Introduction to Copula Functions
part 2

Mahdi Pakdaman
Intelligent System Program

Outline

- Previously on Copula
- Constructing copulas
- Copula Estimation

2

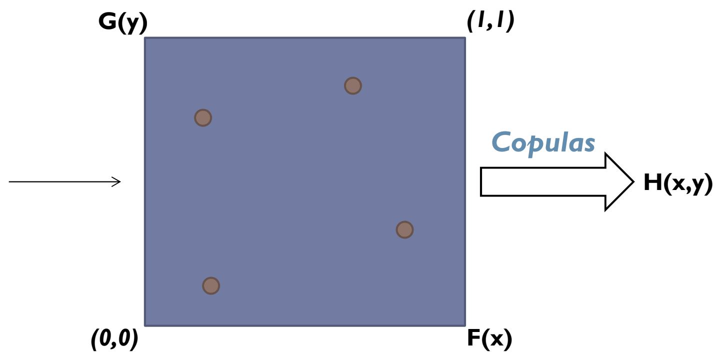
Copula : Definition

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]

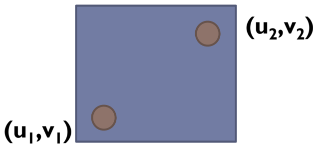
Formal Definition

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

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

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

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



5

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 $[u_1,u_2]\times[v_1,v_2]$

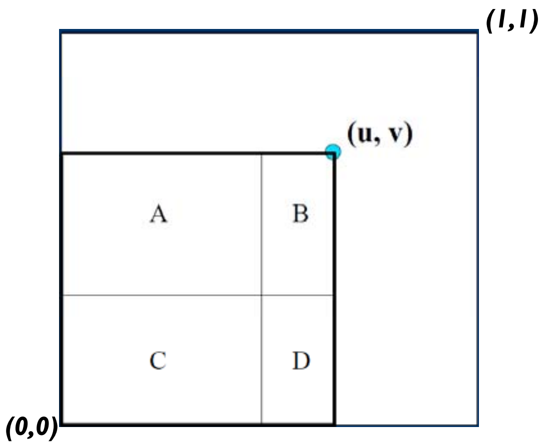
- Copula is the C-Volume of rectangle [0,u]*[0,v]

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

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

6

Geometrical Property



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

7

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 $x \in R^n$

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

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 $F_1,\ldots,F_n$

8

Some Basic Bivariate Copulas

- **Fréchet Lower bound Copula**

  $$C_L(u_1,\ldots,u_n) = max\{0, u_1+\ldots+u_n - \text{n} + 1\}, \boldsymbol{u_i} \in [0,1]^2$$

- **Fréchet Upper bound Copula**

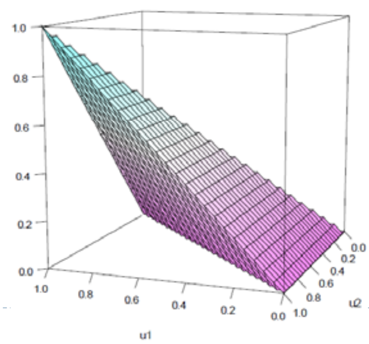
  $$C_U(u_1,\ldots, u_n) = min\{u_1, \ldots, u_n\}, \boldsymbol{u_i} \in [0,1]^2$$

9

Copula Property

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

  $$C_L(u_1,u_2) \leq C(u_1,u_2) \leq C_U(u_1,u_2)\ \forall \boldsymbol{u} \in [0,1]^2$$



10

Additional Properties

- Survival copula

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

Dependence Measure

- Pearson's Correlation coefficient: $\rho XY = \frac{cov[X,Y]}{\partial_X \partial_Y}$
- Ranking Correlation:
  - Spearman's rho $\rho s(X,Y) = \rho(F_1(X), F_2(Y))$
  - Kendall's tau

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

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

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!

13

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 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**
  $Pr[U_2<v|U_1\ v]$ (= $Pr[U_2>v|U_1>v]$)

14

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 $2 - 2^{\theta}$

15

Positive Quadrant Dependence

- Two random variables X,Y are said to exhibit PQD if their copula is greater than their product, i.e.,

  $$C(u_1,u_2) > u_1 u_2 \text{ or } C > C^{\perp}$$

- PQD implies $F(x,y) \ge F_1(x)F_2(y)$ for all (x,y) in $\mathbb{R}^2$
- PQD implies nonnegative correlation and nonnegative rank correlation
- LTD and RTI properties imply the property of PQD

16

Some Popular Copulas

Farlie-Gumbel-Morgentern (FGM)

$$C(u_1, u_2; \theta) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2))$$

- Proposed in 1956
- It is the perturbation of product copula
- Prieger(2002) used it for modeling selection into health insurance plan.
- Restrictive since it is useful when the dependence between two marginal is modest in magnitude

Gaussian (Normal) copula

- $C(u_1, u_2; \theta) = \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta),$

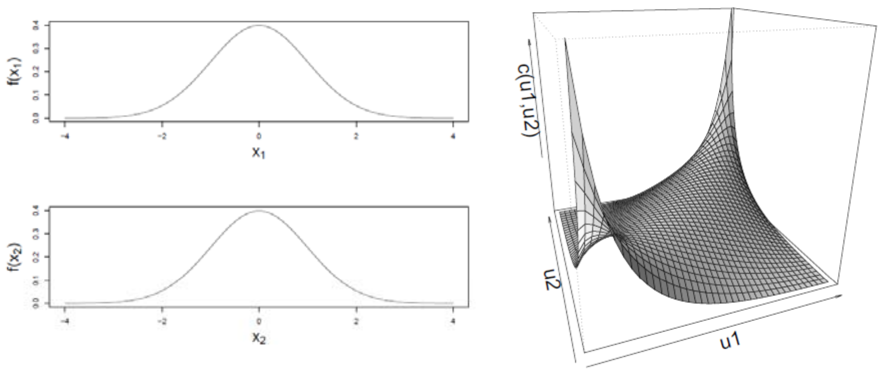
$$= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}}$$

$$\times \left\{ \frac{-s^2 - 2\theta st + t^2}{2(1-\theta^2)} \right\} \text{dsdt}$$

- where Φ is the cdf of the standard normal distribution, and $\Phi_G(u_1,u_2)$ is the standard bivariate normal distribution with correlation parameter θ restricted to the interval $(-1, 1)$.
- Proposed by Lee(1983)
- Flexibility: The normal copula allows for equal degrees of positive and negative dependence and as dependence parameter approaches -1 and 1, it attains the Frechet lower and upper bound.

19

Gaussian Copula



20

Student's t-copula

$$C^t(u_1,u_2;\theta_1,\theta_2) = \int_{-\infty}^{t_{\theta_1}^{-1}(u_1)} \int_{-\infty}^{t_{\theta_2}^{-1}(u_2)} \frac{1}{2\pi(1-\theta_2^2)^{1/2}} \times \left\{1+\frac{s^2-2\theta_2 st+t^2}{v(1-\theta_2^2)}\right\}^{-(\theta_1+2)/2} dsdt$$

- $t_{\theta_1}^{-1}(u_1)$ is the inverse of the cdf of standard univariate t-distribution with θ_1 degree of freedom
- θ_1 controls the heaviness of the tails.
- As $\theta_1 \to \infty$ it will behave like Gaussian copula

21

Clayton Copula

$$C(u_1,u_2;\theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

- Dependence parameter θ restricted $(0,\infty)$
- when θ approaches zeros the marginal become independent . As θ approaches infinity the copula attain Frechet upper bound but for no value attain its lower bound
- When correlation between two events strongest in the left tail of joint distribution it is appropriate for modeling (e.g. performance of two funds or spouses' ages at death)

22

Frank Copula

$$C(u_1, u_2; \theta) = -\theta^{-1} \log \{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\}$$

- **Pros:**
  - Values to -∞ , 0, ∞ correspond to
    Frechet lower bound, independence, upper bound
  - Permits negative dependence between the marginal
  - Dependence is symmetric in both tails
- **Cons:**
  - Although covers all range of dependency. dependence in the tails of Frank copulas are weak w.r.t. Gaussian Copula. (Meester and MacKay 1994)

Gumbel copula

$$C(u_1, u_2; \theta) = \exp(-(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta})^{1/\theta})$$

$$\tilde{u}_j = -\log u_j$$

- θ Is in the range [1 , ∞) correspond to independence and Frechet upper bound.
- Proper for modeling the outcomes that have strong correlation at high values and have weak correlation at low values.

Generating Copulas

- Method of inversion
- Algebraic Methods
- Mixtures and Convex sum
  - Mixture of Powers
- Archimedean Copulas
- Geometric

Method of Inversion

- Considering $F(y_1, y_2) = C(F_1(y_1), F_2(y_2))$

Using inverse transformation $U_1 = F_1^{-1}(y_1)$, $U_2 = F_2^{-1}(y_2)$ we will have

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))$$

- Also the survival copula is given by:

$$\bar{C}(u_1, u_2) = \bar{F}(\overline{F_1}^{-1}(u_1), \overline{F_2}^{-1}(u_2))$$

Where $\overline{F_1}$, $\overline{F_2}$ are marginal survival function.

Method of Inversion: Example

- $F(y_1,y_2) = \exp\{-[e^{-y_1} + e^{-y_2} - (e^{-\theta y_1} + e^{-\theta y_1})^{-1/\theta}]\}$

$$-\infty < y_1, y_2 < \infty, \theta \geq 0$$

$$\lim_{y_2 \to \infty} F(y_1, y_2) = F_1(y_1) = \exp(e^{-y_1}) \equiv u_1$$

$$\lim_{y_1 \to \infty} F(y_1, y_2) = F_2(y_2) = \exp(e^{-y_2}) \equiv u_2$$

$$y_1 = -\log(-\log(u_1)) \; and \; y_2 = -\log(-\log(u_2))$$

$$C(u_1,u_2) = u_1 u_2 \exp\{[(-\log(u_1))^{-\theta} + (-\log(u_2))^{-\theta}]^{-1/\theta}\}$$

This expression can be written as

$$C(u_1,u_2) = u_1 u_2 \varphi^{-1}\{[(-\varphi(u_1))^{-\theta} + (-\varphi(u_2))^{-\theta}]^{-1/\theta}\}$$

Which will be seen to be a member of Archimedean class

Other Copulas : Using Inversion Method

Case	Joint distribution: $F(y_1, y_2)$	Margins: $F(y_1)$, $F(y_2)$	Copula: $C(u_1, u_2)$
1	$1 - (e^{-\theta y_1} + e^{-2\theta y_2} - e^{-\theta(y_1+2y_2)})^{1/\theta}$	$F(y_1) = 1 - e^{-y_1}$	$1 - \{(1-(1-u_2)^{\theta})(1-u_1)^{\theta}$
	$\theta \geq 0$	$F(y_2) = 1 - e^{-2y_2}$	$+(1-u_2)^{\theta}\}^{1/\theta}$
2.	$\exp\{-(e^{-\theta y_1} + e^{-\theta y_2})^{1/\theta}\}$	$F(y_1) = \exp(-e^{-y_1})$;	$\exp\{-(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\}^{1/\theta}$
	$-\infty < y_1, y_2 < \infty, \theta \geq 1$	$F(y_2) = \exp(-e^{-y_2})$	
3.	$(1 + e^{-y_1} + e^{-y_2})^{-1}$	$F(y_1) = (1 + e^{-y_1})^{-1}$; $F(y_2) = (1 + e^{-y_2})^{-1}$	$u_1 u_2/(u_2 + u_1 - u_1 u_2)$

Algebraic Method

- Some derivations of copulas begin with a relationship between marginals based on independence. Then this relationship is modified by introducing a dependence parameter and the corresponding copula is obtained.
- Example 3 in the previous Table is Gumbel's bivariate logistic distribution denoted $F(y_1, y_2)$

29

Algebraic Method: Example

- Let $(1 - F(y_1,y_2)) / F(y_1,y_2)$ denote the bivariate survival odds ratio

$$\frac{1 - F(y_1, y_2)}{F(y_1, y_2)} = e^{-y_1} + e^{-y_2}$$
$$= \frac{1 - F_1(y_1)}{F_1(y_1)} + \frac{1 - F_2(y_2)}{F_2(y_2)}$$

- Observe that in this case there is no explicit dependence parameter

30

Algebraic method: Example(Cntd)

- In the case of independence, since $F(y_1,y_2) = F_1(y_1)F_2(y_2)$, so we can write the ratio as:

$$\frac{1-F(y_1,y_2)}{F(y_1,y_2)} = \frac{1-F_1(y_1)F_2(y_2)}{F_1(y_1)F_2(y_2)}$$

$$= \frac{1-F_1(y_1)}{F_1(y_1)} + \frac{1-F_2(y_2)}{F_2(y_2)} + \frac{1}{}$$

- Noting the similarity between the bivariate odds ratio in the dependence and independence cases, Ali, Mikhail, and Haq proposed a generalized bivariate ratio with a dependence parameter θ

Algebraic method: Example(Cntd)

$$\frac{1-F(y_1,y_2)}{F(y_1,y_2)} =$$

$$= \frac{1-F_1(y_1)}{F_1(y_1)} + \frac{1-F_2(y_2)}{F_2(y_2)} + (\text{:}$$

$$\frac{1-C(u_1,u_2;\theta)}{C(u_1,u_2;\theta)} = \frac{1-u_1}{u_1} + \frac{1-u_2}{u_2} + (1-\theta)\frac{1-u_1}{u_1}\frac{1-u_2}{u_2}$$

whence

$$C(u_1,u_2;\theta) = \frac{u_1 u_2}{1-\theta(1-u_1)(1-u_2)}$$

Convex Sum

- We can obtain new copulas using a convex combination of copula. E.g.
  - The class of Frechet Copulas, denoted by C^F defined as:

$$C^F = \pi_1 C_L + (1 - \pi_1 - \pi_2) C^{\perp} + \pi_2 C_U$$

- We can generalize this idea by averaging over **infinite collection** of copulas indexed by a continuous variable η with a distribution function $\Lambda_\theta(\eta)$ with parameter θ.so the copula obtained:

$$C_\theta(u_1, u_2) = E_\eta[C_\eta(u_1, u_2)] = \int_{R(\eta)} C_\eta(u_1, u_2) d\Lambda_\theta(\eta)$$

33

Mixture of Powers

- Marshall and Olkin(1998) consider the mixture as:

$$H(y) = \int [F(y)]^\eta d\Lambda(\eta), \quad \eta > 0 \qquad (Eq_1)$$

- And they showed that for every specified pair {H(y), Λ(η)}, $\bar{\Lambda}(0) = 1$ there exist F(y).
- A well known example from Marshall and Olkin(1998) shows how convex sum or mixtures lead to copulas constructed from Laplace Transform of distribution functions.

34

Mixture of Powers(2)

- Lets $\varphi(t)$ defines the Laplace transform of a positive random variable η distributed as Λ

$$\varphi(t) = \int_0^\infty e^{-\eta t} d\Lambda(\eta)$$

- Now the rhs of equation Eq_1 Can be written as φ[-ln F(y)] and so F(y) = exp[-φ^{-1}H(y)]
- An inverse Laplace transform could be a copula generator. How?
- Let $F_1(y_1)$ = exp[-$\varphi^{-1}H_1(y_1)$] and $F_2(y_2)$ = exp[-$\varphi^{-1}H_2(y_2)$] be some benchmark distribution function for y_1 and y_2
- generally y_1 and y_2 are not independent. What we are doing is to introduce unobserved heterogeneity term (latent Random Variable) η with the distribution $\Lambda_\theta(\eta)$ and now assume that y_1 and y_2 are independent condition on η.

35

Mixture of Powers(2)

- Let the conditional distribution given the (laten) random variable η, η>0, be $F_1(y_1| \eta) = [F_1(y_1)]^\eta$ and $F_2(y_2| \eta) = [F_2(y_2)]^\eta$ then

$$H(y_1, y_2; \theta) = \int_0^\infty [F_1(y_1)]^\eta [F_2(y_2)]^\eta d\Lambda_\theta(\eta),$$
$$= \int_0^\infty \exp[-\eta[\varphi^{-1}(H_1(y_1)) + \varphi^{-1}(H_2(y_2))]] d\Lambda_\theta(\eta),$$
$$= \varphi[\varphi^{-1}(H_1(y_1)) + \varphi^{-1}(H_2(y_2)); \theta]$$

- Is shown to be joint distribution of y_1 and y_2 . And it is also a (Archimedean) copula.
- $H_1(y_1)$ and $H_2(y_2)$ are the marginal distributions of $H(y_1, y_2; \theta)$

36

Convex Sum: Example

- Dependence Between Stock indexes
- Hu(2004) studies the dependence of monthly return between four stock indexes.
- She uses monthly averages from January 1970 to September 2003.
- She Modeled dependence on a pair-wise basis using a mixture of three copulas {Gaussian(C_G), Gumbel(C_{gumbel}) and Gumbel-Survival(C_{GS})}

$$C_{mix}(u,v;\rho,\alpha,\theta) = \pi_1 C_{Gauss}(u,v;\rho) + \pi_2 C_{Gumbel}(u,v;\alpha) + (1-\pi_1-\pi_2) C_{GS}(u,v;\theta)$$

37

Convex Sum : Example(2)

- Such a mixture imparts additional flexibility and also allows one to capture left and/or right tail dependence
- Hu uses a two-step semi-parametric approach to estimate the model parameters.
- Note that pairwise modeling of dependence can be potentially misleading if dependence is more appropriately captured by a higher dimensional model

38

Archimedean Copulas

Archimedean Copulas

- Consider the function φ:[0,1]→ [0, ∞] with the following properties: it has continuous derivative, it is decreasing and convex
- Such functions are called generator functions. E.g. φ(t)=-ln(t) , $\varphi(t)= t^{-\theta}$, $\theta > 1$

$$C(u_1,u_2;\theta) = \phi^{-1}(\phi(u_1) + \phi(u_2))$$

Archimedean Copulas(cntd)

- If $\varphi(0) = \infty$ the generator is called strict and inverse φ^{-1} exist. In this case we have $C(u_1,u_2) > 0$ except u_1=0 or u_2=0.
- If $\varphi(0) < \infty$ it is not strict and its pseudo-inverse $\varphi^{[-1]}$ exist.

$$\varphi^{[-1]}(t) = f(x) = \begin{cases} \varphi^{-1}(t) & 0 \le t \le \varphi(0) \\ 0 & \varphi(t) \le t \le +\infty \end{cases}$$

In this case the copula has singular component and takes the form $C(u_1,u_2)$= max[(.),0]

- E.g.: $\varphi(t) = (1-t)^{\theta}, \theta \in [1,\infty)$
  - $C(u_1,u_2) = \max[1 - [(1-u_1)^{\theta} + (1-u_1)^{\theta}]^{\frac{1}{\theta}}, 0]$

Archimedean Copulas: Example

- **Let $\varphi(t) = 1-t$, $t \in [0,1]$**

  $\varphi^{-1}(t) = max(1-t,0)$

  $C(u,v) = max(u+v-1,0)$

- Let $\phi(t) = -\ln(t)$, $t \in [0,1]$ then $\phi(0) = \infty$,

  $\phi^{[-1]}(t) = \exp(-t)$.

  $C(u1,u2) = uv$, the product copula

Archimedean Copulas: Properties

- **Archimedean Copula behaves like a binary operation**
- $C(u,v) = C(v,u)$, $\forall u, v \in [0,1]$
- $C(C(u,v),w) = C(u,C(v,w))$, $\forall u, v, w \in [0,1]$
- **Order preserving:**

  $C(u1, v1) \leq C(u2, v2)$, $u1 \leq u2$, $v1 \leq v2$, $\in [0,1]$

Archimedean Copulas: Properties

- The properties of the generator affect tail dependency of the Archimedean copula. If $\phi'(0) < \infty$ and $\phi'(0) \neq 0$, then $C(u1,u2)$ does not have the RTD property. If $C(\cdot)$ has the RTD property then $1/\phi'(0) = -\infty$
- Quantifying dependence is relatively straightforward for Archimedean copulas because Kendall's tau simplifies to a function of the generator function,

$$\Gamma = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt$$

Archimedean Copulas: extended by transformation

- Let ϕ be a generator then:
  - $g : [0, 1] \rightarrow [0, 1]$ be a strictly increasing concave function with $g(1) = 1$. Then $\phi \circ g$ is a generator.
  - $f : [0,\infty] \rightarrow [0,\infty]$ be a strictly increasing convex function with $f(0) = 0$, then $f \circ \phi$ is a generator.

Multivariate Archimedean Copulas

- Let ϕ be a continuous, strictly decreasing function from [0 1] to $[0, \infty)$ such that $\phi(0)= \infty$ and $\phi(1) = 0$ and let ϕ^{-1}, be the inverse of ϕ. Then
  - $C^n(u) = \phi^{-1} (\phi(u_1) + \phi(u_2)+\ldots+ \phi(u_n))$

  Is a n-copula iff ϕ^{-1} is completely monotonic on $[0, \infty)$

$$(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0 \text{ for all } t \in \text{int}([0,\infty)) \text{and } k = 0,1,2,\ldots$$

Multivariate Archimedean Copulas

- Clayton Family

$$C_\theta^n(u) = (u_1^{-\theta} + u_2^{-\theta} + \cdots + u_n^{-\theta} - n + 1)^{-1/\theta} \quad \varphi_\theta(t) = t^{-\theta} - 1 \;\; for \;\; \theta > 0$$

- Frank Family

$$C_\theta^n(u) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)\ldots(e^{-\theta u_n} - 1)}{(e^{-\theta} - 1)^{n-1}}\right)$$

$$\varphi_\theta(t) = -\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right) \quad for\; \theta > 0$$

47

Multivariate Archimedean Copulas

- Gumbel-Hougaard Family

$$C_\theta^n(u) = \exp\left(-[(-\ln u_1)^\theta + (-\ln u_1)^\theta + \cdots + (-\ln u_1)^\theta]^{\frac{1}{\theta}}\right)$$

$$\varphi_\theta(t) = (-\ln t)^\theta, \theta \geq 1$$

- A 2-parameter Multivariate Copula

$$C_{\alpha,\beta}^n(u) = \left\{\left[(u_1^{-\alpha} - 1)^\beta + (u_2^{-\alpha} - 1)^\beta + \cdots + (u_n^{-\alpha} - 1)^\beta\right]^{1/\beta} + 1\right\}^{-1/\alpha}$$

$$\varphi_{\alpha,\beta}(t) = (t^{-\alpha} - 1)^\beta \quad for\; \alpha > 0, \beta \geq 1$$

48

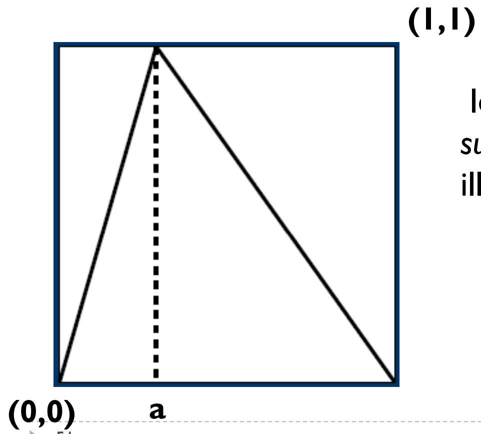
Archimedean Copulas: summery

- Archimedean Copulas have a wide range of applications for some reasons:
  - Easy to be constructed
  - Easy to extend to high dimension
  - Capable of capturing wide range of dependence
  - Many families of copulas belong to it

Generating Copula: Geometric method

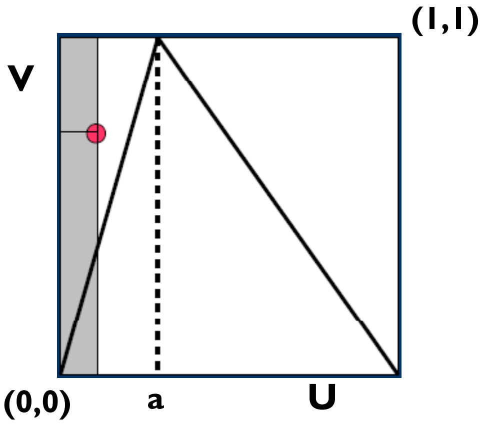
Without reference to distribution functions or random variables, we can obtain the copula via the C-Volume of rectangles in [0, 1]*[0, 1]

Geometric method



let C_a *denote the copula* with *support as the line* segments illustrated in the graph.

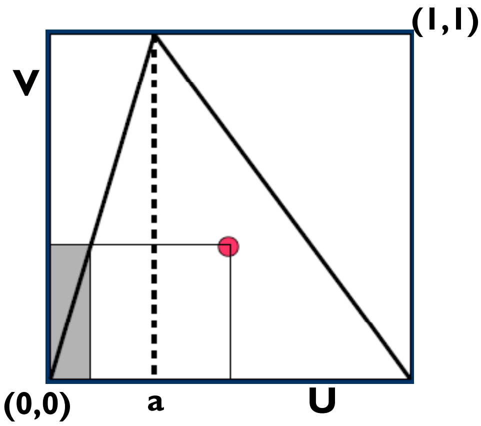
Geometric Method



When $u \le av$

$C_a(u,v) = V_{ca}([0,u] \times [0,1]) = u$

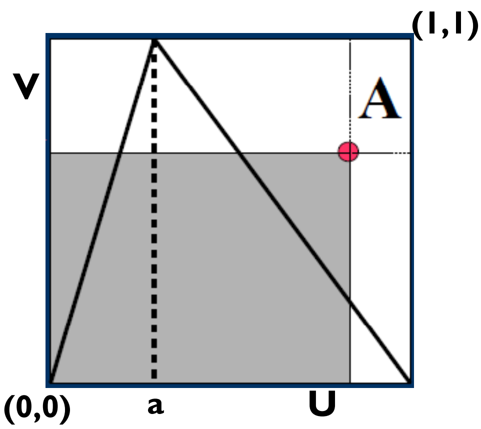
Geometric Method



When
$1-(1-a)v > u > av$

$C_a(u,v)= C_a(av,v)= av$

Geometric Method



When $u > 1-(1-a)v$
$V_{ca}(A) = 0 \quad \rightarrow$

$C_a(u,v)= u + v - 1$

Geometric Method

$C_a(u,v) = u \qquad 0 \le u \le av \le a$

$C_a(u,v) = av \qquad 0 \le av < u < 1-(1-a)v$

$C_a(u,v) = u + v - 1 \qquad a \le 1-(1-a)v \le u \le 1$

- **C_1 :Fréchet Upper bound Copula**
- **C_0: Fréchet Lower bound Copula**

Copula Estimation

Copula Estimation

- Full Maximum Likelihood Estimate(FML)
- 2-Step Maximum Like likelihood (TSML)

Copula likelihood function

- Let we want to estimate the parameters of copula model in which we have parametric marginal as well as parametric copula
- Let marginal density function $f_j(y_j|x_j;\beta_j) = \frac{\partial F_j(y_j|x_j;\beta_j)}{\partial y_j}$
- Let Copula density:
- $c(F_1(.), F_2(.))= \frac{d}{dy_1 dy_2} C(F_1(.), F_2(.)) = C_{12}(F_1(.), F_2(.))f_1(.)f_2(.)$
- $C_{12}((F_1|x_1;\beta_1), (F_2|x_2;\beta_2));\ \theta) = \frac{\partial C((F_1|x_1;\beta_1), (F_2|x_2;\beta_2)); \theta)}{\partial F_1\, \partial F_2}$

Copula Likelihood Function

- $L_N((y_1|x_1;\beta_1),\ (y_2|x_2;\beta_2));\ \theta) =$

$$\sum_{i=1}^{N}\sum_{j=1}^{2} \ln f_{ji}(y_{ji}|x_{ji};\beta_j) + \sum_{i=1}^{N} C_{12}(F_1(y_{1i}|x_{1i};\beta_1), F_2(y_{2i}\ |x_{2i};\beta_2);\ \theta)$$

$$L_N(\beta_1,\beta_2,\theta) = L_{1,N}(\beta_1,\beta_2) + L_{2,N}(\beta_1,\beta_2,\theta)$$

Generate Archimedean Copula

- Let $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ random sample of bivariate observations
- Assume that the distribution function has an Archimedean copula C_φ
- *Consider an intermediate pseduo-observation Z_i with the distribution function $K(z) = P[Z_i \leq z]$*
- *Genest and Rivies(1993) showed that K is related to φ through*

$$K(z) = z - \frac{\phi(z)}{\acute{\phi}(z)}$$

Generating Archimedean Copula: Algorithm

1. **Estimate Kendall's correlation coefficient using the usual estimate**

$$\tau_n = \binom{n}{2}^{-1} \sum_{i<j} Sign[(X_{1i} - X_{1j})(X_{2i} - X_{2j})]$$

2. **Construct a nonparametric estimate of K as follows:**
   a) define the pseudo-observations

$$Z_i = \frac{\{number\ of\ (X_{1j}, X_{2j})\ such\ that\ X_{1i} > X_{1j}\ and\ X_{2i} > X_{2j}\}}{n-1}$$

   a) construct the estimate K_n of K as $K_n(z)$ = proportion of $Z_i \leq z$

Generating Archimedean Copula: Algorithm

3) Since K has to satisfy the relation

$$K(z) = z - \frac{\phi(z)}{\acute{\phi}(z)}$$

we obtain an estimate of ϕ_n of ϕ, by solving the equation

$$z - \frac{\phi_n(z)}{\acute{\phi}_n(z)} = K_n(z)$$

Drawbacks of using the copula

- Few parametric copula can be generalized beyond the bivariate case
- The same is true for copula model selection where most goodness-of-fit tests are devised for a bivariate copula and cannot be extended to higher dimensionality
- intuitive interpretation of copula-parameter(s) is not always available

63

Copula in Machine Learning

- The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs, **H. Liu, J. Lafferty, L. Wasserman**(JMLR 2009)
- Kernel-based Copula Processes, S. Jaimungal, E. K. H. Ng, *Machine Learning and Knowledge Discovery in Databases (2009)*
- Copula Bayesian Networks, G. Elidan (NIPS 2010)
- Copula Process, A. G. Wilson, Z. Ghahramani (NIPS 2010)

64

NIPS 2011 Workshop

- Copula in Machine Learning
- Abstract submission deadline, October 21st, 2011

- **Organizers**
  - Gal Elidan, The Hebrew University of Jerusalem
  - Zoubin Ghahramani Cambridge University and Carnegie Mellon University
  - John Lafferty, University of Chicago and Carnegie Mellon University

Link: http://pluto.huji.ac.il/~galelidan/CopulaWorkshop/index.html

Thank you