Quantitative Risk Management (QRM) 2025/2026

Lecture notes

Rasmus Frigaard Lemvig (rfl@math.ku.dk)

Third edition

Preface

These lecture notes are written for the course *Quantitative Risk Management (QRM)* at the University of Copenhagen and began to take form in the academic year 2023/2024. The first two editions were based almost entirely on the lectures given by Jeffrey F. Collamore with few of my own additions, primarily exercises and supplementary examples. Since then, much supplementary material and many additional exercises have been added.

The primary purpose of the notes is to provide background and understanding of the topics covered in the course. The most essential results and techniques of the course are presented at the lectures, and the lectures should also provide an overview of what to dive into in the notes. In terms of topics, the chapters follow the course almost one to one with additional material and concepts to help your understanding. All sections marked with * are purely supplementary and not part of the curriculum. These notes also contain five appendices. It will be clear in the text when it can be a good idea to consult certain parts of the appendix. Certain techniques, such as integration with respect to functions of finite variation, are taken for granted throughout the course, and taking a look at the appendix before the beginning of the course may thus be fruitful.

Feedback in general is very appreciated. The exercises constitute a large part of this book, and any suggestions in regards to difficulty and relevance are very welcome. Last but not least, there is likely still some typos and mistakes remaining. Please don't hesitate to let me know if you find any.

Rasmus Frigaard Lemvig
January 2026

Table of contents

Table of contents

1 Risk measures **1**
- 1.1 Introduction . . . 1
- 1.2 Loss random variables . . . 2
- 1.3 Risk measures . . . 7
- 1.4 Computational methods . . . 17
- 1.5 More theory for coherent risk measures* . . . 22
- 1.6 Exercises . . . 28

2 Extreme value theory **35**
- 2.1 Introduction . . . 35
- 2.2 Regular variation . . . 36
- 2.3 The Hill estimator . . . 40
- 2.4 The Peaks over threshold (POT) method . . . 44
- 2.5 Maximum domains of attraction* . . . 53
- 2.6 Exercises . . . 61

3 Spherical and elliptical distributions **75**
- 3.1 Spherical and elliptical distributions . . . 77
- 3.2 Properties of elliptical distributions . . . 79
- 3.3 Exercises . . . 85

4 Copulas **89**
- 4.1 What are copulas? . . . 89
- 4.2 Examples of copulas . . . 97
- 4.3 Archimedean copulas . . . 102
- 4.4 Archimedean copulas in higher dimensions . . . 107
- 4.5 Extreme value copulas* . . . 110
- 4.6 Constructing new copulas* . . . 115
- 4.7 Additional proofs and results* . . . 118
- 4.8 Exercises . . . 121

5 Dependence concepts and fitting copulas to data **138**
- 5.1 Introduction . . . 138
- 5.2 Kendall's τ . . . 138
- 5.3 Spearman's ρ . . . 142

5.4 Tail dependence . . . 148
5.5 Fitting copulas to data . . . 150
5.6 Geometric properties of Archimedean copulas* . . . 159
5.7 Additional proofs* . . . 164
5.8 Exercises . . . 167

6 Credit risk **183**
6.1 Portfolio credit risk . . . 183
6.2 Operational risk . . . 193
6.3 Exercises . . . 199

A Generalised inverses **202**
A.1 Definition and basic properties . . . 202
A.2 Exercises . . . 204

B Probability theory **205**
B.1 Random variables and distribution functions . . . 205
B.2 Characteristic functions and moment-generating functions . . . 207
B.3 Some important distributions . . . 209
B.4 Conditional expectations . . . 213
B.5 Exercises . . . 218

C Convergence concepts and results **221**
C.1 Convergence almost surely and in probability . . . 221
C.2 Convergence in distribution . . . 223
C.3 Exercises . . . 225

D Stochastic processes **228**
D.1 Basic definitions . . . 228
D.2 Essential examples . . . 229
D.3 SDE's . . . 231
D.4 Exercises . . . 234

E Finite variation calculus **236**
E.1 Functions of bounded variation . . . 236
E.2 Integration . . . 237
E.3 Exercises . . . 242

Bibliography **243**

Index **254**

Chapter 1

Risk measures

1.1 Introduction

What is quantitative risk management about?

Quantitative risk management aims at describing and understanding risk in a financial context. To motivate our discussion, let us start with a basic example. Suppose we have a stock with value S_n at time n (discrete time units). Assume the Bernoulli model

$$S_0 = 1, \quad P(S_{n+1} = 2S_n) = \frac{2}{3} \quad \text{and} \quad P(S_{n+1} = 0.5S_n) = \frac{1}{3} \quad \text{for} \quad n > 0.$$

In this model there are some basic questions we may ask. For example, what is the risk in this investment policy? What are the expected returns? We can compute

$$E[S_n \mid S_{n-1}] = \frac{2}{3} \cdot 2S_{n-1} + \frac{1}{3} \cdot \frac{1}{2} S_{n-1} = \frac{3}{2} S_{n-1}$$

and it follows that

$$M_n = \left(\frac{2}{3}\right)^n S_n$$

is a martingale, and

$$1 = M_0 = E[M_n] = \left(\frac{2}{3}\right)^n E[S_n], \quad \text{implying} \quad E[S_n] = \left(\frac{3}{2}\right)^n \to \infty.$$

However, there is still a risk that we lose money in this investment policy. Note that we can write

$$S_n = R_1 \cdots R_n$$

for Bernoulli variables R_i with $P(R_i = 2) = 2/3$ and $P(R_i = 1/2) = 1/3$. Taking log yields

$$\log S_n = \sum_{i=1}^{n} \log R_i$$

which is a risk process that can be studied using ruin theory. If $E[e^{-\alpha \log S_1}] = 1$ for some $\alpha > 0$, we obtain the Cramér–Lundberg estimate (for some threshold $u > 0$)

$$P(\log S_n < -u \text{ for some } n) \sim Ce^{-\alpha u}$$

for a constant C. In Cramér–Lundberg theory, there is a tradeoff between profit and risk and the same rule applies to risk management in finance. This tradeoff can be studied from several viewpoints. Sometimes this tradeoff is handled using a utility function U and then maximising the expected utility $E[U(S_n)]$. In this course we follow another approach via so-called risk measures. This approach allows us to compute loss probabilities directly. All these terms will get a more rigorous definition later on.

Some historical remarks

An essential part of risk management is to determine the capital needed to withstand shocks for financial institutions. History has many examples of banks and other institutions failing due to insufficient financial coverage. We go through some of these examples to add more context to the following (more mathematical) discussion.

Example 1.1.1. Barings Bank was founded in 1762 and was one of the UK's oldest and most respected banks. In 1995 the bank collapsed despite having more than $900 million in capital. This was due to unauthorized trading by a single employee. He bought straddles (selling both one call option and put option) which in a typical market will usually expire and provide a gain. However, market instability was caused by a Japanese earthquake, and the bank suffered a loss of more than $1 billion, resulting in bankruptcy. ○

Example 1.1.2. In 1994, the hedge fund LTCM (Long-Term Capital Management) was founded. The employees were experienced traders and academics. $1.3 billion were invested with returns after two years close to 40 %. Early in 1998, net assets were $4 billion, but by the end of the year the fund was close to default. The U.S. Federal Reserve managed a $3.5 billion rescue package to avoid a systematic crisis in the world financial system. The triggering event for this disaster was the devaluation of the ruble by Russia. ○

Example 1.1.3. In 2023, Silicon Valley Bank collapsed. The bank owned low interest bonds and paid even lower interest to the depositors. When the depositors withdrew their capital, this required selling the low-interest treasury bonds, whose market prices had decreased sharply. ○

Example 1.1.4. The final example concerns a particular individual, Jesse Livermore. Livermore is considered the greatest short-seller in history. He succesfully shorted:

- The 1906 earthquake (through a railway investment).
- The 1907 market crash ("panic of 1907").
- The 1929 market crash (through ca. 100 shorts, netting about $100 million).

He also had numerous successful "long" investments. He went bankrupt a number of times however. He went bankrupt in 1901, 1908 and again in 1934. His book on trading remains a classic to this day. ○

1.2 Loss random variables

In describing risk, we focus on the loss instead of the profit. The following definition sets the stage for the most important examples of risks.

Definition 1.2.1. Let $t_0, t_1, ...$ denote discrete timepoints (days, weeks, months or years for example) and let $\Delta t_n = t_{n+1} - t_n$ denote the time passed between timepoint $n+1$ and n. We let V_n denote the *capital* at time t_n, and we let L_{n+1} define the loss between t_n and t_{n+1} e.g.

$$L_{n+1} = -(V_{n+1} - V_n).$$

We let a general loss random variable be denoted by L. Many models consist of assumptions on L. Let us consider some motivating examples. Note that in these examples, it makes sense to think of V_n as a portfolio.

Example 1.2.2 (**Stock investment**)**.** Let $V_n = S_n$ with S_n the price of a certain stock at time t_n. We let $X_{n+1} = \log S_{n+1} - \log S_n$ denote the log returns. Hence

$$e^{X_{n+1}} = \frac{S_{n+1}}{S_n}.$$

Historically, X_{n+1} has often been given a normal distribution. This is motivated by the (Standard) Black–Scholes model, see the end of the examples for a concise explanation of this model. In this model, the change of the stock price (in continuous time t) is described by the dynamics

$$\frac{dS(t)}{S(t)} = rdt + \sigma dW(t)$$

where r and σ are constants with $W(t)$ denoting a Brownian motion. Solving the equation explicitly, assuming that we are currently at time t so that $S(t)$ is known, yields the expression for the stock price at time $T > t$,

$$S(T) = S(t)e^{\left(r - \frac{\sigma^2}{2}\right)(T-t) + \sigma(W(T) - W(t))}.$$

The discrete time analogue is

$$S_{n+1} = S_n e^{\left(r - \frac{\sigma^2}{2}\right)(t_{n+1} - t_n) + \sigma\sqrt{t_{n+1} - t_n}Z}$$

with $Z \sim \mathcal{N}(0,1)$. The log return becomes

$$X_{n+1} = \log S_{n+1} - \log S_n = \left(r - \frac{\sigma^2}{2}\right)(t_{n+1} - t_n) + \sigma\sqrt{t_{n+1} - t_n}Z$$

which is normally distributed. Note that the loss L_{n+1} can be written as

$$L_{n+1} = -(S_{n+1} - S_n) = -S_n(e^{X_{n+1}} - 1).$$

If we are at time t_n, the value S_n is known. A key goal of this course is to model the unknown part X_{n+1} and thereby make inference about the behaviour of the process V_n at a future time. Note also how this approach differs from the one in classical ruin theory where the entire positive timeline is considered. Here we model the change in capital one step at a time. ∘

Example 1.2.3 (**Stock investment with more assets**)**.** The preceding example can be generalised. Assume that we have d stocks. We can then form a portfolio at time t_n by

$$V_n = \sum_{i=1}^{d} \alpha_i S_n^{(i)}$$

with α_i the number of units bought of stock i and $S_n^{(i)}$ the value of stock i at time t_n. Using the previous example, the loss is given by

$$L_{n+1} = -\sum_{i=1}^{d} \alpha_i S_n^{(i)}(e^{X_{n+1}^{(i)}} - 1), \quad X_{n+1}^{(i)} = \log S_{n+1}^{(i)} - \log S_n^{(i)}.$$

Again we need to come up with a model for the log returns $X^{(1)}, ..., X^{(d)}$, where we can write

$$\mathbf{X}_{n+1} = (X_{n+1}^{(1)}, ..., X_{n+1}^{(d)}).$$

Very often the variables in $\mathbf{X}_{n+1}$ are dependent. Think for example of a portfolio of stocks in the same type of companies. This dependence structure is crucial to understand and capture when estimating risk in a portfolio.

○

Example 1.2.4 (**Bond investment**)**.** Consider a *zero-coupon bond*. Such an asset pays one unit at a fixed time T. Let us briefly consider such a bond in continuous time. We have an interest rate r_t at time t, and we let B_t denote the price of the bond at time t (this price will depend on T). The price of the bond can be described by the dynamics

$$dB_t = r_t B_t dt,$$

and using the boundary condition $B_T = 1$, we can solve the above equation and get

$$1 = B_T = B_t e^{\int_t^T r_s ds},$$

and we can rewrite this expression in terms of B_t as

$$B_t = e^{-\int_t^T r_s ds} = e^{-(T-t)y(t,T)}, \quad \text{where} \quad y(t,T) = \frac{1}{T-t}\int_t^T r_s ds$$

is called the *yield* of the bond. Let us now consider discrete time and say that the current time is t_n. The loss is

$$L_{n+1} = -(B_{t_{n+1}} - B_{t_n}) = -B_{t_n}\left(\frac{B_{t_{n+1}}}{B_{t_n}} - 1\right).$$

Let us fix some notation. Denote the yield at time t_n by $Z_n = y(t_n, T)$, and let $X_{n+1} = Z_{n+1} - Z_n$. Note that Z_n is known at time t_n while X_{n+1} again needs to be modelled. We can rewrite the expression $\frac{B_{t_{n+1}}}{B_{t_n}}$ as

$$\begin{aligned}\frac{B_{t_{n+1}}}{B_{t_n}} &= e^{-(T-t_{n+1})y(t_{n+1},T)+(T-t_n)y(t_n,T)} \\ &= e^{-(T-t_n-\Delta t_n)(Z_n+X_{n+1})+(T-t_n)Z_n} \\ &= e^{\Delta t_n Z_n-(T-t_{n+1})X_{n+1}}.\end{aligned}$$

This expression makes it clear how the unknown variable X_{n+1} enters the loss.

○

In the above examples, we ended up with an expression containing a term of the form $e^{(\cdots)} - 1$. This makes it tempting to use a Taylor approximation since $e^x \approx 1 + x$ for small x, and historically, such an approximation was often considered to ease computations. Taking the example with the stock portfolio, we let

$$L_{n+1}^{\Delta} = -\sum_{i=1}^{d} \alpha_i S_n^{(i)} X_{n+1}^{(i)}$$

denote the *linerarized log returns*. This approximation can often be problematic however since it is of interest to consider large losses (which we will do in the next chapter).

A brief rundown of the Black–Scholes model

Consider a risk free asset with price process denoted by B (a bank account) given by the continuous dynamics

$$dB_t = r_t B_t dt$$

where it is often assumed that $B_0 = 1$. We can explicitly solve for the price and obtain

$$B_t = B_0 e^{\int_0^t r_s ds}.$$

r_t is called the interest rate and is assumed to be an adapted process. We model a risky asset (such as a stock) with price process S_t by a *stochastic differential equation* (SDE) of the form

$$dS_t = \mu(t, S_t)dt + \sigma(t, S_t)dW_t$$

with deterministic functions μ and σ and a Brownian motion W. μ is called the *local mean rate of return* for S_t while σ is called the *volatility* of S_t. For all the necessary results on SDEs, consult chapter 4 and 5 of [9]. For our purposes, it suffices to know the model on an intuitive basis. The Black–Scholes model is a special case of the above model.

Definition 1.2.5. The (Standard) Black–Scholes model consists of two assets with dynamics

$$\begin{aligned} dB_t &= rB_t dt, \\ dS_t &= \mu S_t dt + \sigma S_t dW_t \end{aligned}$$

with r, μ and σ constants.

In the language of SDEs, a process with the dynamics of S_t is called a *Geometric Brownian motion* (GBM). Such an SDE can be solved explicitly. In our case, we may write

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}.$$

The Black–Scholes model can of course be extended to include more risky assets with the same type of dynamics as S_t. Such a model is naturally referred to as a multidimensional Black–Scholes model.

The goal of arbitrage theory is to price financial *derivatives*, that is, products based on the price of underlying assets. We think intuitively of an arbitrage as a money machine/free

lunch, namely as a portfolio of assets that costs nothing and produces a positive amount of money with probability one. It turns out that the absence of arbitrage is the same as the existence of a so-called *equivalent martingale measure* (EMM), i.e. a measure Q equivalent to the underlying measure P (Q and P have the same null sets) and such that the discounted price process

$$\frac{S_t}{B_t}$$

is a martingale under Q. Q is also referred to as a *risk neutral measure*. Under the measure Q, the dynamics of S_t change to

$$dS_t = rS_t dt + \sigma S_t dW_t^Q$$

where W^Q is a Brownian motion under the Q measure. Note that the volatility is unchanged while the local mean rate of return becomes the interest rate times S_t. Say we have a derivative which expires at time T, the current time is $t < T$ and that the derivative pays $X = \Phi(S_T)$ at time T. Note that the payout is a function of the price of the risky asset at time T (such a derivative is called *simple*). If $\Pi_t[X]$ denotes the (arbitrage free) price of X at time t, arbitrage theory yields the following formula.

Theorem 1.2.6 (**Risk Neutral Valuation**). *The arbitrage free price of $\Phi(S_T)$ at time $t < T$ is given by*

$$\Pi_t[X] = e^{-r(T-t)} E^Q[\Phi(S_T) \mid \mathcal{F}_t].$$

By an arbitrage free price we mean a price process that doesn't introduce an arbitrage into the market. Note that the above formula says that the price is given by the discounted expected value under the risk neutral measure given the information we currently have available, $\mathcal{F}_t$. It is typical to let $\mathcal{F}_t = \sigma(W_s : s \leq t)$ so that $\mathcal{F}_t$ is the information generated by the underlying Brownian motion.

Example 1.2.7 (**European call option**). A European call option gives the holder the right (but not the obligation) to buy one stock at time T at price K (the *strike price*). The payout is $(S_T - K)^+ = \max\{S_T - K, 0\}$ since if $S_T > K$, we get the payout $S_T - K$ while if $S_T \leq K$, the option is worthless. In the Black–Scholes model, we can solve for S_T as

$$S_T = S_t e^{\left(r - \frac{\sigma^2}{2}\right)(T-t) + \sigma(W_T^Q - W_t^Q)}$$

as seen earlier. If $C(t,T)$ denotes the price at time t, we have by the Risk Neutral Valuation formula that

$$C(t,T) = e^{-r(T-t)} E^Q[(S_T - K)^+ \mid \mathcal{F}_t]$$

and this can be computed explicitly[1]. The result is known as the *Black–Scholes formula*. It says that

$$C(t,T) = S_t \Phi(u) - K e^{-r(T-t)} \Phi(v)$$

where

$$u = \frac{\log(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad v = u - \sigma\sqrt{T-t}.$$

[1]The brave reader can carry out this computation. It involves a lot of integration by substitution.

Say we have a portfolio consisting of one European call option, so that the value is $V_n = C(t_n, T)$. We note that the risk in this portfolio can be explained by the three quantities

$$\mathbf{Z}_n = (Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}) := (\log S_{t_n}, r_n, \sigma_n)$$

since the volatility and interest rate are not constant in real life. One typically needs to model the change in these factors (called risk factors, see the discussion below), namely $\mathbf{X}_{n+1} = \mathbf{Z}_{n+1} - \mathbf{Z}_n$. We can for example consider the linearized loss

$$L_{n+1}^{\Delta} = -\left(\frac{\partial C}{\partial t}\Delta t + \frac{\partial C}{\partial S}X_{n+1}^{(1)} + \frac{\partial C}{\partial r}X_{n+1}^{(2)} + \frac{\partial C}{\partial \sigma}X_{n+1}^{(3)}\right).$$

In mathematical finance, these first order derivatives have names. $\partial C/\partial t$ is called "theta", $\partial C/\partial S$ "delta", $\partial C/\partial r$ "rho" and $\partial C/\partial \sigma$ "vega". Together these quantities are referred to as the *greeks*, see chapter 10 in [9].

∘

A general risk model

All examples considered above had the same form for V_n. We could write $V_n = f(t_n, \mathbf{Z}_n)$ for some (measurable) function f and $\mathbf{Z}_n$ suitable random variables.

Definition 1.2.8. For a portfolio of the form $V_n = f(t_n, \mathbf{Z}_n)$ with $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ a measurable function and $\mathbf{Z}_n = (Z_{n,1}, ..., Z_{n,d})$ a random vector, we call the variables in $\mathbf{Z}_n$ *risk factors*. We call $\mathbf{X}_{n+1} = \mathbf{Z}_{n+1} - \mathbf{Z}_n$ the *risk-factor changes* at time t_{n+1}.

In the previous examples we stressed that we only need to model the change in risk factors $\mathbf{X}_{n+1}$ since the value $\mathbf{Z}_n$ is already known at time t_n. Hence we can write the loss entirely in terms of the change in risk factors. Explicitly,

$$L_{n+1} = -(V_{n+1} - V_n) = -(f(t_{n+1}, \mathbf{Z}_n + \mathbf{X}_{n+1}) - f(t_n, \mathbf{Z}_n)) =: l_{[n]}(\mathbf{X}_{n+1}).$$

We refer to $l_{[n]}$ as the *loss operator*. By considering the linearized loss

$$L_{n+1}^{\Delta} = -\frac{\partial f}{\partial t}(t_n, \mathbf{Z}_n)\Delta t - \sum_{i=1}^{d} \frac{\partial f}{\partial z_i}(t_n, \mathbf{Z}_n)X_{n+1}^{(i)}$$

obtained by applying a first order Taylor expansion, we can similarly define the *linearized loss operator*

$$l_{[n]}^{\Delta}(\mathbf{x}) := -\frac{\partial f}{\partial t}(t_n, \mathbf{Z}_n)\Delta t - \sum_{i=1}^{d} \frac{\partial f}{\partial z_i}(t_n, \mathbf{Z}_n)x^{(i)}.$$

1.3 Risk measures

We need some notion of the "size" of a risk in order to quantify the risk of a loss.

Definition 1.3.1. Let L denote a loss. A *risk measure* ρ associates a real number to L denoted by $\rho(L)$.

A way to make the definition more formal is to let $\mathcal{G}$ denote the set of all measurable real-valued functions on the background probability space. A risk measure is then a mapping $\rho : \mathcal{G} \to \mathbb{R}$. We will generally not worry about these details in this course. We now go through some essential examples of risk measures. In the following, let L denote some loss random variable.

Value at Risk (VaR)

Definition 1.3.2. Let $\alpha \in (0,1)$ and L a loss variable. We define the *Value at Risk* (VaR) at level $\alpha \in (0,1)$ as

$$\mathrm{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : P(L > x) \leq 1 - \alpha\}.$$

One can intuitively think of $\mathrm{VaR}_\alpha(L)$ as the smallest value of x such that $P(L \leq x) \geq \alpha$. We can rewrite

$$\begin{aligned}\mathrm{VaR}_\alpha(L) &= \inf\{x \in \mathbb{R} : 1 - P(L \leq x) \leq 1 - \alpha\} \\ &= \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\} \\ &= F_L^{\leftarrow}(\alpha) =: q_\alpha(F_L)\end{aligned}$$

with F_L denoting the distribution function of L and $F_L^{\leftarrow}$ the generalised inverse of F_L. Since F_L is a distribution function, the generalised inverse coincides with the quantile function $q_{(\cdot)}(L)$. So a statistician would simply call $\mathrm{VaR}_\alpha(L)$ the α-quantile of L. See the appendix for more information on generalised inverses and their properties.

Expected Shortfall

For $\alpha > 0$ and F_L continuous and strictly increasing, we define

$$\overline{\mathrm{ES}}_\alpha(L) = E[L \mid L \geq \mathrm{VaR}_\alpha(L)]$$

called the *Expected Shortfall* at level α. This is the expected loss given that the loss has surpassed the Value at Risk. We immediately see that $\overline{\mathrm{ES}}_\alpha(L) \geq \mathrm{VaR}_\alpha(L)$ and that $\overline{\mathrm{ES}}_\alpha$ takes into account the severity of the loss in comparison to VaR_α. We want to generalise the Expected Shortfall to also be valid for non-continuous distribution functions. We will need the following lemma.

Lemma 1.3.3. *If U is uniformly distributed on $(0,1)$ and L is a random variable with distribution function F_L, then $L \overset{\mathrm{d}}{=} F_L^{\leftarrow}(U)$.*

Proof. Let $Y = F_L^{\leftarrow}(U)$. A property of generalised inverses (see the appendix) is that $u \leq F_L(x)$ if and only if $F_L^{\leftarrow}(u) \leq x$. Hence

$$P(Y \leq x) = P(F_L^{\leftarrow}(U) \leq x) = P(U \leq F_L(x)) = F_L(x)$$

since U is uniformly distributed on $(0,1)$. This proves $Y \overset{\mathrm{d}}{=} L$ as desired. ■

The following proposition tells us how to generalise the notion of Expected Shortfall.

Proposition 1.3.4. *Let L be a loss variable with F_L continuous and strictly increasing. Then*

$$\overline{\mathrm{ES}}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du.$$

Proof. Because F_L is continuous and strictly increasing, the generalised inverse is a proper inverse. Hence

$$P(L \geq \mathrm{VaR}_\alpha(L)) = 1 - \alpha$$

so by the previous lemma,

$$\begin{aligned}\overline{\mathrm{ES}}_\alpha &= \frac{1}{P(L \geq \mathrm{VaR}_\alpha(L))} E[L1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}] = \frac{1}{1-\alpha} E[L; L \geq \mathrm{VaR}_\alpha(L)] \\ &= \frac{1}{1-\alpha} E[F_L^{\leftarrow}(U); F_L^{\leftarrow}(U) \geq F_L^{\leftarrow}(\alpha)] = \frac{1}{1-\alpha} E[F_L^{\leftarrow}(U); U \geq \alpha] \\ &= \frac{1}{1-\alpha} E[\mathrm{VaR}_U(L); U \geq \alpha] = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L) du.\end{aligned}$$

■

Most problems include distributions that have continuous and strictly increasing distribution functions, but it is also useful to have a definition of Expected Shortfall for general distributions. The following definition summarises the above considerations.

Definition 1.3.5. If $E[|L|] < \infty$, we define

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L) du$$

called the Expected Shortfall at level α.

Example 1.3.6 (**Stock investment**)**.** Recall the example on stock investments from earlier. If S_n denotes the price of the stock at time t_n, we had the loss variable $L_{n+1} = -S_n(e^{X_{n+1}} - 1)$ where $X_{n+1} = \log S_{n+1} - \log S_n$ denotes the log return. We assume $X_{n+1} \sim \mathcal{N}(\mu, \sigma^2)$. In this case, L_{n+1} has a distribution function which is continuous and strictly increasing, so we may compute $\mathrm{VaR}_\alpha(L_{n+1})$ by solving the equation $P(L_{n+1} > x) = 1 - \alpha$:

$$\begin{aligned}1 - \alpha &= P(-S_n(e^{X_{n+1}} - 1) > x) = P\left(X_{n+1} < \log\left(1 - \frac{x}{S_n}\right)\right) \\ &= P\left(\frac{X_{n+1} - \mu}{\sigma} < \frac{\log\left(1 - \frac{x}{S_n}\right) - \mu}{\sigma}\right) = \Phi\left(\frac{\log\left(1 - \frac{x}{S_n}\right) - \mu}{\sigma}\right)\end{aligned}$$

with Φ denoting the distribution function of a standard normal variable. Hence

$$\frac{\log\left(1 - \frac{x}{S_n}\right) - \mu}{\sigma} = \Phi^{-1}(1 - \alpha)$$

and we can solve for x explicitly as follows:

$$\begin{aligned}\sigma\Phi^{-1}(1-\alpha) + \mu = \log\left(1 - \frac{x}{S_n}\right) \quad &\Leftrightarrow \quad e^{\sigma\Phi^{-1}(1-\alpha)+\mu} = 1 - \frac{x}{S_n} \\ &\Leftrightarrow \quad x = S_n - S_n e^{\sigma\Phi^{-1}(1-\alpha)+\mu}.\end{aligned}$$

If we consider more stocks, the expression becomes a lot more complicated. It gets even worse with a more diverse portfolio (with stocks, bonds and call options for example).

○

Example 1.3.7. This is from Example 2.11 and 2.14 from [88]. Let $\alpha \in (0,1)$ and assume that the loss distribution F_L is normally distributed with mean μ and variance σ^2. Since F_L is continuous and strictly increasing, we have

$$\mathrm{VaR}_\alpha(L) = \mu + \sigma\Phi^{-1}(\alpha)$$

using the properties of Value at Risk from before. We can now compute the Expected Shortfall, where we again use that F_L is continuous and strictly increasing,

$$\begin{aligned}\mathrm{ES}_\alpha(L) &= E[L \mid L \geq q_\alpha(L)] = \mu + \sigma E\left[\frac{L-\mu}{\sigma} \mid \frac{L-\mu}{\sigma} \geq q_\alpha\left(\frac{L-\mu}{\sigma}\right)\right] \\ &= \mu + \sigma E\left[\frac{L-\mu}{\sigma} \mid \frac{L-\mu}{\sigma} \geq \Phi^{-1}(\alpha)\right] \\ &= \mu + \frac{\sigma}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} x\varphi(x)dx\end{aligned}$$

with φ denoting the density of a standard normal variable. Note that

$$\varphi'(x) = \frac{d}{dx}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right) = -x\varphi(x)$$

so that $x\varphi(x)$ has $-\varphi(x)$ as antiderivative. Hence

$$\mathrm{ES}_\alpha(L) = \mu + \frac{\sigma}{1-\alpha}[-\varphi(x)]_{\Phi^{-1}(\alpha)}^{\infty} = \mu + \sigma\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

∘

Properties/axioms of risk measures

It is natural to ask what characterises a good risk measure. We think intuitively of a risk measure as an amount of capital needed by a financial institution to withstand large shocks. In the above examples with VaR and ES, we had a level α, and we often think of α as large, for example $0.95, 0.99$ or 0.995. In an article by Artzner, Delbaen, Eber and Heath [3], certain desirable properties of risk measures are suggested. They are as follows.

Definition 1.3.8. For a risk measure ρ and loss variables L, L_1, L_2, we consider the following axioms/properties:

1. *Translation invariance*: $\rho(L+a) = \rho(L) + a$ for every constant $a \in \mathbb{R}$.
2. *Subadditivity*: $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$.
3. *Positive homogeneity*: $\rho(\lambda L) = \lambda\rho(L)$ for all $\lambda > 0$.
4. *Monotonicity*: $L_1 \leq L_2$ implies $\rho(L_1) \leq \rho(L_2)$.

A risk measure satisfying all these axioms is called *coherent*.

The rationale for translation invariance is that adding a deterministic quantity to the loss should increase the capital we need to set aside by exactly that amount. The rationale for subadditivity is that diversification should reduce risk. Positive homogeneity makes sense since, if we invest more money into the samme asset, the amount of capital we need to set aside should be multiplied by the same factor. Monotonocity also clearly makes sense. Note also the similarity to the pricing principles from non-life insurance.

Properties of VaR and ES

VaR and ES are the two most popular choices of risk measures. One reason many prefer ES over VaR is that ES in general satisfies all the above axioms, while VaR only satisfies three. Nevertheless, it should be mentioned that the VaR is very easy to interpret, a property which should not be underestimated. The following proposition shows that VaR satisfies translation invariance, positive homogeneity and monotonicity.

Proposition 1.3.9. *Let $\alpha \in (0,1)$ and let L be a loss variable.*

(i) For constants $a > 0, b \in \mathbb{R}$, we have

$$\mathrm{VaR}_\alpha(aL+b) = a\,\mathrm{VaR}_\alpha(L) + b$$

so that in particular, VaR_α satisfies translation invariance and positive homogeneity.

(ii) VaR_α satisfies monotonicity.

Proof. For (i), we compute

$$\begin{aligned}\mathrm{VaR}_\alpha(aL+b) &= \inf\{x \in \mathbb{R} : P(aL+b > x) \le 1-\alpha\}\\ &= \inf\{x \in \mathbb{R} : P(L > (x-b)/a) \le 1-\alpha\}\\ &= \inf\{ax+b \in \mathbb{R} : P(L > x) \le 1-\alpha\}\\ &= a\inf\{x \in \mathbb{R} : P(L > x) \le 1-\alpha\} + b = a\,\mathrm{VaR}_\alpha(L) + b.\end{aligned}$$

(ii) If $L_1 \le L_2$ then $\{x \in \mathbb{R} : P(L_2 > x) \le 1-\alpha\} \subseteq \{x \in \mathbb{R} : P(L_1 > x) \le 1-\alpha\}$ and the claim follows. ■

Value at Risk is not a coherent risk measure in general, since subadditivity may fail. We provide two examples below. The first is from [3] page 217, and the second is Example 2.25 of [88]. Both examples are discrete. A continuous counterexample is provided in the exercises.

Example 1.3.10. Consider two *digital options* on a stock S. A digital option is an option which pays a fixed amount to the holder of the option, if the value of the stock goes below or above a certain threshold. Both our options (call them A and B) have the exercise time T. A pays 1000 if $S_T > U$ and nothing otherwise and B pays 1000 if $S_T < L$ and nothing otherwise where $L < U$ are fixed strike prices. Say A has price u and B price l with $u + l < 1000$. Say we wish to sell these options. If we sell a single A option, we gain u with probability 0.992, while we gain $u - 1000$ with probability 0.008. In other words, the loss L_A of this trade has distribution $P(L_A = -u) = 0.992$ and $P(L_A = 1000 - u) = 0.008$. Similarly, the loss L_B of selling a single option B has distribution given by the point probabilities $P(L_B = -l) = 0.992$ and $P(L_B = 1000 - l) = 0.008$. Hence

$$\mathrm{VaR}_{0.99}(L_A) + \mathrm{VaR}_{0.99}(L_B) = -u - l,$$

but $L_A + L_B$ has probability 0.016 of giving $1000 - u - l$, so $\mathrm{VaR}_{0.99}(L_A + L_B) = 1000 - u - l > 0$. Hence subadditivity fails. ∘

Example 1.3.11. Consider two independent bonds with face value 100 and an interest rate of 5%. After one year, the bonds mature and thus give a payout of 105. We assume no recovery on the bonds, so if the bond fails, the investor receives nothing. Letting L_i denote the loss of bond i, we assume

$$P(L_i = -5) = 1 - p = 0.991 \quad \text{and} \quad P(L_i = 100) = p = 0.009.$$

Consider a portfolio consisting of one of each bond with loss $L = L_1 + L_2$. Then by independence,

$$\begin{aligned} P(L = -10) &= (1-p)^2 = 0.982, \\ P(L = 95) &= 2p(1-p) = 0.017838, \\ P(L = 200) &= p^2 = 0.000081 \end{aligned}$$

corresponding to no defaults, a single default and two defaults, respectively. As the reader may verify, $\mathrm{VaR}_{0.99}(L) = 95$ and $\mathrm{VaR}_{0.99}(L_i) = -5$, which shows that subadditivity fails. Furthermore, note that

$$\mathrm{VaR}_{0.99}(0.5L_1 + 0.5L_2) = 0.5\,\mathrm{Var}_{0.99}(L) = 47.5 > \mathrm{VaR}_{0.99}(L_1)$$

so in this case, diversification is strongly discouraged when using VaR. ◦

In order to prove that the Expected Shortfall is coherent, we need a characterisation of coherent risk measures. This can be done in several ways. One such approach is the topic of the next subsection. Before presenting this theory, we prove subadditivity of the Expected Shortfall when the loss variables involved have continuous distributions. The general case requires more theory, and we will return to it later in the chapter.

Theorem 1.3.12 (**Coherence of ES (continuous case)**). *Expected Shortfall is a coherent risk measure.*

Proof. Translation invariance, positive homogeneity and monotonicity follow immediately by the previous proposition and the definition of Expected Shortfall in terms of the Value at Risk. Subadditivity is harder to prove. We only prove it in the case where L_1 and L_2 have continuous distribution functions. The proof of the general case can be found in [88], see Theorem 8.14. The following proof is from Example 2.26 of the same book. Recall from earlier computations that for L with a continuous distribution function,

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} E[L 1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}].$$

Define $I_i := 1_{\{L_i \geq \mathrm{VaR}_\alpha(L_i)\}}$ for $i = 1, 2$ and $I_{12} := 1_{\{L_1+L_2 \geq \mathrm{VaR}_\alpha(L_1+L_2)\}}$. We compute

$$\begin{aligned} (1-\alpha)(\mathrm{ES}_\alpha(L_1) + \mathrm{ES}_\alpha(L_2) - \mathrm{ES}_\alpha(L_1+L_2)) &= E[L_1 I_1] + E[L_2 I_2] - E[(L_1+L_2) I_{12}] \\ &= E[(L_1(I_1 - I_{12}))] + E[(L_2(I_2 - I_{12}))]. \end{aligned}$$

We now consider two cases for L_1. If $L_1 \geq \mathrm{VaR}_\alpha(L_1)$, then $I_1 - I_{12} \geq 0$ and hence $L_1(I_1 - I_{12}) \geq \mathrm{VaR}_\alpha(L_1)(I_1 - I_{12})$. If $L_1 < \mathrm{VaR}_\alpha(L_1)$, then $I_1 - I_{12} \leq 0$ so again we have

$L_1(I_1 - I_{12}) \geq \mathrm{VaR}_\alpha(L_1)(I_1 - I_{12})$. Applying the same reasoning to L_2, we get

$$\begin{aligned}(1-\alpha)(\mathrm{ES}_\alpha(L_1) + \mathrm{ES}_\alpha(L_2) - \mathrm{ES}_\alpha(L_1+L_2)) &\geq E[\mathrm{VaR}_\alpha(L_1)(I_1 - I_{12}) + \mathrm{VaR}_\alpha(L_2)(I_2 - I_{12})] \\ &= \mathrm{VaR}_\alpha(L_1)E[I_1 - I_{12}] + \mathrm{VaR}_\alpha(L_2)E[I_2 - I_{12}] \\ &= \mathrm{VaR}_\alpha(L_1)((1-\alpha) - (1-\alpha)) \\ &\quad + \mathrm{VaR}_\alpha(L_2)((1-\alpha) - (1-\alpha)) \\ &= 0\end{aligned}$$

implying that $\mathrm{ES}_\alpha(L_1) + \mathrm{ES}_\alpha(L_2) \geq \mathrm{ES}_\alpha(L_1 + L_2)$ which is the desired statement.

■

More examples and techniques

We here present more elaborate examples illustrating how to compute the Value at Risk. In particular, we would like to gain some experience computing VaR for "mixed" distributions, namely distributions that are neither continuous nor discrete but rather a mix of both. Before seeing concrete examples, we provide some theoretical but useful results. The proofs are not particularly illuminating and may be skipped at first reading.

It is intuitively clear that applying an increasing function to a loss variable L should yield a nice expression for the Value at Risk, since the ordering is at least preserved up to ties. This preservation turns out to be the case under a few technical conditions.

Lemma 1.3.13. *Let L be a loss variable and let g be non-decreasing and left-continuous. Then* $\mathrm{VaR}_\alpha(g(L)) = g(\mathrm{VaR}_\alpha(L))$.

Proof. As $F_{g(L)}$ is right-continuous, Proposition A.1.3 from the appendix gives that

$$F^{\leftarrow}_{g(L)}(\alpha) \leq x \quad \Leftrightarrow \quad F_{g(L)}(x) \geq \alpha$$

for all $x \in \mathbb{R}$. Now let $y_0(x) = \sup\{y : g(y) \leq x\}$. We claim that

$$g(y) \leq x \quad \Leftrightarrow \quad y \leq y_0(x).$$

Indeed, $\Rightarrow$ always holds by definition of $y_0(x)$. For the converse, choose a sequence $\{y_n\} \subseteq \{y : g(y) \leq x\}$ with $y_n \uparrow y_0(x)$. Then by left-continuity of g, we have

$$g(y_0(x)) = \lim_{n\to\infty} g(y_n) \leq x.$$

As g is non-decreasing, it follows for any $y \leq y_0(x)$ that $g(y) \leq g(y_0(x)) \leq x$, and the converse implication is proved. We thus have $\{g(L) \leq x\} = \{L \leq y_0(x)\}$ and by taking probabilities,

$$F_{g(L)}(x) \geq \alpha \quad \Leftrightarrow \quad F_L(y_0(x)) \geq \alpha.$$

Applying Proposition A.1.3 again, we thus conclude

$$F_L(y_0(x)) \geq \alpha \quad \Leftrightarrow \quad F^{\leftarrow}_L(\alpha) \leq y_0(x) \quad \Leftrightarrow \quad g(F^{\leftarrow}_L(\alpha)) \leq x.$$

Combining all the equivalences, we obtain the result. ■

It is natural to ask what happens to the Value at Risk when applying a decreasing function. The situation turns out to be more complicated, and one should be careful when the distribution of L is not sufficiently nice. We think of the Value at Risk at level α as the smallest value x where the probability of $L \leq x$ is larger than or equal to α. One may then ask when this coincides with the largest value of x such that the probability of $L < x$ is smaller than or equal to α. To be precise, when does it hold that

$$\inf\{x : P(L \leq x) \geq \alpha\} = \sup\{x : P(L < x) \leq \alpha\}?$$

The following technical lemma provides an answer. Define

$$\mathrm{VaR}_\alpha^+(L) := \sup\{x : P(L < x) \leq \alpha\}.$$

Lemma 1.3.14. *Let L be a loss variable and consider for a fixed $\alpha \in (0,1)$ the set $S_\alpha = \{x : F_L(x) = \alpha\}$. If S_α is either empty or a singleton set, it holds that*

$$\mathrm{VaR}_\alpha(L) = \mathrm{VaR}_\alpha^+(L).$$

Proof. The proof is almost immediate by looking at the figure below, but for completeness, we provide a formal argument also.

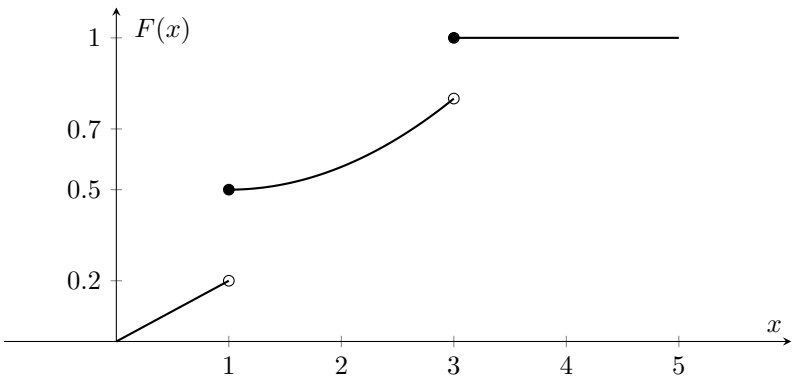


Figure 1: The distribution function for a generic loss variable L. If no x-value hits α (e.g. $\alpha = 0.3$), then it doesn't matter if you take the largest x such that $F_L(x-) \leq \alpha$ or the smallest x such that $F_L(x) \geq \alpha$. The same happens in any region where $F_L(x) = \alpha$ has a unique solution.

We consider both cases for S_α. Assume first that $S_\alpha = \emptyset$. Then there exists a unique point x_0 such that $F_L(x) < \alpha$ for $x < x_0$ and $F_L(x) > \alpha$ for $x \geq x_0$. It is clear that $\mathrm{VaR}_\alpha(L) = x_0$. If $x \leq x_0$, then $F_L(x-) = \lim_{y \uparrow x} F_L(y) \leq \alpha$, so $\mathrm{VaR}_\alpha^+(L) \geq x_0$. If $x > x_0$, then $F_L(x-) \geq F_L(x_0) > \alpha$, so $\mathrm{VaR}_\alpha^+(L) \leq x_0$. We conclude that $\mathrm{VaR}_\alpha^+(L) = x_0$. The second case is $S_\alpha = \{x_0\}$. In this case $F_L(x_0) = \alpha$, $F_L(x) < \alpha$ for $x < x_0$ and $F_L(x) > \alpha$ for $x > x_0$. Again it is clear that $\mathrm{VaR}_\alpha(L) = x_0$. We also see that $(-\infty, x_0] \subseteq \{x : F_L(x) \leq \alpha\}$. If $x > x_0$ and $F_L(x-) = \alpha$, then for any $x' \in (x_0, x)$, we have $F_L(x') = \alpha$, contradicting that S_α is a singleton set. Hence it must hold that $F_L(x-) > \alpha$ for any $x > x_0$, and we conclude that $\mathrm{VaR}_\alpha^+(L) = x_0$. ∎

We can now present a result for the Value at Risk when applying a strictly decreasing function.

Proposition 1.3.15. *Let L be a loss variable and let g be strictly decreasing and continuous. Then for any $\alpha \in (0,1)$,*

$$\mathrm{VaR}_\alpha(g(L)) = g(\mathrm{VaR}^+_{1-\alpha}(L)).$$

In particular, for any α where $F_L(x) = 1-\alpha$ has either one or no solution, it holds that

$$\mathrm{VaR}_\alpha(g(L)) = g(\mathrm{VaR}_{1-\alpha}(L)).$$

Proof. Since g is strictly decreasing, g has a proper inverse. Furthermore, g maps intervals to intervals by continuity. Hence

$$\begin{aligned}\mathrm{VaR}_\alpha(g(L)) &= \inf\{x : P(g(L) \le x) \ge \alpha\} = \inf\{x : P(L \ge g^{-1}(x)) \ge \alpha\} \\ &= \inf\{x : F_L(g^{-1}(x)) \le 1-\alpha\} = \inf\{g(y) : F_L(y-) \le 1-\alpha\} \\ &= g(\sup\{y : F_L(y-) \le 1-\alpha\}) = g(\mathrm{VaR}^+_{1-\alpha}(L)).\end{aligned}$$

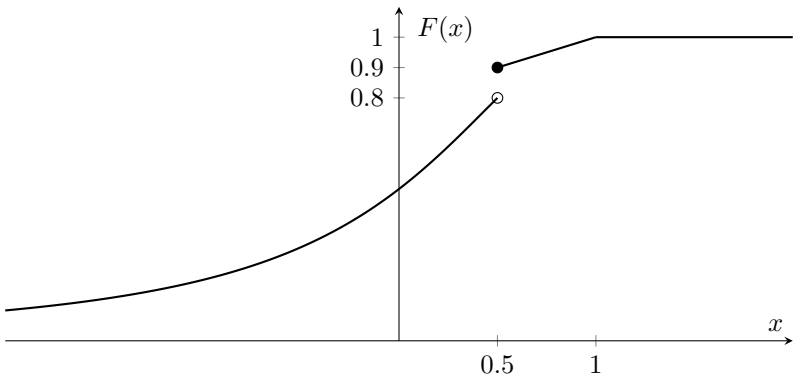
The second part follows immediately from the previous lemma. ■

In the exercises you are asked to prove the results presented here in some nice special cases. Let us now consider two examples.

Example 1.3.16. Assume that the log return X of a stock follows the distribution function

$$F_X(x) = \begin{cases} \frac{1}{1+(x-1)^2}, & x \in (-\infty, 1/2) \\ \frac{x+4}{5}, & x \in [1/2, 1) \\ 1, & x \in [1, \infty) \end{cases}$$

The distribution function looks as follows:



From the figure, we see that we need to consider three cases for α. If $\alpha \in (0, 4/5)$, we need to solve $1/(1+(x-1)^2) = \alpha$ for x, yielding $x = 1 \pm \sqrt{1/\alpha - 1}$. What is the correct sign? If we choose $+$, we get $x > 1$ which is clearly not the case from looking at the drawing. Hence the correct sign must be $-$. When $\alpha \in [4/5, 9/10]$, we have $\mathrm{VaR}_\alpha(X) = 1/2$. Finally,

when $\alpha \in (9/10, 1)$, we have to solve $(x+4)/5 = \alpha$ for x, yielding $\mathrm{VaR}_\alpha(X) = 5\alpha - 4$. We conclude that

$$\mathrm{VaR}_\alpha(X) = \begin{cases} 1 - \sqrt{\frac{1}{\alpha} - 1}, & \alpha \in (0, 4/5) \\ \frac{1}{2}, & \alpha \in [4/5, 9/10] \\ 5\alpha - 4, & \alpha \in (9/10, 1) \end{cases}.$$

The loss for the stock is given by $L = -s(e^X - 1)$ with $s > 0$ some known initial value. L is a strictly decreasing and continuous transformation of X. We also see that $F_X(x) = \alpha$ has at most one solution for any α, and so the above proposition applies to show that

$$\mathrm{VaR}_\alpha(L) = -s(e^{\mathrm{VaR}_{1-\alpha}(X)} - 1), \quad \alpha \in (0, 1).$$

In particular, if $\alpha = 0.95$, $1 - \alpha = 0.05$ so

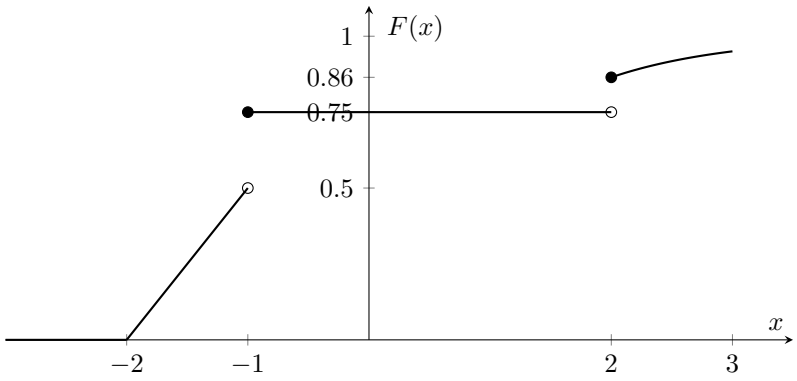
$$\mathrm{VaR}_\alpha(L) = -s(e^{1-\sqrt{1/\alpha - 1}} - 1).$$

$\circ$

Example 1.3.17. Assume that the log return X of a stock follows the distribution function

$$F_X(x) = \begin{cases} 0, & x \in (-\infty, -2) \\ \frac{x}{2} + 1, & x \in [-2, -1) \\ \frac{3}{4}, & x \in [-1, 2) \\ 1 - e^{-x}, & x \in [2, \infty) \end{cases}$$

A drawing of the distribution function is as follows:



A bit of thinking (do this!) shows that

$$\mathrm{VaR}_\alpha(X) = \begin{cases} 2(\alpha - 1), & \alpha \in (0, 1/2) \\ -1, & \alpha \in [1/2, 3/4] \\ 2, & \alpha \in (3/4, 1 - e^{-2}) \\ -\log(1 - \alpha), & \alpha \in [1 - e^{-2}, 1) \end{cases}$$

We have by Proposition 1.3.15 that for the loss $L = -s(e^X - 1)$, we have

$$\mathrm{VaR}_\alpha(L) = -s(e^{\mathrm{VaR}_{1-\alpha}} - 1), \quad \alpha \neq \frac{1}{4}.$$

As for the case $\alpha = 1/4$, we have $\mathrm{VaR}_{1-\alpha}(X) = \mathrm{VaR}_{3/4}(X) = -1$ while $\mathrm{VaR}^+_{1-\alpha}(X) = 2$. Hence

$$\mathrm{VaR}_{1/4}(L) = -s(e^2 - 1) \neq -s(e^{-1} - 1) = -s(e^{\mathrm{VaR}_{1-1/4}(X)} - 1).$$

Note that the sign even shifts. The correct Value at Risk at $\alpha = 1/4$ for L is negative (indicating a gain), while it is positive (a loss!) if one simply plugged in $\mathrm{VaR}_{3/4}(X)$ instead of $\mathrm{VaR}^+_{3/4}(X)$. ∘

1.4 Computational methods

Estimating VaR and ES from data

We now consider methods for computing VaR and ES directly from data. Up to time t_n we have the empirical observations $L_1, ..., L_n$. For this discussion, we make the (not so realistic) assumption that $L_1, ..., L_n$ is an iid sample. Order the sample and let

$$L_{1,n} \geq L_{2,n} \geq ... \geq L_{n,n}$$

denote the corresponding order statistics. We define the empirical distribution function

$$F_L^{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{[L_i,\infty)}(x)$$

that assigns equal weight to each observation. To estimate the risk measures of interest, an idea is to replace F_L (which is unknown) with the empirical distribution function $F_L^{(n)}$. This idea is justified by e.g. the Strong Law of Large Numbers which implies that

$$F_L^{(n)}(x) \to P(L_1 \leq x) \quad \text{a.s. as} \quad n \to \infty$$

for each x or the even stronger result known as the *Glivenko–Cantelli Theorem* which states that

$$\sup_{x \in \mathbb{R}} |F_L^{(n)}(x) - F_L(x)| \to 0 \quad \text{a.s. as} \quad n \to \infty.$$

We can form the natural estimator of the Value at Risk given by

$$\widehat{\mathrm{VaR}}_\alpha(L) = \inf\{x \in \mathbb{R} : 1 - F_L^{(n)} \leq 1 - \alpha\}.$$

A problem with this approach is that it is often very difficult to infer the tail behaviour of F_L from $F_L^{(n)}$ since we often do not have enough data in the tail of F_L. However, one nice thing about this approach is that we can explicitly solve for the estimator. We have

$$\widehat{\mathrm{VaR}}_\alpha(L) = L_{[n(1-\alpha)]+1,n} = \widehat{q}_\alpha(F_L)$$

i.e. the empirical quantile. Here $[n(1-\alpha)]$ denotes the largest integer less than or equal to $n(1-\alpha)$. Similarly, if F_L is continuous and strictly increasing, we may compute an estimate for $\mathrm{ES}_\alpha(L)$, namely

$$\widehat{\mathrm{ES}}_\alpha(L) = \frac{\sum_{i=1}^{[n(1-\alpha)]+1} L_{i,n}}{[n(1-\alpha)] + 1}$$

i.e. the average of the observations greater than or equal to $\widehat{\mathrm{VaR}}_\alpha(L)$. Again we stress that this approximation is often insufficient since we usually do not have many observations in

the tails of F_L. We now turn to the problem of computing confidence intervals where we focus on the Value at Risk. Let $\beta \in (0,1)$ denote some "small" value. Our approach is to find $\widehat{A}$ and $\widehat{B}$ such that

$$P(\widehat{A} < \text{VaR}_\alpha(L) < \widehat{B}) \geq 1 - \beta$$

by determining $\widehat{A}$ and $\widehat{B}$ in such a way that

$$P(\text{VaR}_\alpha(L) \leq \widehat{A}) \leq \frac{\beta}{2} \quad \text{and} \quad P(\text{VaR}_\alpha(L) \geq \widehat{B}) \leq \frac{\beta}{2}.$$

Assume that L has a density so that for the true $\text{VaR}_\alpha(L)$ we have $P(L > \text{VaR}_\alpha(L)) = 1-\alpha$. For the observed data, each data point can land on either side of $\text{VaR}_\alpha(L)$. Hence we have a sequence of Bernoulli trials with probability $1 - \alpha$ of landing to the right of $\text{VaR}_\alpha(L)$ (a success) and probability α of landing to the left of $\text{VaR}_\alpha(L)$ (a failure). Formally, let $Z = 1_{\{L > \text{VaR}_\alpha(L)\}}$, then

$$q := P(Z = 0) = \alpha, \quad p := P(Z = 1) = 1 - \alpha.$$

Let Y denote the number of successes in n trials i.e. $Y = \sum_{i=1}^n Z_i$ for $Z_i = 1_{\{L_i > \text{VaR}_\alpha(L)\}}$ where L_i is the ith loss. Then $Y \sim \text{Bin}(n, p)$ and

$$P(Y \geq j) = \sum_{k=j}^{n} \binom{n}{k} p^k q^{n-k}.$$

Note that $L_{j,n} \geq \text{VaR}_\alpha(L)$ if and only if $Y \geq j$ so using the above, we can find the smallest j such that $P(Y \geq j) \leq \beta/2$. Then $P(L_{j,n} \geq \text{VaR}_\alpha(L)) = P(Y \geq j) \leq \beta/2$ and so we set $\widehat{A} = L_{j,n}$. Conversely, we can find the largest i such that $P(Y \leq i) \leq \beta/2$. Then $P(L_{i,n} \leq \text{VaR}_\alpha(L)) \leq \beta/2$ so set $\widehat{B} = L_{i,n}$.

Other methods for computing VaR

Several methods exist for computing the Value at Risk. We have just seen how one estimates VaR from a sample of losses. We now introduce four additional methods for computing VaR that are used in practice. These are the following:

(i) The Variance-Covariance (Var-Cov) method,

(ii) Monte Carlo simulation,

(iii) Importance Sampling and

(iv) Bootstrapping.

The motivating example to keep in mind is the one with d investments from earlier. Recall that the loss was given by

$$L_{n+1} = -\sum_{i=1}^{d} \alpha_i S_n^{(i)} \left(e^{X_{n+1}^{(i)}} - 1 \right)$$

with $S_n^{(i)}$ the value of the ith asset at time t_n, $X_{n+1} = \log S_{n+1} - \log S_n$ the log return and α_i the number of assets bought of asset i. We now go through the different methods.

The Var-Cov method

Consider the linearized loss

$$L_{n+1}^{\Delta} = -\sum_{i=1}^{d} \alpha_i S_n^{(i)} X_{n+1}^{(i)}$$

obtained by the Taylor approximation $e^x \approx 1 + x$. In the Variance-Covariance method, we assume that $\mathbf{X}_{n+1} \sim \mathcal{N}(\mathbf{m}, \Sigma)$, that is, a multivariate normal distribution with mean vector $\mathbf{m} \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. Letting

$$\mathbf{X}_{n+1} = \begin{pmatrix} X_{n+1}^{(1)} \\ \vdots \\ X_{n+1}^{(d)} \end{pmatrix}, \quad \mathbf{w}_n = \begin{pmatrix} \alpha_1 S_n^{(1)} \\ \vdots \\ \alpha_d S_n^{(d)} \end{pmatrix},$$

we may rewrite $L_{n+1}^{\Delta} = -\langle \mathbf{w}_n, \mathbf{X}_{n+1} \rangle = -\mathbf{w}_n^T \mathbf{X}_{n+1}$. By the properties of the multivariate normal distribution (see the appendix), we have

$$L_{n+1}^{\Delta} \sim \mathcal{N}(-\mathbf{w}_n^T \mathbf{m}, \mathbf{w}_n^T \Sigma \mathbf{w}_n).$$

Let $\mu_n = -\mathbf{w}_n^T \mathbf{m}$ and $\sigma_n^2 = \mathbf{w}_n^T \Sigma \mathbf{w}_n$, so that we may write $L_{n+1}^{\Delta} \overset{\mathrm{d}}{=} \mu_n + \sigma_n Z$ for $Z \sim \mathcal{N}(0,1)$. From last week, we can compute $\mathrm{VaR}_\alpha(L_{n+1}^{\Delta})$ as

$$\mathrm{VaR}_\alpha(L_{n+1}^{\Delta}) = \mu_n + \sigma_n \Phi^{-1}(\alpha).$$

Finally, we can use data to obtain estimates of μ and σ^2 i.e. of $\mathbf{m}$ and Σ. A virtue of this method is how simple it is to use. We get an exact analytical expression for $\mathrm{VaR}_\alpha(L_{n+1}^{\Delta})$. The problem is the approximation $e^x \approx 1 + x$ behind the method. This is not very precise for large losses, and often we are interested in the tail of the distribution where x is large. Also, the normality assumption is often problematic with real data.

Monte Carlo simulation

Monte Carlo simulation is a purely computational method. Suppose we simulate N samples of $\mathbf{X}_{n+1}$ and call these $\mathbf{x}_1, ..., \mathbf{x}_N$. We work with these simulated values as if they were empirical samples. From these we can form "empirical" samples of L_{n+1}. Call these $l_1, ..., l_N$. We use these to compute the empirical Value at Risk. Formally, start by ordering the samples to obtain the order statistics

$$l_{1,N} \geq ... \geq l_{N,N}.$$

Then we can compute the estimate

$$\mathrm{VaR}_\alpha(L_{n+1}) \approx \widehat{\mathrm{VaR}}_\alpha(L_{n+1}) = l_{[N(1-\alpha)]+1,N}$$

per the discussion last week. This method is a lot more precise compared to the Variance-Covariance method since it does not rely on the approximation $e^x \approx 1 + x$. Another obvious advantage is flexibility. The method works for any distribution (at least if we can efficiently simulate large samples from that distribution). One problem is that the rate of convergence of the estimate will be slow since we are working with the tail of the loss distribution. Hence we often have to generate a very large number of samples to obtain a robust estimate. This is especially problematic if we have a large number of assets.

Rare event simulation

Before moving on to importance sampling, we take a brief detour and consider a problem closely related to the one above for the Monte Carlo method, namely estimating $p_x = P(L > x)$ for large values of x. To estimate p_x, we generate a computational iid sample $l_1, ..., l_N$ of L. We can then form the indicators $1_{\{l_1>x\}}, ..., 1_{\{l_N>x\}}$. By the SLLN,

$$\frac{1}{N}\sum_{i=1}^{N} 1_{\{l_i>x\}} \to E[1_{\{L>x\}}] = P(L > x) = p_x \quad \text{a.s.}$$

Define the natural estimator

$$\widehat{p}_x^{(N)} = \frac{1}{N}\sum_{i=1}^{N} 1_{\{l_i>x\}}.$$

As shown above, this estimator is consistent (it converges in probability to the true underlying parameter, in this case p_x). A natural question to ask is how fast $\widehat{p}_x^{(N)}$ converges. If

$$S_N = \sum_{i=1}^{N} 1_{\{l_i>x\}},$$

then the CLT applies to show that

$$\frac{S_N - Np_x}{\sigma_x\sqrt{N}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

where σ_x denotes the variance of $1_{\{l_1>x\}}$. Note that $N\widehat{p}_x^{(N)} = S_N$ so for large N, we get (in distribution) that

$$Z \approx \frac{S_N - Np_x}{\sigma_x\sqrt{N}} = \frac{\widehat{p}_x^{(N)} - p_x}{\sigma_x/\sqrt{N}}.$$

We can rewrite this relation and obtain the approximation (in distribution)

$$\widehat{p}_x^{(N)} \approx p_x + \frac{\sigma_x}{\sqrt{N}}Z.$$

We can form an asymptotic confidence interval for p_x as follows. Let $\beta \in (0,1)$ be some "small" value and let $z_{\beta/2}$ denote the $\beta/2$-quantile of $\mathcal{N}(0,1)$ so that $P(Z \le z_{\beta/2}) = \beta/2$. Thus, with the "high" probability $1-\beta$, we have

$$p_x \in \left(\widehat{p}_x^{(N)} - \frac{\sigma_x}{\sqrt{N}}z_{1-\beta/2}, \widehat{p}_x^{(N)} + \frac{\sigma_x}{\sqrt{N}}z_{1-\beta/2}\right).$$

While the error $\sigma_x/\sqrt{N}$ goes to zero for $N \to \infty$, the probability p_x is often also small, so it can happen that the error still dominates the estimate $\widehat{p}_x^{(N)}$, even when N is very large. To make this precise, define the relative error

$$\text{RE} = \frac{\sigma_x z_{\beta/2}}{\sqrt{N}p_x} = \frac{\sigma_x}{p_x}C(N)$$

with $C(N)$ some constant depending on N. We compute

$$\sigma_x^2 = \text{Var}[1_{\{L>x\}}] = E[1_{\{L>x\}}^2] - E[1_{\{L>x\}}]^2 = p_x - p_x^2.$$

We now have for the relative error that

$$\frac{\sigma_x}{p_x} = \frac{\sqrt{p_x - p_x^2}}{p_x} = \sqrt{\frac{1}{p_x} - 1} \to \infty \quad \text{as} \quad x \to \infty.$$

So the relative error explodes as x gets large. Hence we need more sophisticated techniques so that the relative error is bounded. We present one such method very briefly now.

Importance sampling

To remedy the issue of diverging relative errors, we briefly introduce the main ideas of importance sampling. Consider the setup $L = f(\mathbf{X})$ where $\mathbf{X}$ is an $\mathbb{R}^d$-valued random variable with distribution function F (see the appendix for a brief discussion on multidimensional distribution functions) and f is a deterministic function. We refer to F as the *true* distribution of $\mathbf{X}$. Define the *moment-generating function* (mgf) of $\mathbf{X}$ as

$$\kappa(\xi) = E\left[e^{\langle \xi, \mathbf{X} \rangle}\right] \quad \text{for} \quad \xi \in \mathbb{R}^d.$$

Consider the *shifted distribution* given by

$$dF_\xi(x_1, ..., x_d) = \frac{e^{\langle \xi, \mathbf{x} \rangle}}{\kappa(\xi)} dF(x_1, ..., x_d)$$

for all $\xi \in \mathbb{R}^d$ such that the moment-generating function is finite. We note the following properties of F_ξ:

(i) F_ξ is a probability distribution.

(ii) $E_\xi[\mathbf{X}] = \nabla \Lambda(\xi)$ where $\Lambda(\xi) = \log \kappa(\xi)$ is the *cumulant-generating function* of $\mathbf{X}$ and E_ξ indicates the expectation taken with respect to the shifted measure.

Property (i) follows easily by the calculation

$$\int dF_\xi(\mathbf{x}) = \frac{1}{\kappa(\xi)} \int e^{\langle \xi, \mathbf{x} \rangle} dF(\mathbf{x}) = \frac{\kappa(\xi)}{\kappa(\xi)} = 1.$$

The idea is now to choose ξ in a good way such that $L > x$ occurs frequently i.e. so that $P_\xi(L > x)$ is large, where P_ξ denotes the probability under the shifted distribution. To relate simulations under the shifted distribution with parameter ξ to the original probability, we apply a representation formula,

$$\begin{aligned} p_x = P(L > x) &= \int_{\{\mathbf{y}: f(\mathbf{y}) > x\}} dF(\mathbf{y}) = \int_{\{\mathbf{y}: f(\mathbf{y}) > x\}} \frac{dF}{dF_\xi}(\mathbf{y}) dF_\xi(\mathbf{y}) \\ &= E_\xi\left[1_{\{f(\mathbf{X}) > x\}} \frac{dF}{dF_\xi}(\mathbf{X})\right] \end{aligned}$$

with $\frac{dF}{dF_\xi}$ the Radon–Nikodym derivative.

Bootstrapping

The idea of bootstrapping is to use the existing data to generate new data by resampling. We sample with replacement and if we have N data points $x_1, ..., x_N$, we choose a member with probability $1/N$ N times to get a new data set of the same size. We can do this procedure b times to obtain the bootstrap samples

$$\begin{gathered} x_1^{(1)}, ..., x_N^{(1)} \\ x_1^{(2)}, ..., x_N^{(2)} \\ \vdots \\ x_1^{(b)}, ..., x_N^{(b)}. \end{gathered}$$

If we have some parameter θ that we want to estimate, we can compute an empirical estimate $\widehat{\theta}$ from the original data and estimates $\widehat{\theta}_j$ from each of the b bootstrap samples. We can then use the $\widehat{\theta}_j$ to say something about the distribution of $\widehat{\theta}$. We may for example generate confidence intervals by computing empirical quantiles using the $\widehat{\theta}_j$. Explicitly, define the residuals $R_j = \widehat{\theta} - \widehat{\theta}_j$ and order them from largest to smallest, $R_{1,b} \geq R_{2,b} \geq ... \geq R_{b,b}$. The confidence bounds are then given by

$$\left[\widehat{\theta} + R_{[b(1-\beta/2)],b}, \widehat{\theta} + R_{[b(\beta/2)]+1,b}\right].$$

These bounds improve classical estimates based on the CLT which converge more slowly.

For small probabilities, one may consider a modification of this idea, namely *smoothed bootstrap*. In smoothed bootstrapping, one smoothes the data around the tail values. Namely, given the original sample $x_1, ..., x_N$, sample from the density

$$g(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h} K\left(\frac{x - x_i}{h}\right)$$

where K is a smooth function, e.g.

$$K(t) = \frac{1}{2\pi} e^{-t^2/2}.$$

One can then employ importance sampling to shift the smoothed distribution so that more samples will be in the tail.

1.5 More theory for coherent risk measures*

Representations for coherent risk measures via scenarios

It is reasonable to ask how one constructs a risk measure, in particular a coherent one. In this subsection, we provide an intuitively appealing recipe, which also turns out to be a characterisation. Following along the lines of [3], we make the simplifying assumption on the background space $(\Omega, \mathcal{F})$ that Ω is finite and that $\mathcal{F}$ is the power set of Ω so that all events are measurable.

Definition 1.5.1. Let $\mathcal{P}$ denote a set of probability measures on $(\Omega, \mathcal{F})$. We then define the mapping $\rho_{\mathcal{P}}$ on the set of loss variables by

$$\rho_{\mathcal{P}}(L) = \sup\{E_P[L] : P \in \mathcal{P}\}$$

where the subscript on E indicates that the expectation is computed with respect to the probability measure P.

One can think of the set $\mathcal{P}$ as a collection of scenarios that the company has to take into consideration. The different probability measures could for example each represent a certain economic scenario.

Lemma 1.5.2. *The function $\rho_{\mathcal{P}}$ in the above definition is a coherent risk measure.*

This lemma is a triviality. It turns out however, that every coherent risk measure on $(\Omega, \mathcal{F})$ can be constructed in this way. The rest of this subsection is dedicated to proving this fact and presenting some applications. The proof of the following proposition is rather technical and can be skipped at first reading.

Proposition 1.5.3. *Let ρ be a risk measure. Then ρ is coherent if and only if $\rho = \rho_{\mathcal{P}}$ for some set of probability measures $\mathcal{P}$ on $(\Omega, \mathcal{F})$.*

Proof. We have already seen that $\rho_{\mathcal{P}}$ is a coherent risk measure, so only the converse statement needs to be proven. To this end, we follow the argument in Proposition 10.1 of [62]. Let ρ be a coherent risk measure. It suffices to show that for every loss variable L_0, there exists a probability measure P such that $E_P[L] \leq \rho(L)$ for every L and $E_P[L_0] = \rho(L_0)$, as we may then choose $\mathcal{P}$ to be the set of these measures. Fix a loss variable L_0. By translation invariance and positive homogeneity of ρ, it holds that

$$\rho(aL_0 + b) = a\rho(L_0) + b. \tag{1.1}$$

for $a \geq 0$. Hence we may assume without loss of generality that $\rho(L_0) = 1$. Consider the set $U = \{L : \rho(L) < 1\}$. We leave it to the reader to verify that U is both open and convex. Since $L_0 \notin U$, we have by the Geometric Hahn–Banach theorem (see chapter 18 of [118] for background) that there exists a linear functional[2] λ separating L_0 from U in the sense that

$$\lambda(L) < \lambda(L_0) \quad \forall L \in U. \tag{1.2}$$

Choosing $L = 0$, we get that $\lambda(L_0) > 0$, so we may normalise to assume that $\lambda(L_0) = 1 = \rho(L_0)$. Hence we may rewrite (1.2) to

$$\rho(L) < 1 \quad \Rightarrow \quad \lambda(L) < 1.$$

Using monotonicity and translation invariance, we also have that $L \leq 0$ implies $\rho(L) \leq \rho(0) = 0$. Using that λ is linear, we get

$$c\lambda(L) = -\lambda(-cL) > -1, \quad \text{implying} \quad \lambda(L) \geq -\frac{1}{c}$$

for every $c > 0$. We conclude that λ is a positive functional. We now claim that $\lambda(1) = 1$. Inceed, if $c < 1$, then $\lambda(c) < 1$, so $\lambda(1) \leq 1$. If $c > 1$, we get $\rho(2L_0 - c) = 2 - c < 1$,

[2] A linear functional is a mapping from a space to the real or complex numbers. In this case a mapping $\mathcal{G} \to \mathbb{R}$ where $\mathcal{G}$ is the space of loss variables on $(\Omega, \mathcal{F})$.

so $\lambda(2L_0 - c) = 2 - c\lambda(1) < 1$ or $\lambda(1) > 1/c$. Hence $\lambda(1) = 1$. Combining the fact $\rho(L) < 1 \Rightarrow \lambda(L) < 1$ with (1.1), we may conclude

$$\rho(L) < c \quad \Rightarrow \quad \lambda(L) < c,$$

implying $\lambda(L) \leq \rho(L)$ for all L. Hence we may choose P by $P(A) = \lambda(1_A)$. ■

This result means intuitively that every coherent risk measure arises as a worst case method for a series of scenarios. Unfortunately, it is difficult to come up with a reasonable set of scenarios, and this may explain why a non-coherent risk measure like VaR is still so popular. To quote [115], "regulators like Value at Risk, because they can regulate it".

We now have the tools to verify that ES is indeed a coherent risk measure. We will need a strengthening of Proposition 1.3.4. Recall that $x^+ = \max\{x, 0\}$.

Proposition 1.5.4. *For $\alpha \in (0, 1)$, it holds that*

$$\begin{aligned}\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha} E[(L - \mathrm{Var}_\alpha(L))^+] + \mathrm{VaR}_\alpha(L) \\ &= \frac{1}{1-\alpha}\Big(E[L; L > \mathrm{VaR}_\alpha(L)] + \mathrm{VaR}_\alpha(L)(1 - \alpha - P(L > \mathrm{VaR}_\alpha(L)))\Big).\end{aligned}$$

Proof. If U is uniformly distributed on $(0, 1)$, then $F_L^{\leftarrow}(U)$ has the same distribution as L. Thus,

$$\begin{aligned}\frac{1}{1-\alpha} E[(L - \mathrm{VaR}_\alpha(L))^+] &= \frac{1}{1-\alpha} \int_0^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha))^+ du \\ &= \frac{1}{1-\alpha} \int_\alpha^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha)) du \\ &= \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L) du - \mathrm{VaR}_\alpha(L),\end{aligned}$$

which proves the first equality. As for the second, use that

$$E[(L - \mathrm{VaR}_\alpha(L))^+] = E[L; L > \mathrm{VaR}_\alpha(L)] - \mathrm{VaR}_\alpha(L) P(L > \mathrm{VaR}_\alpha(L))$$

and plug this into the first expression. ■

Remark 1.5.5. If F_L is continuous and strictly increasing, we have $P(L > \mathrm{VaR}_\alpha(L)) = P(L \geq \mathrm{VaR}_\alpha(L)) = 1 - \alpha$ and we recover the result of Proposition 1.3.4.

Recall that a Q is absolutely continuous with respect to P if $P(A) = 0$ implies $Q(A) = 0$ for every measurable set A. In that case, we write $Q \ll P$. Recall also the celebrated Radon–Nikodym Theorem. For probability measures P and Q, it holds that $Q \ll P$ if and only if Q has density with respect to P (see Theorem 4.11 in [38]). We denote this density by dQ/dP. This density is often called the Radon–Nikodym derivative.

Theorem 1.5.6. *Let $\mathcal{G}$ denote the set of loss variables on the probability space $(\Omega, \mathcal{F}, P)$. For $\alpha \in (0, 1)$, we have the following expression for the Expected Shortfall,*

$$\mathrm{ES}_\alpha(L) = \sup\{E_Q[L] : Q \in \mathcal{P}_\alpha\}$$

where $\mathcal{P}_\alpha$ is the set of probability measures Q on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to P and for which $dQ/dP \leq (1 - \alpha)^{-1}$. In particular, Expected Shortfall is a coherent risk measure.

Proof. By translation invariance, we may assume without loss of generality that $\mathrm{VaR}_\alpha(L) > 0$. We may also assume that $L \geq 0$, since otherwise we may replace L by L^+ (note that $\mathrm{ES}_\alpha(L) = \mathrm{ES}_\alpha(L^+)$). Define the coherent risk measure

$$\rho_\alpha(L) = \sup\{E_Q[L] : Q \in \mathcal{P}_\alpha\}.$$

By describing the measures in $\mathcal{P}_\alpha$ via their Radon–Nikodym derivatives, we may identify $\mathcal{P}_\alpha$ with the set of random variables $\{\psi : 0 \leq \psi \leq (1-\alpha)^{-1}, E[\psi] = 1\}$. Thus

$$\rho_\alpha(L) = \sup\{E[L\psi] : 0 \leq \psi \leq (1-\alpha)^{-1}, E[\psi] = 1\}.$$

Now apply the transform $\phi = (1-\alpha)\psi$, then

$$\rho_\alpha(L) = \frac{E[L]}{1-\alpha} \sup\Big\{E\Big[\frac{L}{E[L]}\phi\Big] : 0 \leq \phi \leq 1, E[\phi] = 1-\alpha\Big\}.$$

Define the measure $\tilde{P}$ by the Radon–Nikodym derivative $d\tilde{P}/dP = L/E[L]$. Then

$$\rho_\alpha(L) = \frac{E[L]}{1-\alpha} \sup\{E_{\tilde{P}}[\phi] : 0 \leq \phi \leq 1, E[\phi] = 1-\alpha\}.$$

We claim that the supremum is attained by the random variable $\phi_0 = 1_{\{L > \mathrm{VaR}_\alpha(L)\}} + k1_{\{L = \mathrm{VaR}_\alpha(L)\}}$ where $k \geq 0$ is a constant chosen such that $E[\phi_0] = 1-\alpha$. Let $0 \leq \phi \leq 1$ be arbitrary with $E[\phi] = 1-\alpha$. It then holds that

$$0 \leq (\phi_0 - \phi)(L - \mathrm{VaR}_\alpha(L)).$$

Take expectations to get

$$\begin{aligned} 0 &\leq E[(\phi_0 - \phi)(L - \mathrm{VaR}_\alpha(L))] = E[L(\phi_0 - \phi)] - \mathrm{VaR}_\alpha(L)E[\phi_0 - \phi] \\ &= E[L(\phi_0 - \phi)] = E[L]E_{\tilde{P}}[\phi_0 - \phi] \end{aligned}$$

since the second term disappears due to $E[\phi_0] = E[\phi] = 1-\alpha$. It follows that $E_{\tilde{P}}[\phi_0 - \phi] \geq 0$ which shows that ϕ_0 is indeed a maximiser. Now insert ϕ_0 in the expression for $\rho_\alpha(L)$ to obtain

$$\begin{aligned} \rho_\alpha(L) &= \frac{E[L]}{1-\alpha} E_{\tilde{P}}[\phi_0] = \frac{1}{1-\alpha} E[L\phi_0] \\ &= \frac{1}{1-\alpha}\Big(E[L; L > \mathrm{VaR}_\alpha(L)] + k\,\mathrm{VaR}_\alpha(L)P(L = \mathrm{VaR}_\alpha(L))\Big). \end{aligned}$$

A bit of algebra reveals that

$$k = \frac{1-\alpha-P(L > \mathrm{VaR}_\alpha(L))}{P(L = \mathrm{VaR}_\alpha(L))}$$

with the convention that $0/0 = 0$. Inserting this expression into the above expression $\rho_\alpha(L)$ yields

$$\rho_\alpha(L) = \frac{1}{1-\alpha}\Big(E[L; L > \mathrm{VaR}_\alpha(L)] + \mathrm{VaR}_\alpha(L)(1-\alpha-P(L > \mathrm{VaR}_\alpha(L)))\Big)$$

which equals ES_α by the above proposition, and the proof is complete. ■

Risk measures and decision sets

In this supplementary subsection, we introduce a different perspective on risk management based on *acceptance sets* and show how it is equivalent to the approach with risk measures introduced above. Recall that we have a background space $(\Omega, \mathcal{F})$ and that $\mathcal{G}$ is the set of all mappings (loss variables) $\Omega \to \mathbb{R}$. We assume that Ω is finite and interpret Ω as all the possible states of nature (of the market). We consider a one-period model with timepoints $t_0 = 0$ (current time) and a future time t_1. Let us introduce some terminology and concepts.

Definition 1.5.7. An *acceptance set* $\mathcal{A}$ is a set of final net worths (a subset of $\mathcal{G}$).

One may interpret an acceptance set as a set of possible positions at the end of the period (t_1) that are acceptable without the need to set aside more capital. We can now describe a correspondence between risk measures and acceptance sets.

Definition 1.5.8. Given an acceptance set $\mathcal{A}$, we define the associated risk measure $\rho_{\mathcal{A}}$ by

$$\rho_{\mathcal{A}}(L) = \inf\{m \in \mathbb{R} : L - m \in \mathcal{A}\}.$$

Given a risk measure ρ, we define its associated acceptance set $\mathcal{A}_\rho$ by

$$\mathcal{A}_\rho = \{L : \rho(L) \leq 0\}.$$

It is natural to ask whether the above constructions are inverses of each other. The answer is no in general, but we can say more with a few additional well-known properties on the risk measure.

Proposition 1.5.9. *If ρ is a translation invariant and monotone risk measure, we have*

$$\rho = \rho_{\mathcal{A}_\rho}.$$

Proof. We have

$$\begin{aligned}\rho_{\mathcal{A}_\rho}(L) &= \inf\{m : L - m \in \mathcal{A}_\rho\} = \inf\{m : m\rho(L - m) \leq 0\} \\ &= \inf\{m : \rho(L) \leq m\} = \rho(L).\end{aligned}$$

■

Given an acceptance set $\mathcal{A}$, can we guarantee some nice properties of $\rho_{\mathcal{A}}$? The following result provides a partial answer. You are asked to prove it in the exercises.

Proposition 1.5.10. *Assume the acceptance set $\mathcal{A}$ satisfies the property*

$$L \in \mathcal{A} \quad \textit{and} \quad L' \leq L \quad \Rightarrow \quad L' \in \mathcal{A}$$

and that $\rho_{\mathcal{A}}(L)$ is finite for all L. Then $\rho_{\mathcal{A}}$ is monotone and translation invariant. Furthermore, $\mathcal{A} \subseteq \mathcal{A}_{\rho_{\mathcal{A}}}$.

One can further investigate when the sets $\mathcal{A}$ and $\mathcal{A}_{\rho_{\mathcal{A}}}$ are equal. In the case where all variables in $\mathcal{G}$ are bounded, equality holds if $\mathcal{A}$ is bounded in supremum norm. We refer to section 4.1 of [37] for details. It turns out that the coherence property of a risk measure translates nicely to properties of acceptance sets. We start by recalling some definitions.

Definition 1.5.11. Let C be a subset of $\mathcal{G}$.

(i) C is *convex* if for every $t \in [0,1]$ and $L, L' \in C$, we have $tL + (1-t)L' \in C$. A loss variable of the form $tL + (1-t)L'$ is called a *convex combination* of L and L'.

(ii) C is a *convex cone* if C is convex and it holds for any $t > 0$ and $L \in C$ that $tL \in C$.

Proposition 1.5.12. *Let ρ be a monotone and translation invariant risk measure. Then ρ is coherent if and only if $\mathcal{A}_\rho$ is a convex cone.*

Proof. Assume ρ is coherent. It is then easy to verify that $\mathcal{A}_\rho$ is convex. Take $L \in \mathcal{A}_\rho$. As ρ is positive homogeneous, we have for any $t > 0$ that $\rho(tL) = t\rho(L) \leq 0$ and thus $tL \in \mathcal{A}_\rho$, so that $\mathcal{A}_\rho$ is a convex cone. Assume conversely that $\mathcal{A}_\rho$ is a convex cone. We need to verify that ρ is subadditive and positive homogeneous. We start by verifying the latter. For $L \in \mathcal{G}$, we have $L - \rho(L)$ by definition of $\mathcal{A}_\rho$ and as $\mathcal{A}_\rho$ is a convex cone, $t(L - \rho(L)) \in \mathcal{A}_\rho$ for any $t > 0$. By translation invariance,

$$0 \geq \rho(tL - t\rho(L)) = \rho(tL) - t\rho(L),$$

proving that $\rho(tL) \leq t\rho(L)$. To show the opposite inequality, note that for $c < \rho(L)$, $L - c \notin \mathcal{A}_\rho$ and so $t(L-c) \notin \mathcal{A}_\rho$ for any $t > 0$. Hence

$$0 < \rho(tL - tc) = \rho(tL) - tc$$

showing that $\rho(tL) > tc$. Taking supremum over all $c < \rho(L)$, we obtain the inequality $\rho(tL) \geq t\rho(L)$, and we have shown that ρ is positive homogeneous. For subadditivity, let $L_1, L_2 \in \mathcal{G}$. Then $L_i - \rho(L_i) \in \mathcal{A}_\rho$ for $i = 1, 2$. Using convexity of $\mathcal{A}_\rho$ and translation invariance, we have

$$\begin{aligned} 0 &\geq \rho\Big(\frac{1}{2}(L_1 - \rho(L)) + \frac{1}{2}(L_2 - \rho(L_2))\Big) \\ &= \rho\Big(\frac{1}{2}(L_1 + L_2)\Big) - \frac{1}{2}(\rho(L_1) + \rho(L_2)), \end{aligned}$$

and by the positive homogeneity just shown, we have $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$ as desired. ∎

Notes and comments

Chapter 1 of [88] contains a longer informal introduction to the field of quantitative risk management as well as more historical perspectives. Section 2.3 contains more perspectives on different types of risk measurements. The supplementary part about scenario-based risk measures is also based on chapter 2 of this book. The article [3] is a classical reference on coherent risk measures. The reader should be aware that we use more modern notation and conventions more akin to those in [88]. [9] is a good reference on the Black–Scholes model and a very readable introduction to pricing of financial derivatives. The idea of bootstrapping was proposed in the famous paper by B. Efron, [28]. For more information on importance sampling and Monte Carlo methods, consult [46] and [4]. A good portion of the exercises below are from *The QRM Exercise Book*, [58].

1.6 Exercises

Theoretical exercises

Exercise 1.1:
Consider a loss variable L with $L \sim \text{Exp}(\lambda)$ for $\lambda > 0$.

1)Compute $\text{VaR}_\alpha(L)$.

2)Compute $\text{ES}_\alpha(L)$.

Exercise 1.2:
This exercise is from [63]. Consider a loss variable L with distribution function $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$ for $x \geq 0$ and $\gamma \in (0, 1)$.

1)Compute $\text{VaR}_\alpha(L)$.

2)Compute $\text{ES}_\alpha(L)$.

Exercise 1.3:
Assume that the loss variable L has a Pareto distribution with parameters $\kappa > 0$ and $\beta > 1$. This means that the distribution function is given by $F(x) = 1 - (\kappa/(\kappa + x))^\beta$ for $x \geq 0$.

1)Compute $\text{VaR}_\alpha(L)$.

2)Verify that

$$\text{ES}_\alpha(L) = \kappa\Big(\frac{\beta}{\beta - 1}(1 - \alpha)^{-1/\beta} - 1\Big).$$

Exercise 1.4:
Consider a loss variable L with distribution function F given by

$$F(x) = \begin{cases} 0, & \text{if } x < 1 \\ 1 - \frac{1}{1+x}, & \text{if } 1 \leq x < 3 \\ 1 - \frac{1}{x^2}, & \text{if } x \geq 3 \end{cases}.$$

1)Compute $\text{VaR}_{0.85}(L)$. Hint: Draw the graph of F.

2)Compute $\text{VaR}_{0.95}(L)$ and $\text{VaR}_{0.99}(L)$.

3)Compute $\text{ES}_{0.85}(L)$.

Exercise 1.5:
Let L be a loss variable with distribution function

$$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{s}}}$$

where $\mu \in \mathbb{R}$ and $s > 0$ parameters. This distribution is called the logistic distribution with location μ and scale s. Let $\alpha \in (0, 1)$.

1)Compute $\text{VaR}_\alpha(L)$.

2)Compute $\text{ES}_\alpha(L)$.

Exercise 1.6:
Consider a loss variable L with the Student t distribution with $\nu > 1$ degrees of freedom. That is, the density of L is given by

$$g_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

Show that

$$\mathrm{ES}_\alpha(L) = \frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1}\right).$$

Here, t_ν denotes the distribution function of L.

Exercise 1.7:
Consider the risk measure $\rho(L) = E[L]$. Show/convince yourself that ρ is a coherent risk measure. Explain why ρ may still be a bad risk measure.

Exercise 1.8:
Consider a binomial model for a stock with current value $S_t = 100$. Each year, the stock either increases in price by 4% with probability 0.8 or decreases in price by 4% with probability 0.2.

1)Determine the distribution of the loss of the stock after two years, $L_{t+2} := -(S_{t+2} - S_t)$. In particular, compute the distribution function.

2)Compute $\mathrm{VaR}_\alpha(L_{t+2})$ for $\alpha \in \{0.7, 0.95, 0.96, 0.99\}$.

Exercise 1.9:
The point of this exercise is to provide continuous examples of how subadditivity can fail for VaR.

1)Let L_1 and L_2 be iid normally distributed. Find an $\alpha \in (0,1)$ such that

$$\mathrm{VaR}_\alpha(L_1 + L_2) > \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2).$$

It is allowed to use software such as R to do numerical computations.

2)Construct another counterexample where L_1 and L_2 are *dependent*.

Exercise 1.10:
Let the log return X for a stock have the distribution function

$$F_X(x) = \begin{cases} \frac{1}{1+\log((x-2)^2+1)}, & x \in (-\infty, 1) \\ \frac{7}{10}, & x \in [1,2) \\ 1 + \frac{7}{10} - \frac{1}{1+(x-2)^2}, & x \in [2, x_F) \\ 1, & x \geq x_F \end{cases}$$

where x_F is the value where the function hits 1.

1)Draw F_X. You need only know that $1/(1+\log(2)) \approx 0.59$ and that $x_F \approx 2.65$.

2)Compute $\mathrm{VaR}_\alpha(X)$ for all $\alpha \in (0,1)$.

3)Now let $L = -s(e^X - 1)$ denote the loss for the stock where $s > 0$ is some fixed value. Compute $\mathrm{VaR}_\alpha(L)$ for all α. Hint: Be careful about a particular value of α!

Exercise 1.11:
Let the random variable X have distribution function

$$F(x) = \begin{cases} e^x, & x < -1 \\ 1 - \frac{1}{2(2+x)^2}, & x \geq -1 \end{cases}$$

1)Determine $\mathrm{VaR}_\alpha(X)$ for all $\alpha \in (0,1)$.

2)Assume X is the log return of a single stock so that the loss is

$$L = -s(e^X - 1)$$

where $s > 0$ is fixed. Determine $\mathrm{VaR}_\alpha(L)$ for all $\alpha \in (0,1)$.

Exercise 1.12:
Let L be a loss variable where F_L is continuous and strictly increasing. Let g be a continuous function.

1)Assume g is strictly increasing. Prove that $\mathrm{VaR}_\alpha(g(L)) = g(\mathrm{VaR}_\alpha(L))$.

2)Assume g is strictly decreasing. Prove that $\mathrm{Var}_\alpha(g(L)) = g(\mathrm{VaR}_{1-\alpha}(L))$.

Exercise 1.13:
In this exercise, we will study a risk measure which can be seen as a generalisation of the variance actuarial premium principle. Let $p > 1$ and let $\alpha \in (0,1]$ be a loading factor. We then define

$$\rho(L) = E[L] + \alpha \|(L - E[L])^+\|_p.$$

Recall that $x^+ = \max\{x, 0\}$ denotes the positive part of x and that $\|X\|_p = E[X^p]^{1/p}$ is the L^p-norm of X. In an actuarial context, we interpret this risk measure as the pure premium plus a risk loading determined by the loading factor α.

1)Check that ρ is translation invariant and positive homogeneous.

2)Show that ρ is subadditive.

3)Show that ρ is monotone and conclude that ρ is coherent.

4)Show the general fact that L^p-norms are increasing in p for $p \geq 1$. Use this to show that ρ is increasing in both α and p. Hint: Jensen's inequality.

Exercise 1.14:
For a loss random variable L, the quantity

$$\lim_{\alpha \to 1^-} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} \tag{1.3}$$

is called the *shortfall-to-quantile ratio*. By using the results derived in previous exercises and the text, compute the shortfall-to-quantile ratio in the following cases and provide an interpretation from a risk management perspective.

1)$L \sim \mathrm{Exp}(\lambda)$.

2)$L \sim \mathcal{N}(0,1)$.

3)L Pareto with distribution function $F(x) = 1 - (\kappa/(\kappa + x))^\beta$ for $x \geq 0, \beta > 1$.

Exercise 1.15: Matching VaR and ES
In this exercise we investigate when one has the relation $\mathrm{VaR}_\beta(L) = \mathrm{ES}_\alpha(L)$ for some $\alpha \leq \beta$.

1)Let L be Pareto distributed with distribution function $F(x) = 1 - (1+x)^{-\theta}$ for $x \geq 0$ and $\theta > 1$. Find (analytically) the value of $\alpha \in (0, \beta]$ such that $\mathrm{ES}_\alpha(L) = \mathrm{VaR}_\beta(L)$. Give the range of β values for which the equation has a solution. When a solution exists, α is given as an increasing function of θ. Why?

2)Now let $L \sim \mathcal{N}(0,1)$. Determine a value (using a numerical method) of α such that $\mathrm{ES}_\alpha(L) = \mathrm{Var}_{0.99}(L)$.

3)Repeat the previous subproblem under the assumption that $L \sim t_3$.

Exercise 1.16:
In this and the next exercise, we investigate the concept of *superadditivity* for VaR_α i.e. when it holds that

$$\mathrm{VaR}_\alpha(L_1 + L_2) > \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2) \tag{1.4}$$

for loss variables L_1 and L_2.

1)Show that VaR_α is superadditive if and only if

$$\alpha > F_{L_1+L_2}(F^{\leftarrow}_{L_1}(\alpha) + F^{\leftarrow}_{L_2}(\alpha)) \tag{1.5}$$

where F_L denotes the distribution function of a loss variable L.

In the rest of the exercise, we consider the case of two iid $\mathrm{Exp}(\lambda)$ variables L_1 and L_2.

2)Show that the density of $L_1 + L_2$ is given by

$$f_{L_1+L_2}(x) = \lambda^2 x e^{-\lambda x}, \quad x \geq 0. \tag{1.6}$$

Hint: Recall the general *density convolution formula*

$$f_{X+Y}(z) = \int_0^z f_X(t) f_Y(z-t) dt \tag{1.7}$$

for independent non-negative random variables X and Y.

3)Show that the distribution function of $L_1 + L_2$ is

$$F_{L_1+L_2}(x) = 1 - e^{-\lambda x}(1 + \lambda x), \quad x \geq 0. \tag{1.8}$$

4)Show that VaR_α in this case is superadditive if and only if

$$\alpha > 1 - (1-\alpha)^2(1 - 2\log(1-\alpha)). \tag{1.9}$$

5)By using a numerical root-finding algorithm, show that VaR_α is superadditive in this case if and only if $\alpha \in (0, \alpha_0)$ where $\alpha_0 \approx 0.7153$.

Exercise 1.17:
Let L_1 and L_2 be iid Pareto distributed with distribution function $F(x) = 1 - x^{-1/2}$ for $x \geq 1$. Note that this distribution is too heavy-tailed to have finite expectation. Let $f_{L_1+L_2}$ denote the density of $L_1 + L_2$.

1)By using the density convolution formula from the previous exercise along with the substitution $u = t - x/2$, show that

$$f_{L_1+L_2}(x) = \frac{1}{4}\int_{1-x/2}^{x/2-1} \frac{1}{(x^2/4-u^2)^{3/2}}du. \tag{1.10}$$

2)Use the substitution $s = \arcsin(2u/x)$ to show that

$$f_{L_1+L_2}(x) = \frac{x-2}{x^2\sqrt{x-1}}, \quad x \geq 2. \tag{1.11}$$

3)Conclude that

$$F_{L_1+L_2}(x) = 1 - \frac{2\sqrt{x-1}}{x}, \quad x \geq 2. \tag{1.12}$$

4)Show that

$$\mathrm{VaR}_\alpha(L_1+L_2) = 2\frac{1+\sqrt{1-(1-\alpha)^2}}{(1-\alpha)^2}. \tag{1.13}$$

5)Show that VaR_α in this case is superadditive for all $\alpha \in (0,1)$, that is,

$$\mathrm{VaR}_\alpha(L_1+L_2) > \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2) \tag{1.14}$$

for all $\alpha \in (0,1)$.

Exercise 1.18: Median shortfall
Let L be a loss random variable with continuous distribution function F_L. For a confidence level $\alpha \in [0,1)$, define the new risk measure

$$\mathrm{MS}_\alpha(L) = F^{\leftarrow}_{L|L>\mathrm{VaR}_\alpha(L)}(1/2), \tag{1.15}$$

denoted the *median shortfall*. Here $F_{L|L>\mathrm{VaR}_\alpha(L)}$ is the distribution function of L conditional on the event $L > \mathrm{VaR}_\alpha(L)$.

1)Compute $F_{L|L>\mathrm{VaR}_\alpha(L)}$ in terms of F_L.

2)Prove that

$$\mathrm{MS}_\alpha(L) = \mathrm{VaR}_{(1+\alpha)/2}(L). \tag{1.16}$$

What is the interpretation?

3)Now let L have the Pareto distribution with distribution function $F(x) = 1-(\kappa/(\kappa+x))^\beta$ for $\beta > 1$ and $x \geq 0$. Compute the limit

$$\lim_{\alpha\to 1^-} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{MS}_\alpha(L)} \tag{1.17}$$

and provide an interpretation of the result in a risk management context.

Exercise 1.19:
Consider the Standard Black–Scholes model and the derivative that pays $X = \log S_T$ at time T. Determine the arbitrage free price of this derivative at time $t < T$. Assume the

natural filtration generated by the Brownian motion. Hint: It may be helpful to consult the subsection in the appendix on stochastic processes.

Exercise 1.20:
For the shifted distribution

$$dF_\xi(x_1, ..., x_d) = \frac{e^{\langle \xi, \mathbf{x} \rangle}}{\kappa(\xi)} dF(x_1, ..., x_d)$$

as presented in the subsection on importance sampling above, prove that

$$E_\xi[\mathbf{X}] = \nabla \Lambda(\xi).$$

Exercise 1.21:
In this exercise, we fill in the gaps of the proof of Proposition 1.5.3. Let ρ be a coherent risk measure.

1)Show that the set $U = \{L : \rho(L) < 1\}$ is open. Hint: Consider $\varepsilon := (1 - \rho(L))/2$.

2)Show that U is convex.

Exercise 1.22:
Prove Proposition 1.5.10.

Practical exercises

Exercise 1.23:
In this exercise, you need the dataset of DAX-returns from Absalon. The dataset consists of daily log returns of the DAX stock index over the period from January 1990 until July 1996. Assume that today's price of the stock-index is $S_n = 100$.

1)Estimate the empirical $\text{VaR}_{0.99}$ and $\text{ES}_{0.99}$ for the one-day loss of a portfolio consisting of one share of the DAX index.

2)Construct an exact 95% confidence interval for $\text{VaR}_{0.99}$ using the method described on page 18.

3)Now compute the same quantities for the 10-day loss. This can, of course, be done in more than one way. Suggest a second approach, and discuss the advantages and disadvantages of your alternative approach.

Exercise 1.24:
In this exercise, we use importance sampling on a specific example. Consider a company which invests equal amounts in three stocks, all assumed to have normally distributed log returns that are independent over time. It is estimated that

$$\mathbf{X}_{n+1} = \log S_{n+1} - \log S_n \sim \mathcal{N}(\boldsymbol{\mu}, S),$$

where

$$1000\boldsymbol{\mu} = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad 1000S = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{4} & 1 & -\frac{1}{5} \\ \frac{1}{3} & -\frac{1}{5} & 1 \end{pmatrix}.$$

1)We are interested in studying the one-period loss from these investments, given that a capital of 1000 is invested in each stock at the current time. By simulating $\mathbf{X}_{n+1}$, compute $\text{VaR}_{0.9999}$ with a confidence level of 99%. Hint: You need a very large sample size to get adequate results.

2)Next suppose we would like to estimate

$$P(L_{n+1} \geq \text{VaR}_{0.9999})$$

more accurately, so we use importance sampling. Instead of sampling $\mathbf{X}_{n+1}$ directly, we simulate according to the probability distribution

$$dF_\xi(x_1, x_2, x_3) = Ce^{\langle \xi, \mathbf{x} \rangle} dF(x_1, x_2, x_3)$$

where F is the distribution of $\mathbf{X}_{n+1}$ and C is a normalisation constant. Prove that F_ξ is $\mathcal{N}(\mu + S\xi, S)$ distributed.

3)By experimenting with some possible choices, find a choice of ξ for which the event $\{L_{n+1} \geq \text{VaR}_{0.9999}\}$ occurs relatively frequently, say with probability 1/2 approximately. Now compute $P(L_{n+1} \geq \text{VaR}_{0.9999})$ with an approximate confidence level of 99%. What does this say about your estimate in (a)?

4)Finally, suppose that the real goal is to estimate the VaR (rather than the previous tail probability). To do so, we use importance sampling via the approach in the article [47]. We introduce the empirical tail distribution function

$$\overline{F}^{(N)}(u) = \frac{1}{N} \sum_{i=1}^{N} C^{-1} e^{-\langle \xi, \mathbf{X}_i \rangle} 1_{\{L_i > u\}},$$

where C and ξ are given as earlier and $\mathbf{X}_i$ has the distribution function F_ξ and where L_i is also computed with respect to F_ξ. It can be seen that $\overline{F}^{(N)}(u) \to \overline{F}(u)$ as $N \to \infty$ where F is the distribution function of L. Now use the empirical distribution function $\overline{F}^{(N)}$ to estimate $\text{VaR}_{0.9999}$. Hint: the R function `uniroot` can be used to solve for roots of univariate functions.

Chapter 2

Extreme value theory

2.1 Introduction

As we saw in the previous chapter, it is essential to be able to estimate high quantiles of data. Unfortunately, there is often very little data in the tails which makes the earlier presented methods inadequate for capturing the tail behaviour. Extreme value theory offers a variety of methods designed to deal with this issue. We start by briefly discussing a method for assessing a distributional assumption on the underlying data, namely the QQ plot. Afterwards, we introduce distributions with regularly varying tails and we discuss methods of doing statistics with such distributions. Highlights include the Hill estimator for the tail index of regularly varying distributions and the Peaks over threshold method.

Data analysis

Let $x_1, ..., x_n$ be observed data. If we want to use this data to make predictions about future values, it is natural to propose a distribution F and assume that $x_1, ..., x_n$ are realizations of iid variables $X_1, ..., X_n$ with distribution F. The QQ plot (quantile-quantile plot) can be used to determine whether F is a reasonable choice of distribution. The QQ plot consists of the points

$$\left\{ \left(F^{\leftarrow} \left(\frac{n-k+1}{n+1} \right), x_{k,n} \right) : k = 1, ...,n \right\}$$

where as usual, $x_{1,n} \geq ... \geq x_{n,n}$ denotes the order statistics of the sample. Recall that

$$F_n^{\leftarrow} \left(\frac{n-k+1}{n+1} \right) = x_{k,n}$$

where F_n is the empirical distribution function. If F is the true underlying distribution function for the data, $F_n \to F$ a.s., so that

$$x_{k,n} = F_n^{\leftarrow} \left(\frac{n-k+1}{n+1} \right) \approx F^{\leftarrow} \left(\frac{n-k+1}{n+1} \right).$$

Hence we expect the QQ plot to be a straight line through 0 and with slope 1 if the data truly comes from the reference distribution. The plot tells a bit more however. If the true underlying distribution is an affine transformation of F, the plot will still be a straight line. If that is the case, we can estimate proper location and scale parameters. The plot also

indicates whether the reference distribution has lighter or heavier tails than the empirical sample. See the plots below.

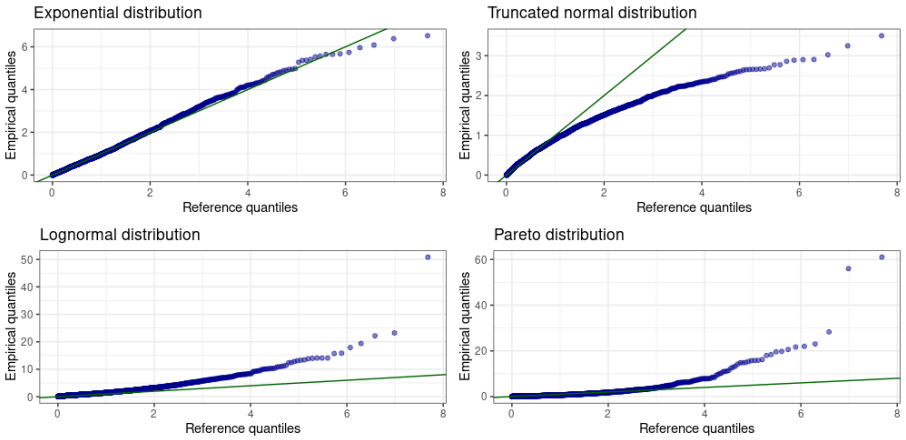


Figure 1: Four examples of QQ plots. We have 500 simulated values from the following distributions: Standard exponential (upper left), folded/truncated normal (i.e. $|X|$ for $X \sim \mathcal{N}(0,1)$) (upper right), standard lognormal (i.e. $\log X \sim \mathcal{N}(0,1)$, lower left) and the Pareto distribution with $\kappa = 1$ and $\alpha = 2$ (lower right). The reference distribution is standard exponential. The green line has slope one and intercept zero.

Consider the plots for a moment. The plot in the upper left corner shows that the data follows a straight line with slope one and intercept zero which is expected, since the underlying law of the empirical distribution equals the reference distribution. For the truncated normal, we see that the data curves downwards, indicating lighter tails than the exponential distribution. For the lognormal and Pareto distributions, the data curves upwards, indicating heavier tails than the exponential distribution.

Standard distributions in statistics include the normal, exponential and gamma distributions. These all have in common that they are light-tailed in the sense that the mgf of these variables exists in some neighbourhood around zero. This is not typical behaviour for log returns in a financial situation. These distributions are usually a lot more heavy-tailed. Heavy-tailed distributions include the regularly varying distributions such as the Pareto along with moderately heavy-tailed distributions such as the lognormal. In these cases, the mgf does not exist. We will now scratch the surface of the theory of regularly varying distributions.

2.2 Regular variation

Regular and slowly varying functions

Definition 2.2.1. A measurable function $h : (0,\infty) \to (0,\infty)$ is called *regularly varying* if there is a $\rho \in \mathbb{R}$ such that for all $t > 0$,

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\rho.$$

In this case, we write $h \in \mathrm{RV}_\rho$. If $h \in \mathrm{RV}_0$, we call h *slowly varying*.

Example 2.2.2. Constant functions and log are examples of slowly varying functions. ∘

The following characterisation of regularly varying functions will prove useful.

Proposition 2.2.3. *$h \in RV_\rho$ if and only if $h(x) = L(x)x^\rho$ for L a slowly varying function.*

Proof. Assume $h \in \mathrm{RV}_\rho$. Define the function $L(x) = h(x)x^{-\rho}$. Then for $t > 0$, we have

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = \lim_{x\to\infty} \frac{h(tx)(xt)^{-\rho}}{h(x)x^{-\rho}} = \lim_{x\to\infty} \frac{h(tx)}{h(x)} t^{-\rho} = t^\rho t^{-\rho} = 1$$

so L is slowly varying and satisfies $h(x) = L(x)x^\rho$. The converse implication is also easy and is left to the reader. ■

Definition 2.2.4. A distribution function F is *regularly varying* if $\overline{F} = 1 - F \in \mathrm{RV}_{-\alpha}$ for some $\alpha > 0$. α is called the *index* of F.

The definition says that F is regularly varying of index $\alpha > 0$ if for all $t > 0$, it holds that

$$\frac{P(X > tx)}{P(X > x)} \to t^{-\alpha} \quad \text{as} \quad x \to \infty$$

where X has distribution function F. Equivalently by the above proposition, $P(X > x) = L(x)x^{-\alpha}$ for a slowly varying function L. A celebrated result for regularly varying functions is the following result by Karamata.

Theorem 2.2.5 (**Karamata's Theorem**). *Let L be slowly varying and locally bounded in $[x_0, \infty)$ for some $x_0 \geq 0$.*

(a) If $\beta > -1$, then

$$\int_{x_0}^{x} t^\beta L(t)dt \sim \frac{1}{\beta+1} x^{\beta+1} L(x), \quad x \to \infty.$$

(b) If $\beta < -1$, then

$$\int_{x}^{\infty} t^\beta L(t)dt \sim -\frac{1}{\beta+1} x^{\beta+1} L(x), \quad x \to \infty.$$

In words, $\int_{x_0}^{u} t^\beta L(t)dt$ and $\int_{u}^{\infty} t^\beta L(t)dt$ behave asymptotically like the integral of a power function. The slowly varying function plays little role asymptotically. A proof of Karamata's theorem can be found in e.g. [102]. The following representation theorem is a direct consequence.

Corollary 2.2.6 (**Karamata's Representation Theorem**). *L is slowly varying if and only if we can write*

$$L(x) = c_0(x) e^{\int_{x_0}^{x} \frac{\epsilon(t)}{t} dt}, \quad x \geq x_0$$

where $c_0(x) \to c_0 > 0$ and $\epsilon(x) \to 0$ for $x \to \infty$ and $x_0 > 0$.

Proof. In the exercises, you will show that a function of the given form is slowly varying. For the converse implication, assume L is slowly varying. Define

$$b(x) := \frac{xL(x)}{\int_{x_0}^{x} L(t)dt}.$$

By part (a) of Karamata's Theorem for $\beta = 0$, we have $b(x) \to 1$ for $x \to \infty$. Define $\epsilon(x) = b(x) - 1$. We see that

$$L(x) = \frac{b(x)}{x} \int_{x_0}^{x} L(t)dt$$

and

$$\begin{aligned}
\int_{1+x_0}^{x} \frac{\epsilon(t)}{t}dt &= \int_{1+x_0}^{x} \frac{L(t)}{\int_{x_0}^{t} L(s)ds}dt - (\log(x) - \log(1+x_0)) \\
&= \int_{1+x_0}^{x} d\left(\log \int_{x_0}^{t} L(s)ds\right) - (\log(x) - \log(1+x_0)) \\
&= \log \int_{x_0}^{x} L(t)dt - \log \int_{x_0}^{1+x_0} L(t)dt - \log \frac{x}{1+x_0} \\
&= \log \frac{\int_{x_0}^{x} L(t)dt}{\int_{x_0}^{1+x_0} L(t)dt} - \log \frac{x}{1+x_0}.
\end{aligned}$$

Take the exponential function to get

$$\exp\left(\int_{1+x_0}^{x} \frac{\epsilon(t)}{t}dt\right) = \frac{1+x_0}{x} \frac{\int_{x_0}^{x} L(t)dt}{\int_{x_0}^{1+x_0} L(t)dt} = L(x) \frac{1+x_0}{b(x)\int_{x_0}^{1+x_0} L(t)dt}.$$

We conclude that the desired representation follows by letting

$$c_0(x) = \frac{b(x)\int_{x_0}^{1+x_0} L(t)dt}{1+x_0}.$$

■

The Generalised Pareto Distribution

The following distribution will play an essential role later on in the chapter.

Definition 2.2.7. The *Generalised Pareto Distribution* (GPD) has distribution function

$$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma}, \quad x > 0$$

where $\gamma > 0$ and $\beta > 0$ are parameters.

Lemma 2.2.8. *The Generalised Pareto Distribution with parameters $\beta, \gamma > 0$ is regularly varying with index $1/\gamma$.*

Proof. Simply note that

$$\overline{G}_{\gamma,\beta}(x) = \left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma}$$

so for $t > 0$, we have

$$\lim_{x\to\infty} \frac{\overline{G}_{\gamma,\beta}(tx)}{\overline{G}_{\gamma,\beta}(x)} = \lim_{x\to\infty} \left(\frac{1 + \frac{\gamma t x}{\beta}}{1 + \frac{\gamma x}{\beta}}\right)^{-1/\gamma} = \lim_{x\to\infty} \left(\frac{\frac{\beta}{\gamma x} + t}{\frac{\beta}{\gamma x} + 1}\right)^{-1/\gamma} = t^{-1/\gamma}.$$

■

The following proposition highlights a few additional properties of the GPD.

Proposition 2.2.9. *Assume X has a GPD with parameters $\gamma > 0$ and $\beta > 0$. Then $E[X] < \infty$ if and only if $\gamma < 1$. Furthermore,*

(i)

$$E\left[\left(1 + \frac{\gamma}{\beta}X\right)^{-r}\right] = \frac{1}{1+\gamma r}, \quad r > -\frac{1}{\gamma},$$

(ii)

$$E\left[\left(\log\left(1 + \frac{\gamma}{\beta}X\right)\right)^{k}\right] = \gamma^k k!, \quad k \in \mathbb{N},$$

(iii)

$$E\left[X\overline{G}_{\gamma,\beta}(X)^k\right] = \frac{\beta}{(k+1-\gamma)(k+1)}, \quad k+1-\gamma > 0.$$

(iv) For $x_1, x_2 > 0$, it holds that

$$\frac{\overline{G}_{\gamma,\beta}(x_1 + x_2)}{\overline{G}_{\gamma,\beta}(x_1)} = \overline{G}_{\gamma,\beta+\gamma x_1}(x_2).$$

(v) If $\gamma < 1/k$ with $k \in \mathbb{N}$, then

$$E[X^k] = \frac{\beta^k}{\gamma^{k+1}} \frac{\Gamma(1/\gamma - k)}{\Gamma(1 + 1/\gamma)} k!.$$

Proof. The proofs of (i) - (iv) are left to the reader, see Exercise 2.10. The proof of (v) goes as follows. The density of X is

$$g_{\gamma,\beta}(x) = -\frac{d}{dx}\overline{G}_{\gamma,\beta}(x) = \frac{1}{\beta}\left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma-1}.$$

Thus, plugging in the density of X and applying the substitution $u = \gamma x/\beta$, we get

$$\begin{aligned} E[X^k] &= \int_0^\infty x^k \frac{1}{\beta}\left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma-1} dx = \int_0^\infty \frac{\beta^k}{\gamma^k} u^k \frac{1}{\beta}(1+u)^{-1/\gamma-1}\frac{\beta}{\gamma} du \\ &= \frac{\beta^k}{\gamma^{k+1}} \int_0^\infty u^k (1+u)^{-1/\gamma-1} du. \end{aligned}$$

The trick is now to apply another substitution, namely $v = (1+u)^{-1}$. Then $u = 1/v - 1$ and $du = -v^{-2}dv$, so

$$\begin{aligned} E[X^k] &= \frac{\beta^k}{\gamma^{k+1}} \int_0^1 (1/v-1)^k v^{1/\gamma+1} v^{-2} dv = \frac{\beta^k}{\gamma^{k+1}} \int_0^1 \left(\frac{1-v}{v}\right)^k v^{1/\gamma-1} dv \\ &= \frac{\beta^k}{\gamma^{k+1}} \int_0^1 (1-v)^{k+1-1} v^{1/\gamma-k-1} dv. \end{aligned}$$

We recognise the integral as the beta function, and we conclude that

$$E[X^k] = \frac{\beta^k}{\gamma^{k+1}} \frac{\Gamma(k+1)\Gamma(1/\gamma-k)}{\Gamma(k+1+1/\gamma-k} = \frac{\beta^k}{\gamma^{k+1}} \frac{\Gamma(1/\gamma-k)}{\Gamma(1+1/\gamma)} k!.$$

■

It turns out that the GPD arises naturally when considering excesses over thresholds, as we shall see later when we discuss the POT method. The full justification for using the GPD in this way is postponed to the supplementary section at the end of the chapter.

2.3 The Hill estimator

Deriving the Hill estimator

From a statistical perspective, it is of interest to determine the tail parameter α if the underlying distribution of the data is assumed to have a regularly varying distribution. We present two such methods. In the following, we assume $\overline{F} \in \mathrm{RV}_{-\alpha}$ so that $\overline{F}(x) = L(x)x^{-\alpha}$, and the goal is to estimate α. A popular estimator is the Hill estimator, which we now derive. Let $\beta = -\alpha - 1$ where $\alpha > 0$. Karamata's Theorem implies

$$\int_u^\infty x^{-\alpha-1} L(x) dx \sim \frac{1}{\alpha} u^{-\alpha} L(u), \quad u \to \infty$$

so for $\overline{F}(x) = L(x)x^{-\alpha}$, we have

$$\int_u^\infty x^{-1} \overline{F}(x) dx \sim \frac{1}{\alpha} \overline{F}(u), \quad u \to \infty.$$

We rewrite the left hand side. First note that $(\log x - \log u)' = 1/x$. We can now apply integration by parts (see the appendix for a review) and obtain

$$\int_u^\infty x^{-1} \overline{F}(x) dx = \left[(\log x - \log u)\overline{F}(x)\right]_u^\infty - \int_u^\infty (\log x - \log u) d\overline{F}(x).$$

$\overline{F}(x)$ decays like $x^{-\alpha}$ and hence decays to zero faster than $\log x$ grows to ∞ and thus the first term is zero. By using that $d\overline{F}(x) = d(1-F(x)) = -dF(x)$, we get

$$\int_u^\infty x^{-1} \overline{F}(x) dx = \int_u^\infty (\log x - \log u) dF(x)$$

and so

$$\frac{1}{\overline{F}(u)} \int_u^\infty (\log x - \log u) dF(x) \to \frac{1}{\alpha}, \quad u \to \infty.$$

This is a theoretical result. To turn this into an estimator, we have to use the empirical distribution function. Suppose we have iid data $x_1, ..., x_n$ distributed according to F. Order the samples, $x_{1,n} \geq ... \geq x_{n,n}$. Let F_n denote the empirical distribution function. We can then approximate F by F_n. Replacing F by F_n in the above expression yields

$$\frac{1}{\alpha} \approx \frac{1}{\overline{F}_n(u)} \int_u^\infty (\log x - \log u) dF_n(x).$$

for sufficiently large u. We want to simplify this expression. Let N_u denote the number of observations greater than u i.e.

$$N_u = \#\{i : x_i > u\},$$

then $\overline{F}_n(u) = N_u/n$. Also recall that $\overline{F}_n(x_{k,n}) = (k-1)/n$. Now choose a "small" k and set $u = x_{k,n}$. Then

$$\begin{aligned}\frac{1}{\alpha} &\approx \frac{n}{k-1} \int_u^\infty (\log y - \log x_{k,n}) dF_n(y) = \frac{n}{k-1} \sum_{j=1}^k \frac{1}{n} (\log x_{j,n} - \log x_{k,n}) \\ &= \frac{1}{k-1} \sum_{j=1}^k (\log x_{j,n} - \log x_{k,n})\end{aligned}$$

since each jump of F_n is of size $1/n$ and the jth jump of F_n occurs at $x_{j,n}$ (see the appendix). Replacing $k-1$ by k gives us the Hill estimator.

Definition 2.3.1. Let $x_1, ..., x_n$ be a sample from a regularly varying distribution with index α. Let $x_{1,n} \geq ... \geq x_{n,n}$ denote the order statistics. We call

$$\widehat{\alpha}_k = \left(\frac{1}{k} \sum_{j=1}^k (\log x_{j,n} - \log x_{k,n}) \right)^{-1}$$

the *Hill estimator* of α.

Remark 2.3.2. Note that $\widehat{\alpha}_k$ depends on k i.e. the threshold.

Choosing the threshold - Hill plots

It is natural to ask how we choose a good value of k. Choosing k small, we get few data points which increases the variance of the estimator, so the estimate is not sufficiently robust. On the other hand, choosing k large makes the approximation based on Karamata's Theorem imprecise. Furthermore, it often happens with real data that the center and the tail have very different distributions, so taking too many data points close to the center makes the estimator biased. One often chooses k based on a *Hill plot*. A Hill plot consists of the value pairs

$$\{(k, \widehat{\alpha}_k) : k = 2, ..., n\}$$

and the value of k is chosen in a region where the estimator looks stable. We stress that this does not always occur for real data.

Example 2.3.3. A very classical data set in extreme value theory is the Danish fire insurance data. The data consists of large Danish fire insurance claims from 1980 to 1990. The data is available in the R package `evir` which has functions to compute estimates and make plots related to extreme value theory. We present some plots of the data below.

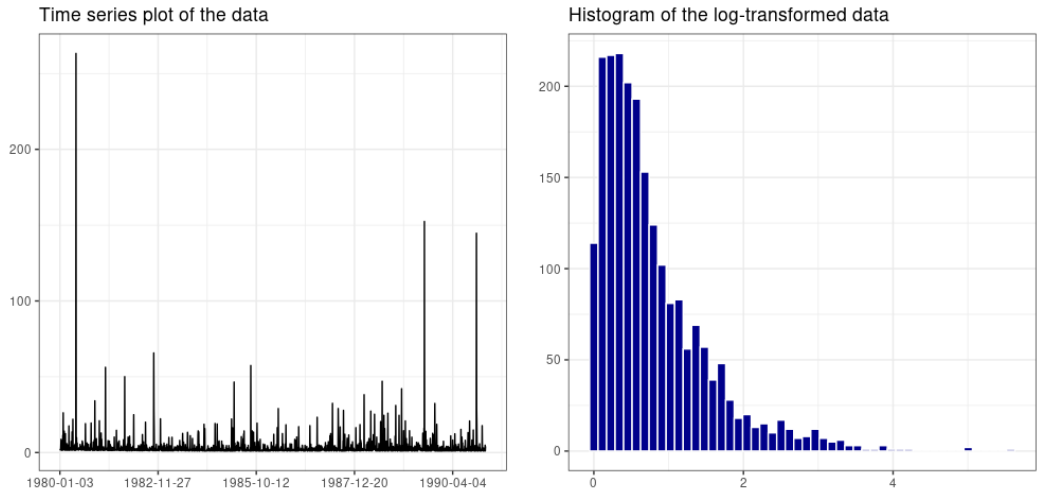


Figure 2: Exploratory plots of the Danish fire insurance data.

Below is a Hill plot of the data:

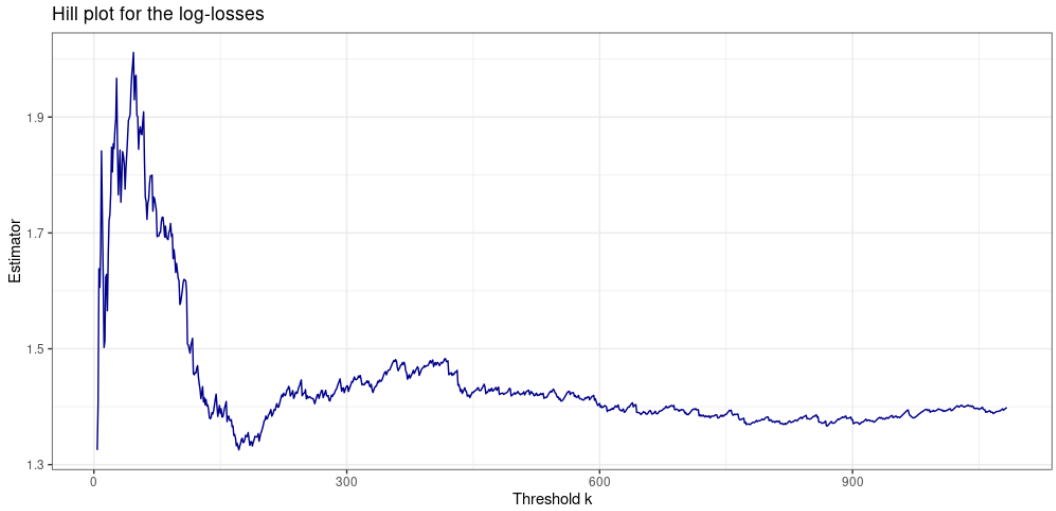


Figure 3: A Hill plot of the Danish fire insurance data.

The plot looks stable from around $k = 300$. We have $\widehat{\alpha}_{300} = 1.4357$, and this is one of many estimates we can choose to report. There is no single correct answer for the choice of k and very often, real life data is much more ugly than this particular data set. This illustrates the importance of applying different tools in extreme value theory before drawing a conclusion. ∘

Estimating quantiles with the Hill estimator

Using the Hill estimator, we can compute the Value at Risk at level β, VaR_β (the letter α is already used). We want to solve for x in the equation $P(X > x) = 1 - \beta$. Choose $k, x_{k,n}$ by looking at the Hill plot. Since X is regularly varying,

$$\overline{F}(x) = \overline{F}\left(\frac{x}{x_{k,n}} x_{k,n}\right) \sim \left(\frac{x}{x_{k,n}}\right)^{-\alpha} \overline{F}(x_{k,n}) \quad \text{as} \quad x \to \infty.$$

Now replace F by the empirical distribution function F_n. Since $\overline{F}_n(x_{k,n}) = (k-1)/n$, we obtain

$$\overline{F}(x) \approx \frac{k-1}{n} \left(\frac{x}{x_{k,n}}\right)^{-\widehat{\alpha}_k}.$$

Now set $\overline{F}(x) = 1 - \beta$ and solve for x. This solution is our estimate of $\mathrm{VaR}_\beta(X)$. You are asked to provide the details of this estimate in Exercise 2.11 where we also consider the Expected Shortfall.

Statistical properties of the Hill estimator

The statistical properties of the Hill estimator have been studied by several authors. The proofs of the following statements are quite technical and rely on a quite extensive background of theory for order statistics. To illustrate the depth needed to tackle these statistical problems, we have provided a path for proving the first assertion in Exercise 2.26 to 2.29 below. See also the notes and comments at the end of the chapter for references. In the following, the threshold k is treated as a function of n.

Theorem 2.3.4. *Let $\{X_n\}$ be an iid sequence of variables with a regularly varying distribution with index $\alpha > 0$. The following holds*

(i) ***(Weak consistency)****: If $k \to \infty$ and $k/n \to 0$, then*

$$\widehat{\alpha}_k \xrightarrow{P} \alpha.$$

(ii) ***(Strong consistency)****: If $k/n \to 0$ and $k/\log\log n \to \infty$, then*

$$\widehat{\alpha}_k \to \alpha \quad a.s.$$

(iii) ***(Weak convergence)****: Under suitable conditions on k and the distribution of X_1, it holds that*

$$\sqrt{k}(\widehat{\alpha}_k - \alpha) \xrightarrow{d} \mathcal{N}(0, \alpha^2).$$

Remark 2.3.5. One should be careful concerning the third assertion on weak convergence. The assumptions on the underlying data are unverifiable in practice, and one should expect an asymptotic bias to enter. For precise details on this asymptotic bias, see e.g. [52] and the references in the notes and comments at the end of the chapter.

Example 2.3.6. Let's consider the Danish fire insurance data again. Using part (iii) of Theorem 2.3.4, we can easily compute asymptotic confidence bands. The following figure shows the Hill plot from before but with these confidence bands added.

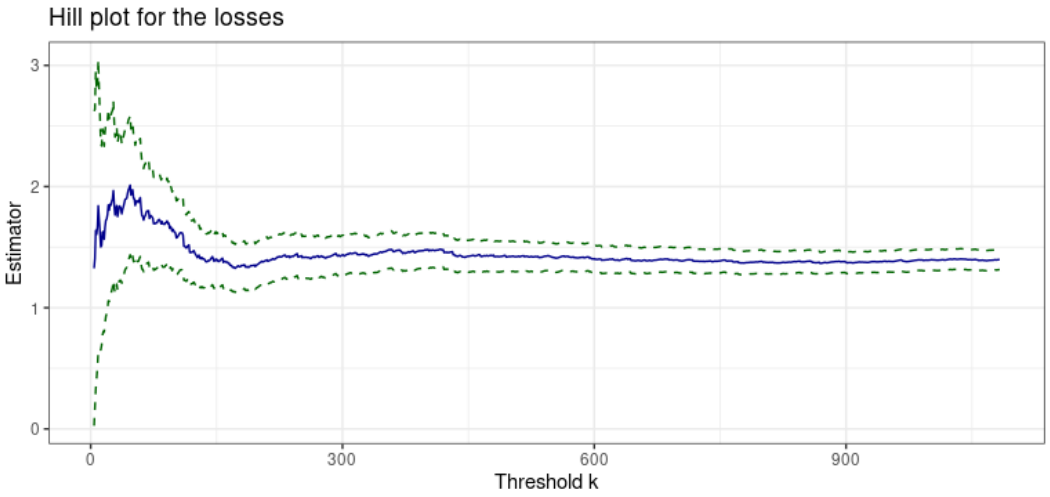


Figure 4: A Hill plot of the Danish fire insurance data with asymptotic 95% confidence bands.

∘

2.4 The Peaks over threshold (POT) method

Overview of the method

Real data often has two "components", namely a center component and a tail component. Often the tail of the data is described by a different distribution than the values close to the center. The following definition captures the idea of considering the tail component.

Definition 2.4.1. Given a distribution function F and a positive threshold u, we define the *excess distribution function* F_u via the tail

$$\overline{F}_u(x) = P(X > u + x \mid X > u) = \frac{\overline{F}(u+x)}{\overline{F}(u)}, \quad x \geq 0.$$

Rearranging, the above definition can be expressed as $\overline{F}(u+x) = \overline{F}(u)\overline{F}_u(x)$ for $x \geq 0$. If we have a sample $x_1, ..., x_n$ and $N_u = \#\{i : x_i > u\}$, then $\overline{F}(u) \approx N_u/n$. To estimate $\overline{F}_u(x)$, we apply the Generalised Pareto Distribution. To justify this, note that by the assumption of regular variation, we have

$$\frac{\overline{F}(tu)}{\overline{F}(u)} \to t^{-\alpha} \quad \text{for} \quad u \to \infty.$$

Hence for large u,

$$\overline{F}_u(x) = \frac{\overline{F}\left(\left(1+\frac{x}{u}\right)u\right)}{\overline{F}(u)} \approx \left(1+\frac{x}{u}\right)^{-\alpha}.$$

There is some cheating involved here. $t = 1 + x/u$ depends on u, but since $t \to 1$ as $u \to \infty$, the approximation is justified. Recall that the GPD with parameters $\beta, \gamma > 0$ has tail

$$\overline{G}_{\gamma,\beta}(x) = \left(1+\frac{\gamma x}{\beta}\right)^{-1/\gamma} = \left(1+\frac{x}{u}\right)^{-\alpha} \quad \text{for} \quad x \geq 0$$

where $\gamma = 1/\alpha$ and $\beta = \beta(u) = u/\alpha$ (while β depends on u, one chooses a fixed u so that β is also fixed). We can now describe the POT method in two steps:

(i) Estimate $\overline{F}(u) \approx N_u/n$.

(ii) Approximate $\overline{F}_u(x)$ by a GPD with parameters $\gamma, \beta > 0$.

There are two things we need to elaborate on concerning step (ii). First of all, we need to choose a threshold u. Second, we need methods for estimating β and γ. Let us first address the second issue. If we have a sample $x_1, ..., x_n$, we start by discarding all $x_i \leq u$ and consider the excesses $z_i := x_i - u$ for the remaining x_i with $x_i > u$. We are then left with a subsample $z_1, ..., z_{N_u}$ for all $i = 1, ..., N_u$. The parameters β and γ are then estimated via maximal likelihood based on the z subsample. The likelihood function is

$$L(\gamma, \beta; z_1, ..., z_{N_u}) = \prod_{i=1}^{N_u} g_{\gamma,\beta}(z_i) \quad \text{for} \quad g_{\gamma,\beta}(x) = \frac{d}{dx} G_{\gamma,\beta}(x).$$

We can be specific and compute

$$g_{\gamma,\beta}(x) = \frac{1}{\beta}\left(1+\frac{\gamma x}{\beta}\right)^{-1/\gamma-1}, \quad x > 0.$$

We can then consider the log-likelihood

$$\begin{aligned} l(\gamma, \beta; z_1, ..., z_{N_u}) &= \log L(\gamma, \beta; z_1, ..., z_{N_u}) = \sum_{i=1}^{N_u} \log g_{\gamma,\beta}(z_i) \\ &= \sum_{i=1}^{N_u} \left(-\log\beta - \frac{\gamma+1}{\gamma}\log\left(1+\frac{\gamma z_i}{\beta}\right)\right) \\ &= -N_u \log\beta - \frac{\gamma+1}{\gamma}\sum_{i=1}^{N_u}\log\left(1+\frac{\gamma z_i}{\beta}\right) \end{aligned}$$

and maximising this equation numerically in terms of β and γ yields the maximal likelihood estimators $\hat{\beta}_n$ and $\hat{\gamma}_n$. It is possible to construct asymptotic confidence intervals by using the following result.

Theorem 2.4.2. *For $\gamma > -1/2$, it holds that*

$$\sqrt{N_u}\left(\widehat{\gamma}_n - \gamma, \frac{\widehat{\beta}_n}{\beta} - 1\right) \xrightarrow{d} \mathcal{N}(0, M^{-1}) \quad \textit{for} \quad N_u \to \infty$$

where

$$M^{-1} = (1+\gamma)\begin{pmatrix} 1+\gamma & -1 \\ -1 & 2 \end{pmatrix}.$$

We now return to the first problem of determining a good value of u. As usual there is a tradeoff between getting sufficiently many datapoints and choosing a value large enough so that the asymptotics "kick in". The main tool for this job is the mean-excess function.

The mean-excess function

Definition 2.4.3. Let X be an integrable random variable with distribution F. The *mean-excess function* of X is defined as

$$e(u) = E[X - u \mid X > u].$$

The following formula is extremely useful for computing the mean-excess function when the distribution function is known.

Proposition 2.4.4. *Let X be an integrable random variable with distribution function F. Then*

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^\infty \overline{F}(x)dx.$$

Proof. We present two proofs. The first is based on integration by parts and goes as follows. We compute

$$e(u) = \int_0^\infty (x-u)dP(X \le x \mid X > u) = \frac{1}{\overline{F}(u)} \int_u^\infty (x-u)dF(x)$$

and using integration by parts, we get

$$\begin{aligned} \int_u^\infty (x-u)dF(x) = -\int_u^\infty (x-u)d\overline{F}(x) &= \left[-(x-u)\overline{F}(x)\right]_u^\infty + \int_u^\infty \overline{F}(x)dx \\ &= \int_u^\infty \overline{F}(x)dx. \end{aligned}$$

To see why $x\overline{F}(x) \to 0$ for $x \to \infty$, note that (assuming $x > 0$),

$$x\overline{F}(x) = E[x1_{\{X>x\}}] \le E[X1_{\{X>x\}}].$$

The integrand is bounded by $|X|$ which is integrable, so the dominated convergence theorem yields $x\overline{F}(x) \to 0$ for $x \to \infty$ as desired. Collecting pieces finishes the first proof. The other proof relies on the formula[1]

$$E[|X|] = \int_0^\infty P(|X| > t)dt.$$

[1] If this formula is unfamiliar to you, try to prove it! Hint: Use Tonelli's theorem.

With respect to the measure $\mu(\cdot) = P(\cdot|X > u)$, the random variable $X - u$ is non-negative almost surely, so with E_μ indicating that the mean is taken with respect to μ, we have

$$\begin{aligned} e(u) &= E_\mu[X-u] = \int_0^\infty \mu(X-u>t)dt = \int_0^\infty P(X-u>t \mid X>u)dt \\ &= \frac{1}{\overline{F}(u)} \int_0^\infty P(X-u>t)dt. \end{aligned}$$

Applying the substitution $x = t + u$ finishes the second proof. ■

Now assume that F is regularly varying and that the tail parameter α satisfies $\alpha > 1$. Using Karamata's Theorem, Theorem 2.2.5, we get

$$\int_u^\infty \overline{F}(x)dx = \int_u^\infty L(x)x^{-\alpha}dx \sim -\frac{L(u)u^{-\alpha+1}}{-\alpha+1} \quad \text{for} \quad u \to \infty$$

so

$$e(u) \sim \frac{1}{L(u)u^{-\alpha}} \frac{L(u)u^{-\alpha+1}}{\alpha-1} = \frac{u}{\alpha-1} \quad \text{as} \quad u \to \infty$$

which shows that $e(u)$ becomes linear asymptotically. This is a crucial observation and hence we state our findings as a proposition.

Proposition 2.4.5. *If F is regularly varying with index $\alpha > 1$, the mean-excess function $e(u)$ satisfies*

$$e(u) \sim \frac{u}{\alpha-1} \quad \textit{as} \quad u \to \infty.$$

To see how the mean-excess function helps in determining a suitable u, we consider the empirical mean-excess function. The idea is to replace F and $\overline{F}$ by their empirical counterparts. The *empirical mean-excess function* is given by

$$\begin{aligned} e_n(u) &= \frac{1}{N_u/n} \int_u^\infty (x-u)dF_n(x) = \frac{n}{N_u} \sum_{j=1}^n (x_{j,n}-u)1_{\{x_{j,n}>u\}} \frac{1}{n} \\ &= \frac{1}{N_u} \sum_{j=1}^n (x_{j,n}-u)1_{\{x_{j,n}>u\}}. \end{aligned}$$

If we set $u = x_{k,n}$ for some $k = 2, 3, ..., n$ ($k = 1$ is excluded since $e_n(u) = 0$ in this case), we can simplify the above expression to

$$e_n(x_{k,n}) = \frac{1}{k-1} \sum_{j=1}^k (x_{j,n} - x_{k,n}).$$

Using the empirical mean-excess function we can construct a *mean-excess plot* by plotting the values

$$\{(x_{k,n}, e_n(x_{k,n})) : k = 2, 3, ..., n\}.$$

If the values $x_1, ..., x_n$ come from a regularly varying distribution, the plot will roughly look like a straight line for large thresholds. For distributions with lighter tails, the mean-excess function will either decrease or remain roughly constant. It is left to the reader to compute

some examples of mean-excess functions in the exercises. Some examples of mean-excess plots are below.

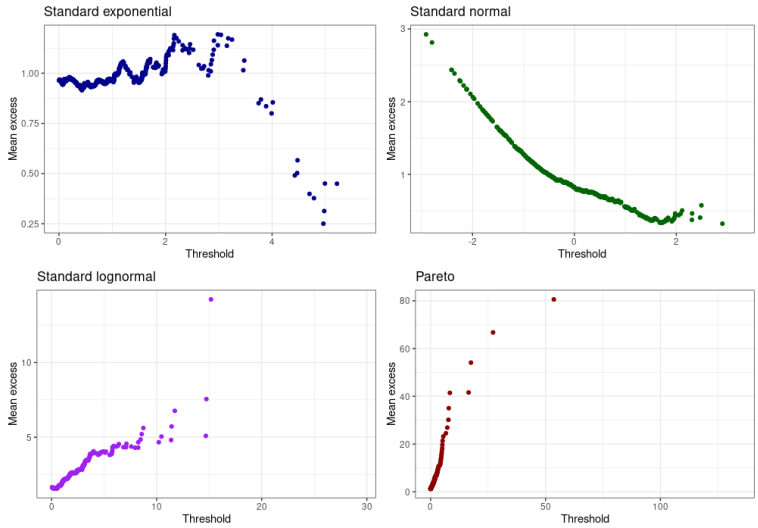


Figure 5: Examples of mean-excess plots based on 500 simulated values from the following distributions: Standard exponential (upper left), standard normal (upper right), standard lognormal (lower left) and the Pareto distribution with $\kappa = 1$ and $\alpha = 2$ (lower right).

The plots should serve not just as an example but also as a warning. The behaviour of the large values in the plots are very chaotic, and one should be cautious in the interpretation of such plots. To illustrate this further, the following plots are made with the exact same distributions but with a different simulated sample.

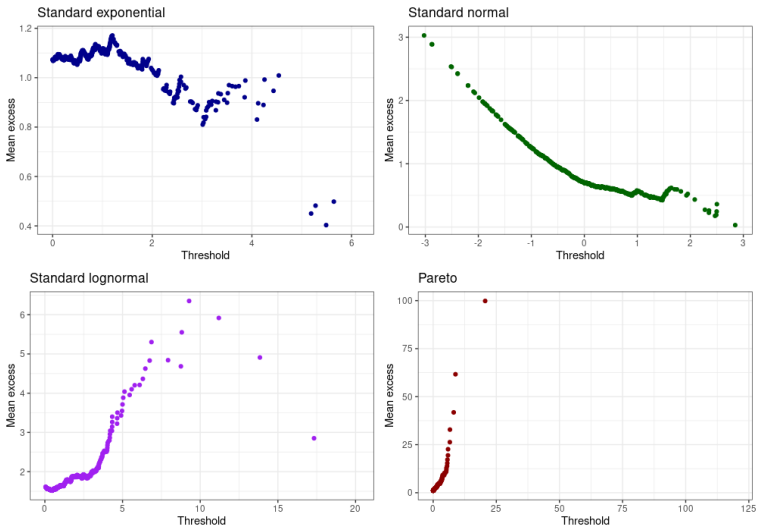


Figure 6: More mean-excess plots with the same distributions as before, but with a new sample.

Some statistical properties of the mean-excess function is covered in the following result. They explain the chaotic behaviour in the plots above.

Theorem 2.4.6. *Let $\{X_n\}$ be an iid sequence of random variables with $X_1 \sim F$ and $E[|X_1|] < \infty$. Fix a threshold u such that $F(u) < 1$. Then*

$$e_n(u) \longrightarrow e(u) \quad a.s. \quad as \quad n \to \infty.$$

If also $E[X_1^2] < \infty$, we have

$$\sqrt{n}(e_n(u) - e(u)) \xrightarrow{d} \mathcal{N}\left(0, \frac{e^{(2)}(u) - e(u)^2}{1 - F(u)}\right)$$

where $e^{(2)}(u) = E[(X-u)^2 \mid X > u]$.

Proof. For the first part, we see that we may rewrite the empirical mean-excess function as

$$e_n(u) = \frac{\sum_{i=1}^n X_i 1_{\{X_i > u\}}}{\sum_{i=1}^n 1_{\{X_i > u\}}} - u.$$

Using the SLLN, we get that a.s.

$$e_n(u) \to \frac{E[(X-u)1_{\{X>u\}}]}{P(X>u)} - u = E[X - u \mid X > u] = e(u)$$

as desired. For the second claim, we know by the CLT that the vector

$$Z_n = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i 1_{\{X_i>u\}} \\ \frac{1}{n}\sum_{i=1}^n 1_{\{X_i>u\}} \end{pmatrix}$$

satisfies

$$\sqrt{n}\left(Z_n - \begin{pmatrix} E[X_1 1_{\{X_1>u\}}] \\ \overline{F}(u) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma(u))$$

where

$$\Sigma(u) = \begin{pmatrix} E[X_1^2 1_{\{X_1>u\}}] - E[X 1_{\{X_1>u\}}]^2 & E[X_1 1_{\{X_1>u\}}]F(u) \\ E[X_1 1_{\{X_1>u\}}]F(u) & F(u)(1-F(u)) \end{pmatrix}.$$

Now note that $e_n(u) = f(Z_n)$ where $f(x, y) = x/y$. The gradient of f is $\nabla f(x, y) = (1/y, -x/y^2)$. Using the delta method (see Theorem C.2.8 in the appendix), the asymptotic variance $\sigma^2(u)$ of $\sqrt{n}(e_n(u) - e(u))$ is

$$\begin{aligned} \sigma^2(u) &= \nabla f\left(E[X_1 1_{\{X_1>u\}}], \overline{F}(u)\right)^T \Sigma(u) \nabla f(E[X_1 1_{\{X_1>u\}}], \overline{F}(u)) \\ &= \frac{e^{(2)}(u) - e(u)^2}{1 - F(u)} \end{aligned}$$

by easy, albeit tedious, calculations. ■

It is clear that for very large thresholds u, the denominator in the asymptotic variance becomes very small, causing the asymptotic variance to explode. This explains the irratic behaviour of the above plots for large values of u. As with any asymptotic normality result, the above proposition can be used to construct confidence intervals for the mean-excess function if one so desires. One should however be cautious in putting too much faith in these confidence bands due to the exploding variance for high thresholds.

It often occurs in practice that the tail and center of the data behave very differently. For example, the tail could behave like a regularly varying distribution while the center acts light-tailed such as a normal or gamma distribution. In that case, the mean-excess plot will probably only become linear for large values of u. In any case, one should use the plot to find a value of u where the points begin to form a straight line. We illustrate this with a (by now well-known) example.

Example 2.4.7. Let us again consider the Danish fire insurance data. We want to model the excesses using the POT method. We wish to determine a proper threshold and therefore make a mean-excess plot:

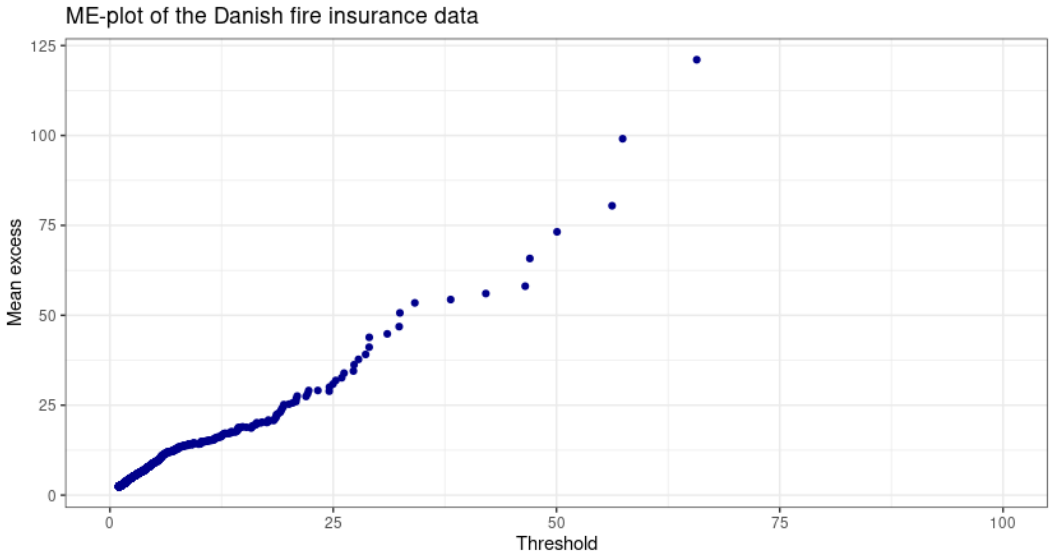


Figure 7: Mean-excess plot of the Danish fire insurance data.

The plot looks very linear for all thresholds, indicating Pareto tails. A word of warning: This is typically not the case for real data! The Danish fire insurance data is in a sense too nice to illustrate this point. Based on the plot, we choose the threshold $u = 4$. While the linear trend may begin for slightly larger values, this choice also gives us sufficient data to work with. To fit the Generalised Pareto Distribution, we apply maximum likelihood estimation using the `gpd` function in the `evir` package ([96]) as follows:

```
data(danish)
u <- 4
gpdfit <- gpd(danish, threshold = u, method = "ml")
gpdfit$par.ests
```

This gives the estimates $\widehat{\gamma} = 0.7209$ and $\widehat{\beta} = 2.6291$. We now plot the empirical tail and the tail from the POT approximation:

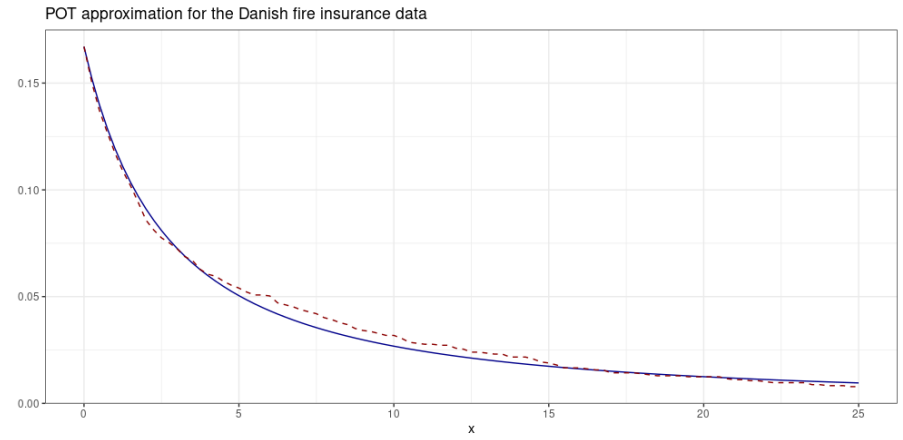


Figure 8: Red/dashed: The empirical tail from the data, $\overline{F}_n(u+x)$. Blue/solid: The corresponding tail from the fitted Generalised Pareto Distribution.

The approximation looks very good. Usually, one is not so lucky. If one is interested in a statistical goodness of fit, a possible way (which is also implemented in `evir` if one uses `plot` around a fitted gpd object) is to consider the excesses $z_i = x_i - u$ which should be approximately $G_{\widehat{\gamma},\widehat{\beta}}$ distributed. Hence $-\log \overline{G}_{\widehat{\gamma},\widehat{\beta}}(z_i)$ (which we call the *generalised residuals*) should be approximately standard exponential distributed. We make a residual plot and a QQ-plot of the $-\log \overline{G}_{\widehat{\gamma},\widehat{\beta}}(z_i)$ against a standard exponential distribution and get the following:

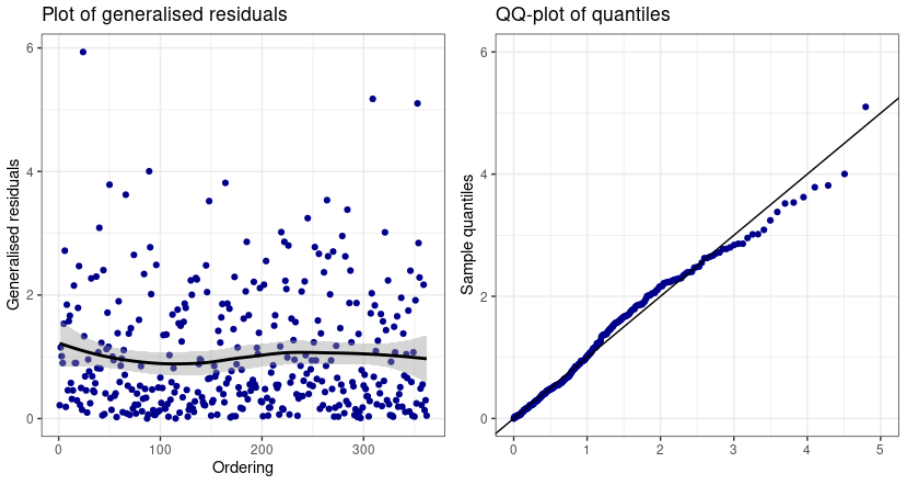


Figure 9: Diagnostic plots for the fitted GPD.

We see no clear tendencies of the generalised residuals. Some of them are quite large, but otherwise the left plot looks good. The right plot also looks good. While some of the sample quantiles are below the line with slope one and intercept zero, the residuals overall seem to

follow a standard exponential distribution. We conclude that the model is an adequate fit. ◦

2.5 Maximum domains of attraction*

The goal of this supplementary section is to provide the theoretical links between regularly varying distributions and the POT method. This also serves to justify many of the choices one explicitly makes when applying these methods in practice such as assuming that the excesses over a certain threshold satisfy a GPD. This theory is elegantly formulated via the *Generalised Extreme Value distribution* (GEV) and so-called *Maximal Domains of Attractions*. Maximal Domains of Attraction (MDA's) are intricately linked to the notion of a max-stable distribution, and this concept is the starting point of our discussion.

The Fisher–Tippett Theorem

Definition 2.5.1. A random variable X (or its distribution function F) is called *max-stable* if there exist sequences $\{c_n\}, \{d_n\} \subseteq \mathbb{R}$ with $c_n > 0$ such that for $\{X_n\}$ iid with $X_1 \sim F$, it holds that

$$\max\{X_1, ..., X_n\} \stackrel{\mathrm{d}}{=} c_n X_1 + d_n.$$

A classical problem in extreme value theory is the following question: If $M_n := \max\{X_1, ..., X_n\}$ for an iid sequence $\{X_n\}$ satisfies

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} Z \quad \text{as} \quad n \to \infty \tag{2.1}$$

for Z non-degenerate, what are the possible distributions of Z? It turns out that these distributions coincide with the max-stable distributions.

Theorem 2.5.2. *The class of possible limit distributions in the setting of (2.1) is the same as the class of max-stable distributions in the sense of Definition 2.5.1.*

The proof of this fact relies on the following technical result, a proof of which can be found in Chapter 0.3 of [102].

Theorem 2.5.3 (Convergence to types theorem). *Let $X, Y, X_1, X_2, ...$ be random variables and $a_n, \alpha_n \in \mathbb{R}$, $b_n, \beta_n > 0$ constants. Assume*

$$\frac{X_n - a_n}{b_n} \xrightarrow{d} X \quad \textit{as} \quad n \to \infty.$$

Then

$$\frac{X_n - \alpha_n}{\beta_n} \xrightarrow{d} Y \quad \textit{as} \quad n \to \infty$$

holds if and only if

$$\lim_{n\to\infty} \frac{b_n}{\beta_n} = b \in [0, \infty) \quad \textit{and} \quad \lim_{n\to\infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R}.$$

In this case, we have $Y \stackrel{\mathrm{d}}{=} bX + a$.

This theorem tells us that the limit distribution of a sequence of random variables is uniquely determined up to location and scale.

Proof of Theorem 2.5.2. It turns out that Z has to be continuous as will follow from the main theorem below. Hence we assume that $Z \sim H$ for a continuous distribution function H. The condition (2.1) is equivalent to (letting $X_1 \sim F$)

$$\lim_{n\to\infty} F^n(c_n x + d_n) = H(x), \quad x \in \mathbb{R}.$$

Taking k'th powers, we get

$$\lim_{n\to\infty} F^{nk}(c_n x + d_n) = (\lim_{n\to\infty} F^n(c_n x + d_n))^k = H^k(x)$$

and

$$\lim_{n\to\infty} F^{nk}(c_{nk} x + d_{nk}) = H(x)$$

for all $x \in \mathbb{R}$. Translated into random variables, these relations tell us that if $Y_1, ..., Y_k \sim H$ are iid, then

$$\frac{\max\{X_1, ..., X_{nk}\} - d_n}{c_n} \xrightarrow{d} \max\{Y_1, ..., Y_k\}$$

and

$$\frac{\max\{X_1, ..., X_{nk}\} - d_{nk}}{c_{nk}} \xrightarrow{d} Y_1$$

for $n \to \infty$. Applying the Convergence to types theorem, there exist limits

$$b_k := \lim_{n\to\infty} \frac{c_{nk}}{c_n} \quad \text{and} \quad a_k := \lim_{n\to\infty} \frac{d_{nk} - d_n}{c_n}$$

such that $\max\{Y_1, ..., Y_k\} \overset{\mathrm{d}}{=} b_n Y_1 + a_n$. This shows that H is a max-stable distribution. The converse implication is trivial, completing the proof. ■

The following theorem, a cornerstone in classical extreme value theory, completely characterises the limit distributions of centered and normalised maxima.

Theorem 2.5.4 (Fisher–Tippett Theorem). *If $\{X_n\}$ is a sequence of iid random variables and $\{c_n\}$,$\{d_n\} \subseteq \mathbb{R}$ are deterministic sequences such that*

$$\frac{\max\{X_1, ..., X_n\} - d_n}{c_n} \xrightarrow{d} H$$

for a non-degenerate distribution function H, then H has to belong to one of the following three classes of distributions ($\alpha > 0$ is a parameter):

(i) ***Fréchet****:*

$$H(x) = \Phi_\alpha(x) = e^{-x^{-\alpha}}, \quad x > 0.$$

(ii) ***Weibull****:*

$$H(x) = \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \le 0 \\ 1, & x > 0 \end{cases}$$

(iii) ***Gumbel****:*

$$H(x) = \Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$$

Remark 2.5.5. The Weibull distribution is typically defined as having the distribution function

$$F_{\lambda,\alpha}(x) = 1 - e^{-(x/\lambda)^{\alpha}}, \quad x > 0$$

for parameters $\lambda, \alpha > 0$. This is not the same distribution function as the one presented above, but they are related through the identity

$$\Psi_{\alpha}(x) = 1 - F_{1,\alpha}(-x), \quad x < 0.$$

In this section, when we refer to the Weibull distribution, we mean the distribution Ψ_{α} from the Fisher–Tippett Theorem.

The proof of this result lies in establishing the form of the distribution function H. This involves solving certain functional equations. For details, we refer to Chapter 0.3 of [102]. The Fisher–Tippet Theorem makes it meaningful to attempt to classify which distributions belong to which limit distribution (if any) when its maxima have been properly normalised.

Definition 2.5.6 (**Maximal Domains of Attraction**)**.** Let H denote one of the possible limit distributions in the Fisher–Tippett Theorem. If X satisfies (2.1) for this H, we say that X is in the *Maximal Domain of Attraction* for H and write $X \in \mathrm{MDA}(H)$.

If X has distribution function F, we will also write $F \in \mathrm{MDA}(H)$ instead of $X \in \mathrm{MDA}(H)$. Theorem 2.5.2 tells us that each of the three possible limit distributions should be max-stable. The normalising constants c_n and d_n are given in the following lemma.

Lemma 2.5.7. *The normalising constants (see Definition 2.5.1) for the limit distributions in the Fisher–Tippett Theorem are given as follows:*

(i) ***Fréchet****:* $c_n = n^{1/\alpha}, d_n = 0$.

(ii) ***Weibull****:* $c_n = n^{-1/\alpha}, d_n = 0$.

(iii) ***Gumbel****:* $c_n = 1, d_n = \log n$.

Proof. Left as an exercise for the reader. ■

We now provide examples of distributions belonging to different MDA's. More examples can be found in the exercises.

Example 2.5.8. Assume X has Pareto tails, $\overline{F}(x) \sim Kx^{-\alpha}$ for some $\alpha > 0$. We claim that F is in $\mathrm{MDA}(\Phi_{\alpha})$. We need to find sequences $\{c_n\}$ and $\{d_n\}$ with $c_n > 0$ such that

$$F^n(c_n x + d_n) \to \Phi_{\alpha}(x).$$

We guess that $d_n = 0$. Then we need to determine c_n such that $(1 - K(c_n x)^{-\alpha})^n \to e^{-x^{\alpha}}$ for every $x > 0$. We see that $c_n = (Kn)^{1/\alpha}$ leads to

$$F^n(c_n x) = \left(1 - \frac{Kx^{-\alpha}}{Kn}\right)^{-\alpha} \to e^{-x^{-\alpha}} = \Phi_{\alpha}(x)$$

as desired. ○

Example 2.5.9. Consider the uniform distribution. In this case,

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

Hence $F^n(x) = x^n$ for $x \in [0,1]$. Now consider some $x < 0$ and let $c_n = n^{-1}, d_n = 1$. Then for large enough n,

$$F^n(c_n x + d_n) = \left(1 - \frac{x}{n}\right)^n \to e^{-x} = \Psi_{-1}(x).$$

This shows that the uniform distribution is in $\mathrm{MDA}(\Psi_{-1})$. ∘

Example 2.5.10. It turns out that the normal distribution is in $\mathrm{MDA}(\Lambda)$, but it is not trivial to show. For further arguments, see Example 3.3.29 in [30]. We state that the normalising constants may be chosen according to

$$c_n \sim (2\log n)^{-1/2}, \quad d_n = \sqrt{2\log n} - \frac{\log\log n + \log 4\pi}{2\sqrt{2\log n}} + o((\log n)^{-1/2}).$$

∘

Characterisations of maximal domains of attraction

We here present some characterisations of the three different maximal domains of attration without proof.

The Fréchet MDA

In the examples above we saw that the Fréchet MDA contains all distributions with Pareto-like tails. It is therefore natural to suggest that this MDA contains all regularly varying distributions. Even more is true as the following theorem says.

Theorem 2.5.11 (**$\mathrm{MDA}(\Phi_\alpha)$**). *A distribution F belongs to $\mathrm{MDA}(\Phi_\alpha)$ if and only if it is of regular variation with index $\alpha > 0$. In this case, one may choose the normalising constants by*

$$c_n = F^{\leftarrow}(1 - n^{-1}), \quad d_n = 0.$$

The case of Pareto tails, $\overline{F}(x) \sim Kx^{-\alpha}$ for $\alpha > 0$, we may solve for generalised inverse of $x \mapsto 1 - Kx^{-\alpha}$, yielding $p \mapsto ((1-p)/K)^{-1/\alpha}$. A choice of normalising sequence is thus

$$c_n = \left(\frac{1 - (1 - n^{-1})}{K}\right)^{-1/\alpha} = (Kn)^{1/\alpha}, \quad d_n = 0.$$

This sequence is identical to the one from before.

The Weibull MDA

We saw earlier that the uniform distribution was in the Weibull MDA. It turns out that all distributions in this maximal domain of attraction have finite right endpoint. A precise classification is the following.

Theorem 2.5.12 (**MDA(Ψ_α)**). *A distribution F belongs to MDA(Ψ_α) for $\alpha > 0$ if and only if $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$ and $x \mapsto \overline{F}(x_F - x^{-1}) \in RV_{-\alpha}$. In this case, one may choose the normalising constants by*

$$c_n = x_F - F^{\leftarrow}(1 - n^{-1}), \quad d_n = x_F.$$

The reader can verify that the normalising constants we found for the uniform distribution coincide with the ones provided by the theorem.

The Gumbel MDA

It turns out that characterising the Gumbel maximum domain of attraction is quite involved, and it takes a bit of work to provide a satisfying answer. Applying the Taylor expansion of e^x to $\Lambda(x) = e^{-e^{-x}}$, we obtain

$$\Lambda(x) = 1 + e^{-x} + o(e^{-x}), \quad \text{i.e.} \quad 1 - \Lambda(x) \sim e^{-x}, \quad x \to \infty.$$

This shows that the tail decays exponentially. Hence we should expect distributions in MDA(Λ) to have all moments. In that sense, MDA(Λ) comprised light-tailed distributions. This will become clearer in the following characterisation using *von Mises functions*. Recall that $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}$ denotes the right endpoint of a distribution function F.

Definition 2.5.13 (**Von Mises function**). Let F be a distribution function, continuous at its right endpoint $x_F \leq \infty$. Suppose that for some $z < x_F$, we may write

$$\overline{F}(x) = c \exp\left(-\int_z^x a(t)^{-1} dt\right), \quad z < x < x_F$$

with $c > 0$ a constant and a a function which is positive and absolutely continuous with respect to Lebesgue measure, where the derivative a' satisfies $a'(x) \to 0$ as $x \uparrow x_F$. Then F is called a *von Mises function* and a the *auxiliary function* of F.

The auxiliary function determines the von Mises function up to multiplication by the constant c. Indeed,

$$\frac{1}{a(x)} = \frac{d}{dx}(-\log \overline{F}(x)) = \frac{f(x)}{\overline{F}(x)}$$

where f is the density of F. Thus $a(x) = \overline{F}(x)/f(x)$. The reader may recognise this function as the inverse hazard. It turns out that every von Mises function is in the maximum domain of attraction of the Gumbel.

Proposition 2.5.14. *If F is a von Mises function with auxiliary function a, then $F \in$ MDA(Λ). The norming constants can be chosen according to*

$$d_n = F^{\leftarrow}(1 - n^{-1}), \quad c_n = a(d_n).$$

This result is not the whole story however, but it is the first step. A complete characterisation can be obtained by slightly altering the von Mises function representation.

Theorem 2.5.15 (**MDA**(Λ))**.** *A distribution F is in MDA(Λ) if and only if there exists $z < x_F$ such that F has representation*

$$\overline{F}(x) = c(x) \exp\left(- \int_z^x \frac{g(t)}{a(t)} dt \right)$$

with c and g measurable functions such that $c(x) \to c > 0$ and $g(x) \to 1$ for $x \uparrow 1$ and a a function satisfying the same requirements as for a von Mises function. The norming constants for F with this representation can be chosen according to

$$d_n = F^{\leftarrow}(1 - n^{-1}), \quad c_n = a(d_n).$$

Furthermore, a choice for the function a is

$$a(x) = \frac{1}{\overline{F}(x)} \int_x^{x_F} \overline{F}(t) dt,$$

that is, the mean-excess function.

Another characterisation can be obtained via the natural notion of tail-equivalence.

Definition 2.5.16. Two distribution functions F and G are called *tail-equivalent* if $x_F = x_G$ and

$$\lim_{x \to x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c$$

for a constant $c \in (0, \infty)$.

We may then provide the following very elegant characterisation.

Theorem 2.5.17 (**MDA**(Λ))**.** *The maximal domain of attraction for the Gumbel distribution consists of von Mises functions and their tail equivalent distribution functions.*

Several famous distributions belong to the Gumbel maximal domain of attraction. It is clear that the exponential distribution is a von Mises function. The same is the case for Weibull distribution with distribution function

$$F(x) = 1 - e^{-(x/\lambda)^{\alpha}}, \quad x > 0$$

for parameters $\lambda, \alpha > 0$. It can also be verified that the gamma distribution is in MDA(Λ). For more examples of distributions and their respective maximal domains of attraction, see Exercise 2.31.

Statistical implications

For the purposes of risk management, we are interested in the implications of the MDA membership of a distribution on our ability to estimate quantities of interest such as high quantiles (VaR). For statistical purposes, it turns out to be fruitful to work with a more general representation of the three limit distributions that appear in the Fisher–Tippett Theorem.

Definition 2.5.18 (**Generalised Extreme Value distribution (GEV)**). The *Generalised Extreme Value distribution* has distribution function given by

$$H_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x > 0$$

for $\gamma \neq 0$ and $H_0(x) = \exp(-e^{-x})$.

We remark a few things. First of all, H_γ is continuous in γ since $\lim_{\gamma\to 0} H_\gamma(x) = H_0(x)$. Second, the support of H_γ is $(-\gamma^{-1}, \infty)$ for $\gamma > 0$, $(-\infty, -\gamma^{-1})$ for $\gamma < 0$ and $\mathbb{R}$ for $\gamma = 0$. The case $\gamma > 0$ corresponds to the Fréchet distribution via the relation

$$H_\gamma(x) = \Phi_\alpha\left(\frac{x+\alpha}{\alpha}\right), \quad \alpha = \frac{1}{\gamma},$$

while the case $\gamma < 0$ corresponds to the Weibull distribution via

$$H_\gamma(x) = \Psi_\alpha\left(\frac{x-\alpha}{\alpha}\right), \quad \alpha = -\frac{1}{\gamma}.$$

And of course $H_0(x) = \Lambda(x)$. The main takeaway from these relations is of statistical nature. The unified representation of the three limit distributions as a single distribution function allows a common statistical treatment. For example, estimating α in a regularly varying distribution is equivalent to estimating γ via $\alpha = 1/\gamma$. This representation also provides further properties linking memberships of maximal domains of attraction to the POT method. If one redefines the GPD to allow for general $\gamma \in \mathbb{R}$, then we obtain the following theorem.

Theorem 2.5.19. *For any $\gamma \in \mathbb{R}$, $F \in MDA(H_\gamma)$ if and only if*

$$\lim_{u\uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\gamma,\beta(u)}(x)| = 0$$

for some positive function β.

In words, F belongs to the maximum domain of attraction of some distribution if and only if the excesses over a sufficiently large threshold u are close to a Generalised Pareto Distribution. This justifies the approximation

$$\overline{F}_u(x) \approx \overline{G}_{\gamma,\beta}(u)$$

from the POT method. There we used maximum likelihood on the excesses to determine the "function" β. As a final note, we mention that the GEV can be extended to also have a scale and location parameter. This distribution has distribution function given by

$$H_{\gamma,\mu,\psi}(x) = \exp\Big(-\Big(1+\gamma\frac{x-\mu}{\psi}\Big)^{-1/\gamma}\Big), \quad 1+\gamma\frac{x-\mu}{\psi} > 0$$

where $\gamma \neq 0$ and $\mu \in \mathbb{R}, \psi > 0$. As for the case of $\gamma = 0$, we let

$$H_{0,\mu,\psi}(x) = \exp\Big(-e^{-(x-\mu)/\psi}\Big), \quad x \in \mathbb{R}.$$

Notes and comments

[8] and [102] contain all the information one could wish for concerning regular variation, including a proof of Karamata's Theorem. [30] is an excellent source for extreme value theory. Chapter 3.4 covers the GPD and chapter 6 covers statistical methods, including the Hill estimator and the POT method. We have chosen to focus entirely on the Hill estimator for estimating the tail index, but several others exist. An older estimator, important for historical reasons, is the *Pickands estimator*, see [64] as well as [21] for asymptotic properties. A more modern estimator which extends the estimation to arbitrary γ for the GEV is the *Dekkers–Einmahl-de Haan estimator*, see [22]. In the remark following Theorem 2.3.4, we mentioned that understanding the exact asymptotic behaviour of the Hill estimator is difficult. The almost sure convergence statement was proved in [94]. The bias has been studied in, among others, [48], [20], [53] and [111]. As for the asymptotic properties of the MLE for the GPD, the limit results presented above can be found in [112]. The POT method has been discussed by several authors. See for example [113], [18] and [19] for thorough introductions to the POT method and [64], [112] and [80] for theoretical background. The relationship between the tail index and the parameters in the GPD allows for the use of the POT method to estimate the tail index, an approach explored by [32]. For applications of the POT method in (re)insurance, we refer to [116], [117] and [101].

2.6 Exercises

Theoretical exercises

Exercise 2.1:
In this exercise, we will get more comfortable with the concept of regular and slow variation.

1)Prove the remaining implication in Proposition 2.2.3.

2)Show that $h(x) = 2 + \cos(1/x)\sin(\pi/2 + 1/x^2)$ for $x > 0$ is slowly varying.

3)Verify that the function $h(x) = \log\log x$ for x sufficiently large is slowly varying.

4)Show that the Pareto distribution with parameters $\kappa > 0$ and $\alpha > 0$ is regularly varying. What is the index? Recall that the distribution function is given by

$$1 - \left(\frac{\kappa}{\kappa + x}\right)^{\alpha}, \quad x > 0.$$

5)Let $\tilde{h}$ be a slowly varying function at zero, i.e.

$$\lim_{x \to 0} \frac{\tilde{h}(xt)}{\tilde{h}(x)} = 1 \tag{2.2}$$

for all $t > 0$. Prove that $h(x) = \tilde{h}(1/x)$ is slowly varying (at ∞ as usual). State and show a similar statement for a function $\tilde{h}$ regularly varying at zero with index $\alpha \in \mathbb{R}$.

Exercise 2.2:
For the following distributions, prove that they are regularly varying and determine the corresponding index.

1)The Cauchy distribution with density

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

2)The Student t distribution with $\nu > 0$ degrees of freedom, i.e. the distribution with density

$$g_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

Exercise 2.3:
Consider a regularly varying distribution function F supported on $[0, \infty)$ of index $\alpha > 0$. Let $X \sim F$.

1)Prove that $E[X^\beta] < \infty$ if $\beta < \alpha$.

2)Prove that $E[X^\beta] = \infty$ when $\beta > \alpha$.
Hint for both subproblems: Apply Karamata's Theorem. Recall also the formula $E[X] = \int_0^\infty P(X > t)dt$ for a positive random variable X.

Exercise 2.4:
In this exercise, we show that the moment conditions in the previous exercise do not characterise regularly varying distributions. Consider the Peter-and-Paul distribution which has

density

$$P(X = 2^k) = 2^{-k}, \quad k \in \mathbb{N}$$

and zero elsewhere.

1)Show that X is not of regular variation.

2)Show that $E[X^\beta] < \infty$ for $\beta < 1$ and that $E[X^\beta] = \infty$ for $\beta \geq 1$.

Exercise 2.5:
By Karamata's Representation Theorem, Corollary 2.2.6, we have that if L is slowly varying, then we can write

$$L(x) = c_0(x)e^{\int_{x_0}^{x} \frac{\varepsilon(t)}{t} dt}, \quad x \geq x_0$$

where $c_0(x) \to c_0 > 0$ for $x \to \infty$ and $\varepsilon(x) \to 0$ for $x \to \infty$ for some $x_0 > 0$. If L is written in this form, we call the above a *Karamata representation* for L.

1)Find a Karamata representation for log.

2)Prove that a function of the form above is slowly varying, completing the proof of Corollary 2.2.6.

3)Use the Karamata representation to show that if L is slowly varying and $\epsilon > 0$, then

$$\lim_{x\to\infty} x^\epsilon L(x) = \infty \quad \text{and} \quad \lim_{x\to\infty} x^{-\epsilon} L(x) = 0.$$

Exercise 2.6:
This exercise is from [102]. Consider the function $f(x) = (1 + x^{-1}) \log x$.

1)Show that f is slowly varying.

2)Find a Karamata representation for f.

3)Repeat the previous two subproblems for the function $g(x) = \exp((\log x)^a)$ for $a \in (0, 1)$.

Exercise 2.7:
This exercise is adapted from Exercise 0.4.2.10 in [102]. As stated in Proposition 0.5 in that book, it can be shown that if $h \in \mathrm{RV}_\rho$, then the convergence $\lim_{x\to\infty} h(tx)/h(x) = t^\rho$ holds *locally uniformly* on $(0, \infty)$. This means that the convergence is uniform on every bounded interval.

1)Prove the following fact from analysis: If $f_n \to f$ uniformly on $[a, b]$ and f is continuous in $t \in [a, b]$, then for every sequence $\{t_n\} \subseteq [a, b]$ with $t_n \to t$, it holds that $f_n(t_n) \to f(t)$.

2)Use the previous subproblem to prove the following result: Assume that we have non-negative random variables $\{X_n\}$ satisfying $X_n/n \xrightarrow{P} X$ as $n \to \infty$, and let $h \in \mathrm{RV}_\rho$ for $\rho \neq 0$ be continuous. Then it holds that

$$\frac{h(X_n)}{h(n)} \xrightarrow{P} X^\rho.$$

Exercise 2.8:
In this exercise, we prove a closure property for regularly varying distributions. Let X_1, X_2 be independent and both regularly varying of index $\alpha > 0$ (but not necessarily with the same slowly varying function!). We want to show that $X_1 + X_2$ is also regularly varying

of index α. Let F_1, F_2 and F denote the distribution functions of X_1, X_2 and $X_1 + X_2$, respectively.

1)Show that $\overline{F}(x) \geq (\overline{F}_1(x) + \overline{F}_2(x))(1 - o(1))$. Hint: $\{X_1 > x\} \cup \{X_2 > x\} \subseteq \{X_1 + X_2 > x\}$.

2)For any $\delta \in (0, 1/2)$, explain why the identity

$$\{X_1 + X_2 > x\} \subseteq \{X_1 > (1-\delta)x\} \cup \{X_2 > (1-\delta)x\} \cup \{X_1 > \delta x, X_2 > \delta x\}$$

holds

3)Show that $\overline{F}(x) \leq (\overline{F}_1((1-\delta)x) + \overline{F}_2((1-\delta)x))(1 + o(1))$.

4)Show that $\overline{F}(x) \sim \overline{F}_1(x) + \overline{F}_2(x)$ and conclude that $X_1 + X_2$ is regularly varying of index α.

We have actually shown something extra. A distribution F satisfying $P(X_1 + X_2 > x)/2\overline{F}(x) \to 1$ for $x \to \infty$ where X_1, X_2 are iid with distribution function F is called *subexponential*. With this definition in place, we have shown that a regularly varying distribution is also subexponential. One can verify that a subexponential distribution does not possess a moment-generating function and so one can think of these distributions as being heavy-tailed. We refer to chapter 1 of [30] for details. See also Exercise 2.20 below.

Exercise 2.9:
Let h be a regularly varying monotone function on $[0, \infty)$. Define the jump $\Delta h(x) := h(x) - h(x-)$ with $h(x-) = \lim_{y \uparrow x} h(y)$. Prove that

$$\lim_{x \to \infty} \frac{\Delta h(x)}{h(x)} = 0.$$

Exercise 2.10:
In this exercise, we fill in the missing parts of the proof of Proposition 2.2.9.

1)Prove part (i). Hint: Recognise a uniform distribution.

2)Prove part (ii). Hint: Recognise an exponential distribution.

3)Prove part (iii). Hint: You need not use the expression for the density.

4)Prove part (iv).

Exercise 2.11:
In this exercise, we derive estimates of the Value at Risk and Expected Shortfall using the Hill estimator. Recall from above that if $\widehat{\alpha}_k$ is the Hill estimator with threshold k, then the tail of X can be approximated by

$$\overline{F}(x) \approx \frac{k-1}{n} \left(\frac{x}{x_{k,n}} \right)^{-\widehat{\alpha}_k}.$$

1)Derive an estimate of VaR_β.

2)Show that an estimate of ES_β is given by

$$\mathrm{ES}_\beta(X) = x_{k,n} \left(\frac{n}{k-1} \right)^{-1/\widehat{\alpha}_k} \frac{\widehat{\alpha}_k}{1 - \widehat{\alpha}_k} (1-\beta)^{-1/\widehat{\alpha}_k}$$

whenever $\widehat{\alpha}_k > 1$.

Exercise 2.12:
In this exercise, we derive the Hill estimator in some special cases.

1)Assume X has tail $\overline{F}(x) = Cx^{-\alpha}$ for $x \geq u > 0$. If C is fully specified, we must have $C = u^\alpha$. Derive the MLE for α. What is the connection to the Hill estimator?

2)Let us generalise the setup a bit and say that $\overline{F}(x) = Cx^{-\alpha}$ for $x \geq u$ without fully specifying C. Letting $N_u = \#\{i : X_{i,n} > u\}$ and conditioning on $N_u = k$, we then need to derive the MLE of α and C by considering the joint density of $(X_{k,n}, ..., X_{1,n})$. You may use without proof that the density of $(X_{k,n}, ..., X_{1,n})$ is given by

$$f_{X_{k,n},...,X_{1,n}}(x_1, ..., x_n) = \frac{n!}{(n-k)!}(1 - Cx_k^{-\alpha})^{n-k} C^k \alpha^k \prod_{i=1}^{k} x_i^{-(\alpha+1)}, \quad u < x_k < \cdots < x_1,$$

see Theorem 4.1.3 in [30] for details. Show that the MLE of α is the Hill estimator.

Exercise 2.13:
Recall that we proved the following formula for the mean-excess function of an integrable random variable X:

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^\infty \overline{F}(x)dx.$$

1)Let $X \sim \text{Exp}(\lambda)$. Prove that $e(u) = 1/\lambda$.

2)Let X be Pareto distributed with parameters $\kappa > 0$ and $\alpha > 1$. Prove that

$$e(u) = \frac{\kappa + u}{\alpha - 1}.$$

3)Show that if X is log-normal with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, then

$$e(u) = e^{\mu + \frac{\sigma^2}{2}} \frac{\Phi\left(\frac{(\mu+\sigma^2) - \log u}{\sigma}\right)}{\Phi\left(\frac{\mu - \log u}{\sigma}\right)} - u,$$

and show that for $u \to \infty$, it holds that

$$e(u) = \frac{\sigma^2 u}{\log u - (\mu + \sigma^2)}(1 + o(1).$$

Hint: For $x \to \infty$, it holds that $\Phi(x) \sim -\phi(x)/x$ with $\phi(x) = \Phi'(x)$ the density of the standard normal distribution.

4)Relate the results to the mean-excess plots in the discussion above.

Exercise 2.14:
In this exercise, we consider the *Burr distribution* which has tail

$$\overline{F}(x) = \left(\frac{\kappa}{\kappa + x^\tau}\right)^\alpha, \quad x > 0$$

where $\kappa, \tau, \alpha > 0$ are parameters.

1)Show that the Burr distribution is regularly varying. What is the index?
2)Show that

$$e(u) = \frac{u}{\alpha\tau - 1}(1 + o(1))$$

when $\alpha\tau > 1$.

Exercise 2.15:
Assume has X a GPD with paramaters $\gamma \in (0, 1)$ and $\beta > 0$. Show that the mean-excess function of X equals

$$e(u) = \frac{\beta + \gamma u}{1 - \gamma}$$

and relate this to Proposition 2.4.5.

Exercise 2.16:
Let X be a non-negative random variable with finite first moment, $E[|X|] < \infty$, and continuous distribution function F. Show that the mean-excess function e completely determines F by verifying the identity

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\left(-\int_0^x \frac{1}{e(t)} dt\right).$$

Exercise 2.17:
Let X be a random variable with distribution function F.
1)Assume $E[|X|^k] < \infty$ for $k \in \mathbb{N}$. Prove that

$$E[|X|^k] = k \int_0^\infty x^{k-1} P(|X| > x) dx.$$

Hint: Use $E[|X|] = \int_0^\infty P(|X| > x)dx$.
2)Assume $E[X^2] < \infty$. Prove the formula

$$e^{(2)}(u) := E[(X - u)^2 \mid X > u] = \frac{2}{\overline{F}(u)} \int_u^\infty (x - u)\overline{F}(x) dx.$$

This formula may be useful for computing the asymptotic variance in Theorem 2.4.6. Hint: Follow the proof of Proposition 2.4.4.
3)Derive a formula for

$$e^{(k)}(u) := E[(X - u)^k \mid X > u]$$

whenever $E[|X|^k] < \infty$.

Exercise 2.18:
Let X be Pareto distributed with parameters $\kappa, \alpha > 0$. Let $Y > 0$ be independent of X with $E[Y^\alpha] < \infty$. Prove that

$$\lim_{x \to \infty} \frac{P(XY > x)}{P(X > x)} = E[Y^\alpha].$$

Use this to show that XY is regularly varying of index α. Hint: Use the tower property to explicitly compute $P(XY > x)$. You may need the formula

$$E[Y^\alpha] = \alpha \int_0^\infty y^{\alpha-1} P(Y > y) dy$$

which was proved in the previous exercise (it also holds when α is not an integer by the same argument).

The result of this exercise is a version of a result known as *Breiman's Lemma* in the litterature. Breiman's Lemma states that the result of the exercise holds whenever X is regularly varying of index α and $E[Y^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, and Y is independent of X.

Exercise 2.19:
The goal of this exercise is to prove that if $X \geq 0$ is a random variable with distribution function F and

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = e^{\gamma y}, \quad y > 0$$

for some $\gamma \in (0, \infty)$, then

$$e(u) \to \gamma^{-1} \quad \text{for} \quad u \to \infty.$$

1)Prove that $\overline{F} \circ \log \in \mathrm{RV}_{-\gamma}$.

2)Prove the above result. Hint: Karamata's Theorem.

3)The above result also holds for $\gamma \in \{0, \infty\}$, and you can use this without proof. Provide another proof of the fact that if X is standard normal, then $e(u) \to 0$. Relate this to the plots in the discussion above.

Exercise 2.20:
We call a positive random variable X (or its distribution function F) *subexponential* whenever it holds for $X_1, X_2 \sim F$ independent that

$$\frac{P(X_1 + X_2 > x)}{2\overline{F}(x)} \to 1 \quad \text{as} \quad x \to \infty.$$

We saw in Exercise 2.8 that a regularly varying distribution is subexponential. In this exercise, we will establish some basic properties of subexponential distributions.

1)Prove that a distribution is subexponential if and only if

$$\frac{P(X_1 + X_2 > x)}{P(\max\{X_1, X_2\} > x)} \to 1 \quad \text{as} \quad x \to \infty.$$

2)Now let F be subexponential. Prove that for any $y > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1.$$

Hints: Recall the *convolution formula*. If X and Y are independent and non-negative with $X \sim G$ and $Y \sim H$,

$$P(X + Y \leq z) = \int_0^z G(z-t) dH(t).$$

Use this formula with $X_1, X_2 \sim F$ and use this to establish the identity

$$\frac{P(X_1 + X_2 > x)}{\overline{F}(x)} = 1 + \int_0^y \frac{\overline{F}(x-t)}{\overline{F}(x)} dF(t) + \int_y^x \frac{\overline{F}(x-t)}{\overline{F}(x)} dF(t).$$

Then explain why

$$\frac{P(X_1 + X_2 > x)}{\overline{F}(x)} \geq 1 + F(y) + \frac{\overline{F}(x-y)}{\overline{F}(x)}(F(x) - F(y)).$$

Isolate $\overline{F}(x-y)/\overline{F}(x)$ and complete the proof.

3)Give an intuitive interpretation of the limit result in the two previous subproblems. What is the implication for an insurer with subexponential claims? Explain also the asymptotic behaviour of the mean-excess function. Hint: Recall the previous exercise.

4)Let F be subexponential. Prove that for any $\varepsilon > 0$, it holds that

$$e^{\varepsilon x}\overline{F}(x) \to \infty \quad \text{as} \quad x \to \infty.$$

Use this to show that a subexponential distribution does not have a finite moment-generating function in any neighbourhood around zero. Hint: Consider $\overline{F}(\log(y))$ and use the result of Exercise 2.5 3).

Exercise 2.21:
In this exercise, we derive estimates of Value at Risk and Expected Shortfall using the POT method. Let $\widehat{\beta}$ and $\widehat{\gamma}$ denote the MLEs of the parameters in the GPD for the excess distribution F_u.

1)Show that we may estimate Value at Risk in the POT method via

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\widehat{\beta}}{\widehat{\gamma}}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\widehat{\gamma}} - 1\right).$$

2)Show that the POT method yields an estimate of the Expected Shortfall given by

$$\widehat{\mathrm{ES}}_\alpha = u + \frac{\widehat{\beta}}{\widehat{\gamma}}\left(\left(\frac{n}{N_u}\right)^{-\widehat{\gamma}} \frac{1}{(1-\alpha)^{\widehat{\gamma}}(1-\widehat{\gamma})} - 1\right)$$

whenever $\widehat{\gamma} < 1$.

Exercise 2.22:
Consider again the density for the GPD with parameters γ, β,

$$f_{\gamma,\beta}(x) = \frac{1}{\beta}\left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma - 1}.$$

Like above, we assume that we have excesses $z_1, ..., z_{N_u}$ which are assumed to follow this distribution.

1)Reparametrise the likelihood via the map $(\gamma, \beta) \mapsto (\gamma, \tau(\gamma, \beta))$ with $\tau(\gamma, \beta) = -\gamma/\beta$.

2)Show that the MLE of γ satisfies

$$\widehat{\gamma} = \frac{1}{N_u}\sum_{i=1}^{N_u}\log(1-\widehat{\tau}z_i)$$

with $\widehat{\tau}$ the MLE of τ.

3)Show that $\widehat{\tau}$ satisfies the equation

$$0 = \frac{1}{\widehat{\tau}} + \frac{1}{N_u}\left(1+\frac{1}{\widehat{\gamma}}\right)\sum_{i=1}^{N_u}\frac{z_i}{1-\widehat{\tau}z_i}.$$

4)Implement a function to simulate from the GPD. You can for example use the inverse transform method or a package. Choose values for γ and β and simulate a sample $z_1, ..., z_n$.

5)Implement a function (for example in R) to solve for $\widehat{\gamma}$ and $\widehat{\tau}$. Compare your MLE solver to the output from a package, for example `evir`.

Exercise 2.23:
The previous exercise concerned parameter estimation in a GPD. In this exercise, we investigate the much simpler case of an iid Pareto distributed sample $X_1, ..., X_n$ with parameters $\kappa, \alpha > 0$ where κ is considered fixed.

1)Write up the log-likelihood $l(\alpha; X_1, ..., X_n)$ and compute the score function $S(\alpha; X_1, ..., X_n) = \frac{\partial}{\partial\alpha}l(\alpha; X_1, ..., X_n)$.

2)Show that the MLE $\widehat{\alpha}$ of α is

$$\widehat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{X_i+\kappa}{\kappa}\right)\right)^{-1}.$$

3)Determine the exact distribution of $\widehat{\alpha}$. Hint: The distribution of $\log((X_1+\kappa)/\kappa)$ should be well-known. What distribution does one obtain when summing n iid of these?

4)Show that $\widehat{\alpha}$ is strongly consistent i.e. $\widehat{\alpha} \to \alpha$ a.s.

5)The Pareto distribution is sufficiently regular to apply classical statistical theory for the asymptotic distribution of $\sqrt{n}(\widehat{\alpha}-\alpha)$. More precisely, it holds that

$$\sqrt{n}(\widehat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}(0, I(\alpha)^{-1})$$

with $I(\alpha)$ the *Fisher information* for one observation given by

$$I(\alpha) = -E\left[\frac{\partial^2}{\partial\alpha^2}l(\alpha; X_1)\right].$$

Derive the asymptotic distribution of $\sqrt{n}(\widehat{\alpha}-\alpha)$. Compare with the asymptotics of the Hill estimator. Does the result surprise you?

Exercise 2.24:
Let N be Poisson distributed with mean $\lambda > 0$, and let $\{X_n\}$ be an iid sequence of random variables independent of N where X_1 has a GPD with parameters $\gamma > 0$ and $\beta > 0$. Define

$M_N := \max\{X_1, ..., X_N\}$. Prove that

$$P(M_N \leq x) = \exp\left(-\lambda\left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma}\right)$$

and show that this equals the distribution function $H_{\gamma,\mu,\psi}$ where $\mu = \beta\gamma^{-1}(\lambda^\gamma - 1)$ and $\psi = \beta\lambda^\gamma$. For the definition of $H_{\gamma,\mu,\psi}$, see the discussion right above the notes and comments.

Exercise 2.25:
In this exercise, we again consider the shortfall-to-quantile ratio as defined in Exercise 1.14. Let F be a continuous and strictly increasing regularly varying distribution function with index $\beta > 1$ and let $L \sim F$. In this exercise, we will verify that

$$\lim_{\alpha \to 1^-} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} = \frac{\beta}{\beta - 1}. \tag{2.3}$$

1)Show that

$$\mathrm{ES}_\alpha(L) = \mathrm{VaR}_\alpha(L) + \frac{1}{\overline{F}(\mathrm{VaR}_\alpha(L))} \int_{\mathrm{VaR}_\alpha(L)}^{\infty} \overline{F}(x)dx. \tag{2.4}$$

Hint: Use Proposition 1.3.4 and integration by parts.

2)Show that

$$\int_{\mathrm{VaR}_\alpha(L)}^{\infty} \overline{F}(x)dx \sim \frac{\mathrm{VaR}_\alpha(L)\overline{F}(\mathrm{VaR}_\alpha(L))}{\beta - 1} \quad \text{as} \quad \alpha \to 1^-.$$

Hint: Use Karamata's Theorem, Theorem 2.2.5.

3)Finish the proof of the result. Provide a risk management interpretation of the result.

4)Show that if L has the Student t distribution with $\nu > 1$ degrees of freedom, then the shortfall-to-quantile ratio is

$$\lim_{\alpha \to 1^-} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} = \frac{\nu}{\nu - 1}.$$

The following exercises contain a a detailed proof of (i) of Theorem 2.3.4. We start by providing the necessary background on order statistics.

Exercise 2.26:
Let $X_1, ..., X_n$ be iid with distribution function F, and let $U_1, ..., U_n$ be iid and uniformly distributed on $(0, 1)$. Prove that

$$(X_{1,n}, ..., X_{n,n}) \stackrel{\mathrm{d}}{=} (F^{\leftarrow}(U_{1,n}), ..., F^{\leftarrow}(U_{n,n})).$$

Exercise 2.27:
Let $X_1, ..., X_n$ be iid where X_1 has density f. In this exercise, we will prove that the density of $(X_{n,n}, ..., X_{1,n})$ is given by

$$f_{X_{n,n},...,X_{1,n}}(x_1, ..., x_n) = n! \prod_{i=1}^{n} f(x_i), \quad x_1 < \cdots < x_n.$$

1)Start by verifying that $X_1, ..., X_n$ has no ties with probability one. In other words, prove that the set

$$A = \{X_{n,n} < \cdots < X_{1,n}\} = \{X_i \neq X_j, i \neq j\}$$

has probability one. Hint: One possible approach is to use the tower property.

2)Now let S_n denote the set of permutations π of $\{1, ..., n\}$ and fix $x_1 < \cdots < x_n$. Introduce

$$A_\pi = \{X_{\pi(i)} = X_{i,n}, i = 1, ..., n\} \cap A \cap \{X_{\pi(1)} \leq x_1, ..., X_{\pi(n)} \leq x_n\}.$$

Argue that

$$P(X_{n,n} \leq x_1, ..., X_{1,n} \leq x_n) = P\left(\bigcup_{\pi \in \Pi_n} A_\pi\right) = \sum_{\pi \in \Pi_n} P(A_\pi).$$

3)Show that

$$P(A_\pi) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f(y_i) 1_{\{y_1 < \cdots y_n\}} dy_n \cdots dy_1.$$

4)Collect the pieces and finish the proof.

Exercise 2.28:
We now verify some useful representation results for the order statistics of iid uniform variables on $(0, 1)$. Let $\{U_n\}$ be iid uniformly distributed on $(0, 1)$, and let $\{E_n\}$ be iid standard exponential. Define $\Gamma_n = E_1 + \cdots + E_n$.

1)Prove that

$$(U_{1,n}, U_{2,n} ..., U_{n,n}) \stackrel{\mathrm{d}}{=} \left(\frac{\Gamma_n}{\Gamma_{n+1}}, \frac{\Gamma_{n-1}}{\Gamma_{n+1}}, ..., \frac{\Gamma_1}{\Gamma_{n+1}}\right).$$

Hints: Show that the densities on the two sides are equal. For the left hand side, use the result of the previous exercise. For the right hand side, write up the density of $(E_1, ..., E_{n+1})$ and apply a suitable density transformation theorem (it may be an idea to do this in several steps).

2)Prove that

$$(1 - U_{1,n}, U_{1,n} - U_{2,n}, ..., U_{n,n}) \stackrel{\mathrm{d}}{=} \left(\frac{E_{n+1}}{\Gamma_{n+1}}, \frac{E_n}{\Gamma_{n+1}}, ..., \frac{E_1}{\Gamma_{n+1}}\right).$$

A different solution to this exercise is provided in [101].

Exercise 2.29:
Now assume X has regularly varying tail $\overline{F}(x) = x^{-\alpha} L_0(x)$ where $L_0 \in \mathrm{RV}_0$. By Exercise 2.26 above, we may assume that the Hill estimator has representation

$$\widehat{\alpha}_k^{-1} = \frac{1}{k} \sum_{i=1}^{k} \log(F^{\leftarrow}(U_{i,n})) - \log(F^{\leftarrow}(U_{k,n})).$$

From Corollary 2.3.4 in [8], we have that

$$F^{\leftarrow}(t) = (1-t)^{-1/\alpha} L((1-t)^{-1})$$

for some $L \in \mathrm{RV}_0$. We will take this result as given.

1)Show that $\widehat{\alpha}_k^{-1}$ has the representation $\widehat{\alpha}_k^{-1} = \widetilde{\alpha}_k^{-1} + r_k$ where

$$\widetilde{\alpha}^{-1} = \frac{1}{\alpha k} \sum_{i=1}^{k} \log\left(\frac{\Gamma_{n+1} - \Gamma_{n-k+1}}{\Gamma_{n+1} - \Gamma_{n-i+1}}\right),$$
$$r_k = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{L(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-i+1}))}{L(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-k+1}))}\right).$$

2)It turns out that $\widetilde{\alpha}^{-1} \xrightarrow{P} \alpha^{-1}$ while r_k is a bias term. Argue, using the previous exercise, that

$$\widetilde{\alpha}^{-1} \stackrel{\mathrm{d}}{=} \frac{1}{\alpha k} \sum_{i=1}^{k} E_i$$

with $E_1, ..., E_k$ iid standard exponential variables. Conclude that $\widetilde{\alpha}^{-1} \xrightarrow{P} \alpha^{-1}$ under the assumption that $k \to \infty$ as $n \to \infty$.

3)Prove that

$$\sqrt{k}(\widetilde{\alpha}^{-1} - \alpha^{-1}) \xrightarrow{d} \mathcal{N}(0, \alpha^2).$$

This shows that the real difficulties in proving Theorem 2.3.4 lie in controlling the bias term r_k.

4)It remains to show that $r_k \to 0$ if $k \to \infty$ and $k/n \to 0$. We know that L has a Karamata representation (see Corollary 2.2.6) given by

$$L(x) = c(x) \exp\left(\int_{x_0}^{x} \frac{\epsilon(t)}{t} dt\right)$$

with $c(x) \to c_0 > 0$ and $\epsilon(x) \to 0$ as $x \to \infty$. Show that r_k has the decomposition $r_k = r_k^{(1)} + r_k^{(2)}$ with

$$r_k^{(1)} = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{c(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-i+1}))}{c(\Gamma_{n+1}/(\Gamma_{n+1} - \Gamma_{n-k+1}))}\right),$$
$$r_k^{(2)} = \frac{1}{k} \sum_{i=1}^{k} \int_{(1-\Gamma_{n-k+1}/\Gamma_{n+1})^{-1}}^{(1-\Gamma_{n-i+1}/\Gamma_{n+1})^{-1}} \frac{\epsilon(t)}{t} dt.$$

5)Show that if $k \to \infty$ and $k/n \to 0$, then

$$\frac{\Gamma_{n-i+1}}{\Gamma_{n+1}} \to 1 \quad \text{a.s.}$$

Conclude that this implies $r_k^{(1)} \to 0$.

6)We now only need to show that $r_k^{(2)} \to 0$. Define

$$C_n = \sup\{|\epsilon(t)| : t \geq (1 - \Gamma_{n-k+1}/\Gamma_{n+1})^{-1}\}.$$

Argue that $C_n \in \mathbb{R}$ for all n. Show that $C_n \to 0$ a.s.

7)Show that $r_k^{(2)} \leq C_n \alpha \widetilde{\alpha}^{-1}$ and finish the proof of Theorem 2.3.4 (i).

The difficulty in establishing asymptotic normality lies in controlling $\sqrt{k} r_k$. This requires conditions on the second order behaviour of L_0.

The remaining exercises concern the supplementary material on Maximal Domains of Attraction.

Exercise 2.30:
Prove Lemma 2.5.7.

Exercise 2.31:
In this exercise, we provide further examples of which MDA some well-known distributions belong to.

1)Show that the exponential distribution is in MDA(Λ) by explicitly determining the normalising constants.

2)Consider the Cauchy distribution with density $f(x) = 1/(\pi(1+x^2))$ for $x \in \mathbb{R}$. Determine the MDA of this distribution.

3)Consider the log-normal distribution. Show that this distribution is in MDA(Λ). Does this surprise you? Hint: Use the normalising constants for the normal distribution provided in the text.

Practical exercises

Exercise 2.32:
Consider the distribution F with tail $\overline{F}(x) = x^{-2}$ for $x > 1$. Then $\overline{F} \in \mathrm{RV}_{-2}$.

1)Implement the Hill estimator (in R for example) and a function that can simulate values from F.

2)Simulate 50 values of F and make a Hill plot using your function from before. Also plot the true value of the index as a line in the plot.

3)Plot the asymptotic confidence bands obtained from Theorem 2.3.4.

4)Repeat for 100, 250 and 1000 values. Comment on the results.

Exercise 2.33: DAX-returns
In this exercise, we again consider the DAX-returns. The dataset consists of daily log-returns of the DAX stock index over the period from January 1990 until July 1996.

1)Make a QQ-plot of the log-return data against the normal distribution and against t distributions with different degrees of freedom. Which distribution seems to give a good fit?

2)Multiply the data by -1 so that negative returns appear along the positive axis. Construct a mean-excess plot. Does it seem to be linear? If so, then above which level u (and corresponding exceedance number N_u) is the mean-excess linear?

3)Extract the excesses over your chosen level u and estimate the parameters (γ, β) of a fitted GPD.

4)Make diagnostic plots. Is your model an adequate fit?

5)Determine asymptotic 95%-confidence intervals for the parameters γ and β, respectively.

6)Give an expression for the empirical counterpart of your fitted GPD approximation of $1 - F(u + x)$, $x > 0$ as a function of x and compare the associated distribution functions in a plot.

7)Use your estimated values for γ and β to estimate the one-day $\text{VaR}_{0.99}$ for the loss of a portfolio consisting of one share of the DAX index, assuming that today's price is 100.

Exercise 2.34: Bitcoin data
In this exercise, we use the POT method on Bitcoin price data.

1)Start by loading the `cryptoQuotes` ([77]) and `quantmod` ([106]) packages into `R`. You can then fetch Bitcoin price data with the following code.

```
library(cryptoQuotes)
library(quantmod)

btc <- get_quote(
  ticker = 'BTCUSDT',
  source = 'binance',
  interval = '1d',
  futures = FALSE,
  from = "2025-01-01",
  to = "2025-11-01"
)
```

You can choose different intervals (`1d`, `1h` etc.) and a different time period if you like. When you have loaded the data, compute the log-returns and select the negative values. This is the data we will work with for the rest of the exercise. Make some exploratory plots. Hint: Use the functions `coredata` and `index` to separate dates from the actual price data.

2)Use a mean-excess plot to find a suitable threshold u where the excesses above u are approximately GPD distributed.

3)Fit the GPD with maximum likelihood.

4)Make diagnostic plots. Is your model an adequate fit?

5)Plot the empirical distribution function of the excess losses in the same plot as the fitted GPD distribution function.

6)Use the POT-based estimator to compute $\text{VaR}_{0.99}$ and $\text{ES}_{0.99}$.

Exercise 2.35: Norwegian fire data
In this exercise, we consider the `norfire` dataset from the R package `CASdatasets`. The dataset contains 9181 fire insurance claims over the period 1972 to 1992 from an unknown Norwegian company. We are interested in the column `Loss2012` which contains the claim sizes in 2012 prices. We wish to analyse this dataset and compute estimates of the Value at Risk and Expected Shortfall using all the methods developed in this chapter.

1)Start by separating the data into two, a vector of claims from before 1992 and one for 1992. Make some exploratory plots to get a feel for the data.

Let us now assume that the data comes from a regularly varying distribution. We are of course interested in estimating the index α.

2)Now make Hill plots for both datasets. What are good choices of k in the two cases? What are your Hill estimates for α?

3)Augment the Hill plots from the previous exercise with the asymptotic confidence bands obtained from Theorem 2.3.4. Is there a significant difference between the two Hill estimates? Are fire losses from before 1992 more or less heavy-tailed?

For the remainder of the exercise, you can choose to work with only one of the two datasets or with both.

4)Compute estimates of the Value at Risk and Expected Shortfall using your Hill estimates of α. Compare these estimates with the ones obtained from a purely empirical estimate. Which estimates do you prefer and why? Hint: Use Exercise 2.11.

Instead of assuming that the data necessarily comes from a regularly varying distribution, we instead apply the POT method.

5)Make mean-excess plots and choose a value of the threshold u that you find proper.

6)Extract the excesses over u and fit a GPD. Report the parameters.

7)Compare the empirical tail(s) of the data with tail(s) from the fitted GPD(s).

8)Make diagnostic plots. Is your model an adequate fit?

9)Provide asymptotic 95% confidence intervals of the parameters γ and β.

10) Use your estimated values of γ and β to compute $\text{VaR}_{0.99}$ and $\text{ES}_{0.99}$ for the data. Compare with the previous two estimates of each. Hint: Use the results of Exercise 2.21.

11) Make a brief executive summary of your findings in this exercise using non-technical lingo. What would you report to the board of directors for a company with this portfolio?

Chapter 3

Spherical and elliptical distributions

Multivariate random vectors: dependence

Consider again the "canonical example" of stock returns where the risk factors are the log returns $\mathbf{X}_{n+1}$ with $X_{n+1}^{(i)} = \log S_{n+1}^{(i)} - \log S_n^{(i)}$. What is the distribution of $\mathbf{X}_{n+1}$? Inspired by the Black–Scholes model, we could assume $\mathbf{X}_{n+1} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. There are several problems with the normal assumption, however. A typical problem is that assets are very often correlated in such a way that high (low) returns for one asset correlates with high (low) returns for another. The normal distribution is very light-tailed, so such correlations are often not captured. The plots below illustrate this issue.

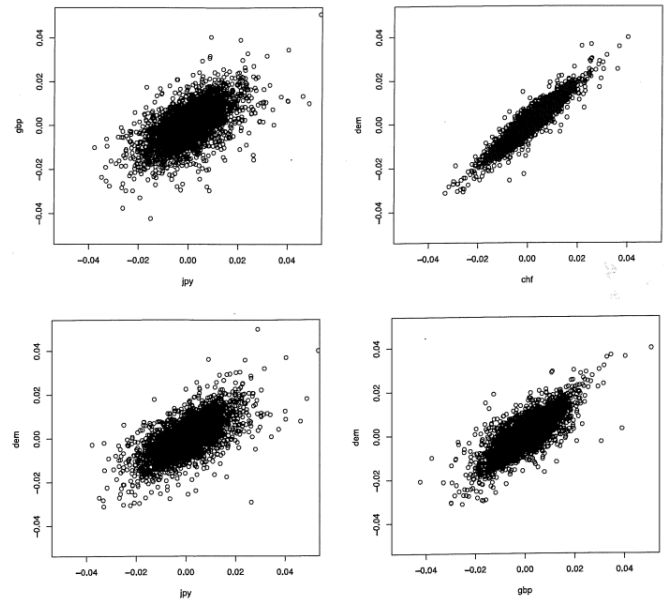


Figure 1: Log returns of foreign exchange rates quotes against the US dollar.

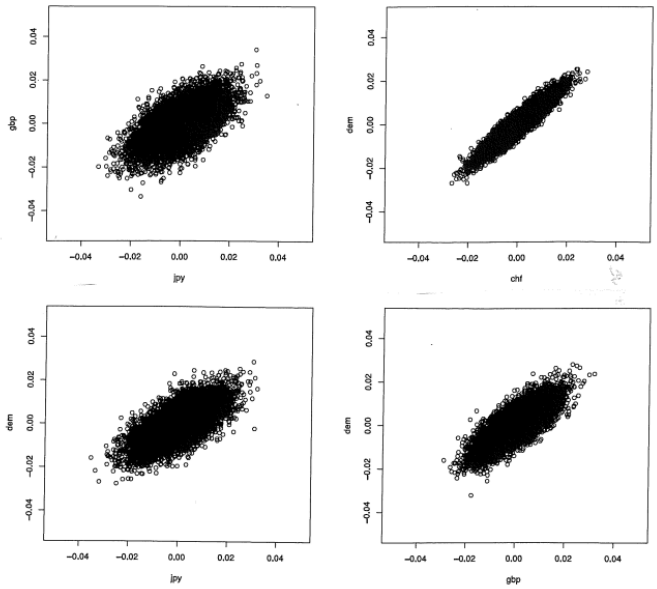


Figure 2: Simulated foreign exchange rates using a bivariate normal distribution with estimated means and covariance matrix.

From the plots, it is evident that the normal distribution fails to capture the dependency in the tails. Furthermore, the probability mass is too concentrated around the mean. The dependency in tails is a very typical phenomenon in financial data as illustrated in the plot below.

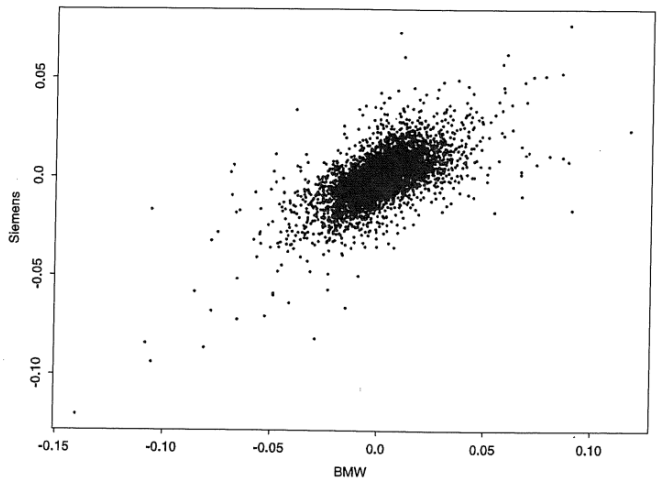


Figure 3: Log returns from BMW and Siemens stocks.

To remedy the issues with the normal distribution, we introduce a class of distributions that in some way resembles the normal distribution and shares a lot of its properties while also

being more flexible in terms of modelling tail behaviour. This is the class of spherical and elliptical distributions.

3.1 Spherical and elliptical distributions

To motivate the spherical and elliptical distributions, we first briefly consider the multivariate normal distribution. If $\mathbf{X} \sim \mathcal{N}(0, I_d)$ (I_d denotes the d-dimensional identity matrix), then $\mathbf{X}$ has density (see the appendix)

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\sum_{i=1}^{d} x_i^2} = \frac{1}{(2\pi)^{d/2}} e^{-r^2/2}$$

with $r^2 = x_1^2 + \cdots + x_d^2$. Hence the density only depends on $\|\mathbf{x}\| = r$. Graphically, the level sets of f are spheres (or circles in two dimensions). One can say that $\mathcal{N}(0, I_d)$ is spherically symmetric/rotationally invariant. Define the random variable R by $R^2 = X_1^2 + \cdots + X_d^2$, then $R^2 \sim \chi^2(d)$ i.e. R^2 is Chi-square distributed with d degrees of freedom. We call R the *radial component* of $\mathbf{X}$. Intuitively, we can decompose $\mathbf{X}$ as $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$ with $\mathbf{S}$ uniformly distributed on the d-dimensional unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^{d} x_i^2 = 1\}$. While this is an informal approach, it gives us the idea on how to proceed formally.

Definition 3.1.1. For a d-dimensional random vector $\mathbf{X}$, we define the *characteristic function* of $\mathbf{X}$ as

$$\Phi_{\mathbf{X}}(\mathbf{t}) = E[e^{i\mathbf{t}^T\mathbf{X}}], \quad \mathbf{t} \in \mathbb{R}^d.$$

Remark 3.1.2. Note the similarity to the moment-generating function $\kappa_{\mathbf{X}}(\mathbf{t}) = E[e^{\mathbf{t}^T\mathbf{X}}]$. These two transforms satisfy similar properties. For example, two random variables have the same distribution if and only if their characteristic functions are equal. See the appendix for more background on these functions. The characteristic function has the advantage that it always exists (since the integrand is bounded by one in norm) but it provides less information about the tail behaviour than the moment-generating function.

Example 3.1.3. If $\mathbf{X} \sim \mathcal{N}(0, I_d)$, simple calculations yield

$$\Phi_{\mathbf{X}}(\mathbf{t}) = e^{-\frac{1}{2}\mathbf{t}^T\mathbf{t}}.$$

Note that we can write $\Phi_{\mathbf{X}}(\mathbf{t}) = \psi(\|\mathbf{t}\|^2)$ for the function $\psi : \mathbb{R} \to \mathbb{R}$ given by $\psi(t) = e^{-t/2}$. This is a formal way of stating that $\Phi_{\mathbf{X}}$ doesn't depend on the direction of $\mathbf{t}$. ∘

We can now introduce spherical distributions.

Definition 3.1.4. A random vector $\mathbf{X}$ in d dimensions has a *spherical distribution* if

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \psi(\|\mathbf{t}\|^2) = \psi(t_1^2 + \cdots + t_d^2), \quad \mathbf{t} \in \mathbb{R}^d$$

for some univariate function ψ. ψ is called the *characteristic generator* of $\mathbf{X}$, and we write $\mathbf{X} \sim S_d(\psi)$.

If $\mathbf{X} \sim S_d(\psi), \mathbf{t} \in \mathbb{R}^d$ and $\mathbf{X}^\theta$ denotes $\mathbf{X}$ rotated by θ (and similarly for $\mathbf{t}^\theta$), we have

$$\Phi_{\mathbf{X}^\theta}(\mathbf{t}^\theta) = E\left[e^{i\langle \mathbf{t}^\theta, \mathbf{X}^\theta\rangle}\right] = E\left[e^{i\langle \mathbf{t}, \mathbf{X}\rangle}\right] = \Phi_{\mathbf{X}}(\mathbf{t}) = \psi(\|\mathbf{t}^\theta\|^2) = \Phi_{\mathbf{X}}(\mathbf{t}^\theta)$$

which is true for all $\mathbf{t}^\theta$. By the uniqueness of the characteristic function, $\mathbf{X}^\theta \overset{\mathrm{d}}{=} \mathbf{X}$. This gives a formal argument for the intuition of $\mathbf{X}$ being rotationally invariant. The following result gives an equivalent formulation of spherical distributions.

Proposition 3.1.5. *The following are equivalent:*

(i) $\mathbf{X}$ *has a spherical distribution.*

(ii) $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$ *with* $\mathbf{S}$ *uniformly distributed on the unit sphere* $\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ *and* R *is a one-dimensional random variable independent of* $\mathbf{S}$.

Proof. We first prove that (ii) implies (i). We have

$$\begin{aligned}\Phi_{\mathbf{X}}(\mathbf{t}) = \Phi_{R\mathbf{S}}(\mathbf{t}) &= E\left[e^{i\langle \mathbf{t}, R\mathbf{S}\rangle}\right] = E\left[E\left[e^{i\langle \mathbf{t}, R\mathbf{S}\rangle} \mid R\right]\right] \\ &= E\left[E\left[e^{i\langle R\mathbf{t}, \mathbf{S}\rangle} \mid R\right]\right] = E[\Phi_{\mathbf{S}}(R\mathbf{t})]\end{aligned}$$

and since $\mathbf{S}$ is uniformly distributed on the unit sphere, $\Phi_{\mathbf{S}}$ only depends on the length and not the direction. Hence $\Phi_{\mathbf{X}}(\mathbf{t})$ also only depends on the length of $\mathbf{t}$ and $\mathbf{X}$ has a spherical distribution. We now show that (i) implies (ii). We have $\Phi_{\mathbf{X}}(\mathbf{t}) = \psi(\|\mathbf{t}\|^2)$. Set $\mathbf{s} = \mathbf{t}/\|\mathbf{t}\|$, then

$$\Phi_{\mathbf{X}}(\mathbf{t}) = E\left[e^{i\|\mathbf{t}\|\langle \mathbf{s}, \mathbf{X}\rangle}\right],$$

and by assumption, this does not depend on $\mathbf{s}$, only $\|\mathbf{t}\|$. Let $\mathbf{S}$ be uniformly distributed on the unit sphere with distribution function $F_{\mathbf{S}}$. Since $\Phi_{\mathbf{X}}(\mathbf{t})$ is constant in $\mathbf{s}$, we have

$$\begin{aligned}\Phi_{\mathbf{X}}(\mathbf{t}) &= \int_{\mathbb{S}^{d-1}} E\left[e^{i\|\mathbf{t}\|\langle \mathbf{s}, \mathbf{X}\rangle}\right] dF_{\mathbf{S}}(\mathbf{s}) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} e^{i\|\mathbf{t}\|\langle \mathbf{s}, \mathbf{x}\rangle} dF_{\mathbf{X}}(\mathbf{x}) dF_{\mathbf{S}}(\mathbf{s}) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} e^{i\|\mathbf{t}\|\langle \mathbf{s}, \mathbf{x}\rangle} dF_{\mathbf{S}}(\mathbf{s}) dF_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} e^{i\langle \mathbf{s}, \|\mathbf{t}\|\mathbf{x}\rangle} dF_{\mathbf{S}}(\mathbf{s}) dF_{\mathbf{X}}(\mathbf{x}) \\ &= \int_{\mathbb{R}^d} E\left[e^{i\langle \|\mathbf{t}\|\mathbf{x}, \mathbf{S}\rangle}\right] dF_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^d} E\left[e^{i\langle \|\mathbf{x}\|\mathbf{t}, \mathbf{S}\rangle}\right] dF_{\mathbf{X}}(\mathbf{x}) \\ &= E\left[E\left[e^{i\langle \|\mathbf{X}\|\mathbf{t}, \mathbf{S}\rangle} \mid \mathbf{X}\right]\right] = E\left[e^{i\langle \mathbf{t}, \|\mathbf{X}\|\mathbf{S}\rangle}\right] = E\left[e^{i\langle \mathbf{t}, R\mathbf{S}\rangle}\right] = \Phi_{R\mathbf{S}}(\mathbf{t})\end{aligned}$$

where we have defined $R := \|\mathbf{X}\|$. Hence $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$ where R and $\mathbf{S}$ have the desired properties. ∎

While characterisation (ii) is more intuitive, it is easier to work with definition (i) when one wants to prove properties of spherical distributions. The following corollary tells how to compute R and $\mathbf{S}$ when we know that $\mathbf{X}$ is spherical.

Corollary 3.1.6. *Let* $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$ *be spherical. Then*

$$\left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|}\right) \overset{\mathrm{d}}{=} (R, \mathbf{S}).$$

Proof. The proof is from [88], see Corollary 6.22. Let $f_1(\mathbf{x}) = \|\mathbf{x}\|$ and $f_2(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$. Since $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$, we have

$$\left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|}\right) = (f_1(\mathbf{X}), f_2(\mathbf{X})) \overset{\mathrm{d}}{=} (f_1(R\mathbf{S}), f_2(R\mathbf{S})) = (R, \mathbf{S})$$

as desired. ∎

We now turn to a generalisation of spherical distributions, namely elliptical distributions.

Definition 3.1.7. A d-dimensional random vector $\mathbf{X}$ has an *elliptical distribution* if $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$ with $\mathbf{Y} \sim S_k(\psi)$ and A is a $d \times k$ matrix. We write $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ for $\Sigma = AA^T$. We call $\boldsymbol{\mu}$ the *location parameter* and Σ the *dispersion matrix*.

As a motivation for this definition, suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ where Σ is positive definite. From linear algebra (see for example chapter 7 in [5]), we know that there exists some matrix A such that $AA^T = \Sigma$. If $Y \sim \mathcal{N}(0, I_d)$, then

$$\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A\mathbf{Y}.$$

There exist methods to find A such that $AA^T = \Sigma$. One such method is the *Cholesky factorisation*. This factorisation determines a lower triangular matrix A such that $AA^T = \Sigma$. In detail,

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{d1} \\ 0 & a_{22} & \cdots & a_{d2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{dd} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1d} \\ \Sigma_{21} & \Sigma_{22} & & \Sigma_{2d} \\ \vdots & \vdots & & \vdots \\ \Sigma_{d1} & \Sigma_{d2} & \cdots & \Sigma_{dd} \end{pmatrix}.$$

The algorithm can (somewhat informally) be described as follows: Since $\Sigma_{11} = a_{11}^2$, Σ_{11} determines a_{11}. Since $\Sigma_{21} = a_{11}a_{21}$, Σ_{21} determines a_{21} and so on. Since we can go back and forth between the matrices A and Σ, the notation $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ makes sense. Using the characterisation of spherical distributions in terms of a radial component, we can also write

$$\mathbf{X} = \boldsymbol{\mu} + RA\mathbf{S}$$

for $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$. We now turn to some properties of elliptical distributions. Afterwards, we take a look at some examples.

3.2 Properties of elliptical distributions

We start by computing the characteristic function for an elliptical distribution.

Lemma 3.2.1. *If $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, then*

$$\Phi_{\mathbf{X}}(\mathbf{t}) = e^{i\langle \mathbf{t}, \boldsymbol{\mu} \rangle} \psi(\mathbf{t}^T \Sigma \mathbf{t}).$$

Proof. The proof is a straightforward computation. Write $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$ with $\mathbf{Y} \sim S_k(\psi)$. We have

$$\Phi_{\mathbf{X}}(\mathbf{t}) = E\left[e^{i\langle \mathbf{t}, \mathbf{X} \rangle}\right] = e^{i\langle \mathbf{t}, \boldsymbol{\mu} \rangle} E\left[e^{i\langle \mathbf{t}, A\mathbf{Y} \rangle}\right]$$

and since $\langle \mathbf{t}, A\mathbf{Y} \rangle = \mathbf{t}^T A\mathbf{Y} = (A^T\mathbf{t})^T\mathbf{Y} = \langle A^T\mathbf{t}, \mathbf{Y} \rangle$, we have

$$\begin{aligned} E\left[e^{i\langle \mathbf{t}, A\mathbf{Y} \rangle}\right] = E\left[e^{i\langle A^T\mathbf{t}, \mathbf{Y} \rangle}\right] &= \psi(\|A^T\mathbf{t}\|^2) = \psi((A^T\mathbf{t})^T A^T\mathbf{t}) \\ &= \psi(\mathbf{t}^T AA^T\mathbf{t}) = \psi(\mathbf{t}^T \Sigma \mathbf{t}) \end{aligned}$$

as desired. ∎

We want to be able to relate covariances to the dispersion. The following proposition shows how these are related for elliptical distributions.

Proposition 3.2.2. *If $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, the covariance between the components is given by*

$$\operatorname{Cov}(X_j, X_l) = -2\psi'(0)\Sigma_{jl}.$$

Proof. Since covariance does not depend on the mean, we can without loss of generality assume that $E[X_j] = 0$ for all $j = 1, ..., d$. We have

$$\begin{aligned}
\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}\Phi_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=0} &= \frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}E\left[e^{i\langle \mathbf{t}, \mathbf{X}\rangle}\right]\Big|_{\mathbf{t}=0} = \frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}E\left[e^{i(t_1X_1+\cdots t_dX_d)}\right]\Big|_{\mathbf{t}=0} \\
&= E\left[(iX_j)(iX_l)e^{i(t_1X_1+\cdots t_dX_d)}\right]\Big|_{\mathbf{t}=0} = -E\left[X_jX_le^{i(t_1X_1+\cdots t_dX_d)}\right]\Big|_{\mathbf{t}=0} \\
&= -E\left[X_jX_l\right] = -\operatorname{Cov}(X_j, X_l)
\end{aligned}$$

and thus by the previous lemma,

$$\operatorname{Cov}(X_j, X_l) = -\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}\Phi_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=0} = -\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}\psi(\mathbf{t}^T\Sigma\mathbf{t})\Big|_{\mathbf{t}=0}.$$

Let us for simplicity assume that $d = 2$. Then

$$\mathbf{t}^T\Sigma\mathbf{t} = \begin{pmatrix} t_1 & t_2 \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = t_1^2\Sigma_{11} + 2t_1t_2\Sigma_{12} + t_2^2\Sigma_{22} =: w(\mathbf{t})$$

and thus

$$\begin{aligned}
\frac{\partial}{\partial t_1}\frac{\partial}{\partial t_2}\psi(\mathbf{t}^T\Sigma\mathbf{t})\Big|_{\mathbf{t}=0} &= \frac{\partial}{\partial t_1}\left(\psi'(w(\mathbf{t}))(2t_1\Sigma_{12} + 2t_2\Sigma_{22})\right)\Big|_{\mathbf{t}=0} \\
&= \psi''(w(\mathbf{t}))(2t_1\Sigma_{11} + 2t_2\Sigma_{12})(2t_1\Sigma_{12} + 2t_2\Sigma_{22}) + \psi'(w(\mathbf{t}))2\Sigma_{12}\Big|_{\mathbf{t}=0} \\
&= 2\psi'(0)\Sigma_{12}.
\end{aligned}$$

This calculation can be generalised so that $\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_l}\psi(\mathbf{t}^T\Sigma\mathbf{t})\Big|_{\mathbf{t}=0} = 2\psi'(0)\Sigma_{jl}$. We conclude that

$$\operatorname{Cov}(X_j, X_l) = -2\psi'(0)\Sigma_{jl}.$$

■

Example 3.2.3. If $\mathbf{Y} \sim \mathcal{N}(0, I_d)$, then $\psi(r) = e^{-r/2}$ as seen earlier. We see that $\psi'(r) = -\frac{1}{2}e^{-r/2}$, so $\psi'(0) = -\frac{1}{2}$. If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, the above proposition tells us that $\operatorname{Cov}(X_j, X_l) = -2\psi'(0)\Sigma_{jl} = \Sigma_{jl}$ as expected. ∘

We list some further properties of elliptical distributions.

Theorem 3.2.4. *Let $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$.*

(i) (Linear combinations). If B is a $k \times d$ matrix and $\mathbf{b} \in \mathbb{R}^k$, then

$$B\mathbf{X} + \mathbf{b} \sim E_k(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B^T, \psi).$$

(ii) *(Marginal distributions). The marginals $X_1, ..., X_d$ also have elliptical distributions with the same characteristic generator. Explicitly, $X_i \sim E_1(\mu_i, \Sigma_{ii}, \psi)$.*

(iii) *(Convolutions). If $\tilde{\mathbf{X}} \sim E_d(\tilde{\boldsymbol{\mu}}, \Sigma, \tilde{\psi})$ is another elliptical distribution independent of $\mathbf{X}$ with the same dimension and dispersion matrix, then $\mathbf{X} + \tilde{\mathbf{X}} \sim E_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma, \bar{\psi})$ with $\bar{\psi}(u) = \psi(u)\tilde{\psi}(u)$.*

(iv) *(Quadratic forms). Whenever A is quadratic, we have*

$$R^2 = \|\mathbf{Y}\|^2 = (\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}).$$

Here, R is called the Mahalanobis distance.

Proof. See the exercises. ■

While elliptical distributions are quite flexible as seen in the examples earlier, the above theorem also illustrates a drawback. The flexibility of elliptical distributions is limited by the fact that if $\mathbf{X}$ is elliptical, so is any coordinate. In real data, it is often the case that the marginals have very different types of distributions. We will see techniques for handling this in the next chapter.

Normal variance mixture models

A useful class of elliptical distributions is the class of normal variance mixtures.

Definition 3.2.5 (**Normal Variance Mixture Models**). Let $\mathbf{Z} \sim \mathcal{N}(0, I_k)$, $W \geq 0$ a random variable and A a fixed $d \times k$ matrix. A *normal variance mixture model* is a model of the form

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{W} A \mathbf{Z}$$

with $\mathbf{Z}$ and W independent.

It follows by construction that, conditional on $W = w$,

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, w\Sigma), \quad \Sigma = AA^T.$$

Thus, $\mathbf{X}$ is obtained by drawing from a collection of normal random variables with random covariance $W\Sigma$. Determining the distribution of a normal variance mixture model is sometimes possible, and we shall see examples shortly. We first make some general considerations. For $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W} A \mathbf{Z}$ a normal variance mixture model, it is easily verified that the covariance $\mathrm{Var}[\mathbf{X}]$ exists whenever $E[W] < \infty$, and in that case,

$$\mathrm{Var}[\mathbf{X}] = E[W]\Sigma. \tag{3.1}$$

We stress that despite the somewhat confusing notation, Σ is in general **not** the covariance matrix of $\mathbf{X}$. Another useful fact about normal variance mixture distributions is the following expression for the density, valid when Σ is invertible and W has density f_W:

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \frac{1}{\sqrt{(2\pi w)^d \det \Sigma}} \exp\left(-\frac{R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)}{2w}\right) f_W(w) dw. \tag{3.2}$$

where $R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$ denotes the Mahalanobis distance and F_W the distribution function of W. All these facts are easily established via conditioning arguments and are left as exercises for the reader.

Example 3.2.6. The multivariate normal distribution is clearly a normal variance mixture model. Here we have $W = 1$, and the identity $\mathrm{Var}[\mathbf{X}] = \Sigma$ concurs with the general formula (3.1) just presented. ∘

Example 3.2.7 (t **distribution**)**.** Here we consider a special case of a normal variance mixture model, namely if we let

$$W \sim \mathrm{Ig}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

where $\mathrm{Ig}(\alpha, \beta)$ denotes the *inverse gamma distribution* with density

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\beta/x}.$$

In particular, we have $\nu/W \sim \chi^2_\nu$, the Chi-square distribution with ν degrees of freedom, and $E[\mathbf{X}]$ exists when $\nu > 1$ and is in this case equal to $\boldsymbol{\mu}$. For $AA^T = \Sigma$, we write

$$\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$$

and we call this distribution the multivariate t distribution. From the above discussion, it follows that the covariance $\mathrm{Var}[\mathbf{X}]$ exists whenever $\nu > 2$, in which case

$$\mathrm{Var}[\mathbf{X}] = \frac{\nu}{\nu - 2}\Sigma$$

and that the density is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma((d+\nu)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}\sqrt{\det \Sigma}} \left(1 + \frac{R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)}{\nu}\right)^{-\frac{d+\nu}{2}}.$$

∘

An application: Portfolio investment theory

Say we want to minimise the risk in a portfolio with d assets with returns $\mathbf{X} = (X_1, ..., X_d)$, $\mu_i = E[X_i]$. The following ideas go back to Markowitz. Let R_p denote the total returns, i.e.

$$R_p = \sum_{i=1}^{d} w_i X_i = \mathbf{w}^T \mathbf{X}$$

where w_i are the weights of the portfolio, $\sum_{i=1}^d w_i = 1$. If we fix the expected total returns $E[R_p] = \sum_{i=1}^d w_i \mu_i = \mathbf{w}^T \boldsymbol{\mu}$, we want to minimise the risk in the sense of minimising the variance (note that this approach is different from the strategy in this course, where we focus on risk measures). If $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$ where $\mathbf{Y} \sim \mathcal{N}(0, I_d)$, then

$$\mathrm{Var}(R_p) = \mathrm{Var}(\mathbf{w}^T(\boldsymbol{\mu} + A\mathbf{Y})) = \mathrm{Var}(\mathbf{w}^T A\mathbf{Y})$$

and since $\mathbf{Y}$ is spherical, the variance is minimised whenever $\|A^T\mathbf{w}\|$ is minimized with respects to the weights w_i.

To transfer these ideas to the setting in this course, let ρ be a risk measure which satisfies monotonicity and translation invariance. Assume more generally that $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$ is

elliptical. The goal now is to minimise $\rho(L)$ where $L = -\mathbf{w}^T\boldsymbol{\mu} - \mathbf{w}^T A\mathbf{Y}$ is the loss. We have

$$\rho(L) = -\mathbf{w}^T\boldsymbol{\mu} + \rho(-\mathbf{w}^T A\mathbf{Y}).$$

As $\mathbf{Y}$ is spherical, we have $\mathbf{Y} \overset{\mathrm{d}}{=} -\mathbf{Y}$, so we can remove the minus in the risk measure and obtain

$$\rho(L) = -\mathbf{w}^T\boldsymbol{\mu} + \rho(\mathbf{w}^T A\mathbf{Y}) = -E[R_p] + \rho(\mathbf{w}^T A\mathbf{Y}).$$

As $E[R_p]$ is fixed, $\rho(L)$ is minimised whenever $\rho(\mathbf{w}^T A\mathbf{Y})$ is minimised. As $\mathbf{Y}$ is spherical, $\rho(\mathbf{w}^T A\mathbf{Y})$ is minimised when $\|A^T\mathbf{w}\|$ is minimised. Thus the answer remains the same as in the classical case above.

Further properties and connections to VaR

The final topic of this chapter is the connection between elliptical distributions and properties of risk measures. The main highlight is the fact that for elliptical distributions, VaR_α is subadditive whenever $\alpha \geq 1/2$. Proving this requires some work. We start by providing further characterisations of spherical distributions. Recall that a square matrix U with real-valued entries is *orthogonal* if $U^TU = I = UU^T$.

Lemma 3.2.8. *Let $\mathbf{X}$ be a d-dimensional random variable.*

(i) $\mathbf{X}$ *has a spherical distribution.*

(ii) For any orthogonal matrix U, it holds that $U\mathbf{X} \overset{\mathrm{d}}{=} \mathbf{X}$.

(iii) For every $\mathbf{a} \in \mathbb{R}^d$, it holds that $\mathbf{a}^T\mathbf{X} \overset{\mathrm{d}}{=} \|\mathbf{a}\|X_1$.

Proof. See Exercise 3.7. ■

Now consider the class of *linear portfolios* given by loss random variables of the form $L = m + \mathbf{a}^T\mathbf{X}$, where $m \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^d$ and $\mathbf{X}$ is a fixed d-dimensional random vector. We let

$$\mathcal{M}_{\mathbf{X}} := \{L : L = m + \mathbf{a}^T\mathbf{X}, m \in \mathbb{R}, \mathbf{a} \in \mathbb{R}^d\}$$

denote the set of linear portfolios belonging to $\mathbf{X}$. The following assumption on a risk measure is natural in most contexts.

Definition 3.2.9. A risk measure ρ is *law-invariant* if whenever $L_1 \overset{\mathrm{d}}{=} L_2$, then $\rho(L_1) = \rho(L_2)$.

We can now state the main result of this subsection.

Theorem 3.2.10. *Assume $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and that ρ is a positive homogeneous, translation-invariant and law-invariant risk measure. Then the following hold.*

(i) For any $L = m + \mathbf{a}^T\mathbf{X} \in \mathcal{M}_{\mathbf{X}}$, it holds that

$$\rho(L) = m + \mathbf{a}^T\boldsymbol{\mu} + \sqrt{\mathbf{a}^T\Sigma\mathbf{a}}\rho(Y)$$

with $Y \sim S_1(\psi)$.

(ii) If $\rho(Y) \geq 0$ with Y as in (i), then ρ is subadditive on $\mathcal{M}_{\mathbf{X}}$.

Proof. For (i), we may write $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$ with $Y \sim S_k(\psi)$ and so

$$L = m + \mathbf{a}^T\boldsymbol{\mu} + \mathbf{a}^T A\mathbf{Y}.$$

Using Lemma 3.2.8, it holds that

$$L \overset{\mathrm{d}}{=} m + \mathbf{a}^T\boldsymbol{\mu} + \|A^T\mathbf{a}\|Y$$

where Y is any coordinate of $\mathbf{Y}$ which has the distribution $Y \sim S_1(\psi)$. Now note that

$$\|A^T\mathbf{a}\| = \sqrt{\langle A^T\mathbf{a}, A^T\mathbf{a}\rangle} = \sqrt{\langle \mathbf{a}, AA^T\mathbf{a}\rangle} = \sqrt{\mathbf{a}^T\Sigma\mathbf{a}}$$

and use translation invariance, law invariance and positive homogeneity of ρ to get

$$\rho(L) = m + \mathbf{a}^T\boldsymbol{\mu} + \sqrt{\mathbf{a}^T\Sigma\mathbf{a}}\rho(Y)$$

as claimed. We now prove (ii). Consider two loss variables $L_1 = m_1 + \mathbf{a}_1^T\mathbf{X}$ and $L_2 = m_2 + \mathbf{a}_2^T\mathbf{X}$ in $\mathcal{M}_{\mathbf{X}}$. As in (i), we have

$$L_1 \overset{\mathrm{d}}{=} m_1 + \mathbf{a}_1^T\boldsymbol{\mu} + \mathbf{a}_1^T A\mathbf{Y}, \quad L_2 \overset{\mathrm{d}}{=} m_2 + \mathbf{a}_2^T\boldsymbol{\mu} + \mathbf{a}_2^T A\mathbf{Y}.$$

Thus

$$\begin{aligned} L_1 + L_2 &= m_1 + \mathbf{a}_1^T\boldsymbol{\mu} + m_2 + \mathbf{a}_2^T\boldsymbol{\mu} + (\mathbf{a}_1 + \mathbf{a}_2)^T A\mathbf{Y} \\ &\overset{\mathrm{d}}{=} m_1 + \mathbf{a}_1^T\boldsymbol{\mu} + m_2 + \mathbf{a}_2^T\boldsymbol{\mu} + \|A^T(\mathbf{a}_1 + \mathbf{a}_2)\|Y \end{aligned}$$

so applying ρ yields

$$\rho(L_1 + L_2) = m_1 + \mathbf{a}_1^T\boldsymbol{\mu} + m_2 + \mathbf{a}_2^T\boldsymbol{\mu} + \|A^T(\mathbf{a}_1 + \mathbf{a}_2)\|\rho(Y),$$

and subadditivity follows from $\|A^T(\mathbf{a}_1 + \mathbf{a}_2)\| \leq \|A^T\mathbf{a}_1\| + \|A^T\mathbf{a}_2\|$. ■

Corollary 3.2.11 (**Subadditivity of VaR for elliptical distributions**). *Consider the Value at Risk measure* VaR_α *on* $\mathcal{M}_{\mathbf{X}}$. *Then* VaR_α *is subadditive whenever* $\alpha \geq 1/2$.

Proof. If $Y \sim S_1(\psi)$, then Y is symmetric around zero, which implies that $\mathrm{VaR}_\alpha(Y) \geq 0$ for $\alpha \geq 1/2$. The claim now follows from (ii) of the above theorem. ■

Notes and comments

Chapter 6 of [88] discusses spherical and elliptical distributions. The chapter contains a few more details and examples. Section 6.3.4 is dedicated to estimation of dispersion and correlation in elliptical distributions. The article [55] discusses normal variance mixture models in more detail and provide additional examples, one of which, namely the Pareto normal variance mixture, is included in the exercises below. The article also discusses methods for parameter estimation.

3.3 Exercises

Theoretical exercises

Exercise 3.1:
Prove Theorem 3.2.4. Hint: Characteristic functions!

Exercise 3.2:
Let $X \sim \Gamma(\alpha, \beta)$ i.e. let X have a gamma distribution with parameters $\alpha, \beta > 0$.

1)Verify that $1/X \sim \text{Ig}(\alpha, \beta)$ by computing the density of $1/X$. This provides an explanation for the name "inverse gamma".

2)Show that

$$E\Big[\frac{1}{X}\Big] = \frac{\beta}{\alpha - 1}$$

when $\alpha > 1$.

3)Show that

$$\text{Var}\left[\frac{1}{X}\right] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

when $\alpha > 2$.

Exercise 3.3:
In this exercise, we fill out some details in the discussion of normal variance mixture models in the main text.

1)Compute the characteristic function of a normal variance mixture model. Conclude that a normal variance mixture model is indeed an elliptical distribution.

2)Verify that for a normal variance mixture model $\mathbf{X}$,

$$\text{Var}[\mathbf{X}] = E[W]\Sigma$$

whenever $E[W] < \infty$.

3)Prove (3.2), that is, show that the density of a normal variance mixture model is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \frac{1}{\sqrt{(2\pi w)^d \det \Sigma}} \exp\left(-\frac{R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)}{2w}\right) f_W(w) dw$$

where $R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$ and f_W is the density of W.

Exercise 3.4:
In this exercise, we elaborate on Example 3.2.7.

1)Verify that when $\nu > 2$,

$$\text{Var}[\mathbf{X}] = \frac{\nu}{\nu - 2}\Sigma.$$

Hint: Use the results of Exercise 3.2.

2)Prove that when Σ is invertible, the density of the multivariate t distribution is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma((d+\nu)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}\sqrt{\det \Sigma}} \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{d+\nu}{2}}.$$

3)Use Stirling's approximation given by

$$\Gamma(x+1) \sim \sqrt{2\pi x}\Big(\frac{x}{e}\Big)^x, \quad x \to \infty$$

to show that the $f_{\mathbf{X}}$ converges to the density of $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ as $\nu \to \infty$.

Exercise 3.5:
Consider the normal variance mixture model $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{Z}$ where W has a Pareto-like distribution with tail

$$\overline{F}_W(w) = \frac{\theta^\alpha}{w^\alpha}, \quad w > \theta$$

where $\theta, \alpha > 0$ are parameters. We shall call this distribution the *Pareto normal variance mixture* with parameters $\theta, \alpha > 0, \boldsymbol{\mu}$ and Σ.

1)Compute the covariance matrix $\mathrm{Var}[\mathbf{X}]$. For what choice of parameters θ and α does this exist?

2)Prove that the density (when Σ is invertible) is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\alpha\theta^\alpha}{\sqrt{(2\pi)^d \det\Sigma}} \left(\frac{R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)}{2}\right)^{-\frac{d}{2}-\alpha} \gamma\left(\alpha + \frac{d}{2}; \frac{R^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)}{2\theta}\right)$$

where

$$\gamma(x, y) = \int_0^y t^{x-1}e^{-t}dt$$

denotes the *lower incomplete gamma function*.

Exercise 3.6:
Let $\mathbf{S}$ be uniformly distributed on the unit sphere in d dimensions $\mathbb{S}^d$.

1)Compute $E[\mathbf{S}]$.

2)Compute $\mathrm{Cov}[\mathbf{S}]$.

3)Let $\mathbf{S}$ be uniformly distributed on the unit circle $\mathbb{S}^1$. Simulate 500 values of $\mathbf{S}$ and plot the result.

Exercise 3.7:
Prove Lemma 3.2.8.

Exercise 3.8:
Recall that the joint density of a normal vector $\mathbf{X} = (X_1, ..., X_d)$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ (where $\det\Sigma \neq 0$) has density

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

Now consider a bivariate normal vector $(X_1, X_2) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

with $\sigma_1^2, \sigma_2^2 > 0$ and $\rho \in (-1, 1)$. Verify that

$$X_2 \mid X_1 = x_1 \sim \mathcal{N}\Big(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\Big).$$

Exercise 3.9: VaR of a sum of jointly normal variables
Let $\mathbf{X} = (X_1, ..., X_d) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Compute $\mathrm{VaR}_\alpha(X_1 + \cdots + X_d)$.

Exercise 3.10:
Let $(X_1, X_2) \sim t_2(\nu, \boldsymbol{\mu}, \Sigma)$ for

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \quad \Sigma = \begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}$$

where $\sigma_1^2, \sigma_2^2 > 0$ and $\rho \in (-1, 1)$. Use the result of Exercise 3.4 to verify that

$$X_2 \mid X_1 = x_1 \sim t\Big(\nu + 1, \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\frac{\nu + ((x_1 - \mu_1)/\sigma_1)^2}{\nu + 1}\Big).$$

Practical exercises

Exercise 3.11:
Without using an R package, write an R function to simulate from the multivariate normal distribution with mean $\boldsymbol{\mu} = (1, 0, 2)$ and covariance matrix

$$\Sigma = \begin{pmatrix}5 & 3 & -2\\ 3 & 4 & -5\\ -2 & -5 & 8\end{pmatrix}.$$

Exercise 3.12:
In this exercise, we simulate from the multivariate normal and t distributions.

1)Without using an R package, simulate 500 values of the multivariate t distribution with

$$\nu = 3, \quad \boldsymbol{\mu} = \begin{pmatrix}2\\ 0\end{pmatrix}, \quad \Sigma = \begin{pmatrix}1 & 1/2\\ 1/2 & 1\end{pmatrix}.$$

Make a plot of the result. Hint: Use the result of Exercise 3.2.

2)Now simulate 500 values of the multivariate normal distribution with the same $\boldsymbol{\mu}$ and Σ and plot the result. Compare the plot to the one for the t distribution.

3)Finally, simulate and plot 500 values of the Pareto normal variance mixture from the previous exercise with the same $\boldsymbol{\mu}$ and Σ as above. Let $\theta = 1$ and experiment with different values of α. Compare with the previous plots.

Exercise 3.13:
For this exercise, you need the BMW-returns dataset available on Absalon. This dataset contains daily log-returns of the BMW stock over the same period as the DAX index considered in earlier exercises. Let $\mathbf{X}_n$ denote the bivariate log-returns of the DAX index and

the BMW stock, respectively. Assume that today's index and today's stock price are both 100.

1)Propose an elliptical (or other) distribution which is appropriate for this data and compute the one-day $\text{VaR}_{0.99}$ and $\text{ES}_{0.99}$.

2)Now use the standard variance-covariance method to compute the one-day $\text{VaR}_{0.99}$. How do your two estimates compare?

Chapter 4

Copulas

4.1 What are copulas?

Motivation and fundamental results

When we have data of the form $\mathbf{X} = (X_1, ..., X_d)$ (for example log returns), a reasonable goal is to find a suitable joint distribution function F for $\mathbf{X}$. In the previous chapter, we used elliptical distributions for this purpose, and we remarked a few drawbacks of this approach. Elliptical distributions force us to also have elliptical marginals, and in some cases, we don't want to exclude the possibility that the X_i have different types of distributions. The topic of this chapter is copulas, a widely used tool for modelling the dependence structure of a random vector while still allowing flexibility in the choice of marginals.

Recall that if X_i has distribution function F_i, then $F_i^{\leftarrow}(U_i) \overset{\mathrm{d}}{=} X_i$ with U_i a Unif$(0,1)$ variable. This is the "inverse transform method" used in simulation. Recall also that $F_i(X_i) \overset{\mathrm{d}}{=} U_i$ whenever F_i is continuous. A refresher on generalised inverses may be found in the appendix.

To study the problem of determining the joint distribution F, we assume that the marginal distribution functions F_i are known. The transformation $U_i := F_i(X_i)$ in the continuous case is illustrated in Figure 4.1 below.

Definition 4.1.1. A *copula* C is a distribution function on $[0,1]^d$ such that all marginals are Unif$(0,1)$ distributed.

To be able to go back and forth between the "original space" and the "copula space", Sklar's Theorem [110] is an essential tool.

Theorem 4.1.2 (Sklar's Theorem). *Let F be the joint distribution function of the random vector $\mathbf{X} = (X_1, ..., X_d)$ with marginal distribution functions $F_1, ..., F_d$. There exists a copula C such that*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{4.1}$$

If $F_1, ..., F_d$ are continuous, C is unique. Conversely, given a copula C and marginal distribution functions $F_1, ..., F_d$, then F as defined in (4.1) is a joint distribution function with marginals $F_1, ..., F_d$.

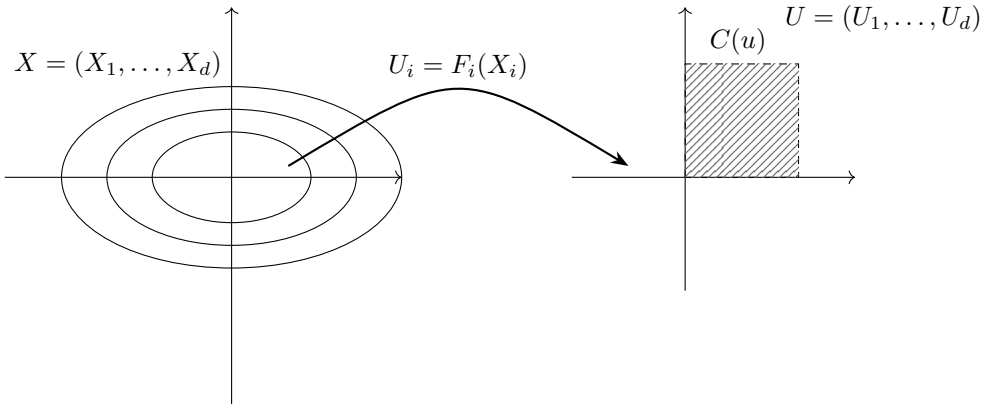


Figure 1: An illustration (in two dimensions) of the idea of a "copula space". The "original space" on the left is where the variables live, and we wish to transform them into a collection of uniform variables with support on $[0,1]^d$.

Proof. For the sake of simplicity, assume that the F_i are continuous. The general statement is proved in the supplementary section. Then $U_i := F_i(X_i) \sim \text{Unif}(0,1)$. Suppose $\mathbf{X} \sim F$ with marginals $X_i \sim F_i$. Let $\mathbf{U} = (F_1(X_1), ..., F_d(X_d))$ and let C be the distribution function of $\mathbf{U}$. By construction and the continuity assumption on the F_i, C is a copula. We compute

$$\begin{aligned} C(F_1(x_1), ..., F_d(x_d)) &= P(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ &= P(F_1(X_1) \leq F_1(x_1), ..., F_d(X_d) \leq F_d(x_d)) \\ &= P(X_1 \leq x_1, ..., X_d \leq x_d) = F(x_1, ..., x_d) \end{aligned}$$

which shows that the copula C has the desired properties. As for uniqueness, by continuity of the F_i, we have $F_i(F_i^{\leftarrow}(u_i)) = u_i$ for all $u_i \in [0,1]$. Letting $x_i = F_i^{\leftarrow}(u_i)$ in the expression above, we get

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d))$$

and any copula $\tilde{C}$ satisfying $\tilde{C}(F_1(x_1), ..., F_d(x_d)) = F(x_1, ..., x_d)$ must satisfy the same relation. Uniqueness now follows. To prove the converse statement, let C be a copula and $F_1, ..., F_d$ univariate distribution functions. Let $\mathbf{U} = (U_1, ..., U_d)$ have distribution function C and define $X_i := F_i^{\leftarrow}(U_i)$, $\mathbf{X} := (X_1, ..., X_d)$. We know that $X_i \sim F_i$, so the marginal distributions are correct. Also,

$$\begin{aligned} C(F_1(x_1), ..., F_d(x_d)) &= P(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ &= P(F_1^{\leftarrow}(U_1) \leq x_1, ..., F_d^{\leftarrow}(U_d) \leq x_d) \\ &= P(X_1 \leq x_1, ..., X_d \leq x_d) \end{aligned}$$

which shows that $F(x_1, ..., x_d) := C(F_1(x_1), ..., F_d(x_d))$ is the distribution function for $\mathbf{X}$. ■

Sklar's Theorem offers an explanation for why copulas are called copulas. The name stems from the fact that a copula "couples" the marginal distributions to the joint distribution.

We will see many examples of copulas shortly.

It is often the case that a copula is unique. So-called implicit copulas are constructed by using the formula

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d))$$

provided in the proof of Sklar's Theorem, valid for continuous $F_1, ..., F_d$. The following example illustrates how uniqueness may fail.

Example 4.1.3. Let (X_1, X_2) be bivariate Bernoulli distributed with

$$P(X_1 = 0, X_2 = 0) = \frac{1}{12}, \quad P(X_1 = 1, X_2 = 0) = \frac{1}{3},$$
$$P(X_1 = 0, X_2 = 1) = \frac{1}{3}, \quad P(X_1 = 1, X_2 = 1) = \frac{1}{4}.$$

Clearly, $P(X_1 = 1) = P(X_2 = 1) = 7/12$ so the marginals are identical, $X_1, X_2 \sim F$ where

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{5}{12}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

From Sklar's Theorem, we know that there exists a copula C such that

$$P(X_1 \leq x_1, X_2 \leq x_2) = C(F(x_1), F(x_2))$$

for all $x_1, x_2 \in \mathbb{R}$. Because C is a copula, the cases where $x_i < 0$ or $x_i \geq 1$ for either $i = 1$ or $i = 2$ are already handled. The only additional restriction comes from the case $x_1, x_2 \in [0, 1)$ which yields the requirement $C(5/12, 5/12) = 1/12$. We conclude that any copula C satisfying $C(5/12, 5/12) = 1/12$ is a copula of (X_1, X_2), and it is not difficult to see that we may construct infinitely many copulas satisfying this relation.

∘

Sklar's Theorem provides a recipe for constructing copulas using a known joint distribution function. We call such copulas *implicit copulas*. Different examples of copulas will be given in the next section. We first consider some more theoretical properties. We start with the following useful characterisation of copulas.

Proposition 4.1.4. *A function $C : [0, 1]^d \to [0, 1]$ is a copula if and only if*

(i) $C(u_1, ..., u_d) = 0$ if $u_i = 0$ for any $i = 1, ..., d$.

(ii) $C(1, ..., 1, u_i, 1, ..., 1) = u_i$ for any $i = 1, ..., d$ and $u_i \in [0, 1]$.

(iii) For all $(a_1, ..., a_d), (b_1, ..., b_d) \in [0, 1]^d$ with $a_i \leq b_i$, we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1 + \cdots + i_d} C(u_{1i_1}, ..., u_{di_d}) \geq 0$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j = 1, ..., d$.

The first two properties are self-explanatory. The third property is called the *rectangle inequality* and can be interpreted as follows: For uniform variables $(U_1, ..., U_d)$, then $P(a_1 \leq U_1 \leq b_1, ..., a_d \leq U_d \leq b_d) \geq 0$.

Proposition 4.1.5 (**Fréchet bounds**). *For every copula C, we have $W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u})$ where*

$$W(\mathbf{u}) = \max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \quad \textit{and} \quad M(\mathbf{u}) = \min_{i=1,...,d} u_i.$$

Proof. Let $\mathbf{U}$ have distribution function C. For every $u_i \in [0,1]$, we have

$$C(\mathbf{u}) = P(U_1 \leq u_1, ..., U_d \leq u_d) \leq P(U_i \leq u_i) = u_i.$$

The bound $C(\mathbf{u}) \leq M(\mathbf{u})$ now follows by minimising over all i. Conversely, for $\mathbf{u} \in [0,1]^d$,

$$\begin{aligned} 1 - C(\mathbf{u}) &= 1 - P(U_1 \leq u_1, ..., U_d \leq u_d) = P\left(\bigcup_{i=1}^{d} \{U_i > u_i\}\right) \\ &\leq \sum_{i=1}^{d} P(U_i > u_i) = \sum_{i=1}^{d} (1 - u_i) = d - \sum_{i=1}^{d} u_i. \end{aligned}$$

Thus $-C(\mathbf{u}) \leq d - 1 - \sum_{i=1}^{d} u_i$ implying $C(\mathbf{u}) \geq \sum_{i=1}^{d} u_i + 1 - d$. Since $C(\mathbf{u}) \geq 0$ always holds, the lower bound follows. ■

What kinds of random vectors produce these upper and lower bounds? It turns out that for $d > 2$, W as given in the proposition is not a copula. This follows from Proposition 4.1.4, since the rectangle inequality does not hold in this case. Indeed, considering the set $[1/2, 1]^d \subseteq [0,1]^d$ and letting $\mathbf{U} \sim W$, we would have

$$\begin{aligned} P(\mathbf{U} \in [1/2,1]^d) &= \max\{1 + \cdots + 1 - d + 1, 0\} - d \max\{1/2 + 1 + \cdots + 1 - d + 1, 0\} \\ &\quad + \binom{d}{2} \max\{1/2 + 1/2 + 1 + \cdots + 1 - d + 1, 0\} - \cdots \\ &\quad + \max\{1/2 + \cdots + 1/2 - d + 1, 0\} = 1 - \frac{d}{2}. \end{aligned}$$

And this is negative whenever $d > 2$. We can however produce M as a copula for any d and W for $d = 2$. To do so, we introduce the concepts of comonotonicity and countermonotonicity.

Comonotonicity and countermonotonicity

Definition 4.1.6. We say that $X_1, ..., X_d$ are *comonotone* if $(X_1, ..., X_d) = (\alpha_1(Z), ..., \alpha_d(Z))$ for some univariate variable Z and non-decreasing functions $\alpha_1, ..., \alpha_d$. We say that (X_1, X_2) are *countermonotonic* if $(X_1, X_2) = (\alpha(Z), \beta(Z))$ for some univariate variable Z, some non-decreasing function α and some non-increasing function β.

The following proposition shows that the lower and upper Fréchet bounds arise from countermonotonic and comonotonic random vectors, respectively.

Proposition 4.1.7. *A comonotone bundle $(X_1, ..., X_d)$ has M as a copula, and a countermonotonic bundle (X_1, X_2) in $d = 2$ dimensions has W as a copula.*

Proof. Consider a comonotone bundle $(X_1, ..., X_d)$ and assume for simplicity that the functions α_i are strictly increasing and continuous. If F is the joint distribution function, we have

$$\begin{aligned}
F(x_1, ..., x_d) &= P(X_1 \leq x_1, ..., X_d \leq x_d) = P(\alpha_1(Z) \leq x_1, ..., \alpha_d(Z) \leq x_d) \\
&= P(Z \leq \alpha_1^{\leftarrow}(x_1), ..., Z \leq \alpha_d^{\leftarrow}(x_d)) = P\left(Z \leq \min_{i=1,...,d} \alpha_i^{\leftarrow}(x_i)\right) \\
&= \min_{i=1,...,d} P(Z_i \leq \alpha_i^{\leftarrow}(x_i)) = \min_{i=1,...,d} P(\alpha_i(Z) \leq x_i) \\
&= \min_{i=1,...,d} P(X_i \leq x_i) = \min_{i=1,...,d} F_i(x_i) = M(F_1(x_1), ..., F_d(x_d))
\end{aligned}$$

so M is a copula for $(X_1, ..., X_d)$. Now consider a countermonotonic bundle $(X_1, X_2) = (\alpha(Z), \beta(Z))$. Assume for simplicity that α is strictly increasing and continuous, β is strictly decreasing and continuous and that Z is continuous. If F is the distribution function of (X_1, X_2), we have

$$\begin{aligned}
F(x_1, x_2) &= P(\alpha(Z) \leq x_1, \beta(Z) \leq x_2) = P(Z \leq \alpha^{\leftarrow}(x_1), Z \geq \beta^{\leftarrow}(x_2)) \\
&= P(Z \leq \alpha^{\leftarrow}(x_1)) - P(Z \leq \alpha^{\leftarrow}(x_1), Z < \beta^{\leftarrow}(x_2)) \\
&= F_1(x_1) - \min\{F_1(x_1), 1 - F_2(x_2)\} \\
&= \max\{F_1(x_1) - F_1(x_1), F_1(x_1) - 1 + F_2(x)\} \\
&= \max\{0, F_1(x_1) + F_2(x_2) + 1 - 2\} = W(F_1(x_1), F_2(x_2))
\end{aligned}$$

so W is a copula for (X_1, X_2). ■

Remark 4.1.8. The implications in the above proposition are biimplications, see the notes and comments at the end of the chapter as well as the exercises afterwards.

In a risk management context, we can interpret comonotonicity as meaning that there is a single source of risk that drives all movements in our portfolio. As the underlying risk varies, all assets move in the same direction. In other words, a comonotone portfolio represents an extreme scenario where risk cannot be diversified. Countermonotonicity has the opposite interpretation. The two assets still depend entirely on the same source of risk, but instead of moving together, whenever one asset decreases in value, the other increases.

Transformation of copulas

Monotone transformations

What happens when we take monotone transformations of a copula? While the distribution itself may change, the copula does not change under strictly increasing transformations as the following result shows.

Proposition 4.1.9. *Consider a random vector $(X_1, ..., X_d)$ with continuous marginal distributions F_i and copula C. Let $T_1, ..., T_d$ be strictly increasing continuous functions. Then $(T_1(X_1), ..., T_d(X_d))$ also has copula C.*

Proof. Let $\tilde{X}_i := T_i(X_i), \tilde{X}_i \sim \tilde{F}_i$ and $(\tilde{X}_1, ..., \tilde{X}_d) \sim \tilde{F}$. Let $\tilde{C}$ be the copula of $(\tilde{X}_1, ..., \tilde{X}_d)$. By Sklar's Theorem,

$$\begin{aligned}\tilde{C}(\tilde{F}_1(x_1), ..., \tilde{F}_d(x_d)) &= \tilde{F}(x_1, ..., x_d) = P(T_1(X_1) \leq x_1, ..., T_d(X_d) \leq x_d) \\ &= P(X_1 \leq T_1^{\leftarrow}(x_1), ..., X_d \leq T_d^{\leftarrow}(x_d)) \\ &= F(T_1^{\leftarrow}(x_1), ..., T_d^{\leftarrow}(X_d)) = C(F_1(T_1^{\leftarrow}(x_1)), ..., F_d(T_d^{\leftarrow}(x_d))).\end{aligned}$$

We now claim that $\tilde{F}_i = F_i \circ T_i^{\leftarrow}$. By definition,

$$\tilde{F}_i(x) = P(\tilde{X}_i \leq x) = P(T_i(X_i) \leq x) = P(X_i \leq T_i^{\leftarrow}(x)) = F_i(T_i^{\leftarrow}(x))$$

as claimed. F_i is continuous by assumption and $T_i^{\leftarrow}$ is continuous since T_i is strictly increasing. Hence $\tilde{F}_i$ is continuous. We have now proved that

$$\tilde{C}(\tilde{F}_1(x_1), ..., \tilde{F}_d(x_d)) = C(\tilde{F}_1(x_1), ..., \tilde{F}_d(x_d))$$

and so $\tilde{C} = C$ by the uniqueness part of Sklar's Theorem. ■

It is natural to ask how the copula changes under decreasing transformations or mixtures of increasing and decreasing transformations on the entries of the vector. This is explored in Exercise 4.16.

Survival copulas

Just like many distributions are nicer to work with when considering the tail, so is it more natural to work with a *survival copula* for some multivariate distributions. From Sklar's Theorem, we know that the relationship between the joint distribution function F of $\mathbf{X} = (X_1, ..., X_d)$ and the copula C of $\mathbf{X}$ is

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

with $F_1, ..., F_d$ the marginal distribution functions of $X_1, ..., X_d$. It is natural to ask whether a similar connection exists between the multivariate tail

$$\overline{F}(x_1, ..., x_d) := P(X_1 > x_1, ..., X_d > x_d)$$

and the marginal tails $\overline{F}_1, ..., \overline{F}_d$. To answer this question, let us specialise to the case $d = 2$. We see that (make a drawing!)

$$\begin{aligned}\overline{F}(x, y) &= 1 - F_1(x) - F_2(y) + F(x, y) = \overline{F}_1(x) + \overline{F}_2(y) - 1 + C(F_1(x), F_2(y)) \\ &= \overline{F}_1(x) + \overline{F}_2(y) - 1 + C(1 - \overline{F}_1(x), 1 - \overline{F}_2(y)).\end{aligned}$$

This leads us to the following definition.

Definition 4.1.10. If (X_1, X_2) is a random vector with copula C, we let

$$\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

denote the *survival copula* of (X_1, X_2).

From Exercise 4.14, it follows that $\widehat{C}$ is also a copula.

Remark 4.1.11. A word of warning. The survival copula $\widehat{C}$ is not the same as the survival function for (U_1, U_2) with distribution function C. The precise relationship is given by

$$\overline{C}(u, v) = 1 - u - v + C(u, v) = \widehat{C}(1 - u, 1 - v).$$

This discrepancy also explains why we don't use the overline notation to denote the survival copula.

One can establish a Sklar's Theorem for survival copulas. We state a simplified version in a moment, but we first need to define what it means to compute the generalised inverse of a non-increasing function. If h is non-increasing, we define

$$h^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : h(x) \leq u\}$$

which is just the definition for a non-decreasing function but with the inequality reversed. Note that if F is a distribution function and $\overline{F} = 1 - F$ the survival function, we have the useful relation

$$\overline{F}^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : \overline{F}(x) \leq u\} = \inf\{x \in \mathbb{R} : F(x) \geq 1 - u\} = F^{\leftarrow}(1 - u).$$

Theorem 4.1.12 (**Sklar's Theorem for survival copulas**). *Let (X, Y) be a bivariate random variable with continuous marginal distribution functions F_X and F_Y and joint survival function $\overline{F}$. Then the unique survival copula $\widehat{C}$ satisfies*

$$\overline{F}(x, y) = \widehat{C}(\overline{F}_X(x), \overline{F}_Y(y)).$$

This copula is given by

$$\widehat{C}(u_1, u_2) = \overline{F}(\overline{F}_X^{\leftarrow}(u_1), \overline{F}_Y^{\leftarrow}(u_2)).$$

Proof. We have for $u_1, u_2 \in [0, 1]$ that

$$\begin{aligned}\overline{F}(\overline{F}_X^{\leftarrow}(u_1), \overline{F}_Y^{\leftarrow}(u_2)) &= P(X > \overline{F}_X^{\leftarrow}(u_1), Y > \overline{F}_Y^{\leftarrow}(u_2)) \\ &= \overline{F}_X(\overline{F}_X^{\leftarrow}(u_1)) + \overline{F}_Y(\overline{F}_Y^{\leftarrow}(u_2)) - 1 \\ &\quad + C(1 - \overline{F}_X(\overline{F}_X^{\leftarrow}(u_1)), 1 - \overline{F}_Y(\overline{F}_Y^{\leftarrow}(u_2))).\end{aligned}$$

By continuity of $\overline{F}_X$ and $\overline{F}_Y$, it follows that $\overline{F}_X(\overline{F}_X^{\leftarrow}(u_1)) = u_1$ and $\overline{F}_Y(\overline{F}_Y^{\leftarrow}(u_2)) = u_2$. Hence

$$\overline{F}(\overline{F}_X^{\leftarrow}(u_1), \overline{F}_Y^{\leftarrow}(u_2)) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) = \widehat{C}(u_1, u_2).$$

■

As a final fact about survival copulas, we mention that if (U_1, U_2) has uniform marginals and copula C, then the survival copula $\widehat{C}$ is realised as the copula of $(1 - U_1, 1 - U_2)$. The proof is simple and is left to the reader.

Symmetry properties

Recall that a random variable X is *symmetric* about $a \in \mathbb{R}$ if $X - a$ and $a - X$ have the same distribution. If the distribution function F of X is continuous, this is equivalent to

$$F(a + x) = \overline{F}(a - x),$$

for all $x \in \mathbb{R}$ as is easily verified (do this!). A classic example is the normal distribution, which is symmetric about its mean. How should we define symmetry for more variables? We here focus on the concept of *radial symmetry*.

Definition 4.1.13. Let $a, b \in \mathbb{R}$. The pair (X, Y) of random variables is said to be *radially symmetric* about (a, b) if the pairs $(X-a, Y-b)$ and $(a-X, b-Y)$ have the same distribution.

Analogous to the univariate case, it holds that whenever X and Y are continuous, then (X, Y) are radially symmetric about (a, b) if and only if the joint distribution function F satisfies

$$F(a+x, b+y) = \overline{F}(a-x, b-y) \quad \text{for all } x, y \in \mathbb{R}.$$

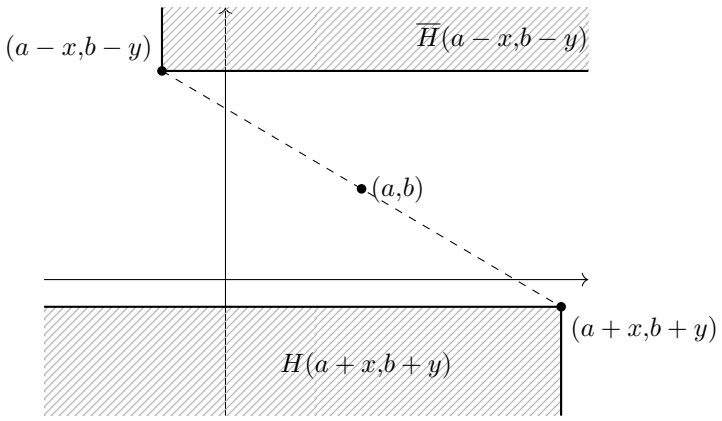


Figure 2: An illustration of radial symmetry for a pair of random variables. The two shaded regions have the same probability mass.

Recall that $\overline{F}$ is the joint survival function of (X, Y) given by $\overline{F}(x, y) = P(X > x, Y > y)$. Again the bivariate normal distribution is an example, where the point (a, b) is the mean vector. Since copulas contain the joint information for the pair (X, Y), it is natural to expect a criterion for radial symmetry in terms of the copula of (X, Y). Such a criterion is given in the following theorem.

Theorem 4.1.14. *Let X and Y be continuous and symmetric around a and b, respectively. Then (X, Y) is radially symmetric around (a, b) if and only if the copula C of (X, Y) satisfies*

$$C(u_1, u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2) \quad \textit{for all } u_1, u_2 \in [0, 1].$$

That is, the copula and survival copula of (X, Y) coincide, $C = \widehat{C}$.

Proof. The proof follows almost immediately by applying Sklar's Theorem and Sklar's Theorem for survival copulas, Theorem 4.1.12. Let X have distribution function F_1 and Y distribution function F_2. The statement for the joint distribution function F

$$F(a+x, b+y) = \overline{F}(a-x, b-y) \quad \text{for all } x, y \in \mathbb{R}.$$

is equivalent to

$$C(F_1(a+x), F_2(b+y)) = \widehat{C}(\overline{F}_1(a-x), \overline{F}_2(b-y)) \quad \text{for all } x, y \in \mathbb{R}.$$

Using the marginal symmetries, $F_1(a+x) = \overline{F}_1(a-x)$ and $F_2(b+y) = \overline{F}_2(b-y)$, it immediately follows that the above is equivalent to $C = \widehat{C}$. ∎

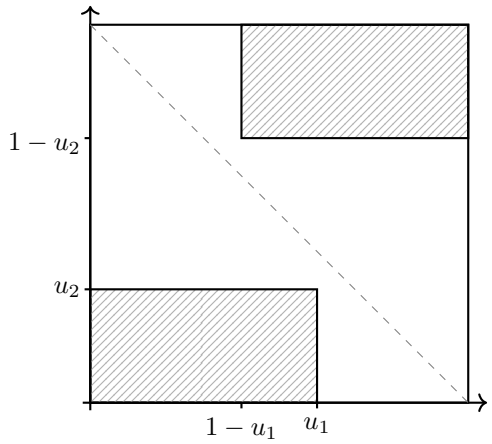


Figure 3: For a radially symmetric pair of random variables, the two areas shaded in grey of the copula space will have same probability.

Different examples of radially symmetric pairs of random variables are scattered throughout the exercises. In Exercise 4.31, a stronger symmetry concept is also explored. Another useful type of symmetry we will not cover here is *exchangeability*. This concept will be explored in Exercise 5.7 in the next chapter. We now present many examples of copulas.

4.2 Examples of copulas

In this section, we consider some examples of copulas with an emphasis on the fundamental and implicit copulas. The two following sections concern a whole class of explicit copulas.

Fundamental copulas

Fundamental copulas arise from theoretical considerations. We have already seen two examples from the Fréchet bounds.

Definition 4.2.1. For any $d > 1$, we call

$$M(\mathbf{u}) = \min_{i=1,\dots,d} u_i$$

the *comonotonicity copula*. For $d = 2$, we call

$$W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$$

the *countermonotonicity copula*.

The Fréchet bound copulas are not the only "theoretical" copulas.

Example 4.2.2 (The independence copula). For any $d > 1$, the *independence copula* is given by

$$\Pi(u_1, \dots, u_d) = \prod_{i=1}^{d} u_i.$$

Unsurprisingly, Π arises from independent variables. Suppose $X_1, ..., X_d$ are independent with continuous distribution functions F_i. Then

$$\begin{aligned}\Pi(F_1(x_1), ..., F_d(x_d)) &= \prod_{i=1}^{d} F_i(x_i) = \prod_{i=1}^{d} P(X_i \leq x_i) \\ &= P(X_1 \leq x_1, ..., X_d \leq x_d) = F(x_1, ..., x_d)\end{aligned}$$

so by the uniqueness from Sklar's Theorem, Π is the copula for $X_1, ..., X_d$.

$\circ$

Implicit copulas

Implicit copulas arise from known joint distribution functions. Let F be a given joint distribution function. If the marginal distribution functions F_i are continuous, we can use Sklar's Theorem to construct a copula C via

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d)).$$

We give a concrete example.

Example 4.2.3 (**The Gaussian copula**)**.** Let $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$ be a d-dimensional normal vector with distribution function Φ_Σ where

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{12} & 1 & \cdots & \\ \vdots & & & \vdots \\ \rho_{1d} & \cdots & \cdots & 1 \end{pmatrix}.$$

One can think of Σ as a correlation matrix. Note that the marginals X_i are standard normal, so that they have common distribution function Φ. We can then construct the *Gaussian copula*

$$C_\Sigma^{\mathrm{Ga}}(u_1, ..., u_d) = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)).$$

What if we have a general mean and covariance matrix? Let $\mathbf{Y} = \boldsymbol{\mu} + B\mathbf{X}$ be an affine transformation of $\mathbf{X}$ where

$$B = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{pmatrix}$$

is a diagonal matrix with $\sigma_i > 0$ for all i. Then $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, B\Sigma B^T)$, and any desired covariance matrix can be written in the form $B\Sigma B^T$. Note that $Y_i = \mu_i + \sigma_i X_i$ is a strictly increasing and continuous transformation of X_i, so Proposition 4.1.9 implies that $\mathbf{Y}$ and $\mathbf{X}$ have the same copula, namely C_Σ^{Ga}. Simulating from this copula is easy. The trick is to follow the construction in Sklar's Theorem. To simulate a sample from C_Σ^{Ga}, follow these steps:

(i) Simulate $\mathbf{X}$ from the multivariate Gaussian distribution $\mathcal{N}(0, \Sigma)$.

(ii) Set $U_i = \Phi(X_i)$ for $i = 1, ..., d$.

(iii) Now $\mathbf{U} = (U_1, ..., U_d)$ has the distribution function C_Σ^{Ga}.

$\circ$

Example 4.2.4 (**Example 2.8 in [91]**)**.** Consider the distribution function

$$F(x,y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1}, & (x,y) \in [-1,1] \times [0,\infty) \\ 1-e^{-y}, & (x,y) \in (1,\infty) \times [0,\infty) \\ 0, & \text{otherwise} \end{cases}$$

for the pair (X,Y). Letting x and y go to infinity, respectively, we obtain that the marginal distribution functions are given by

$$F_X(x) = \begin{cases} \frac{x+1}{2}, & x \in [-1,1] \\ 1, & x > 1 \\ 0, & \text{otherwise} \end{cases}, \quad F_Y(y) = \begin{cases} 0, & y < 0 \\ 1-e^{-y}, & y \geq 0 \end{cases}$$

We see that X is uniform on $[-1,1]$ while Y is standard exponential. It is easy to solve for the inverses F_X^{-1} and F_Y^{-1}, and we obtain

$$F_X^{-1}(u) = 2u - 1, \quad F_Y^{-1}(v) = -\log(1-v),$$

which yields that the unique copula for the pair (X,Y) is given by

$$\begin{aligned} C(u_1,u_2) = F(F_X^{-1}(u_1), F_Y^{-1}(u_2)) &= \frac{(2u_1 - 1 + 1)(e^{-\log(1-u_2)} - 1)}{2u_1 - 1 + 2e^{-\log(1-u_2)} - 1} \\ &= \frac{2u_1(1/(1-u_2) - 1)}{2(u_1 + 1/(1-u_2) - 1)} = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}. \end{aligned}$$

$\circ$

Example 4.2.5 (**Bivariate Pareto distribution, Example 2.14 in [91]**)**.** Let (X,Y) have the joint survival function

$$\overline{F}(x,y) = \begin{cases} (1+x+y)^{-\alpha}, & x,y \geq 0 \\ (1+x)^{-\alpha}, & x \geq 0, y < 0 \\ (1+y)^{-\alpha}, & x < 0, y \geq 0 \\ 1, & x,y < 0 \end{cases}$$

where $\alpha > 0$ is a parameter. This distribution can be interpreted as a bivariate Pareto distribution, and it is clear that the marginals are themselves Pareto,

$$\overline{F}_X(x) = \begin{cases} (1+x)^{-\alpha}, & x \geq 0 \\ 1, & x < 0 \end{cases}, \quad \overline{F}_Y(y) = \begin{cases} (1+y)^{-\alpha}, & y \geq 0 \\ 1, & y < 0 \end{cases}$$

The generalised inverses of the marginals are equal to the ordinary inverses and are given by

$$\overline{F}_X^{-1}(u) = \overline{F}_Y^{-1}(u) = u^{-1/\alpha} - 1.$$

Using the survival version of Sklar's Theorem, Theorem 4.1.12, we obtain that the survival copula is given by

$$\widehat{C}(u_1, u_2) = (1 + u_1^{-1/\alpha} - 1 + u_2^{-1/\alpha} - 1)^{-\alpha} = (u_1^{-1/\alpha} + u_2^{-1/\alpha} - 1)^{-\alpha}.$$

Using the relation

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2),$$

one can obtain a (less nice) expression for the copula C. ∘

Explicit copulas

Explicit copulas are given by a concrete formula. Some well-known examples are the following.

Example 4.2.6 (**Gumbel copula**)**.** The Gumbel copula is given by

$$C_\theta^{\mathrm{Gu}}(u_1, u_2) = \exp\left(-\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{1/\theta}\right)$$

where $1 \leq \theta < \infty$ is a parameter. ∘

Example 4.2.7 (**Clayton copula**)**.** The Clayton copula is given by

$$C_\theta^{\mathrm{Cl}}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

where $0 < \theta < \infty$ is a parameter. ∘

Simulation methods

Simulation from a copula can be done in the same way as for any multidimensional distribution. Let us simplify to the case $d = 2$. Recall that if we want to simulate $(X_1, X_2) \sim F$, where F is known, we can follow these steps:

(i) Simulate X_1.

(ii) Determine the conditional distribution $X_2 \mid X_1 = x$. Then simulate X_2 from this distribution, plugging in the simulated X_1 value for x.

(iii) The vector (X_1, X_2) then has distribution F.

Let C be a two-dimensional copula and let $(U_1, U_2) \sim C$. In the exercises, you are asked to verify that the conditional distribution function of U_2 given $U_1 = u_1$ is given by

$$C_{U_2|U_1}(u_2 \mid u_1) = \frac{\partial}{\partial u_1} C(u_1, u_2).$$

This partial derivative is usually not too hard to compute for explicit examples. An algorithm to simulate from C then goes as follows:

Algorithm 4.2.1 (**Simulation from general two-dimensional copula**)**.** For a two-dimensional copula C:

1. Simulate independent $U_1, V \sim \mathrm{Unif}(0, 1)$.

2. Set
$$U_2 \leftarrow C^{\leftarrow}_{U_2|U_1}(V \mid U_1)$$
with $C^{\leftarrow}_{U_2|U_1}(\cdot \mid u_1)$ denoting the generalised inverse of $C_{U_2|U_1}(\cdot \mid u_1)$.

3. The pair (U_1, U_2) is now a sample from C.

Sampling from a copula yields another method to sample from any multidimensional distribution by using Sklar's Theorem. If $(X_1, ..., X_d)$ has distribution function F, marginals $F_1, ..., F_d$ and copula C, we can simulate from F as follows:

1. Simulate $(U_1, ..., U_d)$ from C.

2. Set $X_i \leftarrow F_i^{\leftarrow}(U_i)$ for $i = 1, ..., d$.

3. The vector $(X_1, ..., X_d)$ then has distribution function F.

Example 4.2.8. In Example 4.2.4, we introduced the copula
$$C(u_1, u_2) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}.$$
We see that
$$\begin{aligned} C_{U_2|U_1}(u_2 \mid u_1) &= \frac{\partial}{\partial u_1}\Big(\frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}\Big) \\ &= \frac{u_2}{u_1 + u_2 - u_1 u_2} - \frac{u_1 u_2}{(u_1 + u_2 - u_1 u_2)^2}(1 - u_2) \\ &= \frac{u_2(u_1 + u_2 - u_1 u_2) - u_1 u_2(1 - u_2)}{(u_1 + u_2 - u_1 u_2)^2} \\ &= \Big(\frac{u_2}{u_1 + u_2 - u_1 u_2}\Big)^2, \end{aligned}$$
and solving for the inverse yields
$$C^{-1}_{U_2|U_1}(v \mid u_1) = \frac{u_1\sqrt{v}}{1 - (1 - u_1)\sqrt{v}}.$$
Hence we may simulate from this copula by first simulating independent $U_1, V \sim \text{Unif}(0, 1)$ and letting
$$U_2 = \frac{U_1\sqrt{V}}{1 - (1 - U_1)\sqrt{V}}.$$
If we want to simulate from the distribution function F in Example 4.2.4, we can then apply the inverse transforms to U_1 and U_2 to obtain a sample (X, Y) via
$$X = 2U_1 - 1, \quad Y = -\log(1 - U_2).$$

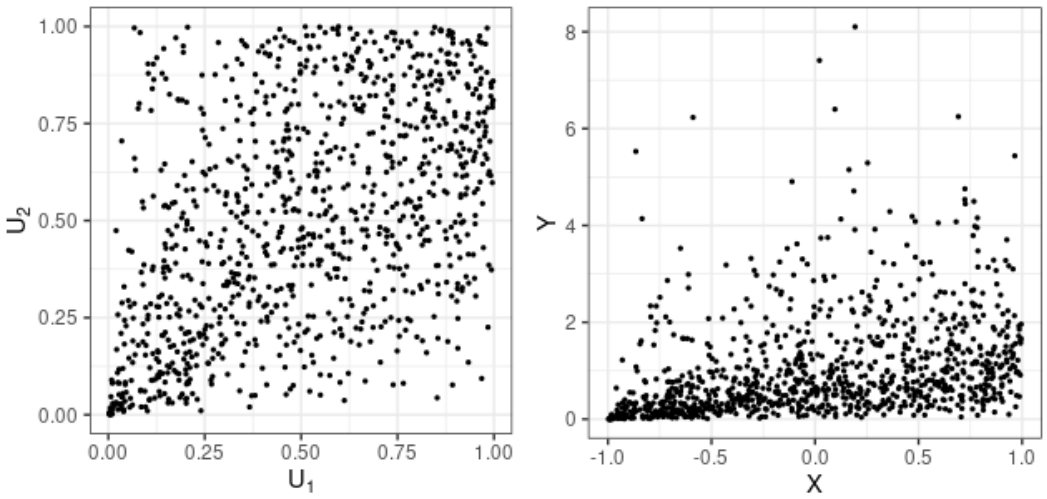


Figure 4: Left: A simulated sample ($n = 1000$) of the copula $C(u_1, u_2) = u_1u_2/(u_1 + u_2 - u_1u_2)$ using the algorithm just described. Right: 1000 simulated samples of the distribution presented in Example 4.2.4 based on the samples in the left plot.

∘

4.3 Archimedean copulas

Basic properties and characterisation

Both of the above examples of explicit copulas are so-called *Archimedean copulas*.

Definition 4.3.1. Let $\varphi : [0, 1] \to [0, \infty]$ be continuous, strictly decreasing and convex with $\varphi(0) = \infty$ and $\varphi(1) = 0$. Then

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

is called the *Archimedean copula* with generator φ.

We see that by choosing $\varphi(t) = (-\log t)^\theta$ for $\theta \geq 1$, we obtain the Gumbel copula. Choosing $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$ gives the Clayton copula. The notion "Archimedean" comes from the fact that an Archimedean copula satisfies a type of Archimedean property analogous to the one for the real numbers, namely that for any $a, b > 0$, there exists an integer n such that $na > b$. See Exercise 4.30 for details.

Example 4.3.2 (Generalised Clayton copula). Choosing $\varphi(t) = \theta^{-\delta}(t^{-\theta} - 1)^\delta$ for $\theta > 0$ and $\delta \geq 1$ gives the *Generalised Clayton copula*. The special case $\delta = 1$ corresponds to the Clayton copula from above. ∘

Example 4.3.3 (Frank copula). The Frank copula is the Archimedean copula with generator

$$\varphi(t) = -\log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$$

where $\theta \in \mathbb{R} \setminus \{0\}$ is a parameter. ∘

It should of course be verified that an Archimedean copula is a copula. The following result characterises Archimedean copulas in two dimensions.

Theorem 4.3.4. *Let $\varphi : [0,1] \to [0,\infty]$ be a continuous, strictly decreasing function such that $\varphi(1) = 0$, $\varphi(0) = \infty$. The function $C : [0,1]^2 \to [0,1]$ given by*

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

is a copula if and only if φ is convex.

The proof of the theorem (which may be skipped at first reading) relies on Proposition 4.1.4. To prove the theorem, the following somewhat technical lemma is required.

Lemma 4.3.5. *Let $\varphi : [0,1]^2 \to [0,\infty]$ be a continuous, strictly decreasing function with $\varphi(1) = 0$ and $\varphi(0) = \infty$. Then the function $C : [0,1]^2 \to [0,1]$ defined by*

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

satisfies (i) and (ii) of Proposition 4.1.4. The rectangle inequality (iii) holds if and only if whenever $u_1 \leq u_2$,

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \tag{4.2}$$

for all $v \in [0,1]$.

Proof. Letting C be the function defined in the lemma, we then have $C(u_1, 0) = \varphi^{-1}(\varphi(u_1) + \varphi(0)) = \varphi^{-1}(\infty) = 0$ and similarly, $C(0, u_2) = 0$. Also, $C(u_1, 1) = \varphi^{-1}(\varphi(u_1) + \varphi(1)) = u_1$ and $C(1, u_2) = u_2$. This shows that (i) and (ii) of Proposition 4.1.4 are satisfied. As for the second assertion of the lemma, we note that (4.2) is a special case of the rectangle inequality (why?), so if the rectangle inequality holds, then certainly (4.2) also holds. Now assume that (4.2) is true for all $v \in [0,1]$. Choose $v_1 \leq v_2$ in $[0,1]$ and note that $C(0, v_2) = 0 \leq v_1 \leq v_2 = C(1, v_2)$. Both φ and φ^{-1} are continuous, hence so is C. Thus there is a $t \in [0,1]$ such that $C(t, v_2) = v_1$, which means that $\varphi(v_2) + \varphi(t) = \varphi(v_1)$. Hence

$$\begin{aligned} C(u_2, v_1) - C(u_1, v_1) &= \varphi^{-1}(\varphi(u_2) + \varphi(v_1)) - \varphi^{-1}(\varphi(u_1) + \varphi(v_1)) \\ &= \varphi^{-1}(\varphi(u_2) + \varphi(v_2) + \varphi(t)) + \varphi^{-1}(\varphi(u_1) + \varphi(v_2) + \varphi(t)) \\ &= C(C(u_2, v_2), t)) - C(C(u_1, v_2), t) \\ &\leq C(u_2, v_2) - C(u_1, v_2). \end{aligned}$$

Rearrange this inequality to obtain the rectangle inequality. ■

Proof of Theorem 4.3.4. Let C be of the form in the previous lemma. We have already shown that (i) and (ii) of Proposition 4.1.4 hold. It remains to show that the rectangle inequality holds if and only if φ is also convex. Assume that the rectangle inequality holds, so that C is a copula. As was established in the previous lemma, the rectangle inequality for Archimedean copulas is equivalent to (4.2). Setting $a = \varphi(u_1), b = \varphi(u_2)$ and $c = \varphi(v)$ for $u_1, u_2, v \in [0,1]$ with $u_1 \leq u_2$, (4.2) can be rewritten to

$$\varphi^{-1}(a) + \varphi^{-1}(b + c) \leq \varphi^{-1}(b) + \varphi^{-1}(a + c) \tag{4.3}$$

for any $a \geq b$ and $c \geq 0$. Since φ is convex if and only if φ^{-1} is (why?), it suffices to show convexity of φ^{-1}. To this end, let $0 \leq s < t$. Setting $a = (s+t)/2$, $b = s$ and $c = (t-s)/2$ and plugging into the inequality from before, we obtain

$$\varphi^{-1}\Big(\frac{s+t}{2}\Big) \leq \frac{\varphi^{-1}(s) + \varphi^{-1}(t)}{2}.$$

Thus φ^{-1} is midconvex[1], so by continuity, φ^{-1} is convex as desired (see Exercise 4.29). To prove the converse, assume φ^{-1} is convex. Fix $a \geq b$ and $c \geq 0$ and let $\gamma = (a-b)/(a-b+c)$. We may then write $a = (1-\gamma)b + \gamma(a+c)$ and $b+c = \gamma b + (1-\gamma)(a+c)$, so from convexity, it follows that

$$\varphi^{-1}(a) \leq (1-\gamma)\varphi^{-1}(b) + \gamma\varphi^{-1}(a+c)$$

and

$$\varphi^{-1}(b+c) \leq \gamma\varphi^{-1}(b) + (1-\gamma)\varphi^{-1}(a+c).$$

Adding these two inequalities yield (4.3). The proof is thus complete.

■

While the method of constructing Archimedean copulas is certainly practical, there are some limitations worth mentioning. An Archimedean copula is seen to be symmetric in the two arguments which is clearly a constraint in modelling. Furthermore, the structure of an Archimedean copula is maybe too simple since it is two-dimensional but has a one-dimensional generator.

Simulating from Archimedean copulas in two dimensions

We here present an algorithm ([41]) for simulating from Archimedean copulas in two dimensions. While simple, it has the drawback that one must compute several quantities related to the generator on a case-by-case basis.

Algorithm 4.3.1. For a two-dimensional Archimedean copula C with generator φ:

1. Simulate independent $U_1, V \sim \text{Unif}(0,1)$.
2. Set
$$W \leftarrow (\varphi')^{-1}\Big(\frac{\varphi'(U_1)}{V}\Big).$$
3. Set $U_2 \leftarrow \varphi^{-1}(\varphi(W) - \varphi(U_1))$.
4. The pair (U_1, U_2) is then a sample from C.

We should of course verify that this algorithm works.

Proposition 4.3.6. *The pair (U_1, U_2) produced by the algorithm above is indeed distributed according to C.*

[1] A function f is *midconvex* if for every s, t in the domain of f, we have $f(s/2 + t/2) \leq (f(s) + f(t))/2$. See also Exercise 4.29.

Proof. Let $u_1, u_2 \in [0,1]$. We are done once we have shown that

$$P(U_1 \leq u_1, U_2 \leq u_2) = C(u_1, u_2).$$

The strategy is to consider $P(U_2 \leq u_2 \mid U_1 = u_1)$. Recall that φ and φ^{-1} are strictly decreasing while φ' is increasing by convexity of φ. Using these facts as well as independence of U_1 and V yields

$$\begin{aligned}
P(U_2 \leq u_2 \mid U_1 = u_1) &= P(\varphi(W) - \varphi(U_1) \geq \varphi(u_2) \mid U_1 = u_1) \\
&= P(\varphi(W) \geq \varphi(u_1) + \varphi(u_2) \mid U_1 = u_1) \\
&= P\Big(\varphi\Big((\varphi')^{-1}\Big(\frac{\varphi'(U_1)}{V}\Big)\Big) \geq \varphi(u_1) + \varphi(u_2) \mid U_1 = u_1\Big) \\
&= P\Big((\varphi')^{-1}\Big(\frac{\varphi'(u_1)}{V}\Big) \leq \varphi^{-1}(\varphi(u_1) + \varphi(u_2))\Big) \\
&= P\Big(\frac{\varphi'(u_1)}{V} \leq \varphi'(\varphi^{-1}(\varphi(u_1) + \varphi(u_2)))\Big) \\
&= P\Big(V \leq \frac{\varphi'(u_1)}{\varphi'(\varphi^{-1}(\varphi(u_1) + \varphi(u_2)))}\Big) \\
&= \frac{\varphi'(u_1)}{\varphi'(\varphi^{-1}(\varphi(u_1) + \varphi(u_2)))}.
\end{aligned}$$

Now, by a conditioning argument

$$P(U_1 \leq u_1, U_2 \leq u_2) = \int_0^{u_1} P(U_2 \leq u_2 \mid U_1 = u) du.$$

We want this to equal $C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$. To see that this is indeed the case, simply differentiate along u_1 to obtain

$$\frac{\partial}{\partial u_1} C(u_1, u_2) = \frac{\varphi'(u_1)}{\varphi'(\varphi^{-1}(\varphi(u_1) + \varphi(u_2)))}.$$

The proof is now complete by the fundamental theorem of calculus. ■

Example 4.3.7. Recall the Clayton family with generator

$$\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1), \quad \theta \in (0, \infty).$$

The inverse of the generator is given by

$$\varphi^{-1}(s) = (1 + \theta s)^{-1/\theta}.$$

The derivative is given by $\varphi'(t) = -t^{-\theta-1}$ with inverse $(\varphi')^{-1}(s) = (-s)^{-1/(1+\theta)}$. Hence, after a few tedious calculations, we obtain the following algorithm to simulate from this copula:

1. Simulate independent $U_1, V \sim \text{Unif}(0,1)$.
2. Set $W \leftarrow U_1 V^{1/(1+\theta)}$.

3. Set
$$U_2 \leftarrow \left(1 + U_1^{-\theta}\left(V^{-\theta/(1+\theta)} - 1\right)\right)^{-1/\theta}.$$

4. The pair (U_1, U_2) is now a sample from the Clayton copula with parameter θ.

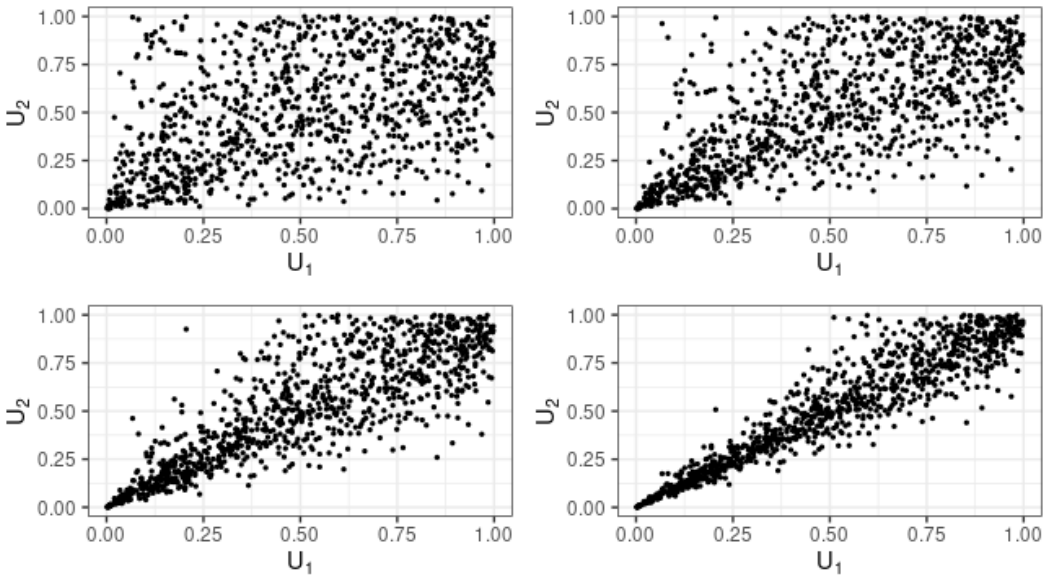


Figure 5: Four simulated samples of size 1000 for the Clayton copula with $\theta = 1$ (upper right), $\theta = 2$ (upper left), $\theta = 4$ (lower left) and $\theta = 8$ (lower right). Note how the samples seem to become more and more "narrow" (dependent) for higher θ values. This phenomenon will be explained in the next chapter on dependence concepts.

In the following section, we will see yet another method to simulate from the Clayton copula, one that works for an arbitrary dimension. In Exercise 4.43, another algorithm is outlined. ∘

Example 4.3.8. Consider the *Joe copula* ([71] and [70]) which has generator
$$\varphi(t) = -\log(1 - (1-t)^\theta)$$
for a parameter $\theta \geq 1$. In this case, the derivative is given by
$$\varphi'(t) = -\frac{\theta(1-t)^{\theta-1}}{1-(1-t)^\theta}$$
and there is no explicit form for the inverse $(\varphi')^{-1}(s)$. But this number satisfies the equation
$$s - s(1 - (\varphi')^{-1}(s))^\theta + \theta(1 - (\varphi')^{-1}(s))^{\theta-1} = 0,$$
and one can use numeric methods to solve for $(\varphi')^{-1}(s)$ in step 2 of the algorithm. ∘

4.4 Archimedean copulas in higher dimensions

Characterisation and complete monotonicity

How can we generalise the Archimedean copula construction to higher dimensions? The logical next step would be to propose

$$C(u_1, ..., u_d) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)) \tag{4.4}$$

where $\varphi : [0,1] \to [0,\infty]$ has the same properties as in the two-dimensional case i.e. φ is strictly decreasing, convex and satisfies $\varphi(0) = \infty$ and $\varphi(1) = 0$. Is C as constructed above a copula? The answer is no in general. One issue is that C is not even a distribution function in general for dimensions higher than two. In order to answer the question of when C is a copula, we need the following definition.

Definition 4.4.1. A decreasing function f is *completely monotonic* on $[a,b]$ if

$$(-1)^k \frac{d^k}{dt^k} f(t) \geq 0 \quad \text{for} \quad k = 0, 1, ... \quad \text{and} \quad t \in (a,b).$$

It turns out that the property of being completely monotonic determines whether C defined by equation (4.4) is a copula.

Theorem 4.4.2. *Let $\varphi : [0,1] \to [0,\infty]$ be a continuous strictly decreasing function such that $\varphi(0) = \infty$ and $\varphi(1) = 0$. C defined by (4.4) is a copula for all $d \geq 2$ if and only if φ^{-1} is completely monotonic on $[0,\infty)$.*

Proof. We refer to [87] for a proof. ■

Checking complete monotonicity for the inverse of a copula generator is often tedious, but there are tools available. The following proposition summarises some of these. A function f is called *absolutely monotonic* on $[a,b]$ if

$$\frac{d^k}{dt^k} f(t) \geq 0 \quad \text{for} \quad k = 0, 1, ... \quad \text{and} \quad t \in (a,b).$$

Proposition 4.4.3. *The following hold.*

(i) If f is absolutely monotonic and g is completely monotonic, then $f \circ g$ is completely monotonic.

(ii) If f and g are completely monotonic, so is fg.

(iii) If f is completely monotonic and g is positive with completely monotone derivative, then $f \circ g$ is completely monotonic.

Proof. Property (i) is covered in Chapter IV of [120]. Properties (ii) and (iii) are covered in [33]. Note that an easy proof of (ii) follows by *Leibniz' formula* for the the k'th derivative of the product,

$$(fg)^{(k)}(t) = \sum_{i=0}^{k} \binom{k}{i} f^{(i)}(t) g^{(k-i)}(t).$$

■

Example 4.4.4. Consider the Gumbel copula with generator $\varphi(t) = (-\log t)^\theta$ for $\theta \geq 1$. The inverse is given by

$$\varphi^{-1}(s) = \exp(-s^{1/\theta}).$$

The function $x \mapsto e^{-x}$ is clearly completely monotone, and $s \mapsto s^{1/\theta}$ is positive with completely monotone derivative. It follows by (iii) of Proposition 4.4.3 that φ^{-1} is completely monotonic. We conclude that the Gumbel copula can be generalised to arbitrary dimensions. Explicitly,

$$C_\theta(u_1, ..., u_d) = \exp\Big(- \Big((-\log u_1)^\theta + \cdots + (-\log u_d)^\theta \Big)^{1/\theta} \Big)$$

is a copula on $[0,1]^d$ for any $d \geq 2$. ∘

Simulation methods

While it is possible to check whether φ^{-1} is completely monotonic, it is a tedious procedure. As an alternative to verifying complete monotonicity, one can apply Laplace transforms of distribution functions. We recall the definition.

Definition 4.4.5. Let G be a distribution function on $[0, \infty)$ with $G(0) = 0$. The Laplace transform of G is

$$\psi(t) = \int_0^\infty e^{-tx} dG(x).$$

Remark 4.4.6. Let G be a distribution function on $[0, \infty)$ with $G(0) = 0$ and Y a random variable distributed according to G. We then have the following relationship between the Laplace transform ψ of G and the moment-generating function κ_Y given by

$$\psi(t) = \kappa_Y(-t).$$

Lemma 4.4.7. *A function ψ on $[0, \infty)$ is the Laplace transform of a distribution function G if and only if ψ is completely monotonic and $\psi(0) = 1$.*

Proof. See Chapter XIII.4 in [33]. ■

The above results provide a strategy to verify that φ^{-1} is completely monotonic. It suffices to show that φ^{-1} is a Laplace transform of some distribution function G.

Example 4.4.8. Consider the Clayton copula with generator $\varphi(t) = (t^{-\theta} - 1)/\theta$. We can solve for φ^{-1} and get

$$\varphi^{-1}(t) = (\theta t + 1)^{-1/\theta} = \frac{(1/\theta)^{1/\theta}}{(t + 1/\theta)^{1/\theta}}.$$

Now recall that the gamma distribution with parameters $\alpha, \beta > 0$ has the moment-generating function

$$\kappa(t) = \left(\frac{\beta}{\beta - t} \right)^\alpha$$

so if we let $\alpha = \beta = 1/\theta$, we have that the Laplace transform of this distribution is given by $\varphi^{-1}(t)$. Explicitly,

$$C_\theta^{\mathrm{Cl}}(u_1, ..., u_d) = (u_1^{-\theta} + u_2^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}$$

is a copula on $[0,1]^d$ for any $d \geq 2$.

∘

Archimedean copulas constructed using Laplace transforms of distributions deserve their own name.

Definition 4.4.9. An *LT-Archimedean copula* is an Archimedean copula where the generator φ satisfies $\varphi^{-1} = \psi$, with ψ the Laplace transform of some distribution function G on $[0, \infty)$ with $G(0) = 0$.

How do we simulate random vectors with a given copula? If we have an LT-Archimedean copula, the following proposition provides a recipe.

Proposition 4.4.10. *Let G be a distribution function on $[0, \infty)$ and $V \sim G$. Let $\psi = \varphi^{-1}$ denote the Laplace transform of V. Suppose we have variables $W_1, ..., W_d$ which are conditionally independent given V with conditional distribution function*

$$F_{W_i|V=v}(u) = e^{-v\varphi(u)}.$$

Then the distribution function of $\mathbf{W}$ satisfies $F_{\mathbf{W}}(\mathbf{u}) = C(\mathbf{u})$.

Proof. The proof is a straightforward computation using a conditioning argument and conditional independence:

$$\begin{aligned} F_{\mathbf{W}}(\mathbf{u}) &= P(W_1 \leq u_1, ..., W_d \leq u_d) = \int_0^\infty P(W_1 \leq u_1, ..., W_d \leq u_d \mid V = v) dG(v) \\ &= \int_0^\infty \prod_{i=1}^d P(W_i \leq u_i \mid V = v) dG(v) = \int_0^\infty \prod_{i=1}^d e^{-v\varphi(u_i)} dG(v) \\ &= \int_0^\infty e^{-v(\varphi(u_1)+\cdots+\varphi(u_d))} dG(v) = \psi(\varphi(u_1) + \cdots + \varphi(u_d)) \\ &= \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)) = C(\mathbf{u}). \end{aligned}$$

■

The proposition tells us that if we want to simulate from the copula $C(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d))$, we should apply the following steps:

1. Identify the distribution G having the Laplace transform $\psi = \varphi^{-1}$.
2. Simulate $V \sim G$.
3. Generate iid $U_1, ..., U_d \sim \text{Unif}(0, 1)$ and apply the inverse transform method i.e. $W_i = F^{\leftarrow}_{W_i|V=v}(U_i)$ with v equal to the simulated value of V from step 2.
4. $W_1, ..., W_d$ is then a sample from C.

We can actually be more specific in step 3. $F_{W_i|V=v}$ has a proper inverse which we solve for as follows:

$$e^{-v\varphi(F^{\leftarrow}_{W_i|V=v}(u))} = u \quad \Leftrightarrow \quad -\log u = v\varphi(F^{\leftarrow}_{W_i|V=v}(u)) \quad \Leftrightarrow \quad F^{\leftarrow}_{W_i|V=v}(u) = \varphi^{-1}\left(-\frac{\log u}{v}\right).$$

Hence W_i in step 3 should be set to

$$W_i = \psi\left(-\frac{\log U_i}{V}\right).$$

To exemplify, let us consider the Clayton copula again. We know that the distribution G in the algorithm is the $\Gamma(1/\theta, 1/\theta)$ distribution. We computed

$$\varphi^{-1}(t) = (\theta t + 1)^{-1/\theta},$$

and so an algorithm for simulating from the d-dimensional Clayton copula is as follows.

Algorithm 4.4.1. To generate a sample from the d-dimensional Clayton copula with parameter $\theta > 0$, do the following:

1. Simulate $V \sim \Gamma(1/\theta, 1/\theta)$.
2. Simulate independent $U_1, ..., U_d \sim \text{Unif}(0, 1)$.
3. Set
$$W_i \leftarrow \left(1 - \theta \frac{\log U_i}{V}\right)^{-1/\theta}.$$
4. $W_1, ..., W_d$ is then a sample from the Clayton copula.

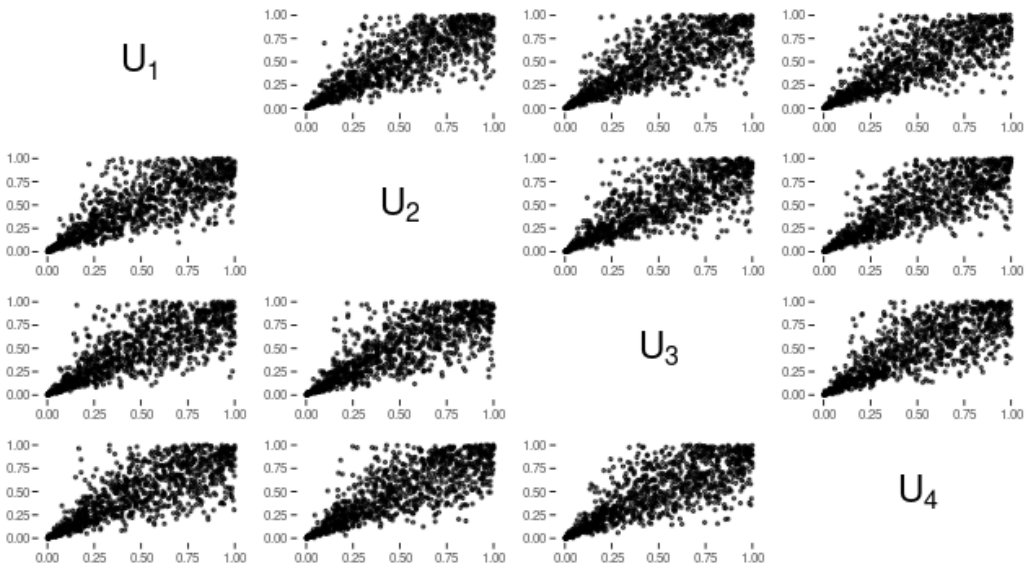


Figure 6: A simulated sample of size 1000 from the four-dimensional Clayton copula with $\theta = 3$.

4.5 Extreme value copulas*

In this supplementary section, we introduce a certain type of copula which arises naturally in the modelling of correlation of extreme events. Extreme events often move together. In a financial context, a large fluctuation in one stock often spills over to other stocks of the same

type, and in nature, extreme events are often caused by a number of factors which reinforce each other. Hence it is vital to be able to capture the dependence structure of exceptional events. We first introduce the concept of a *max-stable* copula, and then we present the definition of an extreme value copula. The remainder of the section is devoted to examples and some essential properties of such copulas. To see the connections to classical extreme value theory, the reader is encouraged to first take a look at the supplementary section in Chapter 2 on maximal domains of attraction if they have not already done so.

Definitions and essential properties

Definition 4.5.1. A copula C i *max-stable* if for all $a \geq 0$ and $u_1, ..., u_d \in [0,1]$, it holds that

$$C(u_1, ..., u_d) = C^a(u_1^{1/a}, ..., u_d^{1/a}).$$

What is the motivation for this definition? The following result (which you are asked to prove in Exercise 4.36) shows how a max-stable copula occurs.

Proposition 4.5.2. *Let $\{X_{1,i}, ..., X_{d,i}\}_{i=1}^n$ be iid random vectors. Assume $(X_{1,1}, ..., X_{d,1})$ has joint distribution function F, copula C and continuous marginals $F_1, ..., F_d$. Let $X_{j,1:n} := \max\{X_{j,1}, ..., X_{j,n}\}$ be the maximum of the observations in the j'th coordinate, $j = 1, ..., d$. Then the copula $C_{1:n}$ of $X_{1,1:n}, ..., X_{d,1:n}$ is given by*

$$C_{1:n}(u_1, ..., u_d) = C^n(u_1^{1/n}, ..., u_d^{1/n}).$$

In other words, a copula is max-stable if and only if the random vector consisting of the componentwise maxima have the same copula as the original joint distribution.

Example 4.5.3. The independence copula $\Pi(u_1, ..., u_d) = u_1 \cdots u_d$ is trivially a max-stable copula. The interpretation of this is that if the sequences $\{X_{1,i}\}_{i=1}^n, ..., \{X_{d,i}\}_{i=1}^n$ are independent, then so are their running maxima. This fact was already clear. ◦

Example 4.5.4. If the random vectors $\{X_{1,i}, ..., X_{d,i}\}_{i=1}^n$ are comonotonic, then so are the componentwise maxima. Indeed, the comonotonicity copula $M(u_1, ..., u_d) = \min\{u_1, ..., u_d\}$ is clearly max-stable. ◦

Example 4.5.5. Consider the Gumbel copula

$$C_\theta^{\mathrm{Gu}}(u_1, u_2) = \exp\Big(- \Big((-\log u_1)^\theta + (-\log u_2)^\theta \Big)^{1/\theta} \Big)$$

for $\theta \geq 1$. Letting $a > 0$, we have

$$\begin{aligned}
(C_\theta^{\mathrm{Gu}})^a(u_1^{1/a}, u_2^{1/a}) &= \exp\Big(- a\Big((-\log u_1^{1/a})^\theta + (-\log u_2^{1/a})^\theta \Big)^{1/\theta} \Big) \\
&= \exp\Big(- a\Big(\frac{1}{a^\theta} \Big((-\log u_1)^\theta + (-\log u_2)^\theta \Big) \Big)^{1/\theta} \Big) \\
&= C_\theta^{\mathrm{Gu}}(u_1, u_2).
\end{aligned}$$

Hence the Gumbel family is max-stable. ◦

We can now introduce the main definition of this section.

Definition 4.5.6. A copula C_* is called an *extreme value copula* if there exists a copula C such that

$$C_*(u_1, ..., u_d) = \lim_{n\to\infty} C^n(u_1^{1/n}, ..., u_d^{1/n})$$

for all $u_1, ..., u_d \in [0, 1]$. In this case, we say that C belongs to the *domain of attraction* of C_*.

The following result is analogous to the result from classical extreme value theory which states that a distribution is an extreme value distribution if and only if it is max-stable (see Theorem 2.5.2).

Theorem 4.5.7. *A copula is an extreme value copula if and only if it is max-stable.*

Proof. If C is max-stable, it is obviously an extreme value copula (and it lies in the domain of attraction of itself). Conversely, if C_* is an extreme value copula which lies in the domain of attraction of C, we have for any $a > 0$ that

$$C_*^a(u_1^{1/a}, ..., u_d^{1/a}) = \lim_{n\to\infty} C^{an}(u_1^{1/an}, ...u_d^{1/an}) = C_*(u_1, ..., u_d),$$

showing that C_* is indeed max-stable. ■

The following theorem provides a constructive procedure to generate extreme value copulas. At the same time, we get another characterisation of extreme value copulas. For simplicity, we only present the result in two dimensions. For a higher-dimensional statement, see [49] and the references therein.

Theorem 4.5.8. *The two-dimensional copula C is an extreme value copula if and only if there exists a function $A : [0, 1] \to [1/2, 1]$ (called the* dependence function*) such that*

$$C(u_1, u_2) = \exp\Big(\log(u_1u_2)A\Big(\frac{\log u_2}{\log(u_1u_2)}\Big)\Big).$$

Given C, A is given by

$$A(t) = -\log C(e^{-(1-t)}, e^{-t}),$$

and a function $A : [0, 1] \to [1/2, 1]$ is a valid dependence function (that is, it generates an extreme value copula) if and only if it is convex and satisfies $A(0) = A(1) = 1$, $\max\{t, 1 - t\} \le A(t) \le 1$.

Proof. Let C be a two-dimensional extreme value copula. Let X and Y be standard exponential with survival copula C. Then the joint survival function of (X, Y) is given by

$$\overline{F}(x, y) = P(X > x, Y > y) = C(e^{-x}, e^{-y}).$$

Using that C is max-stable, we get for any $a > 0$ that

$$\overline{F}(ax, ay) = C^a(e^{-x}, e^{-y}) = \overline{F}^a(x, y).$$

Now define $A : [0, 1] \to [1/2, 1]$ as in the theorem, i.e. $A(t) = -\log C(e^{-(1-t)}, e^{-t})$. Then $C(e^{-(1-t)}, e^{-t}) = \exp(-A(t))$. Now make the change of variables for $a > 0, t \in (0, 1)$

$$(x, y) = (a(1 - t), at) \quad \Leftrightarrow \quad (a, t) = \Big(x + y, \frac{y}{x + y}\Big).$$

Then the joint survival distribution may be written as

$$\begin{aligned}\overline{F}(x,y) &= \overline{F}(a(1-t),at) = \overline{F}^a(1-t,t)\\ &= C^a(e^{-(1-t)},e^{-t}) = \exp(-aA(t))\\ &= \exp\Big(-(x+y)A\Big(\frac{y}{x+y}\Big)\Big).\end{aligned}$$

Now use that $C(u_1,u_2) = \overline{F}(-\log u_1, -\log u_2)$, we have

$$C(u_1,u_2) = \exp\Big(\log(u_1u_2)A\Big(\frac{\log u_2}{\log(u_1u_2)}\Big)\Big)$$

as desired. For the converse, we refer to [70]. ■

Examples of extreme value copulas

We have already seen some examples of extreme value copulas above. For the independence copula, it is clear that $A(t) = 1$ for all t, while for the comonotonicity copula, $A(t) = \max\{1-t,t\}$. We shall study the Gumbel family in more depth, but before doing so, we introduce a new family of copulas.

The Marshall–Olkin family of copulas

The Marshall–Olkin family (see [85] and [86]) is a multivariate distribution with uniform marginals that are all exponential. We here consider the bivariate version, where one can think of the distribution as modelling a two-component system susceptible to three types of independent shocks, namely one which knocks out one component, the other component or both at the same time. The time until a shock happens is modelled by an exponential distribution. To make this precise, denote by T_i the time until the shock of type i happens, $i = 1,2,3$. Then we assume that T_1, T_2 and T_3 are independent and exponentially distributed with parameters λ_1, λ_2 and λ_3, respectively. The joint survival function $\overline{F}$ for the lifetime of the two components $X_1 := \min\{T_1,T_3\}, X_2 := \min\{T_2,T_3\}$ is thus given by

$$\begin{aligned}\overline{F}(x,y) = P(X_1 > x, X_2 > y) &= P(T_1 > x, T_2 > x, T_3 > \max\{x,y\})\\ &= \exp(-\lambda_1 x - \lambda_2 y - \lambda_3 \max\{x,y\}).\end{aligned}$$

The lifetime of the components are both exponential with rates $\lambda_1+\lambda_3$ and $\lambda_2+\lambda_3$, respectively, since the survival functions $\overline{F}_1$ and $\overline{F}_2$ of X_1 and X_2 are easily seen to be given by

$$\overline{F}_1(x) = e^{-(\lambda_1+\lambda_3)x} \quad \text{and} \quad \overline{F}_2(x) = e^{-(\lambda_2+\lambda_3)x}.$$

We wish to determine the survival copula of (X_1,X_2). To do so, rewrite $\max\{x,y\} = x+y-\min\{x,y\}$. Then

$$\begin{aligned}\overline{F}(x,y) &= \exp(-(\lambda_1+\lambda_3)x - (\lambda_2+\lambda_3)y + \lambda_3\min\{x,y\})\\ &= \overline{F}_1(x)\overline{F}_2(y)\min\{\exp(\lambda_3 x), \exp(\lambda_3 y)\}.\end{aligned}$$

Recall that in order to find the survival copula, we need to plug in $\overline{F}_1^{-1}(u_1)$ and $\overline{F}_2^{-1}(u_2)$ for x and y. Before doing so, it makes sense to reparametrise the problem. Let

$$\alpha = \frac{\lambda_1}{\lambda_1+\lambda_3}, \quad \beta = \frac{\lambda_2}{\lambda_2+\lambda_3}.$$

Then with a bit of algebra, we obtain that the survival copula $\widehat{C}$ of (X_1, X_2) is

$$\widehat{C}(u_1, u_2) = u_1 u_2 \min\{u_1^{-\alpha}, u_2^{-\beta}\} = \min\{u_1^{1-\alpha} u_2, u_1 u_2^{1-\beta}\}.$$

While $\alpha, \beta \in (0, 1)$ by construction, one can show that this function is still a copula for $\alpha, \beta \in \{0, 1\}$, see Exercise 4.38.

Definition 4.5.9. The *Marshall–Olkin family of copulas* is the two-parameter family of copulas given by

$$C_{\alpha,\beta}(u_1, u_2) = \min\{u_1^{1-\alpha} u_2, u_1 u_2^{1-\beta}\} = \begin{cases} u_1^{1-\alpha} u_2, & u_1^\alpha \geq u_2^\beta \\ u_1 u_2^{1-\beta}, & u_1^\alpha \leq u_2^\beta, \end{cases}$$

where $\alpha, \beta \in [0, 1]$.

Note that this family is a generalisation of the Cuadras–Augé family of copulas considered in Exercise 4.4. Note that the function $A : [0, 1] \to [1/2, 1]$ given by

$$A(t) = 1 - \min\{\beta t, \alpha(1 - t)\}$$

is a valid dependence function and that it gives rise to the Marshall–Olkin copula. That the Marshall–Olkin family is an extreme value copula family can also be verified directly. Below is an illustration of how samples from this copula and the bivariate exponential distribution above look like.

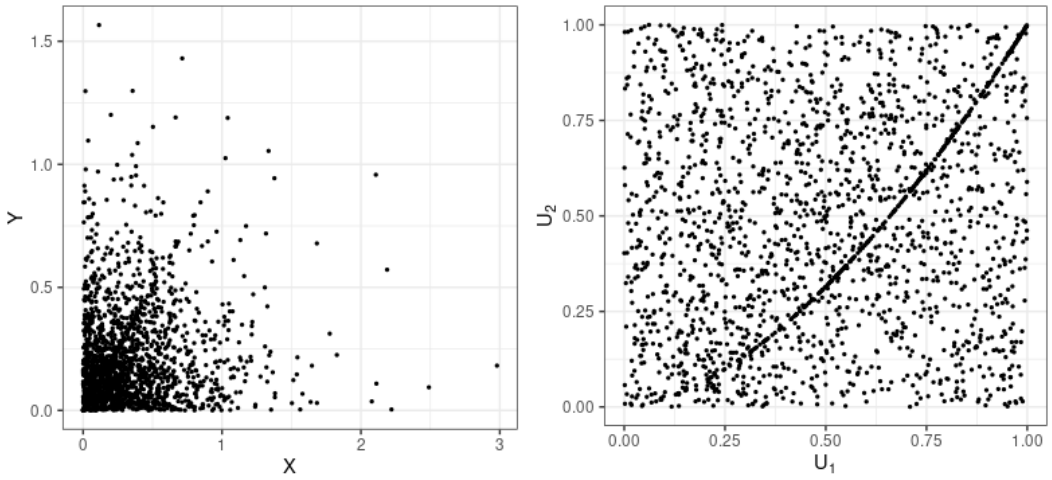


Figure 7: A simulated sample of size 1000 from the Marshall–Olkin copula (right) and the corresponding sample of (X_1, X_2) (left). Here the parameters are chosen to be $\lambda_1 = 2, \lambda_2 = 4$ and $\lambda_3 = 1$, that is, $\alpha = 2/3$ and $\beta = 4/5$.

Note that the copula sample has a curve through the plot. This is due to the fact that the Marshall–Olkin copula has a singular component, see Exercise 4.38 for details. For more information on singular and continuous components of copulas in general, consult [91].

Which extreme value copulas are Archimedean?

We saw above that the Gumbel copula is an extreme value copula. The dependence function is given by

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

It is natural to ask if there are other Archimedean copulas which are extreme value copulas. The answer turns out to be no ([43]).

Theorem 4.5.10. *The Gumbel copula is the only Archimedean extreme value copula.*

Proof. Assume φ generates an Archimedean extreme value copula C. Fix $s > 0$. For any $t \in [0,1]$, we claim that $\varphi(t^s) = c_s\varphi(t)$ for a constant c_s only depending on s. To verify this claim, it helps to introduce the function $h_s : [0,\infty] \to [0,\infty]$ given by $h_s(x) = \varphi(\varphi^{-1}(x)^s)$. We claim that h_s is linear. Let $x, y \in [0,\infty]$. Since φ is surjective, we may find $u_1, u_2 \in [0,1]$ such that $x = \varphi(u_1)$ and $y = \varphi(u_2)$. Using that C is max-stable, we obtain

$$\begin{aligned} h_s(x) + h_s(y) &= \varphi(u_1^s) + \varphi(u_2^s) = \varphi(C(u_1^s, u_2^s)) \\ &= \varphi(C(u_1,u_2)^s) = \varphi(\varphi^{-1}(\varphi(u_1) + \varphi(u_2))^s) \\ &= \varphi(\varphi^{-1}(x+y)^s) = h_s(x+y), \end{aligned}$$

which shows that h_s is additive. Since h is also continuous, being composed of continuous functions, we have that h_s is linear, i.e. $h_s(x) = c_s x$. Translating back to φ, if we let $t = \varphi^{-1}(x)$ (again possible by surjectivity), we get $\varphi(t^s) = c_s\varphi(t)$ as desired. As you will show in an exercise, $c\varphi$ is also a generator for C for any constant $c > 0$, so we may normalise φ in such a way that $\varphi(1/e) = 1$. Letting $s = -\log t$, then $\varphi(e^{-xs}) = c_s\varphi(e^{-x})$. Hence, if we let $g(x) = \varphi(e^{-x})$, then $g(sx) = c_s g(x)$. Because $g(1) = 1$, $c_s = g(s)$, which shows that g satisfies the functional equation $g(xy) = g(x)g(y)$ for $x, y > 0$. By continuity of g, it holds that g is of the form $g(x) = x^\theta$ (see Exercise 4.41). Translating back to the generator, this means that $\varphi(t) = g(-\log t) = (-\log t)^\theta$ which is the generator for the Gumbel copula. This completes the proof. ■

4.6 Constructing new copulas*

There exists a variety of methods to construct copulas. These vary in nature from purely geometric methods, where one attempts to construct copulas based on certain geometric restrictions (diagonal structure, quadratic or cubic sections etc.) to algebraic methods, where certain probabilistic properties are satisfied. We refer to Sections 3.2 and 3.3 in [91], respectively, for a comprehensive introduction. In this supplementary section, we present a few of these results. Other examples can be found in Exercises 4.33, 4.34 and 4.39. For background on the examples and methods presented shortly, see the notes and comments at the end of the chapter.

Constructing new copulas from existing ones

In this short subsection, we present a general method for constructing a new copula given an existing one.

Theorem 4.6.1. *Let $\gamma : [0,1] \to [0,1]$ be a continuous and strictly increasing function with $\gamma(0) = 0$ and $\gamma(1) = 1$. Let C be an arbitrary two-dimensional copula. Then the function $C_\gamma : [0,1]^2 \to [0,1]$ given by*

$$C_\gamma(u_1, u_2) = \gamma^{-1}(C(\gamma(u_1), \gamma(u_2)))$$

is a copula if and only if γ is concave (equivalently: if γ^{-1} is convex).

Proof. Assume first that γ^{-1} is convex. We will show that C_γ is a copula by verifying (i), (ii) and (iii) of Proposition 4.1.4. Since $\gamma(0) = 0 = \gamma^{-1}(0)$ and $\gamma(1) = 1 = \gamma^{-1}(1)$, it follows that C_γ satisfies (i) and (ii). For the rectangle inequality, assume $a_1, a_2, b_1, b_2 \in [0,1]$ with $a_1 \leq b_1$ and $a_2 \leq b_2$. Define

$$\begin{aligned} s &:= C(\gamma(a_1), \gamma(a_2)), \quad t := C(\gamma(a_1), \gamma(b_2)), \\ v &:= C(\gamma(b_1), \gamma(a_2)), \quad w := C(\gamma(b_1), \gamma(b_2)). \end{aligned}$$

Since γ is increasing, $\gamma(a_1) \leq \gamma(b_1)$ and $\gamma(a_2) \leq \gamma(b_2)$, and because C is a copula, we have $s - t - v + w \geq 0$ by (iii) of Proposition 4.1.4. It remains to show that

$$\gamma^{-1}(s) - \gamma^{-1}(t) - \gamma^{-1}(v) + \gamma^{-1}(w) \geq 0.$$

Since C is increasing in both arguments, we have either $s \leq t \leq v \leq w$ or $s \leq v \leq t \leq w$. If the four numbers are distinct, then from [104],

$$\frac{\gamma^{-1}(t) - \gamma^{-1}(s)}{t - s} \leq \frac{\gamma^{-1}(w) - \gamma^{-1}(v)}{w - v}$$

because γ^{-1} is convex. But we know that $t - s \leq w - v$, and so $\gamma^{-1}(t) - \gamma^{-1}(s) \leq \gamma^{-1}(w) - \gamma^{-1}(v)$ as desired. The proof is similar for the case where two or three of the numbers s, t, v, w are equal. Conversely, assume C_γ is a copula for any copula C. Let $a, b \in [0,1]$ with $a \leq b$ and define $u_1 = v_1 = \gamma^{-1}((a+1)/2)$ and $u_2 = v_2 = \gamma^{-1}((b+1)/2)$ so that $\gamma(u_1) = \gamma(v_1) = (a+1)/2$ and $\gamma(u_2) = \gamma(v_2) = (b+1)/2$. Now consider the special case $C = W$. Then

$$W(\gamma(u_1), \gamma(v_1)) = \max\left\{\frac{a+1}{2} + \frac{a+1}{2} - 1, 0\right\} = a$$

and similarly, $W(\gamma(u_1), \gamma(v_2)) = W(\gamma(u_2), \gamma(v_1)) = (a+b)/2$ and $W(\gamma(u_2), \gamma(v_2)) = b$. By assumption W_γ is a copula and thus

$$\gamma^{-1}(a) - 2\gamma^{-1}\Big(\frac{a+b}{2}\Big) + \gamma^{-1}(b) \geq 0$$

which is readily rearranged to

$$\gamma^{-1}\Big(\frac{a+b}{2}\Big) \leq \frac{1}{2}(\gamma^{-1}(a) + \gamma^{-1}(b)),$$

that is, γ^{-1} is midconvex. By continuity, γ^{-1} is convex as desired (Exercise 4.29). ■

One may generalise the theorem to also include the case where $\gamma(0)$ is not necessarily zero, and γ^{-1} is defined on $[0,1]$ as a *pseudo-inverse* instead (see Exercise 4.27). This can be found in [26].

Example 4.6.2. In Proposition 4.5.2, we saw that the copula of the componentwise maxima for a sequence $\{X_{i,1}, X_{i,2}\}_{i=1}^n$ was given by

$$C(u_1, u_2) = C^n(u_1^{1/n}, u_2^{1/n}).$$

That this is a copula is a special case of the above theorem with the function $\gamma(x) = x^{1/n}$, which is readily seen to satisfy the requirements in the theorem. ∘

Generating two-parameter families via Archimedean copulas

All Archimedean copula families presented in this chapter depend on a single parameter. We here present a simple method for extending these families to one with two parameters. The theoretical machinery lies in the following theorem.

Theorem 4.6.3. *Let φ be a generator for a two-dimensional Archimedean copula, and let $\alpha, \beta > 0$ be fixed. Define the functions*

$$\varphi_{\alpha,1}(t) = \varphi(t^\alpha), \quad \varphi_{1,\beta}(t) = \varphi(t)^\beta.$$

Then the following assertions hold:

1. *If $\beta \geq 1$, then $\varphi_{1,\beta}$ is a valid generator.*
2. *If $\alpha \leq 1$, then $\varphi_{\alpha,1}$ is a valid generator.*
3. *If φ is twice differentiable and $t\varphi'(t)$ is non-decreasing on $(0,1)$, then $\varphi_{\alpha,1}$ is a valid generator for all $\alpha > 0$.*

The proof of this theorem is left to the reader, see Exercise 4.40. The result leads to the following definition ([93]).

Definition 4.6.4. Let $\mathcal{A}_2$ denote the set of valid generators for a two-dimensional Archimedean copula. For a fixed $\varphi \in \mathcal{A}_2$, we call the family

$$\{\varphi_{\alpha,1} \in \mathcal{A}_2 : \varphi_{\alpha,1}(t) = \varphi(t^\alpha)\}$$

the *interior power family* associated with φ. We call

$$\{\varphi_{1,\beta} \in \mathcal{A}_2 : \varphi_{1,\beta}(t) = \varphi(t)^\beta\}$$

the *exterior power family* associated with φ.

Example 4.6.5. The exterior power family of $\varphi(t) = -\log t$ is the Gumbel family. The interior power family of $\varphi(t) = 1/t - 1$ is the Clayton family. ∘

Example 4.6.6. Choosing $\theta \in [-1, 1]$ and defining $\varphi(t) = \log((1-\theta(1-t))/t)$ generates an Archimedean copula, the Ali–Mikhail–Haq family. This family will be considered throughout several exercises, e.g. Exercise 4.6. We have

$$t\varphi'(t) = t\frac{(\theta-1)/t^2}{(1-\theta(1-t))/t} = \frac{\theta-1}{1-\theta(1-t)},$$

and since

$$(t\varphi'(t))' = \frac{\theta(1-\theta)}{(1-\theta(1-t))^2} \geq 0$$

whenever $\theta \geq 0$, we obtain that for $\theta \in [0,1]$, the interior power family associated with φ is given by the collection of functions

$$\varphi_{\alpha,1}(t) = \log\Big(\frac{1-\theta(1-t^\alpha)}{t^\alpha}\Big), \quad \alpha > 0.$$

The copulas in this family are explicitly given by

$$C_{\theta;\alpha,1}(u_1,u_2) = \frac{u_1u_2}{(1-\theta(1-u_1^{1/\alpha})(1-u_2^{1/\alpha}))^\alpha}.$$

∘

By combining the interior and exterior power families, one can obtain valid Archimedean copula generators on the form

$$\varphi_{\alpha,\beta}(t) = (\varphi(t^\alpha))^\beta.$$

The range of α will of course depend on the starting generator φ, see points 2. and 3. in Theorem 4.6.3.

Example 4.6.7. We know that the Clayton generator for $\theta = 1$, $\varphi(t) = 1/t - 1$, generates the copula

$$C(u_1,u_2) = \frac{u_1u_2}{u_1+u_2-u_1u_2}.$$

Clearly, $t\varphi'(t) = -1/t$ is non-decreasing for $t \in (0,1)$, and so we may construct the two-parameter family with generator

$$\varphi_{\alpha,\beta}(t) = (t^{-\alpha}-1)^\beta, \quad \alpha > 0, \beta \geq 1.$$

One can verify that the copula family is explicitly given by

$$C_{\alpha,\beta}(u_1,u_2) = \Big(1+\Big((u_1^{-\alpha}-1)^\beta + (u_2^{-\alpha}-1)^\beta\Big)^{1/\beta}\Big)^{-1/\alpha}.$$

∘

4.7 Additional proofs and results*

In this section, we provide additional proofs and technicalities that didn't fit the theme of the main text. We start by providing a general proof of Sklar's Theorem without any continuity restrictions on the marginals. An elegant general proof is provided by [107] who uses the *distributional transform* as the main tool.

Definition 4.7.1. Let X be a random variable with distribution function F. The *modified distribution function* $\tilde{F}(x,t)$ is defined by

$$\tilde{F}(x,t) = P(X<x) + tP(X=x) = F(x-) + t\Delta F(x), \quad t \in [0,1].$$

The *distributional transform* of X is given by

$$U := \tilde{F}(X,V)$$

where V is a Unif$[0,1]$ random variable independent of X.

If F is continuous, then $\tilde{F}(x,t) = F(x)$, and in particular, $\tilde{F}$ does not depend on the t parameter. In this case, it is well-known (see the appendix for details) that U as defined above is Unif$[0,1]$. This turns out to hold more generally as the following lemma shows.

Lemma 4.7.2. *Let $U = \tilde{F}(X,V)$ denote the distributional transform of $X \sim F$. Then U is uniform on $[0,1]$ and $X = F^{\leftarrow}(U)$ a.s.*

Proof. Let $u \in [0,1]$ and define $q(u) := P(X < F^{\leftarrow}(u))$ and $p(u) := P(X = F^{\leftarrow}(u))$. Consider the event

$$\{U \leq u\} = \{F(X-) + V(F(X) - F(X-)) \leq u\}.$$

For this event to hold, we need either $X < F^{\leftarrow}(u)$ or $X = F^{\leftarrow}(u)$. In the former case, the event clearly holds. If $X = F^{\leftarrow}(u)$, we also need that

$$F(F^{\leftarrow}(u)-) + V(F(F^{\leftarrow}(u)) - F(F^{\leftarrow}(u)-)) \leq u.$$

But since $F(F^{\leftarrow}(u)) = P(X \leq F^{\leftarrow}(u)) = p(u)+q(u)$ and $F(F^{\leftarrow}(u)-) = P(X < F^{\leftarrow}(u)) = q(u)$, we can rewrite the event to

$$\{U \leq u\} = \{X < F^{\leftarrow}(u)\} \cup \{X = F^{\leftarrow}(u), q(u) + Vp(u) \leq u\}.$$

Now divide into two cases. If F is continuous at $F^{\leftarrow}(u)$, then $p(u) = 0$ and $q(u) = u$, implying

$$P(U \leq u) = P(X \leq F^{\leftarrow}(u)) = q(u) = u.$$

If F is discontinuous at $F^{\leftarrow}(u)$, F has a jump in this quantile, and thus $p(u) > 0$ and

$$\begin{aligned} P(U \leq u) &= P(X < F^{\leftarrow}(u)) + P\Big(X = F^{\leftarrow}(u), V \leq \frac{u - q(u)}{p(u)}\Big) \\ &= q(u) + p(u)P\Big(V \leq \frac{u - q(u)}{p(u)}\Big) = u. \end{aligned}$$

This shows that U is distributed according to Unif$[0,1]$. We now prove the second assertion, namely that $X = F^{\leftarrow}(U)$ a.s. Fix an $\omega \in \Omega$ (the background space) and consider $x = X(\omega)$ and $u = U(\omega) = \tilde{F}(X(\omega), V(\omega))$. By definition, we have $F(x-) \leq u \leq F(x)$. If $F(x-) < u \leq F(x)$, then $x = F^{\leftarrow}(u)$. If $F(x-) = u$, it may occur that $F^{\leftarrow}(u) < x$, but since the values of ω where this occurs constitute a null set, we may conclude that the second assertion holds. ■

Proof of Theorem 4.1.2. Let $\mathbf{X} = (X_1, ..., X_d)$ have marginal distribution functions $F_1, ..., F_d$ and joint distribution function F. Let V be independent of $\mathbf{X}$ and uniformly distributed on $[0,1]$. Define $U_i := \tilde{F}_i(X_i, V)$. By the lemma above, we know that each U_i is uniformly distributed on $[0,1]$, and we let C be the distribution function of $\mathbf{U} = (U_1, ..., U_d)$. Then C is a copula and

$$\begin{aligned} C(F_1(x_1), ..., F_d(x_d)) &= P(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ &= P(F_1^{\leftarrow}(U_1) \leq x_1, ..., F_d^{\leftarrow}(U_d) \leq x_d) \\ &= P(X_1 \leq x_1, ..., X_d \leq x_d) = F(x_1, ..., x_d), \end{aligned}$$

which shows that C is a copula of $\mathbf{X}$ as desired. The rest of the proof is identical to the one presented above. ■

Remark 4.7.3. The full proof of Sklar's Theorem presented here offers some intuition for why uniqueness fails in general for non-continuous marginals. There is a lot of freedom in how V is chosen when defining the distributional transform of X_i. We could for example choose the same V for every X_i, or every X_i could get its own V. This of course influences the dependence structure between the U_i and therefore the copula C.

Notes and comments

The book by Nelsen, [91], contains all the information about copulas that we use in this course. The interested reader can find many supplementary examples there, and most of the proofs of the results in this chapter can be found there as well, and if not, references are given. Chapter 2 describes additional properties of copulas, including ordering which was not covered here. Chapter 3 covers a wide range of methods to construct copulas with both geometric and purely algebraic methods. Chapter 4 contains all the theoretical details on Archimedean copulas as well as a table of copulas with corresponding generators, see pages 116-119. Quite a few of the exercises below are also borrowed from [91]. Chapter 7 in [88] covers copulas in more generality than here. In particular, see section 7.2.1 for a complete characterisation of the copulas of countermonotonic and comonotonic random vectors. If the reader reads the section on Archimedean copulas (section 7.4), they should be aware that φ and φ^{-1} are switched. For a more detailed survery on extreme value copulas, see [49]. For more on statistics of multivariate extremes, see chapter 7-9 in [67]. For more on bivariate extremes, consult [6]. The construction of the dependence function for extreme value copulas presented above is due to [97]. The method of transforming a copula to obtain a new copula presented in Theorem 4.6.1 is due to [75] and [44].

4.8 Exercises

Theoretical exercises

Exercise 4.1:
Consider the random vector (X_1, X_2) where X_1 is Bernoulli with success probability $p \in (0,1)$ and $X_2 \mid X_1 = x \sim \mathcal{N}(2x, 1)$.

1)Determine the marginal distributions F_1, F_2 of X_1 and X_2.

2)Determine the joint distribution of (X_1, X_2).

3)Find a copula of (X_1, X_2). Is there a unique copula?

Exercise 4.2:
Let $X \sim \mathcal{N}(0,1)$. Determine the copula of $(X, |X|)$.

Exercise 4.3:
Write out the rectangle inequality (see Proposition 4.1.4) in the case where $d = 2$. Use it to verify that the *Morgenstern copula* given by

$$C(u_1, u_2) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2)),$$

for $\theta \in [-1, 1]$ a parameter, is indeed a copula.

Exercise 4.4:
Verify that the *Cuadras–Augé copula* (see [16]) given by

$$C_\theta(u_1, u_2) = \min\{u_1, u_2\}^\theta (u_1 u_2)^{1-\theta} = \begin{cases} u_1 u_2^{1-\theta}, & u_1 \leq u_2 \\ u_1^{1-\theta} u_2, & u_1 \geq u_2 \end{cases}$$

for a parameter $\theta \in [0, 1]$ is indeed a copula. We will see this copula again in Exercise 4.32.

Exercise 4.5:
Let C and $\tilde{C}$ be copulas and $\theta \in [0, 1]$.

1)Verify that $\theta C + (1 - \theta)\tilde{C}$ is also a copula.

2)Say (U_1, U_2) and $(\tilde{U}_1, \tilde{U}_2)$ are pairs of random variables with uniform marginals and copulas C and $\tilde{C}$, respectively. Explain in detail how to construct a pair of random variables with copula $\theta C + (1 - \theta)\tilde{C}$.

We have shown that any convex combination of copulas is also a copula. One such example is the two-parameter *Fréchet family* of copulas ([36]): If $a, b \in [0, 1]$ with $a + b \leq 1$, these copulas are given by

$$C_{a,b}(u_1, u_2) = aM(u_1, u_2) + (1 - a - b)\Pi(u_1, u_2) + bW(u_1, u_2),$$

where Π is the independence copula and M, W the upper and lower Fréchet bounds, respectively. A family of copulas that includes M, W and Π is called *comprehensive*. Another example is the one-parameter family

$$C_\theta(u_1, u_2) = \frac{\theta^2(1 + \theta)}{2} M(u_1, u_2) + (1 - \theta^2)\Pi(u_1, u_2) + \frac{\theta^2(1 - \theta)}{2} W(u_1, u_2)$$

from [84].

Exercise 4.6:
Let (X, Y) have joint distribution function

$$G(x, y) = \frac{1}{1 + e^{-x} + e^{-y}}$$

for all $(x, y) \in \mathbb{R}^2$. This distribution is called *Gumbel's bivariate logistic distribution* ([51]).
1)Verify that X and Y both have the logistic distribution i.e.

$$F_X(x) = (1 + e^{-x})^{-1}, \quad F_Y(y) = (1 + e^{-y})^{-1}.$$

2)Show that (X, Y) has the copula

$$C(u_1, u_2) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}.$$

This distribution can be generalised by adding a parameter $\theta \in [-1, 1]$ by defining ([79] and [15])

$$F_\theta(x, y) = \frac{1}{1 + e^{-x} + e^{-y} + (1 - \theta)e^{-x-y}}.$$

It is easy to verify that the marginals still have the logistic distribution and that $\theta = 1$ yields Gumbels bivariate logistic distribution.
3)Show that the copula of F_θ is

$$C_\theta(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}.$$

Comment on the case $\theta = 0$.

The family of copulas $\{C_\theta\}_{\theta \in [-1,1]}$ is called the *Ali–Mikhail–Haq* family ([2]).

Exercise 4.7:
In this exercise, we consider *Gumbel's bivariate exponential distribution* ([50]) given by

$$F_\theta(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}, & x,y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\theta \in [0, 1]$ is a parameter.
1)Determine the marginal distributions.
2)Verify that the copula of F_θ is given by

$$C_\theta(u_1, u_2) = u_1 + u_2 - 1 + (1 - u_1)(1 - u_2)e^{-\theta \log(1-u_1) \log(1-u_2)}.$$

Comment on the case $\theta = 0$.
3)Determine the survival copula of F_θ.

Exercise 4.8:
Let (X, Y) be a pair of non-negative random variables with joint survival function

$$\overline{F}(x, y) = \frac{1}{e^x + e^y - 1}.$$

1)Determine the marginal distributions of X and Y.
2)Show that the survival copula of (X, Y) is

$$\widehat{C}(u_1, u_2) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}.$$

Compare with Exercise 4.6.

Exercise 4.9:
Consider the so-called *Type B bivariate extreme value distribution* (see [73]) defined by the distribution function

$$F_\theta(x, y) = \exp\Big(- \Big(e^{-\theta x} + e^{-\theta y}\Big)^{1/\theta}\Big),$$

where $\theta \geq 1$ is a parameter. Show that the copula C_θ of (X_1, X_2) is the Gumbel copula,

$$C_\theta(u_1, u_2) = \exp\Big(- \Big((-\log u_1)^\theta + (-\log u_2)^\theta\Big)^{1/\theta}\Big).$$

Exercise 4.10:
In this exercise, we consider *Raftery's bivariate exponential distribution* ([99] and [100]). The distribution is constructed by letting Z_1, Z_2, Z_3 be iid exponential variables with parameter $\lambda > 0$ and letting I be Bernoulli with success probability $\theta \in (0, 1)$ independent of Z_1, Z_2 and Z_3. Then let

$$X_1 = (1 - \theta)Z_1 + IZ_3 \quad \text{and} \quad X_2 = (1 - \theta)Z_2 + IZ_3.$$

1)Let $x, y \geq 0$. Show that the joint survival function F of (X_1, X_2) is

$$\overline{F}(x, y) = \exp(-\lambda \max\{x, y\}) + \frac{1 - \theta}{1 + \theta} \exp\Big(- \frac{\lambda}{1 - \theta}(x + y)\Big)\Big(1 - \exp\Big(\lambda \frac{1 + \theta}{1 - \theta} \min\{x, y\}\Big)\Big).$$

2)Prove that X_1 and X_2 are both exponential with parameter $\lambda > 0$.
3)Show that the survival copula of (X_1, X_2) is given by

$$\widehat{C}_\theta(u_1, u_2) = M(u_1, u_2) + \frac{1 - \theta}{1 + \theta}(u_1 u_2)^{1/(1-\theta)}\Big(1 - \max\{u_1, u_2\}^{-(1+\theta)/(1-\theta)}\Big).$$

Comment on the cases where $\theta \uparrow 1$ and $\theta \downarrow 0$.

Exercise 4.11:
Let (X, Y) be random variables with copula C and marginal distribution functions F_X and F_Y.

1)Verify that $P(\{X \leq x\} \cup \{Y \leq y\}) = \tilde{C}(F_X(x), F_Y(y))$ where $\tilde{C}(u_1, u_2) = u_1 + u_2 - C(u_1, u_2)$ is called the *dual copula* of (X, Y).

2)Verify that $P(\{X > x\} \cup \{Y > y\}) = C^*(\overline{F}_X(x), \overline{F}_Y(y))$ where $C^*(u_1, u_2) = 1 - C(1 - u_1, 1 - u_2)$ is called the *co-copula* of (X, Y).

It is important to remark that neither the dual copula or the co-copula are copulas.

Exercise 4.12: Exercise 2.8 in [91]
For a two-dimensional copula C, the function $\delta_C(t) = C(t, t)$ for $t \in [0, 1]$ is called the *diagonal section* of C.

1)Prove that $\max\{2t - 1, 0\} \leq \delta_C(t) \leq t$ for all $t \in [0, 1]$.

2)Show that $\delta_C(t) = \delta_M(t)$ for all $t \in [0, 1]$ implies $C = M$.

3)Show that $\delta_C(t) = \delta_W(t)$ for all $t \in [0, 1]$ does not necessarily imply that $C = W$. Hint: Consider a pair (U_1, U_2) uniformly distributed on

$$\{(u, u + 1/2) : u \in [0, 1/2]\} \cup \{(u, u - 1/2) : u \in [1/2, 1]\}.$$

Exercise 4.13:
For a copula C, the function $t \mapsto C(t, 1 - t)$ for $t \in [0, 1]$ is called the *secondary diagonal section* of C. In this exercise, we show that $C(t, 1 - t) = 0$ for all t holds if and only if $C = W$, the lower Fréchet bound given by $W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$.

1)Show that $W(t, 1 - t) = 0$ for all $t \in [0, 1]$.

Now let C be a copula such that $C(t, 1 - t) = 0$ for all $t \in [0, 1]$.

2)Assume $u_1 + u_2 \leq 1$. Show that $C(u_1, u_2) = 0$.

3)Now assume $u_1 + u_2 > 1$. Show that $C(u_1, u_2) \leq u_1 + u_2 - 1$. Hint: Use the rectangle inequality from Proposition 4.1.4.

4)Collect the pieces and show that $C = W$.

Exercise 4.14:
For any function $H : \mathbb{R}^2 \to \mathbb{R}$, we define the *H-volume* V_H of a box $[a, b] \times [c, d]$ to be

$$V_H([a, b] \times [c, d]) := H(b, d) - H(b, c) - H(a, d) + H(a, c).$$

If H is a distribution function for (X_1, X_2), this would be the probability $P(X_1 \in [a, b], X_2 \in [c, d])$ which can be seen by making a drawing. Now suppose that C is a two-dimensional copula, and let $a, b \in [0, 1]$ be fixed. Define the function $K_{a,b}$ on $[0, 1]^2$ by

$$K_{a,b}(u, v) = V_C([a(1 - u), u + a(1 - u)] \times [b(1 - v), v + b(1 - v)]).$$

Show that $K_{a,b}$ is a copula. Comment on the special cases $a, b \in \{0, 1\}$. Hint: If (U_1, U_2) has copula C, consider the transformed variables

$$V_1 = \begin{cases} \frac{a - U_1}{a}, & U_1 < a \\ \frac{U_1 - a}{1 - a}, & U_1 \geq a, \end{cases} \qquad V_2 = \begin{cases} \frac{b - U_2}{b}, & U_2 < b \\ \frac{U_2 - b}{1 - b}, & U_2 \geq b. \end{cases}$$

Exercise 4.15:
In this exercise, we explore the concept of comonotonicity further.

1)Prove the following converse of the first part of Proposition 4.1.7: If $(X_1, ..., X_d)$ has M as a copula, then we may write

$$(X_1, ..., X_d) \overset{\mathrm{d}}{=} (\alpha_1(Z), ..., \alpha_d(Z))$$

for a univariate variable Z and non-decreasing functions $\alpha_1, ..., \alpha_d$. Hint: Let $Z = U \sim \mathrm{Unif}(0, 1)$.

2)Assume that $X_1, ..., X_d$ are continuous. Show that $(X_1, ..., X_d)$ is comonotone if and only if for every pair (i, j), we have $X_j = \alpha_{ji}(X_i)$ a.s. for some non-decreasing function α_{ji}. Hint: Start by arguing why it holds (in general) that

$$P(F^{\leftarrow}(F(X)) = X) = 1$$

whenever $X \sim F$. Use the proof from the previous subproblem.

3)Show that if $(X_1, ..., X_d)$ is comonotone with marginal distribution functions $F_1, ..., F_d$, then

$$F^{\leftarrow}_{X_1+\cdots+X_d}(p) = F_1^{\leftarrow}(p) + \cdots + F_d^{\leftarrow}(p), \quad p \in (0, 1).$$

Hint: Use Lemma 1.3.13.

Exercise 4.16:
Let X and Y be random variables with continuous distribution functions F_X and F_Y. Let α, β be functions, C the copula of (X, Y) and $\tilde{C}$ the copula of $(\alpha(X), \beta(Y))$.

1)If α is strictly increasing and β strictly decreasing, prove that

$$\tilde{C}(u_1, u_2) = u_1 - C(u_1, 1 - u_2).$$

2)If α is strictly decreasing and β is strictly increasing, prove that

$$\tilde{C}(u_1, u_2) = u_2 - C(1 - u_1, u_2).$$

3)If α and β are strictly decreasing, prove that

$$\tilde{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

Exercise 4.17:
Let $X_1, ..., X_d$ be random variables with continuous distribution functions $F_1, ..., F_d$. Let $X_{1,n} = \max\{X_1, ..., X_d\}$ and $X_{n,n} = \min\{X_1, ..., X_d\}$. Let $F_{1,n}$ and $F_{n,n}$ denote the distribution function of $X_{1,n}$ and $X_{n,n}$, respectively.

1)Prove that

$$\max\{F_1(x), ..., F_d(x)\} \leq F_{n,n}(x) \leq \min\Big\{\sum_{j=1}^{d} F_j(x), 1\Big\}.$$

2)Prove that

$$\max\Big\{\sum_{j=1}^{d} F_j(x) - d + 1, 0\Big\} \leq F_{1,n}(x) \leq \min\{F_1(x), ..., F_d(x)\}.$$

Exercise 4.18:
In this exercise, we consider some further properties of copulas.

1)Let (U_1, U_2) have uniform marginals and copula C and consider the conditional distribution function $C_{U_2|U_1}(u_2 \mid u_1) = P(U_2 \leq u_2 \mid U_1 = u_1)$. Show that

$$C_{U_2|U_1}(u_2 \mid u_1) = \frac{\partial}{\partial u_1} C(u_1, u_2).$$

You should also argue why this derivative exists (at least almost everywhere).

2)Recall that if we have a multivariate distribution function $F(x_1, ..., x_d)$ with density f, then

$$f(x_1, ..., x_d) = \frac{\partial^d F(x_1, ..., x_d)}{\partial x_1 \cdots \partial x_d}.$$

Use this to verify that if C is the (implicit) copula for the variables $(X_1, ..., X_d)$ with continuous and strictly increasing marginal distribution functions F_i, and f is the joint density, then the density of C is

$$c(u_1, ..., u_d) = \frac{f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}.$$

Here the f_i are the marginal densities.

Exercise 4.19:
Let (U_1, U_2) be a pair of uniform variables on $[0, 1]$.

1)Say the distribution function of (U_1, U_2) is the copula $M(u_1, u_2) = \min\{u_1, u_2\}$ from the upper Fréchet bound. Prove that $P(U_1 = U_2) = 1$ in at least two different ways.

2)Say the distribution function is instead given by $W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$, the lower Fréchet bound. Prove in at least two different ways that $P(U_1 + U_2 = 1) = 1$.

Exercise 4.20:
In this exercise, we fill the small gaps in the subsection on symmetry.

1)Let X have continuous distribution function F. Show that X is symmetric about a if and only if

$$F(a + x) = \overline{F}(a - x)$$

for all $x \in \mathbb{R}$.

2)Let X and Y be continuous with joint distribution function F. Show that (X, Y) is radially symmetric about (a, b) if and only if

$$F(a + x, b + y) = \overline{F}(a - x, b - y) \quad \text{for all } x, y \in \mathbb{R}.$$

Exercise 4.21:
In this exercise, we establish some basic properties of Archimedean copulas. Let C be a two-dimensional Archimedean copula with generator φ.

1)Show that C is *symmetric* i.e. $C(u_1, u_2) = C(u_2, u_1)$ for all $u_1, u_2 \in [0, 1]$.

2)Show that C is *associative* i.e. $C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3))$ for all $u_1, u_2, u_3 \in [0, 1]$.

3)If $c > 0$ is any constant, then $c\varphi$ is also a generator of C.

4)Say C is Archimedean with generator φ. Assume φ^{-1} is d times differentiable and that φ is differentiable. Show that the density is

$$c(u_1, ..., u_d) = (\varphi^{-1})^{(d)}\Big(\sum_{j=1}^{d} \varphi(u_j)\Big) \prod_{j=1}^{d} \varphi'(u_j),$$

where $(\varphi^{-1})^{(d)}$ denotes the d'th derivative of φ^{-1}.

Exercise 4.22:
Recall that the Frank copula has generator

$$\varphi(t) = -\log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$$

with $\theta \in \mathbb{R} \setminus \{0\}$ a parameter.

1)Verify that φ generates a two-dimensional copula by using Theorem 4.3.4.

2)Write the Frank copula in explicit form.

3)Now assume that $\theta > 0$. By using that $\varphi^{-1}(t) = f(g(t))$ for

$$f(x) = -\frac{1}{\theta}\log x, \quad g(x) = 1 + (e^{-\theta} - 1)e^{-x},$$

show that φ^{-1} is completely monotonic and conclude by Theorem 4.4.2 that φ generates an Archimedean copula in any dimension.

Exercise 4.23:
This exercise is a continuation of the previous. Let C_θ denote the Frank copula.

1)Show that the Frank copula satisfies $C_{-\theta}(u_1, u_2) = u_1 - C_\theta(u_1, 1-u_2) = u_2 - C_\theta(1-u_1, u_2)$.

2)Use the previous subproblem to verify that C_θ satisfies the requirement $C_\theta = \widehat{C}_\theta$ for radial symmetry.

Exercise 4.24:
Consider the function $\varphi : [0, 1] \to [0, \infty]$ given by

$$\varphi(t) = \log(1 - \theta \log t)$$

for $\theta \in (0, 1]$ a parameter.

1)Verify that φ is a valid generator for a two-dimensional Archimedean copula.

2)Write the corresponding copula in explicit form.

3)Does φ generate an Archimedean copula of arbitrary dimension $d \geq 2$? Hint: Show that

$$(\varphi^{-1})'''(0) = \frac{-\theta^2 + 3\theta - 1}{\theta^3}.$$

4)In the rest of the exercise, we let $\theta = 1$. For which values of $\alpha \in (0, \infty)$ is the interior power family generator $\varphi_{\alpha,1}(t) = \varphi(t^\alpha)$ valid?

5)Write out the two-parameter family parametrised by α, β (defined where?) via the generator $\varphi_{\alpha,\beta}(t) = \varphi(t^\alpha)^\beta$.

Exercise 4.25:
Consider the function $\varphi : [0, 1] \to [0, \infty]$ given by

$$\varphi(t) = \Big(\frac{1}{t} - 1\Big)^\theta,$$

where $\theta \in [1, \infty)$ is a parameter.

1)Verify that φ is a valid generator for a two-dimensional Archimedean copula.

2)Show that φ in fact generates an Archimedean copula of any dimension $d \geq 2$.

3)Write the copula in explicit form for any $d \geq 2$.

Exercise 4.26:
Consider the discrete random variable N with probabilities

$$P(N = k) = \frac{(1 - e^{-\theta})^k}{k\theta}, \quad k = 1, 2, ...$$

for $\theta > 0$ a fixed parameter.

1)Verify that

$$\sum_{k=1}^{\infty} P(N = k) = 1$$

so that the point probabilities do indeed specify a valid distribution. Hint: Recall the Taylor expansion for $\log(1 - x)$ when $x < 1$.

2)Compute the Laplace transform $\psi(t) = E[e^{-tN}]$.

3)Invert $\psi(t)$. From Lemma 4.4.7, we know that $\psi^{-1}(t)$ is a valid generator for a d-dimensional Archimedean copula. Which one?

4)Implement a function to simulate from this copula. You can use the following algorithms in your code:

Algorithm. Let N be a discrete variable with point probabilities $\{p_k\}_{k\in\mathbb{N}}$ and distribution function F. To simulate from N, do the following:

1. Simulate $U \sim \text{Unif}(0, 1)$. Set $k = 1$.
2. While $F(k) \leq U$, let $k \leftarrow k + 1$.
3. $N = k$ is then a sample from the distribution of N.

Algorithm. Let N be a discrete variable with point probabilities $\{p_k\}_{k\in\mathbb{N}}$, distribution function F and upper bound[2] M. To simulate from N, do the following:

1. Simulate $U \sim \text{Unif}(0,1)$. Set $L \leftarrow 0$, $U \leftarrow M$.
2. While $L < U - 1$, do the following:
   (i) Set $k \leftarrow [(L+U)/2]$ (here $[x]$ denotes the integer closest to x).
   (ii) If $U > F(k)$, then $L \leftarrow k$. Else $U \leftarrow k$.
3. $N = k$ is then a sample from the distribution of N.

If you want, you can provide an explanation for why these algorithms work. What is the rationale for choosing the second algorithm over the first? For more background, consult chapter 3.3 of [103].

Exercise 4.27: Non-strict Archimedean copulas
In this book, we only consider *strict* Archimedean copulas. An Archimedean copula is called strict if the generator φ satisfies $\varphi(0) = \infty$. This assumption can however be relaxed in the following way. If $\varphi : [0,1] \to [0,\infty]$ is a continuous and strictly decreasing function with $\varphi(1) = 0$, we define the *pseudo-inverse* $\varphi^{[-1]}$ by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & t \in [0, \varphi(0)] \\ 0, & t \in [\varphi(0), \infty] \end{cases}.$$

1)Check that $\varphi^{[-1]}(\varphi(u)) = u$ for all $u \in [0,1]$ and that $\varphi(\varphi^{[-1]}(t)) = \min\{t, \varphi(0)\}$.

You can verify that Theorem 4.3.4 still applies to this generalised construction. Just replace the inverse of φ with the pseudo-inverse.

2)Show that W is an Archimedean copula in the non-strict sense by finding the generator explicitly.

3)Show that an Archimedean copula C is strict if and only if $C(u_1, u_2) > 0$ for all $(u_1, u_2) \in (0,1]^2$.

4)Show that $\varphi(t) = (1-t)^\theta$ for $\theta \in [1, \infty)$ generates a non-strict two-dimensional Archimedean copula. Write down this copula explicitly.

Many results presented for Archimedean copulas in this text can be extended to non-strict Archimedean copulas. [91] a priori presents all Archimedean copula results in the more general setting of non-strictness.

Exercise 4.28:
In the previous exercise, we showed that the lower Fréchet bound W is an Archimedean copula in a slightly more general sense than presented in this book. In this exercise, we investigate the remaining theoretical copulas.

1)Show that the independence copula Π is an Archimedean copula.

[2]In practice, this M will have to be chosen very large for unbounded distributions.

Now consider the upper Fréchet bound $M(u_1, u_2) = \min\{u_1, u_2\}$. The following exercises will show that M is not an Archimedean copula.

2)Recall that the diagonal section for a two-dimensional copula is given by $\delta_C(t) = C(t, t)$, see Exercise 4.12. Show that for a general two-dimensional Archimedean copula C with generator φ, it holds that $\delta_C(t) = \varphi^{-1}(2\varphi(t))$. Show that this implies that $\delta_C(t) < t$ for all $t \in [0, 1]$ when C is Archimedean.

3)Show that M is not Archimedean.

Exercise 4.29:
In this exercise, we fill in the blanks of the proof of Theorem 4.3.4.

1)Let f be continuous and strictly decreasing. Show that f is convex if and only if f^{-1} is.

2)Recall that a function g is *midconvex* if for all s, t in the domain of g,

$$g\Big(\frac{s}{2} + \frac{t}{2}\Big) \leq \frac{1}{2}(g(s) + g(t)).$$

Show that if g is midconvex and continuous, then g is convex. Hint: Show and use that every real number can be approximated by rational numbers of the form $m/2^n$ for $m \in \mathbb{Z}$ and $n \in \mathbb{N}_0$.

Exercise 4.30:
In this exercise, we provide the details for why an Archimedean copula is called "Archimedean". Let $u \in [0, 1]$ and let C be a copula. The *C-powers* u_C^n of u are defined recursively by

$$u_C^1 := u \quad \text{and} \quad u_C^{n+1} = C(u, u_C^n).$$

The Archimedean property of an Archimedean copula then states that for any $u, v \in (0, 1)$, there exists a positive integer n such that $u_C^n < v$. Prove this property.

Exercise 4.31: Joint symmetry
In this exercise, we explore a stronger symmetry concept than radial symmetry, namely *joint symmetry*. A pair of random variables (X, Y) is said to be jointly symmetric about (a, b) if the four pairs of random variables $(X - a, Y - b), (X - a, b - Y), (a - X, Y - b)$ and $(a - X, b - Y)$ all have the same distribution.

1)Let X and Y have continuous distribution functions F_1 and F_2, and let F denote the joint distribution function of (X, Y). Show that (X, Y) is jointly symmetric about (a, b) if and only if

$$F(a + x, b + y) = F_1(a + x) - F(a + x, b - y)$$

and

$$F(a + x, b + y) = F_2(b + y) - F(a - x, b + y)$$

for all $x, y \in \mathbb{R} \cup \{\infty\}$. This establishes a criterion similar in spirit to the one for radial symmetry.

2)Let X and Y be as in the previous exercise and assume further that X is symmetric about a and that Y is symmetric about b. Prove that (X, Y) is jointly symmetric about (a, b) if and only if the copula C of (X, Y) satisfies

$$C(u_1, u_2) = u_1 - C(u_1, 1 - u_2) \quad \text{and} \quad C(u_1, u_2) = u_2 - C(1 - u_1, u_2)$$

for all $u_1, u_2 \in [0,1]$.

Exercise 4.32: Homogeneous copulas
We say that a copula C is *homogeneous* of degree $a \in \mathbb{R}$ if for every $\lambda, u_1, u_2 \in [0,1]$, it holds that

$$C(\lambda u_1, \lambda u_2) = \lambda^a C(u_1, u_2).$$

1)Show that M and Π are homogeneous and determine their degrees.

2)Consider the Cuadras–Augé family from Exercise 4.4. Show that this family is homogeneous. What is the degree?

3)It turns out that we have already exhausted all homogeneous copulas:

Theorem 4.8.1. *Suppose C is a homogeneous copula of degree a. Then $a \in [1,2]$, and C is a member of the Cuadras–Augé family with $\theta = 2 - a$.*

Prove this theorem. Hint: Use Exercise 4.12.

Exercise 4.33: Convex sums of copulas
Let $\{C_\theta\}$ be a family of copulas indexed by a parameter θ. We know from Exercise 4.5 that if the family is finite, then a convex sum of copulas from $\{C_\theta\}$ is again a copula.

Now let $\{C_\theta\}$ be any family of copulas indexed by θ. Say θ is a realisation of a random variable Θ with distribution function F_Θ (called the *mixing distribution*). We then define the *convex sum* of $\{C_\theta\}$ by

$$C(u_1, ..., u_d) = \int C_\theta(u_1, ..., u_d) dF_\Theta(\theta).$$

1)Prove that C is again a copula.

2)Let (U_1, U_2) be a pair of uniform variables jointly uniformly distributed on the rectangle in $[0,1]^2$ with vertices $(\theta, 0), (0, \theta), (1-\theta, 1), (1, 1-\theta)$. Sketch the support of (U_1, U_2) and show that the copula of (U_1, U_2) is given by

$$C_\theta(u_1, u_2) = \begin{cases} M(u_1, u_2), & |u_1 - u_2| \geq \theta \\ W(u_1, u_2), & |u_1 + u_2 - 1| \geq 1 - \theta \\ (u_1 + u_2 - \theta)/2, & \text{otherwise.} \end{cases}$$

3)Let $F_\Theta(\theta) = \theta^\alpha$ for $\theta \in [0,1]$ and $\alpha > 0$ a fixed parameter. Show that the convex sum, denoted C_α, of the copula family from the previous subexercise is given by

$$C_\alpha(u_1, u_2) = W(u_1, u_2) + \frac{1}{2(\alpha+1)}\Big((1 - |u_1 + u_2 - 1|)^{\alpha+1} - |u_1 - u_2|^{\alpha+1}\Big).$$

4)Show that C_α obtained in the previous subexercise is radially symmetric, i.e. $\widehat{C}_\alpha = C_\alpha$.

For more convex sums of $\{C_\theta\}$, see [34].

5)Consider now the family of copulas

$$C_\theta(u_1,u_2)=\begin{cases}\max\{u_1+u_2-1,\theta\}, & (u_1,u_2)\in[\theta,1]^2\\ \min\{u_1,u_2\}, & \text{otherwise}\end{cases}$$

where $\theta\in[0,1]$ a parameter. Let $F_\Theta(\theta)=\theta^\alpha$ for $\alpha>0$ a parameter. Show that the convex sum of $\{C_\theta\}$ for this mixing distribution is given by

$$C_\alpha(u_1,u_2)=M(u_1,u_2)-\frac{1}{\alpha+1}\Big(M(u_1,u_2)^{\alpha+1}-W(u_1,u_2)^{\alpha+1}\Big).$$

Comment on the cases $\alpha\downarrow 0$ and $\alpha\to\infty$.

Exercise 4.34: Ordinal sums
Let $\{C_i\}_{i\in\mathcal{I}}$ be an arbitrary family of copulas and $\{I_i\}_{i\in\mathcal{I}}$ a collection of subsets $I_i=[a_i,b_i]$ $(b_i\neq a_i)$ which only overlap at common endpoints such that $\cup_{i\in\mathcal{I}}I_i=[0,1]$. The *ordinal sum of* $\{C_i\}$ *with respect to* $\{I_i\}$ is the copula given by

$$C(u_1,u_2)=\begin{cases}a_i+(b_i-a_i)C_i\Big(\frac{u_1-a_i}{b_i-a_i},\frac{u_2-a_i}{b_i-a_i}\Big), & (u_1,u_2)\in I_i^2\\ M(u_1,u_2), & \text{otherwise.}\end{cases}$$

One can think of C as being pasted together of appropriately scaled copies of $\{C_i\}_{i\in\mathcal{I}}$.
1)Let $\theta\in(0,1)$. Show that the ordinal sum of $\{W,W\}$ with respect to $\{[0,\theta],[\theta,1]\}$ is

$$C_\theta(u_1,u_2)=\begin{cases}\max\{0,u_1+u_2-\theta\}, & (u_1,u_2)\in[0,\theta]^2\\ \max\{\theta,u_1+u_2-1\}, & (u_1,u_2)\in(\theta,1]^2\\ M(u_1,u_2), & \text{otherwise.}\end{cases}$$

2)Determine the ordinal sum of $\{\Pi,W\}$ with respect to $\{[0,\theta],[\theta,1]\}$ where $\theta\in(0,1)$.
3)Prove the following result: A copula C is an ordinal sum if and only if $C(t,t)=t$ for some $t\in(0,1)$. Hint: For the non-trivial direction, consider the functions

$$C_1(u_1,u_2)=\frac{C(tu_1,tu_2)}{t},\quad C_2(u_1,u_2)=\frac{C(t+(1-t)u_1,(1-t)u_2-t)}{1-t}.$$

4)Let (U_1,U_2) be a pair of Unif$[0,1]$ variables. Show that the following are equivalent:

(i) $C(t,t)=t$ for some $t\in(0,1)$,

(ii) $P(\max\{U_1,U_2\}\leq t)=t$ for some $t\in(0,1)$,

(iii) $P(\min\{U_1,U_2\}\leq t)=t$ for some $t\in(0,1)$,

(iv) $P(\max\{U_1,U_2\}\leq t)=P(\min\{U_1,U_2\}\leq t)$ for some $t\in(0,1)$,

(v) $P((U_1-t)(U_2-t)\geq t)=1$ for some $t\in(0,1)$.

Exercise 4.35: Vaswani's bivariate normal distribution
Let V be a Unif$[0,1]$ variable and define

$$X_1 = \Phi^{-1}(V), \quad X_2 = \begin{cases} -\Phi^{-1}(V + 1/2), & V \in [0, 1/2] \\ -\Phi^{-1}(V - 1/2), & V \in (1/2, 1]. \end{cases}$$

1)Show that X_1 and X_2 are both standard normal.

2)Show that $P(\Phi(X_1) \oplus \Phi(X_2) = 1/2) = 1$ where $\oplus$ denotes addition modulo 1.

3)Show that the copula C of (X_1, X_2) is equal to C_θ of the previous exercise for $\theta = 1/2$, i.e.

$$C(u_1, u_2) = \begin{cases} \max\{0, u_1 + u_2 - 1/2\}, & (u_1, u_2) \in [0, 1/2]^2 \\ \max\{1/2, u_1 + u_2 - 1\}, & (u_1, u_2) \in (1/2, 1]^2 \\ M(u_1, u_2), & \text{otherwise.} \end{cases}$$

This copula is from [119].

Exercise 4.36: The copula of the componentwise maxima
Prove Proposition 4.5.2.

Exercise 4.37: The copula of the extreme value statistics
The goal of this exercise is to determine the copula of the minimal and maximal value of an iid sample, following along the lines of [108]. Let $X_1, ..., X_n$ be an iid sample with $X_1 \sim F$ for F continuous. Let $X_{1,n} = \max\{X_1, ..., X_n\}$ and $X_{n,n} = \min\{X_1, ..., X_n\}$.

1)Show that the joint distribution function F^Δ of $(-X_{n,n}, X_{1,n})$ is given by

$$F^\Delta(t_1, t_2) = \max\{F(t_2) - F(-t_1), 0\}^n.$$

2)Show that the copula C^Δ of $(-X_{n,n}, X_{1,n})$ is the Clayton copula[3]

$$C^\Delta(u_1, u_2) = \max\{u_1^{1/n} + u_2^{1/n} - 1, 0\}^n.$$

3)Conclude that the copula C of $(X_{n,n}, X_{1,n})$ is given by

$$C(u_1, u_2) = u_2 - \max\{(1 - u_1)^{1/n} + u_2^{1/n} - 1, 0\}^n.$$

4)Show that $X_{n,n}$ and $X_{1,n}$ are asymptotically independent.

Exercise 4.38: The Marshall–Olkin family
In this exercise, we investigate further the Marshall–Olkin family of copulas

$$C_{\alpha,\beta}(u_1, u_2) = \min\{u_1^{1-\alpha} u_2, u_1 u_2^{1-\beta}\} = \begin{cases} u_1^{1-\alpha} u_2, & u_1^\alpha \geq u_2^\beta \\ u_1 u_2^{1-\beta}, & u_1^\alpha \leq u_2^\beta, \end{cases}$$

[3]The Clayton copula as given in e.g. [91] is equal to $C_\theta^{\mathrm{Cl}}(u_1, u_2) = \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-1/\theta}$ which allows the domain of θ to be extended to $[-1, \infty) \setminus \{0\}$.

for $\alpha, \beta \in [0,1]$.

1)Consider the construction of the Marshall–Olkin copula in the text and recall that

$$\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_3}, \quad \beta = \frac{\lambda_2}{\lambda_2 + \lambda_3}.$$

If $Z_1 \sim \text{Exp}(\lambda)$ and $Z_2 \sim \text{Exp}(\lambda')$ are independent, verify that

$$P(Z_1 \geq Z_2) = \frac{\lambda'}{\lambda + \lambda'}.$$

Use this to provide an interpretation of the α and β parameters.

2)In the text, we only derived this copula for $\alpha, \beta \in (0,1)$. Show that $C_{\alpha,\beta}$ still defines a copula for $\alpha, \beta \in \{0,1\}$.

3)As presented in the text, the Marshall–Olkin copula is an extreme value copula with dependence function

$$A(t) = 1 - \min\{\beta t, \alpha(1-t)\}.$$

Verify that this is indeed a valid dependence function, and that it is the dependence function for the Marshall–Olkin copula. Verify also that the Marshall–Olkin copula is an extreme value copula by showing directly that it is max-stable.

4)The figures in the text illustrated that the Marshall–Olkin copula has a singular component. In this subproblem, we investigate why. Let (U_1, U_2) have distribution function equal to $C_{\alpha,\beta}$. Start by computing $P(U_2 \leq u_2 \mid U_1 = u_1)$ in all points such that $u_1^\alpha \neq u_2^\beta$. Then show that

$$P(U_1^\alpha = U_2^\beta) = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta},$$

and explain what we saw in the plots. Hint: Depending on your approach, you may need to compute $P(U_2 = u_1^{\alpha/\beta} \mid U_1 = u_1)$. To do so, consider the difference of two limits, one from above, the other from below.

5)Derive an algorithm to simulate from the Marshall–Olkin copula. Then reproduce the plots in the text.

Exercise 4.39: The Plackett family

In this exercise we consider the *Plackett* family of copulas given by

$$C_\theta(u_1, u_2) = \frac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{(1 + (\theta - 1)(u_1 + u_2))^2 - 4u_1u_2\theta(\theta - 1)}}{2(\theta - 1)}$$

for a parameter $\theta \in (0, \infty) \setminus \{1\}$. We may extend to all $\theta > 0$ by letting $C_1 = \Pi$. This family arises by considering a continuous extension of the *odds ratio* for 2×2 contingency tables and letting this odds ratio be constant and equal to θ. We refer to Section 3.3.1 of [91] and [98] for details.

1)Show that

$$\lim_{\theta \to 0^+} C_\theta(u_1, u_2) = W(u_1, u_2).$$

2)Show that

$$\lim_{\theta \to \infty} C_\theta(u_1, u_2) = M(u_1, u_2).$$

3)Prove that

$$C_{1/\theta}(u_1, u_2) = u_1 - C_\theta(u_1, 1 - u_2) = u_2 - C_\theta(1 - u_1, u_2).$$

Use this to conclude that the Plackett family satisfies the requirement $C_\theta = \widehat{C}_\theta$ for radial symmetry, see Theorem 4.1.14.

4)Prove that the following algorithm is correct.

Algorithm 4.8.1. To simulate a sample from the Plackett family with parameter $\theta > 0$, do the following:

1. Generate independent $U_1, V \sim \text{Unif}(0, 1)$.
2. Set
$$\begin{aligned} Z_1 &\leftarrow V(1 - V), \\ Z_2 &\leftarrow \theta + Z_1(\theta - 1)^2, \\ Z_3 &\leftarrow 2Z_1(U_1\theta^2 + 1 - U_1) + \theta(1 - 2Z_1), \\ Z_4 &\leftarrow \sqrt{\theta(\theta + 4Z_1U_1(1 - U_1)(1 - \theta)^2)}. \end{aligned}$$
3. Set
$$U_2 \leftarrow \frac{Z_3 - (1 - 2V)Z_4}{2Z_2}.$$
4. Now (U_1, U_2) is a sample from the Plackett copula.

5)Simulate some samples using this algorithm and plot them for different values of θ. We will revisit this copula in the exercises in the next chapter on dependence concepts.

Exercise 4.40:
Prove Theorem 4.6.3. Hint: The requirements in Theorem 4.3.4 can be checked directly.

Exercise 4.41:
In this exercise, we provide the details of the claims in the proof of Theorem 4.5.10 involving functional equations. A function f satisfies *Cauchy's functional equation* if

$$f(x + y) = f(x) + f(y)$$

for all x, y in the domain of f. Such a function is also called *additive*.

1)Show that if f is additive, then it is $\mathbb{Q}$-*linear*, i.e. f is linear when restricted to $\mathbb{Q}$.

2)Show that if f is additive and continuous, then it is linear.

3)Assume $f \geq 0$ satisfies the multiplicative Cauchy equation $f(xy) = f(x)f(y)$. If f is also continuous, show that $f(x) = x^a$ for some a.

In fact, if f is additive and continuous at just a single point, then it is continuous everywhere on its domain. A list of interesting properties of additive functions can be found at https://math.stackexchange.com/questions/423492/overview-of-basic-facts-about-cauchy-functional-equation. A standard reference for functional equations is [1].

Practical exercises

Exercise 4.42:
Prove that the following algorithm (from [72]) indeed produces a sample from the Ali–Mikhail–Haq family

$$C_\theta(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1-u_1)(1-u_2)}$$

with parameter $\theta \in [-1, 1]$ as given in Exercise 4.6

Algorithm. To simulate a sample from the Ali–Mikhail–Haq copula, do the following:

1. Generate independent $U_1, V \sim \text{Unif}(0, 1)$.
2. Set
$$\begin{aligned} W_1 &\leftarrow 1 - U_1, \\ W_2 &\leftarrow -\theta(2W_1V + 1) + 2\theta^2 W_1^2 V + 1, \\ W_3 &\leftarrow \theta^2(4W_1^2V - 4W_1V + 1) - \theta(4W_1V - 4V + 2) + 1. \end{aligned}$$
3. Set
$$U_2 \leftarrow \frac{2V(W_1\theta - 1)^2}{W_2 + \sqrt{W_3}}.$$
4. (U_1, U_2) is then a sample from the Ali–Mikhail–Haq copula.

Implement this algorithm and simulate some samples.

Exercise 4.43:
Prove that the algorithm below (from [25]) does indeed produce a sample from the two-dimensional Clayton copula.

Algorithm. To sample from the two-dimensional Clayton copula with parameter $\theta > 0$, do the following:

1. Simulate two independent standard exponential variables X_1, X_2. Also simulate a $\Gamma(\theta, 1)$ distributed variable Z independent of X_1 and X_2.
2. Set
$$U_1 \leftarrow \left(1 + \frac{X_1}{Z}\right)^{-\theta} \quad \text{and} \quad U_2 \leftarrow \left(1 + \frac{X_2}{Z}\right)^{-\theta}.$$
3. The pair (U_1, U_2) is then a sample from the Clayton copula.

Then implement the algorithm and produce some samples.

Exercise 4.44:
The d-dimensional t copula is given by

$$C^t_{\nu,\Sigma}(u_1, ..., u_d) = t_{\nu,\Sigma}(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d))$$

with $t_{\nu,\Sigma}$ the distribution function of $t(\nu, 0, \Sigma)$ and t_ν the univariate t distribution function with ν degrees of freedom.

1)Describe a procedure to simulate from the t copula.

2)Simulate 1000 samples from the t copula with $\nu = 3$ and

$$\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}.$$

3)Simulate 1000 samples from the Gaussian copula with Σ above. Plot the two simulated samples and compare. Try choosing different marginal distributions for these two copulas and see what happens.

Exercise 4.45:
In this exercise we consider the problem of simulating from the Gumbel copula.

1)Read the part of subsection B.3 in the appendix about stable distributions.

2)Let $\theta > 1$. Show that the moment-generating function of the $S(1/\theta, 1, \gamma, 0)$ distribution for

$$\gamma = \cos\left(\frac{\pi}{2\theta}\right)^{\theta}$$

is $\exp(-t^{1/\theta})$, the inverse of the Gumbel generator. Hint: Use the relation between the characteristic function and the moment-generating functon. It may be useful to recall the formulas

$$\cos t = \frac{e^{it} + e^{-it}}{2}, \quad \sin t = \frac{e^{it} - e^{-it}}{2i}.$$

3)Describe an explicit algorithm to simulate samples from the d-dimensional Gumbel copula.

4)Implement the algorithm, you derived in the previous subproblem. Simulate some samples for different values of θ.

Chapter 5

Dependence concepts and fitting copulas to data

5.1 Introduction

This chapter is about measures of dependence between random variables. Such measures can take many forms, but a desirable property for such measures is that they should be invariant under strictly increasing transformations. This means that copulas are a vital component in the computation of such measures, and we shall indeed see that most of the results in this section are framed in terms of the copula of the variables involved.

The motivation for studying dependence concepts is clear. Understanding how one risk factor behaves in comparison to another is vital in financial modelling, and the ability to compress such information into a single number is efficient both from a statistical perspective, but also when it comes to communicating risk in layman's terms. The first sections of this chapter study different dependence concepts with an emphasis on Kendall's τ, Spearman's ρ and tail dependence. We establish the most important mathematical properties of these measures of dependence with a strong emphasis on how they are computed in different contexts, in particular for interesting special cases such as Archimedean copulas and implicit copulas like the normal and t distributions. After covering these dependence concepts, we explain how they may be used to choose a certain copula for real data and how to do estimation in practice. The chapter ends with supplementary material containing additional proofs.

5.2 Kendall's τ

Definition and basic properties

Definition 5.2.1. *Kendall's* τ for (X_1, X_2) is defined as

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

where (Y_1, Y_2) is independent of (X_1, X_2) with $(Y_1, Y_2) \overset{\mathrm{d}}{=} (X_1, X_2)$.

Kendall's τ gives an indication of whether X_1 and X_2 get large together or if one tends to get larger when the other gets smaller. If X_1 and X_2 tend to move together, the event

$\{(X_1 - Y_1)(X_2 - Y_2) > 0\}$ has high probability, resulting in a value of $\rho_\tau(X_1, X_2)$ close to 1, and if X_1 and X_2 move in opposite directions, $\{(X_1 - Y_1)(X_2 - Y_2) < 0\}$ will have high probability so that $\rho_\tau(X_1, X_2)$ is close to -1. To make this precise, we can make the following definition.

Definition 5.2.2. Consider two points $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$. We say that (x_1, x_2) and (y_1, y_2) are *concordant* if $(x_1 - y_1)(x_2 - y_2) > 0$ and *discordant* if $(x_1 - y_1)(x_2 - y_2) < 0$.

Writing out the definition, we see that (x_1, x_2) and (y_1, y_2) are concordant if $x_1 > y_1$ and $x_2 > y_2$ or $X_1 < y_1$ and $x_2 < y_2$. In other words, if and only if the line connecting the two points has positive slope. The situation is of course reversed for a discordant pair.

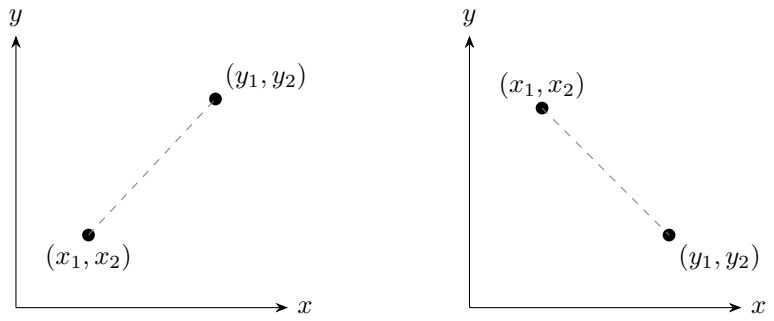


Figure 1: An illustration of the concepts of concordance (left) and discordance (right) for a pair of points (x_1, x_2) and (y_1, y_2).

Intuitively, $\rho_\tau(X_1, X_2) = 1$ should correspond to comonotonicity while $\rho_\tau(X_1, X_2) = -1$ should correspond to countermonotonicity. This is indeed the case as you are asked to show in Exercise 5.4. We have the following results on Kendall's τ for copulas.

Theorem 5.2.3. *The following hold:*

(i) Let (X_1, X_2) and (Y_1, Y_2) be two independent pairs of continuous variables with copulas C_1 and C_2, respectively. Assume that the marginal distribution of the two pairs are the same. Then the difference in the concordance and discordance probability of the pairs is given by

$$P((X_1-Y_1)(X_2-Y_2) > 0) - P((X_1-Y_1)(X_2-Y_2) < 0) = 4 \int_{[0,1]^2} C_2(u_1, u_2) dC_1(u_1, u_2).$$

In particular, we have the formula for Kendall's τ given by

$$\rho_\tau(X_1, X_2) = 4 \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1$$

where C is the copula of (X_1, X_2).

(ii) Let (X_1, X_2) be random variables with Archimedean copula with generator φ. Then

$$\rho_\tau(X_1, X_2) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

Proof. To prove (i), let F_1 and F_2 denote the distribution functions of X_1, Y_1 and X_2, Y_2, respectively. By continuity, we have that the difference in the two probabilities equals

$$2P((X_1 - Y_1)(X_2 - Y_2) > 0) - 1.$$

It thus remains to compute $P((X_1 - Y_1)(X_2 - Y_2) > 0)$. But we have

$$P((X_1 - Y_1)(X_2 - Y_2) > 0) = P(X_1 > Y_1, X_2 > Y_2) + P(X_1 < Y_1, X_2 < Y_2),$$

and so we compute each of these two probabilities. As for $P(X_1 > Y_1, X_2 > Y_2)$, we have by Sklar's Theorem that the joint distribution function of (X_1, X_2) is given by $C_1(F_1, F_2)$. Hence by independence and another application of Sklar's Theorem,

$$\begin{aligned} P(X_1 > Y_1, X_2 > Y_2) &= \int_{\mathbb{R}^2} P(Y_1 \leq y_1, Y_2 \leq y_2) dC_1(F_1(y_1), F_2(y_2)) \\ &= \int_{\mathbb{R}^2} C_2(F_1(y_1), F_2(y_2)) dC_1(F_1(y_1), F_2(y_2)) \\ &= \int_{[0,1]^2} C_2(u_1, u_2) dC_1(u_1, u_2), \end{aligned}$$

where the last step follows by the substitution $u_1 = F_1(y_1), u_2 = F_2(y_2)$. The probability $P(X_1 < Y_1, X_2 < Y_2)$ is calculated similarly as

$$\begin{aligned} P(X_1 < Y_1, X_2 < Y_2) &= \int_{\mathbb{R}^2} P(Y_1 > y_1, Y_2 > y_2) dC_1(F_1(y_1), F_2(y_2)) \\ &= \int_{\mathbb{R}^2} (1 - F_1(y_1) - F_2(y_2) + C_2(F_1(y_1), F_2(y_2))) dC_1(F_1(y_1), F_2(y_2)) \\ &= \int_{[0,1]^2} (1 - u_1 - u_2 + C_2(u_1, u_2)) dC_1(u_1, u_2) \\ &= 1 - \frac{1}{2} - \frac{1}{2} + \int_{[0,1]^2} C_2(u_1, u_2) dC_1(u_1, u_2) \\ &= \int_{[0,1]^2} C_2(u_1, u_2) dC_1(u_1, u_2), \end{aligned}$$

where the second to last equality follows because the marginal distributions of C_1 are uniform on $[0, 1]$ and thus have expectation $1/2$ each. Hence

$$P((X_1 - Y_1)(X_2 - Y_2) > 0) = 2 \int_{[0,1]^2} C_2(u_1, u_2) dC_1(u_1, u_2).$$

Substituting back into the first equation of the argument finishes the proof of (i). The proof of (ii) relies on certain geometric properties of Archimedean copulas. All details may be found in the supplementary section. ■

We leave it as an exercise for the reader to verify that if (X_1, X_2) has the Clayton copula, $\rho_\tau(X_1, X_2) = \theta/(\theta + 2)$ while for the Gumbel copula, $\rho_\tau(X_1, X_2) = 1 - 1/\theta$. Other examples for Archimedean copulas are outlined in the exercises. The following result provides a general formula for Kendall's τ for an elliptical distribution in terms of its ordinary linear correlation, the only requirement being no point mass at the origin.

Theorem 5.2.4. *Let $(X_1, X_2) \sim E_2(0, \Sigma, \psi)$ where Σ is a correlation matrix,*

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and assume $P((X_1, X_2) = (0, 0)) = 0$. Then

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \rho.$$

The proof builds on the following lemma, which is interesting in itself.

Lemma 5.2.5. *Let $(X_1, X_2) \sim E_2(0, \Sigma, \psi)$ be as in Theorem 5.2.4. Then*

$$P(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}.$$

Proof. We claim that

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} (Z_1, \rho Z_1 + \sqrt{1 - \rho^2} Z_2)$$

for $(Z_1, Z_2) \sim S_2(\psi)$, a spherical pair of variables with generator ψ. To see why, simply compute the characteristic function of the right hand side as follows:

$$\begin{aligned} \Psi_{(Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2)}(\mathbf{t}) &= E\Big[e^{t_1 Z_1 + t_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2)}\Big] = E\Big[e^{(t_1 + \rho t_2) Z_1 + t_2 \sqrt{1-\rho^2} Z_2}\Big] \\ &= \psi((t_1 + \rho t_2)^2 + t_2^{1-\rho^2)} = \psi(t_1^2 + t_2^2 + 2\rho t_1 t_2) \\ &= \psi\Big(\mathbf{t}^T \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \mathbf{t}\Big) = \Psi_{(X_1, X_2)}(\mathbf{t}). \end{aligned}$$

Since the characteristic functions coincide, the claim follows. Recall that we may write $(Z_1, Z_2) = R\mathbf{S}$ where $\mathbf{S}$ is uniformly distributed on the unit circle. Hence we can write $\mathbf{S} = (\cos \Theta, \sin \Theta)$ for $\Theta \sim \mathrm{Unif}[-\pi, \pi)$. Thus,

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} R(\cos \Theta, \rho \cos \Theta + \sqrt{1 - \rho^2} \sin \Theta).$$

Let $\phi = \arcsin \rho$. Then $\sin \phi = \rho$ and $\cos \phi = \sqrt{1 - \rho^2}$. Since $P(R = 0) = P((X_1, X_2) = (0, 0)) = 0$, we have (recall that $\phi \in [-\pi/2, \pi/2]$)

$$\begin{aligned} P(X_1 > 0, X_2 > 0) &= P(\cos \Theta > 0, \sin \phi \cos \Theta + \cos \phi \sin \Theta > 0) \\ &= P(\cos \Theta > 0, \sin(\Theta + \phi) > 0) \\ &= P(\Theta \in (-\pi/2, \pi/2), \Theta + \phi \in (0, \pi)) \\ &= P(\Theta \in (-\phi, \pi/2)) = \frac{1}{2\pi}\Big(\frac{\pi}{2} + \phi\Big) \\ &= \frac{1}{4} + \frac{\arcsin \rho}{2\pi}. \end{aligned}$$

■

Proof of Theorem 5.2.4. The proof relies on the convolution property of Elliptical distributions, see Theorem 3.2.4 (iii). Let $\mathbf{X} = (X_1, X_2)$ and let $\mathbf{Y} = (Y_1, Y_2)$. We know that $\mathbf{Z} := \mathbf{X} - \mathbf{Y}$ is again elliptical. We have

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 2P((X_1 - Y_1)(X_2 - Y_2) > 0) - 1 = 2P(Z_1 Z_2 > 0) - 1 \\ &= 2(P(Z_1 > 0, Z_2 > 0) + P(Z_1 < 0, Z_2 < 0)) - 1 = 4P(Z_1 > 0, Z_2 > 0) - 1. \end{aligned}$$

Now plug in the result from the lemma and conclude. ■

Estimation

We want to be able to estimate Kendall's τ from data $\{(X_{t,1}, X_{t,2}) : t = 1, ..., n\}$. To do so, we need to compare all pairs $(X_{t,1}, X_{t,2})$ and $(X_{s,1}, X_{s,2})$ with one another. In the case of concordance, the pair should yield the value 1 (indicating a positive sign) and -1 in the case of discordance. Since there are $\binom{n}{2}$ pairs in total, we obtain the estimator

$$\hat{\rho}_\tau(X_1, X_2) = \binom{n}{2}^{-1} \sum_{1 \le t < s \le n} \text{sign}((X_{t,1} - X_{s,1})(X_{t,2} - X_{s,2}))$$

where

$$\text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}.$$

If (X_1, X_2) has an elliptical distribution, we can estimate the correlation as usual by

$$\hat{\rho}(X_1, X_2) = \frac{\sum_{t=1}^n (X_{t,1} - \bar{X}_1)(X_{t,2} - \bar{X}_2)}{\sqrt{\sum_{t=1}^n (X_{t,1} - \bar{X}_1)^2} \sqrt{\sum_{t=1}^n (X_{t,2} - \bar{X}_2)^2}}$$

and then apply Theorem 5.2.4 to obtain the natural estimator

$$\hat{\rho}_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \hat{\rho}(X_1, X_2).$$

5.3 Spearman's ρ

Definition and basic properties

Another measure of rank correlation is Spearman's ρ defined as follows.

Definition 5.3.1. Consider a pair of variables (X_1, X_2) and introduce independent copies $(Y_1, Y_2) \overset{\text{d}}{=} (Z_1, Z_2) \overset{\text{d}}{=} (X_1, X_2)$. *Spearman's* ρ of (X_1, X_2) is given by

$$\rho_S(X_1, X_2) = 3(P((X_1 - Y_1)(X_2 - Z_2) > 0) - P((X_1 - Y_1)(X_2 - Z_2) < 0)).$$

The intuition for Spearman's ρ is the same as for Kendall's τ. Spearman's ρ is proportional to the difference in the probability of concordance and the probability of discordance between the pairs (X_1, X_2) and (Y_1, Z_2). In other words, this means that Spearman's ρ of (X_1, X_2) is given by 3 times the difference in the probability of concordance and discordance of the pairs (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$, where the latter pair has the independence copula Π but the same marginal distributions as (X_1, X_2). We will use this observation several times in the following. Why is there a "3" in the definition? We can provide an answer after presenting the following theorem.

Theorem 5.3.2. *Let X_1 and X_2 be continuous variables with copula C. Then*

$$\rho_S(X_1, X_2) = 12 \int_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 3 = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3.$$

Proof. Let the pair $(\tilde{X}_1, \tilde{X}_2)$ have the same marginal distributions as (X_1, X_2) but copula Π. Using Theorem 5.2.3 (i), the difference in the concordance and discordance probability equals

$$4 \int_{[0,1]^2} \Pi(u_1, u_2) dC(u_1, u_2) - 1 = 4 \int_{[0,1]^2} u_1 u_2 dC_1(u_1, u_2) - 1.$$

Multiplying by 3 and using the observation in the paragraph above completes the proof of the first equality. The second follows immediately by symmetry. ■

Corollary 5.3.3. *For a pair (X_1, X_2) of continuous random variables with copula C, we have*

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2.$$

We can now provide an answer to why we multiply by 3 in the definition of Spearman's ρ. Consider the integral

$$\int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2$$

which appears in the theorem just presented. Using the Fréchet bounds, we have that a lower bound for the integral is

$$\begin{aligned} \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 &\geq \int_0^1 \int_0^1 \max\{u_1 + u_2 - 1, 0\} du_1 du_2 \\ &= \int_0^1 \int_{1-u_2}^1 u_1 + u_2 - 1 du_1 du_2 = \frac{1}{6}, \end{aligned}$$

which means that $\rho_S \geq -1$. Similarly, by plugging in the upper Fréchet bound,

$$\int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 \leq \int_0^1 \int_0^1 \min\{u_1, u_2\} du_1 du_2 = \frac{1}{3}$$

so that $\rho_S \leq 1$. This means that with the multiplication by 3, ρ_S becomes a proper correlation coefficient. This is also clear from the following result, which states that Spearman's ρ for (X_1, X_2) equals the classical correlation coefficent for their copula (at least when X_1 and X_2 are continuous).

Proposition 5.3.4. *Let (X_1, X_2) be a pair of continuous random variables with marginal distribution functions F_1 and F_2. Then*

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)).$$

Proof. Define $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$. From continuity and Sklar's Theorem, (U_1, U_2) has distribution function C, the copula of (X_1, X_2). Using Theorem 5.3.2, we see that $\rho_S(X_1, X_2) = 12E[U_1 U_2] - 3$. But $E[U_1] = E[U_2] = 1/2$ since U_1, U_2 are standard uniform, and so

$$\rho_S(X_1, X_2) = 12 \operatorname{Cov}(U_1, U_2) = \frac{\operatorname{Cov}(U_1, U_2)}{\sqrt{\operatorname{Var}[U_1] \operatorname{Var}[U_2]}} = \rho(U_1, U_2),$$

where we recall that $\operatorname{Var}[U_1] = \operatorname{Var}[U_2] = 1/12$. ■

The following formula, which is an immediate consequence of the result just presented, is often useful for computations.

Corollary 5.3.5. *Let (X_1, X_2) be a pair of continuous random variables with marginal distribution functions F_1 and F_2. If $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$, then*

$$\rho_S(X_1, X_2) = E[C(U_1, U_2)] - 3.$$

Unfortunately, we do not have a nice general formula for Spearman's ρ in terms of the ordinary linear correlation ρ which applies to any elliptical distribution. But for the special case of the Gaussian copula, a closed formula exists.

Proposition 5.3.6. *Let $\mathbf{X} = (X_1, X_2)$ have the Gaussian copula C_Σ^{Ga} and continuous marginal distributions. Then*

$$\rho_S(X_1, X_2) = \frac{6}{\pi} \arcsin \frac{\rho(X_1, X_2)}{2}.$$

Proof. Let (Y_1, Y_2) be a vector of independent standard Gaussian variables independent of (X_1, X_2). We then compute

$$\begin{aligned}
\rho_S(X_1, X_2) &= 3(2P((X_1 - Y_1)(X_2 - Y_2) > 0) - 1) \\
&= 3(4P(X_1 - Y_1 > 0, X_2 - Y_2 > 0) - 1) \\
&= 3(4P(Z_1 > 0, Z_2 > 0) - 1)
\end{aligned}$$

where $\mathbf{Z} = \mathbf{X} - \mathbf{Y} \sim \mathcal{N}(0, \Sigma + I_2)$. Now use the fact that $\rho(Z_1, Z_2) = \rho/2$ and Lemma 5.2.5 to conclude

$$\rho_S(X_1, X_2) = 3\Big(4\Big(\frac{1}{4} + \frac{\arcsin \rho/2}{2\pi}\Big) - 1\Big) = \frac{6}{\pi} \arcsin \frac{\rho}{2}.$$

■

The figure below illustrates the three discussed dependence concepts for the Gaussian copula.

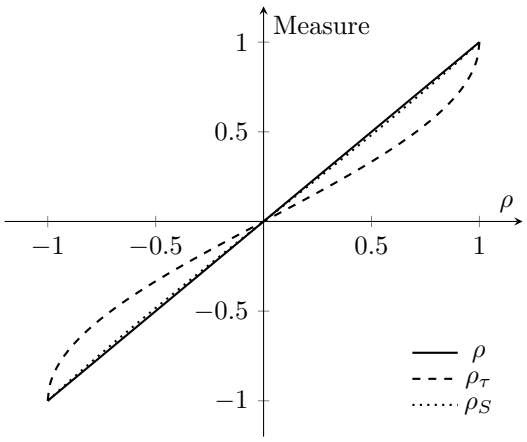


Figure 2: An illustration of the linear correlation ρ, Kendall's τ, ρ_τ, and Spearman's ρ, ρ_S, for the Gaussian copula.

From the figure, one can see that ρ and ρ_S are very close, and one may thus approximate ρ_S by ρ for the Gaussian copula. One way to see this mathematically is to do a Taylor expansion of arcsin, which yields

$$\rho_S(X_1, X_2) = \frac{6}{\pi}\left(\frac{\rho}{2} + \frac{(\rho/2)^3}{6} + O(\rho^5)\right) = \frac{3}{\pi}\rho + \frac{\rho^3}{8\pi} + O(\rho^5).$$

While no general closed formula for Spearman's ρ in terms of the linear correlation ρ exists, we do have a useful result for normal variance mixture models.

Theorem 5.3.7. *Let $\mathbf{X} = (X_1, X_2)$ follow a normal variance mixture model with mean zero of the form*

$$\mathbf{X} = \sqrt{W}\mathbf{Z}$$

where $W \geq 0$ is a univariate random variable and $\mathbf{Z} \sim \mathcal{N}(0, I_2)$. Write $\rho = \rho(X_1, X_2)$ for the linear correlation and assume $P(\mathbf{X} = 0) = 0$. Then Spearman's ρ can be computed according to

$$\rho_S(X_1, X_2) = \frac{6}{\pi} E\left[\arcsin\left(\frac{\rho W}{\sqrt{(W + \tilde{W})(W + \bar{W})}}\right)\right]$$

where $\tilde{W}$ and $\bar{W}$ are independent copies of W.

Proof. Let $\tilde{Z}$ and $\bar{Z}$ be standard Gaussian variables and assume $\mathbf{Z}, \tilde{Z}, \bar{Z}, W, \tilde{W}$ and $\bar{W}$ are all independent. Define $\tilde{X} := \sqrt{\tilde{W}}\tilde{Z}$ and $\bar{X} := \sqrt{\bar{W}}\bar{Z}$ and

$$Y_1 := X_1 - \tilde{X}, \quad Y_2 := X_2 - \bar{X}.$$

Then the vector $\mathbf{Y} = (Y_1, Y_2)$ is normal distributed conditional on $W, \tilde{W}$ and $\bar{W}$. Specifically,

$$\mathbf{Y} \mid W, \tilde{W}, \bar{W} \sim \mathcal{N}\left(0, \begin{pmatrix} W + \tilde{W} & \rho W \\ \rho W & W + \bar{W} \end{pmatrix}\right).$$

We may now compute using the tower property

$$\begin{aligned}
\rho_S(X_1, X_2) &= 3(2P((X_1 - \bar{X})(X_2 - \tilde{X}) > 0) - 1) \\
&= 3(2E[P(Y_1 Y_2 > 0 \mid W, \tilde{W}, \bar{W})] - 1) \\
&= 3E[4P(Y_1 > 0, Y_2 > 0 \mid W, \tilde{W}, \bar{W}) - 1].
\end{aligned}$$

Now apply Lemma 5.2.5 and use that conditionally on $W, \tilde{W}, \bar{W}$, we have

$$\rho(Y_1, Y_2) = \frac{\rho W}{\sqrt{(W + \tilde{W})(W + \bar{W})}}$$

to obtain

$$\begin{aligned}
\rho_S(X_1, X_2) &= 3E\left[4\left(\frac{1}{4} + \frac{\arcsin \rho(Y_1, Y_2)}{2\pi}\right) - 1\right] \\
&= \frac{6}{\pi} E\left[\arcsin\left(\frac{\rho W}{\sqrt{(W + \tilde{W})(W + \bar{W})}}\right)\right].
\end{aligned}$$

■

Estimation and computational methods

Proposition 5.3.4 suggests a method for estimating Spearman's ρ when (X_1, X_2) is continuous. Assume we are given data $\{(X_{t,1}, X_{t,2}) : t = 1, ..., n\}$ and let $F_n^{(1)}$ and $F_n^{(2)}$ denote the empirical distribution functions based on the marginal samples $\{X_{t,1}\}$ and $\{X_{t,2}\}$, respectively. We can then consider the transformed variables $U_{i,j} := F_n^{(j)}(X_{i,j})$ for $j = 1, 2$ and $i = 1, ..., n$ and construct the estimator

$$\hat{\rho}_S(X_1, X_2) = \frac{\frac{1}{n}\sum_{i=1}^n U_{i,1}U_{i,2} - \overline{U}_1\overline{U}_2}{\sqrt{\frac{1}{n}\sum_{i=1}^n (U_{i,1} - \overline{U}_1)^2 \frac{1}{n}\sum_{i=1}^n (U_{i,2} - \overline{U}_2)^2}}$$

where

$$\overline{U}_1 := \frac{1}{n}\sum_{i=1}^n U_{i,1}, \quad \overline{U}_2 := \frac{1}{n}\sum_{i=1}^n U_{i,2}$$

are the empirical means of the transformed variables. The quantity $R_{i,j} := nF_n^{(j)}(X_{i,j}) = nU_{i,j}$ is called the *rank* of $X_{i,j}$ in the sample $X_{1,j}, ..., X_{n,j}$. With this definition, letting

$$\overline{R}_j := \frac{1}{n}\sum_{i=1}^n R_{i,j}, \quad \sigma_j^2 := \frac{1}{n}\sum_{i=1}^n (R_{i,j} - \overline{R}_j)^2,$$

we may rewrite the estimator above to

$$\hat{\rho}_S(X_1, X_2) = \frac{\frac{1}{n}\sum_{i=1}^n R_{i,1}R_{i,2} - \overline{R}_1\overline{R}_2}{\sigma_1\sigma_2}. \tag{5.1}$$

With a bit more work (see Exercise 5.16), one can show the following.

Proposition 5.3.8. *The estimator for Spearman's ρ given by (5.1) can in the case of no ties in the two samples $\{X_{t,1}\}$ and $\{X_{t,2}\}$ be written as*

$$\hat{\rho}_S(X_1, X_2) = 1 - \frac{6\sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

where $d_i = R_{i,1} - R_{i,2}$ is the difference in rank for observation i in the two samples.

This estimator is very popular and can be found in several places.

In the case where we have chosen a normal variance mixture model, we may compute ρ_S using Monte Carlo methods via Theorem 5.3.7. Indeed, choose a high number N and simulate independent iid sequences $\{W_i\}_{i=1}^N, \{\tilde{W}_i\}_{i=1}^N$ and $\{\bar{W}_i\}_{i=1}^N$, all distributed according to W. The Monte Carlo estimate for Spearman's ρ is then

$$\rho_S(X_1, X_2) \approx \frac{6}{\pi N}\sum_{i=1}^N \arcsin\left(\frac{\rho W_i}{\sqrt{(W_i + \tilde{W}_i)(W_i + \bar{W}_i)}}\right).$$

Comparison of Kendall's τ and Spearman's ρ

We now present some results on the relationship between Kendall's τ and Spearman's ρ. The proofs can be found in a supplementary section below. For the background on these results, consult the notes and comments at the end of the chapter.

Theorem 5.3.9. *Let (X_1, X_2) be a pair of continuous random variables. Then*

$$-1 \leq 3\rho_\tau(X_1, X_2) - 2\rho_S(X_1, X_2) \leq 1.$$

The next theorem provides a different set of inequalities.

Theorem 5.3.10. *Let (X_1, X_2) be a pair of continuous random variables. Then*

$$\frac{1 + \rho_S(X_1, X_2)}{2} \geq \left(\frac{1 + \rho_\tau(X_1, X_2)}{2}\right)^2$$

and

$$\frac{1 - \rho_S(X_1, X_2)}{2} \geq \left(\frac{1 - \rho_\tau(X_1, X_2)}{2}\right)^2.$$

We may now combine the inequalities in the two preceding theorems to obtain the following result.

Corollary 5.3.11. *Let (X_1, X_2) be a pair of continuous random variables. Then*

$$\frac{3\rho_\tau(X_1, X_2) - 1}{2} \leq \rho_S(X_1, X_2) \leq \frac{1 + 2\rho_\tau(X_1, X_2) - \rho_\tau(X_1, X_2)^2}{2}, \quad \rho_\tau(X_1, X_2) \geq 0$$

and

$$\frac{\rho_\tau(X_1, X_2)^2 + 2\rho_\tau(X_1, X_2) - 1}{2} \leq \rho_S(X_1, X_2) \leq \frac{1 + 3\rho_\tau(X_1, X_2)}{2}, \quad \rho_\tau(X_1, X_2) \leq 0.$$

The following figure illustrates these bounds.

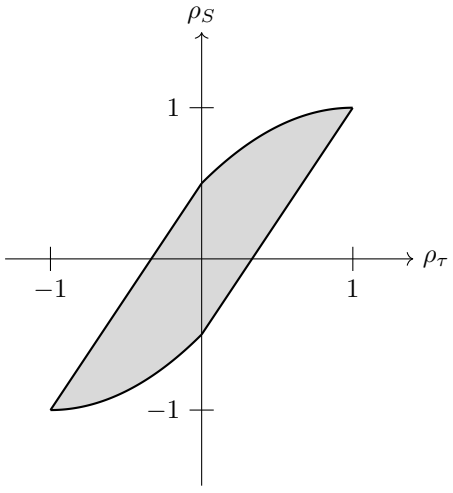


Figure 3: The shaded area contains the possible values of (ρ_τ, ρ_S) for a pair of continuous random variables. The exact bounds are presented in Corollary 5.3.11.

In the exercises below, we investigate how tight these bounds are by considering some concrete copulas.

5.4 Tail dependence

Definition 5.4.1. The *coefficient of upper tail dependence* of X_1 and X_2 is given by

$$\lambda_U(X_1, X_2) = \lim_{u \uparrow 1} P(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u))$$

where F_i denotes the distribution function of X_i, $i = 1, 2$.

λ_U essentially measures how X_2 behaves when X_1 gets large. We leave it as an exercise for the reader to verify that in the extreme case where X_1 and X_2 are comonotone, then $\lambda_U(X_1, X_2) = 1$. To compute the coefficient of upper tail dependence, the following result is often useful.

Proposition 5.4.2. *Assume X_1 and X_2 have continuous distribution functions F_1 and F_2 and unique copula C. Then*

$$\lambda_U(X_1, X_2) = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}.$$

Proof. Since F_1 and F_2 are continuous, we have

$$\lambda_U(X_1, X_2) = \lim_{u \uparrow 1} P(F_2(X_2) > u \mid F_1(X_1) > u) = \lim_{u \uparrow 1} P(U_2 > u \mid U_1 > u).$$

Note that we have the following:

$$P(U_1 > u) + P(U_2 > u) + P(U_1 \leq u, U_2 \leq u) = 1 + P(U_1 > u, U_2 > u).$$

To see this, it may help to draw a figure. The equation can be rewritten as

$$(1 - u) + (1 - u) + C(u, u) = 1 + P(U_1 > u, U_2 > u)$$

so that

$$P(U_2 > u \mid U_1 > u) = \frac{P(U_1 > u, U_2 > u)}{P(U_1 > u)} = \frac{1 - 2u + C(u, u)}{1 - u}$$

from which the claim follows. ■

In the exercises below and in the mandatory assignments, we will see examples of computing the coefficient of upper tail dependence. We now introduce the corresponding notion for the lower tail.

Definition 5.4.3. The *coefficient of lower tail dependence* of X_1 and X_2 is given by

$$\lambda_L(X_1, X_2) = \lim_{u \downarrow 0} P(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)).$$

We have an analogous result to the one above for continuous marginal distribution functions.

Proposition 5.4.4. *Assume X_1 and X_2 have continuous distribution functions F_1 and F_2 and unique copula C. Then*

$$\lambda_L(X_1, X_2) = \lim_{u \downarrow 0} \frac{C(u, u)}{u}.$$

Proof. The proof is essentially a simpler version of the one given for the coefficient of upper tail dependence. We compute

$$\begin{aligned}\lambda_L(X_1,X_2) &= \lim_{u\downarrow 0} P(X_2 \le F_2^{\leftarrow}(u) \mid X_1 \le F_1^{\leftarrow}(u)) = \lim_{u\downarrow 0} P(F_2(X_2) \le u \mid F_1(X_1) \le u) \\ &= \lim_{u\downarrow 0} P(U_2 \le u \mid U_1 \le u) = \lim_{u\downarrow 0} \frac{P(U_2 \le u, U_1 \le u)}{P(U_1 \le u)} = \lim_{u\downarrow 0} \frac{C(u,u)}{u}\end{aligned}$$

as desired. ■

Example 5.4.5. Let $(X_1, X_2) \sim \mathcal{N}(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Assume $\rho \in (-1, 1)$. We want to show that the coefficient of upper tail dependence $\lambda_U(X_1, X_2)$ is zero. Consider a $t > 1$ such that $t\rho < 1$ and make the bound

$$\begin{aligned}P(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)) \le{}& P(X_2 > F_2^{\leftarrow}(u), X_1 \le tF_1^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)) \\ &+ P(X_1 > tF_1^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)).\end{aligned}$$

Consider the second term first and let $v := F_1^{\leftarrow}(u)$. We have by L'Hospital's rule that

$$\begin{aligned}\lim_{u\uparrow 1} P(X_1 > tF_1^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)) &= \lim_{v\to\infty} \frac{P(X_1 > tv)}{P(X_1 > v)} = \lim_{v\to\infty} \frac{\int_{tv}^{\infty} e^{-x^2/2} dx}{\int_v^{\infty} e^{-x^2/2} dx} \\ &= \lim_{v\to\infty} \frac{e^{-(tv)^2/2}}{e^{-v^2/2}} = 0.\end{aligned}$$

We now consider the limit of the first term i.e.

$$\lim_{u\uparrow 1} P(X_2 > F_2^{\leftarrow}(u), X_1 \le tF_1^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)) = \lim_{v\to\infty} \frac{\int_v^{tv} P(X_2 > v \mid X_1 = x) d\Phi(x)}{\int_v^{\infty} d\Phi(x)}.$$

We now use that $X_2 \mid X_1 = x \sim \mathcal{N}(\rho x, 1 - \rho^2)$. Letting $Y \sim \mathcal{N}(\rho x, 1 - \rho^2)$ and setting $Z = (Y - \rho x)/\sqrt{1 - \rho^2} \sim \mathcal{N}(0, 1)$, we can rewrite

$$P(X_2 > v \mid X_1 = x) = P\left(Z > \frac{v - \rho x}{\sqrt{1 - \rho^2}}\right).$$

Consider $x \in [v, tv]$. If $x = v$, then $v - \rho x = v - \rho v > 0$. Similarly, if $x = tv$, we have (due to the assumption $t\rho < 1$) that $v - \rho x = v - t\rho v > 0$. We conclude that

$$\lim_{v\to\infty} P\left(Z > \frac{v - \rho x}{\sqrt{1 - \rho^2}}\right) = 0$$

uniformly for $x \in [v, tv]$. It follows that the limit of the first term goes to zero as well which establishes $\lambda_U(X_1, X_2) = 0$. We refer to this property as *asymptotic independence* in the tails. No matter how large the correlation ρ is in the range $(-1, 1)$, if we go far enough into the tails, extreme events occur independently in X_1 and X_2. This illustrates a potential problem with the Gaussian copula for modelling financial data. In such data, large losses in different variables are often correlated, and the normal copula may fail to capture this. ∘

5.5 Fitting copulas to data

In this section, we discuss different approaches to fitting a copula to data. We start by discussing some of the considerations one should make in choosing a particular copula, and we then move on to the more technical aspect of fitting a particular copula using maximum likelihood.

The problem

Let us set the stage. We are given continuous multivariate iid observations $\mathbf{X}_t = (X_{t,1}, ..., X_{t,d})$ for $t = 1, ..., n$, and the goal is to find a suitable model for the distribution of $\mathbf{X}$ (we use $\mathbf{X} = (X_1, ..., X_d)$ to denote a generic observation with the distribution of one $\mathbf{X}_t$). This is a very difficult task, since the complexity in choosing the marginal distributions and the possible dependence structure explodes as the dimension d increases. Nevertheless, we have developed copulas as a tool to make this easier by separating the modelling of the dependence structure from the marginals. The general approach to multivariate modelling in this chapter can be outlined as follows:

1. Determine the marginal distributions of $\mathbf{X}$, either through a parametric model, nonparametrically or a mix of the two (a *semiparametric* aproach). Call the fitted distribution functions $F_{1,n}, ..., F_{d,n}$.
2. Let $\mathbf{U}_t = (U_{t,1}, ..., U_{t,d}) := (F_{1,n}(X_{t,1}), ..., F_{d,n}(X_{t,d}))$ for $t = 1, ..., d$. We refer to $\mathbf{U}_t$ as a *pseudo-sample*.
3. Choose a copula C by some method. While it is possible to proceed completely nonparametrically, we choose to let C depend on a parameter θ and determine $\hat{\theta}$ by either maximum likelihood or calibration based on dependence concepts.

Points 1 and 3 deserve elaboration.

Generating the pseudo-sample - choosing the marginal distributions

For modelling the marginal distributions, several choices are available. In line with [88], we highlight the following three options.

(i) **Parametric models**: Based on some criterion such as the value of the log-likelihood, AIC or something entirely different, we choose parametric models for each X_i, $i = 1, ..., d$. The distribution functions are then estimated via maximum likelihood, a moment estimator or by other means. In a financial context, we may choose e.g. Student's t distribution or a Pareto distribution for heavy-tailed quantities or a gamma, truncated normal etc. for light-tailed losses.

(ii) **Nonparametric estimation by (a variant of) the empirical distribution function**: We could estimate the marginal distribution functions $F_1, ..., F_d$ purely nonparametrically by

$$F_{j,n}(x) = \frac{1}{n+1} \sum_{t=1}^{n} 1_{\{X_{t,j} \leq x\}}, \quad j = 1, ..., d$$

The reason for choosing $n + 1$ instead of n in the denominator is to avoid boundary issues. The density of a copula is often infinite at the boundary, and we avoid this

problem entirely by having $U_{t,j} < 1$ for all t and j. Of course, this makes little to no difference asymptotically.

(iii) **Semiparametric tail estimation**: As discussed in Chapter 2, the empirical tail often fails to capture the tail behaviour properly. An alternative to (ii) is therefore to choose a sufficiently high threshold u (as in the POT method) and letting $F_j(x)$ for $x < u$ be modelled by the empirical distribution function, while $F_j(x)$ for $x \geq u$ is modelled using a Generalised Pareto Distribution.

Choosing a particular copula

To determine a proper copula for a data set, we should think of properties like correlation, tail dependence and symmetry. The tail behaviour is especially relevant in a risk management context since we want to model large losses. Large losses also tend to move together. When one loss is large, the other losses tend to be large as well. This can for example happen in a portfolio with stocks in similar companies. A concrete approach to choosing a copula is therefore to first consider the problem from a top-down perspective. For example, if we are interested in modelling large financial losses via e.g. Value at Risk, we should choose a copula which does not exhibit asymptotic independence in the tails, so that we don't get too optimistic estimates for the loss. A simple model-driven choice is to fit several copulas to the pseudo-samples with maximum likelihood and choose the one with the highest log-likelihood.

Maximum likelihood and calibration

If a copula C has been chosen for the data, one can apply maximum likelihood methods in the case where C depends on a parameter θ. If c_θ denotes the density (assuming this exists) of C, then the maximum likelihood estimator $\hat{\theta}$ would maximise the log-likelihood

$$\log L(\theta; \mathbf{U}_1, ..., \mathbf{U}_n) = \sum_{i=1}^{n} \log c_\theta(\mathbf{U}_i)$$

for the pseudo-sample $\mathbf{U}_1, ..., \mathbf{U}_n$. An entirely different approach would be to compute a nonparametric estimate for e.g. Kendall's τ or Spearman's ρ, when these are given as explicit invertible functions of θ, and then apply the inverse to get an estimator for θ. We refer to this method as *calibration*. We start by discussing the Gaussian and t copulas in detail, since these are widely used. We follow the presentation in [88].

Example 5.5.1 (**Fitting the Gaussian copula**)**.** Say $(X_1, ..., X_d)$ has the Gaussian copula with correlation matrix P. Recall from earlier (Proposition 5.3.6) that for the Gaussian copula, we have the very precise approximation

$$\rho_S(X_i, X_j) = \frac{6}{\pi} \arcsin \frac{\rho(X_i, X_j)}{2} \approx \rho(X_i, X_j).$$

There are two advantages to this approximation, the first being simplicity. Simply compute an estimate for Spearman's ρ for each pair of variables in $\mathbf{X}$ to estimate the correlation matrix P in C_P^{Ga}. The second advantage is that the resulting estimated correlation matrix

is actually (very close to) an actual correlation matrix. If we choose not to rely on the approximation and instead estimate $\rho(X_i, X_j)$ via

$$\hat{\rho}(X_i, X_j) = \sin \frac{\pi \hat{\rho}_S(X_i, X_j)}{6},$$

we run into the potential problem that the resulting estimator may not be a valid correlation matrix (it may not be positive definite). This problem can be addressed by the *eigenvalue method*, which we present shortly.

Let us turn to maximum likelihood estimation. Let c_P^{Ga} denote the density of a Gaussian copula with correlation matrix P. Using Exercise 4.18 and letting f_P denote the density of $\mathcal{N}(0, P)$ and ϕ the density of $\mathcal{N}(0, 1)$, we have

$$\log c_P^{\mathrm{Ga}}(\mathbf{U}_1, ..., \mathbf{U}_n) = \sum_{t=1}^{n} \log f_P(\Phi^{-1}(U_{t,1}), ..., \Phi^{-1}(U_{t,d})) - \sum_{t=1}^{n} \sum_{j=1}^{d} \log \phi(\Phi^{-1}(U_{t,j})).$$

The last part plays no role in determining P, and we thus have

$$\hat{P} = \operatorname*{argmax}_{P \in \mathcal{C}} \sum_{t=1}^{n} \log f_P(\Phi^{-1}(U_{t,1}), ..., \Phi^{-1}(U_{t,d}))$$

where $\mathcal{C}$ denotes the set of all possible correlation matrices. Since the number of parameters is $O(d^2)$, an exhaustive search over this set does not scale well when increasing the dimension. An approximate solution can instead be to compute the MLE for the *covariance* matrix Σ which is given by

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} \mathbf{Y}_t \mathbf{Y}_t^T$$

where $\mathbf{Y}_t := (\Phi^{-1}(U_{t,1}), ..., \Phi^{-1}(U_{t,d}))$. One can then normalise $\hat{\Sigma}$ to obtain an approximation to the desired correlation matrix P. ∘

Example 5.5.2 (Fitting the t copula). Say $(X_1, ..., X_d)$ has the t copula with correlation matrix P and degree of freedom ν, $C_{\nu,P}^t$. From Theorem 5.2.4, we have

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi} \arcsin \rho(X_i, X_j)$$

and so we can compute a nonparametric estimate for $\rho_\tau(X_i, X_j)$ and then approximate the correlation matrix by

$$\hat{\rho}(X_i, X_j) \approx \sin \frac{\pi \hat{\rho}_\tau(X_i, X_j)}{2}.$$

But as we already alluded to in the previous example, this may not be positive definite. If this is the case, we can fix it by applying the algorithm after the example.

Letting $g_{\nu,P}$ denote the density of the t distribution with ν degrees of freedom and covariance matrix P. The log-likelihood of the t copula is then given by

$$\log c_{\nu,P}^t(\mathbf{U}_1, ..., \mathbf{U}_n) = \sum_{t=1}^{n} \log g_{\nu,P}(t_\nu^{-1}(U_{t,1}), ..., t_\nu^{-1}(U_{t,d})) - \sum_{t=1}^{n} \sum_{j=1}^{d} \log g_\nu(t_\nu^{-1}(U_{t,j}))$$

where g_ν denotes the density of the one-dimensional t distribution $t(\nu, 0, 1)$. A method of maximising this equation is to first maximise the first term over all correlation matrices P. To do so, we can use the method of calibrating via Kendall's τ as described above or do an exhaustive search in low dimensions. The degrees of freedom ν may then be estimated using maximum likelihood. ◦

If one has a matrix, which should be close to a correlation matrix, but is not positive definite, the following algorithm is useful for transforming the matrix slightly to become positive definite.

Algorithm 5.5.1 (**Eigenvalue method**). Let R^* be a symmetric matrix of pairwise correlation estimates with unit diagonal entries and off-diagonal entries in $[-1, 1]$ which is not positive semi-definite. The following procedure constructs a valid correlation matrix R.

1. Compute the eigenvalue decomposition $R^* = \Lambda D \Lambda^T$ where D is a diagonal matrix of the eigenvalues of R^*, and Λ is a matrix with columns equal to the eigenvectors of R^*. This decomposition is always possible by the spectral theorems from linear algebra.

2. Replace the negative eigenvalues in D by small strictly positive numbers. Call the new matrix $\tilde{D}$.

3. Calculate $\tilde{R} = \Lambda \tilde{D} \Lambda^T$. This matrix will be positive definite, but not necessarily a correlation matrix, since the diagonal elements will not necessarily all equal one.

4. If $\tilde{R} = \{\tilde{r}_{ij}\}$, define $\Delta\tilde{R} := \operatorname{diag}(\sqrt{\tilde{r}_{11}}, ..., \sqrt{\tilde{r}_{dd}})$. Normalising $\tilde{R}$ via

$$R = (\Delta\tilde{R})^{-1} \tilde{R} (\Delta\tilde{R})^{-1},$$

   then R is the desired output.

We now provide some examples of Archimedean copulas. Since these copulas are explicitly given, the densities can be derived in a straightforward, albeit somewhat tedious manner. From Exercise 4.18, we immediately get that if C is Archimedean with differentiable generator φ, where φ^{-1} is d times differentiable, then

$$\log c(\mathbf{U}_1, ..., \mathbf{U}_n) = \sum_{t=1}^{n} \log(\varphi^{-1})^{(d)} \Big(\sum_{j=1}^{d} \varphi(U_{t,j}) \Big) + \sum_{t=1}^{n} \log \prod_{j=1}^{d} \varphi'(U_{t,j}).$$

Hence the difficult part boils down to computing (by hand or using a CAS tool) the d'th derivative of φ^{-1}, if this is even possible. When it is, this is the most straightforward and robust method of fitting a copula. Often one is interested in the special case of $d = 2$, where we get

$$\log c(\mathbf{U}_1, ..., \mathbf{U}_n) = \sum_{t=1}^{n} \log(\varphi^{-1})''(\varphi(U_{t,1}) + \varphi(U_{t,2})) + \sum_{t=1}^{n} \log(\varphi'(U_{t,1})\varphi'(U_{t,2}).$$

Example 5.5.3 (**Fitting the Clayton copula**). Let C_θ denote the Clayton copula with generator $\varphi(t) = (t^{-\theta} - 1)/\theta$ for $\theta > 0$. We have $\varphi'(t) = -t^{-\theta-1}$ and

$$\varphi^{-1}(s) = (1 + \theta s)^{-1/\theta}, \quad \text{and thus} \quad (\varphi^{-1})''(s) = (1 + \theta)(1 + \theta s)^{-(2+\frac{1}{\theta})}.$$

Plugging into the formula for the log-likelihood in two dimensions, we obtain

$$\log c_\theta(\mathbf{U}_1, ..., \mathbf{U}_n) = n\log(1+\theta) - \left(2 + \frac{1}{\theta}\right)\sum_{t=1}^{n}\log(U_{t,1}^{-\theta} + U_{t,2}^{-\theta} - 1)$$
$$- (1+\theta)\sum_{t=1}^{n}(\log U_{t,1} + \log U_{t,2}).$$

∘

In the exercises, you are asked to derive the log-likelihood for more copulas.

Some tips and tricks for fitting copulas in practice

When fitting copulas in practice, the following things are worth considering.

- What is the dependence structure in the data? If Kendall's τ, Spearman's ρ or the tail dependence coefficients are far from zero, it is important to choose a copula which captures this dependence.
- What is the purpose of the analysis? In a risk management context where we are interested in estimating risk measures such as Value at Risk and Expected Shortfall for financial data, it is essential to choose a copula which provides sufficiently conservative estimates.
- Always fit several copulas. An easy metric for choosing the best fit is to compare the values of the log-likelihood in the optimal parameters (when using the maximum likelihood approach).

As for the implementation aspect, the `R` Package `copula` ([74] and [66]) contains a plethora of useful functions for fitting different copulas. It also contains functions for estimating Kendall's τ and Spearman's ρ. For a manual, see [59]. Many of the essential functions are also showcased in the following case study.

Case study: Novo Nordisk, Pfizer and Eli Lilly

Data exploration and pseudo-samples

To showcase the methods discussed in this section, we analyse stock prices for three large medicine companies, namely Novo Nordisk, Pfizer and Eli Lilly from the period 2020-11-01 to 2025-11-01 (1256 observations in total, yielding 1255 log returns). Such financial data can easily be imported into `R` with the `quantmod` package (see [106]) as follows.

```
library(quantmod)
stocks_env <- new.env()

getSymbols("NVO", env = stocks_env, src = "yahoo", from = as.Date("
    2020-11-01"), to = as.Date("2025-11-01"), warnings = FALSE)
novo <- stocks_env$NVO
```

We start by defining an environment into which the different stocks are put. They are then extracted using `$`. We can repeat for Pfizer ("PFE") and Eli Lilly ("LLY"). Below are two exploratory plots of the data.

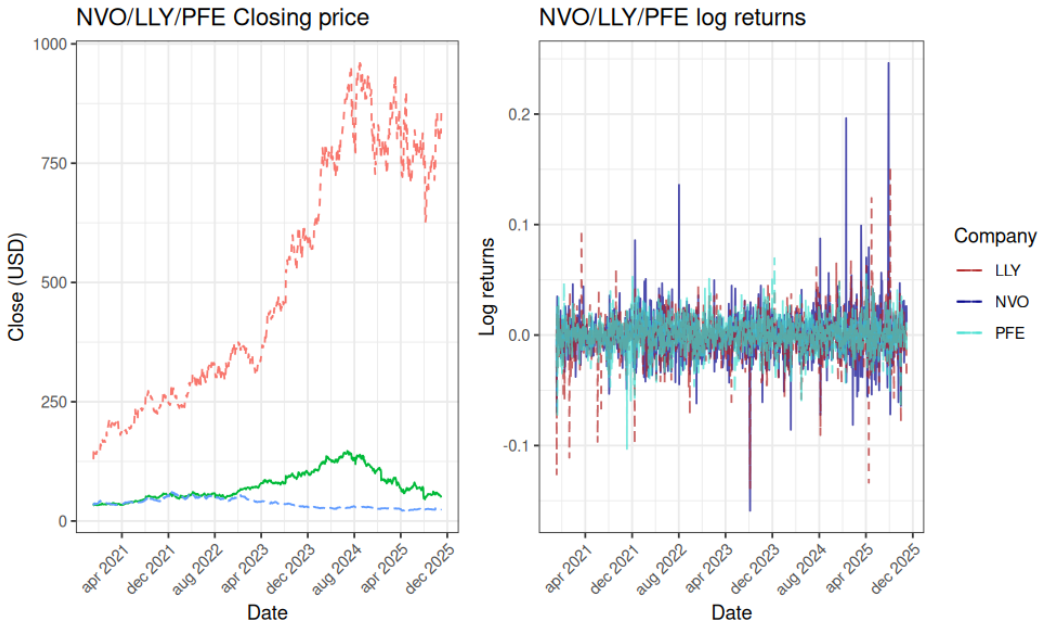


Figure 4: Closing prices (USD) of three medicine companies (left) and the corresponding log returns (right) in the period 2020-11-01 to 2025-11-01.

Say we are interested in portfolios consisting of pairs of these stocks, and we wish to choose a fitting copula for each pair. We therefore work with the log returns as displayed in the right plot. We start by choosing a model for the marginal distributions. For simplicity, we choose the nonparametric approach where each marginal distribution is simply modelled by its empirical distribution function, corresponding to (ii) above. The following code snippet illustrates this for Novo Nordisk.

```
# collect the data in a data.frame
df <- data.frame(date = index(novo), NVO = novo$NVO.Close,
                 LLY = el$LLY.Close, PFE = pf$PFE.Close)

# compute data.frame of log returns
n <- length(df$NVO.Close)
df_returns <- data.frame(date = df$date[1:(n - 1)],
                         NVO = df$NVO.Close[1:(n - 1)] / df$NVO.Close
                             [2:n],
                         LLY = df$LLY.Close[1:(n - 1)] / df$LLY.Close
                             [2:n],
                         PFE = df$PFE.Close[1:(n - 1)] / df$PFE.Close
                             [2:n])

df_log_returns <- data.frame(date = df_returns$date, NVO = log(df_
    returns$NVO),
                             LLY = log(df_returns$LLY), PFE = log(df_
                                 returns$PFE))
```

```r
# define the empirical distribution function as in (ii)
edf_NVO <- function(x) {
  n * mean(df_log_returns$NVO <= x) / (n + 1)
}
edf_NVO <- Vectorize(edf_NVO)

# similarly for LLY and PFE...

# generate (completely nonparametric) pseudo-samples
df_pseudo <- data.frame(NVO = edf_NVO(df_log_returns$NVO),
                        LLY = edf_LLY(df_log_returns$LLY),
                        PFE = edf_PFE(df_log_returns$PFE))
```

The `Vectorize` function transforms a function to be able to take vectors as input and compute function values entry-wise. The following plot displays the pseudo-samples in a grid.

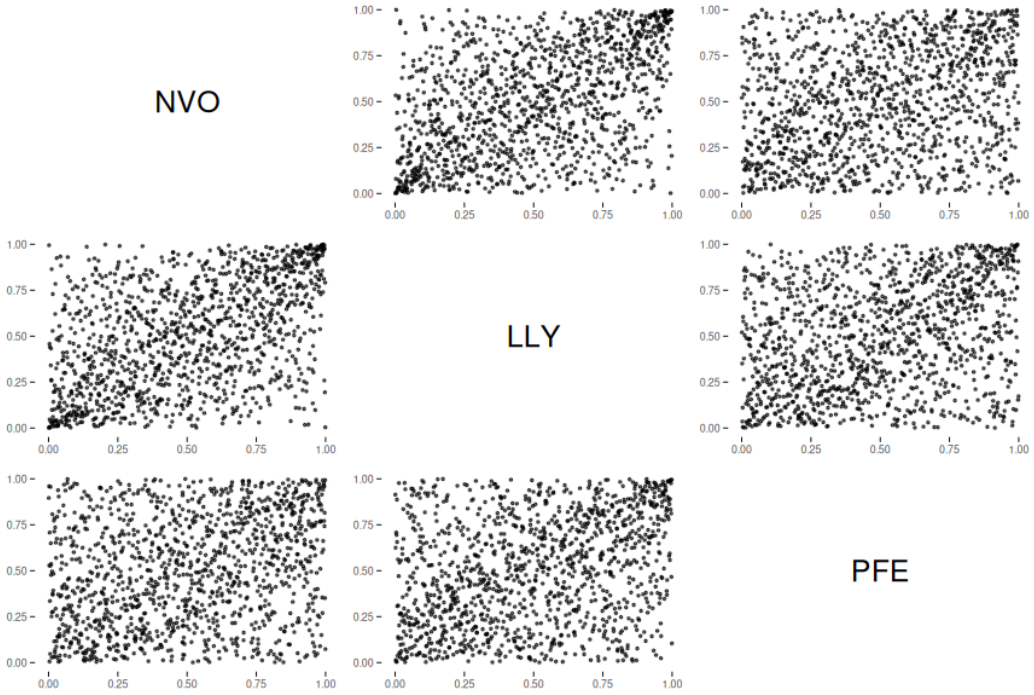


Figure 5: Pairwise scatterplots of pseudo-samples from a copula for trivariate Novo Nordisk, Eli Lilly and Pfizer log returns.

Fitting copulas

We choose to consider four different copulas, namely

- The Gaussian copula,
- the t copula,
- the Clayton copula and
- the Gumbel copula.

We fit the copulas using the `copula` package and maximum likelihood and compare to the method of calibrating via Kendall's τ and Spearman's ρ. We of course consider each of the three pairs of stocks, but we only showcase the code for Novo Nordisk and Eli Lilly, the other pairs being treated in the same way. We start by computing the different notions of correlation as follows:

```
# extract relevant pseudo-samples
U <- df_pseudo[,1:2]

# compute dependence statistics
kendall <- cor(U, method = "kendall")[1,2]     # 0.3257994
spearman <- cor(U, method = "spearman")[1,2]   # 0.4608413
cor(U)[1,2]                                    # 0.4608476
```

We see that $\rho_\tau \approx 0.3258$, $\rho_S \approx 0.4608$ and the ordinary correlation ρ is estimated to be ≈ 0.4608. The Gaussian copula is then fitted as follows.

```
fit_norm <- fitCopula(normalCopula(), data = U, method = "ml")
# rho: 0.479, log-likelihood: 161.3

# calibrate via Spearman's rho and ordinary correlation
spearman                      # 0.4608413
2 * sin(pi * spearman / 6)    # 0.4779224
```

Note how the `fitCopula` function returns the estimate $\rho \approx 0.4790$ which is close to the value obtained by calibrating via Spearman's ρ, but the estimated value of Spearman's ρ (0.4608) is not very close to the maximum likelihood estimate for the correlation. This means that the Gaussian copula is likely not a very good fit as per Proposition 5.3.6. The t copula is fitted in the following way.

```
fit_t <- fitCopula(tCopula(), data = U, method = "ml")
# rho: 0.4911, df: 4.1512, log-likelihood: 198.2
```

Here we have simply used maximum likelihood directly to obtain the model parameters. If we want to calibrate with Kendall's τ, we can do so with the following code.

```
# calibrate via Kendall's tau
rho_K <- sin(pi * kendall/2) # 0.4897164 (very close to 0.4911)
```

We obtain an estimate of the correlation very close to the one obtained from maximum likelihood. Fitting the Clayton copula is done in the following code snippet.

```
# Clayton copula
fit_clayton <- fitCopula(claytonCopula(), data = U, method = "ml")
fit_clayton # theta = 0.9665, log-likelihood = 114.1

# calibration via Kendall's tau
2*kendall/(1 - kendall)  # theta = 0.966476 (very close)
```

We can repeat for Gumbel, using the `gumbelCopula` function instead and obtain an estimate of $\theta \approx 1.468$ from maximum likelihood and $\theta \approx 1.483$ from calibration via Kendall's τ. The log-likelihood is 177.9 for the Gumbel copula. So far, we can conclude that the Gaussian and Clayton copulas fit the worst. This can be explained from the plot of the pseudo-samples above. There definitely seems to be quite strong dependence in the tails for Novo Nordisk and Eli Lilly. We know that $\lambda_U = \lambda_L = 0$ for the Gaussian copula, and as you will verify in the exercises, $\lambda_U = 0$ for the Clayton copula and $\lambda_U = 2 - 2^{1/\theta}$ for the Gumbel copula. Hence the poor fit from the Clayton and Gaussian copulas can be explained by a lack of ability to capture tail behaviour. On the other hand, the t copula and the Gumbel copula seem to fit much better with the t copula having the highest log-likelihood.

Conclusions

Repeating all of the above for the two other pairs of stocks, we can summarise our findings in the following tables.

Pair	Gaussian ($\hat{\rho}$)	t ($\hat{\rho}, \hat{\nu}$)	Clayton ($\hat{\theta}$)	Gumbel ($\hat{\theta}$)
NVO/LLY	0.479	(0.4911, 4.1512)	0.9665	1.468
NVO/PFE	0.2392	(0.2613, 8.1575)	0.4253	1.181
LLY/PFE	0.301	(0.3229, 5.7620)	0.5352	1.252

Table 5.1: Fitted values for each of the pairs of medicine stocks. All values are obtained via maximum likelihood using the `fitCopula` function from the `copula` package.

Pair	Gaussian	t	Clayton	Gumbel
NVO/LLY	161.3	198.2	114.1	177.9
NVO/PFE	36.31	44.76	17.06	40.39
LLY/PFE	58.57	74.14	28.1	71.29

Table 5.2: The log-likelihood at the maximum likelihood estimates for each of the pairs of medicine stocks.

Across all the different copulas, we see that the t copula gives the best fit, although the Gumbel copula comes quite close to the t copula. It also seems that Novo Nordisk and Eli Lilly are the most strongly correlated stocks and Novo Nordisk and Pfizer the least correlated.

Further considerations

The presented analysis can be extended considerably. The following questions come to mind:

- Is there a copula which fits better than the t copula?
- Do the results of the analysis change if we work with other marginal distributions?
- What is the estimated VaR and ES for a portfolio of the different stocks? And how do the estimates change when using different copulas and marginal distributions.

- How would the results of the analysis look if we just fit a single three-dimensional copula and consider a single portfolio consisting of all three stocks?

All these questions are addressed in some of the practical exercises below.

5.6 Geometric properties of Archimedean copulas*

In this supplementary section, we investigate geometric properties of Archimedean copulas. Geometric properties for copulas include level sets and curves and their probabilistic interpretations. Ultimately, these considerations will allow us to prove part (ii) of Theorem 5.2.3.

Level curves and C-measure

Definition 5.6.1. The *level set* of a copula C is the set of points

$$\{(u_1, ..., u_d) \in [0,1]^d : C(u_1, ..., u_d) = t\}$$

for $t \in [0,1]$. When $t = 0$, the set is called the *zero set* of C and is denoted by $Z(C)$.

Now consider a two-dimensional Archimedean copula C with generator φ. In that case, the level set for $t \in [0,1)$ consists of the $u_1, u_2 \in [0,1]$ such that $\varphi(u_1) + \varphi(u_2) = \varphi(t)$. We see that this set can be realised as the *level curve* parametrised by the map

$$u_2 = L_t(u_1) = \varphi^{-1}(\varphi(t) - \varphi(u_1)).$$

When $t = 0$, this map is called the *zero curve* of C. The following result shows that the level curve of an Archimedean copula in two dimensions is convex. This is not the case for all copulas, see Exercise 5.17.

Proposition 5.6.2. *An Archimedean copula in two dimensions has convex level curves.*

Proof. Let C be an Archimedean copula with generator φ and consider the level curve L_t given above, where $t \in [0,1)$. By Exercise 4.29, it suffices to show that L_t is midconvex, since L_t is continuous. Since φ is convex,

$$\begin{aligned}\varphi(t) - \varphi\left(\frac{u_1+u_2}{2}\right) &\geq \varphi(t) - \frac{\varphi(u_1)+\varphi(u_2)}{2} \\ &= \frac{(\varphi(t)-\varphi(u_1)) + (\varphi(t)-\varphi(u_2))}{2},\end{aligned}$$

and as φ^{-1} is convex and decreasing, we get

$$\begin{aligned}L_t\left(\frac{u_1+u_2}{2}\right) &= \varphi^{-1}\left(\varphi(t) - \varphi\left(\frac{u_1+u_2}{2}\right)\right) \\ &\leq \varphi^{-1}\left(\frac{(\varphi(t)-\varphi(u_1)) + (\varphi(t)-\varphi(u_2))}{2}\right) \\ &\leq \frac{1}{2}(\varphi^{-1}(\varphi(t)-\varphi(u_1)) + \varphi^{-1}(\varphi(t)-\varphi(u_2))) \\ &= \frac{L_t(u_1)+L_t(u_2)}{2},\end{aligned}$$

completing the proof. ■

Another geometric property with a nice probabilistic interpretation is the concept of the C-measure.

Definition 5.6.3. For a d-dimensional copula C, we define the *C-measure* of C as the measure V_C defined on subsets of $[0,1]^d$ by

$$V_C([0,u_1] \times \cdots \times [0,u_d]) = C(u_1, ..., u_d).$$

The collection of subsets of the form $[0,u_1] \times \cdots \times [0,u_d]$ for $u_1, ..., u_d \in [0,1]$ is clearly an intersection stable class, and so the set function V_C extends uniquely to a measure on all Borel subsets of $[0,1]^d$. If $A \in \mathcal{B}([0,1]^d)$, the probabilistic interpretation of $V_C(A)$ is the following: If $(U_1, ..., U_d)$ is distributed according to C, then $V_C(A) = P((U_1, ..., U_d) \in A)$. We will in the following focus on the case $d = 2$. As with all measures, theorems are usually proven by showing properties on members of a class generating the relevant sigma-algebra and then applying general properties of measures. The following theorem provides a formula for the C-measure of the level curve of an Archimedean copula. Recall that we denote by $f(x-)$ and $f(x+)$ the left and right limits, respectively, of the function f in x,

$$f(x-) = \lim_{y \uparrow x} f(y), \quad f(x+) = \lim_{y \downarrow x} f(y).$$

Theorem 5.6.4. *For an Archimedean copula C with generator φ, the C-measure of the level curve $\varphi(u_1) + \varphi(u_2) = \varphi(t)$ for $t \in (0,1)$ is given by*

$$\varphi(t) \left(\frac{1}{\varphi'(t-)} - \frac{1}{\varphi'(t+)} \right).$$

In particular, in every point t where $\varphi'(t)$ exists, the C-measure of the level curve at t is zero. This is the case for all but a countable number of t.

Proof. Since φ is convex, the one-sided derivatives $\varphi'(t-)$ and $\varphi'(t+)$ exist in $(0,1]$ and $[0,1)$, respectively (see [104] for details). Let $n \in \mathbb{N}$, $t \in (0,1)$ and $w = \varphi(t)$. Now consider the following partition. First make the equidistant partition $0, w/n, 2w/n, ..., w$ of $[0,w]$ and let the corresponding partition of $[t,1]$, $t = t_0, t_1, ..., t_n = 1$ be defined by

$$t_{n-k} = \varphi^{-1}\left(\frac{kw}{n}\right), \quad k = 0, 1, ..., n.$$

We then have

$$\begin{aligned} C(t_j, t_k) &= \varphi^{-1}(\varphi(t_j) + \varphi(t_k)) = \varphi^{-1}\left(\frac{n-j}{n}w + \frac{n-k}{n}w\right) \\ &= \varphi^{-1}\left(w + \frac{n-j-k}{n}w\right). \end{aligned}$$

Note that $C(t_j, t_{n-j}) = \varphi^{-1}(w) = t$, and let $R_k := [t_{k-1}, t_k] \times [t_{n-k}, t_{n-k+1}]$, $S_n = \cup_{k=1}^n R_k$. Then S_n is an approximation to the level curve as the following figure illustrates.

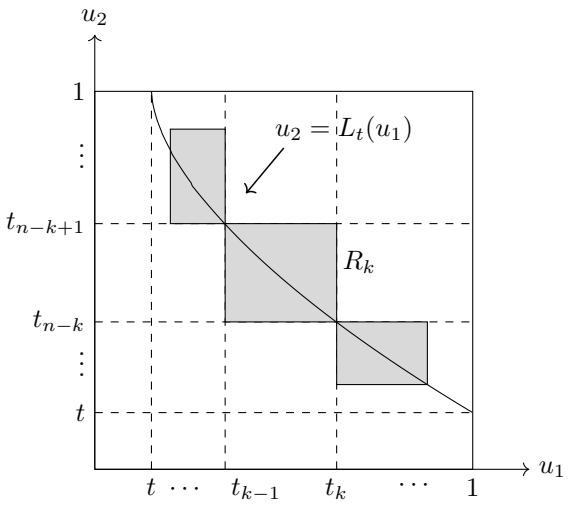


Figure 6: The rectangles R_k in the proof. The grey area is the set S_n which approximates the curve.

Since φ^{-1} is convex, it holds that

$$0 \leq t_1 - t_0 \leq t_2 - t_1 \leq \cdots \leq t_n - t_{n-1} = 1 - t_{n-1},$$

and $\lim_{n\to\infty}(1 - t_{n-1}) = 1 - \varphi^{-1}(0) = 0$. This shows that S_n actually approximates the level curve, so using continuity from above, the C-measure of the level curve is given by $\lim_{n\to\infty} V_C(S_n)$. We compute the C-measure of each rectangle R_k to be

$$\begin{aligned} V_C(R_k) &= C(t_k, t_{n-k+1}) + C_{t_{k-1},t_{n-k}} - C(t_k, t_{n-k}) - C(t_{k-1}, t_{n-k+1}) \\ &= C(t_{k-1}, t_{n-k}) + C_{t_k,t_{n-k+1}} - 2t \\ &= \left(\varphi^{-1}\left(w + \frac{w}{n}\right) - \varphi^{-1}(w)\right) - \left(\varphi^{-1}(w) - \varphi^{-1}\left(w - \frac{w}{n}\right)\right). \end{aligned}$$

Using that the R_k are disjoint, the C-volume of S_n is then

$$V_C(S_n) = \sum_{k=1}^{n} V_C(R_k) = w\left(\frac{\varphi^{-1}(w + w/n) - \varphi^{-1}(w)}{w/n} - \frac{\varphi^{-1}(w) - \varphi^{-1}(w - w/n)}{w/n}\right).$$

Now take the limit $n \to \infty$ to obtain the result of the theorem. ■

Example 5.6.5. Many Archimedean copulas have a generator which is C^1 (Gumbel, Clayton etc.). For all these copulas, the result shows that the C-measure of their level curves are zero. ∘

Example 5.6.6 (Example 4.5 in [91])**.** Consider the Archimedean copula C with generator φ given by

$$\varphi(t) = \begin{cases} -\frac{2-\theta}{\theta}t + 2 - \theta, & t \in [0, \theta/2] \\ 1 - t, & t \in [\theta/2, 1] \end{cases}$$

for a parameter $\theta \in (0, 1]$. The astute reader will see that this generator does not satisfy the usual requirements. Specifically, we do not have $\varphi(0) = \infty$. A generator which satisfies all the requirements for an Archimedean copula generator except for $\varphi(0) = \infty$ is called *non-strict*. See Exercise 4.27 for more details. The generator is drawn in the figure below.

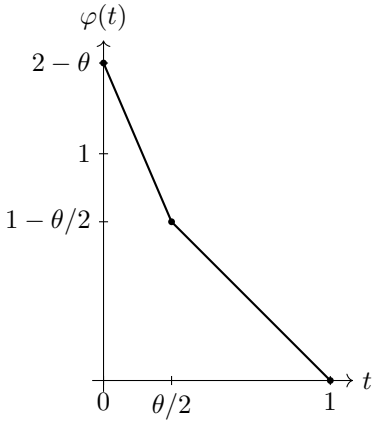


Figure 7: The graph of the generator.

We see that φ is the piecewise linear function which connects $(0, 2-\theta)$ to $(\theta/2, 1-\theta/2)$ to $(1, 0)$. Clearly, the generator is not C^1 in $t = \theta/2$, and we get that the C-measure of the level curve at this point is

$$\left(1 - \frac{\theta}{2}\right)\left(-\frac{\theta}{2-\theta} + 1\right) = 1 - \theta.$$

$\circ$

The following theorem provides a formula for the C-measure to the left of a level curve. The proof is very similar to the one for Theorem 5.6.4.

Theorem 5.6.7. *Let C be an Archimedean copula with generator φ, and let $K_C(t)$ denote the C-measure of the set $\{(u_1, u_2) \in [0, 1]^2 : C(u_1, u_2) \leq t\}$. It holds for any $t \in [0, 1]$ that*

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t+)}. \tag{5.2}$$

Proof. Let $t \in (0, 1)$ and $w = \varphi(t)$. Let $n \in \mathbb{N}$ and define the partition $t = t_0, t_1, ..., t_n = 1$ in the same way as in the proof of Theorem 5.6.4. Define the rectangle $R'_k = [t_{k-1}, t_k] \times [0, t_{n-k+1}]$ and set $S'_n = \cup_{k=1}^{n} R'_k$.

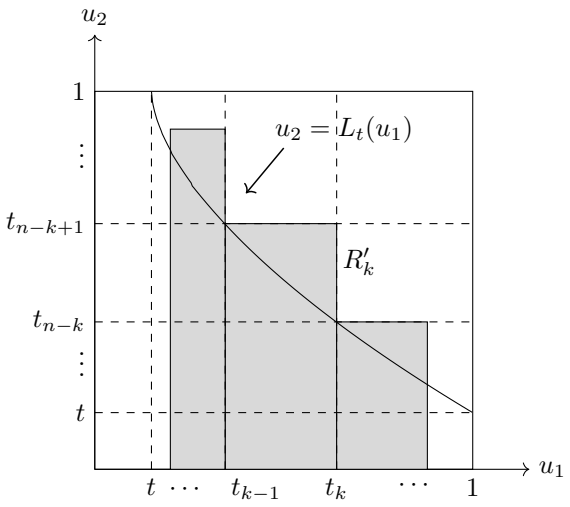


Figure 8: The rectangles R'_k in the proof. The grey area is the set S'_n which we use to approximate the C-measure of the set below the level curve $C(u_1, u_2) = t$.

From the figure, it is clear that $K_C(t)$ is given by the sum of t (the probability mass to the left of t) and $\lim_{n\to\infty} V_C(S'_n)$ (again by continuity from above). We compute

$$V_C(R'_k) = C(t_k, t_{n-k+1}) - t = \varphi^{-1}\left(w - \frac{w}{n}\right) - \varphi^{-1}(w).$$

From this, it follows that

$$V_C(S'_n) = \sum_{k=1}^{n} V_C(R'_k) = -w\frac{\varphi^{-1}(w) - \varphi^{-1}(w - w/n)}{w/n},$$

and taking the limit $n \to \infty$ completes the proof. ■

Corollary 5.6.8. *Let (U_1, U_2) be a pair of variables, each uniform on $[0, 1]$, with distribution function C, an Archimedean copula with generator φ. The function K_C given by (5.2) is the distribution function of $C(U_1, U_2)$.*

This corollary will allow us to prove Theorem 5.2.3 (ii).

The proof of Theorem 5.2.3 (ii)

Proof of Theorem 5.2.3 (ii). Let (X_1, X_2) have copula C, assumed to be Archimedean with generator φ. Let (U_1, U_2) be distributed according to C. By (i) and Corollary 5.6.8, we have

$$\rho_\tau(X_1, X_2) = 4E[C(U_1, U_2)] - 1 = 4\int_0^1 t dK_C(t) - 1.$$

Now apply integration by parts to obtain

$$4\int_0^1 t dK_C(t) - 1 = 4[tK_C(t)]_0^1 - 4\int_0^1 K_C(t)dt - 1 = 3 - 4\int_0^1 K_C(t)dt.$$

Using the result of Theorem 5.6.7, we get

$$\rho_\tau(X_1, X_2) = 3 - 4\int_0^1 \left(t - \frac{\varphi(t)}{\varphi'(t+)}\right) dt = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt,$$

where we may replace $\varphi'(t+)$ by $\varphi'(t)$ in the denominator since concave functions are differentiable almost everywhere. ■

5.7 Additional proofs*

Proof of Theorem 5.3.9. Let $(X_1, X_2), (Y_1, Y_2)$ and (Z_1, Z_2) be independent random variables with the same distribution. Due to continuity,

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 2P((X_1 - Y_1)(X_2 - Y_2) > 0) - 1,\\ \rho_S(X_1, X_2) &= 6P((X_1 - Y_1)(X_2 - Z_2) > 0) - 3.\end{aligned}$$

By symmetry of the three pairs, we may write

$$\begin{aligned}\rho_\tau(X_1, X_2) = \frac{2}{3}\Big(&P((X_1 - Y_1)(X_2 - Y_2) > 0) + P((Y_1 - Z_1)(Y_2 - Z_2) > 0)\\ &+ P((Z_1 - X_1)(Z_2 - X_2) > 0)\Big) - 1\end{aligned}$$

and

$$\begin{aligned}\rho_S(X_1, X_2) = \Big(&P((X_1 - Y_1)(X_2 - Z_2) > 0) + P((X_1 - Z_1)(X_2 - Y_2) > 0)\\ &+ P((Y_1 - X_1)(Y_2 - Z_2) > 0) + P((Z_1 - Y_1)(Z_2 - X_2) > 0)\\ &+ P((Y_1 - Z_1)(Y_2 - X_2) > 0) + P((Z_1 - X_1)(Z_2 - Y_2) > 0)\Big) - 3.\end{aligned}$$

Both expressions above are now invariant under permutations of the subscripts, and thus both expressions may be evaluated under the assumption $X_1 < Y_1 < Z_1$. For $\rho_\tau(X_1, X_2)$, this yields

$$\rho_\tau(X_1, X_2) = \frac{2}{3}(P(X_2 < Y_2) + P(Y_2 < Z_2) + P(X_2 < Z_2)) - 1.$$

We provide a formal justification for this identity for $\rho_\tau(X_1, X_2)$. The argument for $\rho_S(X_1, X_2)$ is analogous. Using the tower property, we may write

$$\rho_\tau(X_1, X_2) = \frac{2}{3}E[f(X_1, Y_1, Z_1)] - 1$$

where

$$\begin{aligned}f(X_1, Y_1, Z_1) = &1_{\{X_1 > Y_1\}}P(X_2 > Y_2 \mid X_1, Y_1) + 1_{\{X_1 < Y_1\}}P(X_2 < Y_2 \mid X_1, Y_1)\\ &+ 1_{\{Y_1 < Z_1\}}P(Y_2 < Z_2 \mid Z_1, Y_1) + 1_{\{Y_1 > Z_1\}}P(Y_2 > Z_2 \mid Y_1, Z_1)\\ &+ 1_{\{Z_1 > X_1\}}P(Z_2 > X_2 \mid X_1, Z_1) + 1_{\{Z_1 < X_1\}}P(Z_2 < X_2 \mid X_1, Z_1).\end{aligned}$$

Note that $f(X_1, Y_1, Z_1)$ is invariant under the ordering of (X_1, Y_1, Z_1). Hence we may compute $f(X_1, Y_1, Z_1)$ under the assumption that $X_1 < Y_1 < Z_1$ which yields

$$f(X_1, Y_1, Z_1) = P(X_2 < Y_2 \mid X_1, Y_1) + P(Y_2 < Z_2 \mid Y_1, Z_1) + P(Z_2 > X_2 \mid X_1, Z_1).$$

Plugging this expression into the one for $\rho_\tau(X_1, X_2)$ yields the desired identity. For $\rho_S(X_1, X_2)$, the corresponding identity reads

$$\begin{aligned}\rho_S(X_1, X_2) =& \Big(P(X_2 < Z_2) + P(X_2 < Y_2) + P(Y_2 > Z_2) + P(Z_2 > X_2) \\ &+ P(Y_2 < X_2) + P(Z_2 > Y_2)\Big) - 3.\end{aligned}$$

Now let $p_{XYZ} := P(X_2 < Y_2 < Z_2 \mid X_1 < Y_1 < Z_1)$, $p_{XZY} := P(X_2 < Z_2 < Y_2 \mid X_1 < Y_1 < Z_1)$ etc. (note that the order on the variables in the condition doesn't change). The six p's sum to one, and we get

$$\begin{aligned}\rho_\tau(X_1, X_2) &= \frac{2}{3}\Big(p_{XYZ} + p_{XZY} + p_{ZXY}) + (p_{XYZ} + p_{YXZ} + p_{YZX}) \\ &\quad + (p_{XYZ} + p_{XZY} + p_{YXZ})\Big) - 1 \\ &= p_{XYZ} + \frac{1}{3}(p_{XZY} + p_{YXZ}) - \frac{1}{3}(p_{YZX} + p_{ZXY}) - p_{ZYX}\end{aligned}$$

and

$$\begin{aligned}\rho_S(X_1, X_2) &= 2(p_{XYZ} + p_{XZY} + p_{YXZ}) - 1 \\ &= p_{XYZ} + p_{XZY} + p_{YXZ} - p_{YZ1} - p_{ZYX} - p_{ZYX},\end{aligned}$$

and a little rearranging yields

$$-1 \leq 3\rho_\tau(X_1, X_2) - 2\rho_S(X_1, X_2) \leq 1.$$

■

Proof of Theorem 5.3.10. Let again (X_1, X_2), (Y_1, Y_2) and (Z_1, Z_2) be independent copies of each other with common distribution function F. Let p denote the probability that two of the pairs are concordant with the third, i.e.

$$p := P((Y_1 - X_1)(Y_2 - X_2) > 0, (Z_1 - X_1)(Y_2 - X_2) > 0).$$

Then we may compute using independence,

$$\begin{aligned}p &= \int_{\mathbb{R}}\int_{\mathbb{R}} P((Y_1 - x_1)(Y_2 - x_2) > 0)P((Z_1 - x_1)(Y_2 - x_2) > 0)dF(x_1, x_2) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} P((Y_1 - x_1)(Y_2 - x_2) > 0)^2 dF(x_1, x_2).\end{aligned}$$

Now apply the conditional Jensen's inequality, the tower property and continuity to obtain

$$\begin{aligned}p &\geq \left(\int_{\mathbb{R}}\int_{\mathbb{R}} P((Y_1 - x_1)(Y_2 - x_2) > 0)dF(x_1, x_2)\right)^2 \\ &= P((Y_1 - X_1)(Y_2 - X_2) > 0)^2 = \left(\frac{1 + \rho_\tau(X_1, X_2)}{2}\right)^2.\end{aligned}$$

Again by symmetry of the variables, we have

$$\begin{aligned} p = \frac{1}{3}\Big(& P((Y_1 - X_1)(Y_2 - X_2) > 0, (Z_1 - X_1)(Z_2 - X_2) > 0) \\ & + P((X_1 - Y_1)(X_2 - Y_2) > 0, (Z_1 - Y_1)(Z_2 - Y_2) > 0) \\ & + P((X_1 - Z_1)(X_2 - Z_2) > 0, (Y_1 - Z_1)(Y_2 - Z_2) > 0)\Big). \end{aligned}$$

Now define the probabilities p_{XYZ}, p_{XZY} etc. as in the proof of Theorem 5.3.9. Then

$$\begin{aligned} p &= \frac{1}{3}\Big((p_{XYZ} + p_{XZY}) + p_{XYZ} + (p_{XYZ} + p_{YXZ})\Big) \\ &= p_{XYZ} + \frac{1}{3}p_{XZY} + \frac{1}{3}p_{YXZ}. \end{aligned}$$

Recall that we in the proof of Theorem 5.3.9 proved the identity

$$\rho_S(X_1, X_2) = 2(p_{XYZ} + p_{XZY} + p_{YXZ}) - 1.$$

Combine with what we have just shown to obtain

$$\frac{1 + \rho_S(X_1, X_2)}{2} = p_{XYZ} + p_{XZY} + p_{YXZ} \geq p \geq \left(\frac{1 + \rho_\tau(X_1, X_2)}{2}\right)^2.$$

This shows the first inequality. To prove the other, replace "concordant" with "discordant" in the beginning of the proof and make similar calculations. ■

Notes and comments

An early treatment of dependence concepts is due to Hoeffding, see [56] and [57]. For modern books treating dependence concepts, consult e.g. [107] and [24]. References for rank correlations include [78] and [69], while [109] discusses the relation between rank correlation and copulas. The formula for Kendall's τ presented in Theorem 5.2.3 (ii) was proved in [40]. The results presented in the text on the relationship between Kendall's τ and Spearman's ρ are all from [91]. Theorem 5.3.9 is originally from [17], while Theorem 5.3.10 is from [27]. The proofs presented are both from [78].

As for estimation and statistical inference using the presented dependence concepts, the estimators for Kendall's τ and Spearman's ρ are both examples of so-called *U-statistics*. More background on such estimators and their asymptotic properties can be found in [81]. The method of calibrating the Gaussian copula using Spearman's ρ can be found in [65]. The eigenvalue method as presented above is due to [105]. Approach (i) of generating pseudo-samples from a copula is described in [70] who also provides asymptotic theory. Approach (ii) (which was used in the case study above) is described by [42] who also prove asymptotic results. Goodness-of-fit was barely discussed above, but several methods exist. A survey article on the subject is [45]. For examples of applications of copulas in actuarial and financial contexts, see [35], [76], [13] and [12].

5.8 Exercises

Theoretical exercises

Exercise 5.1:
Recall that the Clayton copula has generator

$$\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$$

and that the copula is explicitly given by

$$C_\theta^{\mathrm{Cl}}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad 0 < \theta < \infty.$$

1)Verify that $\rho_\tau = \theta/(\theta + 2)$.
2)Show that $\lambda_U = 0$.
3)Show that $\lambda_L = 2^{-1/\theta}$.

Exercise 5.2:
Recall that the Gumbel copula has generator

$$\varphi(t) = (-\log t)^\theta,$$

and that the copula is explicitly given by

$$C_\theta^{\mathrm{Gu}}(u_1, u_2) = \exp\left(-\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{1/\theta}\right), \quad 1 \leq \theta < \infty.$$

1)Verify that $\rho_\tau = 1 - 1/\theta$.
2)Show that $\lambda_U = 2 - 2^{1/\theta}$.
3)Show that $\lambda_L = 0$.

Exercise 5.3:
Prove the following formula ([70]) for Kendall's τ for an Archimedean copula with generator φ:

$$\rho_\tau = 1 - 4\int_0^\infty s\left(\frac{d}{ds}\varphi^{-1}(s)\right)^2 ds.$$

Exercise 5.4:
In this exercise, we verify the claims in the text on concordance and discordance. Let (X_1, X_2) be a pair of continuous random variables.
1)Show that $\rho_\tau(X_1, X_2) = 1$ if and only if (X_1, X_2) is comonotonic.
2)Show that $\rho_\tau(X_1, X_2) = -1$ if and only if (X_1, X_2) is countermonotonic.

Exercise 5.5:
Let (X_1, X_2) be comonotone. Show that $\lambda_U(X_1, X_2) = 1$. What is $\lambda_U(X_1, X_2)$ when (X_1, X_2) is countermonotonic?

Exercise 5.6:
In this exercise, we consider the Morgenstern copula introduced in Exercise 4.3,

$$C_\theta(u_1, u_2) = u_1 u_2(1 + \theta(1 - u_1)(1 - u_2)),$$

where $\theta \in [-1, 1]$ is a parameter.

1)Show that

$$dC_\theta(u_1, u_2) = (1 + \theta(1 - 2u_1)(1 - 2u_2))du_1 du_2.$$

Hint: We have

$$dC_\theta(u_1, u_2) = \frac{\partial^2 C_\theta(u_1, u_2)}{\partial u_1 \partial u_2} du_1 du_2.$$

2)Show that Kendall's τ is

$$\rho_\tau = \frac{2\theta}{9}.$$

3)Show that Spearman's ρ is

$$\rho_S = \frac{\theta}{3}.$$

4)Compute λ_U and λ_L for this copula.

Exercise 5.7:
A collection of random variables $(X_1, ..., X_d)$ is said to be *exchangeable* if for every permutation Π of $1, ..., d$, we have $(X_1,..., X_d) \overset{\mathrm{d}}{=} (X_{\Pi(1)}, ..., X_{\Pi(d)})$.

1)Assume (U_1, U_2) is exchangeable. Verify that

$$P(U_2 \leq u \mid U_1 = u) = P(U_1 \leq u \mid U_2 = u).$$

Now assume (U_1, U_2) has distribution function equal to a copula for a pair of continuous exchangeable random variables (X_1, X_2). Show that the coefficient of lower tail dependence λ_L satisfies

$$\lambda_L(X_1, X_2) = 2 \lim_{u \downarrow 0} P(U_2 \leq u \mid U_1 = u).$$

2)Let X_1 and X_2 be continuous random variables with distribution functions F_1 and F_2. Prove that (X_1, X_2) is exchangeable if and only if $F_1 = F_2$ and the copula C of (X_1, X_2) satisfies

$$C(u_1, u_2) = C(u_2, u_1)$$

for all $u_1, u_2 \in [0, 1]$. Such a copula is called *symmetric*.

3)Provide an example which shows that it does not suffice to have $F_1 = F_2$ in order to conclude that (X_1, X_2) is exchangeable.

4)Assume (X_1, X_2) has identical marginal distributions and has an Archimedean copula. Show that (X_1, X_2) is exchangeable.

5)Show that $(X_1, ..., X_d)$ is exchangeable if and only if the characteristic function Φ is symmetric in its input: $\Phi(t_1, ..., t_d) = \Phi(t_{\Pi(1)}, ..., t_{\Pi(d)})$ for any permutation Π of $1, ..., d$.

6)Show that any elliptical distribution with $\Sigma_{ij} = \rho$ for all $i \neq j$ and $\sigma_{ii} = \sigma^2$ and $\boldsymbol{\mu}_i = \mu$ for all i is exchangeable.

Exercise 5.8: Exercise 3.7 in [91]
Consider the copula

$$C(u_1,u_2)=\begin{cases}u_1, & 0\le u_1\le u_2/2\le 1/2\\ u_2/2, & 0\le u_2/2<u_1<1-u_2/2\\ u_1+u_2-1, & 1/2\le 1-u_2/2\le u_1\le 1.\end{cases}$$

Assume (U_1,U_2) has distribution function C.

1)Show that $P(U_2=1-|2U_1-1|)=1$ and $\mathrm{Cov}(U_1,U_2)=0$. Hence two random variables can be uncorrelated, even though one of them can be perfectly determined as a function of the other.

2)Show that C is not symmetric. Hence two random variables can be identically distributed and uncorrelated but not exchangeable.

3)Show that $P(U_2-U_1>0)=2/3$. Hence two random variables can be identically distributed, even though their difference is not symmetric about zero.

4)Let $X_1=2U_1-1$ and $X_2=2U_2-1$. Then X_1 and X_2 are both Unif$(-1,1)$. Show that $P(X_1+X_2>0)=2/3$. Hence two random variables can both be symmetric about zero, even though their sum is not.

Exercise 5.9:
Consider again the Frank copula with generator

$$\varphi(t)=-\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right),$$

where $\theta\in\mathbb{R}\setminus\{0\}$.

1)Compute λ_U.

2)Compute λ_L.

It is possible to obtain almost explicit expressions for Kendall's τ and Spearman's ρ for the Frank copula. These are

$$\rho_\tau=1-\frac{4}{\theta}(1-D_1(\theta))\quad\text{and}\quad\rho_S=1-\frac{12}{\theta}(D_1(\theta)-D_2(\theta)),$$

where D_k is the *Debye function* given by

$$D_k(x)=\frac{k}{x^k}\int_0^x\frac{t^k}{e^t-1}dt.$$

These formulas are provided in [39] and [90].

Exercise 5.10: The log-copula
Consider the function $\varphi:[0,1]\to[0,\infty]$ given by

$$\varphi(t)=\left(1-\frac{\log t}{\theta\psi}\right)^{\theta+1}-1$$

where $\theta,\psi\in(0,\infty)$ are parameters.

1)Show that φ is a valid generator for a two-dimensional Archimedean copula family. Denote the copula by $C_{\theta,\psi}$.

2)Show that

$$C_{\theta,\psi}(u_1,u_2) = \exp\Big(\theta\psi\Big(1-\Big(\Big(1-\frac{\log u_1}{\theta\psi}\Big)^{\theta+1}+\Big(1-\frac{\log u_2}{\theta\psi}\Big)^{\theta+1}-1\Big)^{1/(1+\theta)}\Big)\Big).$$

3)Show that $\lambda_L = 0$.

One can show that also $\lambda_U = 0$. You are welcome to try and verify this. In the remaining exercises, we compute Kendall's τ for this copula as explicitly as possible.

4)Show that

$$\frac{\varphi(t)}{\varphi'(t)} = -\frac{\theta\psi}{\theta+1}\Big(t-\frac{t\log t}{\theta\psi}-\frac{t}{(1-(\log t)/\theta\psi)^{\theta}}\Big).$$

5)Show that

$$\int t\log t dt = \frac{t^2\log t}{2}-\frac{t^2}{4}$$

and conclude that

$$\int_0^1 t\log t dt = -\frac{1}{4}.$$

6)Show that

$$\int_0^1 \frac{t}{(1-(\log t)/\theta\psi)^{\theta}}dt = e^{2\theta\psi}(\theta\psi)^{\theta}2^{\theta-1}\Gamma(1-\theta,2\theta\psi)$$

where $\Gamma(x,y)$ is the *upper incomplete gamma function* given by

$$\Gamma(x,y) = \int_y^{\infty} t^{x-1}e^{-t}dt.$$

Hint: Use substitution several times. It is easiest to start with $s = -(\log t)/\theta\psi$.

7)Combine the pieces and show that

$$\rho_\tau = \frac{e^{2\theta\psi}2^{\theta+1}(\theta\psi)^{\theta+1}\Gamma(1-\theta,2\theta\psi)+\theta-2\theta\psi}{\theta+1}$$

This copula is sometimes referred to as the *log-copula* and is considered in e.g. [61], [42] and [60].

Exercise 5.11:
Consider the function $\varphi : [0,1] \to [0,\infty]$ given by

$$\varphi(t) = \log\left(\frac{1-\theta(1-t)}{t}\right)$$

for $\theta \in [-1,1)$.

1)Prove that φ is a valid generator for a two-dimensional Archimedean copula. We denote this copula by C_θ.

2)Show that

$$C_\theta(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1-u_1)(1-u_2)}.$$

This is the Ali–Mikhail–Haq copula from Exercise 4.6.

3)Compute the coefficients of lower and upper tail dependence λ_L and λ_U for C_θ.

4)Show that Kendall's τ for this copula is given by

$$\rho_\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2}\log(1-\theta).$$

5)[Bonus] Show that Spearman's ρ for this copula is given by

$$\rho_S = \frac{12(1+\theta)}{\theta^2}\text{dilog}(1-\theta) - \frac{24(1-\theta)}{\theta^2}\log(1-\theta) - \frac{3(\theta+12)}{\theta}$$

where dilog is the *dilogarithm function* given by

$$\text{dilog}(x) = \int_1^x \frac{\log t}{1-t} dt.$$

In the next exercise, we show that this copula extends to arbitrary dimensions.

Exercise 5.12:
Consider again the Ali–Mikhail–Haq copula generator from the previous exercise,

$$\varphi(t) = \log\left(\frac{1-\theta(1-t)}{t}\right)$$

where $\theta \in [-1, 1)$. In this exercise, we show that the Ali–Mikhail–Haq copula extends to any dimension in two different ways for appropriate restrictions on θ.

1)The inverse $\varphi^{-1}(t)$ can be realised as a Laplace transform of a certain well-known discrete distribution. Which one? And for which θ is this distribution well-defined?

2)Describe and implement an algorithm for simulating from the d-dimensional Ali–Mikhail–Haq copula.

3)One can also prove directly that φ^{-1} is completely monotonic. Do so. Hint: Write φ^{-1} as an infinite series. Think geometric series!

Exercise 5.13:
In the text, we proved that if $(X_1, X_2) \sim \mathcal{N}(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \rho \in (-1, 1),$$

then $\lambda_U(X_1, X_2) = 0$. In this exercise, we go through a different proof. Due to the symmetry of the Gaussian copula, it holds that $\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2)$. Now use the result of Exercise 5.7 to verify that $\lambda_U(X_1, X_2) = 0$.

Exercise 5.14:
Suppose that $(X_1, X_2) \sim t_2(\nu, \mathbf{0}, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Verify that

$$\lambda_L(X_1, X_2) = \lambda_U(X_1, X_2) = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right).$$

Recall that t_ν denotes the distribution function of the Student t-distribution with ν degrees of freedom. Hint: You may use without proof that conditional on $X_1 = x$,

$$\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{X_2 - x\rho}{\sqrt{1-\rho^2}} \sim t_{\nu+1}.$$

Also, the t distribution is symmetric. Use Exercise 5.7.

Exercise 5.15:
Let $\lambda_L(C)$ and $\lambda_U(C)$ denote the coefficients of lower and upper tail dependence for a copula C, and recall that $\widehat{C}$ denotes the survival copula of C given by

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2).$$

Show that $\lambda_L(\widehat{C}) = \lambda_U(C)$ and $\lambda_U(\widehat{C}) = \lambda_L(C)$.

Exercise 5.16:
Recall the estimator for Spearman's ρ given in (5.1),

$$\hat{\rho}_S(X_1, X_2) = \frac{\frac{1}{n}\sum_{i=1}^n R_{i,1}R_{i,2} - \overline{R}_1\overline{R}_2}{\sigma_1\sigma_2}.$$

In this exercise, we prove Proposition 5.3.8 in the following steps.

1)Argue why we may think of $R_{i,j}$ as being distributed like a discrete uniform random variable U on the set $\{1, ..., n\}$. Argue that $\overline{R}_j = E[U]$ and $\sigma_j^2 = \mathrm{Var}[U]$ for $j = 1, 2$.

2)Show that

$$E[U] = \frac{n+1}{2} \quad \text{and} \quad \mathrm{Var}[U] = \frac{n^2-1}{12}.$$

3)Show the identity

$$\frac{1}{n}\sum_{i=1}^n R_{i,1}R_{i,2} - \overline{R}_1\overline{R}_2 = \sigma_1\sigma_2 - \frac{1}{2n}\sum_{i=1}^n d_i^2,$$

where we recall that $d_i := R_{i,1} - R_{i,2}$.

4)Finish the proof of the proposition.

Exercise 5.17:
Consider the copula $C = (M + W)/2$. Show that this copula does not have convex level curves. Is C Archimedean?

Exercise 5.18:
Show that for a non-strict Archimedean copula with generator φ (see Exercise 4.27), the C-measure of the zero curve $\varphi(u_1) + \varphi(u_2) = \varphi(0)$ equals

$$-\frac{\varphi(0)}{\varphi'(0+)}.$$

In particular, if $\varphi'(0+) = -\infty$, the C-measure is zero. What is the C-measure of the zero curve when C is the copula in Example 5.6.6? Hint: Study the proof of Theorem 5.6.4. What changes when φ is not strict?

In the following four exercises (all taken from [91]), we investigate the bounds presented in Corollary 5.3.11, describing the relation between Kendall's τ and Spearman's ρ.

Exercise 5.19:
Let $U_1 \sim \text{Unif}(0,1)$ and $U_2 := U_1 \oplus \theta$, where $\oplus$ denotes addition modulo 1. Here $\theta \in (0,1)$ is a parameter.

1)Show that the copula of (U_1, U_2) is

$$C_\theta(u_1,u_2) = \begin{cases} \min\{u_1, u_2 - \theta\}, & (u_1,u_2) \in [0, 1-\theta] \times [\theta, 1] \\ \min\{u_1 + \theta - 1, u_2\}, & (u_1,u_2) \in [1-\theta, 1] \times [0, \theta] \\ W(u_1,u_2), & \text{otherwise} \end{cases}$$

and illustrate the support of C_θ.

2)Prove that for this copula,

$$\rho_\tau = (1-2\theta)^2$$

3)Show that

$$\rho_S = 1 - 6\theta(1-\theta).$$

4)Conclude that

$$\rho_S = \frac{3\rho_\tau - 1}{2}, \quad \rho_\tau \geq 0,$$

so that the left linear segment in Figure 5.3 is actually attainable for some pair of random variables.

Exercise 5.20:
Let $\theta \in (0,1)$ and consider the copula

$$C_\theta(u_1,u_2) = \begin{cases} \max\{0, u_1 + u_2 - \theta\}, & (u_1,u_2) \in [0,\theta]^2 \\ \max\{\theta, u_1 + u_2 - 1\}, & (u_1,u_2) \in (\theta,1]^2 \\ M(u_1,u_2), & \text{otherwise.} \end{cases}$$

This copula is constructed in Example 3.4 in [91] and was also considered in Exercise 4.34. It has the interpretation that (U_1, U_2) is uniformly distributed on the two line segments joining $(0,\theta)$ to $(\theta,0)$ with mass θ and $(\theta,1)$ to $(1,\theta)$ with mass $1-\theta$. One can verify that $U_1 \oplus U_2 = \theta$, $\oplus$ denoting addition modulo 1.

1)Show that

$$\rho_\tau = -(1-2\theta)^2.$$

2)Show that

$$\rho_S = 6\theta(1-\theta) - 1.$$

3)Verify that

$$\rho_S = \frac{1+3\rho_\tau}{2}, \quad \rho_\tau \leq 0,$$

so that the right linear segment in Figure 5.3 is actually attainable for some pair of random variables.

Exercise 5.21:
Let $n \in \mathbb{N}$ and define the copula C_n as the ordinal sum (see Exercise 4.34) of n copies of W with respect to the equidistant partition of $[0,1]$ into n subintervals, that is,

$$C_n(u_1,u_2) = \begin{cases} \max\left\{\frac{k-1}{n}, u_1+u_2-\frac{k}{n}\right\}, & (u_1,u_2) \in \left[\frac{k-1}{n}, \frac{k}{n}\right]^2, \quad k = 1,2,...,n \\ \min\{u_1,u_2\}, & \text{otherwise.} \end{cases}$$

The support of C_n consists of n line segments connecting the points $((k-1)/n, k/n)$ and $(k/n, (k-1)/n)$ for $k = 1,2,...,n$. This is illustrated in the figure below for $n = 4$.

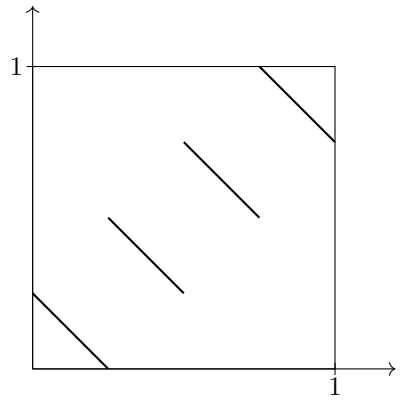


Figure 9: The support of the copula C_4 in Exercise 5.21.

1)Show that Kendall's τ for this copula is given by

$$\rho_\tau = 1 - \frac{2}{n}$$

2)Show that Spearman's ρ for this copula is given by

$$\rho_S = 1 - \frac{2}{n^2}.$$

3)Verify that

$$\rho_S = \frac{1 + 2\rho_\tau - \rho_\tau^2}{2}$$

and conclude that the upper bound in Corollary 5.3.11 for $\rho_\tau \geq 0$ is attainable for some pair of random variables.

Exercise 5.22:
Let $n \in \mathbb{N}$ and define the copula C_n by

$$C_n(u_1, u_2) = \begin{cases} \min\left\{u_1 - \frac{k-1}{n}, u_2 - \frac{n-k}{n}\right\}, & (u_1, u_2) \in \left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{n-k+1}{n}, \frac{n-k}{n}\right], \\ & k = 1, 2, ..., n \\ \max\{u_1 + u_2 - 1, 0\}, & \text{otherwise.} \end{cases}$$

The support of C_n consists of n line segments connecting the points $((k-1)/n, (n-k)/n)$ and $(k/n, (n-k+1)/n)$, $k = 1, 2, ..., n$. This is illustrated in the figure below for $n = 4$.

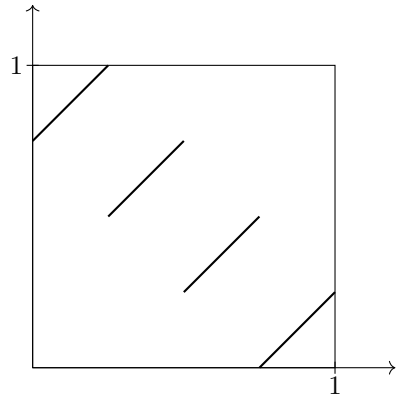


Figure 10: The support of the copula C_4 in Exercise 5.22.

1)Show that Kendall's τ for this copula is given by

$$\rho_\tau = \frac{2}{n} - 1.$$

2)Show that Spearman's ρ for this copula is given by

$$\rho_S = \frac{2}{n^2} - 1.$$

3)Verify that

$$\rho_S = \frac{\rho_\tau^2 + 2\rho_\tau - 1}{2}$$

and conclude that the lower bound of Corollary 5.3.11 for $\rho_\tau \leq 0$ is attainable for some pair of random variables.

Exercise 5.23:
Recall from Exercise 4.39 the Plackett copula

$$C_\theta(u_1, u_2) = \frac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{(1 + (\theta - 1)(u_1 + u_2))^2 - 4u_1u_2\theta(\theta - 1)}}{2(\theta - 1)}$$

for $\theta \in (0, \infty) \setminus \{1\}$ and $C_1 = \Pi$. Show that Spearman's ρ for C_θ is

$$\rho_S = \frac{\theta + 1}{\theta - 1} - \frac{2\theta}{(\theta - 1)^2} \log \theta.$$

According to [91], no closed form expression for Kendall's τ exists for this family.

Exercise 5.24:
In this exercise, we derive expressions for Kendall's τ and Spearman's ρ for ordinal sums of copulas. It is highly recommended that you take a look at Exercise 4.34 if you have not done so yet.

Let C_1 and C_2 be copulas and let $\rho_\tau^{(1)}, \rho_\tau^{(2)}, \rho_S^{(1)}, \rho_S^{(2)}$ denote Kendall's τ and Spearman's ρ for these. Let C_θ be the ordinal sum of $\{C_2, C_1\}$ with respect to $\{[0, \theta], [\theta, 1]\}$ for $\theta \in [0, 1]$ (note the ordering!). Let ρ_τ and ρ_S denote Kendall's τ and Spearman's ρ for C_θ.

1)Show that

$$\rho_\tau = \theta^2 \rho_\tau^{(2)} + (1 - \theta)^2 \rho_\tau^{(1)} + 2\theta(1 - \theta).$$

2)Show that

$$\rho_S = \theta^3 \rho_S^{(2)} + (1 - \theta)^3 \rho_S^{(1)} + 3\theta(1 - \theta).$$

Exercise 5.25:
Consider a copula of the form

$$C(u_1, u_2) = u_1u_2 + u_1u_2(1-u_1)(1-u_2)\Big(A_1u_2(1-u_1) + A_2(1-u_1)(1-u_2) + B_1u_1u_2 + B_2u_1(1-u_2)\Big)$$

for suitable constants A_1, A_2, B_1, B_2 satisfying certain constraints. A copula of this form is called a *copula with cubic sections*, see Section 3.2.5 in [91] for details.

1)Prove that Kendall's τ for such a copula satisfies

$$\rho_\tau = \frac{A_1 + A_2 + B_1 + B_2}{18} + \frac{A_2B_1 - A_1B_2}{450}.$$

2)Prove that Spearman's ρ for such a copula satisfies

$$\rho_S = \frac{A_1 + A_2 + B_1 + B_2}{12}.$$

The following two exercises concern another measure of concordance.

Exercise 5.26: Blomqvist's β
In this exercise we introduce another concordance concept known as the *medial correlation coefficient* or *Blomqvist's* β, see [10]. For a pair (X_1, X_2) of random variables, it is defined by

$$\beta(X_1, X_2) := P((X_1 - M[X_1])(X_2 - M[X_2]) > 0) - P((X_1 - M[X_1])(X_2 - M[X_2]) < 0)$$

where $M[X_j]$ is a median of X_j. Assume in the following that (X_1, X_2) is continuous with joint distribution function F.

1)Show that

$$\beta(X_1, X_2) = 4F(M[X_1], M[X_2]) - 1.$$

2)If the copula of (X_1, X_2) is C, show that

$$\beta(X_1, X_2) = 4C\Big(\frac{1}{2}, \frac{1}{2}\Big) - 1.$$

A strength of Blomqvist's β is that it can be used to approximate Kendall's τ and Spearman's ρ.

3)Consider the Ali–Mikhail–Haq copula from Exercise 5.11 above,

$$C_\theta(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}$$

where $\theta \in [-1, 1)$. Show that for this copula,

$$\beta = \frac{\theta}{4 - \theta} \in \Big[-\frac{1}{5}, \frac{1}{3}\Big],$$

and that we may reparametrise the copula in terms of β via $\theta = 4\beta/(1 + \beta)$.

Using this reparametrisation along with some Taylor calculations, one can show using the expressions for ρ_τ and ρ_S in Exercise 5.11 above that

$$\rho_\tau = \frac{8}{9}\beta + \frac{8}{15}\beta^3 + \frac{16}{45}\beta^4 + \cdots,$$
$$\rho_S = \frac{4}{3}\beta + \frac{44}{75}\beta^3 + \frac{8}{25}\beta^4 + \cdots.$$

Since Blomqvist's β is very easy to compute for this copula and is in $[-1/5, 1/3]$, these formulas provide very precise approximations to ρ_τ and ρ_S.

Exercise 5.27:
Consider again the Plackett family from Exercise 4.39 and Exercise 5.23 above,

$$C_\theta(u_1, u_2) = \frac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{(1 + (\theta - 1)(u_1 + u_2))^2 - 4u_1 u_2 \theta(\theta - 1)}}{2(\theta - 1)}$$

for $\theta \in (0, \infty) \setminus \{1\}$ and $C_1 = \Pi$.

1)Show that Blomqvist's β for this copula is given by

$$\beta = \frac{\sqrt{\theta} - 1}{\sqrt{\theta} + 1}.$$

2)Show that a second-order approximation to ρ_S (see the result of Exercise 5.23) in terms of β is given by

$$\rho_S \approx \frac{4\beta}{3}.$$

Note: When θ is considered as an odds ratio for a 2×2 contingency table as in the construction of this copula, the expression for β is called *Yule's Y* or *Yule's coefficient of colligation*.

Exercise 5.28: Dependence concepts for jointly symmetric copulas
Recall the concept of joint symmetry for a pair of random variables from Exercise 4.31. Show that if (X_1, X_2) is a pair of continuous random variables whose copula C satisfies either of the two identities for joint symmetry (Exercise 4.31 2)), then

$$\rho_\tau = \rho_S = \beta = 0.$$

This shows that joint symmetry is indeed a very strong requirement!

Exercise 5.29:
Let C be an extreme value copula with dependence function A (see Theorem 4.5.8).

1)Prove that

$$\lambda_L = \begin{cases} 0, & A(1/2) > 1/2 \\ 1, & A(1/2) = 1/2. \end{cases}$$

2)Prove that

$$\lambda_U = 2(1 - A(1/2)).$$

Exercise 5.30:
Let C_1 and C_2 be copulas that are both absolutely continuous, so that

$$\frac{\partial^2 C_1(u_1, u_2)}{\partial u_1 \partial u_2} \quad \text{and} \quad \frac{\partial^2 C_2(u_1, u_2)}{\partial u_1 \partial u_2}$$

both exist. Using integration by parts, prove the result

$$\int_{[0,1]^2} C_1(u_1, u_2) dC_2(u_1, u_2) = \frac{1}{2} - \int_{[0,1]^2} \frac{\partial}{\partial u_1} C_1(u_1, u_2) \frac{\partial}{\partial u_2} C_2(u_1, u_2) du_1 du_2.$$

This result is useful for computing Kendall's τ as is illustrated in the following exercise.

Note: The right hand side of the formula only involves first derivatives of the two copulas, which may lead one to conjecture that the formula should still hold even when the two copulas are not absolutely continuous. This is indeed true. We refer to [83] for a proof.

Exercise 5.31:
Consider the Marshall–Olkin family of copulas

$$C_{\alpha,\beta}(u_1,u_2)=\begin{cases}u_1^{1-\alpha}u_2, & u_1^\alpha \geq u_2^\beta\\ u_1u_2^{1-\beta}, & u_1^\alpha \leq u_2^\beta,\end{cases}$$

where $0<\alpha,\beta<1$ are parameters. For background on this copula, consult the supplementary section of the previous chapter along with Exercise 4.38.

1)Verify that

$$\frac{\partial}{\partial u_1}C_{\alpha,\beta}(u_1,u_2)\frac{\partial}{\partial u_2}C_{\alpha,\beta}(u_1,u_2)=\begin{cases}(1-\alpha)u_1^{1-2\alpha}u_2, & u_1^\alpha > u_2^\beta\\ (1-\beta)u_1u_2^{1-2\beta}, & u_1^\alpha < u_2^\beta.\end{cases}$$

2)Prove that Kendall's τ is given by

$$\rho_\tau=\frac{\alpha\beta}{\alpha+\beta-\alpha\beta}$$

by using the result of the previous exercise.

3)Show that Spearman's ρ is given by

$$\rho_S=\frac{3\alpha\beta}{2\alpha+2\beta-\alpha\beta}.$$

4)Compute λ_L and λ_U for this copula. Hint: Use the result of Exercise 5.29.

Practical exercises

Exercise 5.32:
Consider the Frank copula C_θ with generator

$$\varphi(t)=-\log\Big(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\Big),\quad \theta\neq 0.$$

1)Show that the log-likelihood for $d=2$ based on the sample $\mathbf{U}_1,...,\mathbf{U}_n$ is given by

$$\begin{aligned}\log c_\theta(\mathbf{U}_1,...,\mathbf{U}_n)&=n\log(\theta(1-e^{-\theta}))-\theta\sum_{t=1}^n(U_{t,1}+U_{t,2})\\&\quad-\sum_{t=1}^n\log\Big(\Big(e^{-\theta(U_{t,1}+U_{t,2})}-e^{-\theta U_{t,1}}-e^{-\theta U_{t,2}}+e^{-\theta}\Big)^2\Big)\end{aligned}$$

2)Show that the following algorithm for simulating from the Frank copula is correct.

Algorithm 5.8.1. To simulate a sample from the two-dimensional Frank copula with $\theta\neq 0$, do the following.

1. Simulate independent $U_1, V \sim \text{Unif}(0,1)$.
2. Set
$$U_2 \leftarrow -\frac{1}{\theta}\log\Big(1+\frac{V(e^{-\theta}-1)}{V+e^{-\theta U_1}(1-V)}\Big).$$
3. The pair (U_1, U_2) is then a sample from the two-dimensional Frank copula.

Hint: Write the copula C_θ explicitly and use Algorithm 4.2.1.

3)Now simulate some samples of the Frank copula and maximise the resulting log-likelihood to obtain an estimator for θ. Compare with the true value. Try varying θ and the number of samples. Compare with the estimates from an `R` package such as `copula`.

Exercise 5.33:
Consider the Gumbel copula C_θ with generator

$$\varphi(t) = (-\log t)^\theta,$$

where $\theta \geq 1$ is a parameter. We have considered this copula many times already.

1)Show that the log-likelihood for the two-dimensional Gumbel copula based on the sample $\mathbf{U}_1, ..., \mathbf{U}_n$ is given by

$$\begin{aligned}\log c_\theta(\mathbf{U}_1, ..., \mathbf{U}_n) &= -\sum_{t=1}^n((-\log U_{t,1})^\theta + (-\log U_{t,2})^\theta)^{1/\theta} \\ &+ \sum_{t=1}^n \log\Big((\theta-1)((-\log U_{t,1})^\theta + (-\log U_{t,2})^\theta)^{1/\theta-2} + ((-\log U_{t,1})^\theta + (-\log U_{t,2})^\theta)^{2/\theta-2}\Big) \\ &- \sum_{t=1}^n(\log U_{t,1} + \log U_{t,2}) + (\theta-1)\sum_{t=1}^n \log((\log U_{t,1})(\log U_{t,2})).\end{aligned}$$

2)Use your algorithm from Exercise 4.45 to simulate some samples from the Gumbel copula for different values of θ and different sample sizes. Compute the maximum likelihood estimate of θ and compare with the true value. Compare with the estimates from an `R` package such as `copula`.

Exercise 5.34:
Suppose $(X_1, ..., X_{50})$ are the daily log-returns for 50 different stocks and suppose that today's stock prices are given by $(S_1, ..., S_{50}) = (100, ..., 100)$. We hold one share of each stock. Hence the portfolio loss over one day is

$$L = -\sum_{i=1}^{50} S_i(e^{X_i} - 1).$$

Suppose $\rho_\tau(X_i, X_j) = 0.4$ for all $i \neq j$. We also assume that, for every i,

$$X_i \overset{\text{d}}{=} \frac{1}{100\sqrt{3}}Z$$

for $Z \sim t_3$.

1)Estimate $\text{VaR}_{0.99}(L)$ and $\text{ES}_{0.99}(L)$ under the assumption that $(X_1, ..., X_{50})$ has a Gaussian copula.

2)Repeat under the assumption that $(X_1, ..., X_{50})$ has a Clayton copula. Hint: Recall that $\rho_\tau = \theta/(\theta + 2)$.

3)Do you see differences in your estimates? If so, explain why.

The following exercises expand upon the case study on Novo Nordisk, Eli Lilly and Pfizer stocks presented in the text.

Exercise 5.35:
We first investigate other copulas than the four used above (Gaussian, t, Clayton and Gumbel).

1)Start by using the code in the case study to load the raw data and compute the log returns.

2)Now fit, using maximum likelihood and the `copula` package, the Frank, Joe, Ali–Mikhail–Haq and the Plackett (see Exercise 4.39) copulas to the three pairs of stocks. Hint: Use the functions `frankCopula`, `joeCopula`, `amhCopula` and `plackettCopula`.

3)Compare the model fits to the results in the text. Can you beat the t copula?

Exercise 5.36:
We now investigate the impact of different marginal distributions. In the case study in the text, we only considered the purely nonparametric approach (ii). Here we consider the alternatives (i) and (iii).

1)Suggest parametric distributions for each of the three log returns. Try to be creative!

2)Using your parametric distributions from 1), redo the analysis from the text. Do the results and conclusions change?

3)If you have done the previous exercise, redo the analysis with these new marginal distributions.

We are now done investigating option (i). For the rest of the exercise, we consider approach (iii).

4)Using methods from extreme value theory, for each of the three stocks, determine an upper limit u such that the excesses above this threshold are approximately Generalised Pareto distributed. Let the "body" of the distributions be modelled nonparametrically.

5)Using these semiparametric distributions, redo the analysis from the texts. Do the results and conclusions change compared to the two previous approaches?

6)Again, if you have done the previous exercise, redo the analysis there with the new semiparametric distributions.

Exercise 5.37:
We again build on top of the case study in the text and the previous exercises. Let S_1 denote Novo Nordisk, S_2 Eli Lilly and S_3 Pfizer. For any pair (S_i, S_j), $i \neq j$, of stocks, we consider a portfolio where we invest 1000\$ in each stock. The portfolio loss over one day is thus

$$L = -1000(e^{X_i} - 1) - 1000(e^{X_j} - 1) \tag{5.3}$$

with X_i and X_j denoting the corresponding log returns. Estimate the VaR and ES for as many combinations of copulas and marginal distributions as you want from the case study and the previous exercises. Summarise your findings.

Hint: To simulate from a copula using the `copula` package, see the function `rCopula`.

Exercise 5.38:
Finally, we investigate how the analysis turns out if we only consider a single portfolio consisting of an initial investment of 1000$ in each stock so that the one day loss is

$$L = -1000(e^{X_1} - 1) - 1000(e^{X_2} - 1) - 1000(e^{X_3} - 1). \tag{5.4}$$

1)Generate pseudo-samples of a three-dimensional copula based on the different marginal distributions from the case study in the text and the previous exercises.

2)Now propose different three-dimensional copulas to fit these pseudo-samples. What copula fits the best? Does the best copula vary over the pseudo-samples?

3)Finally, estimate the VaR and ES for the loss of the portfolio.

Chapter 6

Credit risk

6.1 Portfolio credit risk

Introduction and setup

Credit risk is the risk associated with loans and other obligations, more precisely, the risk that a financial party is not able to pay what it owes to another party (for a loan, this is called *default*). Simply put, we have a situation where a financial institution (such as a bank) called the *lender* lends money to another party called the *obligor*.

$$\text{Lendor} \longrightarrow \text{Obligor}$$

In this course, we consider a very simple case. We assume the following:

- We have a one-period model. Namely, we will consider the loss that occurs over a single timestep such as a year, a month etc.
- We have n total loans.
- We have a probability of default p_i for loan i. p_i will depend on external factors, and these should be incorporated into the model.

Definition 6.1.1. The total one-period loss (for the bank, say) is given by

$$L = \sum_{i=1}^{n} X_i L_i (1 - \lambda_i)$$

where L_i is the size of the ith loan, λ_i is the *recovery rate* for the ith loan and

$$X_i = \begin{cases} 1, & \text{if default} \\ 0, & \text{if no default} \end{cases}.$$

The recovery rate λ_i is a number between $[0, 1]$ that describes how much the bank can recover in case of default. If, for example, a third of the loan is already repaid by the time of default ($\lambda_i = 1/3$), the bank will only lose two thirds of the loan. Before considering concrete models for credit risk, we make two important remarks.

(i) Defaults will be dependent since defaults are often triggered by external factors, for example increases in interest rates.

(ii) Large losses are usually not caused by one default but by a large number of defaults, even though the individual losses are often not large.

We will study the following models:

- The Merton model/KMV model.
- The Probit normal mixture model.
- The Bernoulli mixture model.
- The Poisson mixture model.

The Merton model is an example of a *structural model*. The three others are examples of *reduced form models*.

The Merton model

Merton's model considers a firm as an obligor. The model consists of a process V_A in continuous time describing the total value of the assets of the firm by time t and a fixed number K called the debt to be paid at time T. According to the model, V_A behaves like the assets in the Black–Scholes model,

$$dV_A(t) = \mu_A V_A(t)dt + \sigma_A V_A(t)dW(t)$$

where W is a standard Brownian motion. The reader can consult the rundown of the Black–Scholes model in Chapter 1. Since V_A is a Geometric Brownian motion, we have the explicit solution for the assets at time T in terms of the value at the current time t, namely

$$V_A(T) = V_A(t)e^{\left(\mu_A - \frac{\sigma_A^2}{2}\right)(T-t)+\sigma_A(W(T)-W(t))}.$$

Note that $Z := W(T) - W(t) \sim \mathcal{N}(0, T-t)$. Default in this model means by definition that $V_A(T) < K$. We can rewrite this as follows:

$$\begin{aligned} V_A(T) < K \quad &\Leftrightarrow \quad \log V_A(t) + \left(\mu_A - \frac{\sigma_A^2}{2}\right)(T-t) + \sigma_A Z < \log K \\ &\Leftrightarrow \quad \frac{\log K - \log V_A(t) + \left(\frac{\sigma_A^2}{2} - \mu_A\right)(T-t)}{\sigma_A} > Z \\ &\Leftrightarrow \quad \frac{\log K - \log V_A(t) + \left(\frac{\sigma_A^2}{2} - \mu_A\right)(T-t)}{\sigma_A\sqrt{T-t}} > Y \end{aligned}$$

where $Y := Z/\sqrt{T-t} \sim \mathcal{N}(0,1)$. Note that in the above expression, Y is the only source of randomness since the value $V_A(t)$ is known at time t. We summarise our findings in the following definition.

Definition 6.1.2. In the above setup of Merton's model, the quantity

$$DD := -\frac{\log K - \log V_A(t) + \left(\frac{\sigma_A^2}{2} - \mu_A\right)(T-t)}{\sigma_A\sqrt{T-t}}$$

is called the *distance to default*. The probability of default can thus be written

$$P(\text{Default}) = P(Y < -DD)$$

with $Y \sim \mathcal{N}(0,1)$.

To estimate $P(\text{Default})$, we need to estimate DD. One problem is that V_A is not observable. It is hard to put a number on the value of every asset of a company (buildings, furniture, machines, patents etc.). However, the *equity* V_E is observable. We define the equity at time T to be

$$V_E(T) = (V_A(T) - K)^+$$

since if $K > V_A(T)$ the company defaults and the equity is zero. We recognise the above as a call option with strike K. If we also assume a constant interest rate r, we can apply the Black–Scholes formula,

$$V_E(t) = V_A(t)\Phi(z) - Ke^{-r(T-t)}\Phi(y)$$

where

$$z = \frac{\log V_A(t) - \log K + \left(r + \frac{\sigma_A^2}{2}\right)(T-t)}{\sigma_A\sqrt{T-t}}, \quad y = z - \sigma_A\sqrt{T-t}.$$

This formula relates V_A to V_E. A specific form of the Merton model used in the industry is the *KMV model*. In this model, one estimates the volatility of V_E, $\sigma_E = g(V_A(t), \sigma_A, r)$. Given μ_A and σ_A, the firm estimates the DD and hence the probability of default. All these estimates may be very unreliable and so in the actual KMV procedure, further steps are taken, namely:

- Consider n other firms.
- Calculate DD_i, $i = 1, ..., n$, where DD_i is the distance to default for firm i.
- Compare with past empirical observations with the same DD_i. Use the empirical frequency of default from these past loans and compare with the predictions from the estimates.

The description of the KMV model above is vague on purpose. Since it is an industry model, the exact procedures are not public and may change from firm to firm.

The Merton model we have considered so far only involves one firm, but the model is easily extended to an arbitrary number of firms.

The multivariate Merton model

In the multivariate Merton model, we consider n firms with asset processes

$$dV_{A,i}(t) = \mu_{A,i}V_{A,i}(t)dt + V_{A,i}(t)\sum_{j=1}^{m}\sigma_{A,i,j}dW_j(t), \quad i = 1, ..., n$$

with m independent Brownian motions $W_1, ..., W_m$. Note that all the Brownian motions appear in all the asset processes. This makes the $V_{A,i}$ dependent. One can think of the Brownian motions as underlying risk factors (for example fluctuations in interest rates). Letting $\sigma_{A,i}^2 = \sum_{j=1}^{m}\sigma_{A,i,j}^2$, we can explicitly solve for each $V_{A,i}$ and obtain

$$V_{A,i}(T) = V_{A,i}(t)e^{\left(\mu_{A,i}-\frac{\sigma_{A,i}^2}{2}\right)(T-t)+\sum_{j=1}^{m}\sigma_{A,i,j}(W_j(T)-W_j(t))}$$

which is very reminiscent of the univariate case. Letting

$$Z_i = \sum_{j=1}^{m}\sigma_{A,i,j}(W_j(T) - W_j(t)),$$

we have $Z_i \sim \mathcal{N}(0, \sigma_{A,i}^2(T-t))$. Like in the one-dimensional case, we can solve for Z_i and get

$$Z_i = \log V_{A,i}(T) - \log V_{A,i}(t) + \left(\frac{\sigma_{A,i}^2}{2} - \mu_{A,i}\right)(T-t).$$

The ith firm defaults if $V_{A,i}(T) < K_i$ where K_i is some threshold which we think of as the debt of the company. We can rewrite $V_{A,i}(T) < K_i$ as

$$Y_i := \frac{Z_i}{\sigma_{A,i}\sqrt{T-t}} < \frac{1}{\sigma_{A,i}\sqrt{T-t}}\left(\log K_i - \log V_{A,i}(t) + \left(\frac{\sigma_{A,i}^2}{2} - \mu_{A,i}\right)(T-t)\right) =: -DD_i$$

DD_i is the distance to default for the ith firm, and hence default for company i means $Y_i < -DD_i$. Since $Y_i \sim \mathcal{N}(0,1)$, we have

$$P(\text{Default for company } i) = P(Y_i < -DD_i) = \Phi(-DD_i).$$

Note that the Y_i are **dependent**. We want to describe this dependence. Note that we can write

$$Y_i = \frac{1}{\sigma_{A,i}\sqrt{T-t}}\sum_{j=1}^{m}\sigma_{A,i,j}(W_j(T) - W_j(t)) = \sum_{j=1}^{m}\frac{\sigma_{A,i,j}}{\sigma_{A,i}}\frac{W_j(T) - W_j(t)}{\sqrt{T-t}}$$

so the Y_i are weighted sums of the same iid $\mathcal{N}(0,1)$-variables. More specifically, we can observe that the Y_i have the form

$$Y_i = \sum_{j=1}^{m}c_{i,j}R_j$$

for constants $c_{i,j}$ specific to the ith loan and $R_1, ..., R_m$ iid $\mathcal{N}(0,1)$. This leads us to consider models for the Y_i called *factor models*. In a factor model, we assume that the Y_i have the form

$$Y_i = \sum_{j=1}^{k}a_{ij}U_j + b_i\mathcal{W}_i.$$

The U_j are common stochastic factors that affect all loans, and we assume that $U_1, ..., U_k$ are iid $\mathcal{N}(0,1)$. The variable $\mathcal{W}_i$ is a firm specific stochastic factor, and the a_{ij} and b_i are constants. In order to estimate the default probability, we will need estimates of a_{ij}, b_i and DD_i. This is not an easy task. These quantities are not observable from data, and hence we cannot apply ordinary statistical methods to estimate them. Another approach is usually to divide the loans into different "classes" where for a specific class, a_{ij} and b_i do not depend on i but only on the class. This means that in a specific class, we have

$$Y_i = \sum_{j=1}^{k} a_j U_j + b\mathcal{W}_i.$$

A standard approach at this point is to normalise the constants so that $\sum_{j=1}^{k} a_j^2 + b^2 = 1$. Then

$$\sum_{j=1}^{k} a_j U_j \sim \mathcal{N}(0, \|\mathbf{a}\|^2), \quad \mathbf{a} = (a_1, ..., a_k)$$

from which it follows that $Y_i \overset{\mathrm{d}}{=} \|\mathbf{a}\| Z + b\mathcal{W}_i$ for $Z \sim \mathcal{N}(0,1)$. Since $\|\mathbf{a}\|^2 + b^2 = 1$, we can write $\|\mathbf{a}\| = \sqrt{\rho}$ and $b = \sqrt{1-\rho}$ for some $\rho \in (0,1)$. Y_i can thus be written as

$$Y_i \overset{\mathrm{d}}{=} \sqrt{\rho} Z + \sqrt{1-\rho}\mathcal{W}_i.$$

A model with Y_i of this form is called a *probit normal mixture model*.

Estimating VaR in the probit normal mixture model

Let $Z, \mathcal{W}_1, ..., \mathcal{W}_n$ be iid $\mathcal{N}(0,1)$ and $Y_i = \sqrt{\rho} Z + \sqrt{1-\rho}\mathcal{W}_i$ for $\rho \in (0,1)$. Default for the ith loan means $Y_i < d_i$ for some threshold d_i (earlier denoted by $-DD_i$). Now let

$$X_i = \begin{cases} 1, & \text{if } Y_i < d_i \\ 0, & \text{if } Y_i \geq d_i \end{cases},$$

then $N_n = \sum_{i=1}^{n} X_i$ is the total number of defaults in the portfolio of loans. We focus on the number of defaults and not the total loss, and the goal is to compute/estimate the VaR. We carry out this computation in several steps. Define $p_i(Z) = P(X_i = 1 \mid Z)$ i.e. the probability that the ith loan defaults given Z. Then

$$\begin{aligned} p_i(Z) &= P(Y_i < d_i \mid Z) = P(\sqrt{\rho} Z + \sqrt{1-\rho}\mathcal{W}_i < d_i \mid Z) \\ &= P\left(\mathcal{W}_i < \frac{d_i - \sqrt{\rho} Z}{\sqrt{1-\rho}} \mid Z\right) = \Phi\left(\frac{d_i - \sqrt{\rho} Z}{\sqrt{1-\rho}}\right) \end{aligned}$$

since $\mathcal{W}_i \sim \mathcal{N}(0,1)$. As $Z \overset{\mathrm{d}}{=} -Z$, we can rewrite the above to

$$\Phi\left(\frac{d_i - \sqrt{\rho} Z}{\sqrt{1-\rho}}\right) = \Phi\left(\frac{d_i}{\sqrt{1-\rho}} + \frac{\sqrt{\rho}}{\sqrt{1-\rho}} Z\right) = \Phi(a_i + bZ)$$

where

$$a_i = \frac{d_i}{\sqrt{1-\rho}}, \quad b = \frac{\sqrt{\rho}}{\sqrt{1-\rho}}.$$

Our goal is to compute VaR_α for N_n. To do so, we first compute VaR_α for $p_i(Z)$. This amounts to solving the equation $1-\alpha = P(p_i(Z) \geq x_i)$ for x_i:

$$\begin{aligned} 1-\alpha = P(p_i(Z) \geq x_i) &= P(\Phi(a_i + bZ) \geq x_i) = P(a_i + bZ \geq \Phi^{-1}(x_i)) \\ &= P\left(Z \geq \frac{\Phi^{-1}(x_i) - a_i}{b}\right) = 1 - \Phi\left(\frac{\Phi^{-1}(x_i) - a_i}{b}\right) \end{aligned}$$

hence we need to solve

$$\alpha = \Phi\left(\frac{\Phi^{-1}(x_i) - a_i}{b}\right) \quad \Leftrightarrow \quad \Phi^{-1}(\alpha) = \frac{\Phi^{-1}(x_i) - a_i}{b} \quad \Leftrightarrow \quad a_i + b\Phi^{-1}(\alpha) = \Phi^{-1}(x_i)$$

which yields the final answer $x_i = \Phi(a_i + b\Phi^{-1}(\alpha))$. Hence

$$\text{VaR}_\alpha(p_i(Z)) = \Phi(a_i + b\Phi^{-1}(\alpha)) = \Phi\left(\frac{d_i}{\sqrt{1-\rho}} + \sqrt{\frac{\rho}{1-\rho}}\Phi^{-1}(\alpha)\right).$$

Let us make the simplifying assumption that $d_i = d$ for all i, i.e. that the distances to default are identical for all loans. This also implies that $p_i(Z)$ is the same for all i since the Y_i are iid conditional on Z. This allows us to write $p(Z)$ without the subscript i. Furthermore, we write $N_n(Z)$ to stress that N_n depends on Z. We claim that, conditional on Z,

$$\frac{N_n(Z)}{n} \xrightarrow{\text{P}} p(Z)$$

where $\xrightarrow{\text{P}}$ denotes convergence in probability. See the appendix for a review if necessary. To be precise, we claim that for every $\varepsilon > 0$,

$$P\left(\left|\frac{N_n(Z)}{n} - p(Z)\right| > \varepsilon \mid Z\right) \to 0$$

uniformly in Z. Let $\varepsilon > 0$ be given. We start by noting that

$$p(Z) = P(X_i = 1 \mid Z) = E[X_i \mid Z] = E\left[\frac{N_n(Z)}{n} \mid Z\right].$$

This comes from the fact that now,

$$X_i = \begin{cases} 1, & \text{if } Y_i < d \\ 0, & \text{if } Y_i \geq d \end{cases}$$

which implies that the X_i are identically distributed so that

$$E[X_1 \mid Z] = E[X_2 \mid Z] = \cdots = E[X_n \mid Z]$$

and thus

$$E[N_n(Z) \mid Z] = \sum_{j=1}^{n} E[X_j \mid Z] = nE[X_i \mid Z]$$

for any i. The rest of the proof is a straightforward application of Chebyshev's inequality. We get

$$\begin{aligned} P\left(\left|\frac{N_n(Z)}{n} - p(Z)\right| > \varepsilon \mid Z\right) &= P(|N_n(Z) - E[N_n(Z) \mid Z]| > n\varepsilon) \\ &\leq \frac{1}{n^2\varepsilon^2}\,\text{Var}(N_n(Z) \mid Z) = \frac{p(Z)(1-p(Z))}{n\varepsilon^2}. \end{aligned}$$

The final inequality is a consequence of the X_i being iid given Z (since this is true for the Y_i) and from observing that conditional on Z, $N_n(Z)$ has a binomial distribution with parameters n and $p(Z)$. The above converges to zero uniformly in Z which proves the claim. We are now close to the goal of providing a formula for estimating the VaR. Assume $f(x) := P(p(Z) > x)$ is continuous. Then we have for $\alpha \in (0,1)$ that

$$1 - \alpha = P(p(Z) > \mathrm{VaR}_\alpha(p(Z)))$$

and using the result $N_n(Z)/n \xrightarrow{\mathrm{P}} p(Z)$ conditional on Z, we have the approximation

$$1 - \alpha \approx P\left(\frac{N_n(Z)}{n} > \mathrm{VaR}_\alpha(p(Z))\right)$$

which yields the *Basel formula*

$$\mathrm{VaR}_\alpha(N_n(Z)) \approx n\,\mathrm{VaR}_\alpha(p(Z)) = n\Phi\left(\frac{d}{\sqrt{1-\rho}} + \sqrt{\frac{\rho}{1-\rho}}\Phi^{-1}(\alpha)\right).$$

The Bernoulli mixture model

In this model, we assume common factors $\mathbf{Z} = (Z_1, ..., Z_m)$. These could for example be economic factors such as interest rates. We again let $p_i(\mathbf{Z}) = P(X_i = 1 \mid \mathbf{Z})$ with $X_i = 1$ when company i defaults and $X_i = 0$ otherwise just like before. We assume that the defaults (i.e. the X_i) are independent given $\mathbf{Z}$. We again let $N_n = \sum_{i=1}^n X_i$ and note that the X_i are Bernoulli variables conditional on $\mathbf{Z}$ with success probability $p_i(\mathbf{Z})$. This also implies that N_n is binomial with parameters n and $p_i(\mathbf{Z})$ conditional on $\mathbf{Z}$.

We consider a simplified version of this model, namely a so-called *one-factor model*. In such a model, we assume that $\mathbf{Z}$ is one-dimensional (so we write Z instead of $\mathbf{Z}$) and that $p_i(Z) = p(Z)$ is the same for all i. Note that this model is a generalisation of the probit normal mixture model. Indeed, the probit normal mixture model has this exact setup but with a specific X_i, namely

$$X_i = \begin{cases} 1, & \text{if } \sqrt{\rho}Z + \sqrt{1-\rho}\mathcal{W}_i < d \\ 0, & \text{if } \sqrt{\rho}Z + \sqrt{1-\rho}\mathcal{W}_i \geq d \end{cases}$$

for $Z, \mathcal{W}_1, ..., \mathcal{W}_n$ iid and standard normal as we saw earlier. Returning to the one-factor Bernoulli mixture model, we have

$$P(N_n = k \mid Z) = \binom{n}{k} p(Z)^k (1 - p(Z))^{n-k}$$

as was also noted before. If Z has distribution function G, we can write

$$P(N_n = k) = \binom{n}{k} \int_{\mathbb{R}} p(z)^k (1 - p(z))^{n-k} dG(z).$$

What are choices of G and $p(Z)$ that make the above expression mathematically tractable? Consider the special case $Z \sim \mathrm{Beta}(a, b)$ and $p(Z) = Z$. This model is called the *Beta mixture model*. The density of Z is

$$g(z) = \frac{1}{\beta(a,b)} z^{a-1}(1-z)^{b-1}, \quad 0 \leq z \leq 1 \quad \text{where}$$

$$\beta(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

With these assumptions, we can compute $P(N_n = k)$ explicitly as follows:

$$\begin{aligned} P(N_n = k) &= \binom{n}{k}\int_0^1 z^k(1-z)^{n-k}g(z)dz = \binom{n}{k}\int_0^1 \frac{1}{\beta(a,b)}z^{k+a-1}(1-z)^{n-k+b-1}dz \\ &= \binom{n}{k}\frac{\beta(a+k,b+n-k)}{\beta(a,b)}. \end{aligned}$$

Since we now have an explicit expression for the density of N_n, we can compute all sorts of risk measures such as VaR, ES etc. explicitly as well. While this is a nice property of the model, we stress that the motivation for choosing $Z \sim \text{Beta}(a,b)$ and $p(Z) = Z$ is purely mathematical. Nevertheless, we continue our study of the model. How do we estimate a and b? While a and b are not directly observed from data, we can relate it to quantities that are observed. We know the number of defaults that have occured at a given time which allows us to estimate the probability of default p, at least in a portfolio of homogeneous loans (think for example of a portfolio consisting only of AAA rated loans, only B rated loans etc.). We can also estimate the linear correlation ρ_L of the X_i. We now determine the relations between these quantities and a and b in the Beta mixture model. We have

$$\begin{aligned} p = P(\text{Default}) = E[P(\text{Default} \mid Z)] = E[p(Z)] = E[Z] = \frac{a}{a+b}, \\ \rho_L = \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i)} = \frac{E[(X_i - p)(X_j - p)]}{p(1-p)} = \frac{E[X_iX_j] - 2pE[X_i] + p^2}{p(1-p)} \end{aligned}$$

and

$$\begin{aligned} E[X_iX_j] &= P(X_iX_j = 1) = P(X_i = 1, X_j = 1) = E[P(X_i = 1, X_j = 1 \mid Z)] \\ &= E[P(X_i = 1 \mid Z)P(X_j = 1 \mid Z)] = E[Z^2] = \frac{a(a+1)}{(a+b)(a+b+1)}. \end{aligned}$$

We hence have two equations in two unknowns. Solving these yields the relations

$$a = (1-p)\frac{1-\rho_L}{\rho_L}, \quad b = p\frac{1-\rho_L}{\rho_L}.$$

To summarise: We can estimate a and b from data by first estimating the default probability and the linear correlation from the data and then apply the formulas above.

The Poisson mixture model

The Poisson mixture model is different from the previous models in the sense that the X_i can attain infinitely many values, namely $X_i \in \{0, 1, ...\}$. We assume that the X_i are independent given $\mathbf{Z}$ and that they have a Poisson distribution conditionally on $\mathbf{Z}$, namely

$$P(X_i = k \mid \mathbf{Z}) = \frac{(\lambda_i(\mathbf{Z}))^k}{k!}e^{-\lambda_i(\mathbf{Z})}$$

for some function λ_i. Just like earlier, $\mathbf{Z}$ is a collection of m variables which we interpret as some underlying factors. We think of $X_i = 1$ as a default of the ith loan and $X_i = 0$ as

no default. The other events, $X_i = k$ for $k \geq 2$, do not have a natural interpretation, but we choose λ_i small enough so that $X_i > 1$ happens with very low probability. We again consider $N_n = \sum_{i=1}^n X_i$, i.e. the number of defaults (at least for small λ_i). We now consider a special case of the Poisson mixture model which goes under the name CreditRisk$^+$. In CreditRisk$^+$, we assume the following:

- $Z_1, ..., Z_m$ are independent.
- $Z_j \sim \Gamma(\alpha_j, \beta_j^{-1})$ with $\alpha_j \beta_j = 1$.
- $\lambda_i(\mathbf{Z}) = \overline{\lambda}_i \sum_{j=1}^m a_{ij} Z_j$ with $a_{ij} \geq 0$ and $\sum_{j=1}^m a_{ij} = 1$. Here $\overline{\lambda}_i > 0$ is a constant for each i.

The assumption $\alpha_j \beta_j = 1$ implies that $E[Z_j] = 1$. With the above assumptions, we have $E[\lambda_i(\mathbf{Z})] = \overline{\lambda}_i$ which gives us control over the λ_i functions. We need to choose them small. With the above assumptions, it is possible to derive the distribution of N_n which is also a motivation for the model. In order to do so, we need to introduce some theory.

Definition 6.1.3. For a discrete random variable Y with values in $\{0, 1, ...\}$, the function

$$g_Y(t) = E[t^Y]$$

is called the *probability-generating function* of Y.

Remark 6.1.4. Notice the relation $g_Y(t) = E[e^{Y \log t}] = \kappa_Y(\log t)$ with κ_Y the moment-generating function of Y. This relation implies that the probability-generating function (if it exists) determines the distribution of Y.

Example 6.1.5. Let N be Poisson distributed with parameter $\lambda > 0$. Then

$$\kappa_N(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}.$$

Hence the probability-generating function is

$$g_N(t) = e^{\lambda(t-1)}.$$

∘

Example 6.1.6. Let N have a negative binomial distribution with parameters r and p i.e.

$$P(N = k) = \binom{k + r - 1}{k} (1 - p)^k p^r, \quad k = 0, 1, 2, ...$$

We leave it as an exercise for the reader to verify that the probability-generating function of N is given by

$$g_N(t) = \left(\frac{p}{1 - (1 - p)t} \right)^r \quad \text{for} \quad |t| < \frac{1}{1 - p}.$$

∘

Theorem 6.1.7. *With the assumptions of the CreditRisk$^+$ model, we have*

$$g_{N_n}(t) = \prod_{j=1}^m \left(\frac{1 - \delta_j}{1 - \delta_j t} \right)^{\alpha_j}, \quad \textit{where} \quad \delta_j = \frac{\beta_j \sum_{i=1}^n \overline{\lambda}_i a_{ij}}{1 + \beta_j \sum_{i=1}^n \overline{\lambda}_i a_{ij}}.$$

Proof. The proof is essentially a computational exercise in the assumptions of the CreditRisk$^+$ model. We start by conditioning on $\mathbf{Z}$. The conditional probability-generating function for N_n given $\mathbf{Z}$ is

$$\begin{aligned} g_{N_n|\mathbf{Z}}(t) &= E\left[t^{N_n} \mid \mathbf{Z}\right] = E\left[t^{X_1+\cdots+X_n} \mid \mathbf{Z}\right] = \prod_{i=1}^{n} E\left[t^{X_i} \mid \mathbf{Z}\right] \\ &= \prod_{i=1}^{n} \kappa_{X_i|\mathbf{Z}}(\log t) = \prod_{i=1}^{n} e^{\lambda_i(\mathbf{Z})(t-1)} \end{aligned}$$

where we used the conditional independence and the above example for the Poisson distribution. By applying the tower property, we can remove the conditioning on $\mathbf{Z}$ by taking an expectation. Let f_j denote the density of Z_j. Then we have by independence of the Z_j that

$$\begin{aligned} g_{N_n}(t) &= E\left[E\left[t^{N_n} \mid \mathbf{Z}\right]\right] = E\left[\prod_{i=1}^{n} e^{\lambda_i(\mathbf{Z})(t-1)}\right] \\ &= \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{n} e^{\lambda_i(\mathbf{z})(t-1)} f_1(z_1)\cdots f_m(z_m)dz_1\cdots dz_m \\ &= \int_0^\infty \cdots \int_0^\infty e^{(t-1)\sum_{i=1}^n \overline{\lambda}_i \sum_{j=1}^m a_{ij}z_j} f_1(z_1)\cdots f_m(z_m)dz_1\cdots dz_m \\ &= \int_0^\infty \cdots \int_0^\infty e^{(t-1)\sum_{j=1}^m \sum_{i=1}^n \overline{\lambda}_i a_{ij}z_j} f_1(z_1)\cdots f_m(z_m)dz_1\cdots dz_m \end{aligned}$$

For the sake of simplicity, let $\mu_j = \sum_{i=1}^n \overline{\lambda}_i a_{ij}$. Then we can continue the computation as follows:

$$\begin{aligned} g_{N_n}(t) &= \int_0^\infty \cdots \int_0^\infty e^{(t-1)\sum_{j=1}^m \mu_j z_j} f_1(z_1)\cdots f_m(z_m)dz_1\cdots dz_m \\ &= \int_0^\infty \cdots \int_0^\infty e^{(t-1)\mu_1 z_1} f(z_1)dz_1 \cdots e^{(t-1)\mu_m z_m} f(z_m)dz_m \\ &= \prod_{j=1}^{m} \int_0^\infty e^{(t-1)\mu_j z} f(z)dz = \prod_{j=1}^{m} \int_0^\infty e^{(t-1)\mu_j z} \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} z^{\alpha_j-1}e^{-z/\beta_j}dz \\ &= \prod_{j=1}^{m} \int_0^\infty \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} z^{\alpha_j-1}e^{-z(\beta_j^{-1}-(t-1)\mu_j)}dz. \end{aligned}$$

We now compute each integral (denoted by I_j) in the product:

$$\begin{aligned} I_j &:= \int_0^\infty \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} z^{\alpha_j-1}e^{-z(\beta_j^{-1}-(t-1)\mu_j)}dz \\ &= \frac{(\beta_j^{-1}-(t-1)\mu_j)^{-\alpha_j}\Gamma(\alpha_j)}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} \int_0^\infty \frac{(\beta_j^{-1}-(t-1)\mu_j)^{\alpha_j}}{\Gamma(\alpha_j)} z^{\alpha_j-1}e^{-z(\beta_j^{-1}-(t-1)\mu_j)}dz \\ &= \frac{1}{\beta_j^{\alpha_j}(\beta_j^{-1}-(t-1)\mu_j)^{\alpha_j}} = \frac{1}{(1-(t-1)\beta_j\mu_j)^{\alpha_j}} = \left(\frac{1-\delta_j}{1-\delta_j t}\right)^{\alpha_j} \end{aligned}$$

by noting that δ_j as defined in the theorem is given by

$$\delta_j = \frac{\beta_j \mu_j}{1 + \beta_j \mu_j}.$$

Plugging this expression back into the one for $g_{N_n}(t)$ completes the proof. ■

Remark 6.1.8. Note that the terms

$$\left(\frac{1 - \delta_j}{1 - \delta_j t} \right)^{\alpha_j}$$

are probability-generating functions for negative binomial variables with parameters $p = 1 - \delta_j$ and $r = \alpha_j$.

In principle, one can invert g_{N_n}. There is a whole litterature dedicated to inverting moment and probability-generating functions. We will not pursue this here. We will only mention that one can make a very crude approximation that relies on Markov's inequality, namely

$$P(N_n > k) \leq \frac{E[t^{N_n}]}{t^k} = \frac{g_{N_n}(t)}{t^k}$$

for every $t > 0$. One can then minimise this expression over t.

6.2 Operational risk

Operational risk can be stated as "loss from failed internal processes, people or systems or from external events". To elaborate a bit on this, we can roughly divide such risks in categories. One category is *repetitive human errors* or *repetitive operational risks* (repetitive OR). These include IT failures, errors in settlements of transactions, litigation and the like. Other types of losses include fraud and external events such as flooding, fires, earthquakes and terrorism (although the latter is extremely hard to model). A difficulty in operational risk is that we often have little data available, and the data is often heavy tailed. The claim arrivals can also be hard to model since they occur randomly in time (and frequently in clusters) and since the frequency changes over time. One can for example imagine that a large traded volume leads to a large number of back office errors.

Approaches in analyzing operational risk

We now discuss the basics of two approaches in analyzing operational risk. These are

- The *basic indicator approach* and
- The *advanced measurement approach*.

Under the basic indicator (BI) approach, the capital requirement to cover OR (operational risk) losses at time n is given by

$$\mathrm{RC}^n_{BI}(\mathrm{OR}) = \frac{1}{Z_n} \sum_{i=1}^{3} \alpha \max(\mathrm{GI}^{n-i}, 0),$$

where $\alpha \approx 0.15$ is a constant,

$$Z_n = \#\text{years where } \mathrm{GI}^{n-i} > 0, i = 1, 2, 3,$$

and GI^s denotes the "gross income" at time s. Under the advanced measurement (AM) approach, we divide into K lines of business (typically $K = 8$ and lines include for example corporate finance, trading and sales). The capital requirement to cover OR losses is given by

$$\mathrm{RC}^n_{AM}(\mathrm{OR}) = \sum_{b=1}^{K} \rho_\alpha(L^{n,b})$$

where $0.99 \leq \alpha \leq 0.999$ and ρ_α is a risk measure such as VaR_α or ES_α.

Mathematical estimates

In this subsection, we investigate methods to analyze the loss via a stochastic process approach. Let us start by recalling the definition of a Poisson process.

Definition 6.2.1. A stochastic process $\{N_t\}$ is called a *Poisson process* with *intensity* $\lambda > 0$ if N_t takes values in $\{0, 1, 2, ...\}$ and

(i) $P(N_h \geq 1) = \lambda h + o(h)$,

(ii) $P(N_h \geq 2) = o(h)$ and

(iii) $\{N_t\}$ has stationary and independent increments.

Recall that stationarity means that for every $s \leq t$, $N_t - N_s \overset{\mathrm{d}}{=} N_{t-s}$. By independent increments, we mean that for every finite partition $0 < t_1 < t_2 < \cdots < t_k$, the variables $\{N_{t_{i+1}} - N_{t_i}\}_{i=1}^{k-1}$ are independent. Intuitively, we think of a Poisson process as a claim number process which satisfies

$$P(1 \text{ claim in } [t, t+h]) = \lambda h + o(h), \quad P(\geq 2 \text{ claims in } [t, t+h]) = o(h).$$

We now model the loss of the company at time t by the process

$$L_t = \sum_{i=1}^{N_t} X_i$$

with $\{N_t\}$ a Poisson process with intensity λ independent of the iid sequence $\{X_i\}$. We discuss the following ideas based on risk theory:

- Laplace transform method.
- Panjer recursion.
- A sophisticated large deviation approach based on the "Arwedson approximation" from risk theory.
- Time-dependent intensity.
- Stochastic processes for market risk.

Let us first discuss the Laplace transform method. Recall that the Laplace transform of a random variable Y is given by $\psi_Y(s) = E[e^{-sY}]$. We can compute the Laplace transform of the loss using the tower property as follows:

$$\psi_{L_t}(s) = E\left[e^{-sL_t}\right] = E\left[E\left[e^{-s\sum_{i=1}^{N_t} X_i} \mid N_t\right]\right] = E\left[\psi_X(s)^{N_t}\right]$$

where we have defined $\psi_X(s) = E\left[e^{-sX_1}\right]$. We continue the computation and get

$$\psi_{L_t}(s) = \sum_{n=0}^{\infty}(\psi_X(s))^n \frac{\lambda^n}{n!}e^{-\lambda} = \sum_{n=0}^{\infty}\frac{(\psi_X(s)\lambda)^n}{n!}e^{-\lambda} = e^{\lambda(\psi_X(s)-1)}.$$

We should note that all the Laplace transforms exist since we are working with non-negative random variables. After obtaining the Laplace-transform, numerical inversion techniques can be applied. The method is more flexible than this however. To illustrate how, consider a Poisson intensity that changes over time. To make this concrete, assume we are considering the time interval $[0, 2]$, and we have Poisson processes N_1 and N_2 on $[0, 1]$ and $(1, 2]$, respectively, along with two claim sequences $\{X_i^{(1)}\}$ and $\{X_i^{(2)}\}$ belonging to each interval. The total loss is obtained by summing losses from each interval i.e. $L = L_1 + L_2$, where L_1 and L_2 are assumed to be independent. Then

$$\begin{aligned}\psi_L(s) &= E\left[e^{-s(L_1+L_2)}\right] = E\left[e^{-sL_1}\right]E\left[e^{-sL_2}\right] = \psi_{L_1}(s)\psi_{L_2}(s)\\ &= e^{-\lambda_1(\psi_{X^{(1)}}(s)-1)}e^{-\lambda_2(\psi_{X^{(2)}}(s)-1)},\end{aligned}$$

and we can invert this function to obtain the distribution. For those familiar with mixture distributions, the above Laplace transform is one of a compound Poisson sum

$$\sum_{i=1}^{\tilde{N}_t} Y_i, \quad t \in [0, 1]$$

with $\{\tilde{N}_t\}$ a Poisson process with intensity $\lambda_1 + \lambda_2$ and Y_i mixture distributed with distribution function

$$G(y) = \frac{\lambda_1}{\lambda_1+\lambda_2}G_1(y) + \frac{\lambda_2}{\lambda_1+\lambda_2}G_2(y)$$

where $X_1^{(1)} \sim G_1$ and $X_1^{(2)} \sim G_2$. In practice, one often observes $E[N_t] < \mathrm{Var}(N_t)$, a phenomenon called *overdispersion*. To remedy this, one can use a *mixed Poisson process* $\{N_t\}$ defined by $N_t = \tilde{N}_{\Lambda t}$ where $\{\tilde{N}_t\}$ is a Poisson process with intensity 1 and $\Lambda > 0$ is a random variable.

Example 6.2.2. Choosing $\Lambda \sim \Gamma(\alpha, \beta)$ leads to the so-called *negative binomial process*. We let the reader verify that $N_t = \tilde{N}_{\Lambda t}$ is indeed negative binomial distributed. ∘

We now leave the world of Laplace transforms and discuss the next topic, namely *Panjer recursion*. For this technique to be applicable, assume N_t satisfies the recursive relationship

$$q_n := P(N_t = n) = \left(a + \frac{b}{n}\right)q_{n-1}, \quad n = 1, 2, ...$$

for constants a and b. We also assume that $X_i \in \{1, 2, ...\}$. Panjer recursion yields the exact recursive formula for $p_n = P(L_t = n)$:

$$p_n = \sum_{i=1}^{n} \left(a + \frac{bi}{n} \right) P(X_1 = i) p_{n-i}, \quad p_0 = q_0.$$

The assumption $X_i \in \{1, 2, ...\}$ is not as restrictive as it may seem. One can always scale the values as necessary.

We now consider the *Arwedson approximation*. For the loss process

$$L_t = \sum_{i=1}^{N_t} X_i,$$

we consider a "small" $\delta \in (0, 1)$ and the probability of a large loss over the small time interval $[0, \delta u]$,

$$\varphi_\delta(u) := P(L_t > u, \text{ some } 0 \le t \le \delta u).$$

Ultimately, we will choose $\delta u = 1$. Those familiar with classical ruin theory will immediately see the connection to the present setting. But one should note that this situation is slightly different since we have no premium payments, and we are working over a finite time interval $[0, \delta u]$ and not $[0, \infty)$. The Arwedson approximation was originally developed in the study of *finite-time* ruin theory. In ruin theory, one studies the *Cramér–Lundberg process*

$$C_t = u + ct - \sum_{i=1}^{N_t} X_i$$

with $u \ge 0$ the initial capital of the company, $c > 0$ a constant premium rate and $\{X_i\}$ the insurance claim sizes. Arwedson considered the finite-time ruin probabilities

$$\Psi_K(u) = P(C_t < 0, \text{ for some } 0 \le t \le Ku)$$

where $0 \le K < \infty$. Under classical Cramér–Lundberg assumptions, Arwedson showed that

$$\Psi_K(u) \sim \begin{cases} \frac{C_K}{\sqrt{u}} e^{-uI(K)}, & \text{if } K \le \rho \\ Ce^{-Ru}, & \text{if } K > \rho \end{cases}.$$

Note that the case $K > \rho$ corresponds to the ordinary Cramér–Lundberg estimate. $I(K)$ would nowadays be called the "large deviation rate function" which describes the exponential decay of a probability as $u \to \infty$. R solves the equation $\Lambda(R) := \log E\left[e^{R(C_1 - u)}\right] = 0$ (called the *adjustment coefficient* in ruin theory). Here, $I(K) > R$ for $K < \rho$ and $\rho = (\Lambda'(R))^{-1}$ (one can show that ρu is the "most likely" time of ruin).

Returning to $\varphi_\delta(u)$, if X_i has exponential moments (i.e. is "light tailed"), one can show that

$$\varphi_\delta(u) \sim \frac{\mathcal{C}(\delta)}{\sqrt{u}} e^{-uJ(\delta)} \quad \text{as} \quad u \to \infty$$

which has connections to Arwedson's original result as well as the exponentially shifted measure from our discussion of importance sampling. If X_1 is *subexponential* (see Exercise 2.20), for example if X_1 is regularly varying, one can prove that

$$\varphi_\delta(u) \sim Du\overline{F}_X(u(1 - \delta\mu)) \quad \text{as} \quad u \to \infty.$$

The proof relies on the concept of "one large jump" in heavy tailed ruin problems. A concept that should be familiar to someone who has studied ruin theory.

We now turn to the point of time-dependent intensities. It makes sense to assume that the intensity of N_t changes over time.

Example 6.2.3. Assume N_t has intensity λ_n in the interval $(n-1, n]$ where

$$\lambda_n = F(Z_n), \quad Z_n = cZ_{n-1} + \xi_n, \quad |c| < 1,$$

where $\{\xi_n\}$ is iid $\mathcal{N}(0,1)$ and Z_n is a so-called AR(1) process. Z_n could for example represent traded volume, which is linked to increases in operational risk. ◦

For time-dependent intensities of ARMA-type, similar "Arwedson" approximations can be derived. Namely, one can likewise show

$$\varphi_\delta(u) \sim Du\overline{F}_X(u(1-\delta\mu)) \quad \text{as} \quad u \to \infty.$$

Similar insurance-based methods are potentially useful for market risk. We end this chapter with a brief discussion on stochastic processes for market risk. Throughout the course, we only considered one-period models, and we assumed iid returns. Real-life data is not iid! Hence stochastic processes (time-dependent models) are called for. This is very complicated because multiple stochastic processes are usually dependent, and this is difficult to model, so most current research either considers dependent processes in one dimension or coordinatewise dependence in one-period models (but not both). A very classical model for dependence is the ARMA(p,q) model:

$$X_n - \sum_{j=1}^{p} \phi_j X_{n-j} = Z_n + \sum_{i=1}^{q} \theta_i Z_{n-i}, \quad t = 1, 2, ...$$

where $\{Z_n\}$ is an iid $\mathcal{N}(0,1)$ sequence and ϕ_j, θ_i constants. This is an example of a *time series model*. For "sufficiently small" ϕ_j and θ_i, we have that

$$X_n \xrightarrow{\text{d}} X$$

i.e. that X_n converges to a stationary distribution where X is normally distributed. In an evolution of log-returns of stocks, one typically observes the following:

- Log-returns contain many "large" values i.e. the data is heavy-tailed.
- Exceedances of high thresholds occur in clusters, i.e. we have dependence in the tails.
- While dependent, returns show little serial correlation.
- Absolute (or squared) returns show strong serial correlation.
- Volatility varies over time.

To address these issues, the GARCH models were introduced. The first of these models, the ARCH(1) model, was introduced by Engle, see [31]. In this model, the log-returns $\{R_n\}$ satisfy

$$R_n^2 = (\phi_0 + \phi_1 R_{n-1}^2) Z_n^2, \quad n = 1, 2, ...$$

where $\{Z_n\}$ is an iid $\mathcal{N}(0,1)$ sequence. A more complicated model is the GARCH(1,1) model introduced by Bollerslev, see [11]. Here the log-returns $\{R_n\}$ satisfy

$$R_n = \sigma_n Z_n, \quad n = 1, 2, ...$$

where $\{Z_n\}$ is iid $\mathcal{N}(0,1)$ and

$$\sigma_n^2 = \alpha_0 + \beta_1 \sigma_{n-1}^2 + \alpha_1 R_{n-1}^2 = \alpha_0 + \sigma_{n-1}^2(\beta_1 + \alpha_1 Z_{n-1}^2).$$

Both ARCH(1) and GARCH(1,1) are examples of *stochastic recursive sequences*. Namely,

$$V_n = A_n V_{n-1} + B_n, \quad n = 1, 2, ...$$

where

$$\begin{aligned} V_n &= R_n^2 \quad \text{for ARCH(1),} \\ V_n &= \sigma_n^2 \quad \text{for GARCH(1, 1).} \end{aligned}$$

Here, $\{(A_n, B_n) : n = 1, 2, ...\}$ is any iid sequence on $(0, \infty) \times \mathbb{R}$. Under certain reasonable conditions, $V_n \xrightarrow{\text{d}} V$, and it is natural to consider $P(V > u)$ for large u. One can apply renewal theoretic methods (such as those presented in the course SkadeStok) to obtain

$$P(V > u) \sim Cu^{-R} \quad \text{as } u \to \infty$$

where

$$\Lambda(R) = 0, \quad \Lambda(\xi) = \log E\left[e^{\xi \log A}\right].$$

This shows that Pareto tails characterise the decay rate. For more complex models (e.g. GARCH(p,q)), one needs to consider *matrix recursions*. This is currently an active research area.

Notes and comments

See section 10.3 in [88] for more on the Merton model. Section 10.1 gives an informal introduction to credit risk. The computation in the proof of Theorem 6.1.7 is inspired by the one in section 12.2 of [63]. For more information about Panjer recursion, we refer to [89], section 3.3.3. In the final discussion on operational risk, many tools from ruin theory were discussed. Ruin theory and related tools such as renewal theory were discussed in the course SkadeStok. See [82] for lecture notes from the last run of the course.

6.3 Exercises

Exercise 6.1:
Upgrade the statement $N_n(Z)/n \xrightarrow{\mathrm{P}} p(Z)$ to $N_n(Z)/n \to p(Z)$ almost surely (conditional on Z). Hint: Use the Markov inequality and Proposition C.1.8. The fourth central moment of the binomial distribution with parameters n and p is given by

$$np(1-p)(1+(3n-6)p(1-p)).$$

Exercise 6.2:
Verify that the probability-generating function for a negative binomial variable N with parameters r and p is given by

$$g_N(t) = \left(\frac{p}{1-(1-p)t}\right)^r \quad \text{for} \quad |t| < \frac{1}{p}.$$

Exercise 6.3:
Let $\{\tilde{N}_t\}$ be a Poisson process with intensity 1 and $\Lambda \sim \Gamma(\alpha, \beta)$. Verify that the mixed Poisson process $N_t = \tilde{N}_{\Lambda t}$ has a negative binomial distribution and determine the parameters.

Exercise 6.4:
Let N be a discrete random variable with $N \in \{0, 1, 2, ...\}$. Define $q_n := P(N = n)$ and consider the relation

$$q_n = \left(a + \frac{b}{n}\right) q_{n-1}, \quad n = 1, 2, ...$$

Prove that N satisfies this relation for proper choices of a and b in the following cases.

1) N Poisson distributed with intensity $\lambda > 0$.

2) N Binomial distributed with parameters k and p.

3) N negative binomial distributed with parameters p and r.

One can prove that these three distributions are the only distributions satisfying this recursive relation.

Exercise 6.5:
The general GARCH(p, q) process is defined by $\mathcal{R}_n = \sigma_n Z_n$, where $\{Z_n\}$ is an iid sequence and

$$\sigma_n^2 = a_0 + \sum_{i=1}^{p} a_i \mathcal{R}_{n-i}^2 + \sum_{j=1}^{q} b_j \sigma_{n-j}^2$$

with $a_0, ..., a_p, b_1, ..., b_q$ constants. For simplicity, we specialise here to the case $p = q = 2$. Let $V_n = (\sigma_{n+1}^2, \sigma_n^2, \mathcal{R}_n^2)^T$ where T denotes the transpose.

1) Show that

$$V_n = A_n V_{n-1} + B_n$$

for some 3×3 matrix A_n and some vector B_n. Argue that A_n and B_n do not depend on n in the sense that $A_n \overset{\mathrm{d}}{=} A$ and $B_n \overset{\mathrm{d}}{=} B$ for some A and B.

2)Assume that V_n converges in distribution to some V. Find a function Φ depending on A and B such that $V \stackrel{\mathrm{d}}{=} \Phi(V)$. Such a relation is called a *stochastic fixed point equation*.

Exercise 6.6:
The objective of this problem is to apply large deviations methods to credit risk, following along the lines of the article [23].

Consider a Bernoulli mixture model where the probability that the i'th obligor defaults is given by $P(X_i = 1) = p$, where p is a "small" constant. The size of the i'th loss given default is Z_i, where $\{Z_i\}$ is an iid sequence with $Z_i \sim \mathrm{Exp}(\theta)$. The total loss is

$$L_n = \sum_{i=1}^{n} X_i Z_i$$

where $X_i = 1$ in the case of default and $X_i = 0$ otherwise. We assume that $\{X_i\}$ and $\{Z_i\}$ are independent. The objective is to understand the exponential decay of

$$P(L_n > na) \quad \text{as} \quad n \to \infty.$$

According to Cramér's large deviation theorem,

$$\lim_{n\to\infty} \frac{1}{n} \log P(L_n > na) = -\Lambda^*(a), \quad \text{where} \quad \Lambda^*(a) = \sup_{\xi\in\mathbb{R}}(\xi a - \Lambda(\xi)). \tag{6.1}$$

Here, Λ denotes the cumulant-generating function of $X_1 Z_1$, that is, $\Lambda(\xi) = \log E[e^{\xi X_1 Z_1}]$.

1)Compute $\Lambda(\xi)$. Hint: Use the law of total expectation with the events $X_1 = 1$ and $X_1 = 0$.

2)Calculate the limit appearing on the left-hand side of (6.1).

3)Now use the approximation $P(L_n > na) \approx e^{-n\Lambda^*(a)}$ to write an approximation for VaR_α as a function of α, θ and p.

Exercise 6.7:
Consider the probit Normal mixture model where we have

$$P\Big(\frac{N_n}{n} > x\Big) \to \Phi(p^{-1}(x)), \tag{6.2}$$

where N_n denotes the number of defaults in a portfolio with n homogeneous loans, $p(z)$ is the probability of default for a single loan conditioned on the common factor $Z = z$, and Φ is the standard Normal distribution function. Our goal is to find a tractable expression for the right-hand side of (6.2) by using an asymptotic approach.

1)Find an explicit expression for $p^{-1}(x)$ (where your answer will depend on Φ^{-1} and the parameters in the probit Normal mixture model).

2)Next, observe that for the Normal distribution function,

$$\Phi(x) \sim \frac{C}{x} e^{-x^2/2} \quad \text{as} \quad x \to \infty,$$

and use this asymptotic expression to describe $P(N_n/n > x)$ as $x \to 1$, where your answer will still depend on Φ^{-1} at this stage.

3)Finally, let s be a positive constant and set $x_n = \Phi(s\sqrt{\log n})$, and then derive an asymptotic estimate for $P(N_n/n > x_n)$ as $n \to \infty$. Using this estimate, develop a Value-at-Risk approximation at level α, where $\alpha \in (0,1)$ is a constant which is close to one. Hint: Consider

$$\frac{1}{\log n} \log P\Big(\frac{N_n}{n} > x_n\Big).$$

4)Now assume instead that Z has a symmetric two-sided regularly varying distribution, so that the tail is given by $L(x)x^{-\beta}$ for some $\beta > 1$, where again we choose $x_n = \Phi(s\sqrt{\log n})$. Verify that

$$P\Big(\frac{N_n}{n} > x_n\Big) \sim L\Big(\frac{\sqrt{1-\rho}s\sqrt{\log n} - d}{\sqrt{\rho}}\Big)\Big(\frac{\sqrt{1-\rho}s\sqrt{\log n} - d}{\sqrt{\rho}}\Big)^{-\beta}.$$

What is, intuitively, the growth rate of the right-hand side?

5)Now let $L \equiv C$ in the previous subproblem. Derive an asymptotic approximation for the Value-at-Risk of N_n.

6)Discuss what needs to be verified (based on the proof of (6.2)) for these asymptotic results to hold in a rigorous mathematical sense.

Appendix A

Generalised inverses

A.1 Definition and basic properties

Definition A.1.1. Let $h : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. We define the *generalised inverse* of h as

$$h^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : h(x) \geq t\}.$$

We have the convention $\inf \emptyset = \infty$.

Proposition A.1.2. *For a non-decreasing function h, $h^{\leftarrow}$ is left-continuous.*

Proof. This proof is from [102]. Assume that $t_n \uparrow t$ but $h^{\leftarrow}(t-) := \lim_{t_n \uparrow t} h^{\leftarrow}(t_n) < h^{\leftarrow}(t)$. Then we can find $x \in \mathbb{R}$ and $\delta > 0$ such that for all n,

$$h^{\leftarrow}(t_n) < x < h^{\leftarrow}(t) - \delta.$$

But by definition of $h^{\leftarrow}$, we have $h(x) \geq t_n$ for all n. Let $n \to \infty$, then $h(x) \geq t$ so by (i) of the proposition below, $h^{\leftarrow}(t) \leq x$. This is in contradiction to $x < h^{\leftarrow}(t) - \delta$. ■

The following properties of generalised inverses will be useful.

Proposition A.1.3. *Let h be non-decreasing.*

(i) $x \geq h^{\leftarrow}(t)$ if and only if $h(x) \geq t$ when h is right-continuous.

(ii) h is continuous if and only if $h^{\leftarrow}$ is strictly increasing.

(iii) $h(h^{\leftarrow}(t)) = t$ for all t if and only if h is continuous.

(iv) h is strictly increasing if and only if $h^{\leftarrow}(h(x)) = x$ for all x.

(v) If h is right-continuous, $h(h^{\leftarrow}(t)) \geq t$.

(vi) It always holds that $h^{\leftarrow}(h(x)) \leq x$.

Proof. Point (i) is left as an exercise. Consider (ii). h is non-decreasing, so any discontinuity is a (positive) jump. Since a jump of h corresponds to a flat region for $h^{\leftarrow}$ (make a drawing!), the claim follows. One can make similar arguments for (iii) and (iv). For complete proofs of these statements and many others concerning generalised inverses, consult [29]. ■

Proposition A.1.4. *Let F be the distribution function of the random variable X.*

(i) $F^{\leftarrow}(U) \stackrel{\mathrm{d}}{=} X$ *for* $U \sim \mathrm{Unif}(0,1)$.

(ii) If F is continuous, $F(X) \stackrel{\mathrm{d}}{=} U$ *for* $U \sim \mathrm{Unif}(0,1)$.

(iii) $P(X \leq x) = P(F(X) \leq F(x))$.

Proof. (i) and (ii) are left as exercises. As for (iii), we note that $X \leq x$ implies $F(X) \leq F(x)$ since F is non-decreasing. Conversely, consider the event $\{F(X) \leq F(x), X > x\}$. If $X > x$, we have $F(X) \geq F(x)$ and hence $\{F(X) \leq F(x), X > x\} \subseteq \{F(X) = F(x), X > x\}$ so that F is flat on $[x, X]$. This implies $P(F(X) \leq F(x), X > x) = 0$, completing the proof. ■

A.2 Exercises

Exercise A.1:
Prove (i) in Proposition A.1.3.

Exercise A.2:
Prove (i) and (ii) in Proposition A.1.4. Give a counterexample which shows that (ii) need not hold for general distribution functions.

Appendix B

Probability theory

B.1 Random variables and distribution functions

Definition B.1.1. If $(\Omega, \mathcal{F}, P)$ is a probability space, a *random variable* is a measurable map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel sigma-algebra on $\mathbb{R}$.

In this course, measurability is a central concept. We also rarely worry about the background space $(\Omega, \mathcal{F}, P)$. We now go through distribution functions in some detail since distribution functions and quantile functions play a central role in the course.

Definition B.1.2. For a random variable X, we define the *distribution function* F of X as

$$F(x) = P(X \leq x).$$

Similarly, if $\mathbf{X} = (X_1, ..., X_d)$ is $\mathbb{R}^d$-valued, the distribution function F is given by

$$F(x_1, ..., x_d) = P(X_1 \leq x_1, ..., X_d \leq x_d).$$

If F is a distribution function for X, we call $\overline{F} = 1 - F$ the *survival function* of X.

In the univariate case, we have a nice characterisation of distribution functions.

Proposition B.1.3. *A function $F : \mathbb{R} \to \mathbb{R}$ is the unique distribution function of a random variable if and only if the following properties hold:*

1. *F is right-continuous.*
2. *F is non-decreasing.*
3. $\lim_{x \to \infty} F(x) = 1$.
4. $\lim_{x \to -\infty} F(x) = 0$.

Proof. Assume that F is a distribution function for the random variable X. For $\varepsilon > 0$, we have $\{X \leq x + \varepsilon\} \downarrow \{X \leq x\}$ for $\varepsilon \downarrow 0$. By continuity from above for measures, $F(x + \varepsilon) \to F(x)$ for $\varepsilon \downarrow 0$, showing that F has property 1. If $x \leq y$, then $\{X \leq x\} \subseteq \{X \leq y\}$ so that $F(x) = P(X \leq x) \leq P(X \leq y) = F(y)$, proving 2. Properties 3 and 4 follow from the fact

that X is real-valued. Conversely, suppose F satisfies properties 1 - 4. Let $X = F^{\leftarrow}(U)$ where U is uniformly distributed on $(0, 1)$. Then

$$P(X \leq x) = P(F^{\leftarrow}(U) \leq x) = P(U \leq F(x)) = F(x)$$

by Proposition A.1.3. Hence X has distribution function F. ■

We recall the following definition.

Definition B.1.4. A function f is called càdlàg (French: continue à droite, limite à gauche) if f is right-continuous and has left limits.

Corollary B.1.5. *Every distribution function is càdlàg.*

Proof. A non-decreasing function has left-limits. As a distribution function is right-continuous by the above result, the corollary follows. ■

The distribution function of a random variable determines its distribution. Note also the useful identity

$$P(a < X \leq b) = F(b) - F(a)$$

for every $a < b$. This generalises to the two-dimensional case as follows: If (X, Y) has distribution function F, then

$$P(a < X \leq b, c < Y \leq d) = F(a, c) + F(b, d) - F(a, d) - F(b, c). \tag{B.1}$$

This is best seen by making a drawing of the rectangle in question. Multivariate distribution functions have similar properties as in the univariate case.

Proposition B.1.6. *Any multivariate distribution function $F : \mathbb{R}^d \to \mathbb{R}$ satisfies the following properties:*

1. *F is non-decreasing in each variable.*
2. *F is right-continuous in each variable.*
3. $\lim_{x_1,...,x_d \to \infty} F(x_1, ..., x_d) = 1$.
4. $0 \leq F(x_1, ..., x_d) \leq 1$.
5. $\lim_{x_i \to -\infty} F(x_1, ..., x_d) = 0$ *for every* $i = 1, ..., d$.

Proof. Left as an exercise for the reader. ■

Note that the above proposition is not an if and only if statement as in the univariate case. A counterexample is given in the exercises. We end this subsection about distribution functions with two tables containing the most important examples for this course.

Distribution	Density	Distribution function	Parameters
Normal, $\mathcal{N}(\mu, \sigma^2)$	$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$	$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$	$(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$
Exponential, Exp(λ)	$\lambda e^{-\lambda x}, x > 0$	$1 - e^{-\lambda x}, x > 0$	$\lambda \in (0, \infty)$
Gamma, $\Gamma(\alpha, \beta)$	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$	$\int_0^x f(t)dt, x > 0$	$(\alpha, \beta) \in (0, \infty)^2$
Student t	$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	$t_\nu(x) = \int_{-\infty}^{x} f(t)dt$	$\nu \in (0, \infty)$
Lognormal(μ, σ^2)	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\log x - \mu)^2/2\sigma^2}, x > 0$	$\int_0^x f(t)dt, x > 0$	$(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$
Pareto	$\frac{\alpha\kappa^\alpha}{(\kappa+x)^{\alpha+1}}, x > 0$	$1 - \left(\frac{\kappa}{\kappa+x}\right)^\alpha, x > 0$	$(\alpha, \kappa) \in (0, \infty)^2$
Beta	$x^{a-1}(1-x)^{b-1}/B(a,b), x \in (0,1)$	$\int_0^x t^{a-1}(1-t)^{b-1}dt/B(a,b)$	$(a, b) \in (0, \infty)^2$

Table B.1: Densities and distribution functions of some common continuous distributions. See below for more details on these distributions.

Distribution	Density $P(N = k)$	Distribution function	Parameters
Poisson(λ)	$\frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, ...$	$\sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda}, k = 0, 1, 2, ...$	$\lambda \in (0, \infty)$
Binomial(n, p)	$\binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, ..., n$	$\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, k = 0, 1, ..., n$	$n \in \mathbb{N}, p \in [0, 1]$
Geometric(p)	$(1-p)^k p, k = 0, 1, 2, ...$	$1 - (1-p)^{k+1}, k = 0, 1, 2, ...$	$p \in [0, 1]$
Negative binomial	$\binom{k+r-1}{k} (1-p)^k p^r, k = 0, 1, 2, ...$	$\sum_{i=0}^{k} \binom{i+r-1}{i} (1-p)^i p^r, k = 0, 1, 2, ...$	$p \in [0, 1]$, $r \in \mathbb{N}$

Table B.2: Densities and distribution functions of some common discrete distributions.

B.2 Characteristic functions and moment-generating functions

An alternative characterisation of distributions is via moment-generating functions and characteristic functions.

Definition B.2.1. For a random variable X, the function

$$\Phi_X(t) = E[e^{itX}]$$

is called the *characteristic function* of X. If there exists a neighbourhood $(-a, a)$ of zero $(a > 0)$ such that

$$\kappa_X(t) = E[e^{tX}], \quad t \in (-a, a)$$

is finite, we call κ_X the *moment-generating function* of X.

For the sake of brevity, we will often write cf for characteristic function and mgf for moment-generating function. Note that the characteristic function of a random variable is always defined. Indeed, the integrand is bounded by 1 in norm.

Example B.2.2. For the $\mathcal{N}(0, 1)$ distribution, the characteristic function is given by

$$\Phi(t) = e^{-t^2/2}.$$

The case for the general normal distribution $\mathcal{N}(\mu, \sigma^2)$ is left as an exercise, see also the lemma below. ∘

The cf and mgf are easily generalised to a multivariate random variable $\mathbf{X} = (X_1, ..., X_d)$ as follows:

$$\Phi_{\mathbf{X}}(\mathbf{t}) = E\left[e^{i\mathbf{t}^T\mathbf{X}}\right], \quad \kappa_{\mathbf{X}}(\mathbf{t}) = E\left[e^{\mathbf{t}^T\mathbf{X}}\right], \quad \mathbf{t} \in \mathbb{R}^d$$

where the mgf is only defined in the neighbourhood of the origin where it is finite.

Lemma B.2.3. *Let $\mathbf{X}$ be a $\mathbb{R}^d$-valued random variable, $\mathbf{a} \in \mathbb{R}^n$ and B a $n \times d$ matrix. Then the $\mathbb{R}^n$-valued random variable $\mathbf{a} + B\mathbf{X}$ has cf*

$$\Phi_{\mathbf{a}+B\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{a}^T\mathbf{t}}\Phi_{\mathbf{X}}(B^T\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^n.$$

Similarly, whenever the mgfs exist,

$$\kappa_{\mathbf{a}+B\mathbf{X}}(\mathbf{t}) = e^{\mathbf{a}^T\mathbf{t}}\kappa_{\mathbf{X}}(B^T\mathbf{t}).$$

Proof. The proof is left to the reader. ■

The cf has the following important properties.

Theorem B.2.4. *If $\mathbf{X}$ and $\mathbf{Y}$ are random variables with the same characteristic functions, $\Phi_{\mathbf{X}} = \Phi_{\mathbf{Y}}$, then $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.*

Proof. See Theorem 14.1 in [68]. ■

Corollary B.2.5. *The variables $X_1, ..., X_d$ are independent if and only if*

$$\Phi_{\mathbf{X}}(t_1, ..., t_d) = \prod_{i=1}^{d} \Phi_{X_i}(t_i)$$

for all $t_1, ..., t_d$ where $\mathbf{X} = (X_1, ..., X_d)$.

Proof. Assume the $X_1, ..., X_d$ are independent. Then

$$\Phi_{\mathbf{X}}(t_1, ..., t_d) = E\left[e^{i(t_1X_1+\cdots t_dX_d)}\right] = E\left[e^{it_1X_1}\right] \cdots E\left[e^{it_dX_d}\right] = \prod_{i=1}^{d} \Phi_{X_i}(t_i).$$

Conversely, if the cf factors as above, it follows immediately from the uniqueness theorem above that the $X_1, ..., X_d$ are independent.

■

Proposition B.2.6. *Let X be a one-dimensional random variable with cf Φ_X. If $E[|X|^k] < \infty$ for some $k \in \mathbb{N}$, then Φ_X is C^k (k times differentiable and the k'th derivative is continuous) and*

$$\Phi_X^{(m)}(0) = i^m E[X^m], \quad m = 1, ..., k.$$

Proof. See Theorem 6.34 in [54] and the paragraph following the theorem. ■

Maybe not surprisingly, these properties more or less carry over to the mgf. A discussion of the result below can be found in [7], chapter 30.

Theorem B.2.7. *If the mgfs of $\mathbf{X}$ and $\mathbf{Y}$ exist in a neighbourhood around zero and are equal, then $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.*

Corollary B.2.8. *Let the variables $X_1, ..., X_d$ have moment-generating functions $\kappa_{X_1}, ..., \kappa_{X_d}$ that exist in a neighbourhood around zero, then $X_1, ..., X_d$ are independent if and only*

$$\kappa_{(X_1,...,X_d)}(t_1, ..., t_d) = \prod_{i=1}^{d} \kappa_{X_i}(t_i).$$

Proposition B.2.9. *Let X be a one-dimensional random variable with mgf κ_X that exists in a neighbourhood $(-c, c)$ around zero. Then X has moments of all orders and for $k \in \mathbb{N}$,*

$$\kappa_X^{(k)}(0) = E[X^k].$$

We end this subsection with tables containing the mgf and cf of the distributions from the tables of distributions above.

Distribution	cf	mgf	Constraint for mgf
Normal, $\mathcal{N}(\mu, \sigma^2)$	$e^{i\mu t - \frac{1}{2}t^2\sigma^2}$	$e^{\mu t + \frac{1}{2}t^2\sigma^2}$	$t \in \mathbb{R}$
Exponential, $\text{Exp}(\lambda)$	$\frac{\lambda}{\lambda - it}$	$\frac{\lambda}{\lambda - t}$	$t \in (-\infty, \lambda)$
Gamma, $\Gamma(\alpha, \beta)$	$\left(\frac{\beta}{\beta - it}\right)^\alpha$	$\left(\frac{\beta}{\beta - t}\right)^\alpha$	$t \in (-\infty, \beta)$
Student t	No explicit form	Doesn't exist	-
Lognormal(μ, σ^2)	No explicit form	Doesn't exist	-
Pareto	No explicit form	Doesn't exist	-
Beta	No explicit form	No explicit form	-

Table B.3: Characteristic functions and moment-generating functions for the distributions in table B.1.

Distribution	cf	mgf	Constraint for mgf
Poisson(λ)	$e^{\lambda(e^{it}-1)}$	$e^{\lambda(e^t-1)}$	$t \in \mathbb{R}$
Binomial(n, p)	$(pe^{it} + 1 - p)^n$	$(pe^t + 1 - p)^n$	$t \in \mathbb{R}$
Geometric(p)	$\frac{p}{1-(1-p)e^{it}}$	$\frac{p}{1-(1-p)e^t}$	$t < -\log(1-p)$
Negative binomial	$\left(\frac{p}{1-(1-p)e^{it}}\right)^r$	$\left(\frac{p}{1-(1-p)e^t}\right)^r$	$t < -\log(1-p)$

Table B.4: Characteristic functions and moment-generating functions for the distributions in table B.2.

B.3 Some important distributions

In the following, whenever we write $x \in ...$, we mean that outside this range, the density is zero.

Normal

The normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ has density

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

This distribution is denoted $\mathcal{N}(\mu, \sigma^2)$. The distribution $\mathcal{N}(0, 1)$ is denoted the *standard normal distribution*. If $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$E[X] = \mu \quad \text{and} \quad \text{Var}[X] = \sigma^2$$

which explains the terminology for these parameters. Some higher order central moments are

$$E[(X - \mu)^k] = 0, \quad k = 1, 3, 5, ... \quad \text{and} \quad E[(X - \mu)^4] = 3\sigma^4$$

and some higher moments are

$$E[X^2] = \sigma^2 + \mu^2, \quad E[X^3] = \mu^3 + 3\mu\sigma^2, \quad E[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4.$$

Uniform

The uniform distribution on (a, b) for $a < b$ has density

$$f(x) = \frac{1}{b - a}, \quad x \in (a, b),$$

and we write $X \sim \text{Unif}(a, b)$ when X has the uniform distribution. One easily shows that the mean and variance of X is then given by

$$E[X] = \frac{b + a}{2}, \quad \text{Var}[X] = \frac{(b - a)^2}{12}.$$

The higher moments are given by

$$E[X^k] = \frac{b^{k+1} - a^{k+1}}{(k + 1)(b - a)}, \quad k \in \mathbb{N}.$$

Exponential

The exponential distribution has density

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

where $\lambda > 0$ is a parameter. When $\lambda = 1$, we talk of the *standard exponential distribution*. The mean and variance of X with this distribution are given by

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}.$$

In the exercises, you are asked to derive all moments for the exponential distribution.

Gamma

To define the density for the gamma distribution, introduce the *gamma function*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha - 1} e^{-x} dx, \quad \alpha > 0.$$

With integration by parts, it is easily verified that this function satisfies the relationship $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. The gamma function can be seen as a generalisation of the factorial

function $n! = n \cdot (n-1) \cdots 2 \cdot 1$ to the positive real numbers.

The *gamma distribution* with parameters $\alpha, \beta > 0$, denoted $\Gamma(\alpha, \beta)$, has density

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

It is seen that the case $\alpha = 1$ corresponds to the exponential distribution. It is easy to verify that if $X \sim \Gamma(\alpha, \beta)$, then

$$E[X] = \frac{\alpha}{\beta}, \quad \mathrm{Var}[X] = \frac{\alpha}{\beta^2}.$$

Pareto

The most typical specification of the Pareto distribution used in this book is the distribution with distribution function

$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^\alpha, \quad x > 0$$

where $\alpha, \kappa > 0$ are parameters. The k'th moment exists if and only if $\alpha > k$. In this case, for X a random variable with this distribution,

$$E[X^k] = \frac{\kappa^k k!}{\prod_{i=1}^k (\alpha - i)}.$$

Beta

The *beta distribution* with parameters $a, b > 0$ has density

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad x \in (0,1),$$

where $B(a,b)$ is the *beta function* given by

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Note that $a = b = 1$ corresponds to the Unif$(0,1)$ distribution. The mean and variance of X with the Beta(a,b) distribution is given by

$$E[X] = \frac{a}{a+b}, \quad \mathrm{Var}[X] = \frac{ab}{(a+b+1)(a+b)^2}.$$

The multivariate normal distribution

Definition B.3.1. An $\mathbb{R}^d$-valued random variable $\mathbf{X} = (X_1, ..., X_d)$ is *multivariate normal* if for every $\mathbf{a} \in \mathbb{R}^d$, the real-valued random variable $\mathbf{a}^T\mathbf{X}$ has a normal distribution.

The definition does not say that being multivariate normal is the same as all marginal variables being normal. A counterexample is provided in the exercises. In order to prove results with the multivariate normal distribution, the following theorem is essential.

Theorem B.3.2. $\mathbf{X}$ *is multivariate normal of dimension* d *if and only if there exists a symmetric positive semi-definite matrix* $\Sigma \in \mathbb{R}^{d\times d}$ *and a vector* $\mu \in \mathbb{R}^d$ *such that*

$$\Phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^d.$$

In this case, Σ *is the covariance matrix of* $\mathbf{X}$ *and* μ *is the mean vector i.e.* $E[X_i] = \mu_i$ *and* $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ *for all* $i, j = 1, ..., d$.

Proof. See Theorem 16.1 in [68]. ■

The theorem allows us to define the following.

Definition B.3.3. For a multivariate normal vector $\mathbf{X}$, we write $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu}$ is the mean vector and Σ is the covariance matrix. If Σ is invertible (i.e. $\det\Sigma \neq 0$), we say that $\mathbf{X}$ has a *regular* multivariate normal distribution. Otherwise, $\mathbf{X}$ is called *singular*.

Theorem B.3.4. *A regular multivariate normal variable* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *in* $\mathbb{R}^d$ *has density*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^d$$

with respect to Lebesgue measure on $\mathbb{R}^d$.

Proof. See Corollary 16.2 in [68]. Note the error in equation (16.5). It should say $(2\pi)^{n/2}$ and not $2\pi^{n/2}$. ■

The following result will be used extensively in the discussion on spherical and elliptical distributions.

Proposition B.3.5. *Let* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *be* d*-dimensional, let* $\mathbf{a} \in \mathbb{R}^n$ *and let* B *be an* $n \times d$*-matrix. Then* $\mathbf{Y} := \mathbf{a} + B\mathbf{X} \sim \mathcal{N}(\mathbf{a} + B\boldsymbol{\mu}, B\Sigma B^T)$.

Proof. Left as an exercise for the reader. ■

Stable distributions

Many parametrisations exist for stable distributions.con We use the following definition as per Definition 1.5 in [92], from which the following results are taken.

Definition B.3.6. A random variable X is *stable* with parameters $\alpha \in (0, 2], \beta \in [-1, 1], \gamma > 0, \delta \in \mathbb{R}$ if the characteristic function of X is given by

$$\Phi_X(t) = \begin{cases} \exp(-\gamma^\alpha|t|^\alpha(1 - i\beta\mathrm{sign}(t)\tan(\pi\alpha/2)) + i\delta t), & \alpha \neq 1 \\ \exp(-\gamma|t|(1 + i\beta\mathrm{sign}(t)\frac{2}{\pi}\log|t|) + i\delta t), & \alpha = 1. \end{cases}$$

In this case, we write $X \sim S(\alpha, \beta, \gamma, \delta; 1)$.

One important fact of this distribution is that if $\alpha < 1$ and $\beta = 1$, the support of X is $[\delta, \infty)$ (this is Lemma 1.1 in [92]). In particular, if $\beta = 1, \alpha < 1$ and $\delta \geq 0$, the support of X is contained in the positive real line. Such a distribution is called *positive stable*. The stable distributions arise as the class of possible limit distributions when considering suitably scaled and normalised sums of iid random variables. Thus it should be no surprise that the normal distribution is contained in this class (see the exercises). The following result shows that stable distributions are preserved under scaling and translation.

Proposition B.3.7 (Proposition 1.4 (a) in [92]). *Let $X \sim S(\alpha, \beta, \gamma, \delta; 1)$ and $a \neq 0, b \in \mathbb{R}$. Then*

$$aX + b \sim \begin{cases} S(\alpha, sign(a)\beta, |a|\gamma, a\delta + b; 1), & \alpha \neq 1 \\ S(1, sign(a)\beta, |a|\gamma, a\delta + b - \frac{2}{\pi}\beta\gamma a \log|a|; 1), & \alpha = 1. \end{cases}$$

Combining this result with the following theorem, we obtain an algorithm to simulate from a stable distribution.

Theorem B.3.8 (Theorem 1.3 in [92]). *Let $\Theta \sim Unif[-\pi/2, \pi/2]$ and $E \sim Exp(1)$ be independent and assume $0 < \alpha \leq 2$. Define*

$$\theta_0 := \frac{1}{\alpha} \arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right), \quad \alpha \neq 1.$$

Then

$$Z = \begin{cases} \frac{\sin(\alpha(\theta_0+\Theta))}{(\cos(\alpha\theta_0)\cos\Theta)^{1/\alpha}} \left(\frac{\cos(\alpha\theta_0+(\alpha-1)\Theta)}{E}\right)^{(1-\alpha)/\alpha}, & \alpha \neq 1 \\ \frac{2}{\pi}\left(\left(\frac{\pi}{2} + \beta\Theta\right)\tan\Theta - \beta\log\left(\frac{\frac{\pi}{2}E\cos\Theta}{\frac{\pi}{2}+\beta\Theta}\right)\right), & \alpha = 1 \end{cases}$$

has a $S(\alpha, \beta, 1, 0; 1)$ distribution.

B.4 Conditional expectations

The presentation in this subsection follows chapter 9 of [54]. Conditional expectations are essential in performing computations in probability theory and statistics. The (measure theoretic) definition of a conditional expectation is somewhat strange at first, but the definition has the advantage that many theoretical properties follow almost trivially.

Definition B.4.1. Let X be a random variable on $(\Omega, \mathcal{F}, P)$, $E[|X|] < \infty$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub-sigma-algebra. The conditional expectation of X with respect to $\mathcal{G}$, denoted $E[X \mid \mathcal{G}]$, is a random variable satisfying the following properties:

(i) $E[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.

(ii) For every $A \in \mathcal{G}$,

$$\int_A X dP = \int_A E[X \mid \mathcal{G}] dP.$$

Intuitively, we think of the conditional expectation $E[X \mid \mathcal{G}]$ as our best guess of the value of X given the information in $\mathcal{G}$. Try to keep this intuition in mind when reading the following examples and theoretical results.

It is by no means trivial that the conditional expectation exists. An elegant construction is via the Radon-Nikodym theorem, see [54] chapter 8 and Theorem 9.1. We also remark that $E[X \mid \mathcal{G}]$ is only unique almost surely. To verify theoretical statements concerning $E[X \mid \mathcal{G}]$, it suffices to verify the two properties above. If another variable Z satisfies the above assumptions, we have $E[X \mid \mathcal{G}] = Z$ a.s. In the following, we omit writing a.s. when considering computations involving conditional expectations. Also, if $\mathcal{G} = \sigma(Y)$ is the smallest sigma-algebra making Y measurable (intuitively, the information Y contains), we write $E[X \mid Y]$ instead of $E[X \mid \sigma(Y)]$.

Example B.4.2. Assume X is $\mathcal{G}$-measurable. We claim that $X = E[X \mid \mathcal{G}]$. X satisfies (i) by assumption and for $A \in \mathcal{G}$, we have

$$\int_A X dP = \int_A E[X \mid \mathcal{G}] dP$$

by definition of $E[X \mid \mathcal{G}]$, verifying (ii). ∘

Example B.4.3. Assume X is independent of $\mathcal{G}$ i.e. $P(A \cap \{X \in B\}) = P(A)P(X \in B)$ for all Borel sets B and $A \in \mathcal{G}$. We claim that $E[X \mid \mathcal{G}] = E[X]$. $E[X]$ is constant and thus trivially $\mathcal{G}$-measurable. Also, we get for $A \in \mathcal{G}$ that

$$\int_A X dP = E[1_A X] = E[1_A]E[X] = P(A)E[X] = \int_A E[X] dP$$

so both (i) and (ii) are satisfied by $E[X]$, proving the claim. ∘

The following proposition allows us to compute a plethora of interesting examples.

Proposition B.4.4. *If $D_1, D_2, ...$ are disjoint sets in $\mathcal{F}$ with $\cup_n D_n = \Omega$ (such a collection is called a* partition*), $P(D_i) > 0$ for all i, $\mathcal{G} = \sigma(D_1, D_2, ...)$ and X is an integrable random variable, then*

$$E[X \mid \mathcal{G}](\omega) = \begin{cases} \frac{1}{P(D_1)} \int_{D_1} X dP, & \omega \in D_1 \\ \frac{1}{P(D_2)} \int_{D_2} X dP, & \omega \in D_2 \\ \vdots \end{cases}$$

Proof. It is not hard to verify that the sigma-algebra $\mathcal{G}$ consists of the sets that are unions of the D_i. Since $E[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable, $E[X \mid \mathcal{G}]$ must be constant when restricted to one of the D_i i.e.

$$E[X \mid \mathcal{G}](\omega) = \begin{cases} c_1, & \omega \in D_1 \\ c_2, & \omega \in D_2 \\ \vdots \end{cases}.$$

Since

$$\int_{D_i} X dP = \int_{D_i} E[X \mid \mathcal{G}] dP = \int_{D_i} c_i dP = c_i P(D_i),$$

the claim follows. ■

Corollary B.4.5. *Let N be a random variable with $N \in \{0, 1, 2, ...\}$. If X is an integrable random variable, then*

$$E[X \mid N] = \begin{cases} \frac{1}{P(N=0)} \int_{\{N=0\}} X dP & on \quad \{N = 0\} \\ \frac{1}{P(N=1)} \int_{\{N=1\}} X dP & on \quad \{N = 1\} \\ \vdots \end{cases}$$

Proof. This follows immediately from the previous proposition by letting $\mathcal{G} = \sigma(N)$ and noting that $\sigma(N)$ is generated by the partition $\{\{N = 0\}, \{N = 1\}, ...\}$. ■

Before computing some interesting examples, we state the following list of properties of conditional expectations.

Theorem B.4.6. *Let X and Y be random variables with finite expectation, $\mathcal{G} \subseteq \mathcal{F}$ a sub-sigma-algebra.*

(i) For $a, b \in \mathbb{R}$, $E[aX + bY \mid \mathcal{G}] = aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]$ (linearity).

(ii) $E[X] = E[E[X \mid \mathcal{G}]]$ (tower property).

(iii) If $X \leq Y$ then $E[X \mid \mathcal{G}] \leq E[Y \mid \mathcal{G}]$ (monotonicity).

(iv) $|E[X \mid \mathcal{G}]| \leq E[|X| \mid \mathcal{G}]$ (triangle inequality).

(v) If X is $\mathcal{G}$-measurable, $E[XY \mid \mathcal{G}] = X[Y \mid \mathcal{G}]$.

Proof. Consider (i). We have to show that $aE[X \mid \mathcal{G}]+bE[Y \mid \mathcal{G}]$ satisfies the two properties of $[aX + bY \mid \mathcal{G}]$. Measurability is obvious. For $A \in \mathcal{G}$, we have

$$\begin{aligned}\int_A aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]dP &= a\int_A E[X \mid \mathcal{G}]dP + b\int_A E[Y \mid \mathcal{G}]dP \\ &= a\int_A XdP + b\int_A YdP \\ &= \int_A aX + bYdP\end{aligned}$$

which proves the claim. As for (ii), simply choose $A = \Omega$ in the second property of conditional expectations to obtain

$$E[X] = \int_\Omega XdP = \int_\Omega E[X \mid \mathcal{G}]dP = E[E[X \mid \mathcal{G}]].$$

(iii) follows immediately from the monotonicity property of integrals. As for (iv), we obviously have $-|X| \leq X \leq |X|$ so from (iii), we get

$$-E[|X| \mid \mathcal{G}] \leq E[X \mid \mathcal{G}] \leq E[|X| \mid \mathcal{G}]$$

which is the desired result. See the exercises for an outline of the proof of (v). ■

The tower property, (ii), has many names. It is also known as the *law of iterated expectations* and the *law of total expectation* to name a few.

Example B.4.7. A typical situation in for example non-life insurance is to have a sum of the form

$$S = \sum_{i=1}^{N} X_i$$

where $\{X_i\}$ is an iid sequence independent of N, a random variable taking values in $\{0, 1, 2,...\}$. Assume both X_1 and N have finite expectation. What is the expectation

of S? Using Corollary B.4.5, we have[1]

$$\begin{aligned}\frac{1}{P(N=n)}\int_{\{N=n\}} S dP &= \frac{1}{P(N=n)}\int_{\{N=n\}} \sum_{i=1}^{n} X_i dP \\ &= \frac{1}{P(N=n)} E[1_{\{N=n\}}] E\left[\sum_{i=1}^{n} X_i\right] \\ &= nE[X_1]\end{aligned}$$

using that N and $\{X_i\}$ are independent. It follows that $E[S \mid N] = NE[X_1]$. From the tower property, it follows that

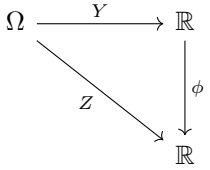
$$E[S] = E[E[S \mid N]] = E[NE[X_1]] = E[N]E[X_1].$$

∘

In practice, one does not proceed as formally as in the above example. The corollary that we applied essentially says that when we condition on a variable, that variable can be treated as a constant. If N was constant equal to n, we would say that $E[S] = nE[X_1]$. Then we just replace n by the random variable N to obtain $E[S \mid N]$. Let us make this more precise by first observing that $E[X \mid Y]$ is a function of Y.

Theorem B.4.8 (Doob–Dynkin lemma). *If a random variable Z is $\sigma(Y)$-measurable, then there exists a measurable function ϕ such that $Z = \phi(Y)$.*

Proof. See Theorem 9.23 in [54]. ■



By definition of a conditional expectation, $E[X \mid Y]$ is $\sigma(Y)$-measurable. Hence $E[X \mid Y] = \phi(Y)$ for some function ϕ. While the Doob-Dynkin lemma does not provide an explicit recipe for ϕ, it is possible to compute ϕ in many situations of interest. The following result shows how to compute $\phi(y) = E[X \mid Y = y]$ in the (quite typical) case where (X, Y) has a density.

Theorem B.4.9. *Let $E[|X|] < \infty$ and assume (X, Y) has density $f(x, y)$. Then*

$$E[X \mid Y = y] = \int_{\mathbb{R}} x \frac{f(x,y)}{g(y)} dx$$

where $g(y) = \int_{\mathbb{R}} f(x, y)dx$ is the density of Y.

Proof. This is Corollary 9.28 in [54] or Proposition 11.11 in [38]. ■

[1] We should in principle first verify that $E[|S|] < \infty$.

Remark B.4.10. By recalling that the conditional density of X given $Y = y$ is defined by

$$f_{X|Y=y}(x) = \frac{f(x,y)}{g(y)},$$

we could also write the above result as

$$E[X \mid Y = y] = \int_{\mathbb{R}} x f_{X|Y=y}(x) dx$$

which also makes sense intuitively. Given densities, the conditional expectation can be computed as an ordinary expectation but with the marginal density replaced by a conditional density.

B.5 Exercises

Exercise B.1:
If $Y \sim \mathcal{N}(\mu, \sigma^2)$, we say that $X = \exp(Y)$ has a lognormal distribution with parameters μ and σ. Using the density

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for the normal distribution, derive the density of the lognormal distribution.

Exercise B.2:
Prove Proposition B.1.6.

Exercise B.3:
Consider the function $F : \mathbb{R}^2 \to \mathbb{R}$ given by

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \text{ or } x + y < 1 \\ 1, & \text{else} \end{cases}.$$

1)Verify that F satisfies all the properties in Proposition B.1.6.

2)Show that F cannot be a distribution function for a pair of random variables (X, Y). Hint: Use equation (B.1). Now consider $a = c = 1/3, b = d = 1$.

Exercise B.4:
Let Y be $\mathcal{N}(0,1)$ and let Z be Bernoulli distributed with success parameter $1/2$. Assume Y and Z are independent. Define $X_1 := Y$ and $X_2 := 1_{\{Z=1\}}Y - 1_{\{Z=0\}}Y$.

1)Verify that X_1 and X_2 are both $\mathcal{N}(0,1)$ variables.

2)Show that (X_1, X_2) is not multivariate normal.

Exercise B.5:
Compute the moment-generating function for the $\Gamma(\alpha, \beta)$ distribution.

Exercise B.6:
In this exercise, we compute the moment-generating functions for some typical discrete distributions.

1)Compute the moment-generating function for the Bernoulli distribution i.e. $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for $p \in [0, 1]$.

2)Compute the moment-generating function for the Binomial(n, p) distribution. Hint: Use the previous exercise.

3)Compute the moment-generating function for the Geometric(p) distribution.

Exercise B.7:

1)Prove Lemma B.2.3.

We know that the standard normal distribution has the characteristic function

$$\Phi(t) = e^{-t^2/2}.$$

2)Compute the characteristic function for the $\mathcal{N}(\mu, \sigma^2)$ distribution.

Exercise B.8:
Compute all moments of the exponential distribution.

Exercise B.9:
The $\Gamma(\lambda, n)$ distribution for $n \in \mathbb{N}$ is called the *Erlang distribution*. Verify that if $X \sim \Gamma(\lambda, n)$ then

$$X \stackrel{\mathrm{d}}{=} Y_1 + \cdots + Y_n$$

with $Y_1, ..., Y_n$ iid exponential distributed with parameter λ.

Exercise B.10:
Prove Proposition B.3.5.

Exercise B.11:
Show that the normal distribution is a stable distribution i.e. that if X_1, X_2 are iid $\mathcal{N}(\mu, \sigma^2)$ and $a, b \geq 0$ are arbitrary, then

$$aX_1 + bX_2 \stackrel{\mathrm{d}}{=} cX + d$$

for suitable constants $c \geq 0$ and $d \in \mathbb{R}$ where $X \sim \mathcal{N}(\mu, \sigma^2)$. In the notation of stable distributions, verify that

$$\mathcal{N}(\mu, \sigma^2) = S(2, 0, \sigma/\sqrt{2}, \mu; 1).$$

Exercise B.12:
Assume N is Poisson distributed with parameter $\lambda > 0$ and $\{X_i\}$ is an iid sequence independent of N where X_1 has moment-generating function κ. Compute the moment-generating function of

$$\sum_{i=1}^{N} X_i.$$

Hint: Tower property.

Exercise B.13:
Prove the following extension of the tower property: If $\mathcal{G} \subseteq \mathcal{H} \subseteq \mathcal{F}$, then

$$E[E[X \mid \mathcal{H}] \mid \mathcal{G}] = E[X \mid \mathcal{G}].$$

Exercise B.14:
In this exercise, we introduce the *conditional variance*. Assume $E[X^2] < \infty$ and that $\mathcal{G}$ is a sub-sigma-algebra, then the conditional variance is defined by

$$\mathrm{Var}(X \mid \mathcal{G}) = E[(X - E[X \mid \mathcal{G}])^2 \mid \mathcal{G}].$$

1)Show that

$$\mathrm{Var}(X \mid \mathcal{G}) = E[X^2 \mid \mathcal{G}] - E[X \mid \mathcal{G}]^2.$$

2)Prove the *law of total variance*,

$$\mathrm{Var}(X) = E[\mathrm{Var}(X \mid \mathcal{G})] + \mathrm{Var}(E[X \mid \mathcal{G}]).$$

3)Assume X is $\mathcal{G}$-measurable. Prove that $\mathrm{Var}(X \mid \mathcal{G}) = 0$.
4)Assume Y is $\mathcal{G}$-measurable. Prove that $\mathrm{Var}(X + Y \mid \mathcal{G}) = \mathrm{Var}(X \mid \mathcal{G})$.
5)Assume X is independent of $\mathcal{G}$. Prove that $\mathrm{Var}(X \mid \mathcal{G}) = \mathrm{Var}(X)$.
6)Give an intuitive interpretation of the previous three subproblems.

Exercise B.15:
Consider Example B.4.7. Assume that the X_i have finite second moment. Show that

$$\mathrm{Var}(S) = E[N]\,\mathrm{Var}(X_1) + \mathrm{Var}(N)E[X_1]^2.$$

Hint: Use the law of total variance from the previous exercise.

Appendix C

Convergence concepts and results

C.1 Convergence almost surely and in probability

In this chapter, we briefly touch upon the convergence concepts that we will use in this course.

Definition C.1.1. Let $X, X_1, X_2, ...$ be random variables. We say that the sequence $\{X_n\}$ *converges almost surely* to X for $n \to \infty$ if the event $\{X_n \to X \text{ for } n \to \infty\}$ has probability one. We write

$$P(X_n \to X) = 1.$$

Example C.1.2. Let $X_1, X_2, ...$ be iid Bernoulli distributed with success probability $p \in (0, 1)$ i.e. $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$ for all i. Consider the product process $Y_n = X_1 \cdots X_n$. We claim that $Y_n \to 0$ a.s. Indeed, note that $Y_n \in \{0, 1\}$ a.s. and

$$P(Y_n \to 0) = P(Y_n = 0 \text{ for some } n \in \mathbb{N}) = 1 - P(Y_n = 1 \text{ for all } n \in \mathbb{N}).$$

By independence, we have for any $N \in \mathbb{N}$ that

$$P(Y_N = 1) = P(X_1 = 1, ..., X_N = 1) = p^N$$

and

$$P(Y_n = 1 \text{ for all } n \in \mathbb{N}) \leq P(Y_N = 1) = p^N.$$

As this equality holds for all N, we can take limits on both sides and obtain $P(Y_n = 1 \text{ for all } n \in \mathbb{N}) = 0$ which yields $P(Y_n \to 0) = 1$ as desired. ∘

Almost sure convergence is a strong form of convergence. A weaker type of convergence is convergence in probability.

Definition C.1.3. Let $X, X_1, X_2, ...$ be random variables. We say that the sequence $\{X_n\}$ *converges in probability* to X for $n \to \infty$ if for every $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0.$$

We write $X_n \xrightarrow{\text{P}} X$.

Lemma C.1.4. *Almost sure convergence implies convergence in probability.*

Proof. This proof is from [54]. Let $\{X_n\}$ be a sequence of random variables converging almost surely to X, and let $\varepsilon > 0$ be given. Consider an ω such that $X_n(\omega) \to X(\omega)$. There exists an $N \in \mathbb{N}$ (depending on ω) such that

$$|X_n(\omega) - X(\omega)| \leq \varepsilon \quad \text{for} \quad n \geq N,$$

implying that $1_{\{|X_n - X| > \varepsilon\}}(\omega) = 0$ for $n \geq N$. It follows that $1_{\{|X_n - X| > \varepsilon\}}(\omega) \to 0$ for $n \to \infty$ and since this holds for almost every ω, we have $1_{\{|X_n - X| > \varepsilon\}} \to 0$ almost surely. As this function is bounded by 1, dominated convergence implies

$$P(|X_n - X| > \varepsilon) = \int 1_{\{|X_n - X| > \varepsilon\}} dP \to 0$$

as desired. ■

It is not immediately clear that almost sure convergence is strictly stronger than convergence in probability. The difference is of a very technical nature. However, counterexamples exist, and we encourage the reader to look them up. See for example [114]. It is useful to have some tools to prove convergence almost surely and in probability. Such tools include the Markov inequality and Chebyshev's inequality.

Lemma C.1.5 (**Markov's inequality**). *Let X be a random variable. Then for any $\varepsilon > 0$,*

$$P(|X| > \varepsilon) \leq \frac{E[|X|]}{\varepsilon}.$$

Proof. Trivially, $\varepsilon 1_{\{|X| > \varepsilon\}} \leq |X|$. Now take expectations on both sides and rearrange. ■

Corollary C.1.6 (**Chebyshev's inequality**). *Let X be a random variable with finite second moment, $E[X^2] < \infty$. Then for any $\varepsilon > 0$,*

$$P(|X - E[X]| > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Proof. Left as an exercise for the reader. ■

Example C.1.7. Consider a sequence of non-negative variables $X_1, X_2, ...$ with finite expectation and $E[X_n] = 1/n$. For any $\varepsilon > 0$, we have by the Markov inequality that

$$P(|X_n - 0| > \varepsilon) \leq \frac{E[X_n]}{\varepsilon} = \frac{1}{n\varepsilon} \to 0$$

so $X_n \xrightarrow{\mathrm{P}} 0$. This result should not be surprising, considering the fact that a non-negative random variable is zero almost surely if and only if it has mean zero. ∘

The above example of an application of the Markov inequality is not exactly interesting, but we want to stress that the inequality, while a triviality, is extremely useful and flexible. The following result shows that almost sure convergence follows if the probability $P(|X_n - X| > \varepsilon)$ goes to zero fast enough.

Proposition C.1.8 (**Borel–Cantelli Criterion for Almost Sure Convergence**). *Let $X, X_1, X_2, ...$ be random variables. If for every $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty,$$

then $X_n \to X$ almost surely.

Proof. The proof is an immediate consequence of the Borel–Cantelli lemma, see Lemma 2.26 and Theorem 2.27 in [54]. Alternatively, consult Lemma 9.11 in [38]. ■

The following result is a cornerstone of probability theory.

Theorem C.1.9 (**Strong Law of Large Numbers**). *Let $\{X_i\}$ be an iid sequence of random variables with $E[|X_1|] < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X_1] \quad a.s.$$

Proof. A proof can be found in [54], Theorem 4.25, or in [38], Theorem 10.8. ■

Many strong convergence statements can be proved using the SLLN without the need to use the definition. A different technique relies on taking continuous transformations of sequences where the asymptotic behaviour is known as the following extremely useful theorem shows.

Theorem C.1.10 (**Continuous mapping theorem**). *Let f be a continuous function. Then*

(i) If $X_n \xrightarrow{P} X$, then $f(X_n) \xrightarrow{P} f(X)$.

(ii) If $X_n \to X$ a.s. then $f(X_n) \to f(X)$ a.s.

C.2 Convergence in distribution

An even weaker form of convergence than convergence in probability is convergence in distribution.

Definition C.2.1. A sequence of random variables $X_1, X_2, ...$ is said to *converge in distribution* to X if for every continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$,

$$\int f(X_n)dP \to \int f(X)dP.$$

We write $X_n \xrightarrow{\mathrm{d}} X$.

Remark C.2.2. Convergence in distribution is also called *weak convergence*.

The above definition is difficult to check in practice. The following results provide much easier ways to check convergence in distribution.

Theorem C.2.3 (**Helly–Bray**). *Let $X, X_1, X_2, ...$ be random variables with distribution functions $F, F_1, F_2, ...$ We have $X_n \xrightarrow{\mathrm{d}} X$ if and only if there exists a dense subset $A \subseteq \mathbb{R}$ such that $F_n(x) \to F(x)$ for $x \in A$. In this case, A can be chosen to be the set of continuity points of F.*

Proof. See Theorem 18.4 in [68] or Theorem 6.18 in [54]. ■

Proposition C.2.4. *If $X_n \xrightarrow{\mathrm{P}} X$, then $X_n \xrightarrow{\mathrm{d}} X$.*

Proof. See Theorem 18.2 in [68] or Lemma 6.12 in [54]. ■

We now state a version of the Central Limit Theorem, often abbreviated CLT.

Theorem C.2.5 (**Central Limit Theorem**). *Let $\{X_i\}$ be iid with $E[X_1^2] < \infty$, $\mu = E[X_1]$ and $\sigma^2 = \mathrm{Var}(X_1)$. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0,1).$$

Proof. See Theorem 21.1 in [68] or Theorem 10.15 in [38]. ■

In this course, this version of the CLT suffices. Nevertheless, it is useful to know that more general versions of the CLT exist.

Theorem C.2.6 (**Multivariate Central Limit Theorem**). *Let $\{\mathbf{X}_i\}$ be iid random vectors in $\mathbb{R}^k$. Assume that $E[\|\mathbf{X}_1\|^2] < \infty$. Then it holds that*

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - E[\mathbf{X}_1]\Big) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathrm{Cov}[\mathbf{X}_1]).$$

The continuous mapping theorem also holds for convergence in distribution.

Theorem C.2.7 (**Continuous mapping theorem**). *Let f be a continuous function and assume $X_n \xrightarrow{d} X$. Then $f(X_n) \xrightarrow{d} f(X)$.*

The results presented here are only a small part of the whole story. More perspectives and versions of the CLT can be found in chapter 7 of [54].

Theorem C.2.8 (**Delta method**). *Let $\mathbf{X}_1, \mathbf{X}_2, ...$ be random vectors in $\mathbb{R}^k$. If $f : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable in ξ with Jacobian $Df(\xi)$ and*

$$\sqrt{n}(\mathbf{X}_n - \xi) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma),$$

then

$$\sqrt{n}(f(\mathbf{X}_n) - f(\xi)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, Df(\xi)\Sigma Df(\xi)^T).$$

C.3 Exercises

Exercise C.1:
Prove Chebyshev's inequality, Corollary C.1.6.

Exercise C.2:
Without using the Strong Law of Large Numbers, prove the *Weak Law of Large Numbers*: If $\{X_i\}$ is an iid sequence of random variables with $E[X_1^2] < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{P}} E[X_1].$$

Exercise C.3:
Assume $X_n \to X$ a.s. and that f is a continuous function. Prove that $f(X_n) \to f(X)$ a.s.

Exercise C.4:
Let $\{X_i\}$ be iid variables with $E[X_1^2] = 4$ and $E[X_1] = 1$. Show that

$$\lim_{n\to\infty} \frac{X_1^2 + \cdots + X_n^2}{X_1 + \cdots + X_n}$$

exists a.s. and determine the value.

Exercise C.5:
Let $p \geq 1$. A sequence of random variables $\{X_i\}$ with $E[|X_i|^p] < \infty$ for all i is said to converge to X in L^p if $E[|X|^p] < \infty$ and

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0.$$

In that case, we write $X_n \xrightarrow{L^p} X$.

1)Prove that if $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{\text{P}} X$.

2)Let $\{X_i\}$ be a sequence of random variables with $E[X_i] = 0$ and $E[X_i^2] < \infty$ for all i. Prove that $(X_1 + \cdots + X_n)/n$ converges to zero in L^2 and in probability.

3)Prove the following dominated convergence statement: If $X_n \to X$ a.s. and there exists some variable Y with $E[|Y|^p] < \infty$ $(p \geq 1)$ such that $|X_n| \leq |Y|$ a.s. for all n, then $X_n \xrightarrow{L^p} X$.

Exercise C.6:
Assume X is a random variable with finite moment-generating function κ in the neighbourhood $(-c, c)$. Prove *Chernoff's bound*

$$P(X > \varepsilon) \leq \inf_{\alpha\in[0,c]} \frac{\kappa(\alpha)}{e^{\alpha\varepsilon}}.$$

Exercise C.7:
In this exercise, we prove (v) in Theorem B.4.6.

1)Prove the result when $X = 1_{A_0}$ is an indicator function.

2)Prove the result when X is a simple function, e.g. $X = \sum_{i=1}^{n} c_i 1_{A_i}$, $A_1, ..., A_n \in \mathcal{G}$.

3)One can show the following dominated convergence statement for conditional expectations: If $X_n \to X$ a.s. and $|X_n| \leq Y$ for a random variable Y with $E[|Y|] < \infty$, then $E[X_n \mid \mathcal{G}] \to E[X \mid \mathcal{G}]$. Using this result, prove (v) for a general $\mathcal{G}$-measurable X. Hint: Recall that there exists a sequence of simple functions $\{X_n\}$ with $X_n \uparrow X$ when $X \geq 0$. Then extend to general X.

Exercise C.8:
Let $X_1, ..., X_n$ be an iid sample of real-valued random variables with common distribution function F. Let

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}$$

be the *empirical distribution function* based on this sample.

1)Prove that

$$F_n(x) \to F(x) \quad \text{a.s.}$$

for every $x \in \mathbb{R}$.

Note: The Glivenko–Cantelli theorem states that this convergence even holds uniformly in x,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \quad \text{a.s.}$$

2)Prove that

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\text{d}} \mathcal{N}(0, F(x)(1 - F(x))).$$

3)Use the result of the previous subproblem to describe how to make asymptotic confidence bands for the empirical distribution function. Simulate some data from a distribution of your choice and plot the empirical distribution function, the true distribution function and the 95% confidence bands.

Exercise C.9: Exercise 7.3 in [54]
Let $X_1, X_2, ...$ and $Y_1, Y_2, ...$ be independent real-valued random variables. Assume that the X_i are iid Exp(λ) distributed and that the Y_i are iid Exp(λ^{-1}) distributed where $\lambda > 0$. Consider the estimator based on $X_1, ..., X_n, Y_1, ..., Y_n$ given by

$$\widehat{\lambda} = \sqrt{\frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i}}.$$

1)Show that

$$\widehat{\lambda} \to \lambda \quad \text{a.s.}$$

2)Show that

$$\sqrt{n}\Big(\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} X_i \\ Y_i \end{pmatrix} - E\Big[\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}\Big]\Big)$$

converges weakly to a normal distribution and determine the asymptotic mean and variance.
3)Show that

$$\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \lambda^2/2).$$

Hint: Use the delta method.

Appendix D

Stochastic processes

In this chapter, we provide a very brief review of stochastic processes in continuous time.

D.1 Basic definitions

Definition D.1.1. A *(continuous time) stochastic process* is a collection of random variables $\{X_t\}$ indexed by $t \in [0, \infty)$.

To model a flow of information in continuous time, we need the notion of a filtration.

Definition D.1.2. A filtration is a sequence $\{\mathcal{F}_t\}$ of sigma-algebras indexed by $t \in [0, \infty)$ such that $s \leq t$ implies $\mathcal{F}_s \subseteq \mathcal{F}_t$. We have $\mathcal{F}_0 = \{\emptyset, \Omega\}$ by convention.

Definition D.1.3. A stochastic process $\{X_t\}$ is called *adapted* to the filtration $\{\mathcal{F}_t\}$ if X_t is $\mathcal{F}_t$-measurable for all t.

Example D.1.4. For any stochastic process $\{X_t\}$, we can create a filtration by letting $\mathcal{F}_t = \sigma(X_s : s \leq t)$. This is the smallest filtration such that $\{X_t\}$ is adapted. This filtration is sometimes called the *natural filtration*. ∘

A particularly nice type of stochastic process is a martingale.

Definition D.1.5. A stochastic process $\{X_t\}$ is called a *martingale* with respect to the filtration $\{\mathcal{F}_t\}$ if the following hold:

- $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$.
- $E[|X_t|] < \infty$ for each t.
- For every $s \leq t$, $E[X_t \mid \mathcal{F}_s] = X_s$.

Martingales do not play a major role in this course. Nevertheless, some examples are provided in the exercises.

D.2 Essential examples

Brownian motion and Poisson processes

The following stochastic process is the cornerstone of most applications of stochastic processes in a finance and risk management context. It is also the building block for more complicated processes.

Definition D.2.1 (**Brownian motion**). A stochastic process $\{X_t\}$ which satisfies the properties

- $X_0 = 0$,
- $X_t - X_s \sim \mathcal{N}(0, t-s)$ for all $0 \leq s < t$ and
- $X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ are independent for any $0 < t_1 < t_2 < \cdots < t_n$

is called a *(standard) Brownian motion*.

The definition more or less tells us how to simulate a Brownian motion. The implementation is easy and is left to the reader.

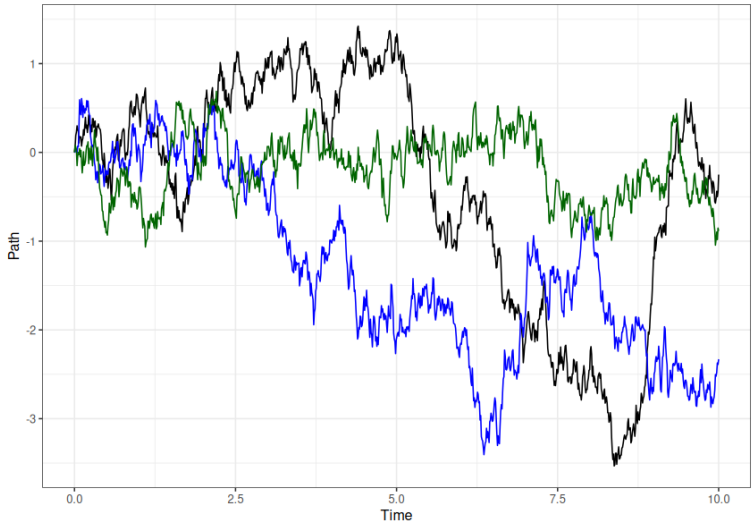


Figure 1: Three simulated paths of a Brownian motion from time 0 to 10.

Definition D.2.2. A stochastic process $\{N_t\}$ satisfying the properties

- $N_0 = 0$,
- $N_t - N_s$ is Poisson distributed with parameter $\lambda(t-s)$ for any $t \geq s$,
- $N_{t_1}, N_{t_2} - N_{t_1}, ..., N_{t_n} - N_{t_{n-1}}$ are independent for $0 < t_1 < t_2 < \cdots < t_n$,
- $\{N_t\}$ has right-continuous sample paths and limits from the left,

then $\{N_t\}$ is called a *Poisson process (homogeneous)* with *intensity* $\lambda > 0$.

In general, a *counting process* $\{N_t\}$ is any continuous time stochastic process which only attains values in $\mathbb{N}_0$ and satisfies $N_t \geq N_s$ for any $t \geq s$. A Poisson process is obviously an example of a counting process. The *arrival times* for a counting process are defined by

$$T_k := \inf\{t \geq 0 : N_t = k\}, \quad k \in \mathbb{N}$$

and $T_0 := 0$. The *interarrival times* are defined by $\tau_i := T_i - T_{i-1}$ for $i = 1, 2, \ldots$. The counting process is uniquely determined by its arrival times since

$$N_t = \sum_{i=0}^{\infty} 1_{\{T_i \leq t\}}.$$

The following characterisation of a Poisson process in terms of its interarrival times makes it trivial to simulate the process.

Theorem D.2.3. *The process $\{N_t\}$ is a homogeneous Poisson process with intensity λ if and only if $\{N_t\}$ is a counting process where the interarrival times for the process are iid Exp(λ).*

Proof. See Theorem 2.1.6 in e.g. [89]. ■

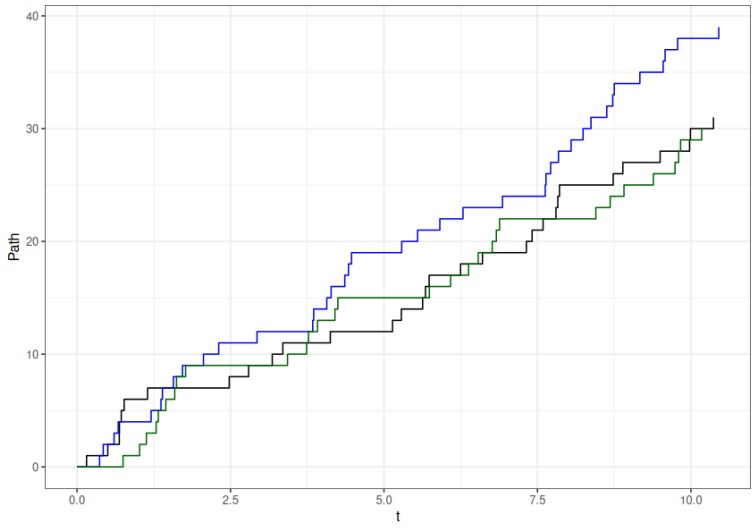


Figure 2: Three simulated paths of a Poisson process with intensity 3. The expected number of jumps is 30 at time 10, which seems to concur at least for two of the paths.

In general, a counting process with iid interarrival times is called a *renewal process*.

D.3 SDE's

Setup and basic properties

In finance and risk management, *stochastic differential equations* are an essential tool in modelling prices of financial assets. We exclusively consider so-called *Itô processes*, which are stochastic processes $\{X_t\}$ of the form

$$X_t = x_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

with $X_0 = x_0 \in \mathbb{R}$ a fixed value and $\{W_t\}$ a Brownian motion. We often use the more compact *dynamic* notation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0.$$

The dynamic notation is simply a reformulation of the integral notation above and thus has no additional mathematical meaning. But it does communicate more clearly how the process $\{X_t\}$ behaves, since we can think of dX_t as the infinitesimal behaviour of the process. It also makes it clear how to simulate from such a process on the interval $[0, t]$. Choose a sufficiently fine discretisation $0 = t_0 < t_1 < \cdots < t_n = t$ and let

$$X_0 = x_0, \quad X_{t_{i+1}} = X_{t_i} + \mu(t_i, X_{t_i})(t_{i+1} - t_i) + \sigma(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i}), \quad i = 0, ..., n-1.$$

Recalling that the increments $W_{t_{i+1}} - W_{t_i}$ are independent and $\mathcal{N}(0, t_{i+1} - t_i)$ distributed, one can simply replace these with an independent sequence $\{Z_i\}$, $Z_i \sim \mathcal{N}(0, t_{i+1} - t_i)$ in the formula above. This scheme is called the *Euler–Maruyama method* or *Euler–Maruyama scheme*.

What is the exact mathematical meaning of the integral with respect to W? While these details are not important in this book, one should be aware that for a process Y the stochastic integral

$$\int_0^t Y_s dW_s$$

is *not* a Lebesgue–Stieltjes integral in the sense of the next chapter. The construction has to be different since a Brownian motion has unbounded variation on every interval a.s. Very roughly, the stochastic integral with respect to a Brownian motion can be constructed in the following way (we here follow [9], for different and more rigorous treatments, see e.g. [121] and [14]). If Y is simple, $Y_s = Y_{t_k}$ for $s \in [t_k, t_{k+1})$ where $a = t_0 < t_1 < \cdots < t_n = b$ is a partition of $[a, b]$, one makes the obvious definition

$$\int_a^b Y_s dW_s = \sum_{k=0}^{n-1} Y_{t_k}(W_{t_{k+1}} - W_{t_k}).$$

Now, given any sufficiently integrable process Y on $[a, b]$, where sufficiently integrable here means that

$$\int_a^b E[Y_s^2]ds < \infty,$$

one approximates Y with a sequence $\{Y^n\}$ of simple processes satisfying

$$\int_a^b E[(Y_s^n - Y_s)^2]ds \to 0.$$

It is then possible to show that the integral $\int_a^b Y_s^n dW_s$ converges in L^2 to some random variable. We then define the stochastic integral $\int_a^b Y_s dW_s$ as the limit of this sequence of integrals. Some useful facts of the stochastic integral are the following.

Proposition D.3.1. *Let $\{Y_t\}$ be a stochastic process adapted to the natural filtration $\{\mathcal{F}_t^W\}$ generated by the Brownian motion $\{W_t\}$ and assume*

$$\int_a^b E[Y_s^2]ds < \infty.$$

Then

1. *(Stochastic integrals have mean zero)*

$$E\Big[\int_a^b Y_s dW_s\Big] = 0,$$

2. *(Itô isometry)*

$$E\Big[\Big(\int_a^b Y_s dW_s\Big)^2\Big] = \int_a^b E[Y_s^2]ds$$

3. *(Measurability of stochastic integrals)*

$$\int_a^b Y_s dW_s \quad \textit{is } \mathcal{F}_b^W\textit{-measurable.}$$

The following deep result is also a core component of the theory of stochastic integration.

Proposition D.3.2. *Modulo integrability conditions, an Itô process with dynamics $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$, $X_0 = 0$ is a martingale with respect to the natural filtration generated by W if and only if $\mu = 0$, that is, X has no dt-term.*

In this book, we will not have to worry about the integrability and measurability conditions which arise in stochastic calculus.

The Itô formula

Given an Itô process

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0$$

and a function $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ which is C^1 in the first coordinate and C^2 in the other, what can we say about $f(t, X_t)$? The Itô formula provides an answer. A function f as just described will be referred to as a $C^{1,2}$-function.

Theorem D.3.3 (**Itô's formula**). *For an Itô process*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0$$

and a $C^{1,2}$-function f, $f(t, X_t)$ is again an Itô process with dynamics

$$df(t, X_t) = \Big(\frac{\partial f}{\partial t}(t, X_t) + \mu(t, X_t)\frac{\partial f}{\partial x}(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2\frac{\partial^2 f}{\partial x^2}(t, X_t)\Big)dt + \sigma(t, X_t)\frac{\partial f}{\partial x}(t, X_t)dW_t.$$

Example D.3.4. Let us compute the dynamics of $Y_t = e^{W_t}$. Here we have $Y_t = f(t, X_t)$ for $f(t, x) = e^x$, and thus

$$\frac{\partial f}{\partial t} = 0, \quad \frac{\partial f}{\partial x} = e^x, \quad \frac{\partial^2 f}{\partial x^2} = e^x.$$

Plugging into the Itô formula, we get (use that $dW_t = 0dt + 1dW_t$)

$$dY_t = \frac{1}{2}1^2 Y_t dt + 1 \cdot Y_t dW_t = \frac{1}{2}Y_t dt + Y_t dW_t, \quad Y_0 = 1.$$

Note how this differs from ordinary calculus. There we have

$$\frac{d}{dx}e^{g(x)} = g'(x)e^{g(x)}$$

which can be written as $d(e^{g(x)}) = e^{g(x)}dg(x)$. Here the "1/2"-term does not show up. ∘

The Itô formula is probably the most important tool in mathematical finance. You can practice it in some of the exercises below. There is also a multidimensional version of the Itô formula, see e.g. Chapter 4 of [9].

Geometric Brownian Motion

For modelling asset processes, the *Geometric Brownian Motion (GBM)* is usually the starting point. This is the Itô process with

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0,$$

where μ and σ are real constants. In the exercises, you are asked to verify that the stochastic process

$$X_t = x_0 e^{(\mu - \sigma^2/2)t + \sigma W_t}$$

is a solution to this SDE. The fact that an exact solution exists is one of the reasons why GBM's are often used in financial models. Another reason is the fact that the above solution is a.s. positive if $x_0 > 0$.

D.4 Exercises

Exercise D.1:
Let $\{X_t\}$ be a Brownian motion

1)Define $Y_t = -X_t$. Show that $\{Y_t\}$ is a Brownian motion.

2)Let $c > 0$ and define $Y_t = cX_{t/c^2}$. Show that $\{Y_t\}$ is a Brownian motion.

Exercise D.2:
A continuous time process $\{X_t\}$ satisfies *continuity in probability* if for every sequence $\{t_n\}$ of non-negative real numbers, we have

$$t_n \to t \quad \Rightarrow \quad X_{t_n} \xrightarrow{\text{P}} X_t.$$

Show that a Brownian motion satisfies continuity in probability.

Exercise D.3:
Let $\{X_t\}$ be a Brownian motion and $\mathcal{F}_t = \sigma(X_s : s \leq t)$ the natural filtration.

1)Show that $\{X_t\}$ is a martingale with respect to $\{\mathcal{F}_t\}$.

2)Show that $\{X_t^2 - t\}$ is a martingale with respect to $\{\mathcal{F}_t\}$.

Exercise D.4:
Let $\{N_t\}$ denote a homogeneous Poisson process with intensity λ. Verify that $\{N_t - \lambda t\}$ is a martingale with respect to the natural filtration.

Exercise D.5:
In this exercise, you will simulate paths of Brownian motion and Poisson processes.

1)Describe and implement an algorithm to simulate from a Brownian motion. Then "recreate" the plot in the text.

2)Describe and implement an algorithm to simulate from a homogeneous Poisson process. Then "recreate" the plot in the text.

Exercise D.6:
In this exercise, we show that the stochastic process given by

$$X_t = x_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$$

is a solution to the stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0.$$

1)Identify the function f such that $X_t = f(t, W_t)$. Compute $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial x}$ and $\frac{\partial^2 f}{\partial x^2}$.

2)Apply the Itô formula to show that X_t satisfies the stochastic differential equation.

3)Prove that the expected value is given by

$$E[X_t] = x_0 e^{\mu t}.$$

4)Simulate 10 paths of X by using the Euler–Maruyama method presented in the text using parameter values of your choice. Plot the 10 paths along with the theoretical mean.

Exercise D.7:
Show that

$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{t}{2}$$

by computing $d(W_t^2)$. How does this differ from ordinary calculus?

Exercise D.8:
Let g be some deterministic function and consider the process

$$X_t = \int_0^t g(s) dW_s.$$

1)Use the Itô formula to compute $d(e^{\alpha X_t})$ for any $\alpha \in \mathbb{R}$.

2)Show that

$$X_t \sim \mathcal{N}\Big(0, \int_0^t g(s)^2 ds\Big).$$

Why does this make sense intuitively?

Exercise D.9:
Suppose X solves the SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0.$$

We know that $E[X_t] = x_0 e^{\mu t}$. In this exercise, we compute the variance of X_t as outlined below.

1)Compute $d(X_t^2)$ using the Itô formula.

2)Compute $E[X_t^2]$ and use this to compute $\text{Var}[X_t]$. Is the value as expected? Hint: For the latter question, what is the distribution of X_t?

Exercise D.10:
Suppose X solves the SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0.$$

Define Y by $Y_t = X_t^\alpha$ where $\alpha \in \mathbb{R}$. Show that Y is again a GBM. What SDE does Y satisfy?

Appendix E

Finite variation calculus

E.1 Functions of bounded variation

During the course, we will occasionally integrate with respect to functions of bounded variation. Examples of such functions include functions that are monotone, in particular distribution functions. Here we introduce the basic theory, following the lines of [95].

Definition E.1.1. Let $f : [0, \infty) \to \mathbb{R}$ be a function. The *variation* of f on the interval $[0, t]$ is given by

$$V^f(t) = \sup \left\{ \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| : 0 = t_0 < t_1 < \cdots < t_n = t \right\}$$

i.e. the supremum of sums of absolute differences over all finite partitions of $[0, t]$. If $V^f(t) < \infty$ for all $t \geq 0$, we call f a *function of (locally) bounded variation.*

Example E.1.2. Let $f : [0, \infty) \to \mathbb{R}$ be monotone. We claim that f is of bounded variation. If f is non-decreasing, this follows immediately from the fact that if $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[0, t]$, we have

$$\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| = \sum_{i=1}^{n} f(t_i) - f(t_{i-1}) = f(t) - f(0)$$

by telescoping. Hence $V^f(t) = f(t) - f(0)$ and f is of bounded variation. An analogous argument works for the case where f is non-increasing. ◦

Recall that for $x \in \mathbb{R}$, $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$. It is easily seen that $x = x^+ - x^-$.

Definition E.1.3. For a function $f : [0, \infty) \to \mathbb{R}$, we define the *positive variation* by

$$V_+^f(t) = \sup \left\{ \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^+ : 0 = t_0 < t_1 < \cdots < t_n = t \right\}$$

and the *negative variation* by

$$V_-^f(t) = \sup \left\{ \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^- : 0 = t_0 < t_1 < \cdots < t_n = t \right\}.$$

Proposition E.1.4. *Let $f, g : [0, \infty) \to \mathbb{R}$ be functions of bounded variation.*

(i) V^f, V^f_+ and V^f_- are non-decreasing.

(ii) $af + bg$ is a function of bounded variation for any $a, b \in \mathbb{R}$.

Proof. Left as an exercise for the reader. ■

The motivation for introducing the negative and positive variation is the following central result.

Theorem E.1.5 (Jordan decomposition). *A function $f : [0, \infty) \to \mathbb{R}$ is of bounded variation if and only if f can be written as the difference of two non-decreasing functions. A possible decomposition is $f(t) - f(0) = V^f_+(t) - V^f_-(t)$.*

Proof. Assume $f = g - h$ for non-decreasing functions g and h. g and h are of bounded variation as shown in the example above, and the previous proposition now implies that f is of bounded variation. Conversely, assume f is of bounded variation. For any partition $0 = t_0 < t_1 < \cdots < t_n = t$ of $[0, t]$, we have

$$f(t) - f(0) = \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^+ - \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^-$$

i.e.

$$\sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^+ = \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^- + f(t) - f(0)$$

and taking the proper supremum on both sides, we obtain the decomposition $f(t) - f(0) = V^f_+(t) - V^f_-(t)$, which is a difference of two non-decreasing functions as desired. ■

Remark E.1.6. Note that V^f has the decomposition $V^f = V^f_+ + V^f_-$ since for any $x \in \mathbb{R}$, $|x| = x^+ + x^-$.

E.2 Integration

We are almost ready to introduce integration. However, we mention that in addition to finite variation, we also require that the functions we integrate with respect to are càdlàg. The Jordan decomposition tells us how to proceed from here. If we define integration with respect to an increasing function, we can use linearity of the integral to define integration with respect to general functions of bounded variation. Let f be a non-decreasing càdlàg function. The function μ^f defined on the intervals $(a, b]$ given by $\mu^f((a, b]) = f(b) - f(a)$ extends to a (positive) measure (called a *Lebesgue–Stieltjes* measure) on all Borel sets. Note how this resembles the Lebesgue measure where f is just the identity. We can now define integration in the same way as in basic measure theory.

Definition E.2.1. Let $f : [0, \infty) \to \mathbb{R}$ be a non-decreasing càdlàg function. The *Lebesgue–Stieltjes integral* of a measurable function g with respect to f is given by

$$\int_{(0,\infty)} g(t) df(t) := \int_0^\infty g(t) d\mu^f(t)$$

given that $\int_0^\infty |g(t)|d\mu^f(t) < \infty$. For any Borel set B, we define

$$\int_B g(t)df(t) := \int_{(0,\infty)} 1_B(t)g(t)df(t).$$

If f is a càdlàg function of bounded variation, we have the Jordan decomposition $f(t) - f(0) = V_+^f(t) - V_-^f(t)$, and we define the Lebesgue–Stieltjes integral of g by

$$\int_0^\infty g(t)df(t) = \int_0^\infty g(t)dV_+^f(t) - \int_0^\infty g(t)dV_-^f(t).$$

The integral is well-defined whenever

$$\int_0^\infty |g(t)|dV^f(t) = \int_0^\infty |g(t)|dV_+^f(t) + \int_0^\infty |g(t)|dV_-^f(t) < \infty.$$

Example E.2.2. Let F be the distribution function for the random variable X. If $B = (a, b]$, then $P(X \in B) = F(b) - F(a) = \int_B dF$. Since the intervals $(a, b]$ are $\cap$-stable and generate the Borel sigma-algebra, we have $P(X \in B) = \int_B dF$ for all Borel sets B. ∘

Example E.2.3. Consider a positive random variable X with finite expectation and distribution function F. We claim that

$$E[g(X)] = \int_0^\infty g(x)dF(x)$$

for any measurable real-valued function g. Since, by the previous example, the image measure of X, P^X, coincides with the measure induced by F, we have

$$E[g(X)] = \int_0^\infty g(x)dP^X(x) = \int_0^\infty g(x)dF(x)$$

as desired. ∘

We now go through some important properties of the Lebesgue–Stieltjes integral.

Proposition E.2.4. *Let f be a càdlàg function of bounded variation. Assume all integrals below are well-defined.*

(i)

$$\int_{(s,t]} df(u) = f(t) - f(s).$$

(ii)

$$\int_{\{t\}} g(u)df(u) = g(t)\Delta f(t), \quad \Delta f(t) := f(t) - f(t-)$$

with $f(t-) = \lim_{s\uparrow t} f(s)$ the limit from the left.

(iii)

$$\int_{(s,t]} g(u)df(u) = 0$$

if f is constant on $(s, t]$.

Proof. Left as an exercise for the reader. ■

These properties will allow us to compute integrals with respect to functions that are piecewise constant. An important example is the empirical distribution function.

Example E.2.5. Let $X_1, ..., X_n$ denote a sample and let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}$$

denote the *empirical distribution function* based on this data. Such a function is sketched below.

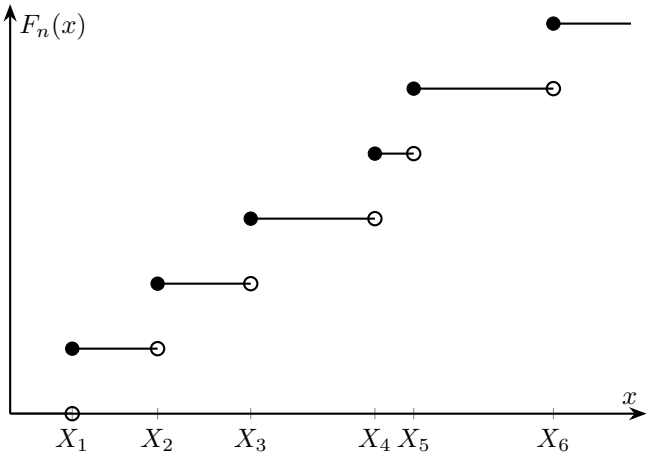


Figure 1: The empirical distribution function based on a sample of size $n = 6$. Each jump has size $1/6$.

This function satisfies all the requirements of a distribution function, and so in particular, it is càdlàg and of bounded variation. Hence it makes sense to consider the integral

$$\int_{-\infty}^{\infty} f(x) dF_n(x)$$

where f is some measurable function. We see that F_n is flat in the regions (X_i, X_{i+1}), and so these regions contribute nothing to the integral. This is part (iii) of the proposition above. All jumps are of size $1/n$ and occur precisely in the X_i, so we get by part (ii) of Proposition E.2.4 that

$$\int_{-\infty}^{\infty} f(x) dF_n(x) = \sum_{i=1}^{n} f(X_i)\Delta F_n(X_i) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

In particular, when f is the identity, we get

$$\int_{-\infty}^{\infty} x dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

which is the empirical mean. ∘

Example E.2.6. The previous example is useful for coming up with natural estimators for quantities of interest. For example, if X is a random variable with finite expectation, the *mean-excess function* is given by

$$e(u) = E[X - u \mid X > u].$$

This function is useful for methods in extreme value theory (see Chapter 2). We may rewrite

$$e(u) = \frac{E[(X-u)1_{\{X>u\}}]}{P(X>u)} = \frac{1}{\overline{F}(u)} \int_{-\infty}^{\infty} (x-u) 1_{\{x>u\}} dF(x)$$

where F is the distribution function of X. A natural estimator is to simply replace F by the empirical distribution function F_n based on a sample $X_1, ..., X_n$. Then we get the estimator

$$e_n(u) = \frac{1}{\overline{F}_n(u)} \int_{-\infty}^{\infty} (x-u) 1_{\{x>u\}} dF_n(x).$$

Using what we just learned about integrating with respect to the empirical distribution function, we get

$$e_n(u) = \frac{\frac{1}{n}\sum_{i=1}^{n} (X_i - u) 1_{\{X_i > u\}}}{\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > u\}}}$$

which is the estimator derived in Chapter 2. ∘

We end this subsection with the following key result.

Theorem E.2.7 (**Integration by parts**). *Let f and g be càdlàg functions of bounded variation. Then (assuming all integrals are well-defined)*

$$\begin{aligned} f(t)g(t) - f(0)g(0) &= \int_{(0,t]} g(s) df(s) + \int_{(0,t]} f(s-) dg(s) \\ &= \int_{(0,t]} g(s-) df(s) + \int_{(0,t]} f(s-) dg(s) + \sum_{0<s\leq t} \Delta g(s) \Delta f(s). \end{aligned}$$

Proof. Note first that

$$f(t)g(t) - f(0)g(0) - f(0)(g(t) - g(0)) - g(0)(f(t) - f(0)) = (f(t) - f(0))(g(t) - g(0)).$$

The result is now a direct consequence of the following computation based on Fubini's theorem:

$$\begin{aligned} (f(t) - f(0))(g(t) - g(0)) &= \int_{(0,t]} \int_{(0,t]} df(u) dg(s) = \int_{(0,t]} \int_{(0,s]} df(u) dg(s) + \int_{(0,t]} \int_{(s,t]} df(u) dg(s) \\ &= \int_{(0,t]} (f(s) - f(0)) dg(s) + \int_{(0,t]} \int_{(0,u)} dg(s) df(u) \\ &= \int_{(0,t]} f(s) dg(s) - f(0)(g(t) - g(0)) + \int_{(0,t]} g(u-) - g(0) df(u) \\ &= \int_{(0,t]} f(s) dg(s) - f(0)(g(t) - g(0)) \\ &+ \int_{(0,t]} g(u-) df(u) - g(0)(f(t) - f(0)). \end{aligned}$$

■

Remark E.2.8. It is not difficult to see that the result also works for other intervals such as $(a, b]$, (t, ∞) etc.

Example E.2.9. In the case of a non-negative random variable X, the formula

$$E[X] = \int_0^\infty P(X > x)dx = \int_0^\infty \overline{F}(x)dx$$

for the case $E[X] < \infty$ should be well-known. One can derive this formula directly using Fubini's theorem, but it can also be derived with integration by parts as long as one makes the additional assumption

$$\lim_{x\to\infty} x\overline{F}(x) = 0.$$

Indeed, from Example E.2.3,

$$E[X] = \int_0^\infty x dF(x).$$

Now use that $dF(x) = -d\overline{F}(x)$ and integration by parts to get

$$E[X] = -\int_0^\infty x d\overline{F}(x) = -[x\overline{F}(x)]_0^\infty + \int_0^\infty \overline{F}(x)dx = \int_0^\infty \overline{F}(x)dx.$$

Of course, the classical proof using Fubini does not rely on the assumption that $x\overline{F}(x) \to 0$ for $x \to \infty$ and is thus more general. Nevertheless, this example illustrates a simple application of integration by parts. ∘

E.3 Exercises

Exercise E.1:
Prove Proposition E.1.4.

Exercise E.2:
Prove Proposition E.2.4.

Exercise E.3:
In this exercise, we consider some classes of functions of bounded variation.

1)Let $f : [0, \infty) \to \mathbb{R}$ be a function which is Lipschitz on every compact interval, i.e. for every $[a, b] \subseteq [0, \infty)$, there exists a constant C (depending on $[a, b]$) such that

$$|f(s) - f(u)| \leq C|s - u|, \quad s, u \in [a, b].$$

Prove that f is of bounded variation.

2)Let $f : [0, \infty) \to \mathbb{R}$ be C^1. Prove that f is Lipschitz on every compact interval and conclude that f is of bounded variation.

Exercise E.4:
Compute the integral $\int_{(0,4]} t df(t)$ in the following cases:

1) $f(t) = k$ for $k - 1 \leq t \leq k$, $k = 1, 2, ...$

2) $f(t) = e^t$.

3) $f(t) = k + e^t$ for $k - 1 \leq t \leq k$, $k = 1, 2, ...$

Exercise E.5:
In this exercise, we provide a proof of the well-known fact

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

based on integration by parts.

1)Let $f : [0, \infty) \to \mathbb{R}$ be a piecewise constant function satisfying the following properties: $f(0) = 0$, f is non-decreasing, $f(s) \to \infty$ for $s \to \infty$ and $\Delta f(s) \in \{0, 1\}$. Choose $t > 0$ such that $f(t) = n + 1$. Show that

$$\int_{(0,t]} f(s-) df(s) = \sum_{i=1}^{n} i.$$

Hint: Make a drawing!

2)Now apply integration by parts to show that

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

3)(**Bonus**) Can you come up with an example of a function satisfying the properties in the first subproblem? Hint: Think (classical) non-life insurance.

Bibliography

[1] J. Aczél. *Lectures on Functional Equations and Their Applications*. Academic Press New York, 1966.

[2] M. M. Ali, N. N. Mikhail, and M. S. Haq. A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8(3):405–412, 1978.

[3] P. Artzner, F. Delbaen, J-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999. doi: https://doi.org/10.1111/1467-9965.00068.

[4] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Stochastic Modelling and Applied Probability. Springer New York, NY, 1 edition, 2007. ISBN 978-0-387-30679-7.

[5] S. Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer Cham, 4 edition, 2023. ISBN 978-3-031-41025-3.

[6] N. Balakrishna and C. D. Lai. *Bivariate Extreme-Value Distributions*. Springer New York, 2009. ISBN 978-0-387-09614-8.

[7] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., anniversary edition, 2012. ISBN 978-1-118-12237-2.

[8] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1987. doi: 10.1017/CBO9780511721434.

[9] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 4 edition, 2020. ISBN 978–0–19–885161–5.

[10] N. Blomqvist. On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, 21(4):593–600, 1950.

[11] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, (31):307–327, 1986. URL `https://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdf`.

[12] W. Breymann, A. Dias, and P. Embrechts. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):2003, 2003.

[13] X. Chen and Y. Fan. Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 33(3):389–414, 2005.

[14] S. N. Cohen and R. J. Elliott. *Stochastic Calculus and Applications*. Probability and Its Applications. Birkhäuser New York, NY, 2 edition, 2015. ISBN 978-1-4939-2866-8.

[15] D. A. Conway. *Multivariate Distributions with Specified Marginals*. Stanford University, 1979.

[16] C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[17] H. E. Daniels. Rank correlation and population models. *Journal of the Royal Statistical Society*, 12(2):171–191, 1950.

[18] A. C. Davidson and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*, 52(3):393–442.

[19] A. C. Davison. *Modelling Excesses over High Thresholds, with an Application*, pages 461–482. Springer Netherlands, 1984. ISBN 978-94-017-3069-3.

[20] L. de Haan and S. Resnick. On asymptotic normality of the hill estimator. *Communications in Statistics. Stochastic Models*, 14(4):849–866, 1998.

[21] A. L. M. Dekkers and L. de Haan. On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*, 17(4):1795–1832, 1989.

[22] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855, 1989.

[23] A. Dembo, J-D. Deuschel, and D. Duffie. Large portfolio losses. *Finance and Stochastics*, 2004. URL `https://link.springer.com/article/10.1007/s00780-003-0107-2`.

[24] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons Inc., 2005. ISBN 978-0-470-01645-9.

[25] L. Devroye. *Non-Uniform Random Variate Generation*. Springer New York, 1 edition, 1986. ISBN 978-1-4613-8643-8.

[26] F. Durante and C. Sempi. Copula and semi-copula transforms. *International Journal of Mathematics and Mathematical Sciences*, 2005(4).

[27] J. Durbin and A. Stuart. Inversions and rank correlation coefficients. *Journal of the Royal Statistical Society*, 12(2):303–309, 1951.

[28] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. URL `http://www.jstor.org/stable/2958830`.

[29] P. Embrechts and M. Hofert. A note on generalized inverses. 2014. URL `https://people.math.ethz.ch/~embrecht/ftp/generalized_inverse.pdf`.

[30] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 1 edition, 1997. ISBN 978-3-540-60931-5.

[31] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912773.

[32] M. Falk. On testing the extreme value index via the pot-method. *The Annals of Statistics*, 23(6):2013–2025, 1995.

[33] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, Inc., 2 edition, 1966.

[34] T. S. Ferguson. A class of symmetric bivariate uniform distributions. *Stat Papers*, 36: 31–40.

[35] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25, 1998.

[36] M. Fréchet. Remarques au sujet de la note précédente, comptes rendus des séances academie de sciences 246, 2719 - 2720, 1958.

[37] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. 4 edition, 2016. ISBN 978-3-110-46345-3.

[38] J-F. Le Gall. *Measure Theory, Probability, and Stochastic Processes*. Graduate Texts in Mathematics. Springer Cham, 1 edition. ISBN 978-3-031-14205-5. URL https://link.springer.com/book/10.1007/978-3-031-14205-5.

[39] C. Genest. Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555, 1987.

[40] C. Genest and R. J. McKay. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–285, 1986.

[41] C. Genest and R. J. McKay. Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *The Canadian Journal of Statistics*, 14:145–159, 1986.

[42] C. Genest and L. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043.

[43] C. Genest and L-P. Rivest. A characterization of gumbel's family of extreme value distributions. *Statistics & Probability Letters*, 8(3):207–211, 1989.

[44] C. Genest and L-P. Rivest. On the multivariate probability integral transform. *Statistics and Probability Letters*, 53:391–399, 2001.

[45] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213, 2009.

[46] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability. Springer New York, NY, 1 edition, 2003. ISBN 978-0-387-00451-8.

[47] P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. 1996. URL https://web.stanford.edu/~glynn/papers/1996/G96.html.

[48] C. M. Goldie and R. L. Smith. Slow variations with remainder: Theory and applications. *The Quarterly Journal of Mathematics*, 38(1):45–71, 1987.

[49] G. Gudendorf and J. Segers. Extreme-value copulas. 2009. URL `https://arxiv.org/pdf/0911.1015`.

[50] E. J. Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707, 1960.

[51] E. J. Gumbel. Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349, 1961.

[52] L. De Haan and L. Peng. Comparison of tail index estimators. *statistica neerlandica*, 52(1):60–70, 2001.

[53] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society*, 44(1):37–42, 1982.

[54] E. Hansen. *Stochastic Processes*. Institut for Matematiske Fag Københavns Universitet, 4 edition, 2023. ISBN 978-87-71252-59-0.

[55] E. Hintz, M. Hofert, and C. Lemieux. Normal variance mixtures: Distribution, density and parameter estimation. *Computational Statistics & Data Analysis*, 157:107175, 2021.

[56] W. Hoeffding. Masstabinvariante korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5(3): 179–233, .

[57] W. Hoeffding. Masstabinvariante korrelationsmasse für diskontinuierliche verteilungen. *Arkiv für matematischen Wirtschaften und Sozialforschung*, 7:49–70, .

[58] M. Hofert, R. Frey, and A. J. McNeil. *The Quantitative Risk Management Exercise Book*. Princeton University Press, 2020. ISBN 978-0-691-20670-7.

[59] Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula: Multivariate Dependence with Copulas*, 2025. URL `https://CRAN.R-project.org/package=copula`. R package version 1.1-6.

[60] P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396, 1986.

[61] P. Hougaard, B. Harvald, and N. V. Holm. Measuring the similarities between the lifetimes of adult danish twins born between 1881-1930. *Journal of the American Statisticial Association*, 87(417):17–24, 1992.

[62] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. John Wiley & Sons Inc., 2 edition. ISBN 978-0-470-12990-6.

[63] H. Hult and F. Lindskog. *Mathematical Modeling and Statistical Methods for Risk Management*. 2007.

[64] J. Pickands III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131, 1975.

[65] R. L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311–334, 1982.

[66] Ivan Kojadinovic and Jun Yan. Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20, 2010. URL `https://www.jstatsoft.org/v34/i09/`.

[67] J. Teugels J. Segers J. Beirlant, Y. Goegebeur. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Ltd., 2004. ISBN 978-0-470-01238-3.

[68] J. Jacod and P. Protter. *Probability Essentials*. Universitext. Springer Berlin, Heidelberg, 2 edition, 2002. ISBN 978-3-642-55682-1.

[69] Kumar Joag-Dev. 4 measures of dependence. 4:79–88, 1984.

[70] H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC. ISBN 978-0-3678-0389-6.

[71] H. Joe. Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46(2):262–282, 1993.

[72] M. E. Johnson. *Multivariate Statistical Simulation*. John Wiley & Sons, Ltd., 1987. ISBN 978-1-118-15074-0.

[73] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley & Sons Inc., 1972. ISBN 9780-471-44370-4.

[74] Jun Yan. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21, 2007. URL `https://www.jstatsoft.org/v21/i04/`.

[75] E. P. Klement, R. Mesiar, and E. Pap. Transformations of copulas and quasi-copulas. *Soft Methodology and Random Information Systems. Advances in Soft Computing*, 26, 2004.

[76] S. A. Klugman and R. Parsa. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 24(1):139–148, 1999.

[77] S. Korkmaz. *cryptoQuotes: Open Access to Cryptocurrency Market Data, Sentiment Indicators and Interactive Charts*, 2025. URL `https://cran.r-project.org/package=cryptoQuotes`. R package version 1.3.3.

[78] W. H. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861, 1958.

[79] C. D. Lai, N. Balakrishnan, and P. Hutchinson. *Continuous Bivariate Distributions*. Springer New York, 2 edition. ISBN 978-0-387-09614-8.

[80] M. R. Leadbetter. On a basis for peaks over threshold modeling. *Statistics and Probability Letters*, 12(4):357–362, 1991.

[81] A. J. Lee. *U-Statistics*. Taylor & Francis, 1 edition, 1990. ISBN 978-0-203-73452-0.

[82] R. F. Lemvig. *Stochastic Processes in Non-Life Insurance (SkadeStok) Lecture Notes*. URL `https://rasmusfl.github.io/projects.html`.

[83] X. Li, P. Mikusiński, and M. D. Taylor. Some integration-by-parts formulas involving 2-copulas. *Cuadras, C.M., Fortiana, J., Rodriguez-Lallena, J.A. (eds) Distributions With Given Marginals and Statistical Modelling*, 2002.

[84] K. V. Mardia. *Families of Bivariate Distributions*. Griffin, 1970. ISBN 978-0-852-64182-8.

[85] A. W. Marshall and I. Olkin. A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4(2):291–302, 1967.

[86] A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44, 1967.

[87] A. J. McNeil and J. Nešlehová. Multivariate archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Annals of Statistics*, 37:3059–3097.

[88] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management - Concepts, Techniques and Tools*. Princeton University Press, revised edition, 2015. ISBN 978-0-691-16627-8.

[89] T. Mikosch. *Non-Life Insurance Mathematics*. Universitext. Springer Berlin, Heidelberg, 2 edition, 2009. ISBN 978-3-540-88232-9.

[90] R. B. Nelsen. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics - Theory and Methods*, 15:3277–3285, 1986.

[91] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, NY, 2 edition, 2006. ISBN 978-0-387-28659-4.

[92] J. P. Nolan. *Univariate Stable Distributions*. Springer Series in Operations Research and Financial Engineering. Springer Cham, 1 edition, 2020. ISBN 978-3-030-52915-4.

[93] D. Oakes. Multivariate survival distributions. *Journal of Nonparametric Statistics*, 3 (3-4):343–354, 1994.

[94] E. Häusler P. Deheuvels and D. M. Mason. Almost sure convergence of the hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(2): 371–381, 1988.

[95] J. L. Pedersen. *Stochastic Processes in Life Insurance: The Dynamic Approach*. Department of Mathematical Sciences University of Copenhagen.

[96] B. Pfaff, A. McNeil, and A. Stephenson. *evir: Extreme Values in R*, 2018. URL `https://cran.r-project.org/package=evir`. R package version 1.7-4.

[97] J. Pickands. Multivariate extreme value distributions. *Proceedings of the 43rd Session ISI (Buenos Aires)*, 1981.

[98] R. L. Plackett. A class of bivariate distributions. *Journal of the American Statistical Association*, 60(310):516–522, 1965.

[99] A. E. Raftery. A continuous multivariate exponential distribution. *Communications in Statistics - Theory and Methods*, 13:947 – 965, 1984.

[100] A. E. Raftery. Some properties of a new continuous bivariate exponential distribution. *Statistics & Decisions Supplement Issue*, (2):53–58, 1984.

[101] R.-D. Reiss. *Approximate Distributions of Order Statistics*. Springer Series in Statistics. Springer New York, NY, 1 edition, 2012. ISBN 978-1-4613-9620-8.

[102] S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer New York, NY, 1 edition, 2007. ISBN 978-0-387-75952-4.

[103] B. D. Ripley. *Stochastic Simulation*. John Wiley & Sons, Inc., 1987. ISBN 0-471-81884-4.

[104] A. W. Roberts and D. E. Varberg. *Convex functions*. 1 edition, 1974. ISBN 978-0-080-87372-5.

[105] P. J. Rousseeuw and G. Molenberghs. Transformation of non positive semidefinite correlation matrices. *Communications in Statistics: Theory and Methods*, 22(4):965–984.

[106] J. A. Ryan, J. M. Ulrich, E. B. Smith, W. Thielen, P. Teetor, and S. Bronder. *quantmod: Quantitative Financial Modelling Framework*, 2025. URL `https://cran.r-project.org/package=quantmod`. R package version 0.4.28.

[107] L. Rüschendorf. *Mathematical Risk Analysis*. Springer, 2013. ISBN 978-0-470-01645-9.

[108] V. Schmitz. Revealing the dependence structure between x(1) and x(n). *Journal of Statistical Planning and Inference*, 123(1):41–47, 2004.

[109] B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *Annals of Statistics*, 9:879–885.

[110] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

[111] R. L. Smith. Uniform rates of convergence in extreme-value theory. *Advances in Applied Probability*, 14(3):600–622, 1982.

[112] R. L. Smith. Estimating tails of probability distributions. *Annals of Statistics*, 15(3): 1174–1207, 1987.

[113] R. L. Smith. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4(4):389–393, 1989.

[114] J. M. Stoyanov. *Counterexamples in Probability*. Dover Books on Mathematics. Dover Publications Inc., 3 edition, 2013. ISBN 978-0-486-49998-7.

[115] R. M. Stulz. Rethinking risk management: presentation at the guest speaker conference, french finance association meeting, geneva", 1996.

[116] J. L. Teugels. Remarks on large claims. *Bulletin of the International Statistical Institute*, 49:1490–1500, 1981.

[117] J. L. Teugels. *Extreme Values in Insurance Mathematics*, pages 253–259. Springer Netherlands, 1984. ISBN 978-94-017-3069-3.

[118] F. Treves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press Inc., 1967. ISBN 0-12-699450-1-1.

[119] S. P. Vaswani. A pitfall in correlation theory. *Nature*, 160, 1947.

[120] D. V. Widder. *The Laplace Transform*. Princeton University Press, 2 edition, 1946.

[121] Bernt Øksendal. *Stochastic Differential Equations*. Springer Berlin, Heidelberg, 6 edition, 2003. ISBN 978-3-642-14394-6.

Index

C-measure, 160
ES_α, 9
VaR_α, 8
ARMA(p, q) model, 197
RV_ρ, 37

absolutely monotonic function, 107
acceptance set, 26
adapted process, 228
additive function, 135
adjustment coefficient, 196
advanced measurement approach, 193
almost sure convergence, 221
ARCH model, 197
Archimedean copula, 102
Arwedson approximation, 196

Basel formula, 189
basic indicator approach, 193
Bernoulli mixture model, 189
beta distribution, 211
beta function, 211
Beta mixture model, 189
Black–Scholes formula, 6
Black–Scholes model, 5
Blomqvist's β, 177
bootstrapping, 22
Borel–Cantelli criterion, 223
bounded variation function, 236
Breiman's Lemma, 66
Brownian motion, 229
Burr distribution, 64

Cauchy distribution, 61
Cauchy's functional equation, 135
Central Limit Theorem, 224
characteristic function, 77, 207
characteristic generator, 77
Chebyshev's inequality, 222
Chernoff's bound, 225
Cholesky factorisation, 79
Clayton copula, 100
    Generalised Clayton copula, 102
CLT, 224
co-copula, 124
coefficient of lower tail dependence, 148
coefficient of upper tail dependence, 148
comonotonicity, 92
comonotonicity copula, 97
completely monotonic function, 107
comprehensive copula family, 121
concordant pair, 139
conditional variance, 219
continuity in probability, 234
converge in distribution, 223
convergence in distribution, 223
convergence in probability, 221
Convergence to types theorem, 53
convex, 27
convex combination, 27
convex cone, 27
convex sum (of a copula), 131
convolution, 66
    density, 31
copula, 89
    explicit copulas, 100
    fundamental copulas, 97
    implicit copulas, 98
countermonotonicity, 92
countermonotonicity copula, 97
counting process, 230
    arrival time, 230
    interarrival time, 230
Cramér–Lundberg process, 196
CreditRisk$^+$, 191

Cuadras–Augé copula, 121
càdlàg, 206

DD, 185
Debye function, 169
default, 183
dependence function (copulas), 112
derivative, 5
diagonal section (of a copula), 124
digital option, 11
dilogarithm function, 171
discordant pair, 139
dispersion matrix, 79
distance to default, 185
distribution function, 205
distributional transform, 118
domain of attraction (copulas), 112
Doob–Dynkin lemma, 216
dual copula, 124

eigenvalue method, 152
elliptical distribution, 79
EMM, 6
empirical distribution function, 226, 239
equity, 185
equivalent martingale measure, 6
Erlang distribution, 219
Euler–Maruyama method, 231
European call option, 6
excess distribution function, 44
Expected Shortfall, 9
exponential distribution, 210
exterior power family, 117
extreme value copula, 112

factor model, 186
filtration, 228
Fisher information, 68
Fisher–Tippett Theorem, 54
Frank copula, 102
Fréchet bounds, 92
Fréchet copula, 121

gamma distribution, 210
gamma function, 210
GARCH model, 198
Gaussian copula, 98
Gaussian distribution, 209
GBM, 5
Generalised Extreme Value distribution, 53
Generalised Extreme Value Distribution (GEV), 59
generalised inverse, 202
Generalised Pareto Distribution, 38
Geometric Brownian Motion, 233
Geometric Brownian motion, 5
Glivenko–Cantelli Theorem, 17
GPD, 38
greeks, 7
Gumbel copula, 100

Hill estimator, 41
  Expected Shortfall, 63
  statistical properties, 43
  Value at Risk, 43, 63
Hill plot, 41
homogeneous copula, 131
homogeneous Poisson process, 230

importance sampling, 21
independence copula, 97
Integration by parts, 240
integration by parts, 40, 46
interior power family, 117
inverse gamma distribution, 82
Itô isometry, 232
Itô process, 231

Joe copula, 106
joint symmetry, 130
Jordan decomposition (of a function of bounded variation), 237

Kendall's τ, 138
KMV model, 185

Laplace transformy, 108
law of iterated expectation, 215
law of total variance, 220
law-invariant risk measure, 83
Lebesgue–Stieltjes integral, 237
Leibniz' formula (derivative), 107
lender, 183
level curve (Archimedean copula), 159
level set (copula), 159
linear portfolios, 83
linearized log returns, 5

linearized loss operator, 7
local mean rate of return, 5
location parameter, 79
log-copula, 170
loss operator, 7
loss random variable, 3
lower incomplete gamma function, 86
LT-Archimedean copula, 109

Mahalanobis distance, 81
Markov's inequality, 222
Marshall–Olkin copula, 114
martingale, 228
max-stable copula, 111
max-stable distribution, 53
Maximal Domain of Attraction, 53
    Fréchet, 56
    Gumbel, 58
    Weibull, 57
Maximal Domain of Attraction (MDA), 55
mean-excess function, 46, 240
    empirical, 47
mean-excess plot, 47
medial correlation coefficient, 177
median shortfall, 32
Merton's model, 184
midconvex, 130
midconvexity, 104
moment-generating function, 21, 207
Monte Carlo simulation, 19
Morgenstern copula, 121
multivariate Gaussian distribution, 82, 211
multivariate normal distribution, 82, 211
    regular, 212
    singular, 212

natural filtration, 228
negative binomial process, 195
normal distribution, 209
Normal Variance Mixture Model, 81

obligor, 183
one-factor model, 189
ordinal sums (copulas), 132
orthogonal matrix, 83
overdispersion, 195

Panjer recursion, 195
Pareto normal variance mixture, 86
Peaks over threshold method, 44
Plackett copula, 134
Poisson mixture model, 190
Poisson process, 194
    mixed Poisson process, 195
POT method, 44
    Expected Shortfall, 67
    Value at Risk, 67
probability-generating function, 191
probit normal mixture model, 187
pseudo-inverse (Archimedean copula), 129
pseudo-sample, 150

QQ plot, 35

radial symmetry, 95
Radon–Nikodym, 24
random variable, 205
rank (of an observation), 146
recovery rate, 183
regular varying function, 36
regularly varying
    for a distribution function, 37
    index for a distribution function, 37
relative error, 20
renewal process, 230
repetitive OR, 193
risk factor, 7
risk measure, 7
    coherent, 10
    monotonicity, 10
    positive homogeneity, 10
    subadditivity, 10
    translation invariance, 10
risk neutral measure, 6
Risk Neutral Valuation formula, 6
risk-factor change, 7

score function, 68
SDE, 5
secondary diagonal section (of a copula), 124
shifted distribution, 21
shortfall-to-quantile ratio, 30
Sklar's Theorem, 89
slowly varying function, 37

smoothed bootstrap, 22
Spearman's ρ, 142
spherical distribution, 77
stable distribution
    positive stable distribution, 212
standard exponential distribution, 210
standard normal distribution, 210
stochastic differential equation, 5, 231
stochastic process, 228
stochastic recursive sequence, 198
strict Archimedean copula, 129
strike price, 6
Strong Law of Large Numbers, 223
Student t distribution, 61
subexponential distribution, 63, 66, 196
superadditivity, 31
survival copula, 94
survival function, 205
symmetry (copulas), 168
symmetry (random variable), 95

tail-equivalence, 58
tower property, 215

uniform distribution, 210
upper incomplete gamma function, 170

Value at Risk, 8
    subadditivity for elliptical distributions, 84
Var-Cov method, 19
variation, 236
    negative variation, 236
    positive variation, 236
volatility, 5
Von Mises function, 57
    auxiliary function, 57

weak convergence, 223
Weak Law of Large Numbers, 225

zero curve (Archimedean copula), 159
zero set (copula), 159
zero-coupon bond, 4