Topic 2: Copula methods

Copulae

Formulation

For the multivariate normal, independence is equivalent to absence of correlation between any two components. In this case the joint cdf is a product of the marginals. When the independence is violated, the relation between the joint multivariate distribution and the marginals is more involved. An interesting concept that can be used to describe this more involved relation is the concept of *copula*. We focus on the two-dimensional case for simplicity. Then the copula is a function $C : [0,1]^2 \rightarrow [0,1]$ with the properties:

1. $C(0,u) = C(u,0) = 0$ for all $u \in [0,1]$.
2. $C(u,1) = C(1,u) = u$ for all $u \in [0,1]$.
3. For all pairs $(u_1, u_2), (v_1, v_2) \in [0,1] \times [0,1]$ with $u_1 \leq v_1, u_2 \leq v_2$:

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

The name is due to the implication that the copula links the multivariate distribution to its marginals. This is explicated in the following theorem:

Theorem 6.1. *Let $F(\cdot,\cdot)$ be a joint cdf with marginal cdf's $F_{X_1}(.)$ and $F_{X_2}(.)$. Then there exists a copula $C(\cdot,\cdot)$ with the property*

$$F(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$

for every pair $(x_1, x_2) \in \mathbb{R}^2$. When $F_{X_1}(\cdot)$ and $F_{X_2}(\cdot)$ are continuous the above copula is unique. Vice versa, if $C(\cdot,\cdot)$ is a copula and $F_{X_1}(\cdot), F_{X_2}(\cdot)$ are cdf then the function $F(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$ is a joint cdf with marginals $F_{X_1}(\cdot)$ and $F_{X_2}(\cdot)$.

Taking derivatives we also get:

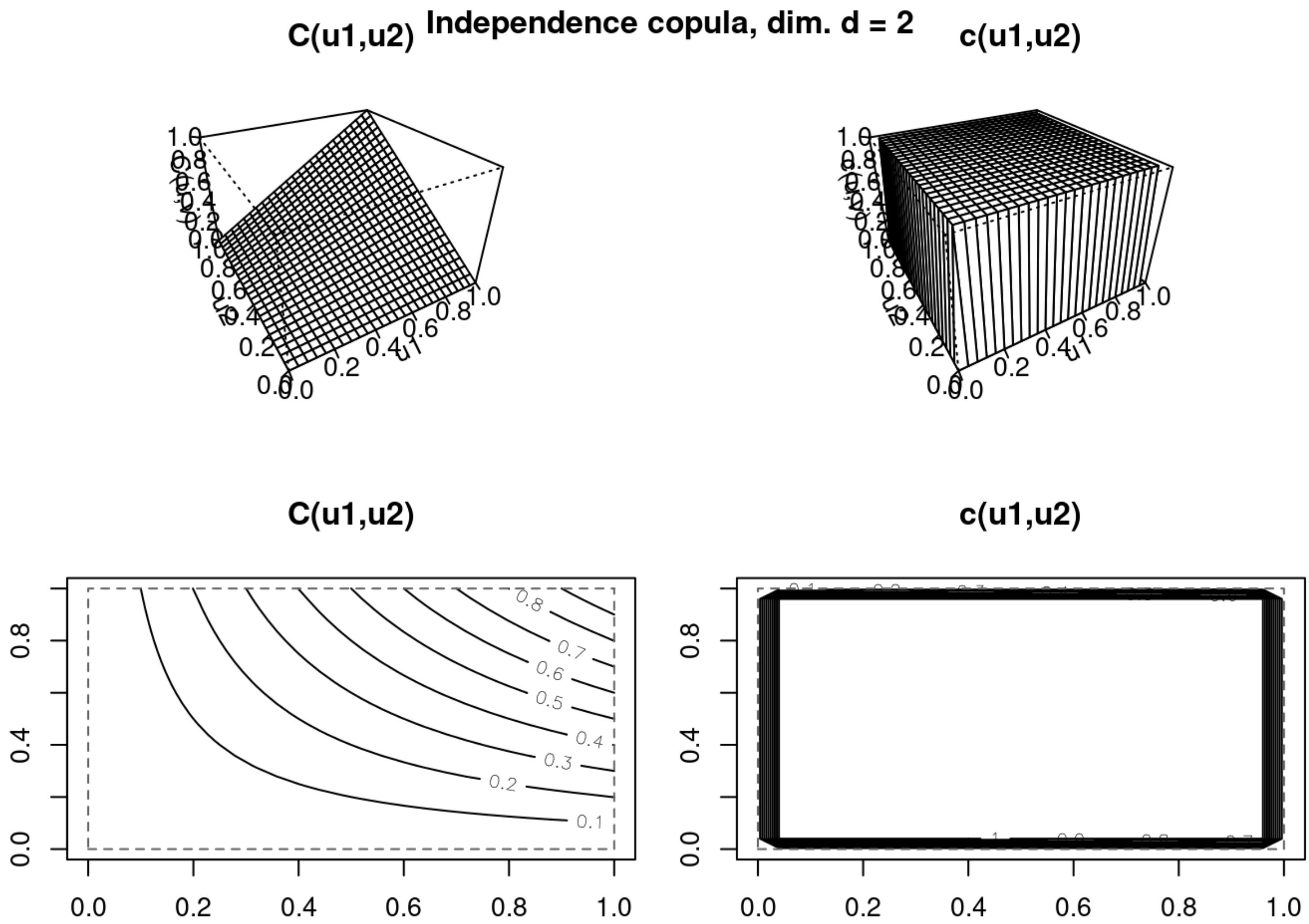
$$f(x_1, x_2) = c(F_{X_1}(x_1), F_{X_2}(x_2)) f_{X_1}(x_1) f_{X_2}(x_2)$$

where

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$$

is the *density* of the copula. This relation clearly shows that the contribution to the joint density of X_1, X_2 comes from two parts: one that comes from the copula and is "responsible" for the dependence ($c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$) and another one which takes into account marginal information only ($f_{X_1}(x_1) f_{X_2}(x_2)$).

It is also clear that the independence implies that the corresponding copula is $\Pi(u,v) = uv$ (this is called the independence copula). Here's what it looks like:



These concepts are generalised also to p dimensions with $p > 2$.

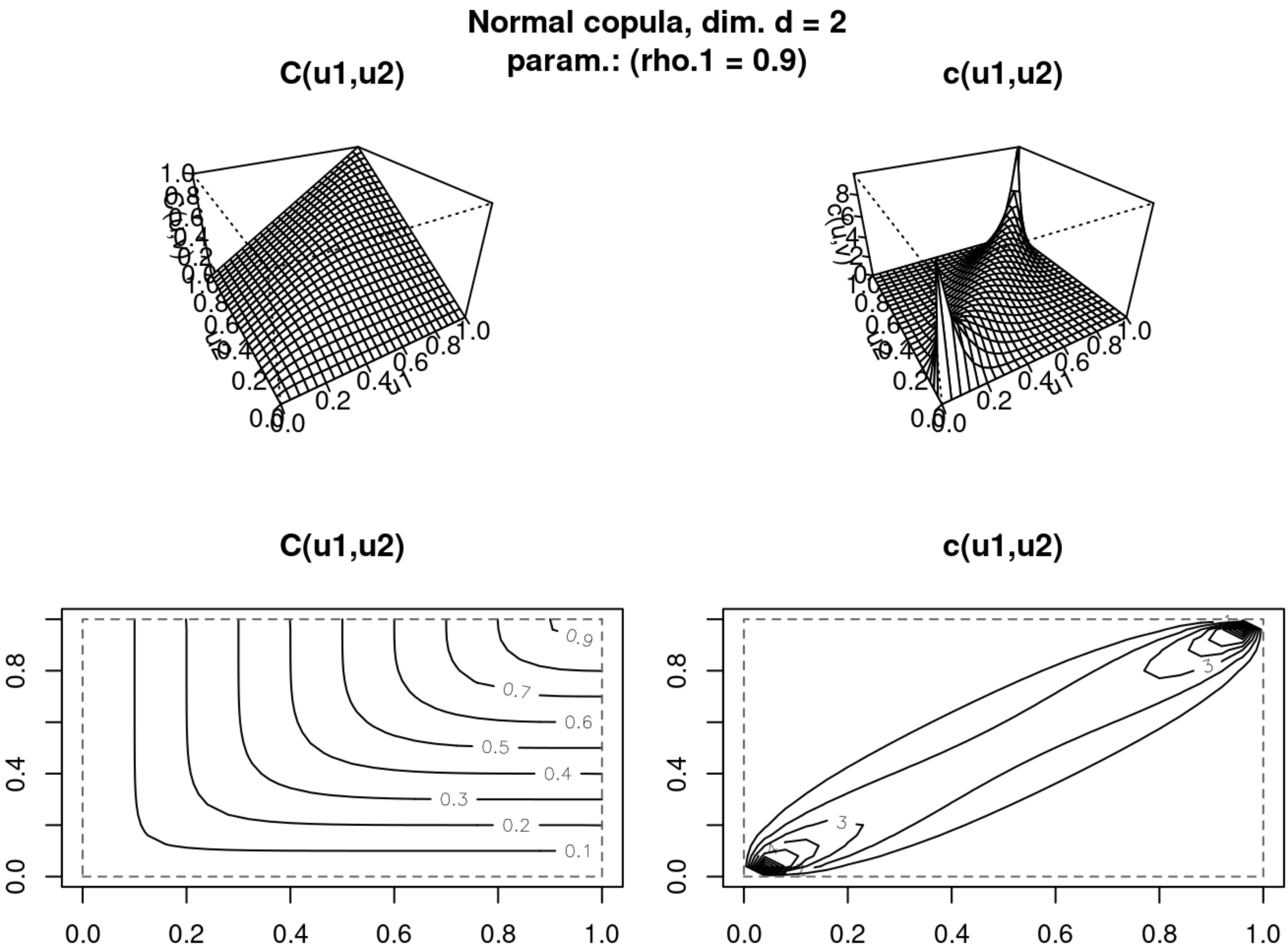
Common copula types

An interesting example is the *Gaussian copula*. For $p = 2$ it is equal to

$$\begin{aligned} C_\rho(u,v) &= \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} f_\rho(x_1, x_2) dx_2 dx_1. \end{aligned}$$

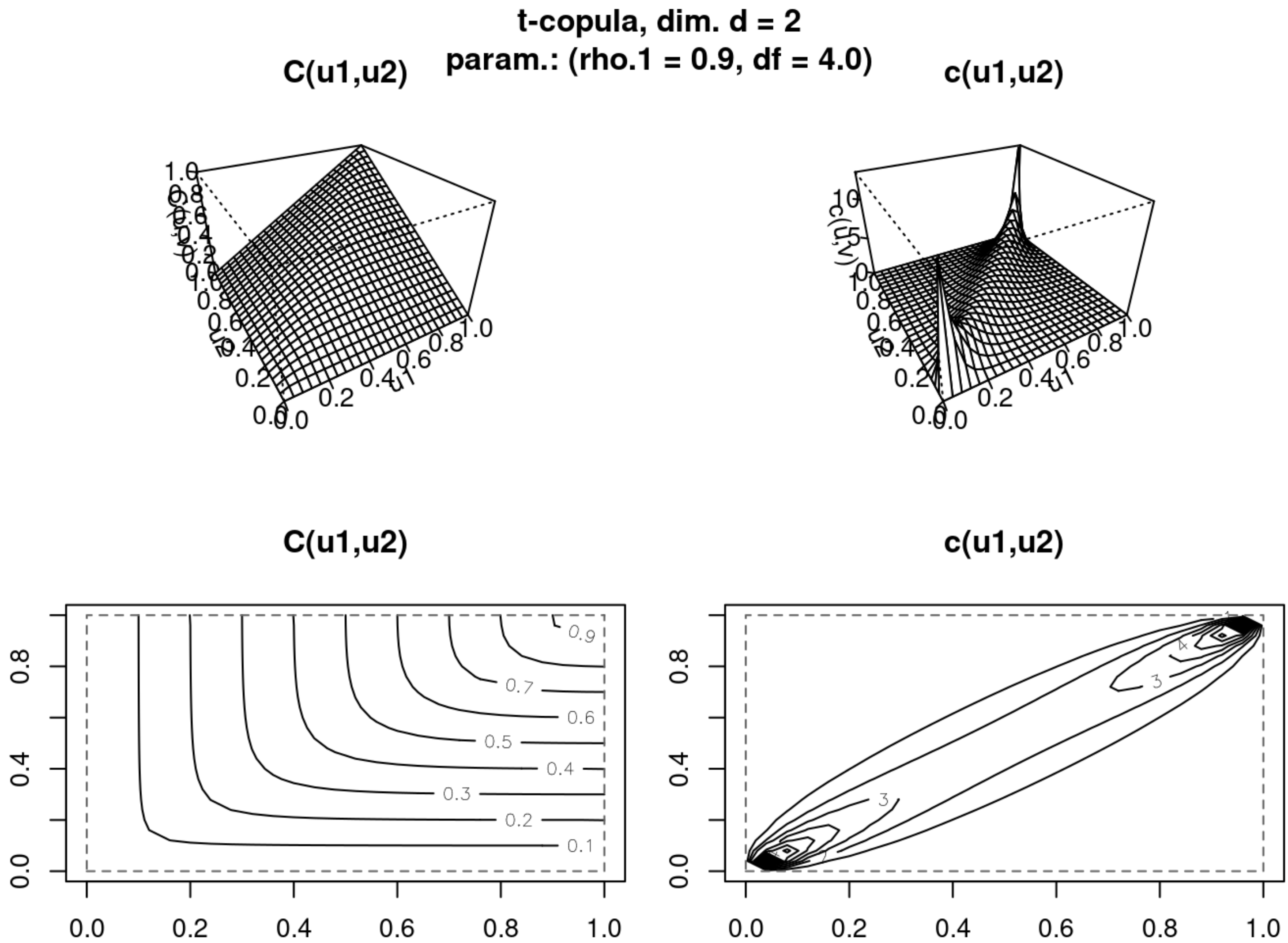
Here $f_\rho(\cdot,\cdot)$ is the joint bivariate normal density with zero mean, unit variances and a correlation ρ, $\Phi_\rho(\cdot,\cdot)$ is its cdf, and $\Phi^{-1}(\cdot)$ is the inverse of the cdf of the standard normal. (This is "The formula that killed Wall street".) When $\rho = 0$ we see that we get $C_0(u,v) = uv$ (as is to be expected).

Here's what it looks like for $\rho = 0.9$.



Non-Gaussian copulae are much more important in practice and inference methods about copulae are a hot topic in Statistics. The reason for importance of non-Gaussian copulae is that Gaussian copulae do not allow us to model reasonably well the tail dependence, that is, joint *extreme* events have virtually a zero probability. Especially in financial applications, it is very important to be able to model dependence in the tails.

A *Multivariate-t copula* uses the multivariate t distribution instead. This distribution has heavier tails, which allows it to model extreme events better:



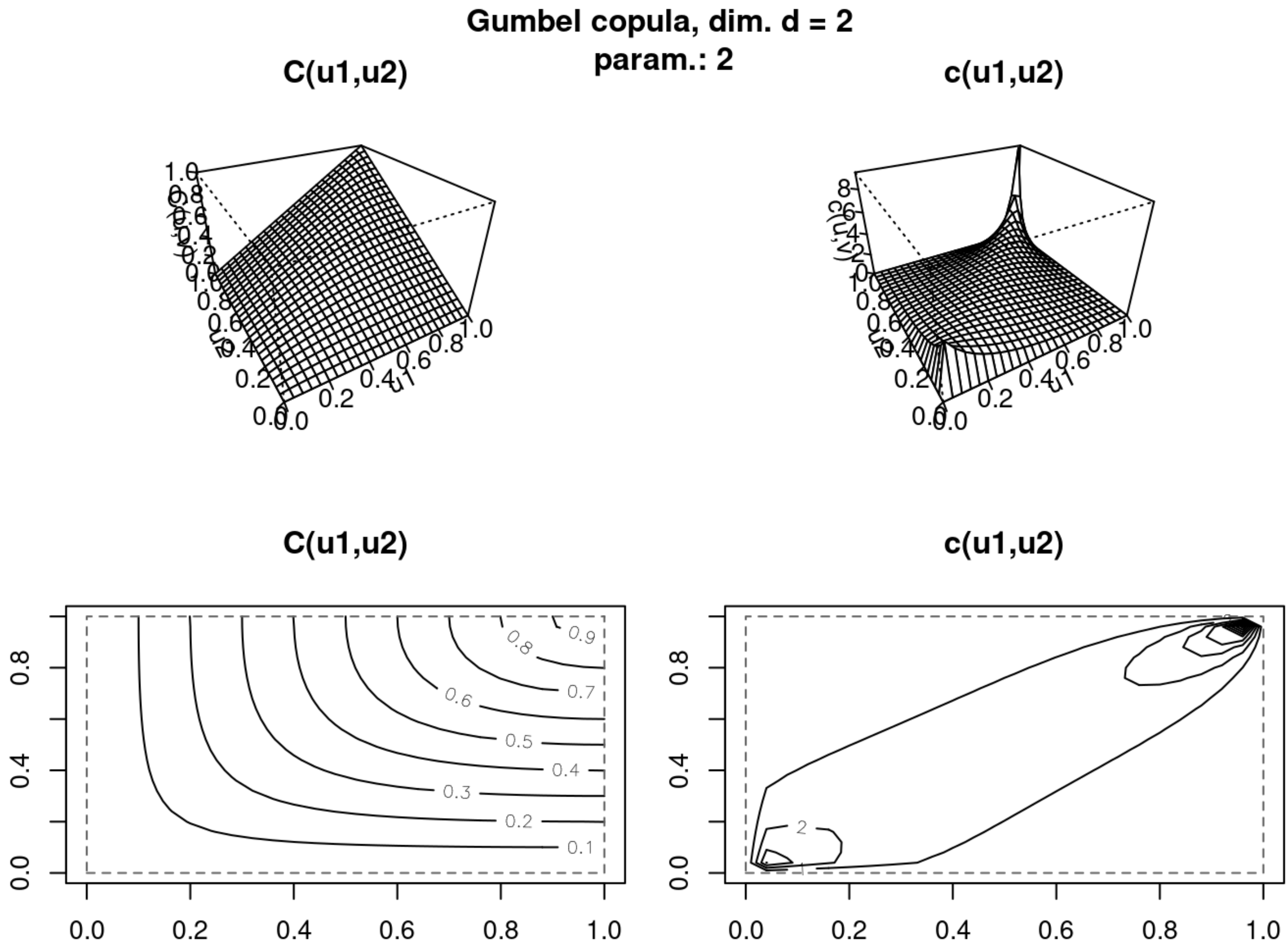
The Gumbel–Hougaard copula is much more flexible in modeling dependence in the upper tails. For an arbitrary dimension p is is defined as

$$C_\theta^{\mathrm{GH}}(u_1, u_2, \ldots, u_p) = \exp\left\{ - \left[\sum_{j=1}^{p} (-\log u_j)^\theta \right]^{1/\theta} \right\},$$

where $\theta \in [1, \infty)$ is a parameter that governs the strength of the dependence. You can easily see that the Gumbell–Hougaard copula reduces to the independence copula when $\theta = 1$ and to the Fréchet–Hoeffding upper bound copula $\min(u_1, \ldots, u_p)$ when $\theta \to \infty$.

Here's how this copula looks for $\theta = 2$:

Note the asymmetry, absent in the Gaussian and t-copulae.



The Gumbel–Hougaard copula is also an example of the so-called *Archimedean* copulae. The latter are characterised by their *generator* $\phi(\cdot)$: a continuous, strictly decreasing, convex function from $[0, 1]$ to $[0, \infty)$ such that $\phi(1) = 0$. Then the Archimedean copula is defined via the generator as follows:

$$C(u_1, u_2, \dots, u_p) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_p)).$$

Here, $\phi^{-1}(t)$ is defined to be 0 if t is not in the image of $\phi(\cdot)$.

Example 6.2. Show that the Gumbell–Hougaard copula is an Archimeden copula with a generator $\phi(t) = (-\log t)^{\theta}$.

The benefit of using the Archimedean copulae is that they allow for simple description of the p-dim dependence by using a function of one argument only (the generator). However, it is seen immediately that the Archimedean copula is symmetric in its arguments and this limits its applicability for modelling dependencies that are not symmetric in their arguments. The so-called *Liouville* copulae are an extension of the Archiemedean copulae and can be used also to model dependencies that are not symmetric in their arguments.

Computing

R: Packages `copula`, `VineCopula`, and others.

Optional viewing: Gaussian copula

Bionic Turtle. (2009). Gaussian copula. Retrieved from: https://youtu.be/z43_pf5Y6A8

Check your understanding

Complete the below exercises to check your understanding of concepts presented so far.

The (p-dimensional) Clayton copula is defined for a given parameter $\theta > 0$ as

$$C_\theta\left(u_1, u_2, \ldots, u_p\right) = \left[\sum_{i=1}^{p} u_i^{-\theta} - p + 1\right]^{-1/\theta}$$

Show that it is an Archimedean copula and that its generator is $\phi(x) = \theta^{-1}(x^{-\theta} - 1)$.

Demonstration: Copula methods

Start by watching the following demonstration by Dr Krivitsky before proceeding to the task.

Transcript

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete this task select the 'Copula_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

Copula_Examples.demo.Rmd

The output of the RMD file is also displayed below:

Challenge: Copula methods

If you choose to complete this task in your own RStudio, upload the following file:

Copula_Examples.challenge.Rmd

Click on the 'Cancor_Examples.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The solution will also be available here on Friday of this week by clicking on the 'Solution' tab in the top right corner.

The output of the RMD file is also displayed below: