

# Dealing with Dependent Risks

Claudia Klüppelberg

Robert Stelzer

In most real life situations we are not only confronted with one single source of risk or one single risk, but with several sources of risk or combinations of risks. An important question is whether individual risks influence each other or not. This may concern the time of their occurrence and/or their severity. In other words, we need to understand how to model and describe the dependence structure of risks. Clearly, if risks influence each other in such a way that they tend to occur together and increase the severity of the overall risk, then the situation may be much more dangerous than otherwise.

We illustrate this with a concrete example. Consider a building which could be hit by an earthquake and a flood. If the building is situated on the Japanese coast, an earthquake may damage the building and cause a tsunami, which in turn floods the building. Hence, it is quite likely that by these two combined sources of risk a particularly disastrous event occurs. In other words, there is a strong positive dependence between these two risks (high damage from an earthquake will often come along with high damage from a flood). This does not mean that they always occur together, since an earthquake does not necessarily cause a tsunami or there may be a flood only due to heavy rain.

## **The facts.**

- Dependence between risks and/or sources of risks are crucial for risk assessment, quantification and management.
- Adequate mathematical measures for the dependence of risks (or more general of random variables) are needed.
- Correlation measures linear dependence, but characterises the full dependence structure only in special parametric models (the multivariate normal distribution is *the* typical example).
- Correlation is also useful in spherical and elliptical distributions.
- Rank correlations are appropriate dependence measures in certain situations.
- Copulae provide a way to characterise the dependence structure completely, but are rather complex objects.
- For risk assessment it is mainly the dependence structure of extreme events that matters. Thus, measures for dependence in extreme observations provide useful dependence measures for combined risks.

# 1 Introduction

In most situations (both in our professional and our daily life) risks are present. Often there exist various sources of risk which, in the end, determine the overall risk of a more or less complex system. This is a common situation in the financial world (i.e., for any bank and insurance company), in any engineering system, when working as a physician or when dealing with environmental consequences. It is then necessary to assess and deal with combinations of risks in an appropriate way.

There is a huge difference between two risks possibly occurring together and risks happening at different times. In one situation you need to be prepared to deal with both risks at the same time, whereas in the other situation it suffices to cope with one risk at a time. For example, if you consider the people needed on stand-by for the emergency services, you will need many more people in the first case. However, in almost all situations life is not even that easy, risks do not have to occur at the same time, instead they may or tend to occur at the same time. Then we need to understand and quantify this tendency. This is exactly what this paper is about, to understand how to model the statistical dependence between different risks.

There are two classical approaches. The first one assesses the single risk factors by some monetary risk measure appropriately, and simply adds the different values of the single risk measures together. The second one is to combine the monetary risks with a multivariate normal model, and assess the dependence via the pairwise correlations.

Both approaches only capture part of the truth, and in this chapter we discuss their appropriateness, other approaches and the pros and cons of different approaches to model and measure risks of complex systems.

We are concerned with risk under dependence and thus we briefly have to make precise what we mean by this. In the end we want to use risk measures (see the chapter Biagini, Meyer-Brandis, and Svindland [4] in this book for a detailed introduction) to quantify risks, as well as to assess the effects of risk management strategies. Essentially, we want to understand the effects the dependence structure has on these risk measures. In models it is of utmost importance to have an appropriate dependence structure capturing all effects relevant for the risk measures. So we want to discuss both how to model dependence and the effects of different ways of modelling on the final risk assessment.

Therefore let us briefly introduce two risk measures and note that we identify risk with a random variable; i.e., the outcome of a risky event.

**Definition 1.1 (Examples of “risk measures”)** *For a random variable  $X$  with distribution function  $F(x) = P(X \leq x)$  for  $x \in \mathbb{R}$  we define the following risk measures:*

(a) Variance:  $\text{var}(X) = E((X - E[X])^2) = E(X^2) - (E(X))^2$  is the mean squared deviation from the mean or expected value of  $X$ .

(b) Value-at-Risk: Define the quantile function of  $F$  as

$$F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}, \quad \alpha \in (0, 1). \quad (1.1)$$

*Note that for strictly increasing  $F$  this is simply the analytic inverse.*

*Then for a large value of  $\alpha$  (usually  $\alpha = 0.95$  or larger)  $\text{VaR}_\alpha = F^{-1}(\alpha)$  is called the Value-at-Risk (for the level  $\alpha$ ).*

The first risk measure, i.e. the variance, gives the average squared difference between a random variable (the realisation of a risk) and its mean outcome. It measures how widely spread various outcomes are. Clearly, it is a very simplistic risk measure, since e.g. it does

not differentiate between values higher and values lower than the mean, as it looks only at the squared distance. Normally, only one direction really matters when considering a particular risk. For instance, if we consider the level of a river in a German city and the flood risk, then it is irrelevant when the level is far smaller than the mean (of course, the “downside” direction may well matter for other risks, e.g. that water becomes scarce) .

The Value-at-Risk or VaR is a very popular risk measure in particular in the financial world. Above it has been assumed that the high realizations of  $X$  are “risky”, but this is only a convention and can be changed to low realizations being risky. Intuitively, the value at risk gives the level which is not exceeded in  $100 \cdot \alpha\%$  of all cases (e.g. if the VaR at the level 0.95 is 500, then the relevant variable, “the risk”, is above 500 in 5% of all cases and in 95% of all cases it is below 500). Moreover, the VaR has been incorporated into the Basel II regulations (the international rules governing how much capital banks must set aside to cover future losses from their business) and Solvency II (similar international rules for insurance companies), and the national legislation which enforces these international standards. VaR is the standard risk measure to be used there (cf. Fassen, Klüppelberg, and Menzel [9] for estimation methods).

We will see later, in particular in Illustration 2.3, that changing the dependence structure usually has tremendous effects on the VaR. But note that VaR has been rightly criticized for various reasons:

- a) VaR takes only the event of large losses into account, but not the size of losses. In this sense the so called Tail-VaR is preferable, which measures the average of all losses exceeding VaR. So if a bank sets aside capital equal to its VaR it certainly goes bankrupt (or needs to be “rescued”), as soon as a loss occurs which is higher than VaR. In contrast to this, if it used the Tail-VaR to determine its risk capital, it has set aside enough capital to withstand such an event on average. So there should be a realistic chance that the capital is sufficient to cover the loss.
- b) VaR is not always a coherent risk measure. For a risk measure to be coherent (cf. Biagini et al. [4]) it is necessary that the risk measure of the sum of two risks is always below the sum of the two risk measures. Since e.g. banks estimate the VaR for each unit and add all resulting VaRs up to estimate the risk of the whole bank, the use of VaR may underestimate the true bank’s VaR considerably.

As a very readable paper on dependence measures and their properties and pitfalls, which goes far beyond the present chapter, we recommend [2].

This paper is structured as follows. In Section 2 we introduce the mathematical definitions of (in)dependence of random variables and illustrate the effects of different dependence structures. In Section 3 we recall the multivariate normal distribution and discuss which kind of dependence it is able to model. We continue this in Section 4 where we consider the correlation as a popular dependence measure discussing in detail its properties, problems, limitations and popular misconceptions. As the next natural step we present spherical and elliptical distributions in Section 5. Thereafter, we turn our focus on alternative dependence measures starting with rank correlations in Section 6. Then in Section 7 we consider a concept – copulae – at length which in principle is able to completely encode all possible dependence structures. As typically extreme events are the really dangerous risks, we indicate how to quantify and model the dependence of extreme events in Section 8. Finally, we give you as our readers some Food for Thought in Section 9 and provide a brief summary in Section 10.

## 2 Independence and dependence

The first simple question to answer is, when exactly do we have dependence between risks? The best answer seems to be a negative one, viz. risks are dependent whenever they are not independent.

Clearly, this means that we have to give a mathematical definition of independence. We do this for two random variables  $X$  and  $Y$  (which represent the risks we are interested in). Think for instance, of our example of an earthquake and a flood at the beginning. Intuitively, independence should mean that whatever happens in one random variable, say  $X$  (the earthquake), should in no way affect what happens in  $Y$  (the flood). If we know the value of  $X$ , this should not change our knowledge of what might happen with which probability in  $Y$ . In proper mathematical terms one says that two random variables are independent if their joint distribution is the product of the two marginal distributions; i.e.,  $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$  for all  $x, y \in \mathbb{R}$  (note:  $P(A)$  means the probability that some event  $A$  occurs.) This implies that the probability distribution of  $Y$  conditional on  $X$  does not depend on  $X$ , but is simply equal to the distribution of  $Y$ ; i.e.,  $P(Y \leq y | X \leq x) := P(X \leq x, Y \leq y) / (P(X \leq x)) = P(Y \leq y)$  for all  $x, y \in \mathbb{R}$ . Obviously this is in line with the intuition given above.

Note that the necessity of a negative definition of dependence tells us that there are (too) many ways in which risks can be dependent. Hence, any mathematical object completely describing the dependence of arbitrary random variables has to be a very complex object. Turned the other way around any simple quantification of dependence – like one real number obtained from the joint distribution of two random variables – will necessarily only reflect a very special aspect of dependence, or describe the dependence completely only in very special situations/set-ups. This should be kept in mind throughout the rest of this chapter and whenever trying to quantify dependence in applications.

In truly realistic situations we are interested in the (in)dependence of more than two random variables. We give the general definition and discuss and illustrate it afterwards.

**Definition 2.1 (Independence)** *Let  $X_1, X_2, \dots, X_n$  for  $n \in \mathbb{N}$  be random variables. Then  $X_1, X_2, \dots, X_n$  are called independent if*

$$P(X_1 \leq x_1, \dots, X_d \leq x_d) = P(X_1 \leq x_1) \cdots P(X_d \leq x_d) \quad (2.1)$$

*holds for all  $x_1, \dots, x_d \in \mathbb{R}$ .*

Let us consider two special cases that are particularly relevant in applications.

- a) Assume the random variables  $X_1, X_2, \dots, X_n$  are discrete; i.e., they can only assume countably many values (e.g. all random variables take only values 0 or 1, or all possible outcomes are natural numbers). Then  $X_1, X_2, \dots, X_n$  are independent if and only if

$$P(X_1 = x_1, \dots, X_d = x_d) = P(X_1 = x_1) \cdots P(X_d = x_d)$$

for all possible values of  $x_1, \dots, x_d$ .

- b) Assume that the random variables  $X_1, X_2, \dots, X_n$  have densities (non-negative functions  $f_i$  such that  $P(X_i \leq x) = \int_{-\infty}^x f_i(t) dt$  for all  $i \in \{1, \dots, n\}$  and  $x \in \mathbb{R}$ ). Provided they have also a joint density; i.e., a non-negative function  $f$  such that  $P(X_1 \leq x_1, \dots, X_d \leq x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_d} f(t_1, t_2, \dots, t_d) dt_1 dt_2 \cdots dt_d$ , then they are independent if and only if

$$f(x_1, x_2, \dots, x_d) = f_1(x_1) f_2(x_2) \cdots f_d(x_d)$$

for all  $x_1, \dots, x_d \in \mathbb{R}$ .

## 2.1 Misconceptions of the independence concept

Unfortunately, there are several popular misunderstandings regarding independence, which we shall discuss now.

### **Misconception 1: Pairwise independence does not necessarily entail independence**

One may be tempted to believe that instead of checking the definition of independence, which involves all random variables, one could check whether all possible pairs of two variables are independent. Unfortunately, such pairwise independence does not imply independence in the sense of Definition 2.1 above. This is illustrated by the following example. Simple random variables (indicator variables) are defined via events  $A, B, C$  by  $1_A, 1_B$  and  $1_C$ , where  $1_A$  is equal to one if the event  $A$  occurs and equal to zero else; analogously for  $B$  and  $C$ . Then our Definition 2.1 is consistent with the usual definition of independent events, which says that events  $A, B, C$  are independent, if  $P(A \cap B \cap C) = P(A)P(B)P(C)$ ,  $P(A \cap B) = P(A)P(B)$ ,  $P(A \cap C) = P(A)P(C)$  and  $P(B \cap C) = P(B)P(C)$  all hold. The following examples shows that independence of all pairs of indicator variables (or events) does not imply independence of all three indicator variables (or events).

**Illustration 2.2** Think of two thunderstorms which we assume to be independent. We only care whether a thunderstorm comes accompanied by hail or not. The probability for a single thunderstorm to come with hail shall be  $1/2$ . Let  $A$  be the event that it hails during the first thunderstorm and  $B$  the event that it hails during the second thunderstorm. Finally, let  $C$  be the event that it either hails or does not hail in both the first and the second thunderstorm. One easily calculates  $P(A) = P(B) = P(C) = 1/2$ . However,  $P(A \cap B \cap C) = P(A \cap B) = 1/4 \neq 1/8 = P(A)P(B)P(C)$ , because if it hails both in the first and the second thunderstorm, the event  $C$  given by no hail in both thunderstorms can no longer occur. So clearly  $A, B, C$  are not independent. However,  $A, B$  are independent by construction and  $P(A \cap C) = P(A \cap B) = P(B \cap C) = 1/4$ ; and thus we have pairwise independence.  $\square$

### **Misconception 2: Total risk is not smallest/largest for independent events**

One cannot conclude in general that the situation of independent risks is particularly (un)favourable from the point of view of the total risk. The reason is that dependence can act both in a risk reducing and risk enhancing way, since typically risk measures are non-linear. We will present some real life examples and discuss the variance and the Value-at-Risk as risk measures.

**Illustration 2.3** Assume that we are confronted with two different risks modelled by two random variables  $X, Y$ . Both random variables are either 0 or 1 (in some monetary unit like 1 million Euros), corresponding to no loss or loss of one monetary unit, each with probability  $1/2$ . For example, in an insurance company  $X, Y$  may describe whether or not damages have been reported for two different insurance contracts and the claim had to be paid (then the corresponding variable is 1, else it is zero). The insurer regards  $X + Y$  as the random variable describing the total risk of both contracts.

When using the variance as risk measure, we simply have to apply the formula

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

Consequently, the risk (in terms of the variance) is equal to the sum of the risks, if  $X$  and  $Y$  are uncorrelated (see Section 4 for the use of the correlation as a dependence measure). The risk of the sum is larger than the sum of risks, if  $X$  and  $Y$  are positively correlated, and likewise the risk of the sum is smaller than the sum of the risks if they are negatively correlated. For the VaR (as well as other more advanced risk measures) the situation is not quite as simple.

Situation 1:  $X$  and  $Y$  are independent (e.g.  $X$  models a life insurance contract and  $Y$  a personal

liability insurance of the same person). Then the loss  $X + Y$  is 0 with probability 1/4, 1 with probability 1/2, or 2 with probability 1/4. When using the Value-at-Risk at the 90% or 70% level as risk measures, one obtains  $\text{VaR}_{0.9}(X + Y) = 2$  and  $\text{VaR}_{0.7}(X + Y) = 1$ .

Situation 2:  $X, Y$  are “completely positive dependent” (e.g.  $X, Y$  are insurance risks for two neighbouring houses of same value against hurricanes; i.e.,  $X = Y$ ). Then the loss  $X + Y$  is 0 with probability 1/2, or 2 with probability 1/2. It can never be 1 and one obtains  $\text{VaR}_{0.9}(X + Y) = \text{VaR}_{0.7}(X + Y) = 2$ .

Situation 3:  $X, Y$  are “completely negative dependent” (e.g.  $X$  is an insurance cover for a farmer against too little rain measured by the annual amount of rain being below a level  $c$ , and  $Y$  is an insurance cover for a holiday resort at the same place against bad weather which pays 1 if the amount of rain is above the same level  $c$ ; i.e.,  $X = 1 - Y$ ). Then the loss  $X + Y$  is 1 with probability 1. It can never be 0 or 2 and one obtains  $\text{VaR}_{0.9}(X + Y) = \text{VaR}_{0.7}(X + Y) = 1$ .

Comparing the values of the VaR for the two different levels in the three examples shows that the risk in the independent situation is neither an upper nor a lower bound on the risk in dependent situations. Note that in the last situation there is actually no risk at all in the sense of an uncertain outcome, because  $X + Y$  is always equal to 1.

Note here also that typical risk measures are non-linear. This is in contrast to the expected value, which for  $X + Y$  is in all situations equal to 1. Hence, our examples illustrate also that the expected value does not at all care about the dependence structure.  $\square$

### 3 Normal distribution

The normal distribution is probably the most widely used probability distribution in applications. Its popularity is due to the facts that it is rather easy to handle, that many properties are known completely explicitly, and often there are arguments that it is a natural distribution to use. By a classical result called the central limit theorem, one can argue that whenever a variable of interest is generated by the averaged results of many different small random effects, then this random variable should be approximately normally distributed. However, this argument has to be used with care and one should always check in detail whether data at hand may reasonably come from a normal distribution.

**Definition 3.1** *A random variable  $X$  is said to be normally distributed with mean  $\mu \in \mathbb{R}$  and variance  $\sigma^2 > 0$ , if it has a probability density given by*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (3.1)$$

If  $\mu = 0$  and  $\sigma^2 = 1$ , we speak of a standard normal random variable.

Dependence issues make sense only for at least two random variables, hence we now turn our focus to multivariate normal distributions. We summarize all risks in a (column) vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$ . We also need the notion of a *positive definite*  $d \times d$  matrix  $\Sigma$  that is a matrix which is symmetric (i.e., the transposed  $\Sigma^\top = \Sigma$ ) and satisfies  $\mathbf{x}^\top \Sigma \mathbf{x} > 0$  for all  $\mathbf{x} \in \mathbb{R}^d$  not equal to the zero vector. We are now ready to define the multivariate normal distribution; cf. the book [11] for many interesting details.

**Definition 3.2** *A  $d$ -dimensional random vector  $\mathbf{X}$  is called normally distributed with mean  $\boldsymbol{\mu} \in \mathbb{R}^d$  and covariance matrix  $\Sigma$  (a positive definite  $d \times d$  matrix), if it has probability density*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d. \quad (3.2)$$

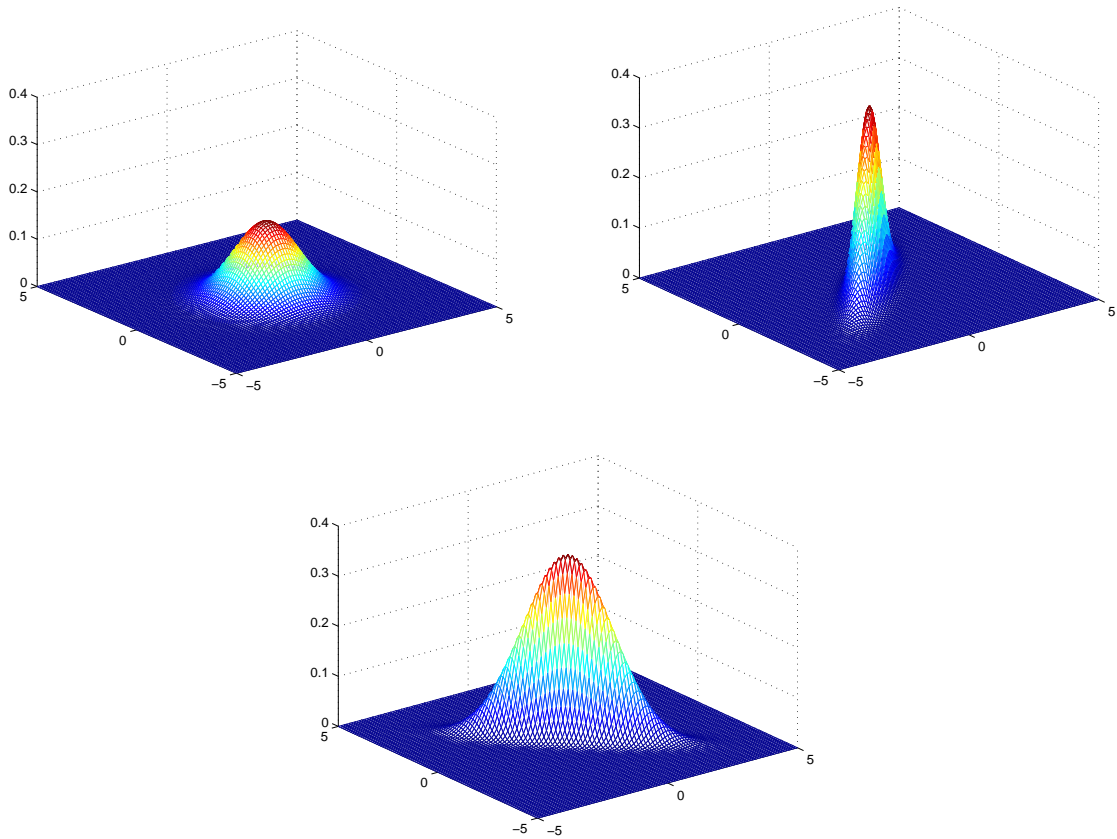


Figure 1: Bivariate normal densities: standard normal (independent components; upper left), normal with variance 1 and covariance  $\rho = 0.9$  (highly positively correlated; upper right), normal with variance 1 and covariance  $\rho = -0.9$  (highly negatively correlated; lower).

If  $\boldsymbol{\mu} = \mathbf{0}$  and  $\Sigma = I_d$  ( $I_d$  being the  $d \times d$ -identity matrix), we speak of a  $d$ -dimensional standard normal vector.

Note that one can also define normal distributions with only a positive semi-definite covariance matrix  $\Sigma$  (i.e., a symmetric matrix satisfying  $\mathbf{x}^\top \Sigma \mathbf{x} \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^d$  not equal to the zero vector). One way to do this is by demanding that  $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$  where  $\mathbf{Y}$  is standard normally distributed (with lower dimension) and  $A$  is chosen such that  $AA^\top = \Sigma$ .

The parameter  $\boldsymbol{\mu}$  is the mean vector of  $\mathbf{X}$  and changing it shifts the distribution (i.e., it only changes the location of the distribution in a non-random way). Hence, it has nothing to do with the dependence structure between the vector components  $X_1, \dots, X_d$ , which therefore must be totally described by  $\Sigma$ .

Each diagonal element  $\Sigma_{ii}$  of the matrix  $\Sigma$  gives the variance of the  $i$ -th coordinate  $X_i$ , whereas the off-diagonal element  $\Sigma_{ij}$  with  $i \neq j$  gives the covariance of  $X_i$  and  $X_j$ , a dependence measure we shall investigate in detail below.

In Figure 1 we depict the densities of several bivariate normal distributions. For the standard normal density the surface is very homogeneous (it is left invariant by rotations), whereas in the two other cases the mass of the distribution (i.e., the area with a high value for the density) is concentrated around the diagonal (i.e., the line where  $x_1 = x_2$ ), or the negative diagonal

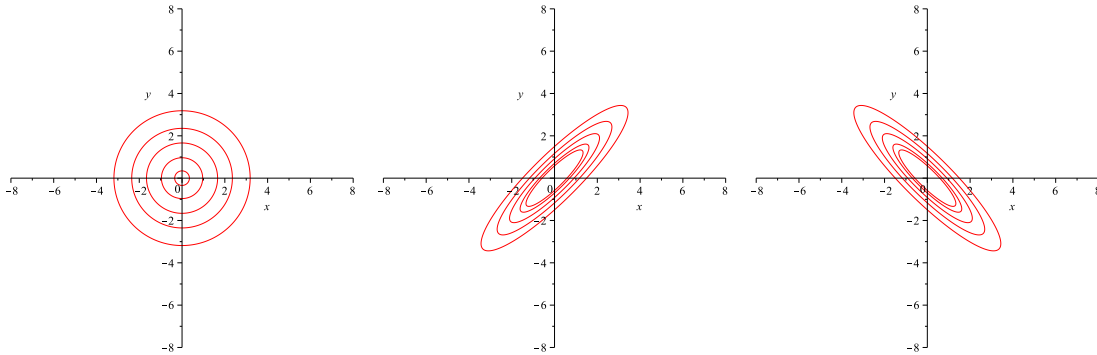


Figure 2: Contour plots of bivariate normal densities: standard normal (independent components; left), normal with variance 1 and covariance  $\rho = 0.9$  (highly positively correlated; middle), normal with variance 1 and covariance  $\rho = -0.9$  (highly negatively correlated; right).

The levels of the contours are 0.15, 0.1, 0.04, 0.01, 0.001

(i.e., the line where  $x_1 = -x_2$ ), respectively. Intuitively it seems that in the standard normal distribution the two components  $X_1$  and  $X_2$  are rather independent, whereas in the other two cases they appear to be rather dependent. This intuition is indeed true.

However, there is more to be learned from these plots. A natural question is, what do the lines look like where the density has a fixed specified value; i.e., what are the sets of possible values  $(x_1, x_2)$  satisfying  $f_{\mathbf{X}}(x_1, x_2) = c$  for some  $c > 0$ . From the plot of the density, we guess that for the standard normal density, these *contour lines* should be circles around the origin. Note that the standard normal density has its maximum at 0 with value  $f_{\mathbf{X}}(0, 0) = 1/(2\pi)$ . We calculate the following for  $c \in (0, 1/(2\pi)]$  from (3.1) (by  $\ln$  we denote the the natural logarithm; i.e., the analytical inverse of the exponential function):

$$\begin{aligned} f_{\mathbf{X}}(x_1, x_2) &= c \\ \Leftrightarrow -\frac{1}{2}(x_1^2 + x_2^2) &= \ln(2\pi c) \\ \Leftrightarrow x_1^2 + x_2^2 &= -2 \ln(2\pi c) \end{aligned}$$

From elementary geometry we recall that this last equation describes the circle around zero with radius  $\sqrt{-2 \ln(2\pi c)}$  (note that  $2\pi c < 1$ , and hence  $\ln(2\pi c) < 0$ ).

In the general case (with arbitrary mean and covariance matrix) we may still assume that  $\boldsymbol{\mu} = 0$ , since the mean changes only the location, not the dependence structure. For arbitrary  $\Sigma$  the sets with equal values for the normal density can also be calculated and we obtain (again only for possible values of  $c$ ) from Definition 3.2 and the formula for the explicit inversion of a  $2 \times 2$  matrix:

$$\begin{aligned} f_{\mathbf{X}}(x_1, x_2) &= c \\ \Leftrightarrow \Sigma_{22}x_1^2 - 2\Sigma_{12}x_1x_2 + \Sigma_{11}x_2^2 &= -2 \det(\Sigma) \ln(2\pi \sqrt{\det(\Sigma)}c). \end{aligned}$$

Since this is again a quadratic equation, elementary geometry tells us that these sets are ellipses centered at the origin. As we shall also discuss in detail later on, the distributions where the contour lines of the density (the lines characterised by the density assuming the same value) are circles or, more generally, ellipses play a special role regarding the description of dependence.



## 4 Correlation as a linear dependence measure

We now discuss the usage of the covariance or correlation as a measure of dependence. We start with a pair  $X, Y$  of random variables representing two different risks. Throughout this section we assume that all random variables have a finite variance; i.e.,  $E(X^2) < \infty$  (equivalently,  $\int_{\mathbb{R}} x^2 f_X(x) dx < \infty$  if  $X$  has a density  $f_X$ ).

Recall that the variance of a random variable  $X$  is given by  $\text{var}(X) = E((X - E(X))^2)$  and can be seen as a measure of the variability of the random variable or, in other words, how much realisations of  $X$  tend to fluctuate around the mean value  $E(X)$ . Note that when  $X$  has a density  $f_X$  then its mean or expectation is  $E(X) = \int_{\mathbb{R}} x f_X(x) dx$ . The covariance of  $X$  and  $Y$  is given by  $\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$ . From the first expression it is obvious that the covariance is a positive number if  $X$  and  $Y$  are “usually” both below or above their mean and negative if “usually” one is above its mean and one below.

The covariance carries information on the dependence, but is also affected by the variability (the typical spread around the mean) of the involved random variables. To get rid of the latter effect and to get a number measuring only dependence aspects one normalises the covariance by dividing the covariance by the product of the involved standard deviations (square roots of the variances).

**Definition 4.1 (Correlation coefficient)** *For two random variables with finite second moment the dependence measure*

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \quad (4.1)$$

is called (Pearson’s) correlation coefficient.

The correlation coefficient is usually estimated by its empirical version: Given independent bivariate data  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  of joint observations from two random variables  $X$  and  $Y$ , respectively, the empirical correlation or correlation estimator is given by

$$\hat{\rho}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (4.2)$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

is the empirical mean of the  $X_i$  and  $\bar{Y}$  the empirical mean of the  $Y_i$ .

Classical results (the Cauchy-Schwarz inequality) ensure that the correlation of any two random variables has to be between  $-1$  and  $1$  (as well as for its empirical estimator), and for independent random variables  $\text{cov}(X, Y) = 0$  and, thus, the correlation  $\rho(X, Y) = 0$  as well.

The correlation is a measure of linear dependence. In particular, perfect linear dependence is equivalent to  $\rho(X, Y) = \pm 1$ .

**Theorem 4.2** *Two random variables  $X, Y$  are perfectly linearly dependent; i.e.,  $Y = aX + b$  with some  $a \neq 0$  and  $b \in \mathbb{R}$ , if and only if  $\rho(X, Y) = \pm 1$ .*

*Proof.* Assume first  $Y = aX + b$ . Then

$$\begin{aligned} \text{cov}(X, Y) &= E((X - E(X))(aX + b - (aE(X) + b))) = aE((X - E(X))^2) = a \text{var}(X) \\ \text{var}(Y) &= \text{var}(aX + b) = a^2 \text{var}(X) \end{aligned}$$

and, hence,  $\rho(X, Y) = a/\sqrt{a^2} = \pm 1$ , depending on the sign of  $a$ .

To ease notation for the converse implication we set  $\tilde{X} = X - E(X)$  and  $\tilde{Y} = Y - E(Y)$  in the following. Assume now that

$$\rho(X, Y) = \frac{E(\tilde{X}\tilde{Y})}{\sqrt{E(\tilde{X}^2)E(\tilde{Y}^2)}} = \pm 1.$$

Then  $E(\tilde{X}^2), E(\tilde{Y}^2) > 0$  and we have that

$$E(\tilde{Y}^2) \left( E(\tilde{X}^2)E(\tilde{Y}^2) - (E(\tilde{X}\tilde{Y}))^2 \right) = 0.$$

However, calculations show that

$$E(\tilde{Y}^2) \left( E(\tilde{X}^2)E(\tilde{Y}^2) - (E(\tilde{X}\tilde{Y}))^2 \right) = E \left( (E(\tilde{Y}^2)\tilde{X} - E(\tilde{X}\tilde{Y})\tilde{Y})^2 \right). \quad (4.3)$$

Since the expectation of a non-negative random variable is zero if and only if the random variable is zero (strictly speaking this has to hold only almost surely, but we ignore such technicalities), (4.3) implies that

$$Y - E(Y) = \frac{E(\tilde{Y}^2)}{E(\tilde{X}\tilde{Y})} (X - E(X))$$

and thus  $Y$  is of the form  $aX + b$  as claimed.  $\square$

**Proposition 4.3 (First properties of correlation)** *Let  $X$  and  $Y$  be two random variables.*

a) *Symmetry:*

$$\rho(X, Y) = \rho(Y, X)$$

b) *Effect of linear transformations:*

*For all  $\alpha, \gamma \neq 0$  and  $\beta, \delta \in \mathbb{R}$ ,*

$$\rho(\alpha X + \beta, \gamma Y + \delta) = \text{sign}(\alpha\gamma)\rho(X, Y),$$

*where  $\text{sign}(x)$  is equal to  $+1$  for  $x > 0$  and  $-1$  for  $x < 0$ . Hence, the correlation is invariant under strictly increasing linear transformations (the case when  $\alpha, \gamma > 0$ ).*

The concepts of covariance and correlation extend to multivariate random vectors as follows.

**Definition 4.4** *Let  $\mathbf{X} = (X_1, \dots, X_d)^\top$  be a  $d$ -dimensional and  $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$  an  $m$ -dimensional random vector. Then we can take covariances and correlations between every pair of components of  $\mathbf{X}$  and  $\mathbf{Y}$  and summarize them in  $d \times m$ -matrices, called the covariance matrix and the correlation matrix:*

$$\begin{aligned} \text{cov}(\mathbf{X}, \mathbf{Y}) &= (\text{cov}(X_i, Y_j))_{1 \leq i \leq d, 1 \leq j \leq m} \\ \text{corr}(\mathbf{X}, \mathbf{Y}) &= (\rho(X_i, Y_j))_{1 \leq i \leq d, 1 \leq j \leq m} \end{aligned}$$

*The covariance matrix of a random vector  $\text{cov}(\mathbf{X}, \mathbf{X})$  with itself is called the covariance matrix of  $\mathbf{X}$  and we write  $\text{var}(\mathbf{X}) := \text{cov}(\mathbf{X}, \mathbf{X})$ .*

**Proposition 4.5 (Further properties of correlations and covariances)**

*Let  $\mathbf{X} = (X_1, \dots, X_d)^\top$  be a  $d$ -dimensional and  $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$  an  $m$ -dimensional random vector.*

a) *Symmetry:*

$\text{var}(\mathbf{X})$  and  $\text{corr}(\mathbf{X}, \mathbf{X})$  are symmetric positive semi-definite matrices (cf. before Definition 3.2).

b) *Linear transformations:*

$$\text{cov}(\mathbf{AX} + a, \mathbf{BY} + b) = A \text{cov}(\mathbf{X}, \mathbf{Y}) B^\top$$

for every  $n \times d$  matrix  $A$ ,  $k \times m$  matrix  $B$  and every  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}^k$ .

c) *Linear combinations:*

For every  $a \in \mathbb{R}^d$  the variance of the linear combination  $a^\top \mathbf{X}$  is given by

$$\text{var}(a^\top \mathbf{X}) = a^\top \text{cov}(\mathbf{X}) a.$$

d) *Additivity:*

$$\text{cov}(\mathbf{X}, \mathbf{Y} + \mathbf{Z}) = \text{cov}(\mathbf{X}, \mathbf{Y}) + \text{cov}(\mathbf{X}, \mathbf{Z})$$

for every  $m$ -dimensional random vector  $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$ .

**Illustration 4.6** Suppose we model the water flow  $R$  (in litres per second) of a river at a certain point and assume that the river is formed by two independent rivers just a bit upstream. The water flow in the first river shall be  $R_1$  and the one in the second river  $R_2$ . Then  $\text{cov}(R_1, R_2) = \rho(R_1, R_2) = 0$  by the assumed independence. Clearly, it should hold that  $R = R_1 + R_2$  (assuming some kind of equilibrium state). Thus  $\text{cov}(R, R_1) = \text{cov}(R_1, R_1) + \text{cov}(R_2, R_1) = \text{var}(R_1)$  and, hence,

$$\rho(R, R_1) = \frac{\text{var}(R_1)}{\sqrt{\text{var}(R_1)\text{var}(R)}} = \frac{\text{var}(R_1)}{\sqrt{\text{var}(R_1)(\text{var}(R_1) + \text{var}(R_2))}} = \sqrt{\frac{\text{var}(R_1)}{(\text{var}(R_1) + \text{var}(R_2))}}$$

and, likewise, if we replace  $R_1$  by  $R_2$ . For example, if both original rivers; i.e.,  $R_1$  and  $R_2$ , have the same variance we get  $\rho(R, R_1) = 1/\sqrt{2}$ .  $\square$

**Illustration 4.7 (Danish Fire)** Throughout this paper we will illustrate the various dependence measures using a data set of Danish fire insurance claims from 1980 to 1990 available from <http://www.ma.hw.ac.uk/~mcneil/data.html>.

The original data set includes data on the losses of the fire insurance coming from the damages of the building, coming from the burnt content of the building, and from losses to profits (of companies in the burnt buildings). Since the last variable is zero in most cases, we consider only the losses of building and content. To avoid strange artefacts due to the fact that the data set considers only events where the total loss (sum of the loss in the three categories) exceeded one million Danish Kroner, we consider only events where both the losses in building and of content individually exceed this threshold.

In Figure 3 we provide a time series plot of the data.

To assess the dependence we provide scatter plots of the loss data as well as the logarithms of the losses in Figure 4. At the original scale it is hard to see what is going on in the majority of the observations, since they form a cloud at the origin and only the extreme events can be seen, for which it is hard to see any clear dependence structure. On the logarithmic scale one sees that there is no clear trend/dependence in the data, but that the two loss variables tend to behave similar and thus should be positively dependent. This can also be seen from the correlations which are 0.51 for the original data and 0.38 after taking logarithms.  $\square$

Correlation is a very popular dependence measure. The reasons are that it can be easily estimated from data by its empirical version, and that it is the natural dependence measure for the multivariate normal distribution. In this model it describes the dependence of the random components completely, and also in the more general class of elliptical distributions.

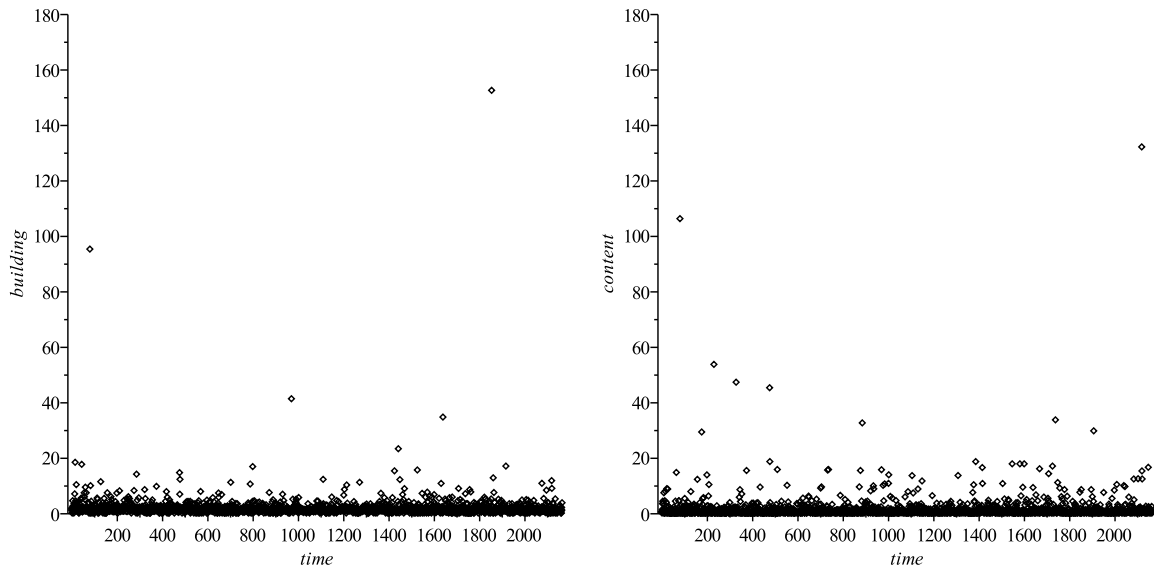


Figure 3: Time series plot of the losses in building (left) and the losses in content of the Danish fire insurance data from 1980 to 1990. The time is in days starting January 3rd, 1980, leaving out weekends and holidays.

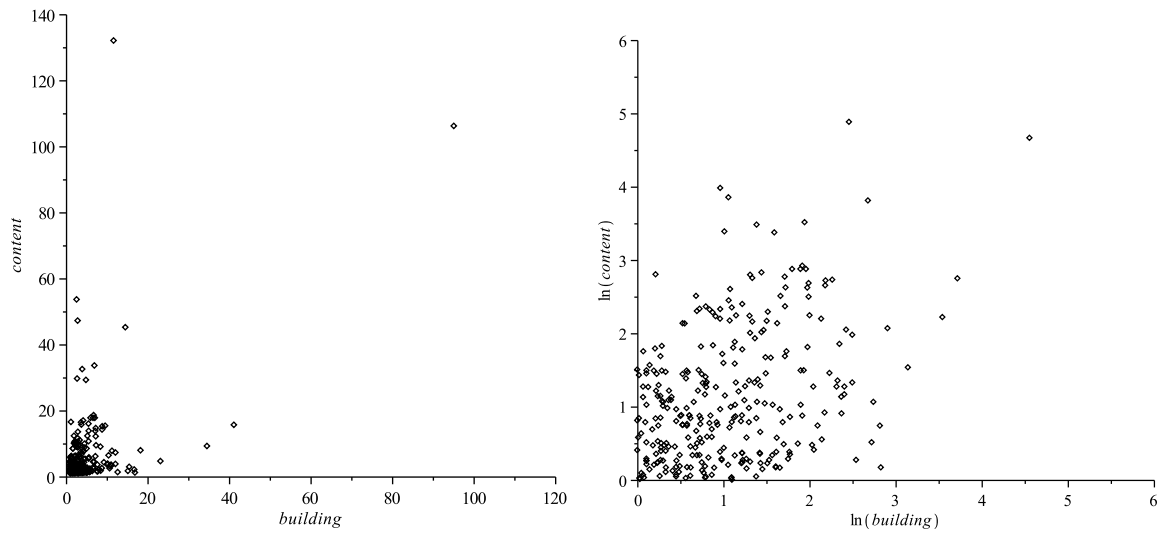


Figure 4: Scatter plot of the Danish fire insurance data: losses of buildings and losses of content, original scale (left) and logarithmic scale (right)

## 4.1 Disadvantages of correlation

Correlation has certain disadvantages that one should be aware of when using it.

- a) It is only defined when the variances of the random variables exist. In particular, for extreme risks this is not always guaranteed. A relevant example in the context of risk is the  $t$ -distribution with  $\nu$  degrees of freedom with density  $f(x) = c(1 + x^2/\nu)^{-(d+\nu)/2}$ ,  $x \in \mathbb{R}$ . For two  $t$ -distributed random variables with  $\nu \leq 2$  the correlation is not defined. Also for two Pareto-distributed random variables with densities  $f_1(x) = \alpha_1/x^{\alpha_1+1}$ ,  $x > 1$ , and  $f_2(x) = \alpha_2/x^{\alpha_2+1}$ ,  $x > 1$ , and shape parameters  $\alpha_1 \leq 2$  or  $\alpha_2 \leq 2$ , the correlation is not defined.
- b) Two independent random variables with finite variances are uncorrelated. However, the converse is not true. There exists an abundance of cases where random variables are uncorrelated, but not independent.

On a simple level, if  $X$  is a standard normal random variable, and  $Y = X^2$ , then  $X$  and  $Y$  are obviously not independent, since  $X^2$  is a function of  $X$ . However,  $\text{cov}(X, Y) = \text{cov}(X, X^2) = E(X^3) - E(X)E(X^2) = 0$ , since all odd moments of a normal random variable are equal to 0.

Examples on a more advanced level include variance mixtures of normal random variables (cf. Example 5.3) and, in a dynamic context, stochastic volatility models in finance and stochastic intermittency models in turbulent and other environmental data.

Only in special parametric models (the multivariate normal distribution is *the* typical example), uncorrelatedness implies independence.

- c) Covariances and correlations depend on the distribution in a highly non-trivial way. For instance, if one knows only the correlation of  $X, Y$ , then nothing can be said about the correlation of  $T(X), T(Y)$  for a non-linear increasing transformation  $T$ .
- d) The correlation depends on the whole distribution. However, in the context of risk one does not really care about the dependence for the “usual outcomes” but about the dependence of the extreme outcomes. The correlation thus typically provides at most very limited information about the dependence of risks.

## 4.2 Misconceptions of correlation

Unfortunately, there are several popular misunderstandings regarding correlation which we shall explain now.

### **Misconception 1: Marginals and correlation matrix determine the distribution:**

It is often wrongly thought that, if one knows the distributions of the random variables  $X_1$  and  $X_2$  and their correlation  $\rho(X_1, X_2)$ , then one knows already the bivariate distribution of the random vector  $\mathbf{X} = (X_1, X_2)^\top$ . This is not only false in general, but already in a normally distributed world. In particular, as we shall see in a moment, if  $X_1$  and  $X_2$  are known to be standard normally distributed with correlation  $\rho$  one cannot conclude that  $(X_1, X_2)^\top$  is bivariate normally distributed with mean zero and covariance matrix  $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ .

**Illustration 4.8** Let  $X_1$  be a standard normally distributed random variable and define  $X_2$  by

$$X_2 = \begin{cases} X_1 & \text{if } |X_1| \leq 1, \\ -X_1 & \text{if } |X_1| > 1. \end{cases}$$

Then  $X_2$  is also standard normally distributed, because  $X_1$  is and the standard normal distribution is symmetric around zero. Since both  $X_1$  and  $X_2$  have a finite variance,  $\rho := \rho(X_1, X_2)$  exists and is some number in  $(-1, 1)$ , which is hard to compute explicitly. Note that it is clear that the correlation is different from  $\pm 1$  because of Theorem 4.2. We now prove that the random vector  $(X_1, X_2)^T$  is not bivariate normally distributed by contradiction. Thus, assume  $(X_1, X_2)^T$  was bivariate normally distributed, then  $X_1 + X_2$  is also normally distributed with mean 0 and variance  $2 + 2\rho > 0$ . However, from the construction of  $X_2$  we see that

$$X_1 + X_2 = \begin{cases} 2X_1 & \text{if } |X_1| \leq 1, \\ 0 & \text{if } |X_1| > 1. \end{cases}$$

Thus, the probability that  $X_1 + X_2$  is strictly bigger than two in absolute value is zero. Since this probability is strictly positive for every normally distributed random variable, we have the desired contradiction. Hence, our assumption that  $(X_1, X_2)^T$  was bivariate normally distributed must be wrong.  $\square$

**Misconception 2: In all multivariate models it is possible to have all values between -1 and 1 as correlation:**

Likewise, the belief is widespread that in every multivariate model one may have all values between  $-1$  and  $1$  for the correlation. Unfortunately, not all combinations of valid pairwise correlations lead to a valid (i.e., positive semi-definite) overall correlation matrix.

However, this is not the only pitfall. Very often the model structure implies additional constraints on the correlation, like having to be non-negative. The following is an example for this.

**Illustration 4.9** Assume that an insurance company has sold insurance policies against damages by storm (S) and heavy rain (R). There are three types of insurance claims, those which regard damages by storm only, those which regard damages by heavy rain only, and those with both types of damages (caused e.g. by a thunderstorm with heavy rain and storm). We now want to model the number of claims regarding storm  $S(t)$  which arrived until time  $t$  (since the initial time 0), and the number of claims regarding heavy rain  $R(t)$  which arrived until time  $t$ .

The classical insurance claim number model is a Poisson process (e.g. Resnick [18]) for the arrivals of insurance claims. A Poisson process with rate (or frequency)  $\lambda > 0$  is a counting process, where the number of claims at any time  $t > 0$  is Poisson distributed with a mean linear in  $t$  with some rate  $\lambda > 0$ . We have  $E(X(t)) = \lambda t$  and  $\text{var}(X(t)) = \lambda t$  for all times  $t > 0$  for a Poisson process  $X$ . An alternative stochastic description of a Poisson process is as follows: it starts at zero at the initial time zero. After an exponentially distributed (with mean  $1/\lambda$ ) waiting time, during which it remains 0, it jumps to one. Afterwards it remains again constant for an exponentially distributed (with mean  $1/\lambda$ ) waiting time and then it jumps to two and so on. The rate  $\lambda$  gives the mean number of jumps (all of height one) in a unit time interval.

We use three independent Poisson processes,  $\{N^R(t)\}_{t \geq 0}$  giving the arrival of claims regarding only heavy rain,  $\{N^S(t)\}_{t \geq 0}$  giving the arrival of claims regarding only storm and  $\{N^B(t)\}_{t \geq 0}$  giving the arrival of claims regarding both. The corresponding rates shall be denoted  $\lambda^R$ ,  $\lambda^S$  and  $\lambda^B$ . Clearly, we have  $R(t) = N^R(t) + N^B(t)$  and  $S(t) = N^S(t) + N^B(t)$  for  $t \geq 0$  and we

want to understand the dependence of  $R(t)$  and  $S(t)$ . The process  $R(t)$  is (as a sum of Poisson processes) again a Poisson process with rate (or frequency)  $\lambda^S + \lambda^B$  and  $S(t)$  is one with rate  $\lambda^R + \lambda^B$ . Hence, for all  $t \geq 0$ , we have

$$\rho(R(t), S(t)) = \frac{\text{cov}(R(t), S(t))}{\sqrt{\text{var}(R(t))\text{var}(S(t))}} = \frac{\text{var}(N^B(t))}{\sqrt{\text{var}(R(t))\text{var}(S(t))}} = \frac{\lambda^B}{\sqrt{(\lambda^B + \lambda^R)(\lambda^B + \lambda^S)}}$$

In this model the correlation can only be between 0 and 1.

Assume further that we have already done univariate modelling of both  $R, S$  and obtained Poisson processes with rates  $\mu^R$  and  $\mu^S$  and then consider the joint model. We must then have that  $\lambda^B + \lambda^R = \mu^R$  and  $\lambda^B + \lambda^S = \mu^S$  to be consistent with the univariate models. Hence,  $\lambda^B \leq \min\{\mu^R, \mu^S\}$  is immediate interpreting the rates as the frequencies of the arrival of claims. Going back to our correlation we get for all  $t \geq 0$

$$\rho(R(t), S(t)) = \frac{\lambda^B}{\sqrt{\mu^R \mu^S}} \leq \min \left\{ \sqrt{\frac{\mu^R}{\mu^S}}, \sqrt{\frac{\mu^S}{\mu^R}} \right\}.$$

If  $\mu^R \neq \mu^S$ , then the possible correlations are thus below an upper bound strictly smaller than one. This result has been obtained in the framework of Operational Risk in Böcker and Klüppelberg [5], Equation 11.  $\square$

For more details on the problematic issues of correlation we refer to [1, 2].

## 5 Spherical and elliptical distributions

We have already seen that the contours of equal density are circles in the standard normal bivariate distribution and they are ellipses in the non-standard normal case. Likewise, one can show that in general dimensions the contours of equal density of the normal distribution are ellipsoids and spheres in the standard normal case (actually whenever all components are independent; i.e., all off-diagonal entries of the covariance matrix are zero, and have the same variance).

The spherical distributions extend the standard normal distribution  $N_d(\mathbf{0}, I_d)$  (i.e., the distribution of  $d$  independent standard normal components). The density of a spherical distribution satisfies

$$f(\mathbf{x}) = \psi(\mathbf{x}^\top \mathbf{x}), \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$$

where  $\psi : \mathbb{R} \rightarrow \mathbb{R}^+$  is an appropriate function.

Examples are the multivariate  $t$ -distribution with  $\nu$  degrees of freedom with density  $f(\mathbf{x}) = c(1 + \mathbf{x}^\top \mathbf{x}/\nu)^{-(d+\nu)/2}$  and the logistic distribution with density  $f(\mathbf{x}) = c \exp(-\mathbf{x}^\top \mathbf{x}) / (1 + \exp(-\mathbf{x}^\top \mathbf{x}))^2$ . Here  $c$  are the norming constants, which guarantee the densities to integrate to 1. It should be noted that random variables with a non-normal joint distribution that is spherical are uncorrelated random variables, which are not independent (see e.g. [13]).

There are various ways how to think about a spherical distribution.

- (i) From the densities above we see that the contours of equal density are circles in the bivariate models; i.e., “spheres” in arbitrary dimensions.
- (ii) Equivalently, we can think of a spherical random vector  $\mathbf{X}$  as having the same distribution under every orthogonal transformation; i.e., if we multiply it by a  $d \times d$  matrix  $M$  with the

property that  $M^\top M = MM^\top = I_d$ , then  $M\mathbf{X}$  has the same distribution as  $\mathbf{X}$ .

(iii) Finally, a spherical random vector  $\mathbf{X}$  has the same distribution as  $R\mathbf{U}$ , where  $\mathbf{U}$  is uniformly distributed on the unit sphere  $\mathcal{S}_{d-1} = \{\mathbf{s} \in \mathbb{R}^d : \mathbf{s}^\top \mathbf{s} = 1\}$ , and  $R$  is a positive random variable, independent of  $\mathbf{U}$ .

Elliptical distributions generalize multivariate normal distributions  $N_d(\boldsymbol{\mu}, \Sigma)$  with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ , and also have contours of equal density which are ellipsoids. Moreover, just as ellipsoids are linear transformations of spheres, elliptical distributions are obtained as linear transformations of spherical distributions.

For a general treatment of elliptical distributions we refer to Fang, Kotz, and Ng [4].

**Definition 5.1** *A random vector  $\mathbf{X} \in \mathbb{R}^d$  has an elliptical distribution if there exist  $\boldsymbol{\mu} \in \mathbb{R}^d$ , a positive semi-definite  $d \times d$  matrix  $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq d}$ , a positive random variable  $G$  and a random vector  $\mathbf{U}^{(d)} \sim \text{unif}\{\mathbf{s} \in \mathbb{R}^d : \mathbf{s}^\top \mathbf{s} = 1\}$  (i.e.,  $\mathbf{U}^{(d)}$  is uniformly distributed on the unit sphere in  $\mathbb{R}^d$ ) independent of  $G$  such that  $\mathbf{X}$  satisfies ( $\stackrel{d}{=}$  means that the distributions of the random variables on both sides are equal)*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + G\mathbf{A}\mathbf{U}^{(d)} \quad \text{with} \quad \mathbf{A} \in \mathbb{R}^{d \times d} \quad \text{and} \quad \mathbf{A}\mathbf{A}^\top = \Sigma. \quad (5.1)$$

We write  $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, G)$ .

The random variable  $G$  is called generating variable. Furthermore, if the first moment exists, then  $E(\mathbf{X}) = \boldsymbol{\mu}$ , and if the second moment exists, then  $G$  can be chosen such that  $\text{var}(\mathbf{X}) = \Sigma$ .

Note that we write  $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma)$  if we consider only quantities, which do not depend on the concrete generating random variable  $G$ , and we denote  $E(\mathbf{X}) = \boldsymbol{\mu}$ ,  $\text{var}(\mathbf{X}) = \Sigma$ , provided they exist.

Furthermore, note that in the following we always call  $\Sigma = \mathbf{A}\mathbf{A}^\top$  the covariance matrix (its elements the covariances) of an elliptical distribution even if the second moments do not exist.

In elliptical models covariances and correlations are natural dependence measures. This is a consequence of the following properties:

**Proposition 5.2 (Properties of elliptical distributions)** *Let  $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma)$  be elliptically distributed.*

- a) *Consider the map  $T(\mathbf{X}) = \mathbf{B}\mathbf{X} + \mathbf{b}$  for a  $q \times d$ -matrix  $\mathbf{B}$  and a vector  $\mathbf{b} \in \mathbb{R}^q$ . Then  $\mathbf{B}\mathbf{X} + \mathbf{b} \sim \mathcal{E}_q(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}^\top)$ .*
- b) *From this follows immediately that all marginal distributions of  $\mathbf{X}$  are elliptical; in particular, the components of  $\mathbf{X}$  are one-dimensional elliptical, which means they are symmetric around their means (or the median, if the mean does not exist).*

*Moreover, for an arbitrary component  $X_i$  there are  $a > 0, b \in \mathbb{R}$  such that  $X_i \stackrel{d}{=} aX_1 + b$ , where instead of  $X_1$  we could also have chosen any other component. Hence, in distribution any component can be realised as a linear transformation of one fixed component.*

*Let  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^\top \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma)$  with  $\mathbf{X}_1 \in \mathbb{R}^p$ ,  $\mathbf{X}_2 \in \mathbb{R}^q$  with  $p + q = d$ . Let  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top$  with  $\boldsymbol{\mu}_1 \in \mathbb{R}^p$ ,  $\boldsymbol{\mu}_2 \in \mathbb{R}^q$ , and  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ . Then*

$$\mathbf{X}_1 \sim \mathcal{E}_p(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \mathbf{X}_2 \sim \mathcal{E}_q(\boldsymbol{\mu}_2, \Sigma_{22}).$$

*Hence, subvectors of elliptically distributed random vectors are again elliptically distributed, and the parameters are known explicitly.*



b

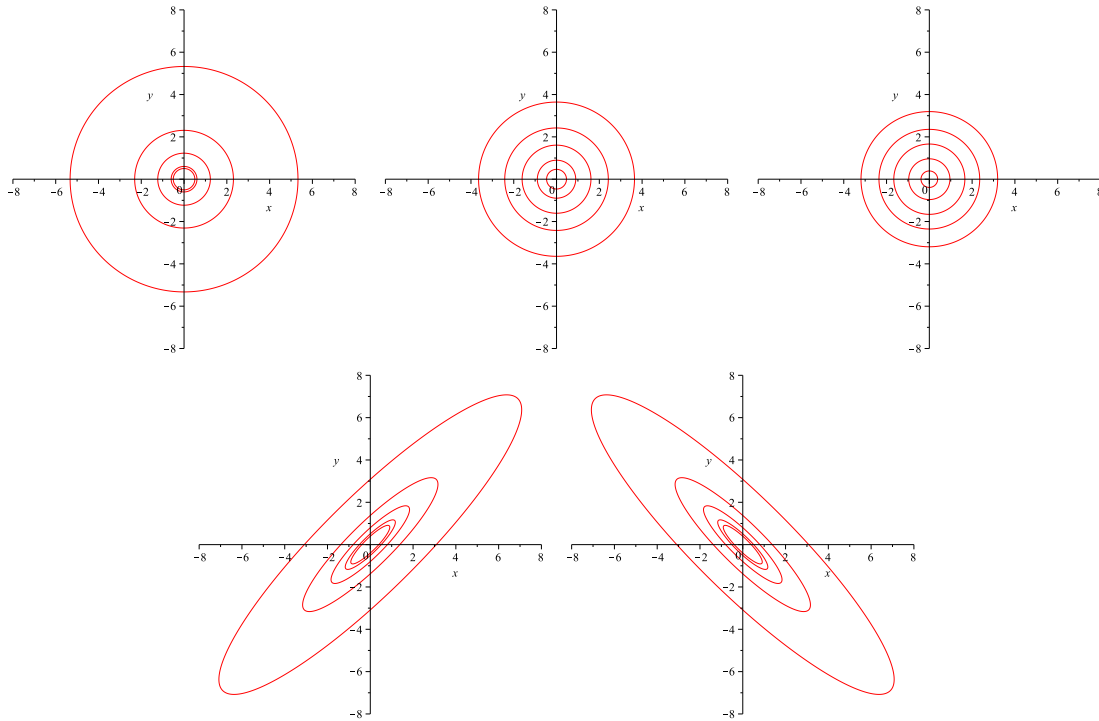


Figure 5: Contour plots of bivariate  $t_\nu$ -densities:

Upper row: uncorrelated components; i.e.,  $\Sigma$  is the identity matrix; different degrees of freedom:  $\nu = 1$  (left),  $\nu = 10$  (middle),  $\nu = 500$  (right).

Lower row ( $\nu = 1$ ): strongly correlated components, with  $\rho = 0.9$  (left), and  $\rho = -0.9$  (right).

The levels for the individual contour lines are the same as in Figure 2.

c) Assume that  $\Sigma$  is positive definite. The conditional distribution of  $\mathbf{X}_1$  given  $\mathbf{X}_2$  is also elliptical:

$$\mathbf{X}_1 \mid \mathbf{X}_2 \sim \mathcal{E}_p(\mu_{1|2}, \Sigma_{11|2}),$$

where  $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)$  and  $\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ .

d) Every elliptical distribution is uniquely determined by the mean, the covariance matrix  $\Sigma$ , and the distribution of the generating random variable  $G$ .

A very important class of elliptical distributions is given by the normal variance mixture models.

**Example 5.3** [Normal variance mixture model]

(a) Let  $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$  with  $\boldsymbol{\mu} \in \mathbb{R}^d$ ,  $\mathbf{A} \in \mathbb{R}^{d \times m}$  a matrix of rank  $d < m$ ,  $\mathbf{Z} \in \mathbb{R}^m$  a standard normal vector and  $W > 0$  a random variable, independent of  $\mathbf{Z}$ . Then  $\mathbf{X}$  is said to follow a normal variance mixture model, and one can show that the contours of equal density are ellipsoids, hence it is an elliptical distribution.

(b) In the situation of part a), if  $W$  has an inverse gamma distribution with parameters  $(\frac{\nu}{2}, \frac{\nu}{2})$ , then for  $\nu$  being an integer  $\nu/W \sim \chi_\nu^2$ , i.e.  $\nu/W$  is  $\chi^2$  (chi-square) distributed with  $\nu$  degrees of freedom. This implies that  $\frac{1}{d}(\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \frac{\nu \chi_d^2}{d \chi_\nu^2}$ , which is  $F(d, \nu)$ -distributed (recall that  $\Sigma = \mathbf{A}\mathbf{A}^\top$ ).

Moreover, we have  $\mathbf{X} - \boldsymbol{\mu} = \frac{AZ}{\sqrt{W}} \sim \mathbf{t}_\nu(0, \Sigma)$ ; i.e.,  $\mathbf{X} - \boldsymbol{\mu}$  is a  $d$ -dimensional  $t$ -distributed vector with  $\nu$  degrees of freedom. Further, if  $\nu > 2$ , then  $\mathbf{X} - \boldsymbol{\mu}$  has covariance matrix  $\frac{\nu}{\nu-2}\Sigma$ . If  $\nu \leq 2$  the covariance matrix does not exist.

Hence, the  $t$ -distribution - occurring frequently in statistics - is an example of a normal variance mixture. It is often used in risk management as an alternative to the normal distribution, because it puts more mass on large events (cf. Figure 5) and, in its multivariate version, it allows for modelling joint large events (cf. Example 8.5(c)).

Some contour plots for the densities of  $t$ -distributions can be found in Figure 5. As can be seen they are quite similar to the corresponding plots for the normal distribution in Figure 2, but especially for small  $\nu$  the density decays much slower than a normal density.  $\square$   $\square$

## 6 Rank correlations

Correlations depend on the underlying distribution, and may even not exist (when there is no finite second moment). Non-parametric and robust alternatives have been proposed, which are based only on the ranks of the observations. Here ranking refers to a data transformation, where numerical or ordinal values are replaced by their ranks. For instance, if numerical data 1.7, 9.3, 7.2 and 5.3 are observed, then the ranks of these data would be 1,4,3,2. The actual size of the data is completely ignored. Obviously, ranking is not unique, when data of equal value are observed. There is a simple way how to deal with these so-called *ties*, and we explain this with an example. Assume that we observe 1.7, 7.2, 9.3, 7.2 and 5.3 then we would take the mean rank for the two equal observations, and obtain ranks 1, 3.5, 5, 3.5, 2. Similarly, one deals with 3 or more equal values.

Often this situation is excluded from the beginning by requiring that the underlying distribution has a density. Then equal values do (with probability 1) not happen in a sample.

**Definition 6.1 (Spearman's rank correlation coefficient)** *Let  $X, Y$  be random variables with continuous distribution functions  $F_1, F_2$  and joint distribution function  $F$ . Let  $\rho$  be Pearson's correlation coefficient from Definition 4.1. Then Spearman's rank correlation is given by*

$$\rho_S(X, Y) = \rho(F_1(X), F_2(Y)).$$

We have to explain, why this is a rank correlation coefficient. Recall that for a distribution function  $F$  we denote by  $F^{-1}$  its generalized inverse function as defined in (1.1) and recall that  $F^{-1}$  is the analytic inverse of  $F$ , if  $F$  is strictly increasing.

First of all note that  $F_1(X)$  is a random variable with values in  $[0, 1]$ . Moreover, since  $F_1$  is continuous,  $P(F_1(X) \leq x) = P(X \leq F_1^{-1}(x)) = F_1(F_1^{-1}(x)) = x$  for  $x \in [0, 1]$ . This implies that  $F_1(X)$  is a standard uniform random variable (i.e., it is uniformly distributed on the interval  $[0, 1]$ ). Consequently,  $\rho_S$  measures the correlation between two uniform random variables, the original sizes of  $X$  and  $Y$  have become irrelevant.

One can say that rank correlations measure the degree of monotone dependence.

Let  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  be bivariate independent observations from two random variables  $X$  and  $Y$  respectively, such that all the values of  $(X_i)$  and  $(Y_i)$  are different (there are no ties).

We estimate Spearman's rank correlation coefficient by its empirical version, which is based on replacing  $F_1(X)$  and  $F_2(Y)$  by its empirical versions. To this end the data  $(X_1, Y_1), \dots, (X_n, Y_n)$

are converted into ranks, which we denote by  $(\text{rank}(X_i), \text{rank}(Y_i))$  and the empirical correlation coefficient as given in (4.2) is calculated for these ranks.

The formula simplifies by the fact that  $\frac{1}{n} \sum_{i=1}^n \text{rank}(X_i) = \frac{1}{n} \sum_{i=1}^n i = \frac{n+1}{2}$ , and

$$\sum_{i=1}^n \left( \text{rank}(X_i) - \frac{n+1}{2} \right)^2 = \sum_{i=1}^n \left( \text{rank}(Y_i) - \frac{n+1}{2} \right)^2 = \sum_{i=1}^n \left( i - \frac{n+1}{2} \right)^2 = \frac{1}{12} n(n^2 - 1).$$

Then the empirical Spearman's rank correlation coefficient is given by

$$\hat{\rho}_S(X, Y) = \frac{1}{2} n(n^2 - 1) \sum_{i=1}^n \left( \text{rank}(X_i) - \frac{n+1}{2} \right) \left( \text{rank}(Y_i) - \frac{n+1}{2} \right).$$

**Definition 6.2 (Kendall's rank correlation)** *Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  be two independent pairs of random vectors with bivariate distribution function  $F$ . Then Kendall's tau is given by*

$$\tau(X, Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0).$$

The dependence Kendall's tau captures is better understood in its empirical version. Let  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  be a sample of bivariate observations from two random variables  $X$  and  $Y$  respectively, such that all the values of  $(X_i)$  and  $(Y_i)$  are different, respectively. Any pair of observations  $(X_i, Y_i)$  and  $(X_j, Y_j)$  are said to be *concordant*, if the ranks for both elements agree: that is, if both  $X_i > X_j$  and  $Y_i > Y_j$  or if both  $X_i < X_j$  and  $Y_i < Y_j$ . They are said to be *discordant*, if  $X_i > X_j$  and  $Y_i < Y_j$  or if  $X_i < X_j$  and  $Y_i > Y_j$ .

**Definition 6.3 (Empirical Kendall's rank correlation coefficient)**

*The empirical version of Kendall's rank correlation is defined as:*

$$\begin{aligned} \hat{\tau} &= \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2} n(n-1)} \\ &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}((X_i - X_j)(Y_i - Y_j)). \end{aligned}$$

Note that the sign is equal to 1, whenever the two pairs are concordant, and it is  $-1$ , whenever the two pairs are discordant.

Rank correlation coefficients share some of the properties of Pearson's correlation coefficient: they are symmetric, lie between  $-1$  and  $1$ , and if  $X$  and  $Y$  are independent, they are equal to  $0$ . Moreover, since they are based on ranks, rank correlations are invariant with respect to increasing transformations; i.e., if  $T(x) \leq T(y)$  for all  $x < y$ , then  $\rho_S(T(X), T(Y)) = \rho_S(X, Y)$  and the same holds for Kendall's tau.

Both Kendall's  $\tau$  and Spearman's  $\rho$  can be calculated from the copula of a bivariate random vector with continuous marginal distributions (for a proof see Section 5.2.3 of McNeil, Frey, and Embrechts [8]); see next section for definitions and discussions of copulae. This means that both rank correlation coefficients are defined by the dependence structure only and not the marginal distributions. Intuitively, both dependence measures check whether the ranks are similar, but there are important differences in what they actually measure, which are rather technical and thus beyond the scope of this introductory chapter (see [10, 16]).

**Illustration 6.4 (Danish Fire continued)** In Figure 6 the ranks of the losses in building are plotted against the ranks of the losses of content. The fact that there are very few points at the lower right and upper left corner hints again at positive dependence. Indeed, we obtain for the empirical versions of Spearman's  $\rho$  the estimate  $\hat{\rho}_S = 0.32$  and of Kendall's  $\tau$  the estimate  $\hat{\tau} = 0.21$ .  $\square$

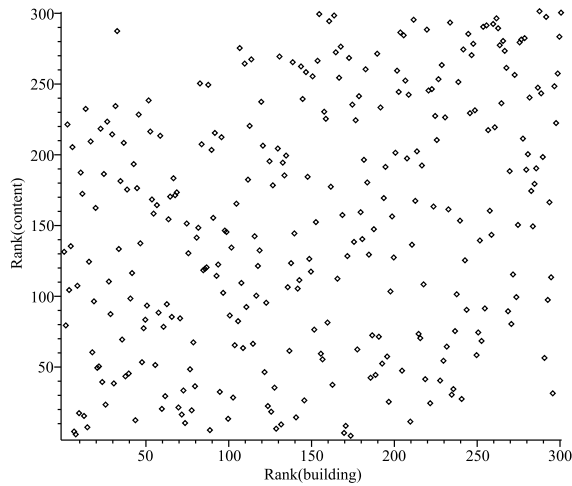


Figure 6: Scatter plot of the Danish fire insurance data losses in building and losses in content after conversion to ranks.

## 7 Copulae

The idea to model dependence in terms of ranks culminates in the concept of a copula. A copula describes completely the dependence structure and, thus, is in general a very complex object.

We start recalling that for a random variable  $X$  with continuous distribution function  $F$  (recall that then we have (with probability 1) no ties in the observations) the transformed random variable  $U := F(X)$  has a standard uniform distribution (i.e., is uniformly distributed on the interval  $[0, 1]$ ).

This concept is now extended to a multivariate distribution as follows. Let  $\mathbf{X} = (X_1, \dots, X_d)^\top$  be a random vector with distribution function  $F$ , and  $F_j$  denote the marginal distribution function of  $X_j$  for  $j = 1, \dots, d$ . If all  $F_j$  are continuous functions, then we can do the same transformation as above componentwise, which yields a random vector  $(F_1(X_1), \dots, F_d(X_d))^\top$  taking values only in the unit cube  $[0, 1]^d$ . Note that all components of this vector are standard uniform random variables. This motivates the following definition.

**Definition 7.1 (Copula)** *A copula is the joint distribution function of marginally uniformly distributed random variables. More precisely, if  $U_1, \dots, U_d$  are  $U(0, 1)$ , then the function  $C : [0, 1]^d \rightarrow [0, 1]$  defined by*

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d)$$

*is a copula.*

Applying this concept to the componentwise transformed random variables above, the vector  $(F_1(X_1), \dots, F_d(X_d))^\top$  has distribution function given by

$$C_F(u_1, \dots, u_d) = P(F_1(X_1) \leq u_1, \dots, F_d(X_d) \leq u_d)$$

for  $(u_1, \dots, u_d)^\top \in [0, 1]^d$ .  $C_F$  is the copula of the vector  $(X_1, \dots, X_d)^\top$ .

In the way we have defined/constructed a copula above, it covers only the continuous case. The case of non-continuous random variables can be covered as well, but this becomes much

more technical. A thorough introduction to copulae can be found in the book by Nelsen [9], for instance, or in [3, 8], which are of special interest in connection with risk modelling.

Before we discuss the use of copulae in risk analysis further, we present some examples. We formulate them for  $d = 2$ , and for most of the models it should be obvious, how they generalize to arbitrary dimension  $d$ .

**Example 7.2** [Bivariate copula families] Let  $u_1, u_2 \in [0, 1]^2$ .

(a) Independence copula:

$$C^{ind}(u_1, u_2) = u_1 u_2.$$

As the name already suggests, this is the copula of two independent random variables. Recall that two random variables (with continuous distributions) are independent if and only if their joint distribution function is the product of the marginals. This is inherited by the copula.

(b) Copula of perfect dependence:

$$C^{dep}(u_1, u_2) = \min(u_1, u_2).$$

This copula models the situation, when the observations are perfectly dependent. For the two uniform random variables corresponding to the copula this means that they are identical. In general two random variables  $X, Y$  have the copula of perfect dependence if and only if there exists a random variable  $Z$  and two increasing functions  $f$  and  $g$  such that  $X = f(Z)$  and  $Y = g(Z)$ . Note that intuitively this means that as soon as you know the value of one variable you also know the value of the other random variable for sure.

(c) Normal copula: for  $\theta \in (-1, 1)$ ,

$$\begin{aligned} C^{No}(u_1, u_2; \theta) &= \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \frac{1}{2\pi\sqrt{1-\theta^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left(\frac{-(x_1^2 - 2\theta x_1 x_2 + x_2^2)}{2(1-\theta^2)}\right) dx_1 dx_2, \end{aligned}$$

where  $\Phi_2$  and  $\Phi$  denote the distribution functions of the bivariate and the univariate standard normal distribution, respectively, and  $\Phi^{-1}$  is the inverse function of the cumulative standard normal distribution function  $\Phi$ .

Again the name already tells the idea behind this copula. It is the copula of two standard normally distributed random variables with correlation  $\theta$  which are also jointly normally distributed. A sample from this copula can easily be obtained by drawing from a bivariate standard normal distribution with correlation  $\theta$  and then applying the function  $\Phi^{-1}$  to every coordinate.

Note that  $\theta = 0$  gives the independence copula, whereas  $\theta = 1$  gives the one of perfect dependence. For  $\theta = -1$  one obtains perfect negative dependence (i.e., the copula  $\max(u_1 + u_2 - 1, 0)$  which, in contrast to the other examples, is a copula only for dimension  $d = 2$ ).  $\square$

As mentioned before and explained in more detail and by examples in Fasen et al. [9], extreme value models are important for risk management. When considering copula models in the context of bivariate extreme value models, then so-called *extreme value copulae* occur. These copulae have to be of a very special form; i.e., their dependence structure can be represented in terms of a so-called Pickands dependence function  $A$ , a convex function satisfying  $\max(s, 1-s) \leq A(s) \leq 1$  for all  $s \in [0, 1]$ ; see e.g., Beirlant, Goegebeur, Segers, and Teugels [3], Chapter 8.2.5. In terms of such a Pickands dependence function an extreme value copula  $C$  has the form

$$C(u_1, u_2) = \exp\left\{\ln(u_1 u_2) A\left(\frac{\ln(u_2)}{\ln(u_1 u_2)}\right)\right\}. \quad (7.1)$$

Note that the right hand side is equal to  $u_1 u_2$  for the Pickands dependence function  $A \equiv 1$ ; this is the independent case. A quantity often considered and estimated is the value in (7.1) for  $u_1 = u_2$ ; i.e.  $A(\frac{1}{2})$ . For symmetric copulae it is the minimum of  $A$ , hence gives a measure of maximal dependence in the model. We come back to this in Section 8.

**Example 7.3** [Extreme value copulae and their Pickands dependence function]  
Throughout  $u_1, u_2 \in [0, 1]^2$  and  $s \in [0, 1]$ .

(a) Gumbel copula:

Using the Pickands dependence function the Gumbel copula with parameter  $\theta \in [1, \infty)$  is given by

$$A^{Gu}(s) = \left( s^\theta + (1-s)^\theta \right)^{1/\theta}.$$

Elementary calculations show that the Gumbel copula is thus

$$C^{Gu}(u_1, u_2) = \exp \left\{ - \left( (-\ln(u_1))^\theta + (-\ln(u_2))^\theta \right)^{1/\theta} \right\}. \quad (7.2)$$

For  $\theta = 1$  the Gumbel copula is actually the independence copula, whereas for  $\theta \rightarrow \infty$  the Gumbel copula converges to the copula of perfect dependence. Thus the Gumbel copula allows modelling a continuum of possible dependencies from independence to perfect positive dependence giving a nice parametric model for different dependence scenarios.

(b)  $t$ -EV copula:

Using the Pickands dependence function the  $t$ -EV copula with parameter  $\boldsymbol{\theta} = (\theta_1, \theta_2) \in (0, \infty) \times (-1, 1)$  is given by

$$A^{t-EV}(s; \boldsymbol{\theta}) = s t_{\theta_1+1} \left( \frac{\left( \frac{s}{1-s} \right)^{1/\theta_1} - \theta_2}{\sqrt{1-\theta_2^2}} \sqrt{\theta_1+1} \right) + (1-s) t_{\theta_1+1} \left( \frac{\left( \frac{1-s}{s} \right)^{1/\theta_1} - \theta_2}{\sqrt{1-\theta_2^2}} \sqrt{\theta_1+1} \right),$$

with  $t_\nu$  for  $\nu \in (0, \infty)$  representing the distribution function of the  $t_\nu$ -distribution (i.e., the  $t$ -distribution with  $\nu$  degrees of freedom). The  $t$ -EV copula (with ‘‘EV’’ standing for ‘‘extreme value’’) arises as the limiting dependence structure of componentwise maxima of independent and identically distributed bivariate  $t_{\theta_1}$ -distributed random variables with the correlation of the underlying bivariate normal distribution being  $\theta_2$ . For more details see e.g. [8].  $\square$   $\square$

Statistically, parametric copulae are rather easy to fit, since it is not necessary to specify marginal models. One can simply take the empirical distribution functions, plug them into a parametric copula model and estimate the copula parameters for instance by likelihood methods. Various copula models are presented in Haug, Klüppelberg, and Peng [5], where also R codes for fitting such copula models are provided. The problem is obviously the choice of the parametric model.

Abstractly speaking a copula encodes the dependence structure of a  $d$ -dimensional random vector by transforming it to a  $d$ -dimensional random vector with standard uniform margins. In principle, one could just as well transform it to any other  $d$ -dimensional random vector with prescribed marginals to encode the dependence structure. So the question arises whether the use of a copula is the best way to transform data. Alternative transformations are indeed used in relation to some special applications. For instance, in reliability theory marginals have been transformed to normal random variables, which is admittedly not as easy as the transformation to uniform, since the normal distribution function is given as an integral, which cannot be calculated explicitly. See [11], [15] and [10] for details.

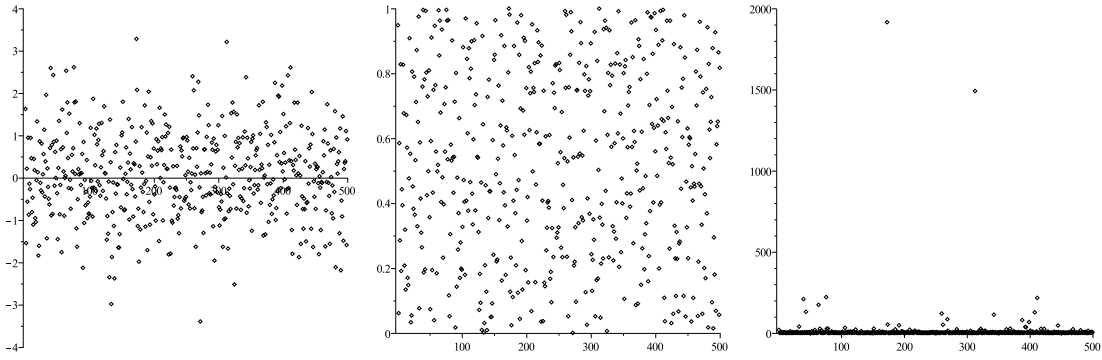


Figure 7: Simulation of 500 independent and identically distributed standard normally distributed random variables (left) and their transformations to standard uniform (middle) and standard Fréchet random variables.

Experts from extreme value theory often normalize marginals to standard extreme value distributions, when interested in the maximum of a sample. Typically the standard Fréchet distribution is used (see e.g. Proposition 5.10 of [18] for more details). Here the transformation is given by  $-1/\ln(F(X))$ , which has distribution function  $P(-1/\ln(F(X)) \leq z) = \exp\{-1/z\}1_{[0,\infty)}(z)$ . When interested in the minimum of a sample, then often the transformation is to the standard exponential distribution given by  $F(x) = (1 - e^{-x})1_{[0,\infty)}(x)$  (i.e., the transformation is  $-\ln(1 - F(X))$ ); cf. e.g. [12] for multivariate exponential distributions.

A Taylor expansion to the standard Fréchet distribution function gives  $P(-1/\ln(F(X)) > z) \sim 1/z$  (equivalently,  $zP(-1/\ln(F(X)) > z) \rightarrow 1$ ) as  $z \rightarrow \infty$ , so that large values of  $z$  happen with substantial probability (in particular compared to the normal distribution where  $P(N(0,1) > z) \sim z\phi(z) = (\sqrt{2\pi}z)^{-1} \exp\{-z^2/2\}$  as  $z \rightarrow \infty$  ( $\phi$  denotes the standard normal density)). Taking  $z = 10$ , one obtains for the Fréchet distribution the probability 0.09516258 and for the standard normal distribution  $7.619853 \cdot 10^{-24}$ . For the uniform distribution, no value larger than 1 can happen (with probability 1). As you can see in Figure 7, it may be advantageous to transform data to Fréchet marginals when interested in the dependence structure of extreme events, as then the extremes really stick out.

**Illustration 7.4** Because of the simple transformation in the marginals the use of copulae to model dependence had a striking success in particular in the financial industry. The mostly applied copula has been the normal copula which means that in the end all dependence is like in a multivariate Gaussian situation and completely described by the correlation matrix of the underlying multivariate Gaussian random variable.

For example, this model was used as a model for the probability of joint defaults - the probability that any two members (say A and B) of a pool of credits will both default within the next year or some other pre-specified period (i.e., the credit taker fails to pay the interest or the credit notional amount back). Denoting by  $T_A$  the time when A defaults and likewise by  $T_B$  that B defaults, this model describes the probability that both credits will default as

$$P(T_A < 1, T_B < 1) = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)); \rho),$$

where  $F_A$  and  $F_B$  are the marginal distribution functions of the default times and  $\rho$  the correlation of the used normal copula.

This model, suggested in [14], was heavily blamed (and obviously before the subprime crisis heavily used) in a by now famous article from 2009 (still to be found on the internet at <http://>

[www.wired.com/techbiz/it/magazine/17-03/wp\\_quant](http://www.wired.com/techbiz/it/magazine/17-03/wp_quant)) entitled “The Formula That Killed Wall Street”. The reason is that in a bivariate (and likewise in higher dimensional) normal model with the correlation different from 1 the probability that both variables  $X$  and  $Y$  are very big at the same time is extremely small: asymptotically for  $z \rightarrow \infty$  the events that  $X > z$  and  $Y > z$  become independent. Now it turned out during the subprime crises that the dependence between different credits is much higher. In the US American subprime credit market it became obvious that many more creditors could not serve their credits than the credit models predicted to be likely. The problem was that these credits had been pooled by the issuing banks and - sliced up into packets - sold to investors all over the world; the prices agreed upon in these sales were based usually on the above model (as were the triple-A ratings of some of these products by all rating agencies). Additionally, many derivatives based upon them - credit default swaps or credit default options were originally designed as insurance against defaults - were traded and very often they were bought or sold not to insure oneself, but for purely speculative reasons. So when many credits started to default, financial institutions all over the world had to realize that their assets were worth much less than what they had thought, which implied tremendous losses in particular for the financial industry. An interesting paper on how to model these risks more realistically is [7].  $\square$

Consequently, the financial crisis of the last years is a clear warning that one should not use models without basic knowledge of what they can model and what they cannot. Model risk is abundant and needs a critical mind concerning the application of various models and the interpretation of their resulting outcome, when applied with care. In the above normal copula model dependence is modelled by the correlation of the underlying normal distribution. It has long been known that a normal copula is by no means a model that captures dependent risks: in a normal copula model very high risks are always independent (see Example 8.5).

As we have seen in Section 5 the elliptical distributions are natural extensions of multivariate normal distributions and are also characterised mainly by their mean and covariance structure, only that additionally a positive generating random variable comes into play. Likewise, we can extend the normal copula to an elliptical copula by using the copula corresponding to a general elliptical distribution.

**Definition 7.5 (Elliptical copula)**

We define an elliptical copula as the copula of  $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, G)$  and write  $\mathcal{EC}_d(R, G)$  for short, where  $R$  is the correlation matrix of the elliptical distribution and  $G$  the generating random variable.

The notation  $\mathcal{EC}_d(R, G)$  for an elliptical copula makes sense, since it is characterized by the generating variable  $G$  (unique up to a multiplicative constant) and the copula correlation matrix  $R$ . This follows as a simple consequence of the definition and the fact that copulae are invariant under strictly increasing transformations.

**Example 7.6** (a) Let  $\mathbf{Z}$  be a  $d$ -dimensional mean  $\mathbf{0}$  normal vector with arbitrary covariance matrix  $\Sigma$ , and denote by  $\Phi$  the one-dimensional standard normal distribution function, then the distribution of  $(\Phi(Z_1), \dots, \Phi(Z_d))$  is a Gaussian copula.

(b) Let  $\mathbf{X} \sim \sqrt{\nu} \frac{\mathbf{Z}}{\sqrt{W}}$  with  $W$  being a  $\chi^2$ -distributed random variable with  $\nu$  degrees of freedom and  $\mathbf{Z}$  a  $d$ -dimensional mean  $\mathbf{0}$  normal vector with arbitrary covariance matrix  $\Sigma$ . So  $\mathbf{X}$  follows a  $d$ -dimensional  $t$ -distribution with  $\nu$  degrees of freedom and we write  $\mathbf{X} \stackrel{d}{=} \mathbf{t}_\nu(\mathbf{0}, \Sigma)$ ; i.e.,  $\mathbf{X}$  is distributed as in Example 5.3 b). Denoting by  $t_\nu$  the one-dimensional  $t$ -distribution with  $\nu$  degrees of freedom, then the distribution of  $(t_\nu(X_1), \dots, t_\nu(X_d))$  is the corresponding copula, which we call a  $\mathbf{t}_\nu$ -copula.  $\square$   $\square$



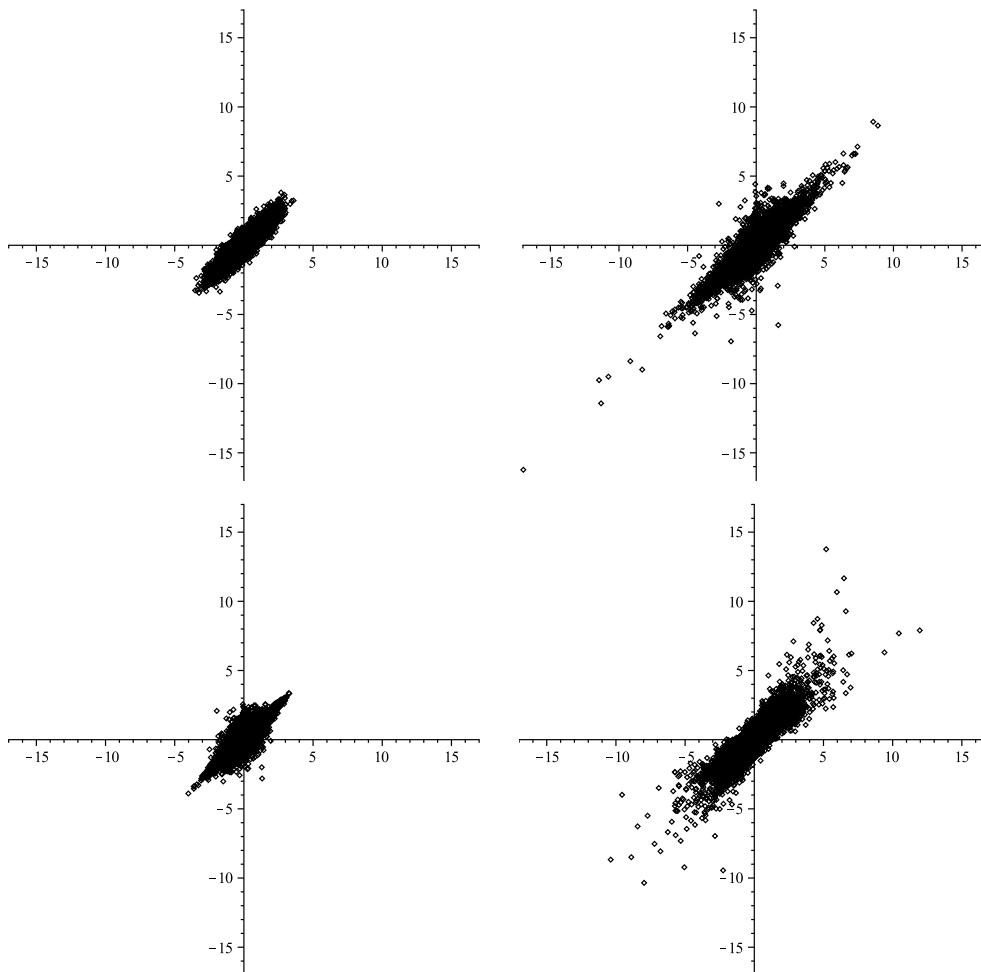


Figure 8: Upper row: simulation of 10 000 bivariate normally distributed random variables (left) and bivariate  $t_4$ -distributed random variables (right)  
 Lower row: simulation of 10 000 bivariate random variables with normal marginal distributions and a  $t_4$ -copula (left), and with  $t_4$  marginal distributions and a normal copula (right).  
 In all cases the correlation parameter was  $\rho = 0.9$ .

In Figure 8 we show the differences between the normal distribution, the  $t_4$ -distribution, and in Figure 9 their copulae. Comparing the figures in the left column we see that, for the same normal margins, the dependence structure given by the  $t_4$  copula yields more data in the left lower and right upper corners. The right column shows first that  $t$ -margins are heavier tailed than normal margins. Furthermore, for the  $t_4$ -distribution we see more data in the left lower and right upper corners than for the normal copula. Moreover, for the  $t_4$ -copula data spread out more in direction of the right lower and left upper corners than for the normal copula.

**Illustration 7.7 (Danish Fire continued)** In Figure 6 the ranks of the losses in building are plotted against the ranks of the losses of content. Up to a normalization this is a plot of the copula (the data transformed to uniform margins as in Figure 9). As we already said, the fact that there are very few points at the lower right and upper left corner hints again at positive dependence.  $\square$

**Illustration 7.8 (Engineering risk analysis)** Engineers often deal with complex systems with a large number of components. Suppose such a system consists of  $d$  components. As

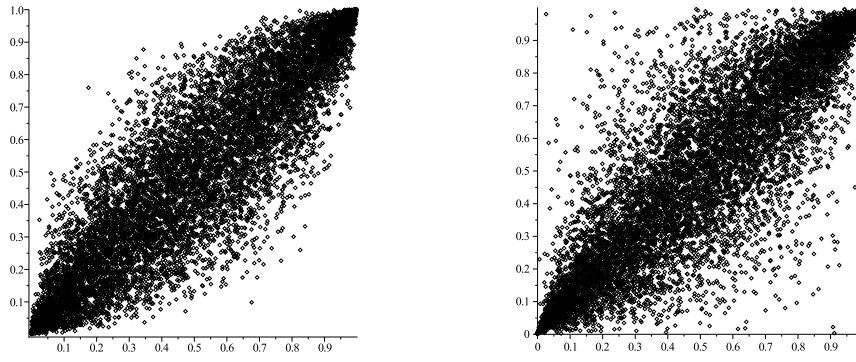


Figure 9: The copulae corresponding to Figure 8; i.e., normal copula (left) and  $t_4$ -copula (right).

the consequence of a risky event  $Y$  (e.g. an accident, an earthquake, a tsunami, a hurricane or a cyber attack) each component can be damaged. Typically the degree of damage will be different for every component.

A realisation  $y$  of  $Y$  would give the strength of such an events above. The damage done to component  $n$  is measured by a random variable  $X_n$  for  $n = 1, \dots, d$  which gives the costs of repairing or replacing when necessary the component. Assume that all damage variables  $X_n$  have continuous distribution functions  $F_n$  with densities  $f_{X_n}$  for  $n = 1, \dots, d$ , and that together with  $Y$  they have a joint density  $f_{X_1, \dots, X_d, Y}$ . Depending on the realised damage due to the risk event  $Y$ , summarized in the vector  $(x_1, \dots, x_d)$ , the monetary amount  $K(x_1, \dots, x_d)$  is needed to repair the system; some components would have to be repaired, some to be replaced. Note that  $K$  could simply be the sum of the  $x_n$ , but we allow for more general functions, since cost reductions or increases occur when you have to repair/replace several components.

In engineering, risk is often calculated as expected costs due to possible damages. We calculate the expected costs for repairing the system as

$$\begin{aligned} E(K) &= \int_0^\infty \dots \int_0^\infty K(x_1, \dots, x_d) f_{X_1, \dots, X_d}(x_1, \dots, x_d) dx_1 \dots dx_d \\ &= \int_0^\infty \left( \int_0^\infty \dots \int_0^\infty K(x_1, \dots, x_d) f_{X_1, \dots, X_d|Y}(x_1, \dots, x_d | y) dx_1 \dots dx_d \right) f_Y(y) dy, \end{aligned}$$

where  $f_Y$  is the density of the risky event variable  $Y$  and  $f_{X_1, \dots, X_d|Y}$  the joint density of the damages to the individual components given the risky event  $Y$ . From this calculation we see immediately that we need a model for the random vector taking the dependence structure between the damages to the different components  $(X_1, \dots, X_d) | Y$  into account.

The dependence structure of  $(X_1, \dots, X_d) | Y$  can be described via a copula. An unrealistic but simple scenario is the independence copula (i.e. we assume that the damages to the individual components are independent given  $Y$ ). If additionally  $K$  is simply the sum of the individual damages, we obtain:

$$E(K) = \int_0^\infty \left( \sum_{n=1}^d \int_0^\infty x_n f_{X_n|Y}(x_n | y) dx_n \right) f_Y(y) dy,$$

where  $f_{X_n|Y}$  is the conditional density of the damage in component  $n$  given the risky event  $Y$ .

Clearly these assumptions will be too simple in most real life applications, because the damages to the individual components are most likely dependent given  $Y$  or the costs of repairing the system are not the sum of the costs of repairing/replacing the individual components.  $\square$

**Remark 7.9** (a) Whereas bivariate copula models are well-known in great detail, higher dimensional models are usually hard to analyse and to fit to real data, not at least due to numerical problems when optimizing the likelihood function. Exceptions are the normal and  $t$ -copula models. A fairly new approach opens up the way to copulae of arbitrary dimension; cf. [1] and the book [7].

(c) In general the usage of copulae seems rather demanding at first and most statistical software does not include functions to handle copulae in their basic distributions. However, for many statistical programmes there are very well implemented and documented extensions available which make the use of copulae rather easy in applications. For example, for the programme R there are the packages `copula` and `fCopulae` available at <http://cran.r-project.org/web/packages/>. They include e.g. functions to handle Archimedean, elliptical and extreme value copulae.

(b) For a parsimonious model with respect to parameters, dimension reduction is an important first step. There exist many well-known methods (as e.g. principal component analysis) in classical multivariate statistics. Therefore, the use of copulae often needs to be combined with such methods. Dimension-reduction methods based on elliptical copula models have been suggested in Klüppelberg and Kuhn [6].

## 8 Extremal dependence measures

As explained in Fasen et al. [9] extremal risks can be modelled and estimated in a stochastic framework. In contrast to the book chapter [9], in the present chapter we are concerned about joint extreme risks, which can be particularly dangerous. Hence, it is of utmost importance to model and assess the joint occurrences of extreme events correctly. In other words it is not important to get the dependence of the “typical” observations right, but one must get the dependence of the extreme events right. One of the first questions for a statistical model is then, if it is likely to model joint extreme events.

In this section we briefly present models and methods to allow for a realistic assessment of the dependence of extremal events. An interesting collection of theoretical results and case studies for further reading is Reiss and Thomas [17]. Another very accessible book on extreme value statistics is Coles [6]; more advanced is Beirlant et al. [3].

One way to consider the question whether extremal events are dependent or not, is by asking the question, what is the probability that a random variable  $Y$  assumes a large value given that we already know that another random variable  $X$  takes a large value. Consequently, one natural way to model extremal dependence is to consider the asymptotic behaviour of the probability that  $Y > z$  given that  $X > z$  as  $z \rightarrow \infty$ . If  $X$  and  $Y$  are independent, we have that  $P(Y > z | X > z) = P(Y > z) \rightarrow 0$  as  $z \rightarrow \infty$ . Thus we call any pair  $X, Y$  of random variables with  $P(Y > z | X > z) \rightarrow 0$  as  $z \rightarrow \infty$  *tail independent*. Intuitively this means that extreme events typically occur only in one variable, provided they occur. In contrast to this, we speak of *tail dependence* whenever the limit is non-zero which implies that with a positive probability extreme events occur in both random variables at the same time. It turns out that this intuitive approach only makes sense when  $X, Y$  have the same distribution (or at least distributions with comparable tails). To account for this, one normalises the tails first using the same trick as we know already from the copulae. To be precise one defines tail dependence coefficients (for the upper tail) as follows. Again we invoke the quantile function from (1.1).

**Definition 8.1 (Tail dependence coefficients)** *Let  $X, Y$  be two random variables with continuous distribution functions  $F_X$  and  $F_Y$ . The upper tail dependence coefficient of  $(X, Y)$  is*

defined by

$$\lambda_U = \lim_{\alpha \uparrow 1} P(F_Y(Y) > \alpha \mid F_X(X) > \alpha) = \lim_{\alpha \uparrow 1} P(Y > F_Y^{-1}(\alpha) \mid X > F_X^{-1}(\alpha)),$$

provided the limit exists ( $\alpha \uparrow 1$  stands for taking the limit for  $\alpha$  going to 1 from below). If  $\lambda_U \in (0, 1]$ , then  $X$  and  $Y$  are called upper tail dependent. If  $\lambda_U = 0$ , they are called upper tail independent.

**Remark 8.2** (i) The assumption of continuous distributions is not really necessary, if one restricts the definition to the right limit.

(ii) Noting that  $P(F_Y(Y) > 1 - t \mid F_X(X) > 1 - t) = \frac{P(F_Y(Y) > 1 - t, F_X(X) > 1 - t)}{P(F_X(X) > 1 - t)}$  and  $P(F_X(X) > 1 - t) = t$ , we obtain the equivalent definition

$$\lambda_U = \lim_{t \rightarrow 0} t^{-1} P(F_X(X) > 1 - t, F_Y(Y) > 1 - t).$$

(iii) The link to the Value-at-Risk as defined in Definition 1.1(b) is obvious:

$$\lambda_U = \lim_{\alpha \uparrow 1} P(Y > \text{VaR}_\alpha(Y) \mid X > \text{VaR}_\alpha(X)).$$

One can show that the tail dependence is a copula property; i.e., the marginal distributions have no effect on the value of  $\lambda_U$ .

**Theorem 8.3** If  $X, Y$  have copula  $C$ , then

$$\lambda_U = \lim_{\alpha \uparrow 1} \frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha}. \quad (8.1)$$

**Remark 8.4** Theorem 8.3 provides a useful link of  $\lambda_U$  to the Pickands dependence function:

$$\frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha} = \frac{1 - 2\alpha + \exp(2 \ln(\alpha) A(\frac{1}{2}))}{1 - \alpha} = 2 \left( 1 - A\left(\frac{1}{2}\right) \frac{-\ln \alpha + o(\ln \alpha)}{1 - \alpha} \right)$$

by a Taylor expansion of the exponential function around 0. Using l'Hospital's rule we calculate

$$\lim_{\alpha \uparrow 1} \frac{-\ln \alpha + o(\ln \alpha)}{1 - \alpha} = \lim_{\alpha \uparrow 1} \frac{\frac{1}{\alpha}(1 + o(1))}{\alpha} = 1$$

giving

$$\lambda_U = 2 \left( 1 - A\left(\frac{1}{2}\right) \right). \quad (8.2)$$

For each copula model we can determine if it allows for tail dependence or not.

**Example 8.5** (a) Since by Proposition 5.2(c) the conditional distribution of a bivariate Gaussian random vector is normal ( $\mathcal{E}_p$  is in this case, of course, the normal distribution), the Gaussian copula (or Gaussian distribution) with correlation  $\rho < 1$  has

$$\lambda_U = 2 \lim_{x \rightarrow \infty} \left( 1 - \Phi\left(\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}x\right) \right) = 0.$$

Hence, when using a Gaussian copula one always has tail independence unless one considers the degenerate situation where  $\rho = 1$ . Therefore, one must never use the Gaussian copula when

one wants to model phenomena where extreme events occur jointly in different variables. The financial industry has learnt this the hard way (see Illustration 7.4).

(b) For a Gumbel copula (8.2) gives  $\lambda_U = 2 - 2^{1/\theta}$ . Hence, whenever  $\theta > 1$  we have a positive tail dependence and the tail dependence coefficient can assume any value in  $(0, 1)$ .

(c) For the bivariate  $t_\nu$ -copula with  $\nu$  degrees of freedom and correlation  $\rho \in [-1, 1]$  one calculates using again (8.2)

$$\lambda_U = 2 \left( 1 - t_{\nu+1} \left( \frac{\sqrt{\nu+1}\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \right)$$

with  $t_{\nu+1}$  being the distribution function of a  $t$ -distributed random variable with  $\nu + 1$  degrees of freedom. This implies that for every  $\rho > -1$  the upper tail dependence coefficient  $\lambda_U > 0$ ; i.e., that even for negative correlation it is far more likely to have both variables large at the same time than in the Gaussian copula.  $\square$

**Illustration 8.6 (Danish Fire continued)** Estimation of the tail dependence coefficient is rather tricky, since it is an asymptotic property (an asymptotic conditional probability). Therefore, it may depend rather strongly on the choice of the threshold approximating this asymptotic. We refer to Haug et al. [5] for a detailed analysis of these issues. For the Danish fire data [5] reports a value of  $\hat{\lambda}_U = 0.416$  for the tail dependence coefficient between the losses in building and content. Therefore, there is a non-negligible tail dependence and thus an insurance company needs to be prepared to meet large losses in its fire insurance for buildings and its insurance for the contents at the same time. Of course, intuitively this is not surprising.

To sum up our simple data example using the fire insurance data, we see that all dependence measures in this context report a positive dependence. But they focus on different aspects and thus the most adequate one should be used in any particular application. In particular, you should be aware that using correlations a simple order preserving transformation like taking logarithms may have a big impact, whereas this has no effect, if the dependence measure only depends on the ranks (or the copula).  $\square$

## 9 Food for Thought

We list some questions, which should be seriously considered for every real risk problem at hand.

- Is my risk problem multivariate? What are the risk factors involved?
- Which techniques do I use to model dependence? Does risk occur from the usual data around the mean or rather from extreme events? Is it important to get the bulk of the data right or the extremes? Should I use all data or only extreme values for a statistical analysis?
- What model should I use? What does the model I use assume about the dependence structure?
- How will I deal with the model risk?
- How sensitive are the outcomes of my research to assumptions about dependence? Should I apply several models and check robustness of the outcomes by a sensitivity analysis?

**Important final call:** We could only give a brief introduction into dependence modelling and some related problems. Likewise, we could only give an overview over some techniques and a very limited number of examples without going into details. Much more can be found in the literature and in the end every application calls for a tailor-made model. Therefore, it may well be necessary to extend and adapt the existing techniques in line with what is needed for a concrete application.

## 10 Summary

In this paper we showed that the dependence structure matters severely when facing different risks. The overall risk may change completely when the dependence changes. We discussed various approaches to model the dependence structure. The most popular measure of dependence is correlation which, however, only covers linear effects and has other drawbacks and limitations. As alternative dependence measures we considered rank correlations and copulae. The latter are theoretically able to encode the complete dependence structure, but for a statistical risk analysis one chooses a certain parametric families which may introduce severe limitations and also model risk; cf. Bannör and Scherer [2]. Furthermore, we explained that elliptical distributions are natural generalisations of the multivariate normal distribution where mainly the correlation structure matters. Finally, we introduced tail dependence and explained that it is of utmost importance in connection with risk modelling, because it captures the dependence of the extremes, which is what typically matters in risk assessment, risk evaluation and consequential risk handling, and which may be rather different from the dependence of the bulk of the observations.

**Acknowledgment.** The authors thank Wilson Lauw for preparing large parts of the figures during his research internship at Ulm University. Financial support of the Institute for Advanced Study of the Technische Universität München (TUM-IAS) is gratefully acknowledged.

## Selected bibliography

- [1] P. Embrechts, A. J. McNeil, and D. Straumann. Correlation: Pitfalls and alternatives. *RISK Magazine*, pages 69–71, May 1999.
- [2] P. Embrechts, A. J. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond (Cambridge, 1998)*, pages 176–223. Cambridge Univ. Press, Cambridge, 2002.
- [3] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. Elsevier, Amsterdam, 2003.
- [4] K.-T. Fang, S. Kotz, and K.-W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London, 1990.
- [5] S. Haug, C. Klüppelberg, and L. Peng. Statistical models and methods for dependent insurance data. *Journal of the Korean Statistical Society*, 40:125–139, 2011.
- [6] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society Series B Statistical Methodology*, 71:737–753, 2009.

- [7] D. Kurowicka and H. Joe, editors. *Dependence Modeling: Vine Copula Handbook*. World Scientific, Singapore, 2010.
- [8] A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton Series in Finance. Princeton University Press, Princeton, 2005.
- [9] R. Nelsen. *An Introduction to Copulas*. Lecture Notes in Statistics, No. 139. Springer, New York, second edition, 2006.
- [10] D. Straub. Engineering risk analysis. In C. Klüppelberg, D. Straub, and I. Welpé, editors, *Risk - A Multidisciplinary Introduction*. 2013.
- [11] Y. Tong. *The Multivariate Normal Distribution*. Springer, New York, 1990.

## Additional literature

- [1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.
- [2] K. Bannör and M. Scherer. Handling model risk and uncertainty; illustrated with examples from mathematical finance. In C. Klüppelberg, D. Straub, and I. Welpé, editors, *Risk - A Multidisciplinary Introduction*. 2012.
- [3] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes*. Wiley, Chichester, 2004.
- [4] F. Biagini, T. Meyer-Brandis, and G. Svindland. The mathematical concept of risk. In *Risk - A Multidisciplinary Introduction*. 2011.
- [5] K. Böcker and C. Klüppelberg. Modelling and measuring multivariate operational risk with Lévy copulas. *Journal of Operational Risk*, 3(2):3–27, 2008.
- [6] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.
- [7] R. Cont and Y. H. Kan. Statistical modeling of credit default swaps. Working paper, 2011. Available at SSRN: <http://ssrn.com/abstract=1771862>.
- [8] S. Demarta and A. McNeil. The  $t$  copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.
- [9] V. Fasen, C. Klüppelberg, and A. Menzel. Quantifying extreme events. In C. Klüppelberg, D. Straub, and I. Welpé, editors, *Risk - A Multidisciplinary Introduction*. 2012.
- [10] G. A. Fredricks and R. B. Nelsen. On the relationship between Spearman’s rho and Kendall’s tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, 137(7):2143–2150, 2007.
- [11] M. Hohenbichler and R. Rackwitz. Non-normal dependent vectors in structural safety. *Journal of the Engineering Mechanics Division, ASCE*, 107(6):1227–1249, 1981.
- [12] H. Joe. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9(1):75–81, 1990.

- [13] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā. The Indian Journal of Statistics. Series A*, 32:419–438, 1970.
- [14] D. X. Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9: 43–54, 2000.
- [15] P.-L. Liu and A. Der Kiureghian. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2):105–112, 1986.
- [16] R. B. Nelsen. On measures of association as measures of positive dependence. *Statistics & Probability Letters*, 14(4):269–274, 1992.
- [17] R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser, Basel, 2 edition, 2001.
- [18] S. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York, 2nd edition, 1987.

Claudia Klüppelberg  
 Center for Mathematical Sciences, and  
 Institute for Advanced Study  
 Technische Universität München  
 Boltzmannstrasse 3  
 85747 Garching b. München  
 Germany  
 Email: [cklu@ma.tum.de](mailto:cklu@ma.tum.de)  
 URL: <http://www-m4.ma.tum.de>

Robert Stelzer  
 Institute of Mathematical Finance  
  
 Ulm University  
 Helmholtzstrasse 18  
 89081 Ulm  
 Germany  
 Email: [robert.stelzer@uni-ulm.de](mailto:robert.stelzer@uni-ulm.de)  
 URL: <http://www.uni-ulm.de/mawi/finmath.html>