Queen Mary
University of London

MTH5126 Statistics for Insurance

Dr Lei Fang

lei.fang@qmul.ac.uk

Week 6

Great credit and thanks to D. Boland, G. Ng and F. Parsa, previous lecturers, for their excellent work in producing the original version of these notes.

Copulas

Marginal and joint distribution

Association, concordance, correlation and tail dependence

Copulas

- Definition
- Sklar's theorem
- Expressions of tail dependence and survival copula
- Types of copula functions

Fundamental copulas

- Independence (or product) copulas
- Co-monotonic (or minimum) copulas
- Counter-monotonic (or maximum) copulas

Explicit copulas

- Archimedean copulas
  - Gumbel copula
  - Clayton copula
  - Frank copula

Implicit copulas

- Gaussian copula
- Student's t copula

Choosing a suitable copula function

Marginal and joint distributions

- For two variables, the joint (cumulative) distribution function (CDF) is:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

- This can be extended from the bivariate case to the multivariate case in d dimensions:

$$F_{X_1,X_2,\ldots,X_d}(x_1,x_2,\ldots,x_d) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d)$$

- In the context of joint distribution functions, the individual distribution of each of the variables in isolation is known as its *marginal distribution*.

Association, concordance, correlation and tail dependence

- Variables are said to be *associated* if there is some form of statistical relationship between them, whether causal or not.
- Note that a positive association between two variables does not necessarily imply that one is dependent on the other. **Example**: Both might be dependent on a third (perhaps unobserved) variable, with neither being directly dependent on the other. So, correlation does not imply causation!
- *Concordance* is a particular form of association. Broadly speaking, two random variables are concordant if small values of one are likely to be associated with small values of the other, and vice versa.

According to an ancient legend, the stork is responsible for bringing babies to new parents. So if you observe a fall in the population of storks and a fall in birth rates, does it imply causation?

Association, concordance, correlation and tail dependence

- Pearson's linear correlation coefficient (also known as Pearson's ρ) measures the degree to which there is a **linear** relationship between two variables:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

- Two commonly used measures of correlation that are more robust than Pearson's ρ are Spearman's ρ (often called the rank correlation) and Kendall's τ.

Association, concordance, correlation and tail dependence

- It is often the case in insurance and investment applications that large losses tend to occur together. For example, a hurricane could result in large losses on several different property insurance policies sold by the same company, or a stock market crash could lead to large losses on a number of investments in the same investment portfolio.
- So, the relationships between the variables at the extremes of the distributions, i.e., in the upper and lower tails, are of particular importance. These can be measured using the coefficients of upper and lower tail dependence.

Association, concordance, correlation and tail dependence

- Coefficient of upper tail dependence:

$$\lambda_U = \lim_{u \to 1^-} P\left(X > F_X^{-1}(u) \middle| Y > F_Y^{-1}(u)\right)$$

- This coefficient is a probability, so it takes a value between 0 and 1.
- The coefficient of upper tail dependence indicates whether high values of one random variable, *X* , tend to be linked with high values of another random variable *Y* .
- It considers the probability of the random variable *X* taking a value in the upper tail of its distribution (e.g., a tail with a probability mass of 5% implies u = 0.95), given that the random variable *Y* takes a value in the same sized upper tail of its distribution.
- Specifically, the coefficient of upper tail dependence is the limiting value of this probability as $u \to 1^-$, i.e., as we move further into the upper tail (from below).

Association, concordance, correlation and tail dependence

- Coefficient of lower tail dependence:

$$\lambda_L = \lim_{u \to 0^+} P\left(X \le F_X^{-1}(u) \mid Y \le F_Y^{-1}(u)\right)$$

- Again, this coefficient is a probability, so it takes a value between 0 and 1.
- The coefficient of lower tail dependence indicates whether low values of one random variable, *X* , tend to be linked with low values of another random variable *Y* .
- It considers the probability of the random variable *X* taking a value in the lower tail of its distribution (e.g., a tail with a probability mass of 5% implies u = 0.05), given that the random variable *Y* takes a value in the same sized lower tail of its distribution.
- Specifically, the coefficient of lower tail dependence is the limiting value of this probability as $u \to 0^+$, i.e., as we move further into the lower tail (from above).

Copulas

Definition

Definition: A copula is a function that expresses a multivariate cumulative distribution function in terms of the individual marginal cumulative distributions.

- It is useful to remember that a copula is a **function**. It takes marginal cumulative distributions of random variables as inputs, and outputs a corresponding joint cumulative distribution function.

Copulas

Definition

Definition of a copula:

For a bivariate distribution, the copula is a function C_{XY} defined by:

$$C_{XY}\left[F_X(x),F_Y(y)\right]=P(X\le x,Y\le y)=F_{X,Y}(x,y)$$

This is often written in the more compact form:

$$C[u,v]=F_{X,Y}(x,y) \quad \text{where } u=F_X(x) \text{ and } v=F_Y(y)$$

This definition can be extended to the multivariate case where we have:

$$C[u_1,u_2,\ldots,u_d]=F_{X_1,X_2,\ldots,X_d}(x_1,x_2,\ldots,x_d) \quad \text{where } u_i=F_{X_i}(x_i)$$

- Note that the arguments u_1, u_2, …, u_d and the output value of the copula function are restricted to the range [0,1], as they correspond to probabilities.

Copulas

Definition: Example

Question:

Explain in words, the meaning of the following copula expression: ***C*[*u*, *v*, *w*]**

Answer:

This gives the probability that random variable 1 is in the bottom *u* percentile, and
random variable 2 is in the bottom *v* percentile, and
random variable 3 is in the bottom *w* percentile.

Copulas

Sklar's theorem

Sklar's theorem:

Let F be a joint (cumulative) distribution function with marginal cumulative distribution functions $F_1, \ldots, F_d$. Then there exists a copula, C, such that for all $x_1, \ldots, x_d \in [-\infty, \infty]$:

$$F(x_1, x_2, \ldots, x_d) = C\,[F_1(x_1), \ldots, F_d(x_d)]$$

- In the case where the variables have a *continuous* distribution, the copula is unique.
- The converse also holds.

Converse of Sklar's theorem:

If C is a copula and $F_1, \ldots, F_d$ are univariate cumulative distribution functions, then the function F defined above is a joint cdf with marginal cdf $F_1, \ldots, F_d$.

Copulas

Expressions of tail dependence and survival copula

Coefficient of lower tail dependence in terms of the copula function:

$$\lambda_L = \lim_{u \to 0^+} P\left(X \le F_X^{-1}(u) \mid Y \le F_Y^{-1}(u)\right) = \lim_{u \to 0^+} \frac{C[u,u]}{u}$$

- i.e., the coefficient of lower tail dependence can be calculated directly from the copula function.
- The coefficient of lower tail dependence can take values between 0 (no dependence) and 1 (full dependence).

Copulas

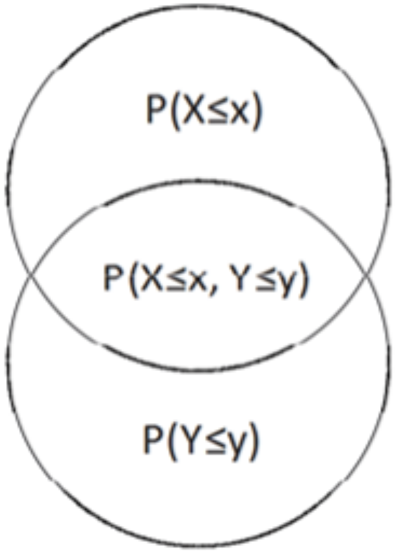
Expressions of tail dependence and survival copula

Survival copula:

To define the upper tail dependence, we need to look at the opposite end of the marginal distributions. Associated with each copula function is a survival copula function (denoted with a bar), which is defined by:

$$\bar{F}(x,y) = P(X > x, Y > y) = \bar{C}\left[\bar{F}_X(x), \bar{F}_Y(y)\right]$$

where $\bar{F}_X(x) = 1 - F_X(x)$ and $\bar{F}_Y(y) = 1 - F_Y(y)$.



By the principle of inclusion/exclusion, we have:

$$P(X \le x \cup Y \le y) = P(X \le x) + P(Y \le y) - P(X \le x, Y \le y)$$

ie $$1 - P(X > x, Y > y) = P(X \le x) + P(Y \le y) - P(X \le x, Y \le y)$$

$\Rightarrow$ $$P(X > x, Y > y) = 1 - P(X \le x) - P(Y \le y) + P(X \le x, Y \le y)$$

So, the survival copula is related to the original copula function by:

$$\bar{C}[1-u, 1-v] = 1 - u - v + C[u,v]$$

Copulas

Expressions of tail dependence and survival copula

Coefficient of upper tail dependence in terms of survival copula function:

$$\lambda_U = \lim_{u\to 1^-} P\left(X > F_X^{-1}(u) \,\middle|\, Y > F_Y^{-1}(u)\right) = \lim_{u\to 0^+} \frac{\bar{C}[u,u]}{u}$$

$$= \lim_{u\to 1^-}\left(\frac{1-2u+C[u,u]}{1-u}\right)$$

Copulas

Types of copula function

3 main families of copula in this module:

- Fundamental copulas
- Explicit copulas
- Implicit copulas

Fundamental Copulas

The 3 copulas below are referred to as the fundamental copulas.

3 fundamental dependencies that a set of variables can display:

- Independence: represented by the independence (or product) copula
- Perfect positive interdependence: represented by co-monotonic (or minimum) copula
- Perfect negative interdependence: represented by counter-monotonic (or maximum) copula

Fundamental Copulas

Independence (or product) copula

The *product copula* is defined in the bivariate case as:

$$C[u, v] = uv$$

Expressing the above in terms of *x* and *y*:

$$F_{X,Y}(x,y) = C[F_X(x), F_Y(y)] = F_X(x)F_Y(y)$$

or: $$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

- This captures the property of the independence of the 2 variables *X* and *Y*. Hence, this copula is also called the *independence copula*.
- The independence copula can be extended to the multivariate case, e.g.:

$$_{ind}C\left[F_{X_1}(x_1),\ldots,F_{X_d}(x_d)\right] = F_{X_1}(x_1)\times\ldots\times F_{X_d}(x_d)$$

Fundamental Copulas

Co-monotonic (or minimum) copula

The *co-monotonic copula* is defined in the bivariate case as:

$$C[u, v] = \min(u, v)$$

Expressing the above in terms of *x* and *y*:

$$C\left[F_X(x), F_Y(y)\right] = \min\left(F_X(x), F_Y(y)\right)$$

or: $$P(X \le x, Y \le y) = \min\left(P(X \le x), P(Y \le y)\right)$$

- The co-monotonic copula captures the relationship between two variables whose values are perfectly positively interdependent on each other.
- The co-monotonic copula can be extended to the multivariate case, e.g.:

$$_{\min}C\left[F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right] = \min\left(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right)$$

Queen Mary University of London

Fundamental Copulas

Counter-monotonic (or maximum) copula

The *counter-monotonic copula* is defined in the bivariate case as:

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

Expressing the above in terms of *x* and *y*:

$$C\left[F_X(x), F_Y(y)\right] = \max\left(F_X(x) + F_Y(y) - 1, 0\right)$$

or: $$P\left(X \leq x, Y \leq y\right) = \max\left(P\left(X \leq x\right) + P\left(Y \leq y\right) - 1, 0\right)$$

- The counter-monotonic copula captures the relationship between two variables whose values are perfectly negatively interdependent on each other.
- The counter-monotonic copula <u>cannot</u> be extended to the multivariate case because it is not possible to have 3 or more variables where each pair has a direct inverse relationship.

Explicit Copulas

Archimedean copulas

- Explicit copulas have simple closed-form expressions.
- Archimedean copulas are a subset of explicit copulas.
- Archimedean copulas take the form:

$$C[u, v] = \psi^{[-1]}(\psi(u) + \psi(v))$$

where $\psi(t)$ is a generator function and $\psi^{[-1]}$ is a pseudo-inverse generator function.

- The pseudo-inverse function $\psi^{[-1]}$ of a function $\psi(t)$ is defined as:

$$\psi^{[-1]}(x) = \begin{cases} \psi^{-1}(x) & \text{if } 0 \leq x \leq \psi(0) \\ 0 & \text{if } \psi(0) < x \leq \infty \end{cases}$$

where $\psi^{(-1)}(x)$ denotes the ordinary inverse function obtained by inverting the equation $x = \psi(y)$ to express y in terms of x.

- If $\psi(0) = \infty$, the pseudo-inverse $\psi^{[-1]}$ is always equal to the 'ordinary' inverse $\psi^{(-1)}$ and the generator function is called a strict generator function.

Explicit Copulas

Archimedean copulas

3 **examples** of Archimedean copulas:

1. The Gumbel copula
2. The Clayton copula
3. The Frank copula

➢ In these examples, α is a parameter whose value can be specified to adjust the strength of dependence between variables.

Explicit Copulas

Gumbel copula

The Gumbel copula is defined in the bivariate case as:

$$C[u, v] = \exp\{ - ((-\ln u)^{\alpha} + (-\ln v)^{\alpha})^{1/\alpha} \} \text{ for } \alpha \geq 1$$

- The generator function is:

$$\psi(t) = (-\ln t)^{\alpha} \text{ where } 1 \leq \alpha < \infty$$

- The Gumbel copula describes an interdependence structure in which there is upper tail dependence (the level of which is determined by the parameter α), but there is no lower tail dependence.
- The Gumbel copula describes **positive** upper tail dependence for $\alpha > 1$.
- The Gumbel copula is often referred to as the Gumbel-Hougaard copula.

Explicit Copulas

Gumbel copula: Example

Question:

Derive an expression for the Gumbel copula for the case where there are 3 variables. The Gumbel copula has a generator function:

$$\psi(t) = (-\ln t)^{\alpha} \text{ where } 1 \leq \alpha < \infty$$

Answer:

- The Gumbel copula is an example of Archimedean copulas.
- For the case where there are 3 variables, Archimedean copulas take the form:

$$C[u, v, w] = \psi^{[-1]}(\psi(u) + \psi(v) + \psi(w))$$

- So, we need to find $\psi^{[-1]}$, the pseudo-inverse generator function.
- Check $\psi(0) = \lim_{t \to 0} (-\ln t)^{\alpha} = \infty$
- So, the pseudo-inverse $\psi^{[-1]}$ is equal to the 'ordinary' inverse $\psi^{(-1)}$.

Explicit Copulas

Gumbel copula: Example

Answer (continued)

Let $y = \psi^{(-1)}(x)$, then $\psi(y) = x$. So $(-\ln y)^{\alpha} = x$, or $-\ln y = x^{1/\alpha}$. We can write

$$\ln y = -x^{1/\alpha}$$

$$y = \exp(-x^{1/\alpha})$$

$$\psi^{(-1)}(x) = \exp(-x^{1/\alpha})$$

So, $C[u, v, w] = \psi^{[-1]}(\psi(u) + \psi(v) + \psi(w))$

$= \psi^{[-1]}([-\ln u]^{\alpha} + [-\ln v]^{\alpha} + [-\ln w]^{\alpha})$

$= \psi^{(-1)}([-\ln u]^{\alpha} + [-\ln v]^{\alpha} + [-\ln w]^{\alpha})$, (since $\psi(0) = \lim_{t \to 0}(-\ln t)^{\alpha} = \infty$)

$= \exp(-\{[-\ln u]^{\alpha} + [-\ln v]^{\alpha} + [-\ln w]^{\alpha}\}^{1/\alpha})$

Explicit Copulas

Clayton copula

The Clayton copula is defined in the bivariate case as:

$$C[u, v] = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha} \text{ for } \alpha > 0$$

- The generator function is:

$$\psi(t) = \frac{1}{\alpha}\left(t^{-\alpha} - 1\right) \text{ where } \alpha \geq -1 \text{ and } \alpha \neq 0$$

- The Clayton copula describes an interdependence structure in which there is lower tail dependence (the level of which is determined by the parameter α), but there is no upper tail dependence.

Explicit Copulas

Frank copula

The Frank copula is defined in the bivariate case as:

$$C[u,v] = -\frac{1}{\alpha}\ln\left(1+\frac{\left(e^{-\alpha u}-1\right)\left(e^{-\alpha v}-1\right)}{\left(e^{-\alpha}-1\right)}\right) \quad \text{for } \alpha \neq 0$$

- The generator function is:

$$\psi(t) = -\ln\left(\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}\right) \quad \text{where } \alpha \neq 0$$

- The Frank copula describes an interdependence structure in which there is no upper or lower tail dependence.

Implicit Copulas

- Implicit copulas do not have simple closed-form expressions.
- They are based on or implied by well-known multivariate distributions.
- 2 examples of implicit copulas:
  1. The Gaussian copula – based on multivariate normal distribution
  2. The Student's *t* copula – based on multivariate Student's *t* distribution

Implicit Copulas

Gaussian copula

The Gaussian copula is defined in the bivariate case as:

$$C[u, v] = \Phi_\rho\,[\Phi^{-1}(u), \Phi^{-1}(v)]$$

sometimes written as

$$C_{Gauss}[u, v] = \Phi_\rho\,[\Phi^{-1}(u), \Phi^{-1}(v)]$$

where

- Φ is the distribution function of the standard normal distribution and
- Φ_ρ is the distribution function of a bivariate normal distribution with correlation ρ.

- Applying this Gaussian copula to normal marginal distributions will result in a bivariate normal distribution with correlation ρ.
- The independence, co-monotonic and counter-monotonic copulas are special cases of the Gaussian copula where $\rho = 0$, $\rho = +1$ and $\rho = -1$ respectively.

Implicit Copulas

Gaussian copula: Example

Question:

Consider a two-dimensional Gaussian copula function, $C_{Gauss}[u, v]$, with parameter $\rho = 0$:

i. Give the solution to the copula function $C_{Gauss}[1, 1]$.

ii. Give the solution to the copula function $C_{Gauss}[1, 0.2]$.

iii. Give the solution to the copula function $C_{Gauss}[0.2, 0.2]$.

iv. Outline how your answers to parts (i), (ii) and (iii) would change if $\rho = 1$.

Implicit Copulas

Gaussian copula: Example

Answer:

The independence copula is a special case of the Gaussian copula when $\rho = 0$.

When $\rho = 0$, $C_{Gauss}[u, v] = uv$

i. $C_{Gauss}[1, 1] = 1*1 = 1$

ii. $C_{Gauss}[1, 0.2] = 1 * 0.2 = 0.2$

iii. $C_{Gauss}[0.2, 0.2] = 0.2 * 0.2 = 0.04$

iv. The co-monotonic copula is a special case of the Gaussian copula when $\rho = 1$.

When $\rho = 1$, $C_{Gauss}[u, v] = \min(u, v)$

$C_{Gauss}[1, 1] = \min(1,1) = 1$

$C_{Gauss}[1, 0.2] = \min(1, 0.2) = 0.2$

$C_{Gauss}[0.2, 0.2] = \min(0.2, 0.2) = 0.2$

- So, if $\rho = 1$, answers to (i) and (ii) will remain the same but answer to (iii) will change to 0.2.

Implicit Copulas

Student's t copula

The Student's *t* copula is defined by:

$$C[u, v] = t_{\gamma,\rho}\,[t_{\gamma}^{-1}(u),\ t_{\gamma}^{-1}(v)]$$

where

- t_{γ} is the distribution function of a random variable with Student's *t* distribution with γ degrees of freedom and
- $t_{\gamma,\rho}$ is the distribution function of a bivariate Student's *t* distribution with correlation ρ.

- The Student's *t* copula allows the dependencies between the variables to be adjusted more finely than the corresponding Gaussian copula.
- This is because the Student's *t* copula has an additional parameter γ.
- In the same way that the standard normal distribution is a limiting case of the Student's t distribution (as the number of degrees of freedom tends to infinity), the Gaussian copula is the limiting case of the Student's *t* copula.

Implicit Copulas

Gaussian copula

The Gaussian copula is defined in the bivariate case as:

$$C[u, v] = \Phi_\rho [\Phi^{-1}(u), \Phi^{-1}(v)]$$

sometimes written as

$$C_{Gauss}[u, v] = \Phi_\rho [\Phi^{-1}(u), \Phi^{-1}(v)]$$

where

- Φ is the distribution function of the standard normal distribution and
- Φ_ρ is the distribution function of a bivariate normal distribution with correlation ρ.

➢ Applying this Gaussian copula to normal marginal distributions will result in a bivariate normal distribution with correlation ρ.

➢ The independence, co-monotonic and counter-monotonic copulas are special cases of the Gaussian copula where $\rho = 0$, $\rho = +1$ and $\rho = -1$ respectively.

Choosing a suitable copula function

- If we want to create a mathematical model to represent real-world phenomena, then we might look at past data and:
  1) select and parameterise marginal distributions for each of the relevant variables, and
  2) describe and quantify the form and extent of the associations between the variables.
- Examination of the form and levels of association between the variables of interest allows us to select a suitable candidate copula from the list of established copulas or to develop a new bespoke copula.
- Different copulas result in different levels of tail dependence. Example:
  - the Gumbel copula has zero lower tail dependence but upper tail dependence of $2 - 2^{1/\alpha}$
  - the Clayton copula, on the other hand, has zero upper tail dependence but lower tail dependence of $2^{-1/\alpha}$
  - the Frank copula has zero dependence in both tails.
  - the Gaussian copula has zero dependence in both tails (unless $\rho = 1$ in which case the Gaussian copula is a co-monotonic copula and thus has positive upper and lower tail dependence)
  - the Student's t copula has equal positive dependence in both tails.

Choosing a suitable copula function

- As we would expect, variables related by the independence (product) copula have a concordance of 0, whereas variables related by the co-monotonic (minimum) or counter-monotonic (maximum) copulas have a concordance of +1 and –1 respectively.
- So, the Gumbel copula, with an appropriate value for the parameter α , might be a suitable copula to use when modelling large general insurance claims resulting from a common underlying cause.

Choosing a suitable copula function

For bivariate cases, we can summarise the coefficients of lower and upper tail dependence:

Copula name	λ_L	λ_U
Independence	0	
Co-monotonic	1	
Counter-monotonic	0	
Gumbel	0	$2-2^{1/\alpha}$
Clayton	$2^{-1/\alpha}$ if $\alpha > 0$ 0 if $-1 \leq \alpha < 0$	0
Frank	0	
Gaussian	0 if $\rho < 1$ 1 if $\rho = 1$	
Student's t	>0 if $\gamma < \infty$, increasing as γ decreases 0 if $\gamma = \infty$ and $\rho \neq 1$ 1 for all γ when $\rho = 1$	

Queen Mary
University of London