

Class: TY BSc

Subject: Statistical and Risk Modelling - 4

Chapter: Unit 1

Chapter Name: Copulas and its application



# Today's Agenda

- 1. Introduction
  - 1. Introduction to terminology
- 2. Copulas
  - 1. Definition
  - 2. Sklar's Theorem
  - 3. Expressions for Tail dependency in terms of copula
- 3. Types of Copula
  - 1. Fundamental copula
  - 2. Graphical representation of copula
  - 3. Explicit copula
  - 4. Implicit copula
- 4. Choosing and Fitting a copula



### 1 Introduction

Insurance and investment companies are often interested in being able to compute the joint probability of events occurring, for example the joint probability of losses on different classes of business or on investments, or the joint probability of default on investments. One way of calculating a joint probability is to use a joint PDF (or probability function in the case of discrete random variables) and then to integrate (or sum) this to find the probability. There are a number of drawbacks to this approach.



An alternative way of calculating a joint probability is to use a copula. A copula is a function that takes as inputs marginal CDFs and outputs a joint CDF.

There are many distinct copula functions, each of which expresses different types and levels of association between the variables.



#### **Association**

Variables are said to be associated if there is some form of statistical relationship between them – whether causal or not. To facilitate comparisons, measures of association can be constructed.

Coefficients of association are generally designed so that their values vary between -1 and +1. Their absolute values increase with the strength of the relationship.

Any one particular type of **coefficient of association measures a particular form of association.** For example, Pearson's correlation coefficient measures the degree to which there is a linear relationship between the variables.



#### **Concordance**

Concordance is another 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.

Spearman's rho and Kendall's tau are two examples of measures of concordance.

### Desirable properties of a measure of concordance / association

A good measure of the concordance (or association) between two variables should have a number of properties.

These include invariance, which requires that the measure of concordance does not change if we apply the same monotone function to the value of each variable.

Pearson's p does not have this property.



### **Tail dependency**

It is often the case in insurance and investment applications that large losses tend to occur together. 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, ie in the upper and lower tails, are of particular importance. These can be measured using the coefficients of upper and lower tail dependence.





### **Coefficient of upper tail dependence**

We can define the coefficient of upper tail dependence as:

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

It considers the probability of the random variable X taking a value in the upper tail of its distribution (eg a tail with a probability mass of  $5\% \rightarrow u = 0.95$ ), given that the random variable Y takes a value in the same sized upper tail of its distribution.

This coefficient is a probability, so it takes a value between 0 and 1.





### **Coefficient of lower tail dependence**

The coefficient of lower tail dependence is defined as:

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

It considers the probability of the random variable X taking a value in the lower tail of its distribution (eg a tail with a probability mass of  $5\% \rightarrow u = 0.05$ ), given that the random variable Y takes a value in the same sized lower tail of its distribution.

This coefficient is a probability, so it takes a value between 0 and 1.



# 2 Copulas

### **Expressing the association between variables explicitly**

The joint distribution combines the information from the marginal distributions and the way in which the variables depend on each another. However, it expresses this dependence implicitly. We cannot immediately see the nature of the interdependence simply by looking at the formula for the joint distribution function.

# Copulas provide an alternative approach that expresses the interdependence between the variables explicitly.

They allow us to deconstruct the joint distribution of a set of variables into components (the marginal distributions plus a copula) that can be adjusted individually.



# 2.1 Copulas



### **Definition of a copula**

A copula is a function that expresses a multivariate cumulative distribution function in terms of the individual marginal cumulative distributions. It is important to remember that a copula is a function.

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

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

This is often written in the more compact form:

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

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

$$c[u_1, u_2, ..., u_d] = F_{X_1, X_2, ..., X_d}(x_1, x_2, ..., x_d)$$
 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.

# 2.1 Copulas

### Three properties of copulas

Copulas must also satisfy three technical properties to ensure that they correctly capture the properties we would expect of a joint distribution in all circumstances.

1. A copula is an increasing function of its inputs:

$$C[u_1, ..., u_i^*, ..., u_d] > C[u_1, ..., u_i, ..., u_d]$$
 for  $u_i^* > u_i$  and  $i = 1, ..., d$ 

2. If all the marginal CDFs are equal to 1 except for one of them, then the copula function is equal to the value of that one marginal CDF:

$$C(1, ..., 1, u_i, 1, ..., 1) = u_i \text{ for } i = 1, 2, ..., d \text{ and } u_i \in [0, 1]$$

3. A copula function always returns a valid probability:

$$C[u_1, u_2, \dots, u_d] \in [0,1]$$

## 2.2 Sklar's Theorem

Sklar demonstrated in 1959 that the dependence structure of a set of random variables can be captured using copulas. The theorem is as follows:

#### Sklar's theorem

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

$$F(x_1, ..., x_d) = C[F_1(x_1), ..., F_d(x_d)]$$

In the case of variables that have a continuous distribution, the copula is unique.

### **Converse of Sklar's theorem**

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



### Lower tail dependence

Recall that the coefficient of lower tail dependence is defined as:

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

Coefficient of lower tail dependence in terms of the copula function

$$\lambda_L = \lim_{u \to 0^+} \frac{c[u, u]}{u}$$

ie 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).



### The 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 (indicated with a bar), which is defined by:

$$ar{F}(x,y) = P(X > x, Y > y) = ar{C}[ar{F}_X(x), ar{F}_Y(y)]$$
 where  $ar{F}_X(x) = 1 - F_X(x)$  and  $ar{F}_Y(y) = 1 - F_Y(y)$ .

Contd.

### The survival copula

By the principle of inclusion / exclusion, we have:

$$P(X \le x \text{ and/or } Y \le y) = P(X \le x) + P(Y \le y) - P(X \le x, Y \le y)$$
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:

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

### **Upper tail dependence**

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

We can then define the coefficient of upper tail dependence as:

$$\lambda_U = \lim_{u \to 1^-} P(X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)) = \lim_{u \to 0^+} \frac{\bar{c}[u, u]}{u}$$

Coefficient of upper tail dependence in terms of the copula function

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C[u, u]}{1 - u}$$



# 3 Types of Copulas

There are three main families of copula that we will go on to consider in the subsequent sections of this chapter:

- (i) fundamental copulas
- (ii) explicit copulas.
- (iii) implicit copulas.



Fundamental copulas represent the three basic (or fundamental) dependencies that a set of variables can display, namely:

- independence
- perfect positive interdependence, and
- perfect negative interdependence.

These copulas are referred to as the:

- independence (or product) copula
- co-monotonic (or minimum) copula
- counter-monotonic (or maximum) copula.

Collectively these three copulas are referred to as fundamental copulas.

### Independence (or product) copula

One example of a bivariate copula is the product copula C[u, v] = uv.

Here we have:

$$F_{X,Y}(x,y) = C[F_X(x), F_Y(y)] = F_X(x)F_Y(y)$$
or: 
$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

This captures the property of independence of the two variables X and Y, and so is also called the independence (or product) copula.

### **Co-monotonic (or minimum) copula**

This copula is used where random variable demonstrate perfect positive interdependence.

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

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

Here we have:

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

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

### Counter-monotonic (or maximum) copula

The co-monotonic copula captures the relationship between two variables whose values are perfectly positively interdependent on each other, while the countermonotonic copula captures the corresponding inverse relationship.

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

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

Here we have:

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

or:

$$P(X \le x, Y \le y) = \max(P(X \le x) + P(Y \le y) - 1,0)$$

#### The multivariate case

The independence and co-monotonic copulas can be extended in the obvious way to the multivariate case. However, the counter-monotonic copula cannot. This is because it is not possible to have three or more variables where each pair has a direct inverse relationship.

In the multivariate case, we can extend the independence and co-monotonic copulas to d dimensions as follows:

$$ind \ C[F_{X_1}(x_1), ..., F_{X_d}(x_d)] = F_{X_1}(x_1) \times \cdots \times F_{X_d}(x_d)$$

$$min C[F_{X_1}(x_1), ..., F_{X_d}(x_d)] = min (F_{X_1}(x_1), ..., F_{X_d}(x_d))$$

However, it is impossible to have three or more variables, eg  $x_1$ ,  $x_2$  and  $x_3$ , each of which always move in the opposite direction to all of the others. This is why the countermonotonicity copula is only defined in two dimensions.



# 3.2 Graphical representation of Copulas

Bivariate copulas can be represented graphically in various ways:

- scatterplots
- 3D (perspective) representations, and corresponding contour plots.

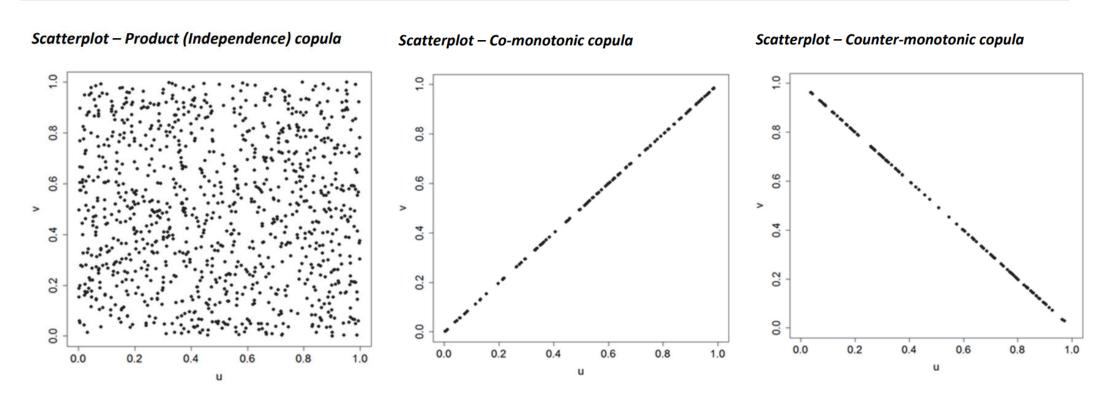
**Scatterplots** - The relationships implied by a copula can be illustrated using a scatterplot of simulated values of u and v.

**3D representations and contour diagrams** - The relationships described by copula functions, illustrated by the scatterplots (above), can also be represented in 3 dimensions: u, v and C[u, v].



# 3.2 Graphical representation of Copulas

### **Scatterplots - Example**

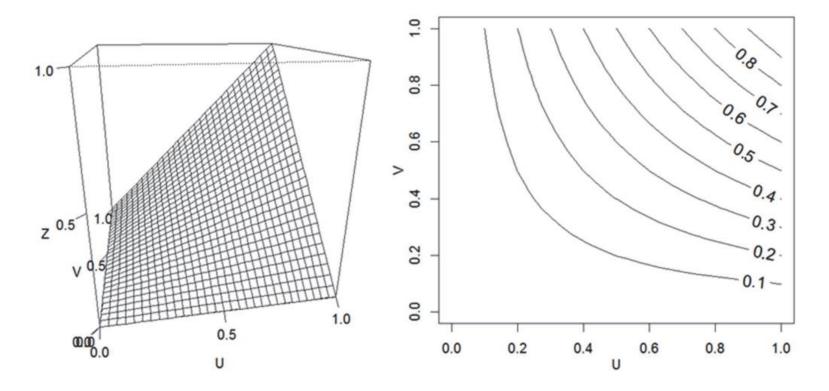




# 3.2 Graphical representation of Copulas

### **3D representations and contour diagrams - Example**

independence (or product) copula:





Explicit copulas have simple closed-form expressions. An important subclass is that of Archimedean copulas.

Several copulas can be specified by a special form of generator function that automatically captures the properties required for a copula. These are called Archimedean copulas.

Archimedean copulas are described by reference to a generator function. In the bivariate case, they take the form:

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

where  $\psi(x)$  is the generator function, and  $\psi^{[-1]}$  is the pseudo-inverse function (explained below).

Three examples of Archimedean copulas are: the Gumbel, Clayton and Frank copulas.

### **The Gumbel Copula**

The Gumbel copula is defined in the bivariate case as:

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

Note that the Gumbel copula is often referred to as the Gumbel-Hougaard copula.

The Gumbel copula can be defined by the generator function:

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

which we can use to deduce an explicit formula for the copula function.

The Gumbel copula describes an interdependence structure in which there is upper tail dependence (the level of which is determined by the parameter  $\alpha$ ), but there is no lower tail dependence.



### The Gumbel Copula – Tail dependency

For the Gumbel copula setting u = v gives:

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

$$= \exp\{-(2(-\ln u)^{\alpha})^{1/\alpha}\}$$

$$= \exp\{-(2^{1/\alpha}(-\ln u))\}$$

$$= \exp\{(2^{1/\alpha}\ln u)\}\}$$

$$= \exp\{\ln u^{2^{1/\alpha}}\}$$

$$= u^{2^{1/\alpha}}$$

The coefficient of upper tail dependence is given by:

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C[u, u]}{1 - u} = \lim_{u \to 1^-} \frac{1 - 2u + u^{2^{1/\alpha}}}{1 - u}$$



### **The Gumbel Copula – Tail dependency**

In the limit this fraction has the form  $\frac{0}{0'}$  which is undefined. However, we can use L'Hôpital's rule,  $\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)'}$  to find the value of the limit:

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + u^{2^{1/\alpha}}}{1 - u} = \lim_{u \to 1^-} \frac{-2 + 2^{1/\alpha} u^{2^{1/\alpha} - 1}}{-1} = 2 - 2^{1/\alpha}$$

As  $\alpha$  increases,  $2^{1/\alpha}$  reduces and hence  $2-2^{1/\alpha}$  increases. So, increasing the value of the parameter  $\alpha$  increases the degree of upper tail dependence of the Gumbel copula.



### **The Clayton Copula**

The Clayton copula is defined in the bivariate case as:

$$c[u,v] = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$$

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

The Clayton copula is defined by the generator:

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



### **The Frank Copula**

The Frank copula is defined in the bivariate case as:

$$c[u,v] = -\frac{1}{\alpha} \ln \left( 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{(e^{-\alpha} - 1)} \right)$$

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

The Frank copula is defined by the generator:

$$\psi(t) = -\ln\left(\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}\right)$$
 where  $-\infty < \alpha < \infty$ 



### Independence (or product) copula

The independence (or product) copula is also Archimedean.

Its generator is  $\psi(t) = -\ln t$ .





## Question

### **CS2A A2021 Q1**

The Frank copula,  $C_F$ , for a bivariate distribution is defined as:

$$C_F(u, v) = -\frac{1}{\alpha} ln \left( 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{(e^{-\alpha} - 1)} \right), \alpha > 0$$

- (i) Determine the probability that two jointly distributed random variables, X and Y, are both less than or equal to their median values where X and Y follow the Frank copula,  $C_F$ , with  $\alpha = 1$ . [2]
- (ii) Determine the revised value of the probability in part (i) when  $\alpha = 0.1$ . [1]
- (iii) Determine the probability that two jointly distributed random variables, *X* and *Y*, are both less than or equal to their median values where *X* and *Y* follow the product copula. [1]
- (iv) Comment on your answers to parts (i), (ii) and (iii) with reference to the sign and level of dependence exhibited by the Frank copula. [2]

  [Total 6]



## Solution

```
(i) The required probability is C_{-}F(0.5, 0.5)
= -\ln (1 + [(e^{-0.5} - 1)^{2}]/[e^{-1} - 1])
= 0.280930
(ii) The required probability is C_{-}F(0.5, 0.5)
= -(1/0.1) * \ln (1 + [(e^{-0.05} - 1)^{2}]/[e^{-0.1} - 1])
= 0.253125
[½]
```



### Solution

```
(iii)
The required probability is
        C_{\text{Product}}(0.5, 0.5)
        = u * v = 0.5 * 0.5
                                                                                                          [\frac{1}{2}]
                                                                                                          [\frac{1}{2}]
        = 0.25
(iv)
The probabilities for the Frank copula in parts (i) and (ii) are higher than for the product
copula in part (iii)
                                                                                                          [\frac{1}{2}]
because the Frank copula exhibits positive dependence
                                                                                                          [\frac{1}{2}]
As α approaches zero, the level of dependence approaches zero
                                                                                                          [\frac{1}{2}]
and so, the probability approaches that under the product copula.
                                                                                                          [\frac{1}{2}]
                                                                                                    [Total 6]
```



# 3.4 Implicit copulas

The final group of copulas that we consider are called implicit copulas. These copulas are based on (or implied by) well-known multivariate distributions, but no simple closed-form expression exists for them.

#### We look at:

- the Gaussian copula (based on the multivariate normal distribution)
- the Student's t copula (based on the multivariate Student's t distribution).

# 3.4 Implicit copulas

### **The Gaussian Copulas**

The bivariate Gaussian copula is defined by:

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

where  $\Phi$  is the distribution function of the standard normal distribution and  $\Phi_{\rho}$  is the distribution function of a bivariate normal distribution with correlation  $\rho$ .

Applying this Gaussian copula to normal marginal distributions will result in a bivariate normal distribution with correlation  $\rho$ .

The formula defining the bivariate Gaussian copula is mathematically equivalent to the following integral form:

$$C[u, v] = \int_0^u \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(t)}{\sqrt{1 - \rho^2}}\right) dt$$



# 3.4 Implicit copulas

### The Student's t copula

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

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

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



# 4 Choosing and fitting a suitable copula function

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:

- select and parameterise marginal distributions for each of the relevant variables, and
- 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.



# 4.1 The upper and lower tail dependence

Copula name	$L^{\lambda}$	$u^{\lambda}$
Independence	0	
Co-monotonic	1	
Counter-monotonic	О	
Gumbel	0	$2 - 2^{1/\alpha}$
Clayton	$2^{-1/\alpha}$ if $\alpha > 0$ $\alpha \leq 0$	0
Frank	0	



# 4.1 The upper and lower tail dependence

Gaussian	o if $ ho < 1$ 1 if $ ho = 1$	
Student's t	$> 0$ if $\gamma < \infty$ , increasing as $\gamma$ decreases $\gamma$ or if $\gamma = \infty$ and $\rho \neq 1$ $\gamma$ for all $\gamma$ when $\rho = 1$	

The degree of concordance and the level of tail dependencies exhibited by a particular set of data helps to indicate which copula(s) might be appropriate to consider using.



## Question

### **CS2A S2022 Q7**

An insurer writes three different classes of insurance business: X, Y and Z. The classes have the following total annual claims distributions:

 $X \sim Exp(0.08)$ 

 $Y \sim Normal(10, 22)$ 

 $Z \sim Normal(20, 32)$ 

The insurer models the level of dependency between the classes' total claim amounts using a Clayton ( $\alpha = 2$ ) copula.

- (i) Calculate the probability, using the Clayton (2) copula, that both X < 3 and Y < 8. [3]
- (ii) Calculate the probability, using the Clayton (2) copula, that all of the following occur: X < 3, Y < 8 and Z < 20. [3]





# Question

(iii) Calculate the probability, using the Clayton (2) copula, that both X > 10 and Y > 12. [3]

A student has noted that copulas are a useful modelling tool as 'they allow us to model different degrees of dependency'.

(iv) Comment on this statement. [3] [Total 12]



### Solution

```
(i)
Using (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}
F_X(3) = 1 - \exp(-0.08 * 3)
= 0:2134,
F_Y(8) = P(Z < -1) where Z \sim N(0, 1)
=0.1587
C(0.2134; 0.1587) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}
= (0.2134^{-2} + 0.1587^{-2} - 1)^{-1/2} = 0.1284
(ii)
Using (u^{-\alpha} + v^{-\alpha} + w^{-\alpha} - 1)^{-1/\alpha}
F Z(20) = P(Z < 0) \text{ where } Z \sim N(0, 1)
= 0.5
C(0.2134; 0.1587; 0.5) = (u^{-\alpha} + v^{-\alpha} + w^{-\alpha} - 1)^{-1/\alpha}
= (0.2134^{-2} + 0.1587^{-2} + 0.5^{-2} - 2)^{-\frac{1}{2}} = 0.1253
```



### Solution

```
(iii)
F X(10) = 1 - \exp(-0.08 * 10) = 0.5507
                                                                                                 [\frac{1}{2}]
F Y(12) = P(Z < 1) where Z \sim N(0, 1)
= 0.8413
                                                                                                 [\frac{1}{2}]
C(0.5507; 0.8413) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}
=(0.5507^{-2}+0.8413^{-2}-1)^{-1/2}=0.5191
                                                                                                 [\frac{1}{2}]
Then using the survival copula, the required probability is
1 - F X(10) - F Y(12) + C(0.5507; 0.8413)
                                                                                                  [1]
= 1 - 0.5507 - 0.8413 + 0.5192 = 0.1271
                                                                                                 [\frac{1}{2}]
(iv)
It is true that part of the benefit of copulas is that they allow us to model different
degrees of dependency
                                                                                                  [1]
However, copulas also exhibit different patterns of dependency
                                                                                                  [1]
As well as different degrees of dependency in the body of the distributions, copulas
exhibit different degrees of dependency in the tails
                                                                                                  [1]
Copulas allow the dependency structure between random variables to be modelled
separately from the marginal distributions
                                                                                                  [1]
                                                                  [Marks available 4, maximum 3]
                                                                                          [Total 12]
```