Mathematical Modeling and Statistical Methods for Risk Management

Lecture Notes

2007

Contents

Preface

These lecture notes aim at giving an introduction to Quantitative Risk Management. We will introduce statistical techniques used for deriving the profit-and-loss distribution for a portfolio of financial instruments and to compute risk measures associated with this distribution. The focus lies on the mathematical/statistical modeling of market- and credit risk. Operational risks and the use of financial time series for risk modeling are not treated in these lecture notes. Financial institutions typically hold portfolios consisting on large number of financial instruments. A careful modeling of the dependence between these instruments is crucial for good risk management in these situations. A large part of these lecture notes is therefore devoted to the issue of dependence modeling.

The reader is assumed to have a mathematical/statistical knowledge corresponding to basic courses in linear algebra, analysis, statistics and an intermediate course in probability. The lecture notes are written with the aim of presenting the material in a fairly rigorous way without any use of measure theory.

The chapters 1-4 in these lecture notes are based on the book [12] which we strongly recommend. More material on the topics presented in remaining chapters can be found in [8] (chapters 5-7), [12] (chapters 8-12) and articles found in the list of references at the end of these lecture notes.

Henrik Hult and Filip Lindskog, 2007

1 Some background to financial risk management

We will now give a brief introduction to the topic of risk management and explain why this may be of importance for a bank or financial institution. We will start with a preliminary example illustrating in a simple way some of the issues encountered when dealing with risks and risk measurements.

1.1 A preliminary example

A player (investor/speculator) is entering a casino with an initial capital of $V_0 = 1$ million Swedish Kroner. All initial capital is used to place bets according to a predetermined gambling strategy. After the game the capital is V_1. We denote the profit(loss) by a random variable $X = V_1 - V_0$. The distribution of X is called the *profit-and-loss distribution* (P&L) and the distribution of $L = -X = V_0 - V_1$ is simply called the *loss distribution*. As the loss may be positive this is a risky position, i.e. there is a risk of losing some of the initial capital.

Suppose a game is constructed so that it gives 1.6 million Swedish Kroner with probability p and 0.6 million Swedish Kroner with probability $1-p$. Hence,

$$X = \begin{cases} 0.6 & \text{with probability } p, \\ -0.4 & \text{with probability } 1 - p. \end{cases} \tag{1.1}$$

The fair price for this game, corresponding to $\mathrm{E}(X) = 0$, is $p = 0.4$. However, even if $p > 0.4$ the player might choose not to participate in the game with the view that not participating is more attractive than playing a game with a small expected profit together with a risk of loosing 0.4 million Swedish Kroner. This attitude is called *risk-averseness*.

Clearly, the choice of whether to participate or not depends on the P&L distribution. However, in most cases (think of investing in instruments on the financial market) the P&L distribution is not known. Then you need to evaluate some aspects of the distribution to decide whether to play or not. For this purpose it is natural to use a *risk measure*. A risk measure ϱ is a mapping from the random variables to the real numbers; to every loss random variable L there is a real number $\varrho(L)$ representing the riskiness of L. To evaluate the loss distribution in terms of a single number is of course a huge simplification of the world but the hope is that it can give us sufficient indication whether to play the game or not.

Consider the game (1.1) described above and suppose that the mean $\mathrm{E}(L) = -0.1$ (i.e. a positive expected profit) and standard deviation $\mathrm{std}(L) = 0.5$ of the loss L is known. In this case the game had only two known possible outcomes so the information about the mean and standard deviation uniquely specifies the P&L distribution, yielding $p = 0.5$. However, the possible outcomes of a typical real-world game are typically not known and mean and standard deviation do

not specify the P&L distribution. A simple example is the following:

$$X = \begin{cases} 0.35 & \text{with probability } 0.8, \\ -0.9 & \text{with probability } 0.2. \end{cases} \qquad (1.2)$$

Here we also have $E(L) = -0.1$ and $std(L) = 0.5$. However, most risk-averse players would agree that the game (1.2) is riskier than the game (1.1) with $p = 0.5$. Using an appropriate quantile of the loss L as a risk measure would classify the game (1.2) as riskier than the game (1.1) with $p = 0.5$. However, evaluating a single risk measure such as a quantile will in general not provide a lot of information about the loss distribution, although it can provide some relevant information. A key to a sound risk management is to look for risk measures that give as much relevant information about the loss distribution as possible.

A risk manager at a financial institution with responsibility for a portfolio consisting of a few up to hundreds or thousands of financial assets and contracts faces a similar problem as the player above entering the casino. Management or investors have also imposed risk preferences that the risk manager is trying to meet. To evaluate the position the risk manager tries to assess the loss distribution to make sure that the current positions is in accordance with imposed risk preferences. If it is not, then the risk manager must rebalance the portfolio until a desirable loss distribution is obtained. We may view a financial investor as a player participating in the game at the financial market and the loss distribution must be evaluated in order to know which game the investor is participating in.

1.2 Why risk management?

The trading volumes on the financial markets have increased tremendously over the last decades. In 1970 the average daily trading volume at the New York Stock Exchange was 3.5 million shares. In 2002 it was 1.4 billion shares. In the last few years we have seen a significant increase in the derivatives markets. There are a huge number of actors on the financial markets taking risky positions

Contracts	1995	1998	2002
FOREX	13	18	18
Interest rate	26	50	102
Total	47	80	142

Table 1: Global market in OTC derivatives (nominal value) in trillion US dollars (1 trillion = 10^{12}).

and to evaluate their positions properly they need quantitative tools from risk management. Recent history also shows several examples where large losses on the financial market are mainly due to the absence of proper risk control.

Example 1.1 (Orange County) On December 6 1994, Orange County, a prosperous district in California, declared bankruptcy after suffering losses of

2

around $1.6 billion from a wrong-way bet on interest rates in one of its principal investment pools. (Source: www.erisk.com) ■

Example 1.2 (Barings bank) Barings bank had a long history of success and was much respected as the UK's oldest merchant bank. But in February 1995, this highly regarded bank, with $900 million in capital, was bankrupted by $1 billion of unauthorized trading losses. (Source: www.erisk.com) ■

Example 1.3 (LTCM) In 1994 a hedge-fund called Long-Term Capital Management (LTCM) was founded and assembled a star team of traders and academics. Investors and investment banks invested $1.3 billion in the fund and after two years returns was running close to 40%. Early 1998 the net asset value stands at $4 billion but at the end of the year the fund had lost substantial amounts of the investors equity capital and the fund was at the brink of default. The US Federal Reserve managed a $3.5 billion rescue package to avoid the threat of a systematic crisis in th world financial system. (Source: www.erisk.com) ■

1.3 Regulators and supervisors

To be able to cover most financial losses most banks and financial institutions put aside a buffer capital, also called *regulatory capital*. The amount of buffer capital needed is of course related to the amount of risk the bank is taking, i.e. to the overall P&L distribution. The amount is regulated by law and the national supervisory authority makes sure that the banks and financial institutions follow the rules.

There is also a strive to develop international standards and methods for computing regulatory capital. This is the main task of the so-called Basel Committee. The Basel Committee, established in 1974, does not possess any formal supernational supervising authority and its conclusions does not have legal force. It formulates supervisory standards, guidelines and recommends statements of best practice. In this way the Basel Committee has large impact on the national supervisory authorities.

- In 1988 the first Basel Accord on Banking Supervision [2] initiated an important step toward an international minimal capital standard. Emphasis was on credit risk.

- In 1996 an amendment to Basel I prescribes a so–called standardized model for market risk with an option for larger banks to use internal Value-at-Risk (VaR) models.

- In 2001 a new consultative process for the new Basel Accord (Basel II) is initiated. The main theme concerns advanced internal approaches to credit risk and also new capital requirements for operational risk. The new Accord aims at an implementation date of 2006-2007. Details of Basel II is still hotly debated.

1.4 Why the government cares about the buffer capital

The following motivation is given in [6].

"Banks collect deposits and play a key role in the payment system. National governments have a very direct interest in ensuring that banks remain capable of meeting their obligations; in effect they act as a guarantor, sometimes also as a lender of last resort. They therefore wish to limit the cost of the safety net in case of a bank failure. By acting as a buffer against unanticipated losses, regulatory capital helps to privatize a burden that would otherwise be borne by national governments."

1.5 Types of risk

Here is a general definition of risk for an organization: *any event or action that may adversely affect an organization to achieve its obligations and execute its strategies*. In financial risk management we try to be a bit more specific and divide most risks into three categories.

- Market risk – risks due to changing markets, market prices, interest rate fluctuations, foreign exchange rate changes, commodity price changes etc.

- Credit risk – the risk carried by the lender that a debtor will not be able to repay his/her debt or that a counterparty in a financial agreement can not fulfill his/her commitments.

- Operational risk – the risk of losses resulting from inadequate of failed internal processes, people and systems of from external events. This includes people risks such as incompetence and fraud, process risk such as transaction and operational control risk and technology risk such as system failure, programming errors etc.

There are also other types of risks such as *liquidity risk* which is risk that concerns the need for well functioning financial markets where one can buy or sell contracts at fair prices. Other types of risks are for instance *legal risk* and *reputational risk*.

1.6 Financial derivatives

Financial derivatives are financial products or contracts derived from some fundamental underlying; a stock price, stock index, interest rate, commodity price to name a few. The key example is the European Call option written on a particular stock. It gives the holder the right but not the obligation at a given date T to buy the stock S for the price K. For this the buyer pays a premium at time zero. The value of the European Call at time T is then

$$C(T) = \max(S_T - K, 0).$$

Financial derivatives are traded not only for the purpose of speculation but is actively used as a risk management tool as they are tailor made for exchanging

risks between actors on the financial market. Although they are of great importance in risk management we will not discuss financial derivatives much in this course but put emphasis on the statistical models and methods for modeling financial data.

2 Loss operators and financial portfolios

Here we follow [12] to introduce the loss operator and give some examples of financial portfolios that fit into this framework.

2.1 Portfolios and the loss operator

Consider a given portfolio such as for instance a collection of stocks, bonds or risky loans, or the overall position of a financial institution.

The *value* of the portfolio at time t is denoted $V(t)$. Given a time horizon Δt the profit over the interval $[t, t + \Delta t]$ is given by $V(t + \Delta t) - V(t)$ and the distribution of $V(t + \Delta t) - V(t)$ is called the *profit-and-loss distribution* (P&L). The *loss* over the interval is then

$$L_{[t,t+\Delta t]} = -(V(t + \Delta t) - V(t))$$

and the distribution of $L_{[t,t+\Delta t]}$ is called the *loss distribution*. Typical values of Δt is one day (or $1/250$ years as we have approximately 250 trading days in one year), ten days, one month or one year.

We may introduce a discrete parameter $n = 0, 1, 2, \ldots$ and use $t_n = n\Delta t$ as the actual time. We will sometimes use the notation

$$L_{n+1} = L_{[t_n, t_{n+1}]} = L_{[n\Delta t, (n+1)\Delta t]} = -(V((n + 1)\Delta t) - V(n\Delta t)).$$

Often we also write V_n for $V(n\Delta t)$.

Example 2.1 Consider a portfolio of d stocks with α_i units of stock number i, $i = 1, \ldots, d$. The stock prices at time n are given by $S_{n,i}$, $i = 1, \ldots, d$, and the value of the portfolio is

$$V_n = \sum_{i=1}^{d} \alpha_i S_{n,i}.$$

In financial statistics one often tries to find a statistical model for the evolution of the stock prices, e.g. a model for $S_{n+1,i} - S_{n,i}$, to be able to compute the loss distribution L_{n+1}. However, it is often the case that the so-called log returns $X_{n+1,i} = \ln S_{n+1,i} - \ln S_{n,i}$ are easier to model than the differences $S_{n+1,i} - S_{n,i}$. With $Z_{n,i} = \ln S_{n,i}$ we have $S_{n,i} = \exp\{Z_{n,i}\}$ so the portfolio loss $L_{n+1} = -(V_{n+1} - V_n)$ may be written as

$$L_{n+1} = -\sum_{i=1}^{d} \alpha_i(\exp\{Z_{n+1,i}\} - \exp\{Z_{n,i}\})$$

$$= -\sum_{i=1}^{d} \alpha_i \exp\{Z_{n,i}\}(\exp\{X_{n+1,i}\} - 1)$$

$$= -\sum_{i=1}^{d} \alpha_i S_{n,i}(\exp\{X_{n+1,i}\} - 1).$$

6

The relation between the modeled variables $X_{n+1,i}$ and the loss L_{n+1} is nonlinear and it is sometimes useful to linearize this relation. In this case this is done by replacing e^x by $1+x$; recall the Taylor expansion $e^x = 1 + x + O(x^2)$. Then the linearized loss is given by

$$L_{n+1}^\Delta = -\sum_{i=1}^d \alpha_i S_{n,i} X_{n+1,i}.$$

■

2.2 The general case

A general portfolio with value V_n is often modeled using a d-dimensional random vector $\mathbf{Z}_n = (Z_{n,1}, \ldots, Z_{n,d})$ of *risk-factors*. The value of the portfolio is then expressed as

$$V_n = f(t_n, \mathbf{Z}_n)$$

for some known function f and t_n is the actual calendar time. As in the example above it is often convenient to model the *risk-factor changes* $\mathbf{X}_{n+1} = \mathbf{Z}_{n+1} - \mathbf{Z}_n$. Then the loss is given by

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

The loss may be viewed as the result of applying an operator $l_{[n]}(\cdot)$ to the risk-factor changes $\mathbf{X}_{n+1}$ so that

$$L_{n+1} = l_{[n]}(\mathbf{X}_{n+1})$$

where

$$l_{[n]}(\mathbf{x}) = -\big(f(t_{n+1}, \mathbf{Z}_n + \mathbf{x}) - f(t_n, \mathbf{Z}_n)\big).$$

The operator $l_{[n]}(\cdot)$ is called the *loss-operator*. If we want to linearize the relation between L_{n+1} and $\mathbf{X}_{n+1}$ then we have to differentiate f to get the linearized loss

$$L_{n+1}^\Delta = -\Big(f_t(t_n, \mathbf{Z}_n)\Delta t + \sum_{i=1}^d f_{z_i}(t_n, \mathbf{Z}_n) X_{n+1,i}\Big).$$

Here $f_t(t, \mathbf{z}) = \partial f(t, \mathbf{z})/\partial t$ and $f_{z_i}(t, \mathbf{z}) = \partial f(t, \mathbf{z})/\partial z_i$. The corresponding operator given by

$$l_{[n]}^\Delta(\mathbf{x}) = -\Big(f_t(t_n, \mathbf{Z}_n)\Delta t + \sum_{i=1}^d f_{z_i}(t_n, \mathbf{Z}_n) x_i\Big)$$

is called the *linearized loss-operator*.

Example 2.2 (Portfolio of stocks continued) In the previous example of a portfolio of stocks the risk-factors are the log-prices of the stocks $Z_{n,i} = \ln S_{n,i}$ and the risk-factor-changes are the log returns $X_{n+1,i} = \ln S_{n+1,i} - \ln S_{n,i}$. The loss is $L_{n+1} = l_{[n]}(\mathbf{X}_{n+1})$ and the linearized loss is $L_{n+1}^{\Delta} = l_{[n]}^{\Delta}(\mathbf{X}_{n+1})$, where

$$l_{[n]}(\mathbf{x}) = -\sum_{i=1}^{d} \alpha_i S_{n,i}(\exp\{x_i\} - 1) \quad \text{and} \quad l_{[n]}^{\Delta}(\mathbf{x}) = -\sum_{i=1}^{d} \alpha_i S_{n,i} x_i$$

are the loss operator and linearized loss operator, respectively. $\blacksquare$

The following examples may convince you that there are many relevant examples from finance that fits into this general framework of risk-factors and loss-operators.

Example 2.3 (A bond portfolio) A *zero-coupon bond* with *maturity T* is a contract which gives the holder of the contract \$1 at time T. The price of the contract at time $t < T$ is denoted $B(t, T)$ and by definition $B(T, T) = 1$. To a zero-coupon bond we associate the *continuously compounded yield*

$$y(t, T) = -\frac{1}{T - t} \ln B(t, T),$$

i.e.

$$B(t, T) = \exp\{-(T - t)y(t, T)\}.$$

To understand the notion of the yield, consider a bank account where we get a constant interest rate r. The bank account evolves according to the differential equation

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = rS_t, \quad S_0 = 1$$

which has the solution $S_t = \exp\{rt\}$. This means that in every infinitesimal time interval $\mathrm{d}t$ we get the interest rate r. Every dollar put into the bank account at time t is then worth $\exp\{r(T - t)\}$ dollars at time T. Hence, if we have $\exp\{-r(T-t)\}$ dollars on the account at time t then we have exactly \$1 at time T. The yield $y(t, T)$ can be identified with r and is interpreted as the constant interest rate contracted at t that we get over the period $[t, T]$. However, the yield may be different for different maturity times T. The function $T \mapsto y(t, T)$ for fixed t is called the *yield-curve* at t. Consider now a portfolio consisting of d different (default free) zero-coupon bonds with maturity times T_i and prices $B(t, T_i)$. We suppose that we have α_i units of the bond with maturity T_i so that the value of the portfolio is

$$V_n = \sum_{i=1}^{d} \alpha_i B(t_n, T_i).$$

It is natural to use the yields $Z_{n,i} = y(t_n, T_i)$ as risk-factors. Then

$$V_n = \sum_{i=1}^{d} \alpha_i \exp\{-(T_i - t_n)Z_{n,i}\} = f(t_n, \mathbf{Z}_n)$$

and the loss is given by (with $X_{n+1,i} = Z_{n+1,i} - Z_{n,i}$ and $\Delta t = t_{n+1} - t_n$)

$$L_{n+1} = -\sum_{i=1}^{d} \alpha_i \Big(\exp\{-(T_i - t_{n+1})(Z_{n,i} + X_{n+1,i})\} - \exp\{-(T_i - t_n)Z_{n,i}\} \Big)$$

$$= -\sum_{i=1}^{d} \alpha_i B(t_n, T_i) \Big(\exp\{Z_{n,i}\Delta t - (T_i - t_{n+1})X_{n+1,i}\} - 1 \Big).$$

The corresponding loss-operator is then

$$l_{[n]}(\mathbf{x}) = -\sum_{i=1}^{d} \alpha_i B(t_n, T_i) \Big(\exp\{Z_{n,i}\Delta t - (T_i - t_{n+1})x_i\} - 1 \Big)$$

and the linearized loss is given by

$$L_{n+1}^{\Delta} = -\sum_{i=1}^{d} \alpha_i B(t_n, T_i) \Big(Z_{n,i}\Delta t - (T_i - t_{n+1})X_{n+1,i} \Big).$$

∎

Example 2.4 (European call and put) In this example we will consider a portfolio consisting of one European call or put option on a nondividend paying stock with price S_t for one share at time t, with maturity date $T > t$ and strike price K. A European call option is a contract which pays $\max(S_T - K, 0)$ to the holder of the contract at time T. A European put option pays the holder $\max(K - S_T, 0)$ at time T. The price at time $t < T$ for the contract is evaluated using a function C (for call) or P (for put) depending on some parameters. In the Black-Scholes model $C = C(t, T, S_t, K, r, \sigma)$ and $P = P(t, T, S_t, K, r, \sigma)$, where the time to maturity $T - t$ is measured in years, r is the continuously compounded interest rate per year and σ is the volatility (corresponding to the standard deviation of the one-year log return for the stock price). We have

$$C(t, T, S_t, K, r, \sigma) = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2),$$

$$P(t, T, S_t, K, r, \sigma) = K e^{-r(T-t)} \Phi(-d_2) - S_t \Phi(-d_1),$$

$$d_1 = \frac{\ln(S_t/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2 = d_1 - \sigma\sqrt{T - t}.$$

In this case with time measured in years we may set $t = t_n = n\Delta t$ and $T = t_{n+k} = (n+k)\Delta t$, where $\Delta t = 1/250$ years (approximately 250 trading days per year). In this case we may put

$$\mathbf{Z}_n = (\ln S_n, r_n, \sigma_n)$$
$$\mathbf{X}_{n+1} = (\ln S_{n+1} - \ln S_n, r_{n+1} - r_n, \sigma_{n+1} - \sigma_n).$$

The value of the portfolio is $V_n = C(t_n, T, S_n, K, r_n, \sigma_n)$ for the call option and $V_n = P(t_n, T, S_n, K, r_n, \sigma_n)$ for the put option. The linearized loss is given by

$$L_{n+1}^{\Delta} = -(C_t \Delta t + C_S X_{n+1,1} + C_r X_{n+1,2} + C_\sigma X_{n+1,3}),$$
$$L_{n+1}^{\Delta} = -(P_t \Delta t + P_S X_{n+1,1} + P_r X_{n+1,2} + P_\sigma X_{n+1,3})$$

for the call and put option, respectively. The partial derivatives are usually called the "Greeks". C_t is called *theta*; C_S is called *delta*; C_r is called *rho*; C_σ is called *vega* (although this is not a Greek letter). ∎

3   Risk measurement

What is the purpose of risk measurement?

- Determination of risk capital – determine the amount of capital a financial institution needs to cover unexpected losses.

- Management tool – Risk measures are used by management to limit the amount of risk a unit within the firm may take.

Next we present some common measures of risk.

3.1   Elementary measures of risk

Notional amount

Risk of a portfolio is measured as the sum of notional values of individual securities, weighted by a factor for each asset class. The notional amount approach is used for instance in the standard approach of the Basel Committee where risk weights 0%, 10%, 20%, 50% and 100% are used (see [2]). Then the regulatory capital should be such that

$$\frac{\text{regulatory capital}}{\text{risk-weighted sum}} \geq 8\%.$$

Example 3.1 Suppose we have a portfolio with three claims each of notional amount \$1 million. The first claim is on an OECD central bank, the second on a multilateral developed bank and the third on the private sector. According to the risk-weights used by the Basel document the first claim is weighted by 0%, the second by 20% and the third by 100%. Thus the risk-weighted sum is

$$0 \times 10^6 + 0.20 \times 10^6 + 1 \times 10^6 = 1\,200\,000,$$

and the regulatory capital should be at least 8% of this amount, i.e. \$96 000. ∎

Advantage: easy to use.
Disadvantage: Does not differentiate between long and short positions. There are no diversification effects; a portfolio with loans to m independent obligors is considered as risky as the same amount lent to a single obligor.

Factor sensitivity measures

Factor sensitivity measures gives the change in portfolio value for a predetermined change in one of the underlying risk factors. If the value of the portfolio is given by

$$V_n = f(t_n, \mathbf{Z}_n)$$

then factor sensitivity measures are given by the partial derivatives

$$f_{z_i}(t_n, \mathbf{Z}_n) = \frac{\partial f}{\partial z_i}(t_n, \mathbf{Z}_n).$$

11

The "Greeks" of a derivative portfolio may be considered as factor sensitivity measures.

Advantage: Factor sensitivity measures provide information about the robustness of the portfolio value with respect to certain events (risk-factor changes).

Disadvantage: It is difficult to aggregate the sensitivity with respect to changes in different risk-factors or aggregate across markets to get an understanding of the overall riskiness of a position.

Scenario based risk measures

In this approach we consider a number of possible scenarios, i.e. a number of possible risk-factor changes. A scenario may be for instance a 10% rise in a relevant exchange rate and a simultaneous 20% drop in a relevant stock index. The risk is then measured as the maximum loss over all possible (predetermined) scenarios. To assess the maximum loss extreme scenarios may be down-weighted in a suitable way.

Formally, this approach may be formulated as follows. Fix a number N of possible risk-factor changes, $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. Each scenario is given a weight, w_i and we write $\mathbf{w} = (w_1, \ldots, w_N)$. We consider a portfolio with loss-operator $l_{[n]}(\cdot)$. The risk of the portfolio is then measured as

$$\psi_{[\mathcal{X}, \mathbf{w}]} = \max\{w_1 l_{[n]}(\mathbf{x}_1), \ldots, w_N l_{[n]}(\mathbf{x}_N)\}$$

These risk measures are frequently used in practice (example: Chicago Mercantile Exchange).

Example 3.2 (The SPAN rules) As an example of a scenario based risk measure we consider the SPAN rules used at the Chicago Mercantile Exchange [1]. We describe how the initial margin is calculated for a simple portfolio consisting of units of a futures contract and of several puts and calls with a common expiration date on this futures contract. The SPAN margin for such a portfolio is compute as follows: First fourteen scenarios are considered. Each scenario is specified by an up or down move of volatility combined with no move, or an up move, or a down move of the futures prices by 1/3, 2/3 or 3/3 of a specific "range". Next, two additional scenarios relate to "extreme" up or down moves of the futures prices. The measure of risk is the maximum loss incurred, using the full loss of the first fourteen scenarios and only 35% of the loss for the last two "extreme" scenarios. A specified model, typically the Black model, is used to generate the corresponding prices for the options under each scenario.

The account of the investor holding a portfolio is required to have sufficient current net worth to support the maximum expected loss. If it does not, then extra cash is required as margin call, an amount equal to the "measure of risk" involved. ∎

Loss distribution approach

This is the approach a statistician would use to compute the risk of a portfolio. Here we try model the loss L_{n+1} using a probability distribution F_L. The

parameters of the loss distribution are estimated using historical data. One can either try to model the loss distribution directly or to model the risk-factors or risk-factor changes as a d-dimensional random vector or a multivariate time series. Risk measures are based on the distribution function F_L. The next section studies this approach to risk measurement.

3.2   Risk measures based on the loss distribution

Standard deviation

The standard deviation of the loss distribution as a measure of risk is frequently used, in particular in portfolio theory. There are however some disadvantages using the standard deviation as a measure of risk. For instance the standard deviation is only defined for distributions with $E(L^2) < \infty$, so it is undefined for random variables with very heavy tails. More important, profits and losses have equal impact on the standard deviation as risk measure and it does not discriminate between distributions with apparently different probability of potentially large losses. In fact, the standard deviation does not provide any information on how large potential losses may be.

Example 3.3 Consider for instance the two loss distributions $L_1 \sim \mathrm{N}(0, 2)$ and $L_2 \sim t_4$ (standard Student's t-distribution with 4 degrees of freedom). Both L_1 and L_2 have standard deviation equal to $\sqrt{2}$. The probability density is illustrated in Figure 1 for the two distributions. Clearly the probability of large losses is much higher for the t_4 than for the normal distribution.
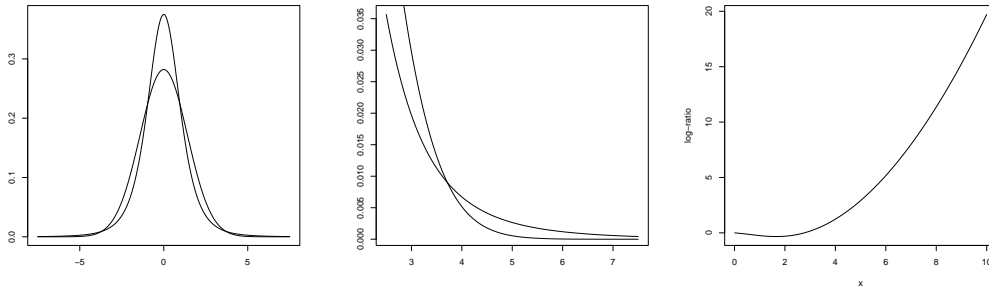
Figure 1: Left/Middle: The density function for a $\mathrm{N}(0, 2)$ and a t_4 distribution. The t_4 is highly peaked around zero and has much heavier tails. Right: The "log-ratio" $\ln[\mathrm{P}(L_2 > x) / \mathrm{P}(L_1 > x)]$ is plotted.

■

Value-at-Risk

We now introduce the widely used risk measure known as Value-at-Risk.

Definition 3.1 *Given a loss L and a confidence level $\alpha \in (0,1)$, $\mathrm{VaR}_\alpha(L)$ is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1 - \alpha$, i.e.*

$$
\begin{aligned}
\mathrm{VaR}_\alpha(L) &= \inf\{l \in \mathbb{R} : \mathrm{P}(L > l) \leq 1 - \alpha\} \\
&= \inf\{l \in \mathbb{R} : 1 - F_L(l) \leq 1 - \alpha\} \\
&= \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.
\end{aligned}
$$

∎
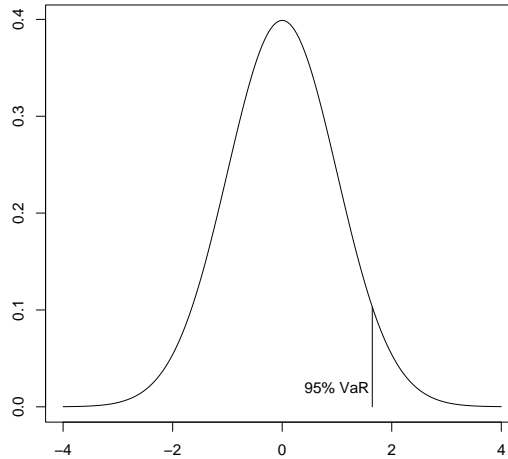
Figure 2: Illustration of $\mathrm{VaR}_{0.95}$.

We might think of L as the (potential) loss resulting from holding a portfolio over some fixed time horizon. In market risk the time horizon is typically one day or ten days. In credit risk the portfolio may consist of loans and the time horizon is often one year. Common confidence levels are 95%, 99% and 99.9% depending on the application. The BIS (Bank of International Settlements) proposes, for market risk, to compute ten-day VaR with confidence level $\alpha = 99\%$. The time horizon of ten days reflects the fact that markets are not perfectly liquid.

Definition 3.2 *Given a nondecreasing function $F : \mathbb{R} \to \mathbb{R}$ the* generalized inverse *of F is given by*

$$
F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}
$$

with the convention $\inf \emptyset = \infty$.

∎

If F is strictly increasing then $F^{\leftarrow} = F^{-1}$, i.e. the usual inverse. Using the generalized inverse we define the *α-quantile* of F by

$$
q_\alpha(F) = F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \quad \alpha \in (0,1).
$$

We note also that $\mathrm{VaR}_\alpha(F) = q_\alpha(F)$, where F is the loss distribution. Notice that for $a > 0$ and $b \in \mathbb{R}$,

$$
\begin{aligned}
\mathrm{VaR}_\alpha(aL + b) &= \inf\{l \in \mathbb{R} : \mathrm{P}(aL + b \le l) \ge \alpha\} \\
&= \inf\{l \in \mathbb{R} : \mathrm{P}(L \le (l - b)/a) \ge \alpha\} \\
&\quad \{\text{let } l' = (l - b)/a\} \\
&= \inf\{al' + b \in \mathbb{R} : \mathrm{P}(L \le l') \ge \alpha\} \\
&= a \inf\{l' \in \mathbb{R} : \mathrm{P}(L \le l') \ge \alpha\} + b \\
&= a \, \mathrm{VaR}_\alpha(L) + b.
\end{aligned}
$$

Hence, the risk measured in VaR for a shares of a portfolio is a times the risk of one share of this portfolio. Moreover, adding $(b < 0)$ or withdrawing $(b > 0)$ an amount $|b|$ of money from the portfolio changes this risk by the same amount.

Example 3.4 Suppose the loss distribution is normal, $L \sim \mathrm{N}(\mu, \sigma^2)$. This means that $L \overset{\mathrm{d}}{=} \mu + \sigma L'$, where $L' \sim \mathrm{N}(0, 1)$. Since $\mathrm{VaR}_\alpha(L') = \Phi^{-1}(\alpha)$, where Φ is the distribution function of a standard normal random variable, we may compute the Value-at-Risk as $\mathrm{VaR}_\alpha(L) = \mu + \sigma \Phi^{-1}(\alpha)$. ∎

Example 3.5 Suppose that the distribution function F is given by

$$
F(x) = \begin{cases} 0 & x < 0, \\ 1/2 & x \in [0, 1), \\ 1 & x \ge 1. \end{cases}
$$

Then $F^\leftarrow(u) = 0$ on $(0, 1/2]$ and $F^\leftarrow(u) = 1$ on $(1/2, 1)$. (F is the distribution function of a random variable X with $\mathrm{P}(X = 0) = \mathrm{P}(X = 1) = 1/2$.) ∎

Example 3.6 You hold a portfolio consisting of a long position of $\alpha = 5$ shares of stock A. The stock price today is $S_0 = 100$. The daily log returns

$$
X_1 = \ln(S_1/S_0), X_2 = \ln(S_2/S_1), \ldots
$$

of stock A are assumed to be normally distributed with zero mean and standard deviation $\sigma = 0.1$. Let L_1 be the portfolio loss from today until tomorrow. For a standard normal random variable $Z \sim \mathrm{N}(0, 1)$ we have $F_Z^{-1}(0.99) \approx 2.3$.

(a) Compute $\mathrm{VaR}_{0.99}(L_1)$.

We have shown that $L_1 = -\alpha S_0(e^{X_1} - 1) = -500(e^{X_1} - 1)$. We have $\mathrm{VaR}_u(L_1) = F_{L_1}^{-1}(u)$ and to compute $F_{L_1}^{-1}(u)$ we use that $F_{L_1}(F_{L_1}^{-1}(u)) = u$.

$$
\begin{aligned}
F_{L_1}(l) &= \mathrm{P}(-500(e^{X_1} - 1) \le l) \\
&= \mathrm{P}(e^{X_1} \ge 1 - l/500) \\
&= \mathrm{P}(X_1 \ge \ln(1 - l/500)) \\
&= 1 - F_{X_1}(\ln(1 - l/500)).
\end{aligned}
$$
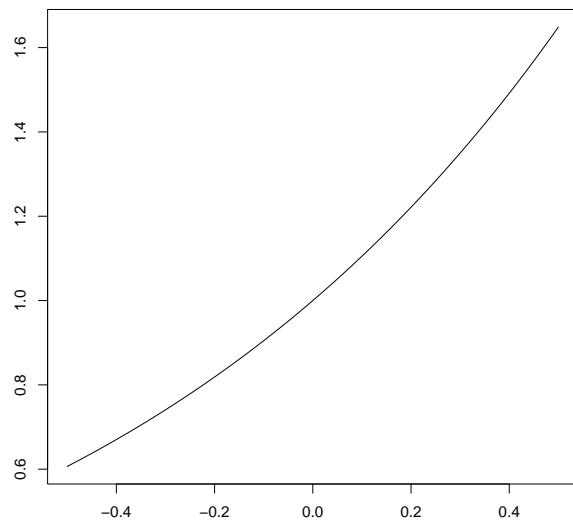
Figure 3: Plot of the function e^x for $x \in [-0.5, 0.5]$.
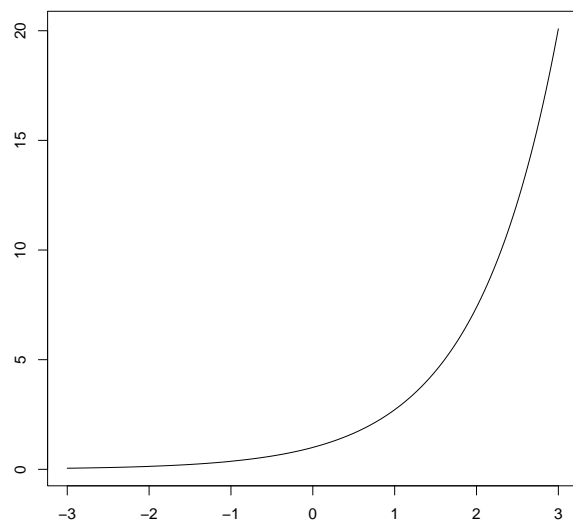
Figure 4: Plot of the function e^x for $x \in [-3, 3]$.

Hence,

$$1 - F_{X_1}(\ln(1 - F_{L_1}^{-1}(u)/500)) = u$$
$$\Leftrightarrow \quad \ln(1 - F_{L_1}^{-1}(u)/500) = F_{X_1}^{-1}(1 - u)$$
$$\Leftrightarrow \quad 1 - F_{L_1}^{-1}(u)/500 = e^{F_{X_1}^{-1}(1-u)}$$
$$\Leftrightarrow \quad F_{L_1}^{-1}(u) = 500(1 - e^{F_{X_1}^{-1}(1-u)}).$$

Since X_1 is symmetric about 0 we have $F_{X_1}^{-1}(1 - u) = -F_{X_1}^{-1}(u)$. Hence, $F_{L_1}^{-1}(u) = 500(1 - e^{-F_{X_1}^{-1}(u)})$. Using that $F_{X_1}^{-1}(0.99) = 0.1 \cdot F_Z^{-1}(0.99) \approx 0.23$ and with the help of Figure 3,

$$F_{L_1}^{-1}(0.99) = 500(1 - e^{-F_{X_1}^{-1}(0.99)})$$
$$\approx 500(1 - e^{-0.23}) \approx 500(1 - 0.8) = 100.$$

Hence, $\text{VaR}_{0.99}(L_1) \approx 100$.

You decide to keep your portfolio for 100 (trading) days before deciding what to do with the portfolio.

(b) Compute $\text{VaR}_{0.99}(L_{100})$ and $\text{VaR}_{0.99}(L_{100}^{\Delta})$, where L_{100} denotes the loss from today until 100 days from today and L_{100}^{Δ} denotes the corresponding linearized 100-day loss.

We have

$$L_{100} = -\alpha S_0(e^{X_{100}} - 1),$$

where X_{100} is the 100-day log return. Notice that

$$X_{100} = \ln S_{100}/S_0 = \ln S_{100} - \ln S_0$$
$$= \ln S_1/S_0 + \cdots + \ln S_{100}/S_{99},$$

i.e. X_{100} is a sum of 100 independent normally distributed random variables with mean zero and standard deviation 0.1. Hence, $X_{100} \overset{d}{=} Z$, where Z is normally distributed with zero mean and standard deviation one. We have

$$\text{VaR}_{0.99}(L_{100}) = 500(1 - e^{-F_Z^{-1}(0.99)})$$
$$\approx 500(1 - e^{-2.3}).$$

Using Figure 4 we find that $e^{-2.3} = 1/e^{2.3} \approx 0.1$. Hence, $\text{VaR}_{0.99}(L_{100}) \approx 500(1 - e^{-2.3}) \approx 450$.

We have $L_{100}^{\Delta} = -500Z$. Hence, $\text{VaR}_{0.99}(L_{100}^{\Delta}) = 500F_Z^{-1}(0.99) \approx 500 \cdot 2.3 = 1150$. One sees that using the linearized loss here gives a very bad risk estimate. ∎

17

Example 3.7 Consider a portfolio consisting of one European put option on a stock market index. If $t = 0$ is the time today, then the Black-Scholes price of the put option is

$$P(T, S_0, K, r, \sigma) = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1),$$

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}$$

with the notation of Example 2.4. What is the Value-at-Risk at confidence level 0.99 for the portfolio loss L from today until time Δt? Let X be the log return for the stock price over this period. If T is much larger than Δt, then we may assume that r and σ remain constant until time Δt and so

$$L = -P(T - \Delta t, S_0 e^X, K, r, \sigma) + P(T, S_0, K, r, \sigma) = l_{[0]}(X).$$

For a European put option we have $\partial P/\partial S < 0$ (negative delta), so the loss operator $l_{[0]}$ above is a continuous and strictly increasing function. This gives

$$F_L(y) = F_X(l_{[0]}^{-1}(y)) \text{ and } F_L^{-1}(p) = l_{[0]}(F_X^{-1}(p)).$$

Hence,

$$\text{VaR}_{0.99}(L) = -P(T - \Delta t, S_0 \exp\{F_X^{-1}(0.99)\}, K, r, \sigma) + P(T, S_0, K, r, \sigma).$$

∎

Expected shortfall

Although Value-at-Risk has become a very popular risk measure among practitioners it has several limitations. For instance, it does not give any information about how bad losses may be when things go wrong. In other words, what is the size of an "average loss" given that the loss exceeds the 99%-Value-at-Risk?

Definition 3.3 *For a loss L with continuous loss distribution function F_L the expected shortfall* at confidence level $\alpha \in (0,1)$ *is given by*

$$\text{ES}_\alpha(L) = \text{E}(L \mid L \geq \text{VaR}_\alpha(L)).$$

∎

We can rewrite this as follows:

$$\begin{aligned}
\text{ES}_\alpha(L) &= \text{E}(L \mid L \geq \text{VaR}_\alpha(L)) \\
&= \frac{\text{E}(L\mathbb{I}_{[q_\alpha(L),\infty)}(L))}{\text{P}(L \geq q_\alpha(L))} \\
&= \frac{1}{1-\alpha}\text{E}(L\mathbb{I}_{[q_\alpha(L),\infty)}(L)) \\
&= \frac{1}{1-\alpha}\int_{q_\alpha(L)}^{\infty} l\,dF_L(l),
\end{aligned}$$

where $\mathbb{I}_A$ is the indicator function: $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise.

For a loss L with continuous distribution function F_L expected shortfall is given by

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_p(L)\mathrm{d}p.$$

To see this we use the facts that $L \overset{\mathrm{d}}{=} F_L^\leftarrow(U)$ if U is uniformly distributed on $(0,1)$, and $F_L^\leftarrow$ is strictly increasing if F_L is continuous.

$$\begin{aligned}
\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha} \mathrm{E}(L\mathbb{I}_{[q_\alpha(L),\infty)}(L)) \\
&= \frac{1}{1-\alpha} \mathrm{E}(F_L^\leftarrow(U)\mathbb{I}_{[F_L^\leftarrow(\alpha),\infty)}(F_L^\leftarrow(U))) \\
&= \frac{1}{1-\alpha} \mathrm{E}(F_L^\leftarrow(U)\mathbb{I}_{[\alpha,1)}(U)) \\
&= \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L)\mathrm{d}u.
\end{aligned}$$

For a discrete distribution there are different possibilities to define expected shortfall. A useful definition called *generalized expected shortfall*, which is a so-called coherent risk measure, is given by

$$\mathrm{GES}_\alpha(L) = \frac{1}{1-\alpha} \Big(\mathrm{E}(L\mathbb{I}_{[q_\alpha(L),\infty)}(L)) + q_\alpha(L)\big(1 - \alpha - \mathrm{P}(L \geq q_\alpha(L))\big)\Big).$$

If the distribution of L is continuous, then the second term vanishes and $\mathrm{GES}_\alpha = \mathrm{ES}_\alpha$.

Exercise 3.1 (a) Let $L \sim \mathrm{Exp}(\lambda)$ and calculate $\mathrm{ES}_\alpha(L)$.
(b) Let L have distribution function $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma \in (0,1)$, and calculate $\mathrm{ES}_\alpha(L)$.

Answer: (a) $\lambda^{-1}(1 - \ln(1-\alpha))$. (b) $\gamma^{-1}[(1-\alpha)^{-\gamma}(1-\gamma)^{-1} - 1]$.

Example 3.8 Suppose that $L \sim \mathrm{N}(0,1)$. Let ϕ and Φ be the density and distribution function of L. Then

$$\begin{aligned}
\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha} \int_{\Phi^{-1}(\alpha)}^\infty l\mathrm{d}\Phi(l) \\
&= \frac{1}{1-\alpha} \int_{\Phi^{-1}(\alpha)}^\infty l\phi(l)\mathrm{d}l \\
&= \frac{1}{1-\alpha} \int_{\Phi^{-1}(\alpha)}^\infty l\frac{1}{\sqrt{2\pi}}e^{-l^2/2}\mathrm{d}l \\
&= \frac{1}{1-\alpha}\Big[-\frac{1}{\sqrt{2\pi}}e^{-l^2/2}\Big]_{\Phi^{-1}(\alpha)}^\infty \\
&= \frac{1}{1-\alpha}\Big[-\phi(l)\Big]_{\Phi^{-1}(\alpha)}^\infty \\
&= \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.
\end{aligned}$$

Suppose that $L' \sim N(\mu, \sigma^2)$. Then

$$
\begin{aligned}
\mathrm{ES}_\alpha(L') &= \mathrm{E}(L' \mid L' \geq \mathrm{VaR}_\alpha(L')) \\
&= \mathrm{E}(\mu + \sigma L \mid \mu + \sigma L \geq \mathrm{VaR}_\alpha(\mu + \sigma L)) \\
&= \mathrm{E}(\mu + \sigma L \mid L \geq \mathrm{VaR}_\alpha(L)) \\
&= \mu + \sigma \, \mathrm{ES}_\alpha(L) \\
&= \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}.
\end{aligned}
$$

∎

Exercise 3.2 Let L have a standard Student's t-distribution with $\nu > 1$ degrees of freedom. Then L has density function

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

Show that

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

where t_ν denotes the distribution function of L.

4 Methods for computing VaR and ES

We will now introduce some standard methods for computing Value-at-Risk and expected shortfall for a portfolio of risky assets. The setup is as follows. We consider a portfolio with value

$$V_m = f(t_m, \mathbf{Z}_m),$$

where f is a known function and $\mathbf{Z}_m$ is a vector of risk-factors. The loss L_{m+1} is then given by

$$L_{m+1} = l_{[m]}(\mathbf{X}_{m+1}).$$

Suppose we have observed the risk factors $\mathbf{Z}_{m-n+1}, \ldots, \mathbf{Z}_m$. These observations will be called historical data. How can we use these observations to compute Value-at-Risk or expected shortfall for the loss L_{m+1} ?

4.1 Empirical VaR and ES

Suppose we have observations $x_1, \ldots, x_n$ of iid random variables $X_1, \ldots, X_n$ with distribution F. The *empirical distribution* function is then given by

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{[X_k, \infty)}(x).$$

The empirical quantile is then given by

$$q_\alpha(F_n) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\} = F_n^{\leftarrow}(\alpha).$$

If we order the sample $X_1, \ldots, X_n$ such that $X_{1,n} \geq \cdots \geq X_{n,n}$ (if F is continuous, then $X_j \neq X_k$ a.s. for $j \neq k$), then the empirical quantile is given by

$$q_\alpha(F_n) = X_{[n(1-\alpha)]+1,n}, \quad \alpha \in (0, 1),$$

where $[y]$ is the integer part of y, $[y] = \sup\{n \in \mathbb{N} : n \leq y\}$ (the largest integer less or equal to y). If F is strictly increasing, then $q_\alpha(F_n) \to q_\alpha(F)$ a.s. as $n \to \infty$ for every $\alpha \in (0, 1)$. Thus, based on the observations $x_1, \ldots, x_n$ we may estimate the quantile $q_\alpha(F)$ by the empirical estimate $\widehat{q}_\alpha(F) = x_{[n(1-\alpha)]+1,n}$. The empirical estimator for expected shortfall is given by

$$\widehat{\mathrm{ES}}_\alpha(F) = \frac{\sum_{k=1}^{[n(1-\alpha)]+1} x_{k,n}}{[n(1-\alpha)]+1}$$

which is the average of the $[n(1-\alpha)]+1$ largest observations.

The reliability of these estimates is of course related to α and to the number of observations. As the true distribution is unknown explicit confidence bounds can in general not be obtained. However, approximate confidence bounds can be obtained using nonparametric techniques. This is described in the next section.

4.2   Confidence intervals

Suppose we have observations $x_1, \ldots, x_n$ of iid random variables $X_1, \ldots, X_n$ from an unknown distribution F and that we want to construct a confidence interval for the risk measure $\varrho(F)$. That is, given $p \in (0,1)$ we want to find a stochastic interval (A, B), where $A = f_A(X_1, \ldots, X_n)$ and $B = f_B(X_1, \ldots, X_n)$ for some functions f_A, f_B, such that

$$\mathrm{P}(A < \varrho(F) < B) = p.$$

The interval (a, b), where $a = f_A(x_1, \ldots, x_n)$ and $b = f_B(x_1, \ldots, x_n)$, is a confidence interval for $\varrho(F)$ with confidence level p. Typically we want a double-sided and centered interval so that

$$\mathrm{P}(A < \varrho(F) < B) = p, \quad \mathrm{P}(A \geq \varrho(F)) = \mathrm{P}(B \leq \varrho(F)) = (1-p)/2.$$

Unfortunately, F is unknown so we cannot find suitable functions f_A, f_B. However, we can construct approximate confidence intervals. Moreover, if $\varrho(F)$ is a quantile of F (Value-at-Risk), then we can actually find exact confidence intervals for $\varrho(F)$, but not for arbitrary choices of confidence levels p.

4.2.1   Exact confidence intervals for Value-at-Risk

Suppose we have observations $x_1, \ldots, x_n$ from iid random variables $X_1, \ldots, X_n$ with common unknown continuous distribution function F. Suppose further that we want to construct a confidence interval (a, b) for the quantile $q_\alpha(F)$, where $a = f_A(x_1, \ldots, x_n)$ and $b = f_B(x_1, \ldots, x_n)$ such that

$$\mathrm{P}(A < q_\alpha(F) < B) = p, \quad \mathrm{P}(A \geq q_\alpha(F)) = \mathrm{P}(B \leq q_\alpha(F)) = (1-p)/2,$$

where p is a confidence level and $A = f_A(X_1, \ldots, X_n)$ and $B = f_B(X_1, \ldots, X_n)$. Since F is unknown we cannot find a and b. However, we can look for $i > j$ and the smallest $p' \geq p$ such that

$$\mathrm{P}(X_{i,n} < q_\alpha(F) < X_{j,n}) = p',$$
$$\mathrm{P}(X_{i,n} \geq q_\alpha(F)) \leq (1-p)/2, \quad \mathrm{P}(X_{j,n} \leq q_\alpha(F)) \leq (1-p)/2. \qquad (4.1)$$

Let $Y_\alpha = \#\{X_k > q_\alpha(F)\}$, i.e. the number of sample points exceeding $q_\alpha(F)$. It is easily seen that Y_α is Binomial$(n, 1-\alpha)$-distributed. Notice that

$$\mathrm{P}(X_{1,n} \leq q_\alpha(F)) = \mathrm{P}(Y_\alpha = 0),$$
$$\mathrm{P}(X_{2,n} \leq q_\alpha(F)) = \mathrm{P}(Y_\alpha \leq 1),$$
$$\ldots$$
$$\mathrm{P}(X_{j,n} \leq q_\alpha(F)) = \mathrm{P}(Y_\alpha \leq j - 1).$$

Similarly, $\mathrm{P}(X_{i,n} \geq q_\alpha(F)) = 1 - \mathrm{P}(Y_\alpha \leq i - 1)$. Hence, we can compute $\mathrm{P}(X_{j,n} \leq q_\alpha(F))$ and $\mathrm{P}(X_{i,n} \geq q_\alpha(F))$ for different i and j until we find indices that satisfy (4.1).
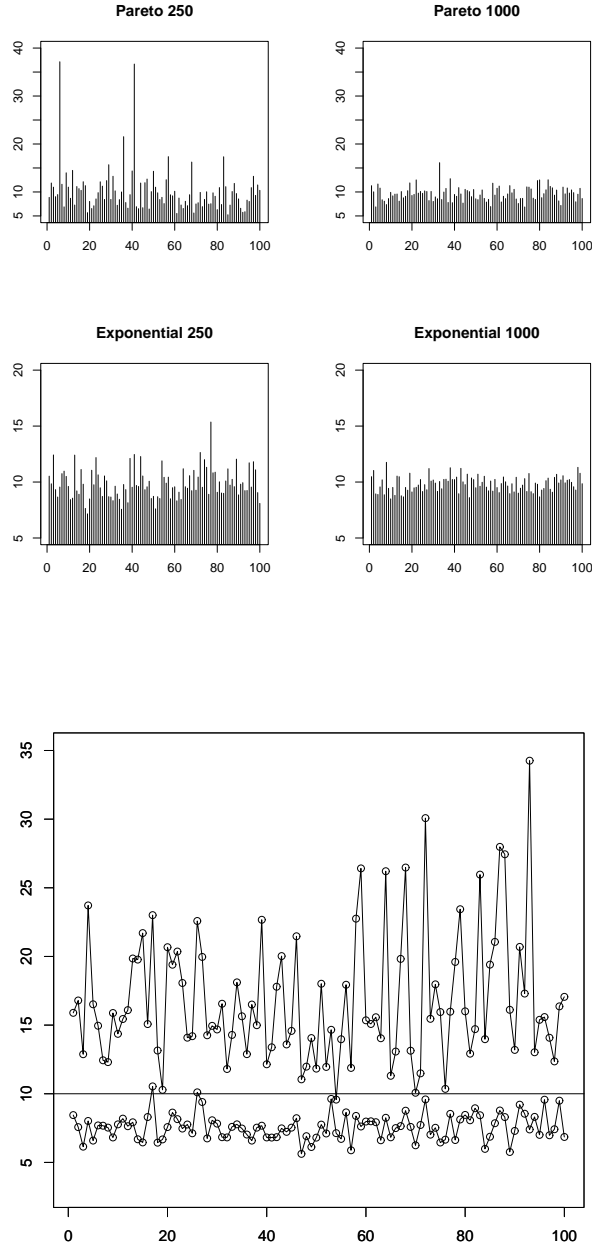
Figure 5: Upper: Empirical estimates of $VaR_{0.99}$ for samples of different sizes and from different distributions with $VaR_{0.99} = 10$. Lower: Simulated 97.6% confidence intervals $(x_{18,1000}, x_{4,1000})$ for $VaR_{0.99}(X) = 10$ based on samples of size 1000 from a Pareto distribution.

Example 4.1 Suppose we have an iid sample $X_1, \ldots, X_{10}$ with common unknown continuous distribution function F and that we want a confidence interval for $q_{0.8}(F)$ with confidence level $p' \geq p = 0.75$. Since $Y_{0.8}$ is Binomial$(10, 0.2)$-

distributed and $P(X_{j,10} \leq q_{0.8}(F)) = P(Y_{0.8} \leq j - 1)$ we find that

$$P(X_{1,10} \leq q_{0.8}(F)) \approx 0.11 \quad \text{and} \quad P(X_{4,10} \geq q_{0.8}(F)) \approx 0.12.$$

Notice that $\max\{0.11, 0.12\} \leq (1 - p)/2 = 0.125$ and $P(X_{4,10} < q_{0.8}(F) < X_{1,10}) \approx 0.77$ so $(x_{4,10}, x_{1,10})$ is a confidence interval for $q_{0.8}(F)$ with confidence level 77%. ∎

4.2.2 Using the bootstrap to obtain confidence intervals

Using the so-called nonparametric bootstrap method we can obtain confidence intervals for e.g. risk measures such as Value-at-Risk and expected shortfall. The nonparametric bootstrap works as follows.

Suppose we have observations $x_1, \ldots, x_n$ of iid random variables $X_1, \ldots, X_n$ and we want to estimate some parameter θ which depends on the unknown distribution F of the X's. For instance θ could be the α-quantile $\theta = q_\alpha(F)$. First an estimator $\widehat{\theta}(x_1, \ldots, x_n)$ of θ is constructed, e.g. $\widehat{\theta}(x_1, \ldots, x_n) = x_{[n(1-\alpha)]+1,n}$. Now we want to construct a confidence interval for θ with confidence level p (for instance $p = 0.95$). To construct a confidence interval we need to know the distribution of $\widehat{\theta}(X_1, \ldots, X_n)$. If F was known this distribution could be approximated arbitrarily well by simulating from F many (N large) times to construct new samples $\widetilde{X}_1^{(i)}, \ldots, \widetilde{X}_n^{(i)}$, $i = 1, \ldots, N$, and compute the estimator for each of these samples to get $\widetilde{\theta}_i = \widehat{\theta}(\widetilde{X}_1^{(i)}, \ldots, \widetilde{X}_n^{(i)})$, $i = 1, \ldots, N$. As $N \to \infty$ the empirical distribution

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{[\widetilde{\theta}_i, \infty)}(x)$$

of $\widehat{\theta}(X_1, \ldots, X_n)$ will converge to the true distribution of $\widehat{\theta}(X_1, \ldots, X_n)$. *The problem is that F is not known.*

What is known is the empirical distribution F_n which puts point masses $1/n$ at the points $X_1, \ldots, X_n$. If n is relatively large we expect that F_n is a good approximation of F. Moreover, we can resample from F_n simply by drawing with replacement among $X_1, \ldots, X_n$. We denote such a resample by $X_1^*, \ldots, X_n^*$. Then we may compute $\theta^* = \widehat{\theta}(X_1^*, \ldots, X_n^*)$. Since F_n approximates the true distribution F we expect the distribution of $\widehat{\theta}(X_1^*, \ldots, X_n^*)$ to approximate the true distribution of $\widehat{\theta}(X_1, \ldots, X_n)$. To obtain the distribution of $\widehat{\theta}(X_1^*, \ldots, X_n^*)$ we resample many (N large) times to create new samples $X_1^{*(i)}, \ldots, X_n^{*(i)}$, $i = 1, \ldots, N$. For each of these samples we compute the corresponding estimate of θ, i.e. $\theta_i^* = \widehat{\theta}(X_1^{*(i)}, \ldots, X_n^{*(i)})$. The empirical distribution F_N^θ, given by

$$F_N^\theta(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{[\theta_i^*, \infty)}(x),$$

is then an approximation of the true distribution of $\widehat{\theta}(X_1, \ldots, X_n)$ denoted F^θ. A confidence interval is then constructed using $A = q_{(1-p)/2}(F_N^\theta)$ and $B =$

$q_{(1+p)/2}(F_N^\theta)$. This means that $A = \theta^*_{[N(1+p)/2]+1,N}$ and $B = \theta^*_{[N(1-p)/2]+1,N}$, where $\theta^*_{1,N} \geq \cdots \geq \theta^*_{N,N}$ is the ordered sample of $\theta^*_1, \ldots, \theta^*_N$.

The nonparametric bootstrap to obtain confidence intervals can be summarized in the following steps: Suppose we have observations $x_1, \ldots, x_n$ of the iid random variables $X_1, \ldots, X_n$ with distribution F and we have an estimator $\widehat{\theta}(x_1, \ldots, x_n)$ of an unknown parameter θ.

- Resample N times among $x_1, \ldots, x_n$ to obtain new samples $x_1^{*(i)}, \ldots, x_n^{*(i)}$, $i = 1, \ldots, N$.

- Compute the estimator for each of the new samples to get

$$\theta_i^* = \widehat{\theta}(x_1^{*(i)}, \ldots, x_n^{*(i)}), \quad i = 1, \ldots, N.$$

- Construct a confidence interval I_p with confidence level p as

$$I_p = (\theta^*_{[N(1+p)/2]+1,N}, \theta^*_{[N(1-p)/2]+1,N}),$$

where $\theta^*_{1,N} \geq \cdots \geq \theta^*_{N,N}$ is the ordered sample of $\theta^*_1, \ldots, \theta^*_N$.

4.3 Historical simulation

In the historical simulation approach we suppose we have observations of risk-factors $\mathbf{Z}_{m-n}, \ldots, \mathbf{Z}_m$ and hence also the risk-factor changes $\mathbf{X}_{m-n+1}, \ldots, \mathbf{X}_m$. We denote these observations by $\mathbf{x}_{m-n+1}, \ldots, \mathbf{x}_m$. Using the loss operator we can compute the corresponding observations of losses $l_k = l_{[m]}(\mathbf{x}_{m-k+1})$, $k = 1, \ldots, n$. Note that l_k is the loss that we will experience if we have the risk factor change $\mathbf{x}_{m-k+1}$ over the next period. This gives us a sample from the loss distribution. It is assumed that the losses during the different time intervals are iid. Then the empirical VaR and ES can be estimated using

$$\widehat{\text{VaR}}_\alpha(L) = \widehat{q}_\alpha(F_{L^n}) = l_{[n(1-\alpha)]+1,n}$$

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

where $l_{1,n} \geq \cdots \geq l_{n,n}$ is the ordered sample.

Similarly we can also aggregate over several days. Say, for instance, that we are interested in the Value-at-Risk for the aggregate loss over ten days. Then we simply use the historical observations given by

$$l_k^{(10)} = l_{[m]}\left(\sum_{j=1}^{10} \mathbf{x}_{m-n+10(k-1)+j}\right), \quad k = 1, \ldots, [n/10],$$

to compute the empirical VaR and ES.

Advantage: This approach is easy to implement and keeps the dependence structure between the components of the vectors of risk-factor changes $\mathbf{X}_{m-k}$.

Disadvantage: The worst case is never worse than what has happened in history. We need a very large sample of relevant historical data to get reliable estimates of Value-at-Risk and expected shortfall.

4.4 Variance–Covariance method

The basic idea of the variance–covariance method is to study the linearized loss

$$L^{\Delta}_{m+1} = l^{\Delta}_{[m]}(\mathbf{X}_{m+1}) = -(c + \sum_{i=1}^{d} w_i X_{m+1,i}) = -(c + \mathbf{w}^{\mathrm{T}}\mathbf{X}_{m+1})$$

where $c = c_m$, $w_i = w_{m,i}$ (i.e. known at time m), $\mathbf{w} = (w_1, \ldots, w_d)^{\mathrm{T}}$ are weights and $\mathbf{X}_{m+1} = (X_{m+1,1}, \ldots, X_{m+1,d})^{\mathrm{T}}$ the risk-factor changes.

It is then assumed that $\mathbf{X}_{m+1} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, i.e. that the risk-factor changes follow a multivariate (d-dimensional) normal distribution. Using the properties of the multivariate normal distribution we have

$$\mathbf{w}^{\mathrm{T}}\mathbf{X}_{m+1} \sim \mathrm{N}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}, \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}).$$

Hence, the loss distribution is normal with mean $-c - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}$ and variance $\mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}$. Suppose that we have $n+1$ historical observations of the risk-factors and the risk-factor changes $\mathbf{X}_{m-n+1}, \ldots, \mathbf{X}_m$. Then (assuming that the risk-factor changes are iid or at least weakly dependent) the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ can be estimated as usual by

$$\widehat{\mu}_i = \frac{1}{n} \sum_{k=1}^{n} X_{m-k+1,i}, \quad i = 1, \ldots, d,$$

$$\widehat{\Sigma}_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (X_{m-k+1,i} - \widehat{\mu}_i)(X_{m-k+1,j} - \widehat{\mu}_j), \quad i,j = 1, \ldots, d.$$

The estimated VaR is then given analytically by

$$\widehat{\mathrm{VaR}}_{\alpha}(L) = -c - \mathbf{w}^{\mathrm{T}}\widehat{\boldsymbol{\mu}} + \sqrt{\mathbf{w}^{\mathrm{T}}\widehat{\Sigma}\mathbf{w}}\,\Phi^{-1}(\alpha).$$

Advantage: Analytic solutions can be obtained: no simulations required. Easy to implement.

Disadvantage: Linearization not always appropriate. We need a short time horizon to justify linearization (see e.g. Example 3.6). The normal distribution may considerably underestimate the risk. We need proper justification that the normal distribution is appropriate before using this approach. *In later chapters we will introduce elliptical distributions, distributions that share many of the nice properties of the multivariate normal distribution. The Variance-Covariance method works well if we replace the assumption of multivariate normality with the weaker assumption of ellipticality. This may provide a model that fits data better.*

4.5 Monte-Carlo methods

Suppose we have observed the risk-factors $\mathbf{Z}_{m-n}, \ldots, \mathbf{Z}_m$ and risk-factor changes $\mathbf{X}_{m-n+1}, \ldots, \mathbf{X}_m$. We suggest a parametric model for $\mathbf{X}_{m+1}$. For instance, that $\mathbf{X}_{m+1}$ has distribution function F and is independent of $\mathbf{X}_{m-n+1}, \ldots, \mathbf{X}_m$.

When an appropriate model for $\mathbf{X}_{m+1}$ is chosen and the parameters estimated we simulate a large number N of outcomes from this distribution to get $\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_N$. For each outcome we can compute the corresponding losses $l_1, \ldots, l_N$ where $l_k = l_{[m]}(\widetilde{\mathbf{x}}_k)$. The true loss distribution of L_{m+1} is then approximated by the empirical distribution F_{L^N} given by

$$F_{L^N}(x) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}_{[l_k, \infty)}(x).$$

Value-at-Risk and expected shortfall can then be estimated by

$$\widehat{\text{VaR}}_\alpha(L) = q_\alpha(F_{L^N}) = l_{[N(1-\alpha)]+1, N}$$

$$\widehat{\text{ES}}_\alpha(L) = \frac{\sum_{k=1}^{[N(1-\alpha)]+1} l_{k,N}}{[N(1-\alpha)] + 1}$$

Advantage: Very flexible. Practically any model that you can simulate from is possible to use. You can also use time series in the Monte-Carlo method which enables you to model time dependence between risk-factor changes.

Disadvantage: Computationally intensive. You need to run a large number of simulations in order to get good estimates. This may take a long time (hours, days) depending on the complexity of the model.

Example 4.2 Consider a portfolio consisting of one share of a stock with stock price S_k and assume that the log returns $X_{k+1} = \ln S_{k+1} - \ln S_k$ are iid with distribution function F_θ, where θ is an unknown parameter. The parameter θ can be estimated from historical data using for instance maximum likelihood and given the information about the stock price S_0 today, the Value-at-Risk for our portfolio loss over the time period today-until-tomorrow is

$$\text{VaR}_\alpha(L_1) = S_0(1 - \exp\{F_\theta^{\leftarrow}(1 - \alpha)\}),$$

i.e. the α-quantile of the distribution of the loss $L_1 = -S_0(\exp\{X_1\} - 1)$. However, the expected shortfall may be difficult to compute explicitly if $F_\theta^{\leftarrow}$ has a complicated expression. Instead of performing numerical integration we may use the Monte-Carlo approach to compute expected shortfall. ∎

Example 4.3 Consider the situation in Example 4.2 with the exception that the log return distribution is given by a GARCH$(1,1)$ model:

$$X_{k+1} = \sigma_{k+1} Z_{k+1}, \quad \sigma_{k+1}^2 = a_0 + a_1 X_k^2 + b_1 \sigma_k^2,$$

where the Z_k's are independent and standard normally distributed and a_0, a_1 and b_1 are parameters to be estimated. Because of the recursive structure it is very easy to simulate from this model. However, analytic computation of the ten-day Value-at-Risk or expected shortfall is very difficult. ∎

5 Extreme value theory for random variables with heavy tails

Given historical loss data a risk manager typically wants to estimate the probability of future large losses to assess the risk of holding a certain portfolio. Extreme Value Theory (EVT) provides the tools for an optimal use of the loss data to obtain accurate estimates of probabilities of large losses or, more generally, of extreme events. Extreme events are particularly frightening because although they are by definition rare, they may cause severe losses to a financial institution or insurance company. Empirical estimates of probabilities of losses larger than what has been observed so far are useless: such an event will be assigned zero probability. Even if the loss lies just within the range of the loss data set, empirical estimates have poor accuracy. However, under certain conditions, EVT methods can extrapolate information from the loss data to obtain meaningful estimates of the probability of rare events such as large losses. This also means that accurate estimates of Value-at-Risk (VaR) and Expected Shortfall (ES) can be obtained.

In this and the following two chapters we will present aspects of and estimators provided by EVT. In order to present the material and derive the expressions of the estimators without a lot of technical details we focus on EVT for distributions with "heavy tails" (see below). Moreover, empirical investigations often support the use of heavy-tailed distributions.

Empirical investigations have shown that daily and higher-frequency returns from financial assets typically have distributions with *heavy tails*. Although there is no definition of the meaning of "heavy tails" it is common to consider the right tail $\overline{F}(x) = 1 - F(x)$, x large, of the distribution function F heavy if

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{e^{-\lambda x}} = \infty \quad \text{for every } \lambda > 0,$$

i.e. if it is heavier than the right tail of every exponential distribution. It is also not unusual to consider a random variable heavy-tailed if not all moments are finite. We will now study the useful class of heavy-tailed distributions with *regularly varying* tails.

5.1 Quantile-quantile plots

In this section we will consider some useful practical methods to study the extremal properties of a data set. To illustrate the methods we will consider a dataset consisting of claims in million Danish Kroner from fire insurance in Denmark. We may observe that there are a few claims much larger the 'everyday' claim. This suggests that the claims have a heavy-tailed distribution. To get an indication of the heaviness of the tails it is useful to use so-called quantile-quantile plots (qq-plots).

Suppose we have a sample $X_1, \ldots, X_n$ of iid random variables but we don't know the distribution of X_1. One would typically suggest a reference distribution F and want to test whether it is reasonable to assume that the data is
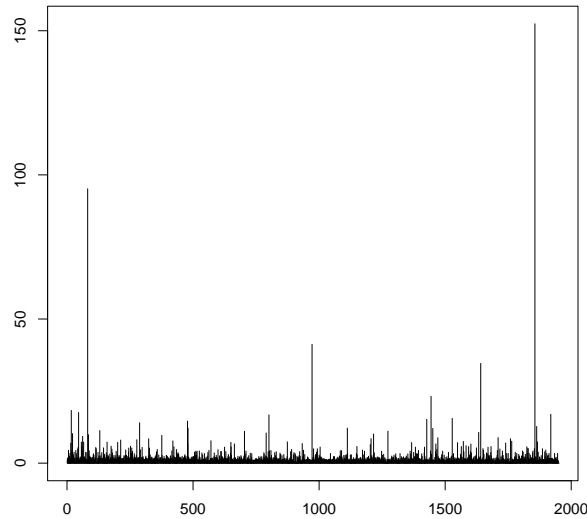
Figure 6: Claims from fire insurance in Denmark in million Danish Kroner.

distributed according to this distribution. If we define the ordered sample as $X_{n,n} \leq X_{n-1,n} \leq \cdots \leq X_{1,n}$, then the qq-plot consists of the points

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

If the data has a similar distribution as the reference distribution then the qq-plot is approximately linear. An important property is that the plot remains approximately linear if the data has a distribution which is a linear transformation of the reference distribution, i.e. from the associated location-scale family $F_{\mu,\sigma}(x) = F((x-\mu)/\sigma)$. Thus, the qq-plot enables us to come up with a suitable class of distributions that fits the data and then we may estimate the location and scale parameters. If the data comes from the reference distribution, with distribution function F, then the qq-plot should be aproximately linear with intercept 0 and slope 1. If the data is a sample from $F_{\mu,\sigma}(x) = F((x-\mu)/\sigma)$, then the qq-plot is still approximately linear since $F_{\mu,\sigma}^{\leftarrow}(p) = \mu + \sigma F^{\leftarrow}(p)$. Moreover, the parameters μ and σ can be estimated from the intercept and slope of the qq-plot.

The qq-plot is particularly useful for studying the tails of the distribution. Given a reference distribution F, if F has heavier tails than the data then the plot will curve down at the left and/or up at the right and the opposite if the reference distribution has too light tails.

Exercise 5.1 Consider the distribution functions F and G given by $F(x) =$

29
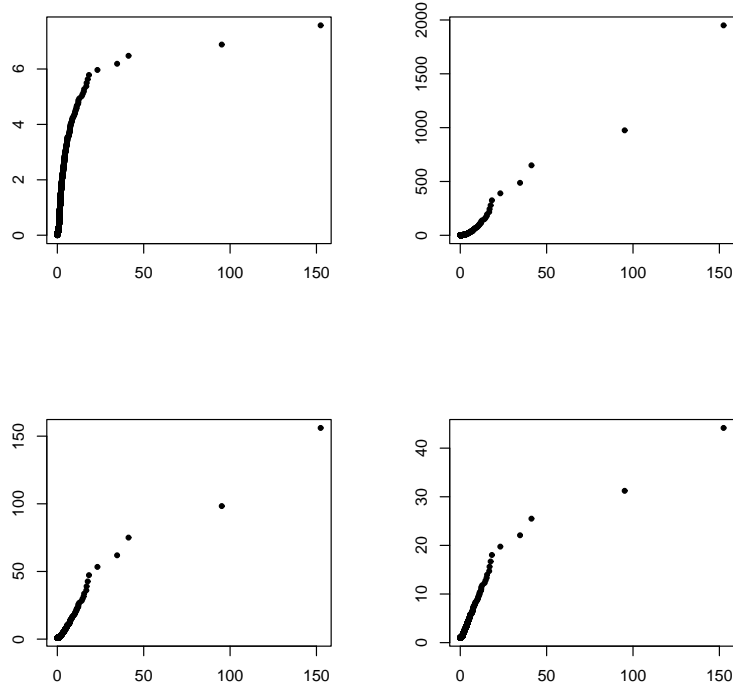
Figure 7: Quantile-quantile plots for the Danish fire insurance claims. Upper left: qq-plot with standard exponential reference distribution. The plot curves down which indicates that data has tails heavier than exponential. Upper right, lower left, lower right: qq-plot against a Pareto(α)-distribution for $\alpha = 1, 1.5, 2$. The plots are approximately linear which indicates that the data may have a Pareto(α)-distribution with $\alpha \in (1, 2)$.

$1 - e^{-x}$ $(x > 0)$ and $G(x) = 1 - x^{-2}$ $(x > 1)$. Plot

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

and interpret the result.

5.2 Regular variation

We start by introducing regularly varying functions.

Definition 5.1 *A function $h : (0, \infty) \to (0, \infty)$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ (written $h \in RV_\rho$) if*

$$\lim_{t \to \infty} \frac{h(tx)}{h(t)} = x^\rho \quad \text{for every } x > 0. \tag{5.1}$$

∎
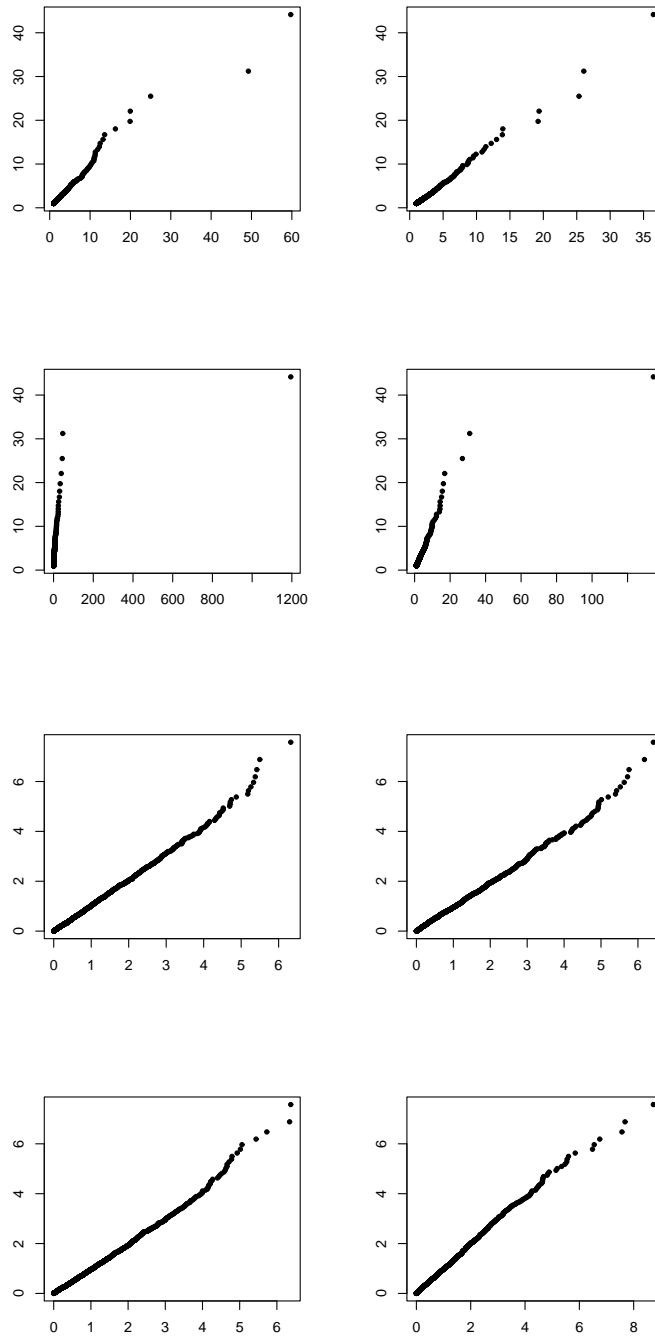
Figure 8: Quantile-quantile plots for 1949 simulated data points from a distribution F with F as reference distribution. The upper four plots: $F = \text{Pareto}(2)$. The lower four plots: $F = \text{Exp}(1)$.

If $\rho = 0$, then we call h slowly varying (at ∞). Slowly varying functions are generically denoted by L. If $h \in \mathrm{RV}_\rho$, then $h(x)/x^\rho \in \mathrm{RV}_0$. Hence, setting $L(x) = h(x)/x^\rho$ we see that a function $h \in \mathrm{RV}_\rho$ can always be represented as $h(x) = L(x)x^\rho$. If $\rho < 0$, then the convergence in (5.1) is uniform on intervals $[b, \infty)$ for $b > 0$, i.e.

$$\lim_{t\to\infty} \sup_{x\in[b,\infty)} \left| \frac{h(tx)}{h(t)} - x^\rho \right| = 0.$$

The most natural examples of slowly varying functions are positive constants and functions converging to positive constants. Other examples are logarithms and iterated logarithms. The following functions L are slowly varying.

(a) $\lim_{x\to\infty} L(x) = c \in (0, \infty)$.

(b) $L(x) = \ln(1 + x)$.

(c) $L(x) = \ln(1 + \ln(1 + x))$.

(d) $L(x) = \ln(e + x) + \sin x$.

Note however that a slowly varying function can have infinite oscillation in the sense that $\liminf_{x\to\infty} L(x) = 0$ and $\limsup_{x\to\infty} L(x) = \infty$.

Example 5.1 (1) Let $F(x) = 1 - x^{-\alpha}$, for $x \geq 1$ and $\alpha > 0$. Then $\overline{F}(tx)/\overline{F}(t) = x^{-\alpha}$ for $t > 0$. Hence $\overline{F} \in \mathrm{RV}_{-\alpha}$.
(2) Let ∎

**Definition 5.2** *A nonnegative random variable is said to be regularly varying if its distribution function F satisfies $\overline{F} \in RV_{-\alpha}$ for some $\alpha \geq 0$.* ∎

Remark 5.1 *If X is a nonnegative random variable with distribution function F satisfying $\overline{F} \in RV_{-\alpha}$ for some $\alpha > 0$, then*

$$\begin{aligned} \mathrm{E}(X^\beta) &< \infty & \text{if } \beta < \alpha, \\ \mathrm{E}(X^\beta) &= \infty & \text{if } \beta > \alpha. \end{aligned}$$

Although the converse does not hold in general, it is useful to think of regularly varying random variables as those random variables for which the β-moment does not exist for β larger than some $\alpha > 0$.

Example 5.2 Consider two risks X_1 and X_2 which are assumed to be nonnegative and iid with common distribution function F. Assume further that F has a regularly varying right tail, i.e. $\overline{F} \in \mathrm{RV}_{-\alpha}$. An investor has bought two shares of the first risky asset and the probability of a portfolio loss greater than l is thus given by $\mathrm{P}(2X_1 > l)$. *Can the loss probability be made smaller by changing to the well diversified portfolio with one share of each risky asset?* To answer this question we study the following ratio of loss probabilities:

$$\frac{\mathrm{P}(X_1 + X_2 > l)}{\mathrm{P}(2X_1 > l)}$$

for large l. We have, for $\varepsilon \in (0, 1/2)$,

$$
\begin{aligned}
\mathrm{P}(X_1 + X_2 &> l) \\
&= 2\,\mathrm{P}(X_1 + X_2 > l, X_1 \le \varepsilon l) + \mathrm{P}(X_1 + X_2 > l, X_1 > \varepsilon l, X_2 > \varepsilon l) \\
&\le 2\,\mathrm{P}(X_2 > (1 - \varepsilon)l) + \mathrm{P}(X_1 > \varepsilon l)^2.
\end{aligned}
$$

and

$$
\mathrm{P}(X_1 + X_2 > l) \ge \mathrm{P}(X_1 > l \text{ or } X_2 > l) = 2\,\mathrm{P}(X_1 > l) - \mathrm{P}(X_1 > l)^2.
$$

Hence,

$$
\underbrace{\frac{2\,\mathrm{P}(X_1 > l) - \mathrm{P}(X_1 > l)^2}{\mathrm{P}(2X_1 > l)}}_{g(\alpha,\epsilon,l)} \le \frac{\mathrm{P}(X_1 + X_2 > l)}{\mathrm{P}(2X_1 > l)}
$$

$$
\le \underbrace{\frac{2\,\mathrm{P}(X_2 > (1 - \varepsilon)l) + \mathrm{P}(X_1 > \varepsilon l)^2}{\mathrm{P}(2X_1 > l)}}_{h(\alpha,\epsilon,l)}.
$$

We have

$$
\lim_{l \to \infty} g(\alpha, \epsilon, l) = 2 \lim_{l \to \infty} \frac{\mathrm{P}(X_1 > l)}{\mathrm{P}(X_1 > l/2)} = 2^{1 - \alpha}
$$

and similarly $\lim_{l \to \infty} h(\alpha, \epsilon, l) = 2^{1-\alpha}(1 - \varepsilon)^{-\alpha}$. Since $\varepsilon > 0$ can be chosen arbitrary small we conclude that

$$
\lim_{l \to \infty} \frac{\mathrm{P}(X_1 + X_2 > l)}{\mathrm{P}(2X_1 > l)} = 2^{1 - \alpha}.
$$

This means that for $\alpha < 1$ (very heavy tails) diversification does not give us a portfolio with smaller probability of large losses. However, for $\alpha > 1$ (the risks have finite means) diversification reduces the probability of large losses. ∎

Example 5.3 Let X_1 and X_2 be as in the previous example, and let $\alpha \in (0, 1)$. We saw that for l sufficiently large we have $\mathrm{P}(X_1 + X_2 > l) > \mathrm{P}(2X_1 > l)$. Hence, for $p \in (0, 1)$ sufficiently large

$$
\begin{aligned}
\mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2) = 2\,\mathrm{VaR}_p(X_1) &= \mathrm{VaR}_p(2X_1) \\
&= \inf\{l \in \mathbb{R} : \mathrm{P}(2X_1 > l) \le 1 - p\} \\
&< \inf\{l \in \mathbb{R} : \mathrm{P}(X_1 + X_2 > l) \le 1 - p\} \\
&= \mathrm{VaR}_p(X_1 + X_2).
\end{aligned}
$$

∎

Example 5.4 Let X and Y be positive random variables representing losses in two lines of business (losses due to fire and car accidents) of an insurance

company. Suppose that X has distribution function F which satisfies $\overline{F} \in \mathrm{RV}_{-\alpha}$ for $\alpha > 0$. Moreover, suppose that $\mathrm{E}(Y^k) < \infty$ for every $k > 0$, i.e. Y has finite moments of all orders.

The insurance company wants to compute $\lim_{x \to \infty} \mathrm{P}(X > x \mid X + Y > x)$ to know the probability of a large loss in the fire insurance line given a large total loss.

We have, for every $\varepsilon \in (0, 1)$ and $x > 0$,

$$\begin{aligned}
\mathrm{P}(X + Y > x) &= \mathrm{P}(X + Y > x, X > (1 - \varepsilon)x) + \mathrm{P}(X + Y > x, X \le (1 - \varepsilon)x) \\
&\le \mathrm{P}(X + Y > x, X > (1 - \varepsilon)x) + \mathrm{P}(X + Y > x, Y > \varepsilon x) \\
&\le \mathrm{P}(X > (1 - \varepsilon)x) + \mathrm{P}(Y > \varepsilon x).
\end{aligned}$$

Hence,

$$\begin{aligned}
1 &\le \frac{\mathrm{P}(X + Y > x)}{\mathrm{P}(X > x)} \\
&\le \frac{\mathrm{P}(X > (1 - \varepsilon)x)}{\mathrm{P}(X > x)} + \frac{\mathrm{P}(Y > \varepsilon x)}{\mathrm{P}(X > x)} \\
&\le \frac{\mathrm{P}(X > (1 - \varepsilon)x)}{\mathrm{P}(X > x)} + \frac{\mathrm{E}(Y^{2\alpha})}{(\varepsilon x)^{2\alpha}\, \mathrm{P}(X > x)} \\
&\to (1 - \varepsilon)^{-\alpha} + 0
\end{aligned}$$

as $x \to \infty$. At the second to last step above, Markov's inequality was used. Since this is true for every $\varepsilon \in (0, 1)$, choosing ε arbitrarily small gives

$$\lim_{x \to \infty} \frac{\mathrm{P}(X + Y > x)}{\mathrm{P}(X > x)} = 1.$$

Hence,

$$\begin{aligned}
\lim_{x \to \infty} \mathrm{P}(X > x \mid X + Y > x) &= \lim_{x \to \infty} \frac{\mathrm{P}(X > x, X + Y > x)}{\mathrm{P}(X + Y > x)} \\
&= \lim_{x \to \infty} \frac{\mathrm{P}(X > x)}{\mathrm{P}(X + Y > x)} = 1.
\end{aligned}$$

We have found that if the insurance company suffers a large loss, it is likely that this is due to a large loss in the fire insurance line only. $\blacksquare$

6 Hill estimation

Suppose we have an iid sample of positive random variables $X_1, \ldots, X_n$ from an unknown distribution function F with a regularly varying right tail. That is, $\overline{F}(x) = x^{-\alpha}L(x)$ for some $\alpha > 0$ and a slowly varying function L. In this section we will describe Hill's method for estimating α.

We will use a result known as the *Karamata Theorem* which says that for $\beta < -1$

$$\int_u^\infty x^\beta L(x) \mathrm{d}x \sim -(\beta+1)^{-1} u^{\beta+1} L(u) \quad \text{as } u \to \infty,$$

where $\sim$ means that the ratio of the left and right sides tends to 1. Using integration by parts we find that

$$\frac{1}{\overline{F}(u)} \int_u^\infty (\ln x - \ln u) \mathrm{d}F(x)$$

$$= \frac{1}{\overline{F}(u)} \left(\left[-(\ln x - \ln u)\overline{F}(x) \right]_u^\infty + \int_u^\infty \frac{\overline{F}(x)}{x} \mathrm{d}x \right)$$

$$= \frac{1}{u^{-\alpha}L(u)} \int_u^\infty x^{-\alpha-1} L(x) \mathrm{d}x.$$

Hence, by the Karamata Theorem,

$$\frac{1}{\overline{F}(u)} \int_u^\infty (\ln x - \ln u) \mathrm{d}F(x) \to \frac{1}{\alpha} \quad \text{as } u \to \infty. \tag{6.1}$$

To turn this into an estimator we replace F by the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{[X_k, \infty)}(x)$$

and replace u by a high data dependent level $X_{k,n}$. Then

$$\frac{1}{\overline{F}_n(X_{k,n})} \int_{X_{k,n}}^\infty (\ln x - \ln X_{k,n}) \mathrm{d}F_n(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln X_{j,n} - \ln X_{k,n}).$$

If $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$, then $X_{k,n} \to \infty$ a.s. as $n \to \infty$ and by (6.1)

$$\frac{1}{k-1} \sum_{j=1}^{k-1} (\ln X_{j,n} - \ln X_{k,n}) \xrightarrow{\text{P}} \frac{1}{\alpha} \quad \text{as } n \to \infty.$$

The same result holds if we replace $k-1$ by k. This gives us the *Hill estimator*

$$\widehat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k,n}) \right)^{-1}.$$

6.1 Selecting the number of upper order statistics

We have seen that if $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$, then

$$\widehat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^{k} (\ln X_{j,n} - \ln X_{k,n}) \right)^{-1} \xrightarrow{P} \alpha \quad \text{as } n \to \infty.$$

In practice however we have a sample of fixed size n and we need to find a suitable k such that $\widehat{\alpha}_{k,n}^{(H)}$ is a good estimate of α. The plot of the pairs

$$\left\{ (k, \widehat{\alpha}_{k,n}^{(H)}) : k = 2, \ldots, n \right\}$$

is called the *Hill plot*. An estimator of α is obtained by graphical inspection of the Hill plot and an estimate of α should be taken for values of k where the plot is stable. On the one hand, k should not be chosen too small since the small number of data points would lead to a high variance for the estimator. On the other hand, k should not be chosen too large so the estimate is based on sample points from the center of the distribution (this introduces a bias). This is illustrated graphically in Figure 9. We now construct an estimator for the
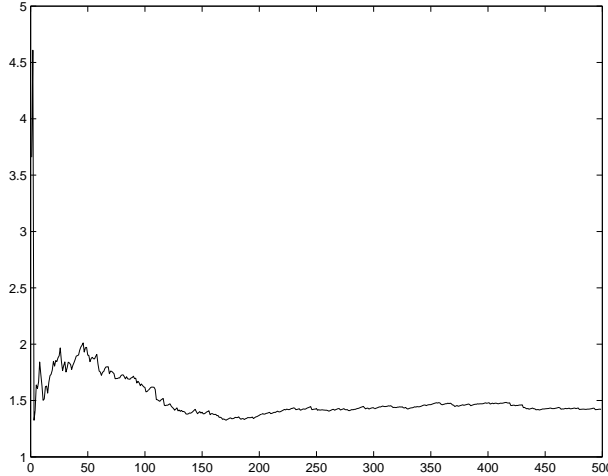
Figure 9: Hill plot of the Danish fire insurance data. The plot looks stable for all k (typically this is NOT the case for heavy-tailed data). We may estimate α by $\widehat{\alpha}_{k,n}^{(H)} = 1.9$ ($k = 50$) or $\widehat{\alpha}_{k,n}^{(H)} = 1.5$ ($k = 250$).

tail probability $\overline{F}(x)$, for x large, based on the sample points $X_1, \ldots, X_n$ and the the Hill estimate $\widehat{\alpha}_{k,n}^{(H)}$. Notice that

$$\overline{F}(x) = \overline{F}\left(\frac{x}{X_{k,n}} X_{k,n} \right) \approx \left(\frac{x}{X_{k,n}} \right)^{-\alpha} \overline{F}(X_{k,n})$$

$$\approx \left(\frac{x}{X_{k,n}} \right)^{-\alpha_{k,n}^{(H)}} \overline{F_n}(X_{k,n}) \approx \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\widehat{\alpha}_{k,n}^{(H)}}.$$

This argument can be made more rigorously. Hence, the following estimator seems reasonable:

$$\widehat{\overline{F}(x)} = \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\widehat{\alpha}_{k,n}^{(H)}}.$$

This leads to an estimator of the quantile $q_p(F) = F^{\leftarrow}(p)$.

$$\widehat{q_p(F)} = \inf \left\{ x \in \mathbb{R} : \widehat{\overline{F}(x)} \le 1 - p \right\}$$

$$= \inf \left\{ x \in \mathbb{R} : \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\widehat{\alpha}_{k,n}^{(H)}} \le 1 - p \right\}$$

$$= \left(\frac{n}{k}(1 - p) \right)^{-1/\widehat{\alpha}_{k,n}^{(H)}} X_{k,n}.$$

More information about Hill estimation can be found in [8] and [12].

7 The Peaks Over Threshold (POT) method

Suppose we have an iid sample of random variables $X_1, \ldots, X_n$ from an unknown distribution function F with a regularly varying right tail. It turns out that the distribution of appropriately scaled excesses $X_k - u$ over a high threshold u is typically well approximated by a distribution called the generalized Pareto distribution. This fact can be used to construct estimates of tail probabilities and quantiles.

For $\gamma > 0$ and $\beta > 0$, the generalized Pareto distribution (GPD) function $G_{\gamma, \beta}$ is given by

$$G_{\gamma, \beta}(x) = 1 - (1 + \gamma x/\beta)^{-1/\gamma} \quad \text{for } x \geq 0.$$

Suppose that X is a random variable with distribution function F that has a regularly varying right tail so that $\lim_{u \to \infty} \overline{F}(\lambda u)/\overline{F}(u) = \lambda^{-\alpha}$ for all $\lambda > 0$ and some $\alpha > 0$. Then

$$\lim_{u \to \infty} \mathrm{P}\left(\frac{X - u}{u/\alpha} > x \mid X > u\right) = \lim_{u \to \infty} \frac{\mathrm{P}(X > u(1 + x/\alpha))}{\mathrm{P}(X > u)}$$
$$= (1 + x/\alpha)^{-\alpha} = \overline{G}_{1/\alpha, 1}(x).$$

The *excess distribution function* of X over the threshold u is given by

$$F_u(x) = \mathrm{P}(X - u \leq x \mid X > u) \quad \text{for } x \geq 0.$$

Notice that

$$\overline{F_u}(x) = \frac{\overline{F}(u + x)}{\overline{F}(u)} = \frac{\overline{F}(u(1 + x/u))}{\overline{F}(u)}. \tag{7.1}$$

Since $\overline{F}$ is regularly varying with index $-\alpha < 0$ it holds that $\overline{F}(\lambda u)/\overline{F}(u) \to \lambda^{-\alpha}$ uniformly in $\lambda \geq 1$ as $u \to \infty$, i.e.

$$\lim_{u \to \infty} \sup_{\lambda \geq 1} |\overline{F}(\lambda u)/\overline{F}(u) - \lambda^{-\alpha}| = 0.$$

Hence, from expression (7.1) we see that

$$\lim_{u \to \infty} \sup_{x > 0} |\overline{F_u}(x) - \overline{G}_{\gamma, \beta(u)}(x)| = 0, \tag{7.2}$$

where $\gamma = 1/\alpha$ and $\beta(u) \sim u/\alpha$ as $u \to \infty$.

We now demonstrate how these findings lead to natural tail- and quantile estimators based on the sample points $X_1, \ldots, X_n$. Choose a high threshold u and let

$$N_u = \#\{i \in \{1, \ldots, n\} : X_i > u\}$$

be the number of exceedances of u by $X_1, \ldots, X_n$. Recall from (7.1) that

$$\overline{F}(u + x) = \overline{F}(u)\overline{F_u}(x). \tag{7.3}$$

If u is not too far out into the tail, then the empirical approximation $\overline{F}(u) \approx \overline{F_n}(u) = N_u/n$ is accurate. Moreover, (7.2) shows that the approximation

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

where $\widehat{\gamma}$ and $\widehat{\beta}$ are the estimated parameters, makes sense. Relation (7.3) then suggests a method for estimating the tail of F by estimating $\overline{F}_u(x)$ and $\overline{F}(u)$ separately. Hence, a natural estimator for $\overline{F}(u + x)$ is

$$\widehat{\overline{F}(u + x)} = \frac{N_u}{n}\left(1 + \widehat{\gamma}\frac{x}{\widehat{\beta}}\right)^{-1/\widehat{\gamma}}. \tag{7.4}$$

Expression (7.4) immediately leads to the following estimator of the quantile $q_p(F) = F^{\leftarrow}(p)$.

$$\widehat{q_p(F)} = \inf\{x \in \mathbb{R} : \widehat{\overline{F}(x)} \le 1 - p\}$$
$$= \inf\{u + x \in \mathbb{R} : \widehat{\overline{F}(u + x)} \le 1 - p\}$$
$$= u + \inf\left\{x \in \mathbb{R} : \frac{N_u}{n}\left(1 + \widehat{\gamma}\frac{x}{\widehat{\beta}}\right)^{-1/\widehat{\gamma}} \le 1 - p\right\}$$
$$= u + \frac{\widehat{\beta}}{\widehat{\gamma}}\left(\left(\frac{n}{N_u}(1 - p)\right)^{-\widehat{\gamma}} - 1\right). \tag{7.5}$$

The POT method for estimating tail probabilities and quantiles can be summarized in the following recipe. Each step will be discussed further below.

(i) Choose a high threshold u using some statistical method and count the number of exceedances N_u.

(ii) Given the sample $Y_1, \ldots, Y_{N_u}$ of excesses, estimate the parameters γ and β.

(ii) Combine steps (i) and (ii) to get estimates of the form (7.4) and (7.5).

The rest of this section will be devoted to step (i) and (ii): How do we choose a high threshold u in a suitable way? and How can one estimate the parameters γ and β?

7.1   How to choose a high threshold.

The choice of a suitable high threshold u is crucial but difficult. If we choose u too large then we will have few observations to use for parameter estimation resulting in poor estimates with large variance. If the threshold is too low then we have more data but on the other hand the approximation $\overline{F}_u(x) \approx \overline{G}_{\gamma,\beta(u)}(x)$

will be questionable. The main idea when choosing the threshold u is to look at the mean-excess plot (see below) and choose u such that the sample mean-excess function is approximately linear above u.

In the example with the Danish fire insurance claims it seems reasonable to choose u between 4.5 and 10. With $u = 4.5$ we get $N_{4.5} = 101$ exceedances which is about 5% of the data. With $u = 10$ we get $N_{10} = 24$ exceedances which is about 1.2% of the data. Given the shape of the mean-excess plot we have selected the threshold $u = 6$ since the shape of the mean-excess plot does not change much with $u \in [4.5, 10]$ we get a reasonable amount of data.

7.2 Mean-excess plot

If $\mathrm{E}(X) < \infty$, then we can define the *mean excess function* as

$$e(u) = \mathrm{E}(X - u \mid X > u) = \frac{\mathrm{E}((X - u)\mathbb{I}_{(u,\infty)}(X))}{\mathrm{E}(\mathbb{I}_{(u,\infty)}(X))}.$$

For a nonnegative random variable Z with distribution function F, integration by parts show that

$$\mathrm{E}(Z) = \int_0^\infty z \mathrm{d}F(z) = \int_0^\infty z \mathrm{d}(1 - \overline{F})(z) = -\int_0^\infty z \mathrm{d}\overline{F}(z)$$

$$= \left[-z\overline{F}(z) \right]_0^\infty + \int_0^\infty \overline{F}(z)\mathrm{d}z = \int_0^\infty \mathrm{P}(Z > z)\mathrm{d}z. \qquad (7.6)$$

We now study the mean excess function for a random variable X with a regularly varying right tail, $\mathrm{P}(X > x) = L(x)x^{-\alpha}$, for $\alpha > 1$.

$$e(u) = \frac{\mathrm{E}(X\mathbb{I}_{(u,\infty)(X)})}{\mathrm{P}(X > u)} - u$$

$$= \frac{1}{\mathrm{P}(X > u)} \left(\int_0^u \mathrm{P}(X\mathbb{I}_{(u,\infty)(X)} > z)\mathrm{d}z - \int_u^\infty \mathrm{P}(X\mathbb{I}_{(u,\infty)(X)} > z)\mathrm{d}z \right) - u$$

$$= \frac{1}{\mathrm{P}(X > u)} \int_u^\infty \mathrm{P}(X > z)\mathrm{d}z$$

$$= \frac{1}{L(u)u^{-\alpha}} \int_u^\infty L(z)z^{-\alpha}\mathrm{d}z$$

$$\sim \frac{1}{L(u)u^{-\alpha}} L(u) \int_u^\infty z^{-\alpha}\mathrm{d}z$$

$$= \frac{u}{\alpha - 1}$$

as $u \to \infty$, where in the second to last step we applied the Karamata theorem. Hence, the mean excess plot is approximately linear for large u.

40

Example 7.1 For the GPD $G_{\gamma,\beta}$, with $\gamma < 1$, integration by parts gives

$$
\mathrm{E}((X - u)\mathbb{I}_{(u,\infty)}(X)) = \left[-(x - u)(1 + \gamma x/\beta)^{-1/\gamma}\right]_u^\infty + \int_u^\infty (1 + \gamma x/\beta)^{-1/\gamma}\mathrm{d}x
$$

$$
= \int_u^\infty (1 + \gamma x/\beta)^{-1/\gamma}\mathrm{d}x
$$

$$
= \frac{\beta}{1 - \gamma}(1 + \gamma u/\beta)^{-1/\gamma+1}.
$$

Moreover, $\mathrm{E}(\mathbb{I}_{(u,\infty)}(X)) = \mathrm{P}(X > u) = (1 + \gamma u/\beta)^{-1/\gamma}$. Hence, $e(u) = (\beta + \gamma u)/(1 - \gamma)$. In particular, the mean excess function is linear for the GPD. $\blacksquare$

A graphical test for assessing the tail behavior may be performed by studying the *sample mean-excess function* based on the sample $X_1, \ldots, X_n$. With N_u being the number of exceedances of u by $X_1, \ldots, X_n$, as above, the sample mean-excess function is given by

$$
e_n(u) = \frac{1}{N_u} \sum_{k=1}^n (X_k - u)\mathbb{I}_{(u,\infty)}(X_k).
$$

The *mean-excess plot* is the plot of the points

$$
\{(X_{k,n}, e_n(X_{k,n})) : k = 2, \ldots, n\}.
$$

If the mean-excess plot is approximately linear with positive slope then X_1 may be assumed to have a heavy-tailed Pareto-like tail.

7.3 Parameter estimation

Given the threshold u we may estimate the parameters γ and β in the GPD based on the observations of excesses $Y_1, \ldots, Y_{N_u}$ over u. We assume that the excesses have distribution function $G_{\gamma,\beta}$ and hence the likelihood function becomes

$$
L(\gamma, \beta; Y_1, \ldots, Y_{N_u}) = \prod_{i=1}^{N_u} g_{\gamma,\beta}(Y_i), \quad g_{\gamma,\beta}(y) = \frac{1}{\beta}\left(1 + \gamma\frac{y}{\beta}\right)^{-1/\gamma-1}.
$$

Instead of maximizing the likelihood function we can maximize the log-likelihood function given by

$$
\ln L(\gamma, \beta; Y_1, \ldots, Y_{N_u}) = -N_u \ln \beta - \left(\frac{1}{\gamma} + 1\right)\sum_{i=1}^{N_u} \ln\left(1 + \frac{\gamma}{\beta}Y_i\right).
$$

Maximizing the log-likelihood numerically gives estimates $\widehat{\gamma}$ and $\widehat{\beta}$. The MLE is approximately normal (for large N_u)

$$
\left(\widehat{\gamma} - \gamma, \frac{\widehat{\beta}}{\beta} - 1\right) \approx N_2(0, \Sigma^{-1}/N_u), \quad \Sigma^{-1} = (1 + \gamma)\begin{pmatrix} 1 + \gamma & -1 \\ -1 & 2 \end{pmatrix}
$$
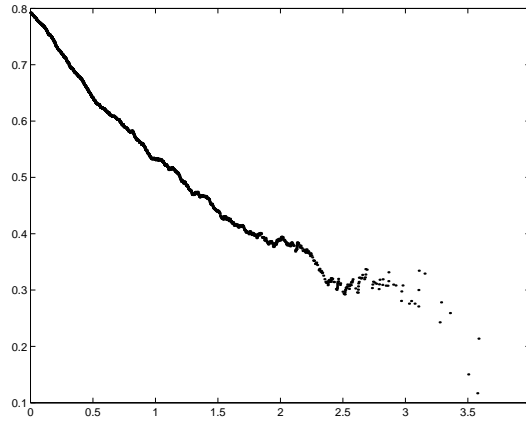
Figure 10: Mean-excess plot of data simulated from a Gaussian distribution.
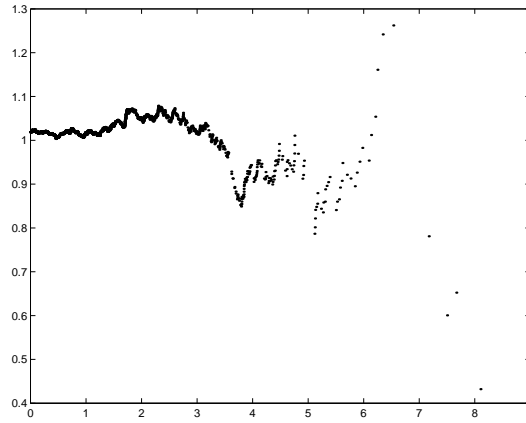
Figure 11: Mean-excess plot of data simulated from an exponential distribution.
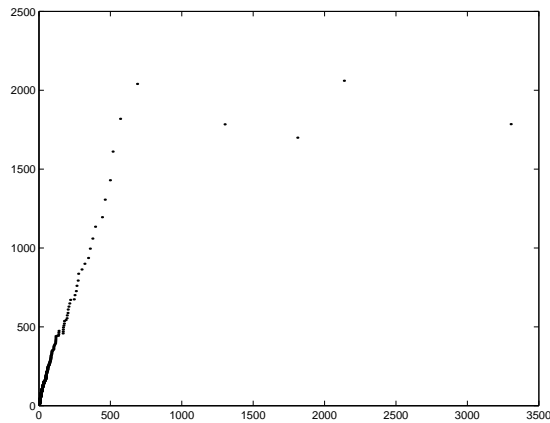
Figure 12: Mean-excess plot of data simulated from a Pareto(1) distribution.

Using the threshold $u = 6$ (gives $N_u = 56$) we obtain the following estimates for the Danish fire insurance data:

$$\widehat{\gamma} = 0.58, \ \widehat{\beta} = 3.60.$$

Having estimated the parameters we can now visually observe how the approximation (7.4) works, see Figure 14.
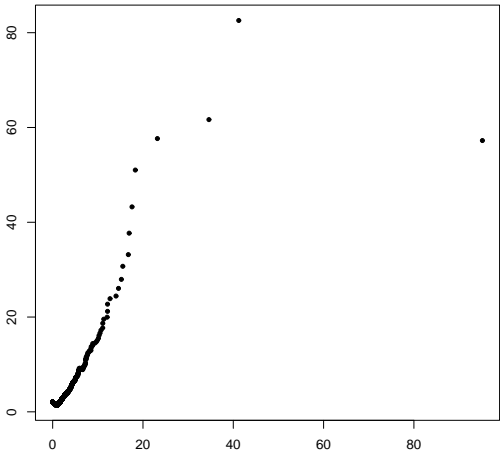
Figure 13: Mean-excess plot of the Danish fire insurance data. The plot looks approximately linear indicating Pareto-like tails.
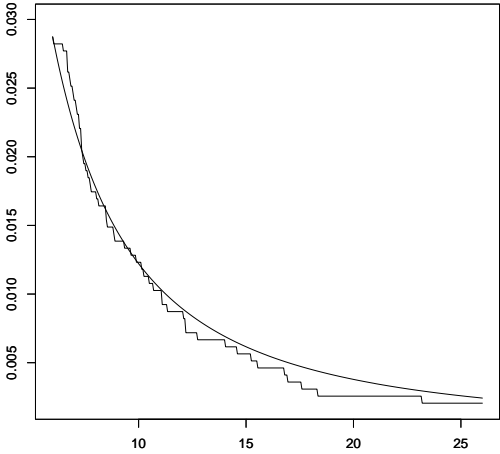
Figure 14: The empirical tail of the Danish data and the POT approximation.

7.4 Estimation of Value-at-Risk and Expected shortfall

Recall that Value-at-Risk at confidence level p for a risk X with distribution function F is, by definition, the quantile $q_p(F)$. Hence, the POT method gives

the Value-at-Risk estimator

$$\widehat{\mathrm{VaR}}_{p,\mathrm{POT}} = u + \frac{\widehat{\beta}}{\widehat{\gamma}} \left(\left(\frac{n}{N_u}(1-p) \right)^{-\widehat{\gamma}} - 1 \right).$$

Similarly, the POT method leads to an estimator for Expected shortfall. Recall that

$$\mathrm{ES}_p(X) = \mathrm{E}(X \mid X > \mathrm{VaR}_p(X)) = \frac{\mathrm{E}(X \mathbb{I}_{(q_p,\infty)}(X))}{\mathrm{E}(\mathbb{I}_{(q_p,\infty)}(X))},$$

where $q_p = \mathrm{VaR}_p(X))$ and we have $\mathrm{E}(\mathbb{I}_{(q_p,\infty)}(X)) = \overline{F}(q_p)$. If p is sufficiently large so that $q_p > u$, then (7.6) can be applied to the nonnegative random variable $X\mathbb{I}_{(q_p,\infty)}(X)$. This gives

$$\mathrm{E}(X\mathbb{I}_{(q_p,\infty)}(X)) = q_p\overline{F}(q_p) + \int_{q_p}^{\infty} \overline{F}(t)\mathrm{d}t$$

$$= q_p\overline{F}(q_p) + \int_{q_p}^{\infty} \overline{F}(u)\overline{F}_u(t-u)\mathrm{d}t.$$

Hence,

$$\mathrm{ES}_p(X) = q_p + \frac{\overline{F}(u)}{\overline{F}(q_p)} \int_{q_p}^{\infty} \overline{F}_u(t-u)\mathrm{d}t = q_p + \frac{\int_{q_p}^{\infty} \overline{F}_u(t-u)\mathrm{d}t}{\overline{F}_u(q_p-u)}.$$

We may now use the estimator $\widehat{\overline{F}_u}(t-u) = \overline{G}_{\widehat{\gamma},\widehat{\beta}}(t-u)$ to obtain, with $\widehat{q}_p = \widehat{\mathrm{VaR}}_{p,\mathrm{POT}}$,

$$\widehat{\mathrm{ES}}_{p,\mathrm{POT}} = \widehat{q}_p + \frac{\int_{\widehat{q}_p}^{\infty}(1+\widehat{\gamma}(t-u)/\widehat{\beta})^{-1/\widehat{\gamma}}\mathrm{d}t}{(1+\widehat{\gamma}(\widehat{q}_p-u)/\widehat{\beta})^{-1/\widehat{\gamma}}} = \widehat{q}_p + \frac{\widehat{\beta} + \widehat{\gamma}(\widehat{q}_p - u)}{1 - \widehat{\gamma}}.$$

More information about the POT method can be found in [8] and [12].

8 Multivariate distributions and dependence

We will now introduce some techniques and concepts for modeling multivariate random vectors. We aim at constructing useful models for the vector of risk factor changes $\mathbf{X}_n$. At this stage we will assume that all vectors of risk factor changes $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ are iid but we allow for dependence between the components of $\mathbf{X}_n$. That is, typically we assume that $X_{n,i}$ and $X_{n,j}$ are dependent whereas $X_{n,i}$ and $X_{n+k,j}$, are independent (for $k \neq 0$).

8.1 Basic properties of random vectors

The probability distribution of d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is completely determined by its joint distribution function F

$$F(\mathbf{x}) = F(x_1, \ldots, x_d) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \mathrm{P}(\mathbf{X} \leq \mathbf{x}).$$

The ith marginal distribution F_i of F is the distribution of X_i and is given by

$$F_i(x_i) = \mathrm{P}(X_i \leq x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty).$$

The distribution F is said to be absolutely continuous if there is a function $f \geq 0$, integrating to one, such that

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, u_d) \mathrm{d}u_1 \ldots \mathrm{d}u_d$$

and then f is called the density of F. The components of $\mathbf{X}$ are independent if and only if

$$F(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i)$$

or equivalently if and only if the joint density f (if the density exists) satisfies

$$f(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i).$$

Recall that the distribution of a random vector $\mathbf{X}$ is completely determined by its characteristic function given by

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}(\exp\{i\,\mathbf{t}^{\mathrm{T}}\mathbf{X}\}), \quad \mathbf{t} \in \mathbb{R}^d.$$

Example 8.1 The multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ has the density (with $|\Sigma|$ being the absolute value of the determinant of Σ)

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Its characteristic function is given by

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{i\,\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t}\right\}, \quad \mathbf{t} \in \mathbb{R}^d.$$

■

8.2 Joint log return distributions

Most classical multivariate models in financial mathematics assume that the joint distribution of log returns is a multivariate normal distribution. However, the main reason for this assumption is mathematical convenience. This assumption is not well supported by empirical findings. To illustrate this fact we study daily log returns of exchange rates. Pairwise log returns of the Swiss franc (chf), German mark (dem), British pound (gbp) and Japanese yen (jpy) quoted against the US dollar are illustrated in Figure 15. Now we ask whether a multivariate normal model would be suitable for the log return data. We estimate the mean and the covariance matrix of each data set (pairwise) and simulate from a bivariate normal distribution with this mean and covariance matrix, see Figure 16. By comparing the two figures we see that although the simulated data resembles the true observations, the simulated data have too few points far away from the mean. That is, the tails of the log return data are heavier than the simulated multivariate normal data.

Another example shows that not only does the multivariate normal distribution have too light tails but also the dependence structure in the normal distribution may be inappropriate when modeling the joint distribution of log returns. Consider for instance the data set consisting of log returns from BMW and Siemens stocks, Figure 17. Notice the strong dependence between large drops in the BMW and Siemens stock prices. The dependence of large drops seems stronger than for ordinary returns. This is something that cannot be modeled by a multivariate normal distribution. To find a good model for the BMW and Siemens data we need to be able to handle more advanced dependence structures than that offered by the multivariate normal distribution.

8.3 Comonotonicity and countermonotonicity

Let (X_1, X_2) be a bivariate random vector and suppose there exist two monotone functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ and a random variable Z such that

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} (\alpha(Z), \beta(Z)).$$

If both α and β are increasing, then X_1 and X_2 are said to be comonotonic. If the distribution functions F_1 and F_2 are continuous, then $X_2 = T(X_1)$ a.s. with $T = F_2^{\leftarrow} \circ F_1$.

If α is increasing and β is decreasing, then X_1 and X_2 are said to be countermonotonic. If the distribution functions F_1 and F_2 are continuous, then $X_2 = T(X_1)$ a.s. with $T = F_2^{\leftarrow} \circ (1 - F_1)$.

8.4 Covariance and linear correlation

Let $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ be a random (column) vector with $\mathrm{E}(X_k^2) < \infty$ for every k. The mean vector of $\mathbf{X}$ is $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X})$ and the covariance matrix of $\mathbf{X}$ is $\mathrm{Cov}(\mathbf{X}) = \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}]$. Here $\mathrm{Cov}(\mathbf{X})$ is a $d \times d$ matrix whose (i, j)-entry (ith row, jth column) is $\mathrm{Cov}(X_i, X_j)$. Notice that $\mathrm{Cov}(X_i, X_i) = \mathrm{var}(X_i)$, i.e. the variance of X_i.
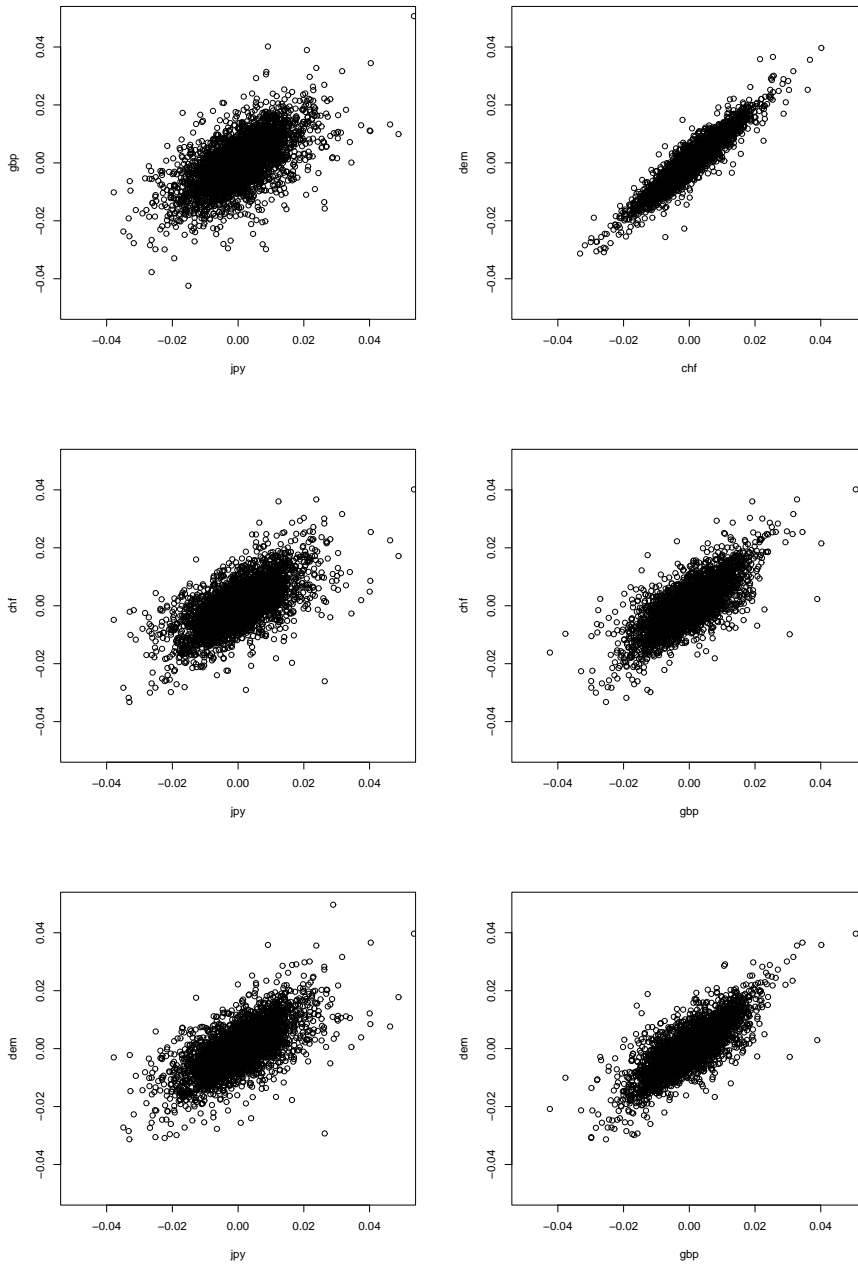
Figure 15: log returns of foreign exchange rates quotes against the US dollar.

Covariance is easily manipulated under linear (affine) transformations. If B is a constant $k \times d$ matrix and $\mathbf{b}$ is a constant k-vector (column vector), then $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$ has mean $\mathrm{E}(\mathbf{Y}) = B\boldsymbol{\mu} + \mathbf{b}$ and covariance matrix

$$\mathrm{Cov}(\mathbf{Y}) = \mathrm{E}[(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))^{\mathrm{T}}] = \mathrm{E}[B(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}B^{\mathrm{T}}]$$
$$= B[\mathrm{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}]B^{\mathrm{T}} = B\mathrm{Cov}(\mathbf{X})B^{\mathrm{T}}.$$
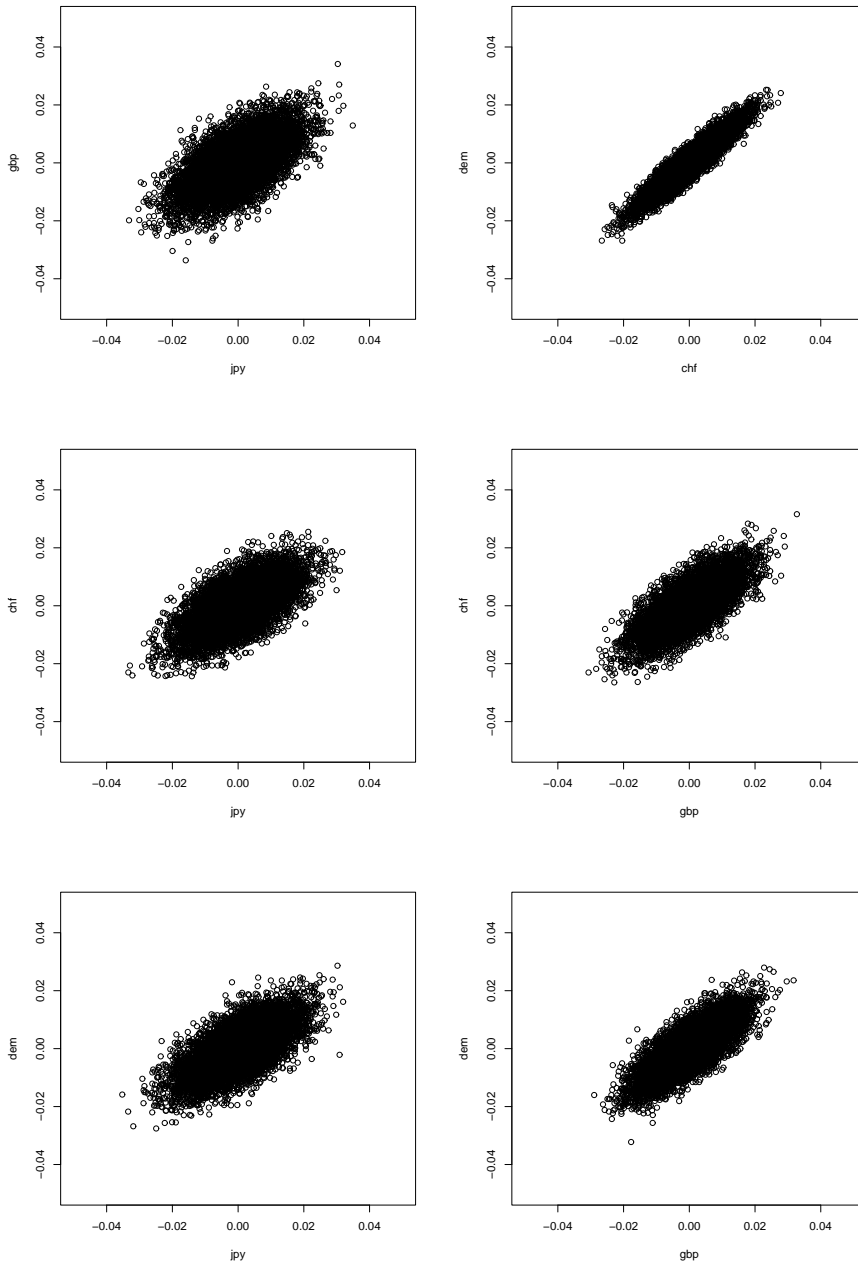
Figure 16: Simulated log returns of foreign exchange rates using bivariate normal distribution with estimated mean and covariance matrix.

If $\text{var}(X_1), \text{var}(X_2) \in (0, \infty)$, then the linear correlation coefficient $\varrho_L(X_1, X_2)$ is

$$\varrho_L(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)\,\text{var}(X_2)}}.$$

If X_1 and X_2 are independent, then $\varrho_L(X_1, X_2) = 0$ but the converse is *false*;

48
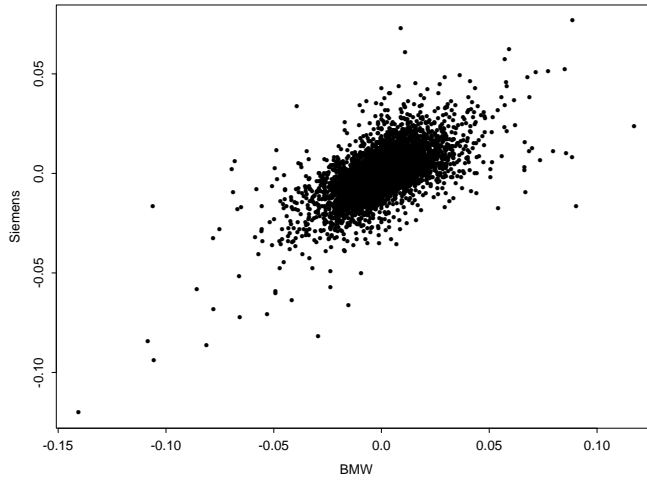
Figure 17: log returns from BMW and Siemens stocks.

$\varrho_L(X_1, X_2) = 0$ need not imply that X_1 and X_2 are independent.

Example 8.2 If $X_1 \sim \mathrm{N}(0,1)$ and $X_2 = X_1^2$, then

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}[(X_1 - 0)(X_2 - 1)] = \mathrm{E}(X_1 X_2) - 0 \cdot 1$$

$$= \mathrm{E}(X_1^3) = \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x = 0.$$

Hence, $\varrho_L(X_1, X_2) = 0$ but clearly X_1 and X_2 are strongly dependent. ∎

Moreover $|\varrho_L(X_1, X_2)| = 1$ if and only if X_1 and X_2 are perfectly linear dependent. That is, if and only if there exist $a \in \mathbb{R}$ and $b \neq 0$ such that $X_2 = a + bX_1$. The linear correlation coefficient is invariant under strictly increasing linear transformations. In particular, for $a_1, a_2 \in \mathbb{R}$ and $b_1, b_2 \neq 0$ we have

$$\varrho_L(a_1 + b_1 X_1, a_2 + b_2 X_2) = \mathrm{sign}(b_1 b_2) \varrho_L(X_1, X_2),$$

where $\mathrm{sign}(b_1 b_2) = b_1 b_2 / |b_1 b_2|$ if $b_1 b_2 \neq 0$ and 0 otherwise. However, linear correlation is not invariant under nonlinear strictly increasing transformations $T : \mathbb{R} \to \mathbb{R}$. That is, for two random variables we have in general

$$\varrho_L(T(X_1), T(X_2)) \neq \varrho_L(X_1, X_2).$$

This is a weakness of linear correlation as a measure of dependence. If we transform X_1 and X_2 by a strictly increasing transformation we have only rescaled the marginal distributions, we have not changed the dependence between X_1 and X_2. However, the linear correlation between $T(X_1)$ and $T(X_2)$ may have changed (falsely) indicating that the dependence has changed.

Proposition 8.1 *Let (X_1, X_2) be a random vector with marginal distribution functions F_1 and F_2, and an unspecified dependence structure. Assume further that $\mathrm{var}(X_1), \mathrm{var}(X_2) \in (0, \infty)$. Then*

(1) The set of possible linear *correlations is a closed interval $[\varrho_{L,\min}, \varrho_{L,\max}]$ with $0 \in [\varrho_{L,\min}, \varrho_{L,\max}]$.*

(2) The minimum linear correlation $\varrho_{L,\min}$ is obtained if and only if X_1 and X_2 are countermonotonic; the maximum $\varrho_{L,\max}$ if and only if X_1 and X_2 are comonotonic.

The following example illustrates Proposition 8.1.

Example 8.3 Let $X_1 \sim \mathrm{Lognormal}(0, 1)$ and $X_2 \sim \mathrm{Lognormal}(0, \sigma^2)$ with $\sigma > 0$. Let $Z \sim \mathrm{N}(0, 1)$ and note that

$$
\begin{aligned}
X_1 &\overset{\mathrm{d}}{=} e^Z, \\
X_2 &\overset{\mathrm{d}}{=} e^{\sigma Z} \overset{\mathrm{d}}{=} e^{-\sigma Z}.
\end{aligned}
$$

Note that e^Z and $e^{\sigma Z}$ are comonotonic and that e^Z and $e^{-\sigma Z}$ are countermonotonic. Hence, by Proposition 8.1,

$$
\begin{aligned}
\varrho_{L,\min} &= \varrho_L(e^Z, e^{-\sigma Z}) = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}}, \\
\varrho_{L,\max} &= \varrho_L(e^Z, e^{\sigma Z}) = \frac{e^{\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}}.
\end{aligned}
$$

In particular, $\varrho_{L,\min} \nearrow 0$ and $\varrho_{L,\max} \searrow 0$ as $\sigma \nearrow \infty$. See Figure 18 for a graphical illustration of these bounds as functions of σ. ∎

8.5 Rank correlation

Rank correlations are measures of concordance for bivariate random vectors. Given two points in $\mathbb{R}^2$, (x_1, x_2) and $(\widetilde{x}_1, \widetilde{x}_2)$, we say that the two points are concordant if $(x_1 - \widetilde{x}_1)(x_2 - \widetilde{x}_2) > 0$ and discordant if $(x_1 - \widetilde{x}_1)(x_2 - \widetilde{x}_2) < 0$. Hence, concordance (discordance) means that the line connecting the two points have a positive (negative) slope. Now consider two independent random vectors (X_1, X_2) and $(\widetilde{X}_1, \widetilde{X}_2)$ with the same bivariate distribution. The Kendall's tau rank correlation is given by

$$
\varrho_\tau(X_1, X_2) = \mathrm{P}\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\right) - \mathrm{P}\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0\right).
$$

If X_2 tend to increase with X_1 we expect the probability of concordance to be high relative to the probability of discordance, giving a high value of $\varrho_\tau(X_1, X_2)$.

Another measure of concordance/discordance is Spearman's rho rank correlation where one introduces a third independent copy of (X_1, X_2) denoted by
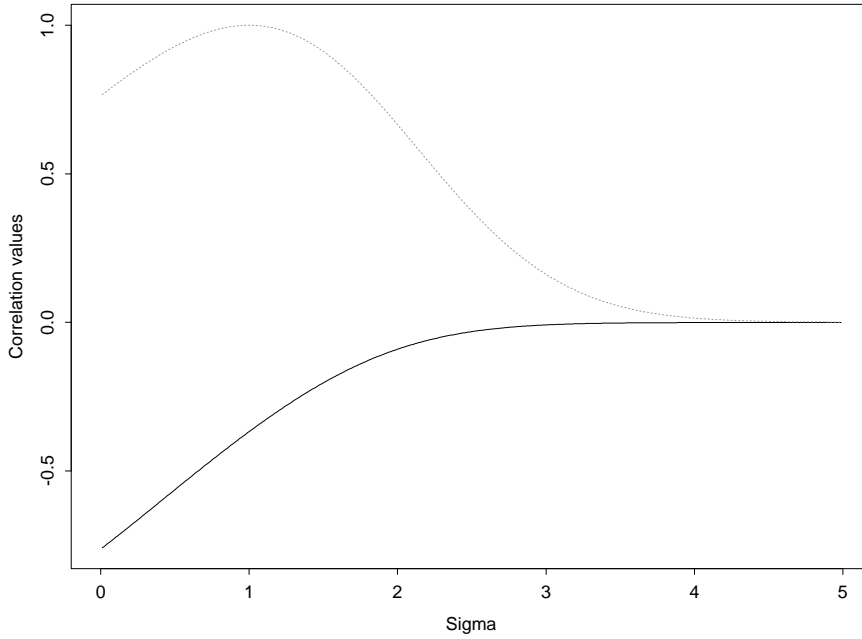
Figure 18: Bounds for the linear correlation coefficient.

$(\widehat{X}_1, \widehat{X}_2)$ and consider the concordance/discordance of the pairs (X_1, X_2) and $(\widetilde{X}_1, \widehat{X}_2)$. Spearman's rho is defined by

$$\varrho_S(X_1, X_2) = 3\Big\{ \mathrm{P}\left((X_1 - \widetilde{X}_1)(X_2 - \widehat{X}_2) > 0 \right) - \mathrm{P}\left((X_1 - \widetilde{X}_1)(X_2 - \widehat{X}_2) < 0 \right) \Big\}.$$

Kendall's tau and Spearman's rho have many properties in common listed below. They also illustrate important differences between Kendall's tau and Spearman's rho on the one hand and the linear correlation coefficient on the other hand.

- $\varrho_\tau(X_1, X_2) \in [-1, 1]$ and $\varrho_S(X_1, X_2) \in [-1, 1]$. All values can be obtained regardless of the marginal distribution functions, if they are continuous.

- If the marginal distribution functions are continuous, then $\varrho_\tau(X_1, X_2) = \varrho_S(X_1, X_2) = 1$ if and only if X_1 and X_2 are comonotonic; $\varrho_\tau(X_1, X_2) = \varrho_S(X_1, X_2) = -1$ if and only if X_1 and X_2 are countermonotonic.

- If X_1 and X_2 are independent then $\varrho_\tau(X_1, X_2) = 0$ and $\varrho_S(X_1, X_2) = 0$, but the converse is not true in general.

- If T_1, T_2 are strictly increasing then $\varrho_\tau(T_1(X_1), T_2(X_2)) = \varrho_\tau(X_1, X_2)$ and $\varrho_S(T_1(X_1), T_2(X_2)) = \varrho_S(X_1, X_2)$.

Estimation of Kendall's tau based on iid bivariate random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$

is easy. Note that

$$\varrho_\tau(X_{k,1}, X_{k,2}) = P((X_{k,1} - X_{l,1})(X_{k,2} - X_{l,2}) > 0)$$
$$- P((X_{k,1} - X_{l,1})(X_{k,2} - X_{l,2}) < 0)$$
$$= E(\text{sign}((X_{k,1} - X_{l,1})(X_{k,2} - X_{l,2}))).$$

Hence, we obtain the following estimator of Kendall's tau $\varrho_\tau = \varrho_\tau(X_{k,1}, X_{k,2})$

$$\widehat{\varrho_\tau} = \binom{n}{2}^{-1} \sum_{1 \le k < l \le n} \text{sign}((X_{k,1} - X_{l,1})(X_{k,2} - X_{l,2}))$$
$$= \binom{n}{2}^{-1} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \text{sign}((X_{k,1} - X_{l,1})(X_{k,2} - X_{l,2})).$$

We will return to the rank correlations when we discuss copulas in Section 10.

8.6 Tail dependence

Motivated by the scatter plot of joint BMW and Siemens log returns above we introduce a notion of dependence of extreme values, called tail dependence.

Let (X_1, X_2) be a random vector with marginal distribution functions F_1 and F_2. The *coefficient of upper tail dependence* of (X_1, X_2) is defined as

$$\lambda_U(X_1, X_2) = \lim_{u \nearrow 1} P(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)),$$

provided that the limit $\lambda_U \in [0, 1]$ exists. The *coefficient of lower tail dependence* is defined as

$$\lambda_L(X_1, X_2) = \lim_{u \searrow 0} P(X_2 \le F_2^{\leftarrow}(u) \mid X_1 \le F_1^{\leftarrow}(u)),$$

provided that the limit $\lambda_L \in [0, 1]$ exists. If $\lambda_U > 0$ ($\lambda_L > 0$), then we say that (X_1, X_2) has upper (lower) tail dependence.

See Figure 19 for an illustration of tail dependence.

Figure 19: Illustration of lower tail dependence.

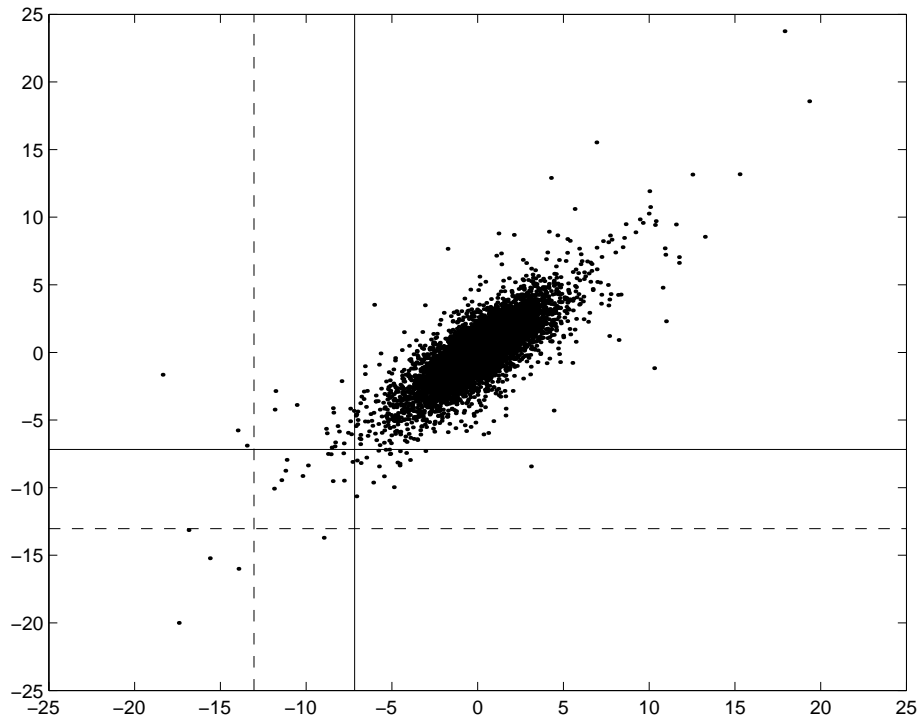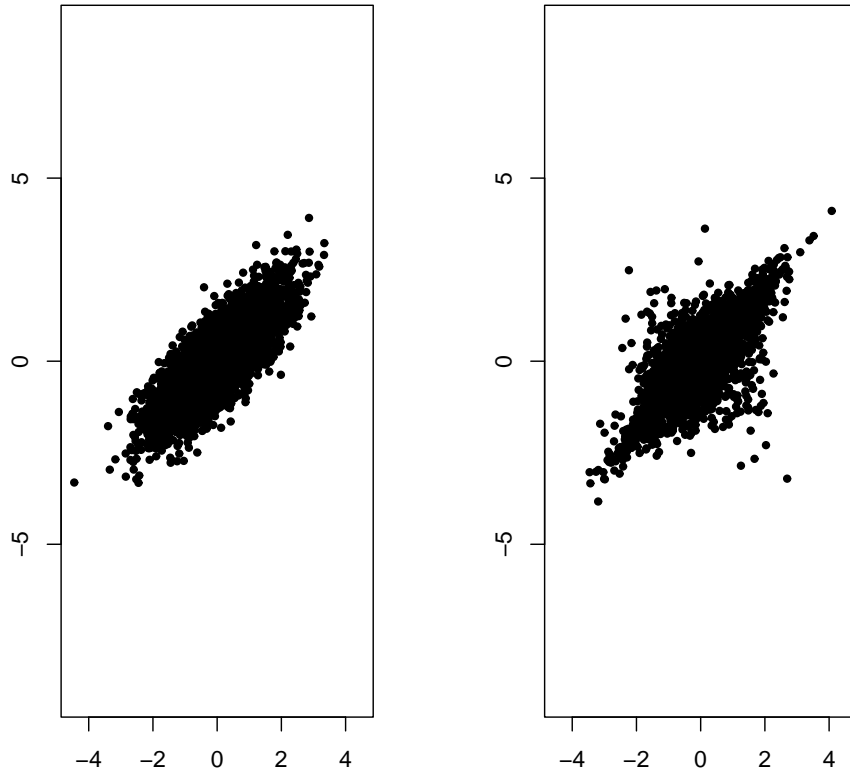Figure 20: Illustration of two bivariate distributions with linear correlation $\varrho_L = 0.8$ and standard normal marginal distributions. The one to the left has tail dependence $\lambda_U = \lambda_L = 0.62$ and the one to the right has $\lambda_U = \lambda_L = 0$.

9 Multivariate elliptical distributions

9.1 The multivariate normal distribution

Recall the following three equivalent definitions of a multivariate normal distribution.

(1) A random vector $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ has a normal distribution if for every vector $\mathbf{a} = (a_1, \ldots, a_d)^{\mathrm{T}}$ the random variable $\mathbf{a}^{\mathrm{T}}\mathbf{X}$ has a normal distribution.

The notation $\mathbf{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ is used to denote that $\mathbf{X}$ has a d-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

(2) $\mathbf{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ if and only if its characteristic function is given by

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}(\exp\{i\mathbf{t}^{\mathrm{T}}\mathbf{X}\}) = \exp\left\{i\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t}\right\}.$$

(3) A random vector $\mathbf{X}$ with $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{X}) = \Sigma$, such that $|\Sigma| > 0$, satisfies $\mathbf{X} \sim \mathrm{N}_d(\mu, \Sigma)$ if and only if it has the density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right\}$$

Next we list some useful properties of the multivariate normal distribution. Let $\mathbf{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.

- *Linear transformations.* For $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$ we have

$$B\mathbf{X} + \mathbf{b} \sim \mathrm{N}_k(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B^{\mathrm{T}}).$$

- *Marginal distributions.* Write $\mathbf{X}^{\mathrm{T}} = (\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})$ with $\mathbf{X}_1 = (X_1, \ldots, X_k)^{\mathrm{T}}$, $\mathbf{X}_2 = (X_{k+1}, \ldots, X_d)^{\mathrm{T}}$ and write

$$\boldsymbol{\mu}^{\mathrm{T}} = (\boldsymbol{\mu}_1^{\mathrm{T}}, \boldsymbol{\mu}_2^{\mathrm{T}}), \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then $\mathbf{X}_1 \sim \mathrm{N}_k(\boldsymbol{\mu}_1, \Sigma_{11})$ and $\mathbf{X}_2 \sim \mathrm{N}_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22})$.

- *Conditional distributions.* If Σ is nonsingular ($|\Sigma| > 0$), then $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \sim \mathrm{N}_{d-k}(\boldsymbol{\mu}_{2,1}, \Sigma_{22,1})$, where

$$\boldsymbol{\mu}_{2,1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \text{ and } \Sigma_{22,1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

- *Quadratic forms.* If Σ is nonsingular, then

$$D^2 = (\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2.$$

The variable D is called the Mahalanobis distance.

- *Convolutions.* If $\mathbf{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{Y} \sim \mathrm{N}_d(\widetilde{\boldsymbol{\mu}}, \widetilde{\Sigma})$ are independent, then $\mathbf{X} + \mathbf{Y} \sim \mathrm{N}_d(\boldsymbol{\mu} + \widetilde{\boldsymbol{\mu}}, \Sigma + \widetilde{\Sigma})$.

9.2 Normal mixtures

Note that if $\mathbf{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, then $\mathbf{X}$ has the representation $\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A\mathbf{Z}$, where $\mathbf{Z} \sim \mathrm{N}_k(\mathbf{0}, \mathbf{I})$ ($\mathbf{I}$ denotes the identity matrix) and $A \in \mathbb{R}^{d \times k}$ is such that $AA^{\mathrm{T}} = \Sigma$. If Σ is nonsingular, then we can take $k = d$.

Definition 9.1 *An $\mathbb{R}^d$-valued random vector $\mathbf{X}$ is said to have a multivariate normal variance mixture distribution if $\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + WA\mathbf{Z}$, where $\mathbf{Z} \sim \mathrm{N}_k(\mathbf{0}, \mathbf{I})$, $W \geq 0$ is a positive random variable independent of $\mathbf{Z}$, and $A \in \mathbb{R}^{d \times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and a vector of constants respectively.* ∎

Note that conditioning on $W = w$ we have that $\mathbf{X}|W = w \sim \mathrm{N}_d(\boldsymbol{\mu}, w^2\Sigma)$, where $\Sigma = AA^{\mathrm{T}}$.

Example 9.1 If we take $W^2 \overset{\mathrm{d}}{=} \nu/S$, where $S \sim \chi_\nu^2$ (Chi-square distribution with ν degrees of freedom), then $\mathbf{X}$ has the multivariate t-distribution with ν degrees of freedom. We use the notation $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$. Note that Σ is not the covariance matrix of $\mathbf{X}$. Since $E(W^2) = \nu/(\nu - 2)$ (if $\nu > 2$) we have $\mathrm{Cov}(\mathbf{X}) = [\nu/(\nu - 2)]\Sigma$. ∎

9.3 Spherical distributions

Many results on spherical (and elliptical) distributions can be found in the book [9].

Definition 9.2 *A random vector $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ has a spherical distribution if there exists a function ψ of a scalar variable such that the characteristic function of $\mathbf{X}$ satisfies $\phi_{\mathbf{X}}(\mathbf{t}) = \psi(\mathbf{t}^{\mathrm{T}}\mathbf{t}) = \psi(t_1^2 + \cdots + t_d^2)$.* ∎

We write $\mathbf{X} \sim S_d(\psi)$ to denote the $\mathbf{X}$ has a spherical distribution with characteristic function $\psi(\mathbf{t}^{\mathrm{T}}\mathbf{t})$.

Proposition 9.1 *The following statements are equivalent.*

(1) $\mathbf{X}$ has a spherical distribution.

(2) For every vector $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{a}^{\mathrm{T}}\mathbf{X} \overset{\mathrm{d}}{=} \|\mathbf{a}\|X_1$ with $\|\mathbf{a}\|^2 = a_1^2 + \ldots a_d^2$.

(3) $\mathbf{X}$ has the stochastic representation $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$, where $\mathbf{S}$ is uniformly distributed on the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ and $R \geq 0$ is independent of $\mathbf{S}$.

The implication $(1) \Rightarrow (2)$ can be shown as follows. Recall that two random variables (and vectors) have the same distribution if and only if their characteristic functions are the same. Let $\mathbf{X} \sim S_d(\psi)$. Then

$$\phi_{\mathbf{a}^{\mathrm{T}}\mathbf{X}}(s) = \mathrm{E}(\exp\{is\mathbf{a}^{\mathrm{T}}\mathbf{X}\}) = \mathrm{E}(\exp\{i(s\mathbf{a})^{\mathrm{T}}\mathbf{X}\}) = \psi(s^2\mathbf{a}^{\mathrm{T}}\mathbf{a}).$$

Note that $\|\mathbf{a}\|X_1 = \mathbf{t}^{\mathrm{T}}\mathbf{X}$ with $\mathbf{t}^{\mathrm{T}} = (\|\mathbf{a}\|, 0, \ldots, 0)$. Hence,

$$\phi_{\|\mathbf{a}\|X_1}(s) = \mathrm{E}(\exp\{is\mathbf{t}^{\mathrm{T}}\mathbf{X}\}) = \mathrm{E}(\exp\{i(s\mathbf{t})^{\mathrm{T}}\mathbf{X}\}) = \psi(s^2\mathbf{t}^{\mathrm{T}}\mathbf{t})$$
$$= \psi(s^2\|\mathbf{a}\|^2) = \psi(s^2\mathbf{a}^{\mathrm{T}}\mathbf{a}).$$

Example 9.2 Let $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I})$. Then $\mathbf{X} \sim S_d(\psi)$ with $\psi(x) = \exp\{-x/2\}$. The characteristic function of the standard normal distribution is

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{i\mathbf{t}^{\mathrm{T}}\mathbf{0} - \frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{I}\mathbf{t}\right\} = \exp\{-\mathbf{t}^{\mathrm{T}}\mathbf{t}/2\} = \psi(\mathbf{t}^{\mathrm{T}}\mathbf{t}).$$

From the stochastic representation $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$ we conclude that $\|\mathbf{X}\|^2 \overset{\mathrm{d}}{=} R^2$ and since the sum of squares of d independent standard normal random variables has a chi-squared distribution with d degrees of freedom we have $R^2 \sim \chi_d^2$. ∎

The stochastic representation in (4) is very useful for simulating from spherical distributions. Simply apply the following algorithm. Suppose the spherically distributed random vector $\mathbf{X}$ has the stochastic representation $\mathbf{X} \overset{\mathrm{d}}{=} R\mathbf{S}$.

(i) Simulate $\mathbf{s}$ from the distribution of $\mathbf{S}$.

(ii) Simulate r from the distribution of R.

(iii) Put $\mathbf{x} = r\mathbf{s}$.

This procedure can then be repeated n times to get a sample of size n from the spherical distribution of $\mathbf{X}$. We illustrate this in Figure 21. Note that a convenient way to simulate a random element $\mathbf{S}$ that has uniform distribution on the unit sphere in $\mathbb{R}^d$ is to simulate a d-dimensional random vector $\mathbf{Y} \sim N_d(\mathbf{0}, \mathbf{I})$ and then put $\mathbf{S} = \mathbf{Y}/\|\mathbf{Y}\|$.

9.4 Elliptical distributions

Definition 9.3 *An $\mathbb{R}^d$-valued random vector $\mathbf{X}$ has an* elliptical distribution *if $\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A\mathbf{Y}$, where $\mathbf{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{d \times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$.* ∎

When $d = 1$ the elliptical distributions coincide with the 1-dimensional symmetric distributions. The characteristic function of an elliptically distributed random vector $\mathbf{X}$ can be written as

$$\begin{aligned}
\phi_{\mathbf{X}}(\mathbf{t}) &= \mathrm{E}(\exp\{i\mathbf{t}^{\mathrm{T}}\mathbf{X}\}) \\
&= \mathrm{E}(\exp\{i\mathbf{t}^{\mathrm{T}}(\boldsymbol{\mu} + A\mathbf{Y})\}) \\
&= \exp\{i\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu}\}\,\mathrm{E}(\exp\{i(A^{\mathrm{T}}\mathbf{t})^{\mathrm{T}}\mathbf{Y}\}) \\
&= \exp\{i\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu}\}\psi(\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t}),
\end{aligned}$$

where $\Sigma = AA^{\mathrm{T}}$. We write $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$. Here $\boldsymbol{\mu}$ is called the location parameter, Σ is called the dispersion matrix and ψ is called the characteristic generator of the elliptical distribution. If $\mathrm{E}(X_k) < \infty$, then $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu}$. If $\mathrm{E}(X_k) < \infty$, then $\mathrm{Cov}(\mathbf{X}) = c\Sigma$ for some $c > 0$. Note that elliptically distributed random vectors are radially symmetric: if $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, then $\mathbf{X} - \boldsymbol{\mu} \overset{\mathrm{d}}{=} \boldsymbol{\mu} - \mathbf{X}$. If $A \in \mathbb{R}^{d \times d}$ is nonsingular with $AA^{\mathrm{T}} = \Sigma$, then we have the following relation between elliptical and spherical distributions:

$$\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi) \iff A^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim S_d(\psi), \quad A \in \mathbb{R}^{d \times d}, AA^{\mathrm{T}} = \Sigma.$$

Figure 21: Simulation from a spherical distribution using the stochastic representation. First we simulate independently n times from the uniform distribution on the unit sphere to obtain $\mathbf{s}_1, \ldots, \mathbf{s}_n$ (above). Then, we simulate $r_1, \ldots, r_n$ from the distribution of R. Finally we put $\mathbf{x}_k = r_k \mathbf{s}_k$ for $k = 1, \ldots, n$ (below).

It follows immediately from the definition that elliptical distributed random vectors have the following stochastic representation. $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ if and only if there exist $\mathbf{S}$, R, and A such that $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + RA\mathbf{S}$ with $\mathbf{S}$ uniformly distributed on the unit sphere, $R \geq 0$ a random variable independent of $\mathbf{S}$, $A \in \mathbb{R}^{d \times k}$ a matrix with $AA^{\mathrm{T}} = \Sigma$ and $\boldsymbol{\mu} \in \mathbb{R}^d$. The stochastic representation is useful when simulating from elliptical distributions. Suppose $\mathbf{X}$ has the stochastic representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + RA\mathbf{S}$.

  (i) Simulate $\mathbf{s}$ from the distribution of $\mathbf{S}$.

 (ii) Multiply $\mathbf{s}$ by the matrix A to get $A\mathbf{s}$.

(iii) Simulate r from the distribution of R and form $rA\mathbf{s}$.

(iv) Put $\mathbf{x} = \boldsymbol{\mu} + rA\mathbf{s}$.

This procedure can then be repeated n times to get a sample of size n from the elliptical distribution of $\mathbf{X}$. We illustrate this in Figure 22.

Example 9.3 An $\mathbb{R}^d$-valued normally distributed random vector with mean $\boldsymbol{\mu}$ and covariance matrix Σ has the representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\mathbf{Z}$, where $AA^{\mathrm{T}} = \Sigma$
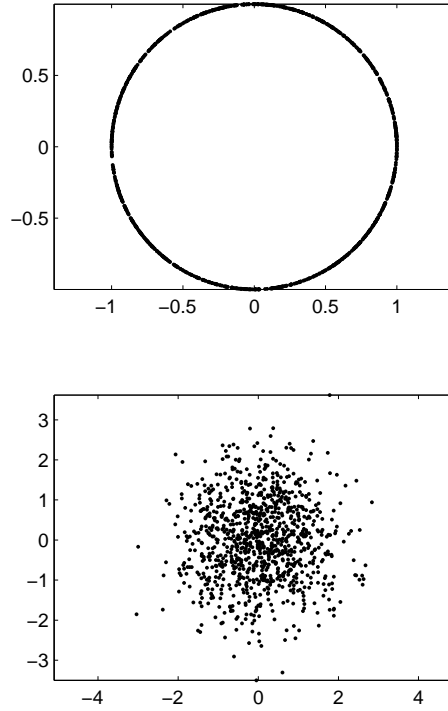
58
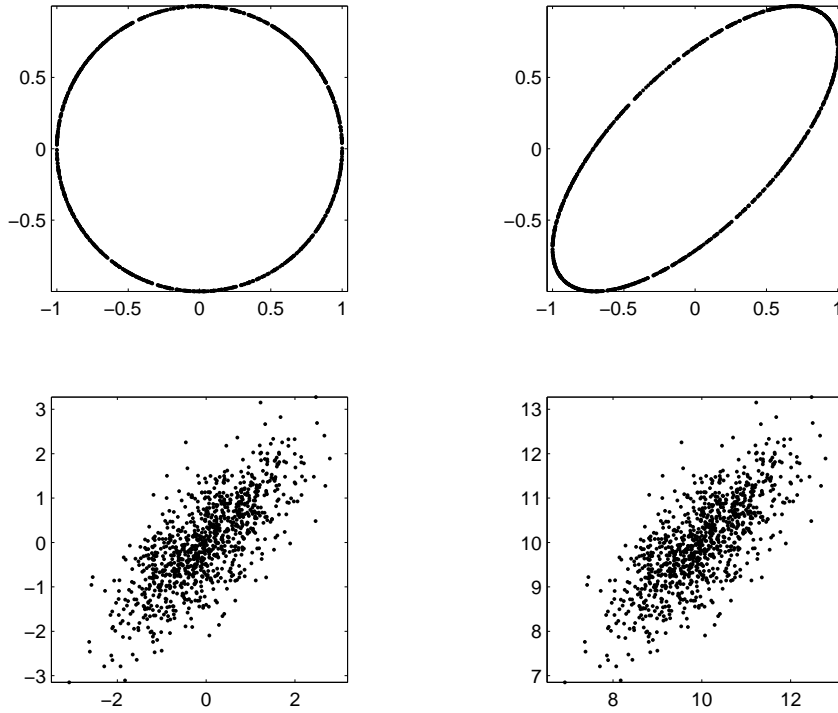
Figure 22: Simulation from a spherical distribution using the stochastic representation. First we simulate independently n times from the uniform distribution on the unit sphere to obtain $\mathbf{s}_1, \ldots, \mathbf{s}_n$ (above, left). Then, we multiply each sample by A to get the points $A\mathbf{s}_1, \ldots, A\mathbf{s}_n$ (above, right). Next, we simulate $r_1, \ldots, r_n$ from the distribution of R to obtain $r_k A\mathbf{s}_k$ for $k = 1, \ldots, n$ (below, left). Finally we add $\boldsymbol{\mu}$ to obtain $\mathbf{x}_k = \boldsymbol{\mu} + r_k A\mathbf{s}_k$ for $k = 1, \ldots, n$ (below, right).

and $\mathbf{Z} \sim \mathrm{N}_k(\mathbf{0}, \mathbf{I})$. Since $\mathbf{Z}$ has a spherical distribution it has representation $\mathbf{Z} \stackrel{\mathrm{d}}{=} R\mathbf{S}$, where $R^2 \sim \chi_k^2$. Hence, $\mathbf{X}$ has the representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + RA\mathbf{S}$ and we conclude that $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with $\psi(x) = \exp\{-x/2\}$. ∎

Example 9.4 A random vector $\mathbf{Z} \sim \mathrm{N}_d(\mathbf{0}, \mathbf{I})$ has a spherical distribution with stochastic representation $\mathbf{Z} \stackrel{\mathrm{d}}{=} V\mathbf{S}$. If $\mathbf{X}$ is a normal variance mixture, then we see from the definition that it has representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + VWA\mathbf{S}$, with $V^2 \sim \chi_d^2$. Hence, it has an elliptical distribution with $R \stackrel{\mathrm{d}}{=} VW$. An example given earlier is the multivariate t_ν-distribution where $W^2 \stackrel{\mathrm{d}}{=} \nu/S$ with $S \sim \chi_\nu^2$. This means that for the multivariate t-distribution $R^2/d \stackrel{\mathrm{d}}{=} V^2W^2/d$ has an $F(d, \nu)$-distribution (see e.g. Problem 10, Chapter 1 in [11]). ∎

9.5 Properties of elliptical distributions

Next we list some useful properties of elliptical distributions. Many of them coincide with the properties of the multivariate normal distribution. This is perhaps the most important argument for using the elliptical distributions in applications. Let $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$.

- *Linear transformations.* For $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$ we have

$$B\mathbf{X} + \mathbf{b} \sim E_k(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B^{\mathrm{T}}, \psi).$$

- *Marginal distributions.* Write $\mathbf{X}^{\mathrm{T}} = (\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})$ with $\mathbf{X}_1 = (X_1, \ldots, X_k)^{\mathrm{T}}$, $\mathbf{X}_2 = (X_{k+1}, \ldots, X_d)^{\mathrm{T}}$ and write

$$\boldsymbol{\mu}^{\mathrm{T}} = (\boldsymbol{\mu}_1^{\mathrm{T}}, \boldsymbol{\mu}_2^{\mathrm{T}}), \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

  Then $\mathbf{X}_1 \sim E_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and $\mathbf{X}_2 \sim E_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$.

- *Conditional distributions.* Assuming that Σ is nonsingular, $\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim E_{d-k}(\boldsymbol{\mu}_{2,1}, \Sigma_{22,1}, \widetilde{\psi})$, where

$$\boldsymbol{\mu}_{2,1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \text{ and } \Sigma_{22,1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

  Typically $\widetilde{\psi}$ is a different characteristic generator than the original ψ.

- *Quadratic forms.* If Σ is nonsingular, then

$$D^2 = (\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \stackrel{\mathrm{d}}{=} R^2.$$

  The variable D is called the Mahalanobis distance.

- *Convolutions.* If $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\mathbf{Y} \sim E_d(\widetilde{\boldsymbol{\mu}}, \Sigma, \widetilde{\psi})$ are independent, then $\mathbf{X} + \mathbf{Y} \sim E_d(\boldsymbol{\mu} + \widetilde{\boldsymbol{\mu}}, \Sigma, \overline{\psi})$, with $\overline{\psi}(x) = \psi(x)\widetilde{\psi}(x)$. Note that the dispersion matrix Σ must (in general) be the same for $\mathbf{X}$ and $\mathbf{Y}$.

IMPORTANT: Contrary to the multivariate normal distribution it is not true that the components of a spherically distributed random vector $\mathbf{X} \sim E_d(\mathbf{0}, \mathbf{I}, \psi)$ are independent. In fact, the components of $\mathbf{X}$ are independent only if $\mathbf{X}$ has a multivariate normal distribution. For instance, assume $\mathbf{X} = (X_1, X_2)^{\mathrm{T}} \sim \mathrm{N}_2(\boldsymbol{\mu}, \mathbf{I})$. Then the linear correlation coefficient $\varrho_L(X_1, X_2) = 0$ and X_1 and X_2 are independent. However, if $\mathbf{X} = (X_1, X_2)^{\mathrm{T}} \sim E_2(\boldsymbol{\mu}, \mathbf{I}, \psi)$ is not normal, then $\varrho_L(X_1, X_2) = 0$ (if $\varrho_L(X_1, X_2)$ exists) but X_1 and X_2 are dependent.

9.6   Elliptical distributions and risk management

Suppose we have the possibility today at time 0 to invest in d risky assets by taking long or short positions. We consider a fixed holding period of length T and let $\mathbf{X} = (X_1, \ldots, X_d)$ be the future asset returns at time T. Suppose that $\mathbf{X}$ has a (finite) nonsingular covariance matrix Σ and mean vector $\mathrm{E}(\mathbf{X})$. Let $\mathcal{P}$ be the set of all linear portfolios $\mathbf{w}^{\mathrm{T}}\mathbf{X}$ for $\mathbf{w} \in \mathbb{R}^d$ and let

$$\mathcal{W}_r = \{\mathbf{w} \in \mathbb{R}^d : \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} = r, \sum_{i=1}^{d} w_i = 1\} \quad \text{and} \quad \mathcal{P}_r = \{Z = \mathbf{w}^{\mathrm{T}}\mathbf{X} : \mathbf{w} \in \mathcal{W}_r\}.$$

Hence, $\mathcal{P}_r$ is the set of of all linear portfolios with expected return r with the normalization constraint $\sum_{i=1}^d w_i = 1$.

How can we find the least risky portfolio with expected portfolio return r? The mean-variance (Markowitz) approach to portfolio optimization solves this problem if we accept variance (or standard deviation) as our measure of risk. The minimum variance portfolio is the portfolio that minimizes the variance $\operatorname{var}(Z) = \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}$ over all $Z \in \mathcal{P}_r$. Hence, the portfolio weights of the minimum-variance portfolio is the solution to the following optimization problem:

$$\min_{\mathbf{w}\in\mathcal{W}_r} \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}.$$

This optimization problem has a unique solution, see e.g. p. 185 in [3]. As already mentioned, the variance is in general not a suitable risk measure. We would typically prefer a risk measure based on the appropriate tail of our portfolio return. It was shown in [7] that the "minimum-risk" portfolio, where risk could be e.g. VaR or ES, coincides with the minimum-variance portfolio for elliptically distributed return vectors $\mathbf{X}$. Hence, the mean-variance approach (Markowitz) to portfolio optimization makes sense in the elliptical world. In fact we can take any risk measure ϱ with the following properties and replace the variance in the Markowitz approach by ϱ. Suppose ϱ is a risk measure that is translation invariant and positively homogeneous, i.e.

$$\varrho(X + a) = \varrho(X) + a \text{ for } r \in \mathbb{R} \text{ and } \varrho(\lambda X) = \lambda\varrho(X) \text{ for } \lambda \geq 0.$$

Moreover, suppose that $\varrho(X)$ depends on X only through its distribution. The risk measure ϱ could be for instance VaR or ES.

Since $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ we have $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\mathbf{Y}$, where $AA^{\mathrm{T}} = \Sigma$ and $\mathbf{Y} \sim S_d(\psi)$. Hence, it follows from Proposition 9.1 that

$$\mathbf{w}^{\mathrm{T}}\mathbf{X} = \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \mathbf{w}^{\mathrm{T}}A\mathbf{Y} = \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + (A^{\mathrm{T}}\mathbf{w})^{\mathrm{T}}\mathbf{Y} \stackrel{\mathrm{d}}{=} \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \|A^{\mathrm{T}}\mathbf{w}\|Y_1. \qquad (9.1)$$

Hence, for any $Z \in \mathcal{P}_r$ it holds that

$$\varrho(Z) = r + \|A^{\mathrm{T}}\mathbf{w}\|\varrho(Y_1) \quad \text{and} \quad \operatorname{var}(Z) = \|A^{\mathrm{T}}\mathbf{w}\|^2 \operatorname{var}(Y_1) = \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}\operatorname{var}(Y_1).$$

Hence, minimization with respect to ϱ and minimization with respect to the variance give the same optimal portfolio weights $\mathbf{w}$:

$$\operatorname{argmin}_{Z\in\mathcal{P}_r}\varrho(Z) = \operatorname{argmin}_{\mathbf{w}\in\mathcal{W}_r}\|A^{\mathrm{T}}\mathbf{w}\|$$
$$= \operatorname{argmin}_{\mathbf{w}\in\mathcal{W}_r}\|A^{\mathrm{T}}\mathbf{w}\|^2 = \operatorname{argmin}_{Z\in\mathcal{P}_r} \operatorname{var}(Z).$$

Example 9.5 Note also that (9.1) implies that for $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^d$ we have

$$\operatorname{VaR}_p(\mathbf{w}_1\mathbf{X} + \mathbf{w}_2\mathbf{X}) = \mathbf{w}_1^{\mathrm{T}}\boldsymbol{\mu} + \mathbf{w}_2^{\mathrm{T}}\boldsymbol{\mu} + \|A^{\mathrm{T}}\mathbf{w}_1 + A^{\mathrm{T}}\mathbf{w}_2\|\operatorname{VaR}_p(Y_1)$$
$$\leq \mathbf{w}_1^{\mathrm{T}}\boldsymbol{\mu} + \mathbf{w}_2^{\mathrm{T}}\boldsymbol{\mu} + \|A^{\mathrm{T}}\mathbf{w}_1\|\operatorname{VaR}_p(Y_1) + \|A^{\mathrm{T}}\mathbf{w}_2\|\operatorname{VaR}_p(Y_1)$$
$$= \operatorname{VaR}_p(\mathbf{w}_1^{\mathrm{T}}\mathbf{X}) + \operatorname{VaR}_p(\mathbf{w}_2^{\mathrm{T}}\mathbf{X}).$$

Hence, Value-at-Risk is subadditive for elliptical distributions. $\blacksquare$

Example 9.6 Let $\mathbf{X} = (X_1, X_2)^T$ be a vector of log returns, for the time period today-until-tomorrow, of two stocks with stock prices $S_1 = S_2 = 1$ today. Suppose that $\mathbf{X} \sim E_2(\boldsymbol{\mu}, \Sigma, \psi)$ (elliptically distributed) with linear correlation coefficient ρ and that

$$\boldsymbol{\mu} = \mathbf{0}, \quad \Sigma = \begin{pmatrix} \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 \end{pmatrix}.$$

Your total capital is 1 which you want to invest fully in the two stocks giving you a linearized portfolio loss $L^\Delta = L^\Delta(w_1, w_2)$ where w_1 and w_2 are portfolio weights. Two investment strategies are available (long positions):
(A) invest your money in equal shares in the two stocks: $w_{A1} = w_{A2} = 1/2$;
(B) invest all your money in the first stock: $w_{B1} = 1, w_{B2} = 0$.
How can we compute the ratio $\mathrm{VaR}_{0.99}(L_A^\Delta)/\mathrm{VaR}_{0.99}(L_B^\Delta)$, where L_A^Δ and L_B^Δ are linearized losses for investment strategies A and B, respectively?
We have

$$L^\Delta = -\mathbf{w}^T \mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{w}^T \mathbf{X}$$

since $\mathbf{X} \sim E_2(\mathbf{0}, \Sigma, \psi)$. Moreover, by (9.1) it holds that

$$\mathbf{w}^T \mathbf{X} \stackrel{\mathrm{d}}{=} \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} Z,$$

where $Z \sim E_1(0, 1, \psi)$. Hence, $\mathrm{VaR}_p(\mathbf{w}^T \mathbf{X}) = \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} \, \mathrm{VaR}_p(Z)$. This yields

$$\frac{\mathrm{VaR}_{0.99}(\mathbf{w}_A^T \mathbf{X})}{\mathrm{VaR}_{0.99}(\mathbf{w}_B^T \mathbf{X})} = \frac{\sqrt{\mathbf{w}_A^T \Sigma \mathbf{w}_A}}{\sqrt{\mathbf{w}_B^T \Sigma \mathbf{w}_B}}.$$

We have $\mathbf{w}_A^T \Sigma \mathbf{w}_A = \sigma^2(1+\rho)/2$ and $\mathbf{w}_B^T \Sigma \mathbf{w}_B = \sigma^2$. Hence

$$\frac{\mathrm{VaR}_{0.99}(L_A^\Delta)}{\mathrm{VaR}_{0.99}(L_B^\Delta)} = \frac{\mathrm{VaR}_{0.99}(\mathbf{w}_A^T \mathbf{X})}{\mathrm{VaR}_{0.99}(\mathbf{w}_B^T \mathbf{X})} = \sqrt{(1+\rho)/2}.$$

$\blacksquare$

10 Copulas

We will now introduce the notion of copula. The main reasons for introducing copulas are: 1) copulas are very useful for building multivariate models with nonstandard (nonGaussian) dependence structures, 2) copulas provide an understanding of dependence beyond linear correlation.

The reader seeking more information about copulas is recommended to consult the book [13].

10.1 Basic properties

Definition 10.1 *A d-dimensional copula is a distribution function on $[0,1]^d$ with standard uniform marginal distributions.* ∎

This means that a copula is the distribution function $P(U_1 \leq u_1, \ldots, U_d \leq u_d)$ of a random vector $(U_1, \ldots, U_d)$ with the property that for all k it holds that $P(U_k \leq u) = u$ for $u \in [0,1]$.

A bivariate distribution function F with marginal distribution functions is a function F that satisfies

(A1) $F(x_1, x_2)$ is nondecreasing in each argument x_k.

(A2) $F(x_1, \infty) = F_1(x_1)$ and $F(\infty, x_2) = F_2(x_2)$.

(A3) For all $(a_1, a_2), (b_1, b_2) \in \mathbb{R}^2$ with $a_k \leq b_k$ we have:

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

Notice that (A3) says that probabilities are always nonnegative. Hence a copula is a function C that satisfies

(B1) $C(u_1, u_2)$ is nondecreasing in each argument u_k.

(B2) $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$.

(B3) For all $(a_1, a_2), (b_1, b_2) \in [0,1]^2$ with $a_k \leq b_k$ we have:

$$C(b_1, b_2) - C(a_1, b_2) - C(b_1, a_2) + C(a_1, a_2) \geq 0.$$

Let $h : \mathbb{R} \to \mathbb{R}$ be nondecreasing. Then the following properties hold for the generalized inverse $h^{\leftarrow}$ of h.

(C1) h is continuous if and only if $h^{\leftarrow}$ is strictly increasing.

(C2) h is strictly increasing if and only if $h^{\leftarrow}$ is continuous.

(C3) If h is continuous, then $h(h^{\leftarrow}(y)) = y$.

(C4) If h is strictly increasing, then $h^{\leftarrow}(h(x)) = x$.

Recall the following important facts for a distribution function G on $\mathbb{R}$.

(D1) Quantile transform. If $U \sim U(0,1)$ (standard uniform distribution), then $P(G^{\leftarrow}(U) \leq x) = G(x)$.

(D2) Probability transform. If Y has distribution function G, where G is continuous, then $G(Y) \sim U(0,1)$.

Let $\mathbf{X}$ be a bivariate random vector with distribution function F that has continuous marginal distribution functions F_1, F_2. Consider the function C given by

$$C(u_1, u_2) = F(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2)).$$

It is clear that C is nondecreasing in each argument so (B1) holds. Moreover,

$$
\begin{aligned}
C(u_1, 1) &= F(F_1^{\leftarrow}(u_1), \infty) = P(X_1 \leq F_1^{\leftarrow}(u_1)) \\
&= P(F_1(X_1) \leq F_1(F_1^{\leftarrow}(u_1))) = P(F_1(X_1) \leq u_1) \\
&= u_1
\end{aligned}
$$

and similarly $C(1, u_2) = u_2$. Hence, (B2) holds. Since F is a bivariate distribution function (B3) holds. Hence, C is a copula.

The following result known as Sklar's Theorem is central to the theory of copulas. It also explains the name "copula": a function that "couples" the joint distribution function to its (univariate) marginal distribution functions.

Theorem 10.1 (Sklar's Theorem) *Let F be a joint distribution function with marginal distribution functions $F_1, \ldots, F_d$. Then there exists a copula C such that for all $x_1, \ldots, x_d \in \bar{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{10.1}$$

If $F_1, \ldots, F_d$ are continuous, then C is unique. Conversely, if C is a copula and $F_1, \ldots, F_d$ are distribution functions, then F defined by (10.1) is a joint distribution function with marginal distribution functions $F_1, \ldots, F_d$.

Definition 10.2 *Let F be a joint distribution function with continuous marginal distribution functions $F_1, \ldots, F_d$. Then the copula C in (10.1) is called the copula of F. If $\mathbf{X}$ is a random vector with distribution function F, then we also call C the copula of $\mathbf{X}$.* ∎

By (C3) above, if F_k is continuous, then $F_k(F_k^{\leftarrow}(u)) = u$. Hence, if F is a joint distribution function with continuous marginal distribution functions $F_1, \ldots, F_d$, then the unique copula of F is given by

$$C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)).$$

Much of the usefulness of copulas is due to the fact that the copula of a random vector with continuous marginal distribution functions is invariant under strictly increasing transformations of the components of the random vector.

Proposition 10.1 *Suppose $(X_1, \ldots, X_d)$ has continuous marginal distribution functions and copula C and let $T_1, \ldots, T_d$ be strictly increasing. Then also the random vector $(T_1(X_1), \ldots, T_d(X_d))$ has copula C.*

Proof. Let F_k denote the distribution function of X_k and let $\widetilde{F}_k$ denote the distribution function of $T_k(X_k)$. Then

$$\widetilde{F}_k(x) = \mathrm{P}(T_k(X_k) \le x) = \mathrm{P}(X_k \le T_k^{\leftarrow}(x)) = F_k(T_k^{\leftarrow}(x)),$$

i.e. $\widetilde{F}_k = F_k \circ T_k^{\leftarrow}$ is continuous. Since $\widetilde{F}_1, \ldots, \widetilde{F}_d$ are continuous the random vector $(T_1(X_1), \ldots, T_d(X_d))$ has a unique copula $\widetilde{C}$. Moreover, for any $\mathbf{x} \in \overline{\mathbb{R}}^d$,

$$
\begin{aligned}
\widetilde{C}(\widetilde{F}_1(x_1), \ldots, \widetilde{F}_d(x_d)) &= \mathrm{P}(T_1(X_1) \le x_1, \ldots, T_d(X_d) \le x_d) \\
&= \mathrm{P}(X_1 \le T_1^{\leftarrow}(x_1), \ldots, X_d \le T_d^{\leftarrow}(x_d)) \\
&= C(F_1 \circ T_1^{\leftarrow}(x_1), \ldots, F_d \circ T_d^{\leftarrow}(x_d)) \\
&= C(\widetilde{F}_1(x_1), \ldots, \widetilde{F}_d(x_d)).
\end{aligned}
$$

Since, for $k = 1, \ldots, d$, $\widetilde{F}_k$ is continuous, $\widetilde{F}_k(\overline{\mathbb{R}}) = [0, 1]$. Hence $\widetilde{C} = C$ on $[0, 1]^d$.

$\square$

Example 10.1 Let C be a d-dimensional copula and let $U_1, \ldots, U_d$ be random variables that are uniformly distributed on $[0, 1]$ with joint distribution function C. Let $F_1, \ldots, F_d$ be univariate continuous distribution functions. Then, for each k, $F_k^{\leftarrow}(U_k)$ has distribution function F_k (the quantile transform). Moreover, for each k, $F_k^{\leftarrow}$ is strictly increasing (C1). Hence, the random vector $(F_1^{\leftarrow}(U_1), \ldots, F_d^{\leftarrow}(U_d))$ has marginal distribution functions $F_1, \ldots, F_d$ and, by Proposition 10.1, copula C. ∎

Example 10.2 Let (X_1, X_2) be a random vector with continuous marginal distribution functions F_1, F_2 and copula C Let g_1 and g_2 be strictly decreasing functions. Determine the copula $\widetilde{C}$ of $(g_1(X_1), g_2(X_2))$?
The distribution- and quantile function of $g_k(X_k)$ is given by

$$
\begin{aligned}
\widetilde{F}_k(x) &= \mathrm{P}(g_k(X_k) \le x) = \mathrm{P}(X_k \ge g_k^{-1}(x)) = 1 - F_k(g_k^{-1}(x)), \\
\widetilde{F}_k^{\leftarrow}(p) &= g_k(F_k^{\leftarrow}(1-p)).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\widetilde{C}(u_1, u_2) &= \mathrm{P}(g_1(X_1) \le \widetilde{F}_1^{-1}(u_1), g_2(X_2) \le \widetilde{F}_2^{-1}(u_2)) \\
&= \mathrm{P}(X_1 \ge F_1^{\leftarrow}(1 - u_1), X_2 \ge F_2^{\leftarrow}(1 - u_2)) \\
&= \mathrm{P}(F_1(X_1) \ge 1 - u_1, F_2(X_2) \ge 1 - u_2) \\
&= 1 - (1 - u_1) - (1 - u_2) + C(1 - u_1, 1 - u_2) \\
&= C(1 - u_1, 1 - u_2) + u_1 + u_2 - 1.
\end{aligned}
$$

Notice that

$$C(u_1, u_2) = P(F_1(X_1) \le u_1, F_2(X_2) \le u_2),$$

$$\widetilde{C}(u_1, u_2) = P(1 - F_1(X_1) \le u_1, 1 - F_2(X_2) \le u_2).$$

See Figure 23 for an illustration with $g_1(x) = g_2(x) = -x$ and $F_1(x) = F_2(x) = \Phi(x)$ (standard normal distribution function). ∎
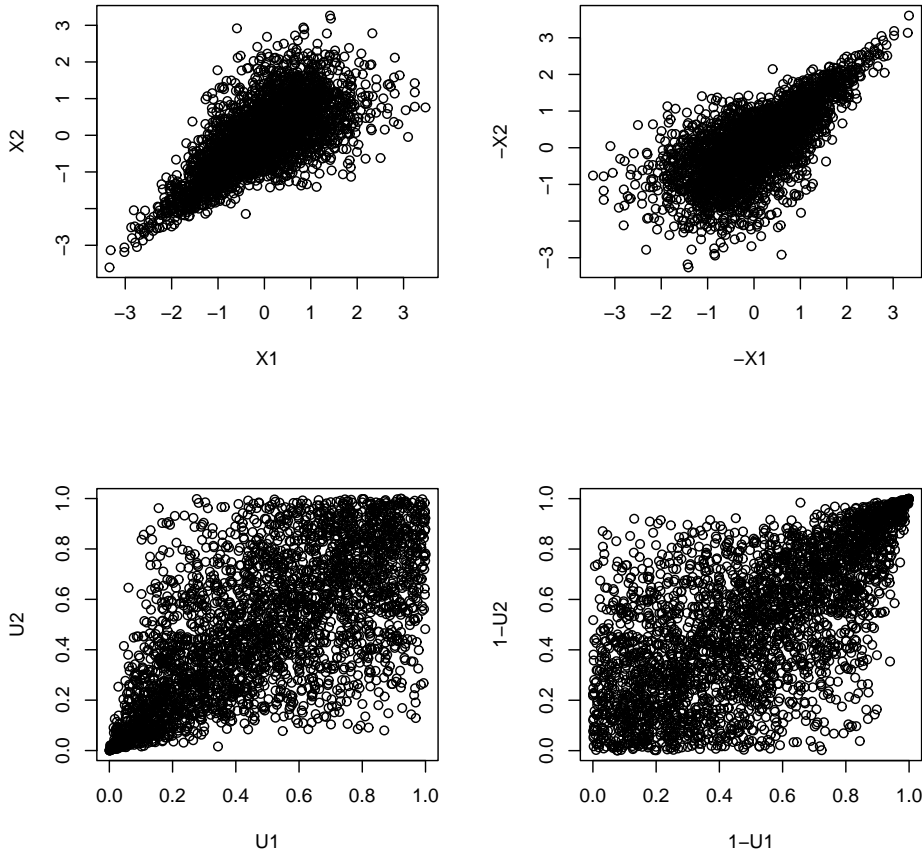
Figure 23: (U_1, U_2) has the copula C as its distribution function. (X_1, X_2) has standard normal marginal distributions and copula C. Samples of size 3000 from (X_1, X_2), $(-X_1, -X_2)$, (U_1, U_2) and $(1 - U_1, 1 - U_2)$.

Example 10.3 Let $\mathbf{X} \sim N_d(\mathbf{0}, R)$ where R is a correlation matrix. Denote by Φ_R and Φ the distribution functions of $\mathbf{X}$ and X_1 respectively (the d-dimensional standard normal distribution function and the 1-dimensional standard normal distribution function). Then $\mathbf{X}$ has the so-called Gaussian copula C_R^{Ga} given by

$$C_R^{\mathrm{Ga}}(\mathbf{u}) = P(\Phi(X_1) \le u_1, \dots, \Phi(X_d) \le u_d) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)). \tag{10.2}$$

Let $\mathbf{Y} = \boldsymbol{\mu} + \Sigma\mathbf{X}$, where for some $\sigma_1, \ldots, \sigma_d > 0$,

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \ldots & & 0 \\ 0 & \sigma_2 & 0 & \ldots & 0 \\ \ldots & & & & \\ 0 & & \ldots & 0 & \sigma_d \end{pmatrix}.$$

Then $\mathbf{Y} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma R \Sigma)$. Note that $\mathbf{Y}$ has linear correlation matrix R. Note also that $T_k(x) = \mu_k + \sigma_k x$ is strictly increasing and that

$$\mathbf{Y} = (T_1(X_1), \ldots, T_d(X_d)).$$

Hence if $C_\mathbf{Y}$ denotes the copula of $\mathbf{Y}$, then by Proposition 10.1 we have $C_\mathbf{Y} = C_R^{\mathrm{Ga}}$. We conclude that the copula of a nondegenerate d-dimensional normal distribution depends only on the linear correlation matrix. For $d = 2$ we see from (10.2) that C_R^{Ga} can be written as

$$C_R^{\mathrm{Ga}}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{\frac{-(x_1^2 - 2\rho x_1 x_2 + x_2^2)}{2(1-\rho^2)}\right\} \mathrm{d}x_1 \mathrm{d}x_2,$$

if $\rho = R_{12} \in (-1, 1)$. ∎

The following result provides us with universal bounds for copulas. See also Example 10.5 below.

Proposition 10.2 (Fréchet bounds) *For every copula C we have the bounds*

$$\max\left\{\sum_{k=1}^{d} u_k - d + 1, 0\right\} \le C(u_1, \ldots, u_d) \le \min\{u_1, \ldots, u_d\}.$$

For $d \ge 2$ we denote by W_d the Fréchet lower bound and by M_d the Fréchet upper bound. For $d = 2$ we drop the subscript of W_2 and M_2, i.e. $W = W_2$ and $M = M_2$.

Example 10.4 Let W_d be the Fréchet lower bound and consider the set function Q given by

$$Q([a_1, b_1] \times \cdots \times [a_d, b_d]) = \sum_{k_1=1}^{2} \cdots \sum_{k_d=1}^{2} (-1)^{k_1 + \cdots + k_d} W_d(u_{1k_1}, \ldots, u_{dk_d}),$$

for all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ with $a_k \le b_k$ where $u_{j1} = a_j$ and $u_{j2} = b_j$ for $j \in \{1, \ldots, d\}$. W_d is a copula (distribution function) if and only if Q is its probability distribution. However,

$$\begin{aligned} Q([1/2, 1]^d) &= \max(1 + \cdots + 1 - d + 1, 0) \\ &\quad - d \max(1/2 + 1 + \cdots + 1 - d + 1, 0) \\ &\quad + \binom{d}{2} \max(1/2 + 1/2 + 1 + \cdots + 1 - d + 1, 0) \\ &\quad \cdots \\ &\quad + \max(1/2 + \cdots + 1/2 - d + 1, 0) \\ &= 1 - d/2. \end{aligned}$$

Hence, Q is not a probability distribution for $d \geq 3$ so W_d is not a copula for $d \geq 3$. ∎

The following result shows that the Fréchet lower bound is the best possible.

Proposition 10.3 *For any $d \geq 3$ and any $\mathbf{u} \in [0,1]^d$, there exists a copula C such that $C(\mathbf{u}) = W_d(\mathbf{u})$.*

Remark 10.1 *For any $d \geq 2$, the Fréchet upper bound M_d is a copula.*

For $d = 2$ the following example shows which random vectors have the Fréchet bounds as their copulas.

Example 10.5 Let X be a random variable with continuous distribution function F_X. Let $Y = T(X)$ for some strictly increasing function T and denote by F_Y the distribution function of Y. Note that

$$F_Y(x) = \mathrm{P}(T(X) \leq x) = F_X(T^{-1}(x))$$

and that F_Y is continuous. Hence the copula of $(X, T(X))$ is

$$\begin{aligned}
\mathrm{P}(F_X(X) \leq u, F_Y(Y) \leq v) &= \mathrm{P}(F_X(X) \leq u, F_X(T^{-1}(T(X))) \leq v) \\
&= \mathrm{P}(F_X(X) \leq \min\{u, v\}) \\
&= \min\{u, v\}.
\end{aligned}$$

By Proposition 10.1 the copula of $(X, T(X))$ is the copula of (U, U), where $U \sim U(0, 1)$. Let $Z = S(X)$ for some strictly decreasing function S and denote by F_Z the distribution function of Z. Note that

$$F_Z(x) = \mathrm{P}(S(X) \leq x) = \mathrm{P}(X > S^{-1}(x)) = 1 - F_X(S^{-1}(x))$$

and that F_Z is continuous. Hence the copula of $(X, S(X))$ is

$$\begin{aligned}
\mathrm{P}(F_X(X) \leq u, F_Z(Z) \leq v) &= \mathrm{P}(F_X(X) \leq u, 1 - F_X(S^{-1}(S(X))) \leq v) \\
&= \mathrm{P}(F_X(X) \leq u, F_X(X) > 1 - v) \\
&= \mathrm{P}(F_X(X) \leq u) - \mathrm{P}(F_X(X) \leq \min\{u, 1 - v\}) \\
&= u - \min\{u, 1 - v\} \\
&= \max\{u + v - 1, 0\}.
\end{aligned}$$

By (a modified version of) Proposition 10.1 the copula of $(X, S(X))$ is the copula of $(U, 1 - U)$, where $U \sim U(0, 1)$. ∎

10.2   Dependence measures

Comonotonicity and countermonotonicity revisited

Proposition 10.4 *Let (X_1, X_2) have one of the copulas W or M (as a possible copula). Then there exist two monotone functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ and a random variable Z so that*

$$(X_1, X_2) \overset{\mathrm{d}}{=} (\alpha(Z), \beta(Z)),$$

with α increasing and β decreasing in the former case (W) and both α and β increasing in the latter case (M). The converse of this result is also true.

Hence, if (X_1, X_2) has the copula M (as a possible copula), then X_1 and X_2 are comonotonic; if it has the copula W (as a possible copula), then they are countermonotonic. Note that if any of F_1 and F_2 (the distribution functions of X_1 and X_2, respectively) have discontinuities, so that the copula is not unique, then W and M are possible copulas. Recall also that if F_1 and F_2 are continuous, then

$$C = W \quad \Leftrightarrow \quad X_2 = T(X_1) \text{ a.s., } T = F_2^{\leftarrow} \circ (1 - F_1) \text{ decreasing,}$$
$$C = M \quad \Leftrightarrow \quad X_2 = T(X_1) \text{ a.s., } T = F_2^{\leftarrow} \circ F_1 \text{ increasing.}$$

Kendall's tau and Spearman's rho revisited

To begin with we recall the definitions of the concordance measures Kendall's tau and Spearman's rho.

Definition 10.3 Kendall's tau *for the random vector (X_1, X_2) is defined as*

$$\varrho_\tau(X_1, X_2) = \mathrm{P}((X_1 - X_1')(X_2 - X_2') > 0) - \mathrm{P}((X_1 - X_1')(X_2 - X_2') < 0),$$

where (X_1', X_2') is an independent copy of (X_1, X_2). Spearman's rho for the random vector (X_1, X_2) is defined as

$$\varrho_S(X_1, X_2) = 3\left(\mathrm{P}((X_1 - X_1')(X_2 - X_2'') > 0) - \mathrm{P}((X_1 - X_1')(X_2 - X_2'') < 0)\right),$$

where (X_1', X_2') and (X_1'', X_2'') are independent copies of (X_1, X_2). ∎

An important property of Kendall's tau and Spearman's rho is that they are invariant under strictly increasing transformations of the underlying random variables. If (X_1, X_2) is a random vector with continuous marginal distribution functions and T_1 and T_2 are strictly increasing transformations on the range of X_1 and X_2 respectively, then $\varrho_\tau(T_1(X_1), T_2(X_2)) = \varrho_\tau(X_1, X_2)$. The same property holds for Spearman's rho. Note that this implies that Kendall's tau and Spearman's rho do not depend on the (marginal) distributions of X_1 and X_2. This is made clear in the following two results.

Proposition 10.5 *Let (X_1, X_2) be a random vector with continuous marginal distribution functions and with copula C. Then*

$$\varrho_\tau(X_1, X_2) = 4 \int_{[0,1]^2} C(u_1, u_2) \mathrm{d}C(u_1, u_2) - 1 = 4\, \mathrm{E}(C(U_1, U_2)) - 1,$$
$$\varrho_S(X_1, X_2) = 12 \int_{[0,1]^2} u_1 u_2 \mathrm{d}C(u_1, u_2) - 3 = 12 \int_{[0,1]^2} C(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 - 3$$
$$= 12\, \mathrm{E}(U_1 U_2) - 3,$$

where (U_1, U_2) has distribution function C.

Remark 10.2 *Note that by Proposition 10.5, if F_1 and F_2 denotes the distribution functions of X_1 and X_2 respectively,*

$$
\begin{aligned}
\varrho_S(X_1, X_2) &= 12 \int_{[0,1]^2} u_1 u_2 \mathrm{d}C(u_1, u_2) - 3 \\
&= 12\,\mathrm{E}(F_1(X_1)F_2(X_2)) - 3 \\
&= \frac{\mathrm{E}(F_1(X_1)F_2(X_2)) - 1/4}{1/12} \\
&= \frac{\mathrm{E}(F_1(X_1)F_2(X_2)) - \mathrm{E}(F_1(X_1))\,\mathrm{E}(F_2(X_2))}{\sqrt{\mathrm{var}(F_1(X_1))}\sqrt{\mathrm{var}(F_2(X_2))}} \\
&= \varrho_l(F_1(X_1), F_2(X_2)).
\end{aligned}
$$

Hence Spearman's rho is simply the linear correlation coefficient of the probability transformed random variables.

Tail dependence revisited

We now return to the dependence concept called tail dependence. The concept of tail dependence is important for the modeling of joint extremes, particularly in portfolio Risk Management. We recall the definition of the coefficient of tail dependence.

Let (X_1, X_2) be a random vector with marginal distribution functions F_1 and F_2. The *coefficient of upper tail dependence* of (X_1, X_2) is defined as

$$
\lambda_U(X_1, X_2) = \lim_{u \nearrow 1} \mathrm{P}(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)),
$$

provided that the limit $\lambda_U \in [0, 1]$ exists. The *coefficient of lower tail dependence* is defined as

$$
\lambda_L(X_1, X_2) = \lim_{u \searrow 0} \mathrm{P}(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)),
$$

provided that the limit $\lambda_L \in [0, 1]$ exists. If $\lambda_U > 0$ ($\lambda_L > 0$), then we say that (X_1, X_2) has upper (lower) tail dependence.

Proposition 10.6 *Let (X_1, X_2) be a random vector with continuous marginal distribution functions and copula C. Then*

$$
\lambda_U(X_1, X_2) = \lim_{u \nearrow 1}(1 - 2u + C(u, u))/(1 - u),
$$

provided that the limit exists, and

$$
\lambda_L(X_1, X_2) = \lim_{u \searrow 0} C(u, u)/u,
$$

provided that the limit exists.

That the limit need not exist is shown by the following example.

Example 10.6 Let Q be a probability measure on $[0,1]^2$ such that for every integer $n \geq 1$, Q assigns mass 2^{-n}, uniformly distributed, to the line segment between $(1 - 2^{-n}, 1 - 2^{-n+1})$ and $(1 - 2^{-n+1}, 1 - 2^{-n})$. Define the distribution function C by $C(u_1, u_2) = Q([0, u_1] \times [0, u_2])$ for $u_1, u_2 \in [0, 1]$. Note that $C(u_1, 0) = 0 = C(0, u_2)$, $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$, i.e. C is a copula. Note also that for every $n \geq 1$ (with $\overline{C}(u, u) = 1 - 2u + C(u, u)$)

$$\overline{C}(1 - 2^{-n+1}, 1 - 2^{-n+1})/2^{-n+1} = 1$$

and

$$\overline{C}(1 - 3/2^{n+1}, 1 - 3/2^{n+1})/(3/2^{n+1}) = 2/3.$$

In particular $\lim_{u \nearrow 1} \overline{C}(u, u)/(1 - u)$ does not exist. $\blacksquare$

Example 10.7 Consider the so-called Gumbel family of copulas given by

$$C_\theta(u_1, u_2) = \exp(-[(- \ln u_1)^\theta + (- \ln u_2)^\theta]^{1/\theta}),$$

for $\theta \geq 1$. Then

$$\frac{1 - 2u + C_\theta(u, u)}{1 - u} = \frac{1 - 2u + \exp(2^{1/\theta} \ln u)}{1 - u} = \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u},$$

and hence by l'Hospitals rule

$$\lim_{u \nearrow 1} (1 - 2u + C_\theta(u, u))/(1 - u) = 2 - \lim_{u \nearrow 1} 2^{1/\theta} u^{2^{1/\theta} - 1} = 2 - 2^{1/\theta}.$$

Thus for $\theta > 1$, C_θ has upper tail dependence: $\lambda_U = 2 - 2^{1/\theta}$. Moreover, again by using l'Hospitals rule,

$$\lim_{u \searrow 0} C_\theta(u, u)/u = \lim_{u \searrow 0} 2^{1/\theta} u^{2^{1/\theta}} = 0,$$

i.e. $\lambda_L = 0$. See Figure 24 for a graphical illustration. $\blacksquare$

Example 10.8 Consider the so-called Clayton family of copulas given by

$$C_\theta(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta},$$

for $\theta > 0$. Then $C_\theta(u, u) = (2u^{-\theta} - 1)^{-1/\theta}$ and hence, by l'Hospitals rule,

$$\begin{aligned}
\lim_{u \searrow 0} C_\theta(u, u)/u &= \lim_{u \searrow 0} (2u^{-\theta} - 1)^{-1/\theta}/u \\
&= \lim_{u \searrow 0} (-\frac{1}{\theta})(2u^{-\theta} - 1)^{-1/\theta-1}(-2\theta(u^\theta)^{-1/\theta-1}) \\
&= \lim_{u \searrow 0} 2(2 - u^\theta)^{-1/\theta-1} \\
&= 2^{-1/\theta},
\end{aligned}$$

i.e. $\lambda_L = 2^{-1/\theta}$. Similarly one shows that $\lambda_U = 0$. See Figure 24 for a graphical illustration. $\blacksquare$
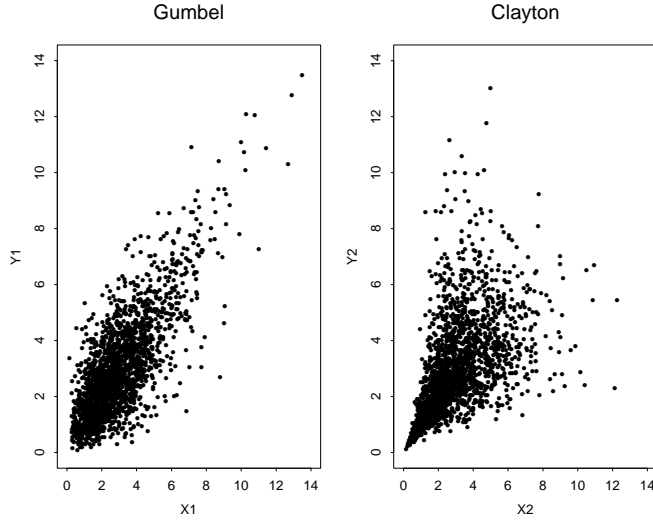
Figure 24: Samples from two distributions with Gamma$(3,1)$ marginal distribution functions, linear correlation 0.5 but different dependence structures. $(X1, Y1)$ has a Gumbel copula and $(X2, Y2)$ has a Clayton copula.

10.3 Elliptical copulas

Definition 10.4 *Let $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with distribution function F and with continuous marginal distribution functions $F_1, \ldots, F_d$. Then the copula C given by $C(\mathbf{u}) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$ is said to be an elliptical copula.* ∎

Note that an elliptical copula is not the distribution function of an elliptical distribution, but rather the copula of an elliptical distribution.

The copula of the d-dimensional normal distribution with linear correlation matrix R is

$$C_R^{\mathrm{Ga}}(\mathbf{u}) = \Phi_R^d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)),$$

where Φ_R^d denotes the joint distribution function of the d-dimensional standard normal distribution function with linear correlation matrix R, and Φ^{-1} denotes the inverse of the distribution function of the univariate standard normal distribution. Copulas of the above form are called Gaussian copulas. In the bivariate case the copula expression can be written as

$$C_R^{\mathrm{Ga}}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ \frac{-(x_1^2 - 2\rho x_1 x_2 + x_2^2)}{2(1-\rho^2)} \right\} \mathrm{d}x_1 \mathrm{d}x_2,$$

if $\rho = R_{12} \in (-1, 1)$.

Proposition 10.7 *(i) If (X_1, X_2) is a normally distributed random vector, then $\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 0$.*
(ii) If (X_1, X_2) has continuous marginal distribution functions and a Gaussian copula, then $\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 0$.

If **X** has the stochastic representation

$$\mathbf{X} =_d \boldsymbol{\mu} + \frac{\sqrt{\nu}}{\sqrt{S}} A\mathbf{Z}, \qquad (10.3)$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ (column vector), $S \sim \chi_\nu^2$ and $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ (column vector) are independent, then **X** has an d-dimensional t_ν-distribution with mean $\boldsymbol{\mu}$ (for $\nu > 1$) and covariance matrix $\frac{\nu}{\nu-2} AA^{\mathrm{T}}$ (for $\nu > 2$). If $\nu \le 2$, then $\mathrm{Cov}(\mathbf{X})$ is not defined. In this case we just interpret $\Sigma = AA^{\mathrm{T}}$ as being the shape parameter of the distribution of **X**. The copula of **X** given by (10.3) can be written as

$$C_{\nu,R}^t(\mathbf{u}) = t_{\nu,R}^d(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)),$$

where $R_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ for $i, j \in \{1, \ldots, d\}$ and where $t_{\nu,R}^d$ denotes the distribution function of $\sqrt{\nu}A\mathbf{Z}/\sqrt{S}$, where $AA^{\mathrm{T}} = R$. Here t_ν denotes the (equal) margins of $t_{\nu,R}^d$, i.e. the distribution function of $\sqrt{\nu}Z_1/\sqrt{S}$. In the bivariate case the copula expression can be written as

$$C_{\nu,R}^t(u_1, u_2) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} \mathrm{d}x_1 \mathrm{d}x_2,$$

if $\rho = R_{12} \in (-1, 1)$. Note that R_{12} is simply the usual linear correlation coefficient of the corresponding bivariate t_ν-distribution if $\nu > 2$.

Proposition 10.8 *(i) If (X_1, X_2) has a t-distribution with ν degrees of freedom and linear correlation matrix R, then*

$$\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-R_{12}}/\sqrt{1+R_{12}} \right). \qquad (10.4)$$

(ii) If (X_1, X_2) has continuous marginal distribution functions and a t-copula with parameters ν and R, then $\lambda_U(X_1, X_2)$ and $\lambda_L(X_1, X_2)$ are as in (10.4).

From this it is also seen that the coefficient of upper tail dependence is increasing in R_{12} and decreasing in ν, as one would expect. Furthermore, the coefficient of upper (lower) tail dependence tends to zero as the number of degrees of freedom tends to infinity for $R_{12} < 1$.

The following result will play an important role in parameter estimation in models with elliptical copulas.

Proposition 10.9 *(i) If $(X_1, X_2) \sim E_2(\boldsymbol{\mu}, \Sigma, \psi)$ with continuous marginal distribution functions, then*

$$\varrho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin R_{12}, \qquad (10.5)$$

where $R_{12} = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$.
(ii) If (X_1, X_2) has continuous marginal distribution functions and the copula of $E_2(\boldsymbol{\mu}, \Sigma, \psi)$, then relation (10.5) holds.

10.4 Simulation from Gaussian and t-copulas

We now address the question of random variate generation from the Gaussian copula C_R^{Ga}. For our purpose, it is sufficient to consider only strictly positive definite matrices R. Write $R = AA^{\mathrm{T}}$ for some $d \times d$ matrix A, and if $Z_1, \ldots, Z_d \sim \mathrm{N}(0,1)$ are independent, then

$$\boldsymbol{\mu} + A\mathbf{Z} \sim \mathrm{N}_d(\boldsymbol{\mu}, R).$$

One natural choice of A is the Cholesky decomposition of R. The Cholesky decomposition of R is the unique lower-triangular matrix L with $LL^{\mathrm{T}} = R$. Furthermore Cholesky decomposition routines are implemented in most mathematical software. This provides an easy algorithm for random variate generation from the d-dimensional Gaussian copula C_R^{Ga}.

Algorithm 10.1

- *Find the Cholesky decomposition A of R: $R = AA^{\mathrm{T}}$.*

- *Simulate d independent random variates $Z_1, \ldots, Z_d$ from $\mathrm{N}(0,1)$.*

- *Set $\mathbf{X} = A\mathbf{Z}$.*

- *Set $U_k = \Phi(X_k)$ for $k = 1, \ldots, d$.*

- *$\mathbf{U} = (U_1, \ldots, U_d)$ has distribution function C_R^{Ga}.*

As usual Φ denotes the univariate standard normal distribution function. □
    Equation (10.3) provides an easy algorithm for random variate generation from the t-copula, $C_{\nu,R}^t$.

Algorithm 10.2

- *Find the Cholesky decomposition A of R: $R = AA^{\mathrm{T}}$.*

- *Simulate d independent random variates $Z_1, \ldots, Z_d$ from $\mathrm{N}(0,1)$.*

- *Simulate a random variate S from χ_ν^2 independent of $Z_1, \ldots, Z_d$.*

- *Set $\mathbf{Y} = A\mathbf{Z}$.*

- *Set $\mathbf{X} = \frac{\sqrt{\nu}}{\sqrt{S}}\mathbf{Y}$.*

- *Set $U_k = t_\nu(X_k)$ for $k = 1, \ldots, d$.*

- *$\mathbf{U} = (U_1, \ldots, U_d)$ has distribution function $C_{\nu,R}^t$.*

□

10.5 Archimedean copulas

As we have seen, elliptical copulas are derived from distribution functions for elliptical distributions using Sklar's Theorem. Since simulation from elliptical distributions is easy, so is simulation from elliptical copulas. There are however drawbacks: elliptical copulas do not have closed form expressions and are restricted to have radial symmetry. In many finance and insurance applications it seems reasonable that there is a stronger dependence between big losses (e.g. a stock market crash) than between big gains. Such asymmetries cannot be modeled with elliptical copulas.

In this section we discuss an important class of copulas called Archimedean copulas. This class of copulas is worth studying for a number of reasons. Many interesting parametric families of copulas are Archimedean and the class of Archimedean copulas allows for a great variety of different dependence structures. Furthermore, in contrast to elliptical copulas, all commonly encountered Archimedean copulas have closed form expressions. Unlike the copulas discussed so far these copulas are not derived from multivariate distribution functions using Sklar's Theorem. A consequence of this is that we need somewhat technical conditions to assert that multivariate extensions of bivariate Archimedean copulas are indeed copulas. A further disadvantage is that multivariate extensions of Archimedean copulas in general suffer from lack of free parameter choice in the sense that some of the entries in the resulting rank correlation matrix are forced to be equal. We begin with a general definition of Archimedean copulas.

Proposition 10.10 *Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(0) = \infty$ and $\varphi(1) = 0$. Let $C : [0,1]^2 \to [0,1]$ be given by*

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)). \tag{10.6}$$

Then C is a copula if and only if φ is convex.

Copulas of the form (10.6) are called Archimedean copulas. The function φ is called a generator of the copula.

Example 10.9 Let $\varphi(t) = (-\ln t)^\theta$, where $\theta \geq 1$. Clearly $\varphi(t)$ is continuous and $\varphi(1) = 0$. $\varphi'(t) = -\theta(-\ln t)^{\theta-1}\frac{1}{t}$, so φ is a strictly decreasing function from $[0,1]$ to $[0,\infty]$. $\varphi''(t) \geq 0$ on $[0,1]$, so φ is convex. Moreover $\varphi(0) = \infty$. From (10.6) we get

$$C_\theta^{\text{Gu}}(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}).$$

Furthermore $C_1 = \Pi$ ($\Pi(u_1, u_2) = u_1 u_2$) and $\lim_{\theta \to \infty} C_\theta = M$ ($M(u_1, u_2) = \min(u_1, u_2)$). This copula family is called the Gumbel family. As shown in Example 10.7 this copula family has upper tail dependence. ∎

Example 10.10 Let $\varphi(t) = t^{-\theta} - 1$, where $\theta > 0$. This gives the Clayton family

$$C_\theta^{\text{Cl}}(u_1, u_2) = \left((u_1^{-\theta} - 1) + (u_2^{-\theta} - 1) + 1\right)^{-1/\theta}. = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}. \tag{10.7}$$

Moreover, $\lim_{\theta\to 0} C_\theta = \Pi$ and $\lim_{\theta\to\infty} C_\theta = M$. As shown in Example 10.8 this copula family has upper tail dependence. $\blacksquare$

Recall that Kendall's tau for a copula C can be expressed as a double integral of C. This double integral is in most cases not straightforward to evaluate. However for an Archimedean copula, Kendall's tau can be expressed as a one-dimensional integral of the generator and its derivative.

Proposition 10.11 *Let (X_1, X_2) be a random vector with continuous marginal distribution functions and with an Archimedean copula C generated by φ. Then Kendall's tau of (X_1, X_2) is given by*

$$\varrho_\tau(X_1, X_2) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

Example 10.11 Consider the Gumbel family with generator $\varphi(t) = (-\ln t)^\theta$, for $\theta \geq 1$. Then $\varphi(t)/\varphi'(t) = (t \ln t)/\theta$. Using Proposition 10.11 we can calculate Kendall's tau for the Gumbel family.

$$\varrho_\tau(\theta) = 1+4 \int_0^1 \frac{t \ln t}{\theta} dt = 1+\frac{4}{\theta}\left(\left[\frac{t^2}{2}\ln t\right]_0^1 - \int_0^1 \frac{t}{2} dt\right) = 1+\frac{4}{\theta}(0-1/4) = 1-\frac{1}{\theta}.$$

As a consequence, in order to have Kendall's tau equal to 0.5 in Figure 25 (the Gumbel case), we put $\theta = 2$. $\blacksquare$

Example 10.12 Consider the Clayton family with generator $\varphi(t) = t^{-\theta} - 1$, for $\theta > 0$. Then $\varphi(t)/\varphi'(t) = (t^{\theta+1} - t)/\theta$. Using Proposition 10.11 we can calculate Kendall's tau for the Clayton family.

$$\varrho_\tau(\theta) = 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} dt = 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) = \frac{\theta}{\theta+2}.$$

$\blacksquare$

A natural question is under which additional conditions on φ we have that the most simple multivariate extension of bivariate Archimedean copulas,

$$\varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)),$$

is a copula for $d \geq 3$. The following results address this question and show why inverses of Laplace transforms are natural choices for generators of Archimedean copulas.

Definition 10.5 *A function $g : [0, \infty) \to [0, \infty)$ is completely monotonic if it is continuous and if for any $t \in (0, \infty)$ and $k = 0, 1, 2, \ldots,$*

$$(-1)^k \left(\frac{d^k}{ds^k} g(s)\right)\Big|_{s=t} \geq 0.$$

$\blacksquare$
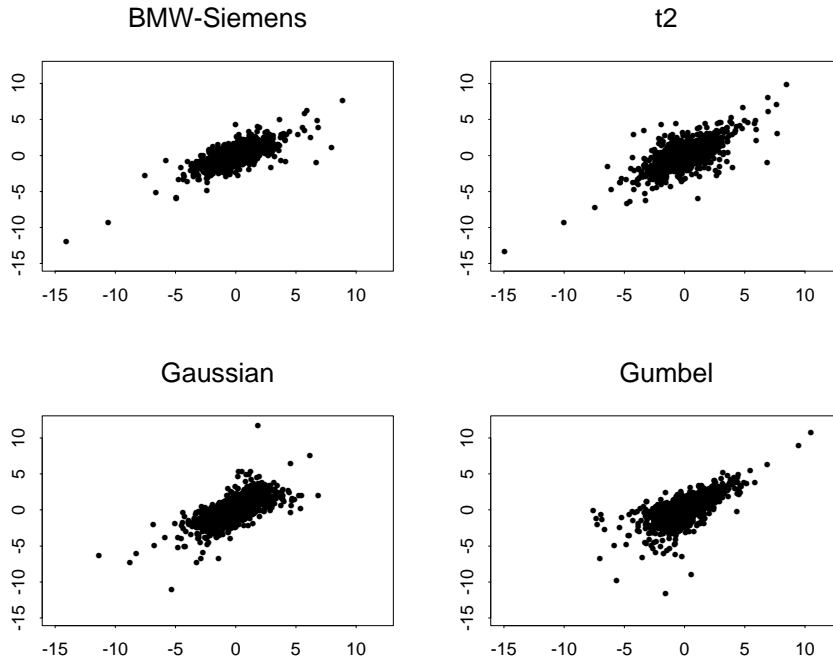
Figure 25: The upper left plot shows BMW-Siemens daily log returns from 1989 to 1996. The other plots show samples from bivariate distributions with t_4-margins and Kendall's tau 0.5.

Proposition 10.12 *Let $\varphi : [0,1] \to [0,\infty]$ be continuous and strictly decreasing such that $\varphi(0) = \infty$ and $\varphi(1) = 0$. Then, for any $d \geq 2$, the function $C : [0,1]^d \to [0,1]$ given by*

$$C(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d))$$

is a copula if and only if φ^{-1} is completely monotonic on $[0,\infty)$.

The following result tells us where to look for generators satisfying the conditions of Proposition 10.12.

Lemma 10.1 *A function $\Psi : [0,\infty) \to [0,\infty)$ is the Laplace transform of a distribution function G on $[0,\infty)$ if and only if Ψ is completely monotonic and $\Psi(0) = 1$.*

Proposition 10.13 *Let G be a distribution function on $[0,\infty)$ with $G(0) = 0$ and Laplace transform*

$$\Psi(s) = \int_0^\infty e^{-sx} \mathrm{d}G(x), \quad s \geq 0.$$

Consider a random variable X with distribution function G and a set of $[0,1]$-valued random variables $U_1, \ldots, U_d$ which are conditionally independent given

X with conditional distribution function given by $F_{U_k|X=x}(u) = \exp(-x\Psi^{-1}(u))$ for $u \in [0,1]$. Then

$$P(U_1 \le u_1, \ldots, U_d \le u_d) = \Psi(\Psi^{-1}(u_1) + \cdots + \Psi^{-1}(u_d)),$$

so that the distribution function of $\mathbf{U}$ is an Archimedean copula with generator Ψ^{-1}.

Proof.

$$P(U_1 \le u_1, \ldots, U_d \le u_d)$$
$$= \int_0^\infty P(U_1 \le u_1, \ldots, U_d \le u_d \mid X = x)\,\mathrm{d}G(x)$$
$$= \int_0^\infty P(U_1 \le u_1 \mid X = x)\ldots P(U_d \le u_d \mid X = x)\,\mathrm{d}G(x)$$
$$= \int_0^\infty \exp\{-x(\Psi^{-1}(u_1) + \cdots + \Psi^{-1}(u_d))\}\,\mathrm{d}G(x)$$
$$= \Psi(\Psi^{-1}(u_1) + \cdots + \Psi^{-1}(u_d)).$$

$\square$

10.6 Simulation from Gumbel and Clayton copulas

As seen from Proposition 10.13, the following algorithm shows how to simulate from an Archimedean copula C of the form

$$C(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)),$$

where φ^{-1} is the Laplace transform of a distribution function G on $[0, \infty)$ with $G(0) = 0$.

Algorithm 10.3

- *Simulate a variate X with distribution function G such that the Laplace transform Ψ of G is the inverse of the generator φ of the required copula C.*

- *Simulate independent standard uniform variates $V_1, \ldots, V_d$.*

- $\mathbf{U} = (\Psi(-\ln(V_1)/X), \ldots, \Psi(-\ln(V_d)/X))$ *has distribution function C.*

To verify that this is correct, notice that with $U_k = \Psi(-\ln(V_k)/X)$ we have

$$P(U_k \le u_k \mid X = x) = P(\Psi(-\ln(V_k)/x) \le u_k)$$
$$= P(-\ln V_k \ge x\Psi^{-1}(u_k))$$
$$= P(V_k \le \exp\{-x\Psi^{-1}(u_k)\})$$
$$= \exp\{-x\Psi^{-1}(u_k)\}.$$

78

As we have seen, the generator $\varphi(t) = t^{-\theta} - 1$, $\theta > 0$, generates the Clayton copula $C_\theta^{\mathrm{Cl}}(\mathbf{u}) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}$. Let $X \sim \mathrm{Gamma}(1/\theta, 1)$, i.e. let X be Gamma-distributed with density function $f_X(x) = x^{1/\theta-1}e^{-x}/\Gamma(1/\theta)$. Then X has Laplace transform

$$\mathrm{E}(e^{-sX}) = \int_0^\infty e^{-sx}\frac{1}{\Gamma(1/\theta)}x^{1/\theta-1}e^{-x}\mathrm{d}x = (s+1)^{-1/\theta} = \varphi^{-1}(s).$$

Hence, the following algorithm can be used for simulation from a Clayton copula.

Algorithm 10.4

- *Simulate a variate $X \sim \mathsf{Gamma}(1/\theta, 1)$.*

- *Simulate independent standard uniform variates $V_1, \ldots, V_d$.*

- *If $\Psi(s) = (s+1)^{-1/\theta}$, then $\mathbf{U} = (\Psi(-\ln(V_1)/X), \ldots, \Psi(-\ln(V_d)/X))$ has distribution function C_θ^{Cl}.*

This approach can also be used for simulation from Gumbel copulas. However, in this case X is a random variable with a nonstandard distribution which is not available in most statistical software. However, one can simulate from bivariate Gumbel copulas using a different method. Take $\theta \geq 1$ and let $\overline{F}(x) = 1 - F(x) = \exp(-x^{1/\theta})$ for $x \geq 0$. If $(Z_1, Z_2) = (VS^\theta, (1-V)S^\theta)$ where V and S are independent, V is standard uniformly distributed and S has density $h(s) = (1 - 1/\theta + (1/\theta)s)\exp(-s)$, then $(\overline{F}(Z_1), \overline{F}(Z_2))$ has distribution function C_θ^{Gu}, where

$$C_\theta^{\mathrm{Gu}}(u_1, u_2) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}).$$

This leads to the following algorithm for simulation from bivariate Gumbel copulas.

Algorithm 10.5

- *Simulate independent random variates $V_1, V_2 \sim U(0,1)$.*

- *Simulate independent random variates W_1, W_2, where $W_k \sim \mathsf{Gamma}(k, 1)$.*

- *Set $S = 1_{\{V_2 \leq 1/\theta\}}W_1 + 1_{\{V_2 > 1/\theta\}}W_2$.*

- *Set $(Z_1, Z_2) = (V_1 S^\theta, (1 - V_1)S^\theta)$.*

- *$\mathbf{U} = (\exp(-Z_1^{1/\theta}), \exp(-Z_2^{1/\theta}))$ has distribution function C_θ^{Gu}.*

10.7 Fitting copulas to data

We now turn to the problem of fitting a parametric copula to data. Clearly, first one has to decide which parametric family to fit. This is perhaps best done by graphical means; if one sees clear signs of asymmetry in the dependence structure as illustrated in Figure 24, then an Archimedean copula with this kind of asymmetry might be a reasonable choice. Otherwise, see e.g. Figure 26, one might go for an elliptical copula. It is also useful to check whether the data shows signs of tail dependence, and depending on whether there are such signs choose a copula with this property.
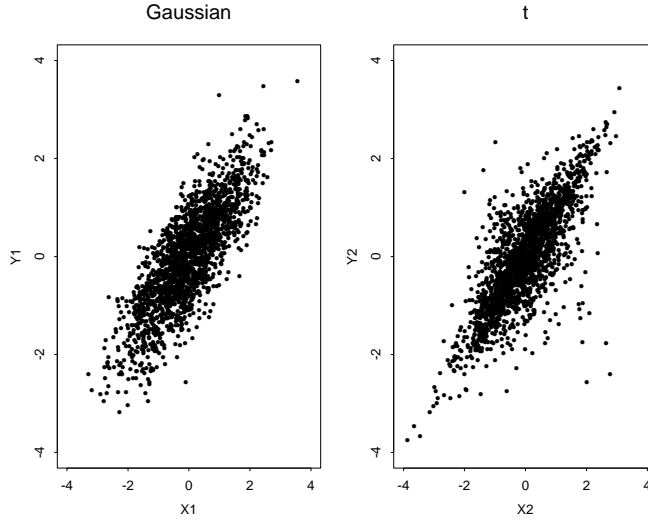
Figure 26: Samples from two distributions with standard normal margins, $R_{12} = 0.8$ but different dependence structures. $(X1, Y1)$ has a Gaussian copula and $(X2, Y2)$ has a t_2-copula.

We will consider the problem of estimating the parameter vector $\boldsymbol{\theta}$ of a copula $C_{\boldsymbol{\theta}}$ given an iid sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ where $\mathbf{X}_k \sim F$ for some distribution function F with continuous marginal distribution functions $F_1, \ldots, F_d$ and hence a unique representation $F(\mathbf{x}) = C_{\boldsymbol{\theta}}(F_1(x_1), \ldots, F_d(x_d))$.

We have seen that for Gaussian-, t-, Gumbel- and Clayton copulas there are simple relations between Kendall's tau and certain copula parameters:

$$
\begin{aligned}
C_R^{\text{Ga}}(\mathbf{u}) &= \Phi_R^d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), & R_{ij} &= \sin(\pi(\varrho_\tau)_{ij}/2), \\
C_R^{\text{t}}(\mathbf{u}) &= t_{\nu,R}^d(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)), & R_{ij} &= \sin(\pi(\varrho_\tau)_{ij}/2), \\
C_\theta^{\text{Gu}}(\mathbf{u}) &= \exp(-[(-\ln u_1)^\theta + \cdots + (-\ln u_d)^\theta]^{1/\theta}), & \theta &= 1/(1 - (\varrho_\tau)_{ij}), \\
C_\theta^{\text{Cl}}(\mathbf{u}) &= (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}, & \theta &= 2(\varrho_\tau)_{ij}/(1 - (\varrho_\tau)_{ij}),
\end{aligned}
$$

where $(\varrho_\tau)_{ij} = \varrho_\tau(X_{k,i}, X_{k,j})$. Hence parameter estimates for the copulas above are obtained by simply replacing $(\varrho_\tau)_{ij}$ by its estimate $(\widehat{\varrho_\tau})_{ij}$ presented in Section 8.5.

10.8 Gaussian and t-copulas

For Gaussian and t-copulas of high dimensions, it might happen that $\widehat{R}$, with $\widehat{R}_{ij} = \sin(\pi \widehat{\varrho}_{\tau ij}/2)$, is not positive definite. If this is the case, then one has to replace $\widehat{R}$ by a linear correlation matrix R^* which is in some sense *close* to $\widehat{R}$. This can be achieved by the so-called eigenvalue method.

Algorithm 10.6

- *Calculate the spectral decomposition $\widehat{R} = \Gamma \Lambda \Gamma^{\mathrm{T}}$, where Λ is a diagonal matrix of eigenvalues of $\widehat{R}$ and Γ is an orthogonal matrix whose columns are are eigenvectors of $\widehat{R}$.*

- *Replace the negative eigenvalues in Λ by some small value $\delta > 0$ to obtain $\widetilde{\Lambda}$.*

- *Calculate $\widetilde{R} = \Gamma \widetilde{\Lambda} \Gamma^{\mathrm{T}}$ which will be symmetric and positive definite but not necessarily a linear correlation matrix since its diagonal elements might differ from one.*

- *Set $\widehat{R} = D\widetilde{R}D$, where D is a diagonal matrix with $D_{k,k} = 1/\sqrt{\widetilde{R}_{k,k}}$.*

After having estimated R (with $\widehat{R}$ possibly modified to assure positive definiteness) it remains to estimate the degrees of freedom parameter. We construct a *pseudo-sample* $\{\widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n\}$ of observations from the copula by componentwise transformation with the estimated marginal distribution functions $\widehat{F}_1, \ldots, \widehat{F}_d$ as follows.

$$\widehat{\mathbf{U}}_k = (\widehat{F}_1(X_{k,1}), \ldots, \widehat{F}_d(X_{k,d})), \quad k = 1, \ldots, n.$$

Either $\widehat{F}_k$ can be taken as a fitted parametric distribution function or as a version of empirical distribution function:

$$\widehat{F}_k(x) = \widehat{F}_k^{(\beta)}(x) = \frac{1}{n+\beta} \sum_{j=1}^{n} 1_{\{X_{j,k} \le x\}},$$

where $\beta \in (0, 1]$ which guarantees that the pseudo-sample data lies within the unit cube, i.e. that $\widehat{\mathbf{U}}_k \in (0, 1)^d$. Given a pseudo-sample from the t-copula, the degrees of freedom parameter ν can be estimated by maximum likelihood estimation (MLE). A ML estimate of ν is obtained by maximizing

$$\ln L(\xi, \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n) = \sum_{k=1}^{n} \ln c_{\xi, \widehat{R}}(\widehat{\mathbf{U}}_k),$$

with respect to ξ, where $c_{\xi, R}$ denotes the density of a t-copula with ξ as degrees of freedom parameter. The log-likelihood function for the t-copula is given by

$$\ln L(\xi, R, \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n)$$
$$= \sum_{k=1}^{n} \ln g_{\xi, R}(t_\xi^{-1}(\widehat{U}_{k,1}), \ldots, t_\xi^{-1}(\widehat{U}_{k,d})) - \sum_{k=1}^{n} \sum_{j=1}^{d} \ln g_\xi(t_\xi^{-1}(\widehat{U}_{k,j})),$$

where $g_{\xi,R}$ denotes the joint density of a standard t-distribution with distribution function $t_{\xi,R}^d$ and g_ξ denotes the density of a univariate standard t-distribution with distribution function t_ξ. Hence an estimate of the degrees of freedom parameter ν is obtained as the $\xi \in (0,\infty)$ that maximizes the log likelihood function $\ln L(\xi, R, \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n)$.

11 Portfolio credit risk modeling

What is credit/default risk? The following explanation is given by Peter Crosbie and Jeffrey R. Bohn in [5]:

Default risk is the uncertainty surrounding a firm's ability to service its debts and obligations. Prior to default, there is no way to discriminate unambiguously between firms that will default and those that won't. At best we can only make probabilistic assessments of the likelihood of default. As a result, firms generally pay a spread over the default-free rate of interest that is proportional to their default probability to compensate lenders for this uncertainty.

If a firm (obligor) can not fulfill its commitments towards a lender, or counterparty in a financial agreement, then we say that the firm is in default. Credit risk also includes the risk related to events other than default such as up- or down moves in credit rating.

The loss suffered by a lender or counterparty in the event of default is usually significant and is determined largely by the details of the particular contract or obligation. In most cases the obligor is able to repay a substantial amount of the loan, so only a certain fraction of the entire loan is lost. For example, typical loss rates in the event of default for senior secured bonds, subordinated bonds and zero coupon bonds are 49%, 68%, and 81%, respectively.

In this chapter we will introduce a general framework for modeling portfolios subject to credit risk.

11.1 A simple model

Consider a portfolio consisting of n loans (or bonds) subject to default. That the loan is subject to default means that with some probability p_i, obligor i will not be able to repay his debt. Each loan has a certain loan size L_i. If there is a default then the lender does not lose the entire amount L_i but rather a proportion $1 - \lambda_i$ of the loan size. We call $\lambda_i \in [0, 1]$ the recovery rate of loan i. The *loss-given-default* for loan number i which is the amount lost by the lender in the case of default is given by

$$\text{LGD}_i = (1 - \lambda_i)L_i.$$

At some time T, say one year from now, each obligor can be in either of two states, default or nondefault. We model the state of each obligor at time T by a Bernoulli random variable

$$X_i = \begin{cases} 1 & \text{if obligor } i \text{ is in default,} \\ 0 & \text{otherwise.} \end{cases}$$

The default probability of obligor i is then given by $p_i = \text{P}(X_i = 1)$. The total loss at time T due to obligors defaulting is then given by

$$L = \sum_{i=1}^{n} X_i \, \text{LGD}_i = \sum_{i=1}^{n} X_i(1 - \lambda_i)L_i.$$

An important issue in quantitative credit risk management is to understand the distribution of the random variable L. Given that we know the size L_i of each loan we need to model the multivariate random vector $(X_1, \ldots, X_n, \lambda_1, \ldots, \lambda_n)$ in order to derive the loss distribution of L. Most commercial models in use today assume the recovery rates λ_i to be independent of $\mathbf{X} = (X_1, \ldots, X_n)$ and independent of each other. This leaves essentially the joint distribution of default indicators $\mathbf{X}$ to be modeled.

The most simple model we may think of is when all loan sizes are equal $L_i = L_1$, all recovery rates are deterministic and equal $\lambda_i = \lambda_1$ and all default indicators X_i are iid with default probability p. Then the loss is given by $L = \mathrm{LGD}_1 N$, where $N = \sum_{i=1}^{n} X_i$ is Binomial(n, p)-distributed. Below we will study some more sophisticated models for the default indicators $\mathbf{X}$.

11.2   Latent variable models

Since it is practically impossible to obtain historical observations of the default indicator X_i for a given obligor i (it is rather unusual that the firm has defaulted many times before) it is a good idea to divide all obligors into m homogeneous groups. Within each group all obligors (firms) have the same default probability. Estimation of default probabilities can then be based upon how many obligors that have defaulted within each group, leading to larger sample sizes. To this end we may introduce a *state variable* $\mathbf{S} = (S_1, \ldots, S_n)$, where S_i represents the state of the obligor i. We suppose that the state is an integer in the set $\{0, \ldots, m\}$ with $S_i = 0$ indicating that obligor i is in the default state. The other states may be thought of as the obligor being in different rating classes. We let X_i denote the default indicator of obligor i, i.e.

$$X_i = \begin{cases} 0 & \text{if } S_i \neq 0, \\ 1 & \text{if } S_i = 0. \end{cases}$$

The vector $\mathbf{X} = (X_1, \ldots, X_n)$ is the vector of default indicators and the default probability is $p_i = \mathrm{P}(X_i = 1)$.

Often the state variables $\mathbf{S} = (S_1, \ldots, S_n)$ are modeled using a vector of so-called latent variables $\mathbf{Y} = (Y_1, \ldots, Y_n)$; Y_i representing for instance the value of the assets, or asset returns, of obligor i. Typically we have a number of thresholds d_{ij}, $i = 1, \ldots, n$, $j = 0, \ldots, m+1$, with $d_{i0} = -\infty$ and $d_{i(m+1)} = \infty$. The state of S_i is then given through Y_i by

$$S_i = j \text{ if } Y_i \in (d_{ij}, d_{i(j+1)}].$$

Let F_i denote the distribution of Y_i. Default occurs if $Y_i \leq d_{i1}$ and hence the default probability is given by $p_i = F_i(d_{i1})$. The probability that the first k obligors, say, default is then given by (the F_is are assumed to be continuous)

$$\begin{aligned} p_{1\ldots k} &= \mathrm{P}(Y_1 \leq d_{11}, \ldots, Y_k \leq d_{k1}) \\ &= C(F_1(d_{11}), \ldots, F_k(d_{k1}), 1, \ldots, 1) \\ &= C(p_1, \ldots, p_k, 1, \ldots, 1), \end{aligned}$$

where C denotes the copula of $\mathbf{Y}$. As the marginal default probabilities $F_i(d_{i1})$ are small, the joint default probability will depend heavily on the choice of copula C.

Example 11.1 Consider a loan portfolio with $n = 100$ obligors where the credit risk is modeled using a latent variable model with copula C. Suppose that C is an exchangeable copula, i.e. that

$$C(u_1, \ldots, u_n) = C(u_{\pi(1)}, \ldots, u_{\pi(n)}),$$

where π is an arbitrary permutation of $\{1, \ldots, n\}$. Suppose further that the individual default probability of each obligor is equal to $p = 0.15$, i.e. $p_i = p = 0.15$. Let N denote the number of defaults and let $\rho_\tau(Y_i, Y_j) = \tau$, $i \neq j$ denote Kendall's tau between any two latent variables (which are assumed to have continuous distribution functions). We assume that $\tau = 0$ and we simulate the number of defaults 10^5 times and illustrate the distribution of the number of defaults in a histogram when
(a) C is a Gaussian copula and (b) C is a t_4 copula. The histograms are shown in Figure 27. One clearly sees that zero correlation is far from independence if the dependence structure is nonGaussian.

11.3 Mixture models

The random vector $\mathbf{X} = (X_1, \ldots, X_n)$ follows a **Bernoulli mixture model** if there is a random vector $\mathbf{Z} = (Z_1, \ldots, Z_m)$, $m < n$, and functions $f_i : \mathbb{R}^m \to [0, 1]$, $i \in \{1, \ldots, n\}$ such that conditional on $\mathbf{Z}$, $\mathbf{X}$ is a vector of independent Bernoulli random variables with

$$P(X_i = 1 \mid \mathbf{Z}) = f_i(\mathbf{Z}), \quad P(X_i = 0 \mid \mathbf{Z}) = 1 - f_i(\mathbf{Z}).$$

For $\mathbf{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$ we then have

$$P(\mathbf{X} = \mathbf{x} \mid \mathbf{Z}) = \prod_{i=1}^{n} f_i(\mathbf{Z})^{x_i} (1 - f_i(\mathbf{Z}))^{1-x_i}.$$

The unconditional distribution is then given by

$$P(\mathbf{X} = \mathbf{x}) = E(P(\mathbf{X} = \mathbf{x} \mid \mathbf{Z})) = E\Big(\prod_{i=1}^{n} f_i(\mathbf{Z})^{x_i} (1 - f_i(\mathbf{Z}))^{1-x_i} \Big).$$

If all the functions f_i are equal, $f_i = f$, then, conditional on $\mathbf{Z}$, the number of defaults $N = \sum_{i=1}^{m} X_i$ is $\mathrm{Bin}(n, f(\mathbf{Z}))$-distributed.

The random vector $\mathbf{X} = (X_1, \ldots, X_n)$ follows a **Poisson mixture model** if there is a random vector $\mathbf{Z} = (Z_1, \ldots, Z_m)$, $m < n$, and functions $\lambda_i : \mathbb{R}^m \to (0, \infty)$, $i \in \{1, \ldots, n\}$ such that conditional on $\mathbf{Z}$, $\mathbf{X}$ is a vector of independent $\mathrm{Po}(\lambda_i(\mathbf{Z}))$-distributed random variables. In this case we have

$$P(X_i = x_i \mid \mathbf{Z}) = \frac{\lambda_i(\mathbf{Z})^{x_i}}{x_i!} e^{-\lambda_i(\mathbf{Z})}, \quad x_i \in \mathbb{N}.$$
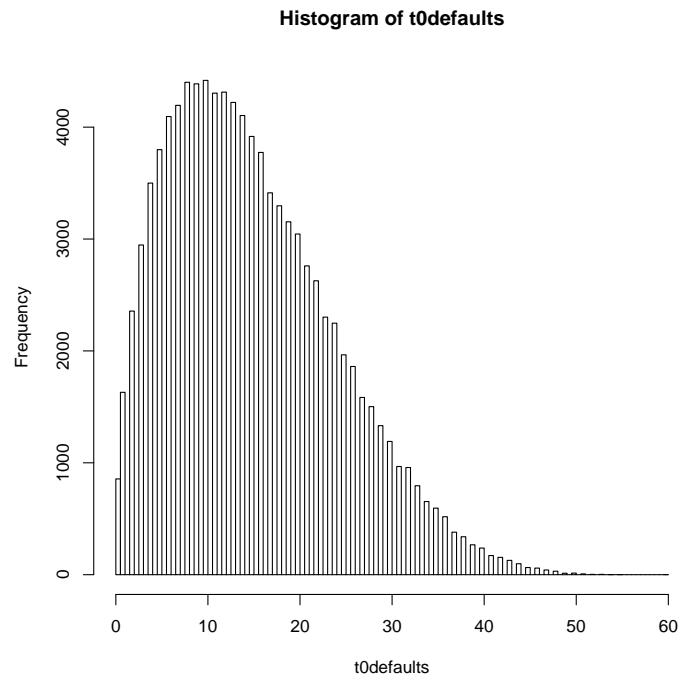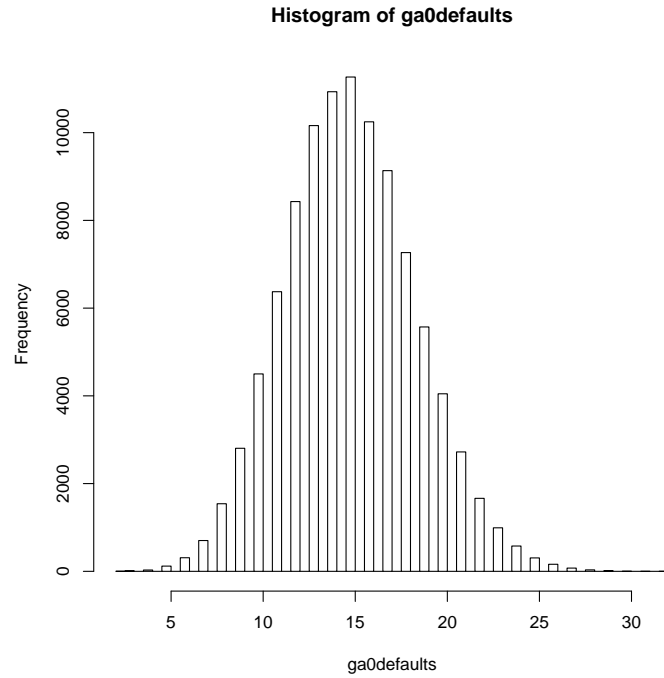
Histogram of ga0defaults

Histogram of t0defaults

Figure 27: Histograms of the number of defaults: (a) Gaussian copula (upper) and (b) t_4 copula (lower).

86

For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{N}^n$ we then have

$$P(\mathbf{X} = \mathbf{x} \mid \mathbf{Z}) = \prod_{i=1}^n \frac{\lambda_i(\mathbf{Z})^{x_i}}{x_i!} e^{-\lambda_i(\mathbf{Z})}.$$

The unconditional distribution is then given by

$$P(\mathbf{X} = \mathbf{x}) = E(P(\mathbf{X} = \mathbf{x} \mid \mathbf{Z})) = E\Big(\prod_{i=1}^n \frac{\lambda_i(\mathbf{Z})^{x_i}}{x_i!} e^{-\lambda_i(\mathbf{Z})} \Big).$$

The use of Poisson mixture models for modeling defaults can be motivated as follows. Suppose that $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_n)$ follows a Poisson mixture model with factors $\mathbf{Z}$. Put $X_i = \mathbb{I}_{[1,\infty)}(\widetilde{X}_i)$. Then $\mathbf{X} = (X_1, \ldots, X_n)$ follows a Bernoulli mixture model with

$$f_i(\mathbf{Z}) = 1 - e^{-\lambda_i(\mathbf{Z})}.$$

If the Poisson parameters $\lambda_i(\mathbf{Z})$ are small then $\widetilde{N} = \sum_{i=1}^n \widetilde{X}_i$ is approximately equal to the number of defaulting obligors and conditional on $\mathbf{Z}$, $\widetilde{N}$ is Poisson($\overline{\lambda}$)-distributed with $\overline{\lambda}(\mathbf{Z}) = \sum_{i=1}^n \lambda_i(\mathbf{Z})$.

Example 11.2 A bank has a loan portfolio of 100 loans. Let X_k be the default indicator for loan k such that $X_k = 1$ in case of default and 0 otherwise. The total number of defaults is $N = X_1 + \cdots + X_{100}$.

(a) Suppose that $X_1, \ldots, X_{100}$ are independent and identically distributed with $P(X_1 = 1) = 0.01$. Compute $E(N)$ and $P(N = k)$ for $k \in \{0, \ldots, 100\}$.

(b) Consider the risk factor Z which reflects the state of the economy. Suppose that conditional on Z, the default indicators are independent and identically distributed with $P(X_1 = 1 \mid Z) = Z$, where

$$P(Z = 0.01) = 0.9 \quad \text{and} \quad P(Z = 0.11) = 0.1.$$

Compute $E(N)$.

(c) Consider the risk factor Z which reflects the state of the economy. Suppose that conditional on Z, the default indicators are independent and identically distributed with

$$P(X_1 = 1 \mid Z) = Z^9,$$

where Z is uniformly distributed on $(0, 1)$. Compute $E(N)$.

Solution (a): We have $N \sim \text{Binomial}(100, 0.01)$. Hence, $E(N) = 100 \cdot 0.01 = 1$ and

$$P(N = k) = \binom{100}{k} 0.01^k 0.99^{100-k}.$$

Solution (b): We have $N \mid Z \sim \text{Binomial}(100, Z)$. Hence,

$$
\begin{aligned}
\mathrm{E}(N) &= \mathrm{E}(\mathrm{E}(N \mid Z)) = \mathrm{E}(100Z) = 100\,\mathrm{E}(Z) \\
&= 100(0.01 \cdot 0.9 + 0.11 \cdot 0.1) = 0.9 + 1.1 = 2.
\end{aligned}
$$

Solution (c): We have $N \mid Z \sim \text{Binomial}(100, Z^9)$. Hence,

$$
\begin{aligned}
\mathrm{E}(N) &= \mathrm{E}(\mathrm{E}(N \mid Z)) = \mathrm{E}(100Z^9) = 100\,\mathrm{E}(Z^9) \\
&= 100 \cdot 0.1 = 10.
\end{aligned}
$$

∎

11.4 One-factor Bernoulli mixture models

In this section we will consider the Bernoulli mixture model where $\mathbf{Z}$ is univariate, $\mathbf{Z} = Z$, i.e. we only have one factor and all the functions f_i are equal, $f_i = f$. This means that all marginal default probabilities are equal and the number of defaults N satisfies $N \mid Z \sim \text{Binomial}(n, f(Z))$. Moreover, the unconditional probability that only the first k obligors defaults is given by

$$
\begin{aligned}
\mathrm{P}(X_1 &= 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) \\
&= \mathrm{E}(\mathrm{P}(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0 \mid Z)) \\
&= \mathrm{E}(f(Z)^k (1 - f(Z))^{n-k}).
\end{aligned}
$$

To determine the unconditional default probabilities, number of defaults, etc. we need to specify the distribution function G of Z. Given G, the unconditional probability that the first k obligors defaults is given by

$$
\mathrm{P}(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_{-\infty}^{\infty} f(z)^k (1 - f(z))^{n-k} G(\mathrm{d}z)
$$

and the number of defaulting obligors N has unconditional distribution

$$
\mathrm{P}(N = k) = \binom{n}{k} \int_{-\infty}^{\infty} f(z)^k (1 - f(z))^{n-k} G(\mathrm{d}z).
$$

Notice also that

$$
\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\,\mathrm{E}(X_j) \\
&= \mathrm{E}(\mathrm{E}(X_i X_j \mid Z)) - \mathrm{E}(\mathrm{E}(X_i \mid Z))\,\mathrm{E}(\mathrm{E}(X_j \mid Z)) \\
&= \mathrm{E}(f(Z)^2) - \mathrm{E}(f(Z))^2 = \mathrm{var}(f(Z)).
\end{aligned}
$$

We have $N = \mathrm{E}(N \mid Z) + N - \mathrm{E}(N \mid Z)$ and

$$
\begin{aligned}
\mathrm{E}(N) &= \mathrm{E}(\mathrm{E}(N \mid Z)) = n\,\mathrm{E}(f(Z)) = np_1, \\
\mathrm{var}(N) &= \mathrm{E}(\mathrm{var}(N \mid Z)) + \mathrm{var}(\mathrm{E}(N \mid Z)) \\
&= \mathrm{E}(nf(Z)(1 - f(Z))) + \mathrm{var}(nf(Z)) \\
&= n\,\mathrm{E}(f(Z)(1 - f(Z))) + n^2\,\mathrm{var}(f(Z)).
\end{aligned}
$$

Notice that by Markov's inequality

$$P(|N/n - f(Z)| > \varepsilon \mid Z) \leq \frac{\mathrm{var}(N/n \mid Z)}{\varepsilon^2} = \frac{f(Z)(1 - f(Z))}{n\varepsilon^2}.$$

Hence, for every $\varepsilon > 0$,

$$P(|N/n - f(Z)| > \varepsilon) = \mathrm{E}(P(|N/n - f(Z)| > \varepsilon \mid Z)) \leq \frac{\mathrm{E}(f(Z)(1 - f(Z)))}{n\varepsilon^2}$$

Hence, $N/n \xrightarrow{\mathrm{P}} f(Z)$ as $n \to \infty$ which justifies the approximation $N/n \approx f(Z)$ for n large. (In fact it hold that $N/n \to f(Z)$ a.s. as $n \to \infty$.)

11.5 Probit normal mixture models

In several portfolio credit risk models (see the section about the KMV model below) the default indicators X_i, $i = 1, \ldots, n$, have the representation $X_i = 1$ if and only if $\sqrt{\varrho}Z + \sqrt{1 - \varrho}W_i \leq d_{i1}$, where $\varrho \in [0, 1]$ and $Z, W_1, \ldots, W_n$ are iid and standard normally distributed. Assuming equal individual default probabilities $p = P(X_i = 1)$ we have $d_{i1} = \Phi^{-1}(p)$ and hence

$$X_i = \mathbb{I}_{(-\infty, \Phi^{-1}(p)]}(\sqrt{\varrho}Z + \sqrt{1 - \varrho}W_i).$$

This gives

$$f(Z) = P(X_i = 1 \mid Z) = P(\sqrt{\varrho}Z + \sqrt{1 - \varrho}W_i \leq \Phi^{-1}(p) \mid Z)$$
$$= \Phi\left(\frac{\Phi^{-1}(p)}{\sqrt{1 - \varrho}} + \frac{\sqrt{\varrho}Z}{\sqrt{1 - \varrho}}\right).$$

This leads to

$$\mathrm{VaR}_q(f(Z)) = \Phi\left(\frac{\sqrt{\varrho}}{\sqrt{1 - \varrho}}\Phi^{-1}(q) + \frac{1}{\sqrt{1 - \varrho}}\Phi^{-1}(p)\right).$$

Setting $q = 0.999$ and using the approximation $N/n \approx f(Z)$ motivated above, we arrive at the "Basel formula" for capital requirement as a fraction of the total exposure for a homogeneous portfolio with individual default probabilities p:

$$\text{Capital requirement} = c_1 c_2 \left[\Phi\left(\frac{\sqrt{\varrho}}{\sqrt{1 - \varrho}}\Phi^{-1}(0.999) + \frac{1}{\sqrt{1 - \varrho}}\Phi^{-1}(p)\right) - p\right],$$

where c_1 is the fraction of the exposure lost in case of default and c_2 is a constant for maturity adjustments. The asset return correlation coefficient ϱ is assigned a value that depends on the asset type and also the size and default probability of the borrowers.

11.6 Beta mixture models

For the Beta mixing distribution we assume that $Z \sim \text{Beta}(a, b)$ and $f(z) = z$. It has density

$$g(z) = \frac{1}{\beta(a, b)} z^{a-1}(1 - z)^{b-1}, a, b > 0, z \in (0, 1),$$

where

$$\beta(a, b) = \int_0^1 z^{a-1}(1 - z)^{b-1} \mathrm{d}z = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

and hence, using that $\Gamma(z + 1) = z\Gamma(z)$,

$$\begin{aligned}
\mathrm{E}(Z) &= \frac{1}{\beta(a, b)} \int_0^1 z^a (1 - z)^{b-1} \mathrm{d}z = \frac{\beta(a + 1, b)}{\beta(a, b)} \\
&= \frac{\Gamma(a + 1)\Gamma(b)}{\Gamma(a + b + 1)} \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a + b}, \\
\mathrm{E}(Z^2) &= \frac{a(a + 1)}{(a + b)(a + b + 1)}.
\end{aligned}$$

We immediately get that the number of defaults N has distribution

$$\begin{aligned}
\mathrm{P}(N = k) &= \binom{n}{k} \int_0^1 z^k (1 - z)^{n-k} g(z) \mathrm{d}z \\
&= \binom{n}{k} \frac{1}{\beta(a, b)} \int_0^1 z^{a+k-1}(1 - z)^{n-k+b-1} \mathrm{d}z \\
&= \binom{n}{k} \frac{\beta(a + k, b + n - k)}{\beta(a, b)},
\end{aligned}$$

which is called the beta-binomial distribution. This probability function is illustrated in Figure 28. The expected number of defaults is easily computed.

$$\mathrm{E}(N) = \mathrm{E}(\mathrm{E}(N \mid Z)) = n\, \mathrm{E}(\mathrm{E}(X_1 \mid Z)) = n\, \mathrm{E}(Z) = n\frac{a}{a + b}.$$

If we have estimated the default probabilities $\mathrm{P}(X_i = 1)$ and $\mathrm{P}(X_i = X_j = 1)$, $i \neq j$, then the parameters a and b can be determined from the relations

$$\mathrm{P}(X_i = 1) = \mathrm{E}(Z) = \frac{a}{a + b}, \quad \mathrm{P}(X_i = X_j = 1) = \mathrm{E}(Z^2) = \frac{a(a + 1)}{(a + b)(a + b + 1)}.$$

Moreover, the linear correlation coefficient is $\varrho_L(X_i, X_j) = (a + b + 1)^{-1}$. If we specify the individual default probability p and the linear correlation coefficient ϱ, then we obtain the parameters a and b of the Beta distribution as functions of (p, ϱ):

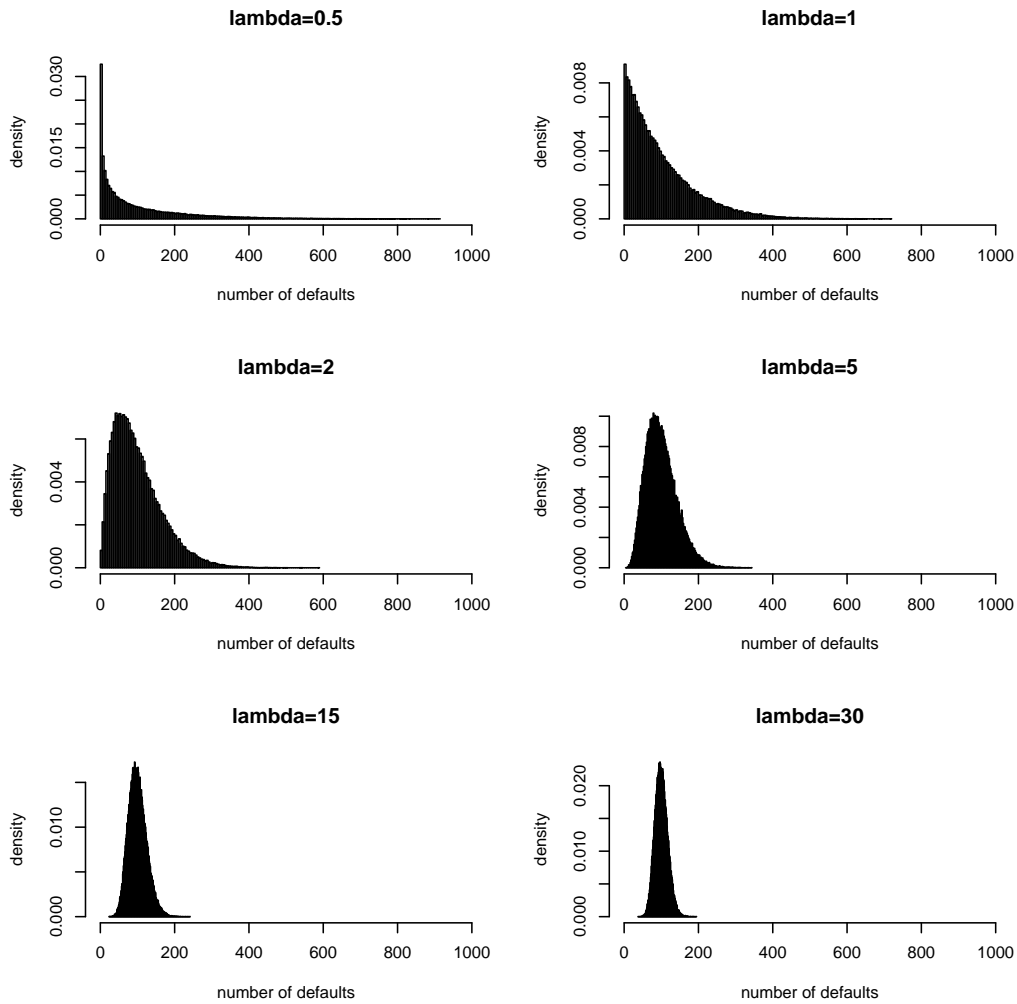$$a = (1 - p)\frac{1 - \varrho}{\varrho}, \quad b = p\frac{1 - \varrho}{\varrho}.$$

Figure 28: The probability function for the number of defaults in a Beta mixture model with $n = 1000$ obligors and $(a, b) = \lambda(1, 9)$.

Example 11.3 Notice that if we only specify the individual default probability $p = P(X_i = 1)$, then we know very little about the model. For example, $p = 0.01 = a/(a+b)$ for $(a,b) = (0.01, 0.99), (0.1, 9.9), (1, 99)$, but the different choices of (a, b) leads to quite different models. This is shown in the table below which considers a portfolio with $n = 1000$ obligors.

(a, b)	p	$\mathrm{corr}(X_i, X_j)$	$\mathrm{VaR}_{0.99[9]}(N)$	$\mathrm{VaR}_{0.99[9]}(nZ)$
(1,99)	0.01	0.01	47 [70]	45 [67]
(0.1,9.9)	0.01	0.09	155 [300]	155 [299]
(0.01,0.99)	0.01	0.5	371 [908]	371 [908]

Notice also how accurate the approximation $N \approx nZ$ is!

Both Example 11.3 and Figure 28 illustrate that only specifying the individual default probability p says very little about the distribution of N. Notice that every choice of $(a, b) = \lambda(1, (1-p)/p)$, $\lambda > 0$, gives default probability p. Let Z_λ be $\mathrm{Beta}(\lambda, \lambda(1-p)/p)$-distributed. Then, for every $\varepsilon > 0$,

$$P(|Z_\lambda - p| > \varepsilon) \leq \frac{\mathrm{var}(Z_\lambda)}{\varepsilon^2} = \frac{p}{\varepsilon^2} \left(\frac{\lambda p + p}{\lambda + p} - p \right) \to 0 \text{ as } \lambda \to \infty.$$

Hence, $Z_\lambda \to p$ in probability as $\lambda \to \infty$. This implies here that, with N_λ being the total number of defaults,

$$P(N_\lambda = m) = E\left(\binom{n}{m} Z_\lambda^m (1 - Z_\lambda)^{n-m} \right) \to \binom{n}{m} p^m (1-p)^{n-m} \text{ as } \lambda \to \infty,$$

or equivalently that N_λ converges in distribution to a $\mathrm{Binom}(n, p)$-distributed random variable as $\lambda \to \infty$. This is also seen in Figure 28.

12 Popular portfolio credit risk models

In this chapter we will present the commercial models currently used by practitioners such as the KMV model and CreditRisk$^+$. Interesting comparisons of these models are given in [6] and [10].

12.1 The KMV model

A popular commercial model for credit risk is the so-called KMV model provided by Moody's KMV (www.moodyskmv.com). It is an example of a latent variable model where the state variables $\mathbf{S} = (S_1, \ldots, S_n)$ can only have two states ($m = 1$). The latent variables $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are related to the value of the assets of each firm in the following way.

The Merton model

It is assumed that the balance sheet of each firm consist of assets and liabilities. The liabilities are divided into debt and equity. The value of the assets of the ith firm at time T is denoted by $V_{A,i}(T)$, the value of the debt by K_i and the value of the equity of the firm at time T by $V_{E,i}(T)$. It is assumed that the future asset value is modeled by a geometric Brownian motion

$$V_{A,i}(T) = V_{A,i}(t) \exp \left\{ \left(\mu_{A,i} - \frac{\sigma_{A,i}^2}{2} \right)(T - t) + \sigma_{A,i} \left(W_i(T) - W_i(t) \right) \right\} \quad (12.1)$$

where $\mu_{A,i}$ is the drift, $\sigma_{A,i}$ the volatility and $(W_i(t); 0 \leq t \leq T)$ a Brownian motion. In particular this means that $W_i(T) - W_i(t) \sim \mathrm{N}(0, T - t)$ and hence that $\ln V_{A,i}(T)$ is normal with mean $\ln V_{A,i}(t) + (\mu_{A,i} - \sigma_{A,i}^2/2)(T - t)$ and variance $\sigma_{A,i}^2(T - t)$. The firm defaults if at time T the value of the assets are less than the value of the debt. That is, the default indicator X_i is given by

$$X_i = \mathbb{I}_{(-\infty, K_i)}(V_{A,i}(T)).$$

Writing

$$Y_i = \frac{W_i(T) - W_i(t)}{\sqrt{T - t}}$$

we get $Y_i \sim \mathrm{N}(0, 1)$ and

$$X_i = \mathbb{I}_{(-\infty, K_i)}(V_{A,i}(T)) = \mathbb{I}_{(-\infty, -\mathrm{DD}_i)}(Y_i)$$

with

$$-\mathrm{DD}_i = \frac{\ln K_i - \ln V_{A,i}(t) + (\sigma_{A,i}^2/2 - \mu_{A,i})(T - t)}{\sigma_{A,i}\sqrt{T - t}}.$$

The quantity DD_i is called the *distance-to-default*. In principle the default probability can then be computed as $\mathrm{P}(V_{A,i}(T) < K_i) = \mathrm{P}(Y_i < -\mathrm{DD}_i)$. Hence, in the general setup of a two state latent variable model we have $Y_i \sim \mathrm{N}(0, 1)$ and default thresholds $d_{i1} = -\mathrm{DD}_i$.

Computing the distance-to-default

To compute the distance-to-default we need to find $V_{A,i}(t)$, $\sigma_{A,i}$ and $\mu_{A,i}$. A problem here is that the value of the firms assets $V_{A,i}(t)$ can not be observed directly. However, the value of the firms equity can be observed by looking at the market stock prices. KMV therefore takes the following viewpoint: the equity holders have the right but not the obligation, at time T, to pay off the holders of the other liabilities and take over the remaining assets of the firm. That is, the debt holders own the firm until the debt is paid off in full by the equity holders. This can be viewed as a call option on the firm's assets with a strike price equal to the debt. That is, at time T we have the relation

$$V_{E,i}(T) = \max(V_{A,i}(T) - K_i, 0).$$

The value of equity at time t, $V_{E,i}(t)$, can then be thought of as the price of a call option with the value of assets as underlying and strike price K_i. Under some simplifying assumptions the price of such an option can be computed using the Black-Scholes option pricing formula. This gives

$$V_{E,i}(t) = C(V_{A,i}(t), \sigma_{A,i}, r),$$

where

$$C(V_{A,i}(t), \sigma_{A,i}, r) = V_{A,i}(t)\Phi(e_1) - K_i e^{-r(T-t)}\Phi(e_2),$$
$$e_1 = \frac{\ln V_{A,i}(t) - \ln K_i + (r + \sigma_{A,i}^2/2)(T-t)}{\sigma_{A,i}\sqrt{T-t}}$$
$$e_2 = e_1 - \sigma_{A,i}\sqrt{T-t},$$

Φ is the distribution function of the standard normal and r is the risk free interest rate (investors use e.g. the interest rate on a three-month U.S. Treasury bill as a proxy for the risk-free rate, since short-term government-issued securities have virtually zero risk of default). KMV also introduces a relation between the volatility $\sigma_{E,i}$ of $V_{E,i}$ and the volatility $\sigma_{A,i}$ of $V_{A,i}$ by

$$\sigma_{E,i} = g(V_{A,i}(t), \sigma_{A,i}, r),$$

where g is some function. Using observed/estimated values of $V_{E,i}(t)$ and $\sigma_{E,i}$ the relation

$$V_{E,i}(t) = C(V_{A,i}(t), \sigma_{A,i}, r)$$
$$\sigma_{E,i} = g(V_{A,i}(t), \sigma_{A,i}, r)$$

is inverted to obtain $V_{A,i}(t)$ and $\sigma_{A,i}$ which enables computation of the distance-to-default DD_i.

The expected default frequency

To find the default probability corresponding to the distance-to-default DD_i KMV do not actually use the probability $\mathrm{P}(Y_i < -\mathrm{DD}_i)$. Instead they use

historical data to search for all companies which at some stage in their history had approximately the same distance-to-default. Then the observed default frequency is converted into an actual probability. In the terminology of KMV this estimated default probability p_i is called *Expected Default Frequency* (EDF).

To summarize: In order to compute the probability of default with the KMV model the following steps are required:

(i) Estimate asset value and volatility. The asset value and asset volatility of the firm is estimated from the market value and volatility of equity and the book of liabilities.

(ii) Calculate the distance-to-default.

(iii) Calculate the default probability using the empirical distribution relating distance-to-default to a default probability.

The multivariate KMV model

In the Merton model above we did not introduce any dependence between the value of the assets for different firms. We only considered each firm separately. To compute joint default probabilities and the distribution of the total credit loss it is natural to introduce dependence between the default indicators by making the asset value processes $V_{A,i}$ dependent. The following methodology is used by KMV. Let $(W_j(t) : 0 \leq t \leq T, j = 1, \ldots, m)$ be m independent standard Brownian motions. The evolution (12.1) of asset i is then replaced by

$$V_{A,i}(T) = V_{A,i}(t) \exp \left\{ \left(\mu_{A,i} - \frac{\sigma_{A,i}^2}{2} \right)(T - t) + \sum_{j=1}^{m} \sigma_{A,i,j} \left(W_j(T) - W_j(t) \right) \right\},$$

where

$$\sigma_{A,i}^2 = \sum_{j=1}^{m} \sigma_{A,i,j}^2.$$

Here, $\sigma_{A,i,j}$ gives the magnitude of which asset i is influenced by the jth Brownian motion. The event $V_{A,i}(T) < K_i$ that company i defaults is equivalent to

$$\sum_{j=1}^{m} \sigma_{A,i,j}(W_j(T) - W_j(t)) < \ln K_i - \ln V_{A,i}(t) + \left(\frac{\sigma_{A,i}^2}{2} - \mu_{A,i} \right)(T - t).$$

If we let

$$Y_i = \frac{\sum_{j=1}^{m} \sigma_{A,i,j}(W_j(T) - W_j(t))}{\sigma_{A,i}\sqrt{T - t}},$$

then $\mathbf{Y} = (Y_1, \ldots, Y_n) \sim \mathrm{N}_n(\mathbf{0}, \Sigma)$ with

$$\Sigma_{ij} = \frac{\sum_{k=1}^{m} \sigma_{A,i,k}\sigma_{A,j,k}}{\sigma_{A,i}\sigma_{A,j}}$$

and the above inequality can be written as

$$Y_i < \underbrace{\frac{\ln K_i - \ln V_{A,i}(t) + \left(\frac{\sigma_{A,i}^2}{2} - \mu_{A,i}\right)(T-t)}{\sigma_{A,i}\sqrt{T-t}}}_{-\,\mathrm{DD}_i}.$$

Hence, in the language of a general latent variable model the probability that the first k firms default is given by

$$\mathrm{P}(X_1 = 1, \ldots, X_k = 1) = \mathrm{P}(Y_1 < -\,\mathrm{DD}_1, \ldots, Y_k < -\,\mathrm{DD}_k)$$
$$= C_\Sigma^{Ga}(\Phi(-\,\mathrm{DD}_1), \ldots, \Phi(-\,\mathrm{DD}_k), 1 \ldots, 1),$$

where C_Σ^{Ga} is the copula of a multivariate normal distribution with covariance matrix Σ. As in the univariate case KMV do not use that default probability resulting from the latent variable model but instead use the expected default frequencies EDF_i. In a similar way KMV use the joint default frequency

$$\mathrm{JDF}_{1\ldots k} = C_\Sigma^{Ga}(\mathrm{EDF}_1, \ldots, \mathrm{EDF}_k, 1 \ldots, 1),$$

as the default probability of the first k firms.

Estimating the correlations

Estimating the correlations of the latent variables in $\mathbf{Y}$ is not particularly easy as the dimension n is typically very large and there is limited available historical data. Moreover, estimating pairwise correlations will rarely give a positive definite correlation matrix if the dimension is large. A way around these problems is to use a factor model where the asset value, or more precisely the latent variables $\mathbf{Y}$ is divided into k key factors and one firm specific factor. The key factors are typically macro-economic factors such as

- Global economic effects

- Regional economic effects

- Sector effects

- Country specific effects

- Industry specific effects

If we write

$$Y_i \stackrel{d}{=} \sum_{j=1}^k a_{ij} Z_j + b_i U_i, \quad i = 1, \ldots, n,$$

where $\mathbf{Z} = (Z_1, \ldots, Z_k) \sim \mathrm{N}_k(\mathbf{0}, \Lambda)$ is independent of $\mathbf{U} = (U_1, \ldots, U_n) \sim \mathrm{N}_n(\mathbf{0}, \mathbf{I})$. then the covariance matrix of the right hand side is given by $A\Lambda A^{\mathrm{T}} + D$ where $A_{ij} = a_{ij}$ and D is a diagonal $(n \times n)$ matrix with entries $D_{ii} = b_i^2$.

96

12.2 CreditRisk$^+$ – a Poisson mixture model

This material presented here on the CreditRisk$^+$ model is based on [6], [10] and [4].

CreditRisk+ is a commercial model for credit risk developed by Credit Suisse First Boston and is an example of a Poisson mixture model. The risk factors $Z_1, \ldots, Z_m$ are assumed to be independent $Z_j \sim \text{Gamma}(\alpha_j, \beta_j)$ and we have

$$\lambda_i(\mathbf{Z}) = \overline{\lambda}_i \sum_{j=1}^{m} a_{ij} Z_j, \quad \sum_{j=1}^{m} a_{ij} = 1, a_{ij} \geq 0.$$

for $i = 1, \ldots, n$. Here $\overline{\lambda}_i > 0$ are constants. The density of Z_j is given by

$$f_j(z) = \frac{z^{\alpha_j - 1} \exp\{-z/\beta_j\}}{\beta_j^{\alpha_j} \Gamma(\alpha_j)}.$$

The parameters α_j, β_j are chosen so that $\alpha_j \beta_j = 1$ and then $\text{E}(Z_j) = 1$ and $E(\lambda_i(\mathbf{Z})) = \overline{\lambda}_i$. Notice that the expected number of defaults, $\text{E}(N)$, is given by

$$\text{E}(N) = \text{E}(\text{E}(N \mid \mathbf{Z})) = \sum_{i=1}^{n} \text{E}(\text{E}(X_i \mid \mathbf{Z}))$$

$$= \sum_{i=1}^{n} \text{E}(\lambda_i(\mathbf{Z})) = \sum_{i=1}^{n} \overline{\lambda}_i \sum_{j=1}^{m} a_{ij} \text{E}(Z_j) = \sum_{i=1}^{n} \overline{\lambda}_i.$$

The loss-given-default LGD_i of obligor i is modeled as a constant fraction $1 - \lambda_i$ of the loan size L_i,

$$\text{LGD}_i = (1 - \lambda_i)L_i, \quad i = 1, \ldots, n.$$

Here λ_i is the (deterministic) expected recovery rate. Each loss amount is then expressed as an integer multiple v_i of a fixed base unit of loss (e.g. one million dollars) denoted L_0. Then we have

$$\text{LGD}_i = (1 - \lambda_i)L_i \approx \left[\frac{(1 - \lambda_i)L_i}{L_0}\right] L_0 = v_i L_0, \quad i = 1, \ldots, n,$$

where $[x]$ denotes the nearest integer of x $(x - [x] \in (-1/2, 1/2])$. In this way every LGD_i can be expressed as a fixed integer multiple v_i of a predefined base unit of loss L_0. The main idea here is to approximate the total loss distribution by a discrete distribution. For this discrete distribution it is possible to compute its probability generating function (pgf) g.

Recall the definition of the pgf for a discrete random variable Y with values in $\{y_1, \ldots, y_m\}$,

$$g_Y(t) = \text{E}(t^Y) = \sum_{i=1}^{m} t^{y_i} \text{P}(Y = y_i)$$

Recall the following formulas for probability generating functions.

(i) If $Y \sim \text{Bernoulli}(p)$ then $g_Y(t) = 1 + p(t-1)$.

(ii) If $Y \sim \text{Poisson}(\lambda)$ then $g_Y(t) = \exp\{\lambda(t-1)\}$.

(iii) If $X_1, \ldots, X_n$ are independent random variables then

$$g_{X_1 + \cdots + X_n}(t) = \prod_{i=1}^n g_{X_i}(t).$$

(iv) Let Y have density f and let $g_{X|Y=y}(t)$ be the pgf of $X|Y = y$. Then

$$g_X(t) = \int g_{X|Y=y}(t) f(y) \mathrm{d}y.$$

(v) If X has pgf $g_X(t)$ then

$$P(X = k) = \frac{1}{k!} g^{(k)}(0), \quad \text{with } g^{(k)}(t) = \frac{\mathrm{d}^k g(t)}{\mathrm{d}t^k}.$$

The pgf of the loss distribution

Let us derive the pgf of the loss distribution

$$L = \sum_{i=1}^n X_i v_i L_0.$$

First we determine the conditional pgf of the number of defaults $N = X_1 + \cdots + X_n$ given $\mathbf{Z} = (Z_1, \ldots, Z_m)$. Given $\mathbf{Z}$ the default intensities $\lambda_1(\mathbf{Z}), \ldots, \lambda_n(\mathbf{Z})$ are known so conditional on $\mathbf{Z}$ the default indicators are independent and $\text{Poisson}(\lambda_i(\mathbf{Z}))$-distributed. Hence

$$g_{X_i|\mathbf{Z}}(t) = \exp\{\lambda_i(\mathbf{Z})(t-1)\}, \quad i = 1, \ldots, n.$$

For N we now obtain

$$g_{N|\mathbf{Z}}(t) = \prod_{i=1}^n g_{X_i|\mathbf{Z}} = \prod_{i=1}^n \exp\{\lambda_i(\mathbf{Z})(t-1)\} = \exp\{\mu(t-1)\},$$

$$\mu = \sum_{i=1}^n \lambda_i(\mathbf{Z}).$$

Next we use (iv) to derive the unconditional distribution of the number of defaults N.

$$g_N(t) = \int_0^\infty \cdots \int_0^\infty g_{N|\mathbf{Z}=(z_1,\ldots,z_m)}(t) f_1(z_1) \cdots f_m(z_m) dz_1 \cdots dz_m$$

$$= \int_0^\infty \cdots \int_0^\infty \exp\left\{(t-1)\sum_{i=1}^n \left(\overline{\lambda}_i \sum_{j=1}^m a_{ij} z_j\right)\right\} f_1(z_1) \cdots f_m(z_m) dz_1 \cdots dz_m$$

$$= \int_0^\infty \cdots \int_0^\infty \exp\left\{(t-1)\sum_{j=1}^m \left(\underbrace{\sum_{i=1}^n \overline{\lambda}_i a_{ij}}_{\mu_j}\right) z_j\right\} f_1(z_1) \cdots f_m(z_m) dz_1 \cdots dz_m$$

$$= \int_0^\infty \cdots \int_0^\infty \exp\{(t-1)\mu_1 z_1\} f_1(z_1) dz_1 \cdots \exp\{(t-1)\mu_m z_m\} f_m(z_m) dz_m$$

$$= \prod_{j=1}^m \int_0^\infty \exp\{z\mu_j(t-1)\} \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} z^{\alpha_j-1} \exp\{-z/\beta_j\} dz.$$

Each of the integrals in the product can be computed as

$$\int_0^\infty \frac{1}{\beta_j^{\alpha_j}} \exp\{z\mu_j(t-1)\} z^{\alpha_j-1} \exp\{-z/\beta_j\} dz$$

$$= \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} \int_0^\infty z^{\alpha_j-1} \exp\{-z(\beta_j^{-1} - \mu_j(t-1)\} dz$$

$$= \left\{ u = z(\beta_j^{-1} - \mu_j(t-1)) \right\}$$

$$= \frac{\Gamma(\alpha_j)}{\beta_j^{\alpha_j}\Gamma(\alpha_j)(\beta_j^{-1} - \mu_j(t-1))^{\alpha_j}} \int_0^\infty \frac{1}{\Gamma(\alpha_j)} u^{\alpha_j-1} \exp\{-u\} du$$

$$= \frac{1}{(1+\beta_j\mu_j(t-1))^{\alpha_j}}$$

$$= \left(\frac{1-\delta_j}{1-\delta_j t}\right)^{\alpha_j},$$

where $\delta_j = \beta_j\mu_j/(1+\beta_j\mu_j)$. Finally we obtain

$$g_N(t) = \prod_{j=1}^m \left(\frac{1-\delta_j}{1-\delta_j t}\right)^{\alpha_j}. \tag{12.2}$$

Similar computations will lead us to the pgf of the loss distribution. Conditional on $\mathbf{Z}$ the loss of obligor i is given by

$$L_i|\mathbf{Z} = v_i(X_i|\mathbf{Z})$$

Since the variables $X_i|\mathbf{Z}$, $i = 1,\ldots,n$, are independent so are the variables $L_i|\mathbf{Z}$, $i = 1,\ldots,n$. The pgf of $L_i|\mathbf{Z}$ is

$$g_{L_i|\mathbf{Z}}(t) = \mathrm{E}(t^{L_i}|\mathbf{Z}) = \mathrm{E}(t^{v_i X_i}|\mathbf{Z}) = g_{X_i|\mathbf{Z}}(t^{v_i}).$$

Hence, the pgf of the total loss conditional on $\mathbf{Z}$ is

$$g_{L|\mathbf{z}}(t) = g_{L_1 + \cdots + L_n | \mathbf{z}}(t) = \prod_{i=1}^{n} g_{L_i | \mathbf{z}}(t) = \prod_{i=1}^{n} g_{X_i | \mathbf{z}}(t^{v_i})$$

$$= \exp\left\{ \sum_{j=1}^{m} Z_j \Big(\sum_{i=1}^{n} \overline{\lambda}_i a_{ij}(t^{v_i} - 1) \Big) \right\}.$$

Similar to the previous computation we obtain

$$g_L(t) = \prod_{j=1}^{m} \Big(\frac{1 - \delta_j}{1 - \delta_j \Lambda_j(t)} \Big)^{\alpha_j}, \qquad \Lambda_j(t) = \frac{1}{\mu_j} \sum_{i=1}^{n} \overline{\lambda}_i a_{ij} t^{v_i}$$

with μ_j and δ_j as above. The loss distribution is then obtained by inverting the pgf. In particular, the pgf $g_L(t)$ uniquely determines the loss distribution.

Example 12.1 Consider a portfolio that consists of $n = 100$ obligors. Recall the probability generating function (12.2) for the number of defaults N in the CreditRisk$^+$ model. In order to compute the probabilities $P(N = k)$, $k \geq 0$, we need to compute the derivatives

$$g_N^{(k)}(0) = \frac{\mathrm{d}^k g_N}{\mathrm{d}t^k}(0).$$

It can be shown that $g_N^{(k)}(0)$ satisfies the recursion formula

$$g_N^{(k)}(0) = \sum_{l=0}^{k-1} \binom{k-1}{l} g_N^{(k-1-l)}(0) \sum_{j=1}^{m} l! \, \alpha_j \, \delta_j^{l+1}, \quad k \geq 1$$

(show this!) Assume that $\overline{\lambda}_i = \overline{\lambda} = 0.15$, $\alpha_j = \alpha = 1$, $\beta_j = \beta = 1$, $a_{ij} = a = 1/m$. To better understand the model we plot the function $P(N = k)$ for $k = 0, \ldots, 100$ when $m = 1$ and when $m = 5$. The result is shown in Figure 29.

One can interpret the plot as follows. With only one risk factor, $m = 1$, to which all default indicators are linked, we either have many default or we have few defaults. Having approximately $E(N) = 15$ defaults is unlikely. With $m = 5$ independent risk factors there is a diversification effect. In this case it is most likely that we have approximately $E(N) = 15$ defaults.

Example 12.2 Consider a homogeneous portfolio with 100 loans and let N be the total number of defaults one year from now. To model the default risk we consider the CreditRisk+ model with one single Gamma(α, β)-distributed risk factor Z with density function

$$f_Z(z) = \frac{z^{\alpha-1} e^{-z/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad z > 0, \alpha > 0, \beta > 0,$$
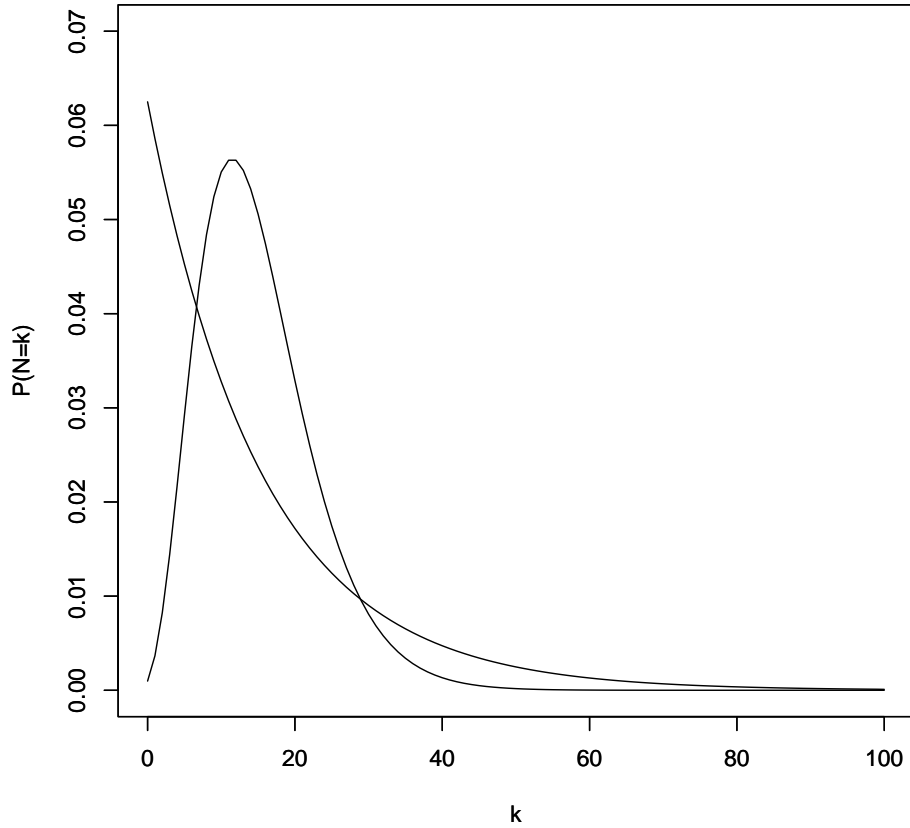
Figure 29: $P(N = k)$ for $k = 0, \ldots, 100$ for $m = 1$ and $m = 5$. For $m = 1$ the probability $P(N = k)$ is decreasing in k, for $m = 5$ the probability $P(N = k)$ first increasing in k and then decreasing.

and mean $E(Z) = \alpha\beta$. We assume that the CreditRisk+ model is chosen so that it is a Poisson mixture model with $\lambda_i(z) = z/100$ for $i = 1, \ldots, 100$. Moreover, $\alpha = 1/\beta$.

Notice that the $X_i \mid Z$ are Poisson$(Z/100)$-distributed and independent.

Hence, $N \mid Z$ is Poisson(Z)-distributed and

$$P(N = k) = E(P(N = k \mid Z)) = \int_0^\infty P(N = k \mid Z = z) f_Z(z) \mathrm{d}z$$

$$= \frac{1}{k!} \int_0^\infty \frac{z^{\alpha + (k-1)} e^{-z(1+1/\beta)}}{\beta^\alpha \Gamma(\alpha)} \mathrm{d}z$$

$$= \frac{1}{k!} \frac{\widetilde{\beta}^{\widetilde{\alpha}} \Gamma(\widetilde{\alpha})}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{z^{\widetilde{\alpha}-1} e^{-z/\widetilde{\beta}}}{\widetilde{\beta}^{\widetilde{\alpha}} \Gamma(\widetilde{\alpha})} \mathrm{d}z$$

$$= \frac{1}{k!} \frac{\widetilde{\beta}^{\widetilde{\alpha}} \Gamma(\widetilde{\alpha})}{\beta^\alpha \Gamma(\alpha)},$$

where $\widetilde{\alpha} = \alpha + k$ and $\widetilde{\beta} = \beta/(1+\beta)$. Hence,

$$P(N = 0) = (1+\beta)^{-\alpha} = (1+\beta)^{-1/\beta}.$$

For $k \geq 1$, $\Gamma(\alpha + k) = (\alpha + k - 1) \cdot \cdots \cdot \alpha \Gamma(\alpha)$ so

$$P(N = k) = \frac{1}{k!} \beta^k (1+\beta)^{-\alpha - k} \prod_{j=0}^{k-1} (\alpha + j)$$

$$\frac{1}{k!} \beta^k (1+\beta)^{-1/\beta - k} \prod_{j=0}^{k-1} [\beta^{-1}(1 + j\beta)]$$

$$= \frac{1}{k!} (1+\beta)^{-1/\beta - k} \prod_{j=0}^{k-1} (1 + j\beta).$$

Similarly, the individual default probability and the probability of joint default is computed as

$$P(X_i \geq 1) = 1 - P(X_i = 0) = 1 - E(P(X_i = 0 \mid Z))$$

$$= 1 - E(e^{-Z/100}) = 1 - \left(\frac{100}{100 + \beta}\right)^{1/\beta},$$

$$P(X_i \geq 1, X_j \geq 1) = E((1 - e^{-Z/100})^2)$$

$$= 1 - 2\left(\frac{100}{100 + \beta}\right)^{1/\beta} + \left(\frac{50}{50 + \beta}\right)^{1/\beta}.$$

If $\widetilde{X}_i = 1_{[1,\infty)}(X_i)$ is the default indicator for obligor i in the corresponding Bernoulli mixture model, then

$$P(\widetilde{X}_i = 1) = P(X_i \geq 1),$$

$$P(\widetilde{X}_i = 1, \widetilde{X}_j = 1) = P(X_i \geq 1, X_j \geq 1),$$

$$\varrho_L(\widetilde{X}_i, \widetilde{X}_j) = \frac{P(\widetilde{X}_i = 1, \widetilde{X}_j = 1) - P(\widetilde{X}_i = 1)^2}{P(\widetilde{X}_i = 1) - P(\widetilde{X}_i = 1)^2},$$

102

and these quantities can be computed from the above expressions. Notice that

$$P(\widetilde{X}_i = 1) = 1 - \left(\frac{100}{100 + \beta}\right)^{1/\beta} \approx 0.01$$

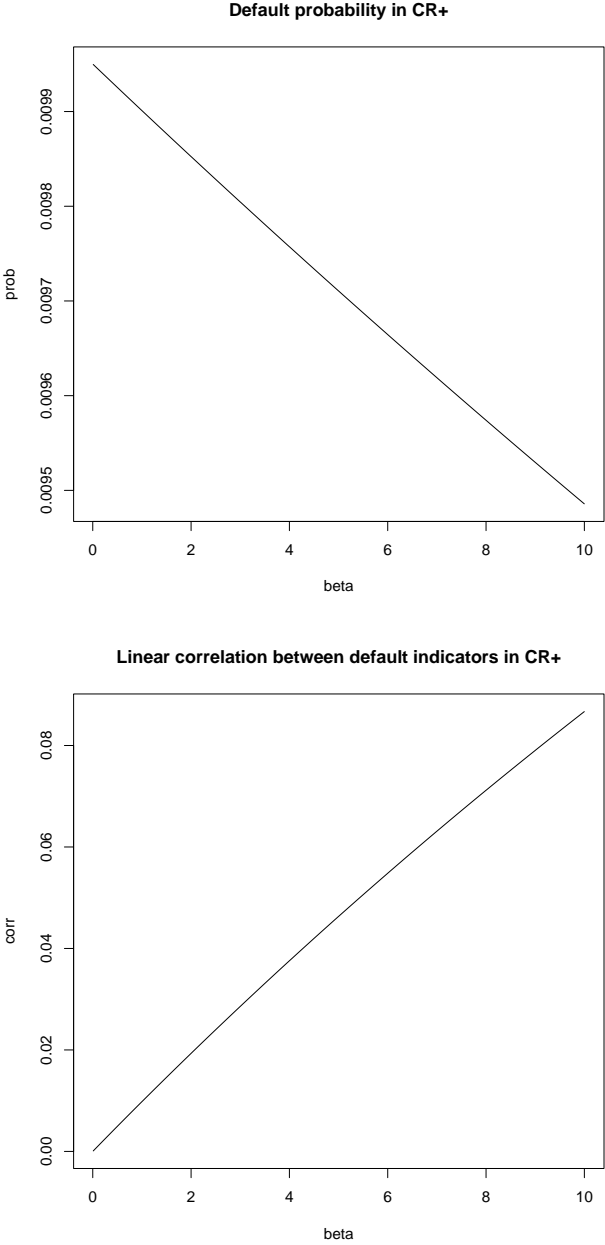and the accuracy of the approximation is better the smaller β is.

Figure 30: The default probability $P(\widetilde{X}_i = 1) = 1 - (100/(100 + \beta))^{1/\beta}$ and linear correlation coefficient $\varrho_L(\widetilde{X}_i, \widetilde{X}_j)$ as functions of β.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999) Coherent Measures of Risk, *Mathematical Finance* 9, 203-228.

[2] Basel Committee on Banking Supervision (1988) International Convergence of Capital Measurement and Capital Standards. Available at http://www.bis.org/publ/bcbs04A.pdf.

[3] Campbell, J.Y., Lo, A.W. and MacKinlay, A. (1997) *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

[4] Credit Suisse First Boston (1997) *CreditRisk$^+$: a Credit Management Framework*. Available at http://www.csfb.com/institutional/research/assets/creditrisk.pdf

[5] Crosbie, P. and Bohn, J. (2001) Modelling Default Risk. K.M.V. Corporation. Available at http://www.moodyskmv.com/research/whitepaper/Modeling_Default_Risk.pdf.

[6] Crouhy, M., Galai, D. and Mark, R. (2000) A comparative analysis of current credit risk models, *Journal of Banking & Finance* 24, 59–117.

[7] Embrechts, P., McNeil, A. and Straumann, D. (2002) Correlation and dependence in risk management: properties and pitfalls. In: *Risk management: value at risk and beyond*, edited by Dempster M, published by Cambridge University Press, Cambridge.

[8] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Berlin.

[9] Fang, K.-T., Kotz, S. and Ng, K.-W. (1987) *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[10] Gordy, M.B. (2000) A comparative anatomy of credit risk models, *Journal of Banking & Finance* 24, 119-149.

[11] Gut, A. (1995) *An Intermediate Course in Probability*, Springer Verlag, New York.

[12] McNeil, A., Frey, R., Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press.

[13] Nelsen, R.B. (1999) *An Introduction to Copulas*, Springer Verlag, New York.

A A few probability facts

A.1 Convergence concepts

Let X and Y be random variables (or vectors) with distribution functions F_X and F_Y. We say that $X = Y$ almost surely (a.s.) if $P(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1$. We say that they are equal in distribution, written $X \stackrel{d}{=} Y$, if $F_X = F_Y$. Notice that $X = Y$ a.s. implies $X \stackrel{d}{=} Y$. However, taking X to be standard normally distributed and $Y = -X$ we see that the converse is false.

Let $X, X_1, X_2, \ldots$ be a sequence of random variables. We say that X_n converges to X almost surely, $X_n \to X$ a.s., if

$$P(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

We say that X_n converges to X in probability, $X_n \stackrel{\mathrm{P}}{\to} X$, if for all $\varepsilon > 0$ it holds that

$$P(|X_n - X| > \varepsilon) \to 0.$$

We say that X_n converges to X in distribution, $X_n \stackrel{d}{\to} X$, if for all continuity points x of F_X it holds that

$$F_{X_n}(x) \to F_X(x).$$

The following implications between the convergence concepts hold:

$$X_n \to X \text{ a.s.} \quad \Rightarrow \quad X_n \stackrel{\mathrm{P}}{\to} X \quad \Rightarrow \quad X_n \stackrel{d}{\to} X.$$

A.2 Limit theorems and inequalities

Let $X_1, X_2, \ldots$ be iid random variables with finite mean $\mathrm{E}(X_1) = \mu$, and let $S_n = X_1 + \cdots + X_n$. The (strong) law of large numbers says that

$$S_n/n \to \mu \text{ a.s. as } n \to \infty.$$

If furthermore $\mathrm{var}(X_1) = \sigma^2 \in (0, \infty)$, then the central limit theorem says that

$$(S_n - n\mu)/(\sigma\sqrt{n}) \stackrel{d}{\to} Z \text{ as } n \to \infty,$$

where the random variable Z has a standard normal distribution.

For a nonnegative random variable V with $\mathrm{E}(V^r) < \infty$ Markov's inequality says that

$$P(V > \varepsilon) \leq \frac{\mathrm{E}(V^r)}{\varepsilon^r} \quad \text{for every } \varepsilon > 0.$$

For a random variable X with finite variance $\mathrm{var}(X)$ this leads to

$$P(|X - \mathrm{E}(X)| > \varepsilon) \leq \frac{\mathrm{E}[(X - \mathrm{E}(X))^2]}{\varepsilon^2} = \frac{\mathrm{var}(X)}{\varepsilon^2} \quad \text{for every } \varepsilon > 0.$$

This inequality is called Chebyshev's inequality.

B Conditional expectations

Suppose that one holds a portfolio or financial contract and let X be the payoff (or loss) one year from today. Let Z be a random variable which represents some information about the state of the economy or relevant interest rate during the next year. We think of Z as a future (and hence unknown) economic scenario which takes values in a set of all possible scenarios. The conditional expectation of X given Z, $\mathrm{E}(X \mid Z)$, represents the best guess of the payoff X given the scenario Z. Notice that Z is unknown today and that $\mathrm{E}(X \mid Z)$ is a function of the future random scenario Z and hence a random variable. If we knew Z then we would also know $g(Z)$ for any given function g. Therefore the property

$$g(Z)\,\mathrm{E}(X \mid Z) = \mathrm{E}(g(Z)X \mid Z)$$

seems natural (whenever the expectations exist finitely). If there were only a finite set of possible scenarios z_k, i.e. values that Z may take, then it is clear that the expected payoff $\mathrm{E}(X)$ may be computed by computing the expected payoff for each scenario z_k and then obtain $\mathrm{E}(X)$ as the weighted sum of these values with the weights $\mathrm{P}(Z = z_k)$. Therefore the property

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid Z))$$

seems natural.

B.1 Definition and properties

If X is a random variable with $\mathrm{E}(X^2) < \infty$, then the conditional expectation $\mathrm{E}(X \mid Z)$ is most naturally defined geometrically as an orthogonal projection of X onto a subspace.

Let L^2 be the space of random variables X with $\mathrm{E}(X^2) < \infty$. Let Z be a random variable and let $L^2(Z)$ be the space of random variables $Y = g(Z)$ for some function g such that $\mathrm{E}(Y^2) < \infty$. Notice that the expected value $\mathrm{E}(X)$ is the number μ that minimizes the expression $\mathrm{E}((X - \mu)^2)$. The conditional expectation $\mathrm{E}(X \mid Z)$ can be defined similarly.

Definition For $X \in L^2$, the conditional expectation $\mathrm{E}(X \mid Z)$ is the random variable $Y \in L^2(Z)$ that minimizes $\mathrm{E}((X - Y)^2)$.

We say that $X, Y \in L^2$ are orthogonal if $\mathrm{E}(XY) = 0$. Then $E(X \mid Z)$ is the orthogonal projection of X onto $L^2(Z)$, i.e. the point in the subspace $L^2(Z)$ that is closest to X. Moreover, $X - \mathrm{E}(X \mid Z)$ is orthogonal to all $Y \in L^2(Z)$, i.e. $\mathrm{E}(Y(X - \mathrm{E}(X \mid Z))) = 0$ for all $Y \in L^2(Z)$. Equivalently, by linearity of the ordinary expectation,

$$\mathrm{E}(Y\,\mathrm{E}(X \mid Z)) = \mathrm{E}(YX) \quad \text{for all } Y \in L^2(Z). \tag{B.1}$$

This relation implies the following three properties:

(i) If $X \in L^2$, then $\mathrm{E}(\mathrm{E}(X \mid Z)) = \mathrm{E}(X)$.

(ii) If $Y \in L^2(Z)$, then $Y \, \mathrm{E}(X \mid Z) = \mathrm{E}(YX \mid Z)$.

(iii) If $X \in L^2$ and we set $\mathrm{var}(X \mid Z) = \mathrm{E}(X^2 \mid Z) - \mathrm{E}(X \mid Z)^2$, then $\mathrm{var}(X) = \mathrm{E}(\mathrm{var}(X \mid Z)) + \mathrm{var}(\mathrm{E}(X \mid Z))$.

This can be shown as follows:

(i) Choosing $Y = 1$ in (B.1) yields $\mathrm{E}(\mathrm{E}(X \mid Z)) = \mathrm{E}(X)$.

(ii) With Y replaced by $\widetilde{Y}Y$ (with $\widetilde{Y} \in L^2(Z)$), (B.1) says that $\mathrm{E}(\widetilde{Y}YX) = \mathrm{E}(\widetilde{Y}Y \, \mathrm{E}(X \mid Z))$. With Y replaced by $\widetilde{Y}$ and X replaced by YX, (B.1) says that $\mathrm{E}(\widetilde{Y}YX) = \mathrm{E}(\widetilde{Y} \, \mathrm{E}(YX \mid Z))$. Hence, for $X \in L^2$ and $Y \in L^2(Z)$,

$$\mathrm{E}(\widetilde{Y}Y \, \mathrm{E}(X \mid Z)) = \mathrm{E}(\widetilde{Y} \, \mathrm{E}(YX \mid Z)) \quad \text{for all } \widetilde{Y} \in L^2(Z).$$

Equivalently, $\mathrm{E}(\widetilde{Y}[Y \, \mathrm{E}(X \mid Z) - \mathrm{E}(YX \mid Z)]) = 0$ for all $\widetilde{Y} \in L^2(Z)$; in particular for $\widetilde{Y} = Y \, \mathrm{E}(X \mid Z) - \mathrm{E}(YX \mid Z)$, which gives $\mathrm{E}((Y \, \mathrm{E}(X \mid Z) - \mathrm{E}(YX \mid Z))^2) = 0$. This implies that $Y \, \mathrm{E}(X \mid Z) = \mathrm{E}(YX \mid Z)$.

(iii) Starting with the right-hand side we obtain

$$\mathrm{E}(\mathrm{var}(X \mid Z)) + \mathrm{var}(\mathrm{E}(X \mid Z))$$
$$= \mathrm{E}\left(\mathrm{E}(X^2 \mid Z) - \mathrm{E}(X \mid Z)^2 \right) + \left(\mathrm{E}(\mathrm{E}(X \mid Z)^2) - \mathrm{E}(\mathrm{E}(X \mid Z))^2 \right)$$
$$= \mathrm{E}(\mathrm{E}(X^2 \mid Z)) - \mathrm{E}(\mathrm{E}(X \mid Z))^2$$
$$= \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \mathrm{var}(X).$$

Hence, we have shown the properties (i)-(iii) above.

As already mentioned there are other ways to introduce the conditional expectation $\mathrm{E}(X \mid Z)$ so that the properties (i) and (ii) hold. In that case the statement in the definition above follows from properties (i) and (ii). This is seen from the following argument. If $W \in L^2(Z)$, then $W \, \mathrm{E}(X \mid Z) = \mathrm{E}(WX \mid Z)$ and hence $\mathrm{E}(W \, \mathrm{E}(X \mid Z)) = \mathrm{E}(\mathrm{E}(WX \mid Z)) = \mathrm{E}(WX)$. Hence,

$$\mathrm{E}(W(X - \mathrm{E}(X \mid Z))) = 0 \quad \text{for all } W \in L^2(Z). \tag{B.2}$$

If $Y \in L^2(Z)$ and $W = Y - \mathrm{E}(X \mid Z)$, then

$$\mathrm{E}((X - Y)^2) = \mathrm{E}((X - \mathrm{E}(X \mid Z) - W)^2)$$
$$= \mathrm{E}((X - \mathrm{E}(X \mid Z))^2) - 2 \, \mathrm{E}(W(X - \mathrm{E}(X \mid Z))) + \mathrm{E}(W^2)$$
$$= \mathrm{E}((X - \mathrm{E}(X \mid Z))^2) + \mathrm{E}(W^2).$$

Hence, $\mathrm{E}((X - Y)^2)$ is minimized when $W = 0$, i.e., when $Y = \mathrm{E}(X \mid Z)$.

B.2   An expression in terms the density of (X, Z)

It is common in introductory texts to assume that X and Z has a joint density and derive the conditional expectation $\mathrm{E}(X \mid Z)$ in terms of this joint density.

Suppose that the random vector (X, Z) has a density $f(x, z)$. Write $h(Z)$ for the conditional expectation $\mathrm{E}(X \mid Z)$. From (B.2) we know that $\mathrm{E}(g(Z)(X - h(Z))) = 0$ for all $g(Z) \in L^2(Z)$, i.e.

$$0 = \iint g(z)(x - h(z))f(x, z)\mathrm{d}x\mathrm{d}z = \int g(z) \left(\int (x - h(z))f(x, z)\mathrm{d}x \right) \mathrm{d}z.$$

Hence,

$$0 = \int xf(x, z)\mathrm{d}x - \int h(z)f(x, z)\mathrm{d}x = \int xf(x, z)\mathrm{d}x - h(z) \int f(x, z)\mathrm{d}x$$

or equivalently $h(z) = \int xf(x, z)\mathrm{d}x / \int f(x, z)\mathrm{d}x$. Hence,

$$\mathrm{E}(X \mid Z) = \frac{\int xf(x, Z)\mathrm{d}x}{\int f(x, Z)\mathrm{d}x}.$$

B.3 Orthogonality and projections in Hilbert spaces

Orthogonality and orthogonal projections onto a subspace in the Euclidean space $\mathbb{R}^d$ is well known from linear algebra. However, these concepts are meaningful also in more general spaces. Such spaces are called Hilbert spaces. The canonical example of a Hilbert space is $\mathbb{R}^3$ and our intuition for orthogonality and projections in $\mathbb{R}^3$ works fine in general Hilbert spaces.

A nonempty set H is called a (real) Hilbert space if H is a linear vector space, so that elements in H may be added and multiplied by real numbers, and there exists a function $(x, y) \mapsto \langle x, y \rangle$ from $H \times H$ to $\mathbb{R}$ with the properties:

(i) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$;

(ii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all $x, y, z \in H$;

(iii) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ for all $x, y \in H$ and $\lambda \in \mathbb{R}$;

(iv) $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in H$;

(v) If $\{x_n\} \subset H$ and $\lim_{m,n \to \infty} \langle x_n - x_m, x_n - x_m \rangle = 0$, then there exists $x \in H$ such that $\lim_{n \to \infty} \langle x_n - x, x_n - x \rangle = 0$.

The function $(x, y) \mapsto \langle x, y \rangle$ is called an *inner product* and $|x|_H = \langle x, x \rangle^{1/2}$ is the *norm* of $x \in H$. For all $x, y \in H$ it holds that $|\langle x, y \rangle| \leq |x|_H |y|_H$. If $x, y \in H$ and $\langle x, y \rangle = 0$, then x and y are said to be *orthogonal* and we write $x \perp y$.

The projection theorem Let M be a closed linear subspace of a Hilbert space H. For every $x_0 \in H$ there exists a unique element $y_0 \in M$ such that $|x_0 - y_0|_H \leq |x_0 - y|$ for all $y \in M$. The element y_0 is called the orthogonal projection of x_0 onto the subspace M, and $x_0 - y_0 \perp y$ for all $y \in M$.

Let L^2 be the Hilbert space of random variables X with $\mathrm{E}(X^2) < \infty$ equipped

with the inner product $(X, Y) \mapsto \mathrm{E}(XY)$. Let Z be a random variable and consider the set of random variables $Y = g(Z)$ for continuous functions g such that $\mathrm{E}(g(Z)^2) < \infty$. We denote the closure of this set by $L^2(Z)$ and note that $L^2(Z)$ is a closed subspace of L^2. The closure is obtained by including elements $X \in L^2$ which satisfy $\lim_{n \to \infty} \mathrm{E}((g_n(Z) - X)^2) = 0$ for some sequence $\{g_n\}$ of continuous functions such that $\mathrm{E}(g_n(Z)^2) < \infty$.