Introduction to Copula Functions

Mahdi Pakdaman
Intelligent System Program

Outline

- Mathematics background
- Copula Definition
- Sklar's Theorem
- Basic Copulas
- Properties
- Dependecy
- Constructing copulas
- Copula Estimation

Basic of joint distributions

- Joint distribution of set of variables $(Y_1, \ldots, Y_m)$ is

$$F(y_1, \ldots, y_m) = Pr[Y_i \le y_i; i = 1, \ldots, m]$$

Survival function corresponding to $F(y_1, \ldots, y_1)$ is

$$\bar{F}(y_1, \ldots, y_m) = \Pr[Y_i > y_i; i = 1, \ldots, m]$$

$$= 1 - F(y_1) \qquad for\ m = 1$$

$$= 1 - F_1(y_1) - F_2(y_2) + F_{12}(y_1, y_2) \qquad for\ m = 2$$

$$= 1 - F_1(y_1) - F_2(y_2) - F_3(y_1) + F_{12}(y_1, y_2)$$

$$+F_{13}(y_1, y_3) + F_{23}(y_2, y_3) - F(y_1, y_2, y_3) \qquad for\ m = 3$$

Basic of joint distributions(2)

- Joint distribution function

$$H(x,y) = P[X \le x, Y \le y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f(X, Y) d_X\, d_Y$$

(1) $lim_{y_j \to -\infty} F(y_1, y_2) = 0, j = 1,2;$

(2) $lim_{y_j \to \infty \forall j} F(y_1, y_2) = 1,$

(3) By the rextangle inequality, for all (a_1, a_2) and (b_1, b_2) with $a_1 \le b_1, a_2 \le b_2$,

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \ge 0.$$

- Condition 3 is referred to as F is 2-increasing . If F has second derivative, then it is equivalent to $\partial^2 F / \partial_{y_1} \partial_{y_2} \ge 0$

Basic of joint distributions(2)

- Marginalization of distributions:

$$F(x) = P[X \le x] = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(X,Y) d_X \, d_Y$$

$$G(y) = P[Y \le y] = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f(X,Y) d_X \, d_Y$$

For each pair (x, y), we can associate three numbers: F(x), G(y) and H(x, y)

- Pseudo-generalized inverse of distribution function F is:

$$F_j^{-1}(t) = \inf[y_j : F_j(y_j) \ge t, 0 < t < 1]$$

Basic of joint distributions(3)

- Probability integral transform Theorem:

Suppose that a random variable X has a continuous distribution for which the cumulative distribution function is F.

Random variable Y defined as $Y = F_x(X)$ has a uniform [0,1] distribution.

History

- 1959: The word *Copula* appeared for the first time (Sklar 1959)
- 1981: The earliest paper relating copulas to the study of dependence among random variables (Schweizer and Wolff 1981)
- 1990's: Copula booster: Joe (1997) and Nelson (1999).
- 1990's +: Academic literatures on how to use copulas in risk management and other applications.
- 2009: Accused of "bringing the world financial system to its knees" (Wired Magazine)

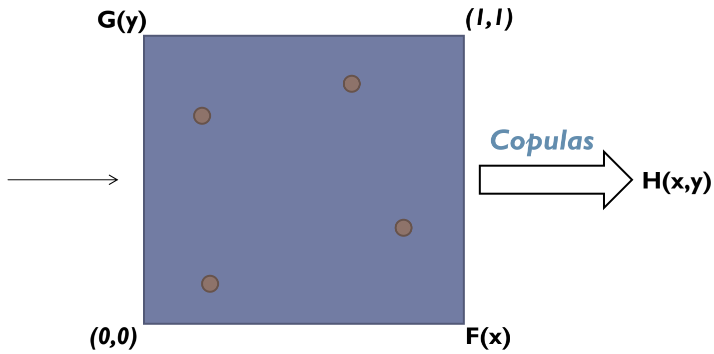
Copula : Definition

The Word *Copula* is a Latin noun that means

"A link, tie, bond"

(Cassell's Latin Dictionary)

Copula: Conceptual



- Each pair of real number (x, y) leads to a point of (F(x), G(y)) in unit square [0, 1]*[0, 1]

Copula: Conceptual

- The mapping, which assigns the value of the joint distribution function to each ordered pair of values of marginal distribution function is indeed a copula.

Copula : informal

- A *2-dimensional copula* is a distribution function on [0, 1]*[0, 1], with standard uniform marginal distributions.

Copula : informal (2)

- If (X, Y) is a pair of continuous random variables with distribution function $H(x,y)$ and marginal distributions $F_x(x)$ and $F_y(y)$ respectively, then $U = F_x(x) \sim U(0, 1)$ and $V = F_y(y) \sim U(0, 1)$ and the distribution function of (U, V) is a copula.

$$C(u,v) = P(U \leq u, V \leq v) = P(X \leq F^{-1}_X(u), Y \leq F^{-1}_Y(v))$$

$$C(u,v) = H(F^{-1}_X(u), F^{-1}_Y(v)) \quad \equiv \quad H(x,y) = C(F_X(x), F_Y(y))$$

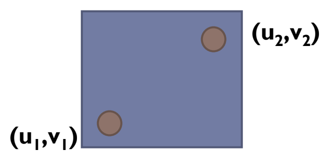
Formal Definition

- $C : [0,1]^2 \rightarrow [0,1]$

1. $C(u,0) = C(0, v) = 0$
2. $C(u,1) = u$, $C(1,v) = v$
3. *C is 2-increasing*

$v_1, v_2, u_1, u_2 \in [0,1]$; $u_2 \geq u_1$, $v_2 \geq v_1$

$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$



Definition (2)

Function

$V_c([u_1,u_2] \times [v_1,v_2]) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$

Is called the C-volume of the rectangle $[u_1,u_2] \times [v_1,v_2]$

So, copula is the restriction to the unit square $[0,1]^2$ of a bivariate distribution function whose margins are uniform on [0,1].

Geometrical Property

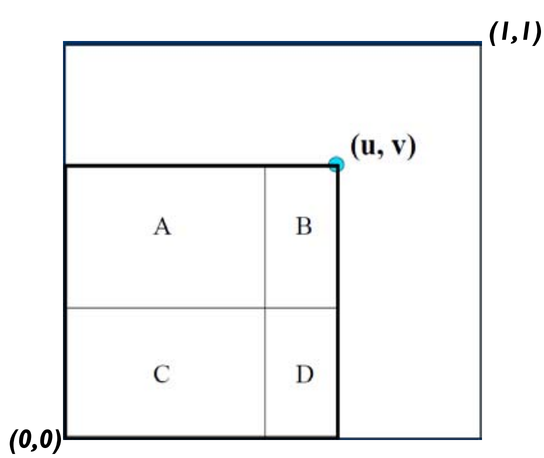
Copula is the C-Volume of rectangle [0,u]*[0,v]

$$C(u,v) = V_c([0,u]\times[0,v])$$

Copula assigns a number to each rectangle in [0,1]*[0,1], which is nonnegative, or

Formally a copula C induces a probability measure on I^2 via $V_c([0,u]\times[0,v]) = C(u,v)$

Geometrical Property



$$C(u,v) = V_c([0,u]\times[0,v]) = V_c(A) + V_c(B) + V_c(C) + V_c(D)$$

Sklar's Theorem(1959)

Let H be a n-dimensional distribution function with margins $F_1,\ldots, F_n$. Then there exists an n-copula *n_copula C* such that for all x □ R^n

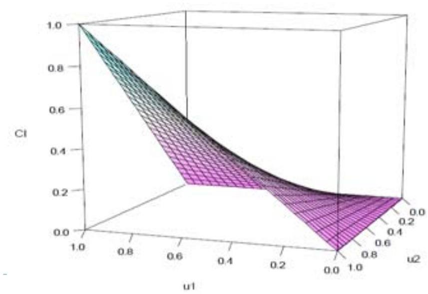
$$H(x_1,\ldots, x_n) = C(F_1(x_1),\ldots, F_n(x_n))$$

C is unique if $F_1,\ldots, F_n$ are all continuous. Conversely, if C is a n-copula and $F_1,\ldots, F_n$ are distribution functions, then H defined above is an n-dimensional distribution function with margins $F_1,\ldots, F_n$

Basic Copulas

Three Basic Bivariate Copulas

- Product Copula
  - $C(u_1, u_2) = u_1 * u_2$
    - It is important as a benchmark because it is correspond to independence



Some Basic Bivariate Copulas

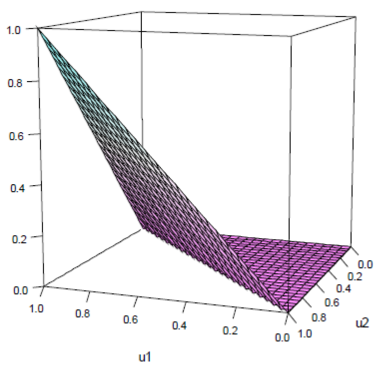
- **Fréchet Lower bound Copula**

  $C_L(u_1,u_2) = max\{0, u_1+u_2-1\}, u \in [0,1]^2$

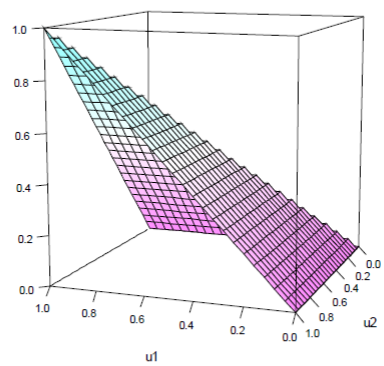
- **Fréchet Upper bound Copula**

  $C_U(u_1,u_2) = min\{u_1, u_2\}, u \in [0,1]^2$

Some Basic Bivariate Copulas



Fréchet Lower bound Copula

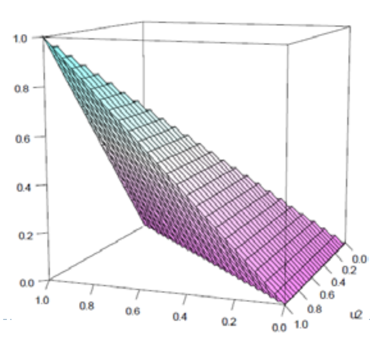


Fréchet Upper bound Copula

Copula Property

- **Any copula will be bounded by Fréchet lower and upper bound copulas**

  $C_L(u_1,u_2) \le C(u_1,u_2) \le C_U(u_1,u_2) \; \forall u \in [0,1]^2$



Additional Properties

- X_1 and X_2 are **independent** iff C is a product copula $C(F_1(x_1), F_2(x_2)) = F_1(x_1)F_2(x_2)$
- X_1 is an increasing function of X_2 iff $C(.) = C_u(.)$
  - Correspond to comonotonicity and perfect positive dependence
- X_1 is a decreasing function of X_2 iff $C(.) = C_L(.)$
  - Correspond to countermonotonicity and perfect negative dependence

Additional Properties

- Perfect positive and negative dependence is defined in terms of **comonotonicity** and **countermonotonicity**
- For any ordered pairs (Y_{1j}, Y_{2j}) and (Y_{1k}, Y_{2k})
  - **Comonotonicity**
    - $\{Y_{1j}<Y_{2j}, Y_{1k}<Y_{2k})\}$ or $\{Y_{1j}>Y_{2j}, Y_{1k}>Y_{2k})\}$
  - **Countermonotonicity**
    - $\{Y_{1j}<Y_{2j}, Y_{1k}>Y_{2k})\}$ or $\{Y_{1j}>Y_{2j}, Y_{1k}<Y_{2k})\}$

Additional Properties

- **Invariance property** : the dependency captured by copula is invariant w.r.t increasing and continuous transformation of marginal distributions.
  - Same copula for (X_1, X_2) as $(\ln X_1 , \ln X_2)$
- If $(U_1,U_2) \sim C$ then there are associated copulas:
  - $(1-U_1, 1-U_2)$, $(U_1, 1-U_2)$, $(1-U_1, U_2)$

Additional Properties

- Survival copula

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) = \Pr[U_1 > u_1, U_1 > u_2]$$

Advantages

- Non-linear dependence
- Be able to measure dependence for heavy tail distributions
- Very flexible: parametric, semi-parametric or non-parametric
- Be able to study asymptotic properties of dependence structures
- Computation is faster and stable with the two-stage estimation

Who cares about dependence?

- **Unsupervised learning**

  –Which observations are dependent / independent
- **Supervised learning**

  –Is there dependence between inputs and outputs?
- **Analysis of stock markets, physical, biological, chemical systems**

  −Dependence between observations?
- **Other applications**

  - Feature selection, Boosting, Clustering

  Information theory, active learning,

  Prediction of protein structure, Drug design, fMRI data processing, Microarray data processing, ICA …

Dependence Measure

(1) $\delta(\mathrm{X},\mathrm{Y}) = \delta(Y,X)$ (symmetry)
(2) -1 $\leq \delta(\mathrm{X},\mathrm{Y}) \leq +1$ $(Normalization)$
(3) $\delta(\mathrm{X},\mathrm{Y})$ = 1 $\leftrightarrow (X,Y)$ $Comonotonic$; $\delta(\mathrm{X},\mathrm{Y})$=-1 $\leftrightarrow (X,Y)$ $countermonotonic$
(4) For any strictly monotonic transformation T: $R \to R$ of **X** :

$$\delta(\mathrm{T}(\mathrm{X}),\mathrm{Y}) = \begin{cases} \delta(\mathrm{X},\mathrm{Y}) & T\ increasing \\ -\delta(\mathrm{X},\mathrm{Y}) & T\ decreasing \end{cases}$$

Correlation

- Correlation coefficient between a pair of variables defined as:

$$\rho XY = \frac{cov[X,Y]}{\partial_X \partial_Y}$$

- it is symetric
- A measure of linear dependence
- Correlation (matrix) gives the all structure of copula in the case of elliptical family distributions (e.g. normal distribution).
- Invariant w.r.t linear transformations

Correlation (2)

- Zero Correlation does not imply independence
- *Cov[X,Y] = 0* $X \sim N(0,1), and\ Y = X^2$
- It is not defined for heavy tail distributions whose second moment do not exit (e.g Student-t with d=1 or 2)
- It is not invariant under strictly increasing nonlinear transformations.

Rank Correlations

- Spearman's rho

$$\rho s(X,Y) = \rho(F_1(X), F_2(Y))$$

- Empirical:

$$\rho_n = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}}$$

$where\ \rho_n \in [-1,1]\ and$

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{n+1}{2} = \frac{1}{n}\sum_{i=1}^{n} S_i = \bar{S}$$

Rank Correlation

- Kendall's tau

$$\rho_\tau = Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

$$\rho_\tau(x, Y) = \Pr[concordance] - \Pr[discordance]$$

- Two pairs (X_i,Y_i) ; (X_j,Y_j) are said to be concordant when $(X_i-X_j)(Y_i - Y_j) > 0$, and discordant when $(X_i-X_j)(Y_i - Y_j) < 0$
- Empirical:
  P_n : # of concordant , Q_n: # of discordant

$$\tau_n = \frac{P_n - Q_n}{C_2^n}$$

Properties

- symmetric
- normalized
- Co- and counter-monotonicity
- Zero when X and Y are independence

$$\rho s(X, Y) = \rho_\tau(X, Y) = -1 \quad \text{iff } C = C_L$$

$$\rho s(X, Y) = \rho_\tau(X, Y) = 1 \quad \text{iff } C = C_U$$

- $C = C_u$ iff $Y=T(X)$ with T increasing
- $C = C_L$ iff $Y=T(X)$ with T decreasing

Properties(2)

- Both $\rho_s(X,Y)$ and $\rho_T(X,Y)$ can be expressed in terms of copulas

$$\rho s(X,Y) = 12\int_0^1\int_0^1 \{C(u_1,u_2) - u_1u_2\}\,du_1du_2$$

$$\rho_\tau(X,Y) = 4\int_0^1\int_0^1 C(u_1,u_2)dC(u_1,u_2) - 1$$

- Are not simple function of moments hence computationally more involved!

Tail dependence

$$\lambda_L = \lim_{u\to 0^+}\frac{C(u,u)}{u}$$

$$\lambda_U = \lim_{u\to 1^-}\frac{S(u,u)}{1-u}$$

- The expression $S(v,v) = \Pr[U_1 > v, U_2 > v]$ represents the joint survival function where $U_1 = F_1^{-1}(X)$, $U_2 = F_2^{-1}(Y)$

Tail dependence

- The dependence measure λ_u is the limiting value of S(v,v)/(1-v) which is the conditional probability
  $Pr[U_1>v|U_2>v]$ $(= Pr[U_2>v|U_1>v])$

- The dependence measure λ_L is the limiting value of conditional probability C(v,v)/v which is the conditional probability
  $Pr[U_2<v|U_1\ v]$ $(= Pr[U_2>v|U_1>v])$

Tail Dependence

- LTD (Left Tail Decreasing)
  - Y is said to be LTD in x, *if* $\Pr[Y \le y / X \le x]$ is decreasing in x for all y
- RTI (Right Tail Increasing)
  - Y is said to be RTI in X if $\Pr[Y > y / X > x]$ is increasing in x for all y
- For copulas with simple analytical expressions, the computation of λ_u can be straight-forward. E.g. for the Gumbel copula λ_u equals $2 - 2^{\theta}$

Tail Dependence (Embrechts et al., 2002)

- the bivariate Gaussian copula has the property of asymptotic independence. Regardless of how high a correlation we choose, if we go far enough into the tail, extreme events appear to occur independently in each margin.
- In contrast, the bivariate *t*-distribution displays asymptotic upper tail dependence even for negative and zero correlations, with dependence rising as the degrees-of-freedom parameter decreases and the marginal distributions become heavy-tailed

Positive Quadrant Dependence

- Two random variables X, Y are said to exhibit PQD if their copula is greater than their product, i.e.,
  $C(u_1, u_2) > u_1 u_2$ or $C > C^{\perp}$
- PQD implies $F(x,y) \geq F_1(x) F_2(y)$ for all (x,y) in R^2
- PQD implies nonnegative correlation and nonnegative rank correlation
- LTD and RTI properties imply the property of PQD