S_2$, where S_1 and S_2 are subsets of I containing 0 and 1, *i.e.*, $\mathrm{I} = [0,1]$.

ii. C' is grounded and 2-increasing.

iii. For every u in S_1 and v in S_2, $C'(u,1) = u$ and $C'(1,v) = v$

4. Basic Terminologies

4.1. 2-increasing Function

The concept of 2-increasing function is basically 2-dimensional analog of non-decreasing functions in one dimension. A 2-dimensional real function is 2-increasing if H-volume of any rectangle B, *i.e.* $V_H(B) \geq 0$ for all rectangles B, whose vertices lie in $DomH$. This is graphically explained in figures1below.

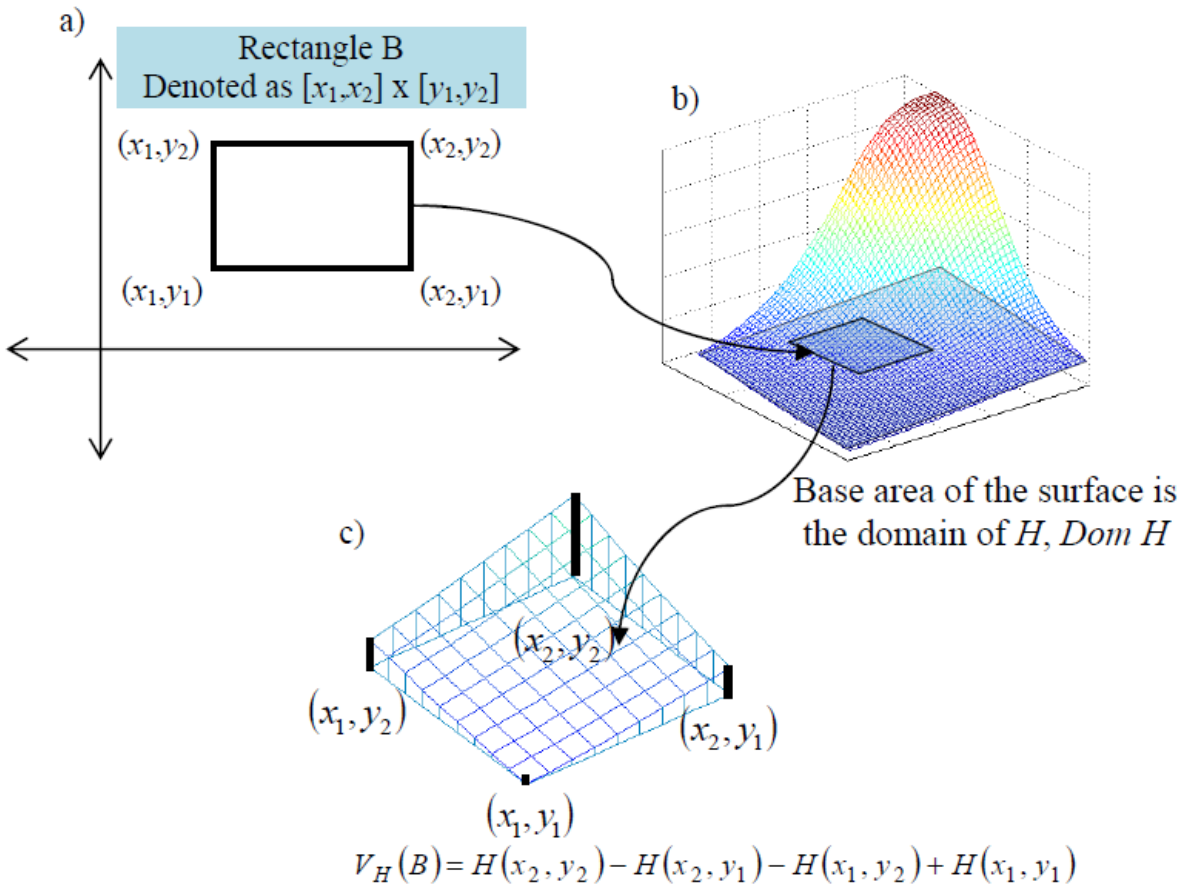


Fig. 1: (a) Rectangle B denoted as $[x_1, y_1]\times[x_2, y_2]$; (b) Joint CDF $H_{X,Y}(x, y)$; (c) H -volume of B denoted as $V_H(B)$ in clockwise direction respectively

4.2. Grounded Function

Copula functions are grounded. Graphical representation of 2-dimensional grounded function is shown in fig. 2. A function is known as zero if one of variable is zero, the value of the function is zero. A 2-increasing function does not imply that H is non-descending in each argument. However, if it is 2-increasing and grounded, then it also can be non-decreasing in each argument.

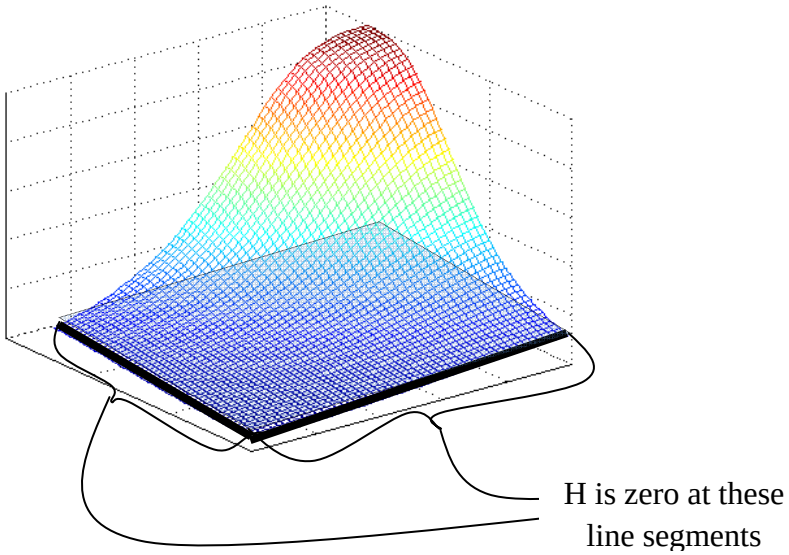


Fig. 2: 2-dimensional Grounded Function

4.3. Properties of a copula function C(u,v)

A copula having domain of I^2 has the following property:

1) For every u_1, u_2, v_1, v_2 in I if such that $u_1 \le u_2$ and $v_1 \le v_2$, then

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \ge 0 \qquad (2)$$

This property indicates that the copula functions are 2-increasing.

2) For every u, v in I,

$$C(u,0) = 0 = C(v,0) \qquad (3 \text{ and } 4)$$

This property indicates that the copula functions are grounded

3) For every u, v in I,

$$C(u,1) = u \qquad (5)$$

$$C(1,v) = v \qquad (6)$$

5. Graphical Representation of Copula

$C(u,v) = uv$ for $0 \le u,v \le 1$ is an example of a copula function. This is the copula function for two independent random variables. This copula function is graphically represented in 3D function plot (Fig. 3) or in 2D contour plots (Fig. 4).

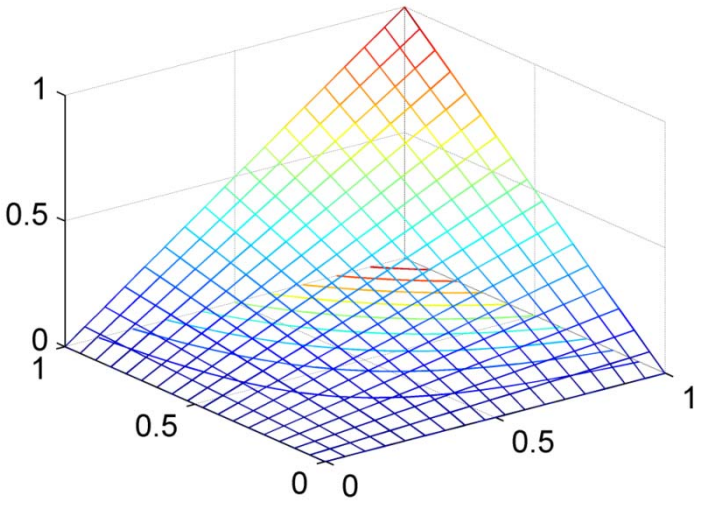


Fig. 3. Copula function in 3D function plot

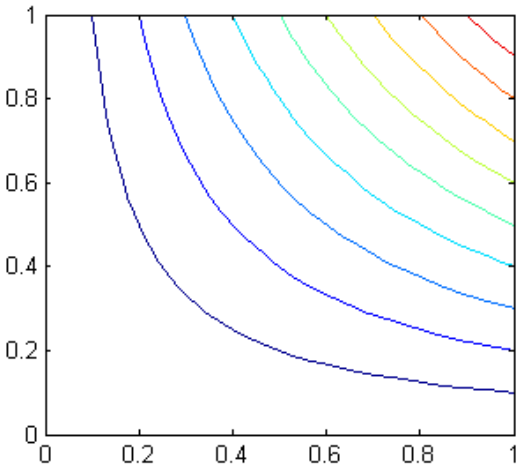


Fig 4. Copula function in contour plot

6. Fréchet-Hoeffding Bounds

Let C' be a copula. Then for every (u,v) in $DomC'$,

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

or

$$W(u,v) \le C'(u,v) \le M(u,v)$$

where $W(u,v) = \max((u+v-1),0)$ and $M(u,v) = \min(u,v)$

The graphical representation for $W(u,v)$ and $M(u,v)$ are given below in figures 5 and 6 respectively.

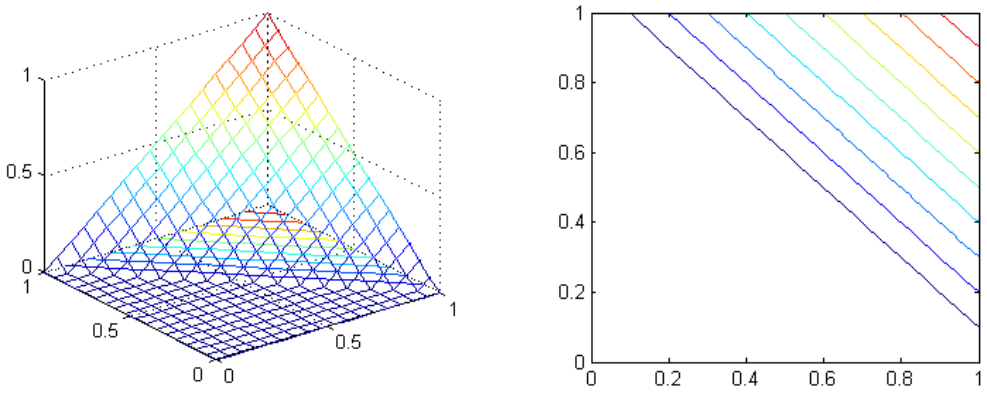


Fig. 5. Graphical representation of $W(u,v)$

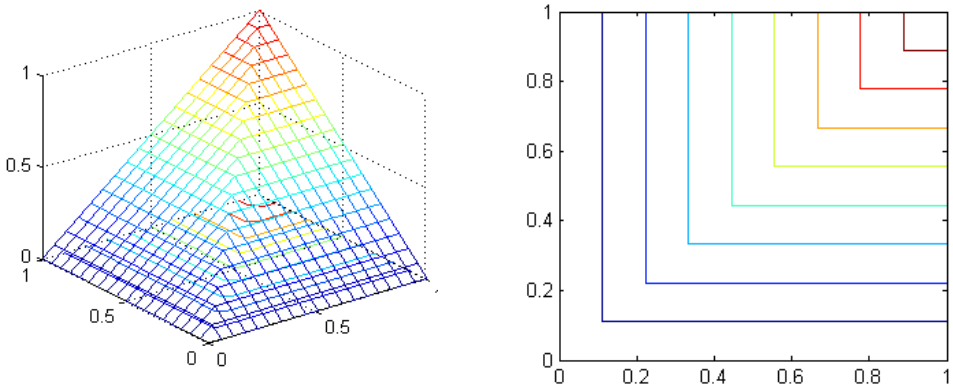


Fig. 6. Graphical representation of $M(u,v)$

7. Nonparametric Measures of Association

Two different nonparametric measures of association are

a) Spearman's rank correlation coefficient or Spearman's rho (ρ_S)

b) Kendall rank correlation coefficient or Kendall's Tau (τ)

For a sample (x_i, y_i) of size n, to compute ρ_S, the samples are first transformed to their respective ranks (R_i^x, R_i^y) and ρ_S is expressed as

$$\rho_S = \frac{\sum_{i=1}^{n}\left(R_i^x - \overline{R}^x\right)\left(R_i^y - \overline{R}^y\right)}{\sqrt{\sum_{i=1}^{n}\left(R_i^x - \overline{R}^x\right)^2 \sum_{i=1}^{n}\left(R_i^y - \overline{R}^y\right)^2}} \tag{7}$$

If there is tie between two or more observations, an average of tie ranks are assigned to all those ties.

The τ is expressed as

$$\tau = \frac{N_c - N_d}{\frac{1}{2}n(n-1)} \tag{8}$$

where N_c and N_d are the number of concordant and discordant pair respectively. If $(x_i - x_j)(y_i - y_j) > 0$ then the pair (x_i, y_i) and (x_j, y_j) is known as concordant pair and if $(x_i - x_j)(y_i - y_j) < 0$ then the pair (x_i, y_i) and (x_j, y_j) is known as discordant pair. If $(x_i - x_j)(y_i - y_j) = 0$, i.e., if there is tie, these pairs are neither concordant or discordant. Further details on these measures can be found elsewhere.

These measures are related to the copula through following expressions:

$$\rho_S = 12\iint uv\,dC(u,v) - 3 \tag{9}$$

and

$$\tau = 4\int C(u,v)\,dC(u,v) - 1 \tag{10}$$

These expressions can be further simplified for Archimedean copulas. These will be discussed in the next lecture.

8. Concluding Remarks

It is important to remember that copula is an effective tool to obtain joint CDF from individual marginal distributions. In this lecture the theory of copula function is introduced. Classical definition is discussed in detail with mathematical expressions and graphical representations. A popular class of copula, known as Archimedean copula, will be discussed in the next lecture.

References:

Nelsen, R. B. (2006), *An introduction to copulas*, Springer series in Statistics, 269 pp., Springer, New York.

Sklar, A. (1959), Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Stat. Univ. Paris, 8, 229–231.