

QUANTITATIVE RISK MANAGEMENT USING COPULAS, FMSN65/MASM33

COMPUTER ASSIGNMENT 2

This assignment is a compulsory part of the course and will be carried out by each group consisting of two students. The report (a PDF file) should be uploaded at the home page of the course in *Canvas* by the corresponding deadline. A text file (with txt-extension in the name) containing the R commands or programs used in each part of the assignment has to be also uploaded separately.

Copulas and their applications in risk management

Exercise 1: Some theoretical questions related to copulas

Answer the following questions:

1. What are the properties of Archimedean copulas and their generators?
2. Write the bivariate distribution functions for Frank, Clayton, Gumbel and AMH copulas.
3. Find Laplace transform of a random variable which is distributed as $\text{Gamma}(a, b)$ (the density is given in (71) in collection of formulas). Set shape parameter $a = 1/\theta$ and rate parameter $b = 1$. Explain how this result is related to Archimedean copulas and find the corresponding copula in the bivariate case.
4. Suppose \mathbf{X} is a d -dimensional random vector. Explain how multivariate normal and multivariate t distributions are defined.
5. Explain how the dependency measures Pearson's correlation ρ , Kendall's τ and Spearman's ρ_S are defined. What is the relationship between Pearson's correlation and two other dependency measures?

Getting started with the computational part:

If you plan to carry out the assignment on your own computer, you need to download and install R from <http://www.r-project.org>.

If you want to carry out the assignment in the computer room MH:230, you just need to start one of the PCs first. Then choose the latest version of R from the **Start** menu.

In this assignment you need to use package `copula`. It is strongly recommended that you download the corresponding PDF file of the manual (`copula.pdf`) from cran.r-project.org to use as a reference in the assignment. Type `library(copula)` to load the package to your R-session. The package is also installed in all computers in the computer room MH:230. If you use your own laptop you need to install package yourself; see the help page for `install.packages` for more information.

In exercises 2, 4 and 5 below you will generate observations from different distributions. As you will be using the simulated values in other parts of this assignment you should save each simulated dataset in an object for later use.

Exercise 2: Random number generation from copulas

In all the exercises below you will plot several datasets. Note that you can use `mfrow` to collect any number of the plots in one single figure.

In each part you need to estimate different measures of dependence and compare them by their theoretical values. Note that for elliptical copulas $\tau = \frac{2}{\pi} \arcsin(\rho)$ and $\rho_S = \frac{6}{\pi} \arcsin(\rho/2)$. Moreover, for Archimedean copulas the relationship between the parameter of copula and Kendall's τ can be found in the attached table. Finally note that Spearman's ρ_S can be calculated as the correlation between the ranks of the observations.

1. Generate 500 observations from a 3-dimensional normal copula with an exchangeable covariance matrix with variance 1 and correlation 0.4. Plot the scatter plot of the simulated observations.

For each pair of the simulated sample calculate the empirical values of Pearson's correlation ρ , Kendall's τ and Spearman's ρ_S . Do the estimates agree with the theoretical values? In addition, calculate the ranks of the observations and verify that the Spearman's ρ_S can be calculated as the correlation between the ranks of the observations.

2. Generate 500 observations from a bivariate t copula with 8 degrees of freedom, an exchangeable characteristic matrix with diagonal 1 and off-diagonal values $\theta = -0.4, 0, 0.4$ (use one value of θ in each simulation). Plot the scatter plot for each dataset and comment on dependence between observations and parameter θ .

For the simulated sample calculate the empirical values of Pearson's correlation ρ , Kendall's τ and Spearman's ρ_S . Do the estimates agree with the theoretical values?

3. Generate 500 observations from a 3-dimensional Frank copula with parameter 3. For each pair of the simulated sample calculate the empirical values of Pearson's correlation ρ , Kendall's τ and Spearman's ρ_S . Do the estimates of Kendall's τ agree with its theoretical value?
4. Use your results in Exercise 1 part 3 above and write a function which generates random observations from the corresponding copula. The function should be able to generate any

number of observations in an arbitrary dimension for a given θ . In addition, it should not contain any `for()`, `while()` or similar commands for one by one iteration. Include your function in your report. Generate 500 observations from this copula in 3-dimensional case with parameters $\theta = 0.5, 1, 5, 50$.

5. Suppose X is random variable uniformly distributed in the interval $[-1, 1]$. Let $Y = T(X) = \operatorname{arctanh}(X)$ and $Z = -T(X) = -\operatorname{arctanh}(X)$.
 - Is function $T(x)$ increasing? Why?
 - Generate 1000 observations from X and plot the corresponding functions for Y and Z .
 - Calculate Spearman's ρ , Kendall's τ and Pearson's correlation for the simulated pairs (x_i, y_i) and (x_i, z_i) . Compare the results and comment on whether all three dependence measures produce reasonable results.

Exercise 3: Graphics for copulas

For two simulated dataset from bivariate t in part 2 of Exercise 2 above plot the simulated values and the corresponding pdf and cdf of each copula. Comment on the plots. Note that you can use `mfrow` to collect all of the plots in one single figure.

Exercise 4: Comparing different copulas

In this exercise we will compare four different copulas. Simulate 2000 observations from four bivariate copulas as the following: the Gaussian copula with parameter $\rho = 0.7$; the Gumbel copula with parameter $\theta = 2$; the Clayton copula with parameter $\theta = 2.2$; the t copula with parameters $df = 4$ and $\rho = 0.71$. You will need to use the simulated samples in the second part of this exercise so make sure you to save all the samples in R. Use `par(mfrow=c(2,2))` and plot all the simulated samples in one single plot. Comment on any differences you see between copulas.

Use quantile transformation and transform all the simulated samples to bivariate distributions with standard normal margins. Make a similar plot as above for the simulated bivariate observations. Note that the parameters of the copulas have been chosen so that all of these distributions have a linear correlation that is roughly 70%. Do you see any differences in the dependence structure of the simulated observations despite the fact that all the simulated samples have the same margins with the same linear correlations? Comment on the results.

Exercise 5: Random number generation from bivariate distributions

In each of the following parts calculate both the linear and Spearman's correlation for each dataset. Theoretically rank correlation should not be affected by change of the margins. At least in one case below calculate both linear and rank correlation for the copula and corresponding bivariate distribution with arbitrary margins and show that the rank correlation does not depend on the marginal distribution.

1. Generate 500 observations from a 2-dimensional distribution with normal copula (parameter 0.75) and $N(0, 2)$ and $\exp(2)$ margins.
2. Generate 500 observations from a 2-dimensional distribution with Gumbel copula (parameter 2) and $\exp(4)$ and $\exp(2)$ margins.

Exercise 6: Graphics for multivariate distributions

For each simulated dataset above plot the simulated values. Plot also contour plots of pdf of the corresponding distributions. Note that you can use `mfrow` to collect all of the plots in one single figure.

Exercise 7: Car data

A collection of measurements were taken from a representative sample of new cars in 1993. Because some of the variables are measured at an ordinal scale, Spearman's ρ is more appropriate than Correlation for measuring monotonic association.

The dataset "`cardata.txt`" can be downloaded from the following locations

- The page **Datasets in R** under the **Pages** in the homepage of the course in *Canvas*,
 - <http://www.maths.lth.se/matstat/kurser/fmsn15masm23/datasetsR.html>.
1. Read the data to *R* (use `read.csv()` function) and plot a scatterplot matrix of the data by using `pairs` function in *R*.
 2. calculate Spearman's ρ corresponding to the scatterplot matrix.
 3. Does your results suggest that
 - vehicles with higher horsepower are more costly?
 - higher fuel economy (MPGcity) meant lower prices in 1993?
 4. Can you suggest a formal test to answer the same questions above?

Hint: You can use `help.search` or `RSiteSearch` to search the help system for key words or phrases.

Exercise 8: Calculating tail probabilities

Consider a portfolio of 2 risks X_1 and X_2 . Let the risks represent potential losses in dependent lines of business for an insurance company and let u_1 and u_2 be some thresholds. Suppose the insurer seeks reinsurance for the situation that both losses exceed their thresholds. In this case these losses will be paid in full by the re-insurer. In the following we will analyze the problem from the re-insurer's point of view so their loss is when both thresholds are exceeded.

Based on historical data we can assume that $X_i \sim \text{Lognormal}(0, 1)$, $i = 1, 2$. Further we assume that $\tau(X_1, X_2) = 0.5$ and we take both thresholds u_i , $i = 1, 2$ equal to 0.95-quantile of each marginal distribution. Answer the following questions based on each of the following assumptions for the joint distribution:

- Gaussian copula
 - t-copula with 2 degrees of freedom
 - Gumbel copula.
1. Calculate probability that both losses exceed their thresholds.

2. Calculate the ratio of the payout probabilities for the reinsurance company for each pair of copulas.
3. How do the ratios change if we increase the threshold to the 0.99-quantile of each margin?
4. Use simulation to estimate expected value of loss for the reinsurance company corresponding to the both choices of thresholds above.
5. Compare also the upper tail dependence for the copulas and comment on the practical consequences of your results.

Exercise 9: Calculating reinsurance premiums

In this exercise we will use copulas to analyze an insurance dataset on indemnity claims. Figure 1 shows observations on 1500 liability claims for an insurance company. The indemnity payment (Loss) and the allocated loss adjustment expense (ALAE) are recorded in USD for each claim. Here, ALAE are types of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims investigation expenses.

The dataset "insuranceData.txt" can be downloaded from the following locations

- The page **Datasets in R** under the **Pages** in the homepage of the course in *Canvas*,
- <http://www.maths.lth.se/matstat/kurser/fmsn15masm23/datasetsR.html>.

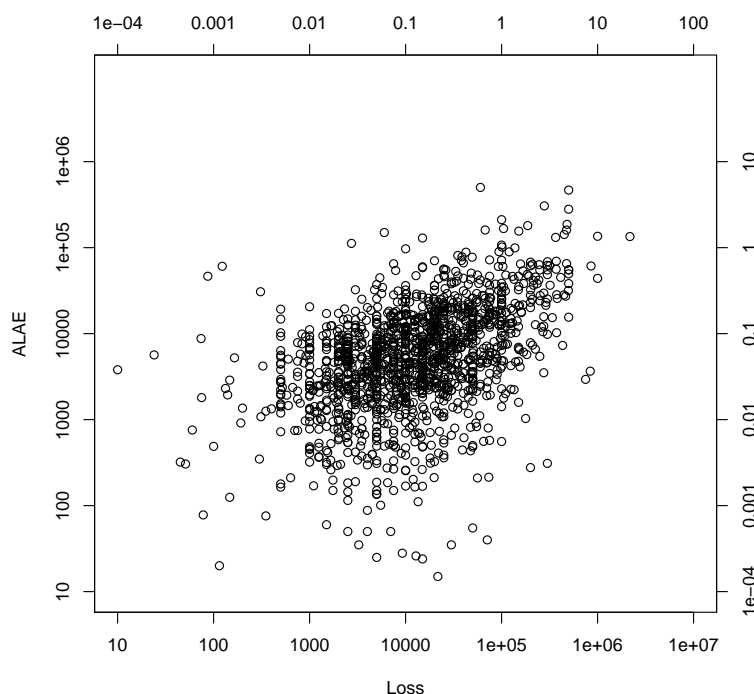


Figure 1: Scatterplot of ALAE versus Loss: The bottom and left axes represent the original data (log-scale). In the analysis we scale the data so that one unit corresponds to 100 000 USD. The corresponding rescaled axes are in the top and right side of the figure.

1. Read the data to *R* (use `read.table("insuranceData.txt",header=T)` function) and plot a figure similar to Figure 1.

2. Calculate the correlation between Loss and ALAE.
3. Fit a bivariate distribution to the dataset by using copulas. For general optimization you can use the function `optim()` in *R*. Note that you need to pass the dataset as a matrix to `optim()` otherwise it will not produce any results (see the function `as.matrix()`). In addition, the package `fitdistrplus` can be used to estimate parameters for several univariate distributions.

Note that you need to provide the details of your calculations here. Specifically, you have to present the details of the univariate distributions you have tried for each margin and justify your final choice of the marginal distributions by providing suitable checks or goodness of fits plots. The same applies to the choice of copula. You should also use different estimation methods (FML, IFM and CML) in fitting the models.

Suppose a reinsurer considers selling reinsurance to this company. As a simple model assume they agree on a policy with limit L for the reinsurer and retention R for the insurance company.

To formulate this policy in mathematical terms let X_1 and X_2 denote the Loss and ALAE, respectively. Assuming a proportional sharing of expenses, the above policy means that the reinsurer's payment, $RP(X_1, X_2)$, can be expressed as the following:

$$RP(X_1, X_2) = \begin{cases} 0 & \text{if } X_1 < R \\ X_1 - R + \frac{X_1 - R}{X_1} X_2 & \text{if } R \leq X_1 < L \\ L - R + \frac{L - R}{L} X_2 & \text{if } X_1 \geq L \end{cases}$$

Suppose the reinsurer wants to estimate $E[RP(X_1, X_2)]$ using simulations.

1. Use the fitted distribution above and estimate $E[RP(X_1, X_2)]$ for all combinations of $L = 10\,000, 500\,000, 1\,000\,000$ and $R/L = 0, 0.25, 0.75, 0.95$.
2. Suppose one makes an unrealistic assumption that Loss and ALAE are independent. Calculate the corresponding values of $E[RP(X_1, X_2)]$ for the same combinations of L and R as in the previous part. In which cases the reinsurance premium is under- or overvalued?

Finishing Off:

When you've finished, close down *R* by typing `q()`. Choose 'Save' when prompted as to whether you want to retain your workspace.

Copula	τ	λ_U	λ_L
Gauss	$\frac{2}{\pi} \arcsin \rho$	0 for $\rho < 1$	0 for $\rho < 1$
	$-1 \leq \tau \leq 1$	1 for $\rho = 1$	1 for $\rho = 1$
t_ν	$\frac{2}{\pi} \arcsin \rho$	$2\bar{t}_{\nu+1} \left(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right)$	λ_U
	$-1 \leq \tau \leq 1$	$-1 \leq \rho_S \leq 1$	
Gumbel	$1 - \frac{1}{\theta}$	$2 - 2^{\frac{1}{\theta}}$	0
	$0 \leq \tau \leq 1$		
Clayton	$\frac{\theta}{\theta+2}$	0	$2^{-\frac{1}{\theta}}$
	$0 \leq \tau \leq 1$		
Frank	$1 - \frac{4}{\theta} \{1 - D_1(\theta)\}$	0	0
	$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$		
	$-1 \leq \tau \leq 1$		

Table 5: Kendall's τ and TDCs for various selected copulae.