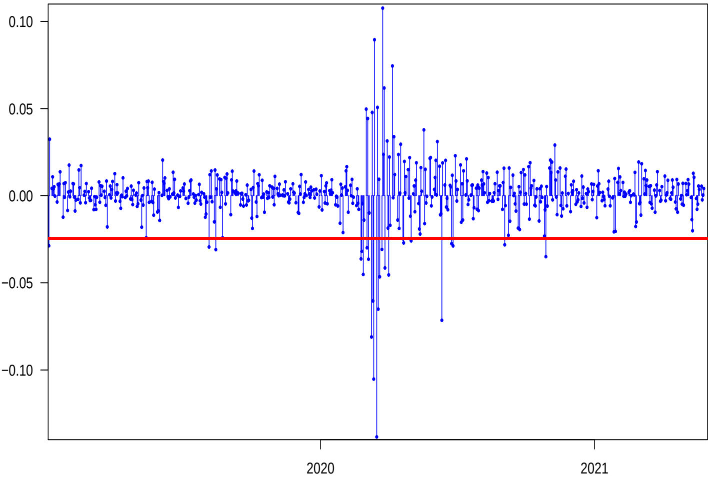
Nicolas Privault

Notes on Financial Risk and Analytics



This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Preface

The topics covered in these notes include an introduction to stochastic modeling with discrete-valued stochastic processes, a basic coverage of Value at Risk and Expected Shortfall, as well as structures of random dependence. Various types of risk, see *e.g.* Gourieroux and Jasiak (2010), can be classified into: market risk, liquidity risk, credit risk, counterparty risk, model risk, estimation risk. For insurance businesses, a more detailed classification can be set as follows.

a) Financial risk

Investment risk
Credit risk,
Market risk (*e.g.* depreciation),
Counterparty risk.

Liability risk
Catastrophe risk,
Non-catastrophe risk (*e.g.* claim volatility).

b) Operational risk

Business risk (*e.g.* lower production),
Event risk (*e.g.* system failure),
Policy risk ...

Part I introduces tools for stochastic modeling, with applications in option pricing, portfolio allocation, and insurance, starting with the use of random walks and geometric Brownian motion for financial modeling in Chapter 1. This is followed by the discrete and continuous-time modeling of time-dependent events using time series and processes with jumps, respectively in Chapters 2 and 3. In particular, the risk theory considered in Chapter 3 is relevant to liability, catastrophe and operational risks such as business or event risk. Correlation and dependence are treated via the used of copulas in Chapter 4.

Financial, investment, market and non-catastrophe risks are dealt with in Part II which focuses on risk measures. This includes the superhedging risk measure in Chapter 5, and Value at Risk and Expected Shortfall in Chapter 6 and 7. Chapter 8 is devoted to credit scoring, using discriminant analysis and logistic regression. Risk theory and credit scoring are presented with illustrative examples in R.

Credit risk is considered Part III, including the structural and reduced-form approaches to credit risk and valuation in Chapters 9 and 10. Credit default is treated via defaultable bonds, Credit Default Swaps (CDS) and collateralized debt obligations (CDOs) in Chapter 11 on credit derivatives.

Parts of this material have been used for teaching in the Masters of Science in Financial Engineering (MFE), Analytics (MSA), and Business Analytics (MSBA) at the Nanyang Technological University in Singapore. The pdf file contains external links and 159 figures, including 10 animated figures and an embedded video in Figure 1.8, that may require using Acrobat Reader for viewing on the complete pdf file. It also contains 14 Python codes and 81 R codes.

This text also includes 70 exercises with solutions. Clicking on an exercise number inside the solution section will send to the original problem text inside the file. Conversely, clicking on the problem number sends the reader to the corresponding solution, however this feature should not be misused.

Nicolas Privault
May 2024

Contents

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

List of Figures

* Animated figures (work in Acrobat Reader).

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

List of Tables

Part I
Stochastic Modeling

Chapter 1
Modeling Market Returns

In an environment subject to risk and randomness, any result or prediction has to rely on a given statistical model. This chapter reviews basic Gaussian modeling for the returns risky assets using Brownian motion and geometric Brownian motion. We also include a statistical benchmarking of such Gaussian-based models to actual market returns.

1.1 Random Walks

Our approach to the modeling of the price of risky assets is based on the use of random walks indexed by discrete or continuous time. We consider the random walk $(S_n)_{n \geqslant 0}$, also called the *Bernoulli* random walk, and defined by letting $S_0 := 0$ and

$$S_n := \sum_{k=1}^{n} X_k = X_1 + \cdots + X_n, \qquad n \geqslant 1,$$

where the increments $(X_k)_{k \geqslant 1}$ form a sequence of *independent and identically distributed* (*i.i.d.*) Bernoulli random variables, with distribution

$$\begin{cases} \mathbb{P}(X_k = +1) = p, \\ \\ \mathbb{P}(X_k = -1) = q, \qquad k \geqslant 1, \end{cases}$$

with $p+q=1$. In other words, the random walk $(S_n)_{n\geqslant 0}$ can only evolve by going up or down by one unit within the finite state space $\{0,1,\ldots,S\}$. We have

$$\mathbb{P}(S_{n+1}=k+1 \mid S_n=k)=p \quad\text{and}\quad \mathbb{P}(S_{n+1}=k-1 \mid S_n=k)=q,$$

$k\in\mathbb{Z}$.

Proposition 1.1. *The mean value of S_n can be expressed as*

$$\mathbb{E}[S_n \mid S_0=0]=\mathbb{E}\left[\sum_{k=1}^n X_k\right]=\sum_{k=1}^n \mathbb{E}\left[X_k\right]=(2p-1)n=(p-q)n,$$

and its variance can be computed as

$$\mathrm{Var}\left[S_n \mid S_0=0\right]=\mathrm{Var}\left[\sum_{k=1}^n X_k\right]=\sum_{k=1}^n \mathrm{Var}\left[X_k\right]=4npq.$$

Figure 1.1 enumerates the $120=\binom{10}{7}=\binom{10}{3}$ possible paths corresponding to $n=5$ and $k=2$, which all have the same probability $p^{n+k}q^{n-k}=p^7q^3$ of occurring.

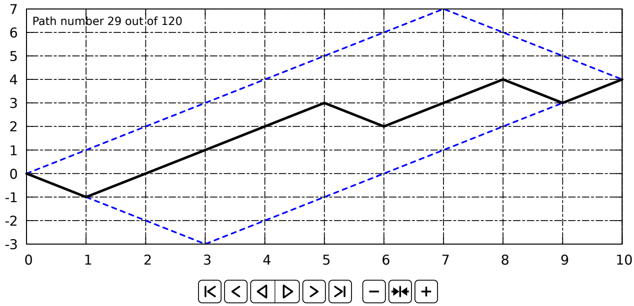


Fig. 1.1: Graph of $120=\binom{10}{7}=\binom{10}{3}$ paths with $n=5$ and $k=2$.*

Proposition 1.2. *The probability distribution of S_{2n}, $n\geqslant 1$, is given by*

* Animated figure (works in Acrobat Reader).

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$\mathbb{P}(S_{2n} = 2k \mid S_0 = 0) = \binom{2n}{n+k} p^{n+k} q^{n-k}, \quad -n \leqslant k \leqslant n. \qquad (1.1)$$

In addition, we note that in an even number of time steps, $(S_n)_{n\in\mathbb{N}}$ *can only reach an even state in* $\mathbb{Z}$ *after starting from* 0.

Similarly, in an odd number of time steps, $(S_n)_{n\in\mathbb{N}}$ can only reach an odd state in $\mathbb{Z}$ after starting from 0.

Brownian Motion

The modeling of random assets in finance is mainly based on stochastic processes, which are families $(X_t)_{t\in I}$ of random variables indexed by a time interval I. Brownian motion is a fundamental example of a stochastic process.

Brown (1828) observed the movement of pollen particles as described in "A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies." Phil. Mag. 4, 161-173, 1828.

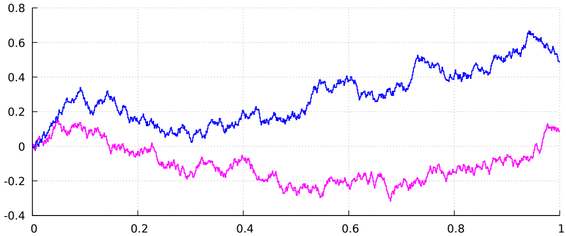


Fig. 1.2: Two sample paths of one-dimensional Brownian motion.

Einstein (1905) received his 1921 Nobel Prize in part for investigations on the theory of Brownian motion: "... in 1905 Einstein founded a kinetic theory to account for this movement", presentation speech by S. Arrhenius, Chairman of the Nobel Committee, Dec. 10, 1922.

Bachelier (1900) used Brownian motion for the modeling of stock prices in his PhD thesis "Théorie de la spéculation", Annales Scientifiques de l'Ecole Normale Supérieure 3 (17): 21-86, 1900.

Wiener (1923) is credited, among other fundamental contributions, for the mathematical foundation of Brownian motion, published in 1923. In particular he constructed the Wiener space and Wiener measure on $\mathcal{C}_0([0,1])$ (the space of continuous functions from $[0,1]$ to $\mathbb{R}$ vanishing at 0).

Itô (1944) constructed the Itô integral with respect to Brownian motion, and the stochastic calculus with respect to Brownian motion, which laid the foundation for the development of calculus for random processes, see Itô (1951) "On stochastic differential equations", *in* Memoirs of the American Mathematical Society.

Definition 1.3. *The standard Brownian motion is a stochastic process* $(B_t)_{t\in\mathbb{R}_+}$ *such that*

1. $B_0 = 0$ *almost surely,*
2. *The sample paths* $t \mapsto B_t$ *are (almost surely) continuous.*
3. *For any finite sequence of times* $t_0 < t_1 < \cdots < t_n$*, the increments*

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

are independent.

4. *For any times* $0 \leqslant s < t$*,* $B_t - B_s$ *is normally distributed with mean zero and variance* $t - s$*.*

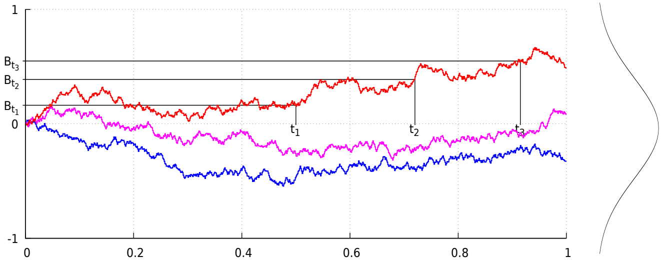


Fig. 1.3: Sample paths of a one-dimensional Brownian motion.

See e.g. Chapter 1 of Revuz and Yor (1994) and Theorem 10.28 in Folland (1999) for proofs of existence of Brownian motion as a stochastic process $(B_t)_{t\in\mathbb{R}_+}$ satisfying the Conditions *1-4* of Definition 1.3.

Approximating Brownian motion as a random walk

We will informally regard Brownian motion as a random walk over infinitesimal time intervals of length Δt, whose increments

$$\Delta B_t := B_{t+\Delta t} - B_t \simeq \mathcal{N}(0, \Delta t)$$

over the time interval $[t, t+\Delta t]$ will be approximated by the Bernoulli random variable

$$\Delta B_t \approx \pm\sqrt{\Delta t} \tag{1.2}$$

with equal probabilities $(1/2, 1/2)$, hence

$$\mathbb{E}[\Delta B_t] = \frac{1}{2}\sqrt{\Delta t} - \frac{1}{2}\sqrt{\Delta t} = 0,$$

and

$$\mathrm{Var}[\Delta B_t] = \mathbb{E}\big[(\Delta B_t)^2\big] = \frac{1}{2}\Delta t + \frac{1}{2}\Delta t = \Delta t.$$

Figure 1.4 presents a simulation of Brownian motion as a random walk with $\Delta t = 0.1$.

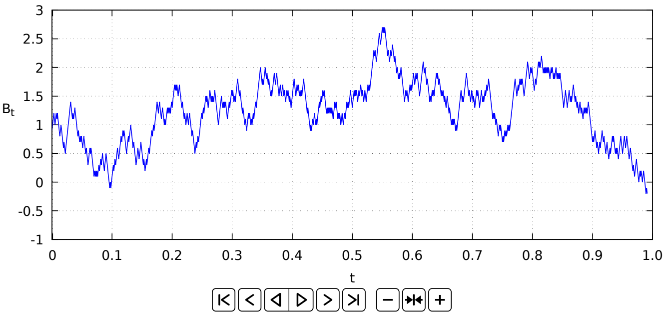


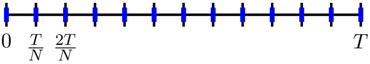
Fig. 1.4: Construction of Brownian motion as a random walk.*

In order to recover the Gaussian distribution property of the random variable B_T, we can split the time interval $[0, T]$ into N subintervals

$$(t_{k-1}, t_k], \qquad k = 1, 2, \ldots, N,$$

* The animation works in Acrobat Reader on the entire pdf file.

of same length $\Delta t = T/N$, where $t_k = kT/N$, $k = 0, 1, \ldots N$, and N is "large".



Letting

$$X_k := \pm\sqrt{T} = \pm\sqrt{N\Delta t}, \qquad k = 1, 2, \ldots N,$$

denote a sequence of symmetric Bernoulli random variables taking values in $\{-\sqrt{T}, \sqrt{T}\}$, with equal probabilities $(1/2, 1/2)$ and

$$\Delta B_{t_k} := \pm\sqrt{\Delta t} = \pm\sqrt{\frac{T}{N}} = \frac{X_k}{\sqrt{N}}, \qquad k = 1, 2, \ldots N,$$

we can write

$$B_T \simeq \sum_{k=1}^{N} \Delta B_{t_k} = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}},$$

hence by the central limit theorem we recover the fact that B_T has the centered Gaussian distribution $\mathcal{N}(0, T)$ with variance T, cf. point 4 of the above Definition 1.3 of Brownian motion, and the illustration given in Figure 1.5.

Remark 1.4.

i) The choice of the square root in (1.2) *is in fact not fortuitous. Indeed, any choice of* $\pm(\Delta t)^\alpha$ *with a power* $\alpha > 1/2$ *would lead to explosion of the process as* dt *tends to zero, whereas a power* $\alpha \in (0, 1/2)$ *would lead to a vanishing process, as can be checked from the R code* below.*

ii) According to this representation, the paths of Brownian motion are not differentiable, although they are continuous by Property 2, *as we have*

$$\frac{\Delta B_t}{\Delta t} \simeq \frac{\pm\sqrt{\Delta t}}{\Delta t} = \pm\frac{1}{\sqrt{\Delta t}} \simeq \pm\infty, \tag{1.3}$$

see e.g. Theorem 10.3 page 153 of Schilling and Partzsch (2014).

* Download the corresponding **IPython notebook** that can be run **here** or **here**.

```
nsim=100; N=1000; t <- 0:N; dt <- 1.0/N; dev.new(width=16,height=7); # Using Bernoulli
    samples
X <- matrix((dt)^0.5*(rbinom( nsim * N, 1, 0.5)-0.5)*2, nsim, N)
X <- cbind(rep(0, nsim), t(apply(X, 1, cumsum))); H<-hist(X[,N],plot=FALSE);
layout(matrix(c(1,2), nrow =1, byrow = TRUE));par(mar=c(2,2,2,0), oma = c(2, 2, 2, 2))
plot(t*dt, X[1, ], xlab = "", ylab = "", type = "l", ylim = c(-2, 2), col = 0,xaxs='i',las=1,
    cex.axis=1.6)
for (i in 1:nsim){lines(t*dt, X[i, ], type = "l", ylim = c(-2, 2), col = i)}
lines(t*dt,sqrt(t*dt),lty=1,col="red",lwd=3);lines(t*dt,-sqrt(t*dt), lty=1, col="red",lwd=3)
lines(t*dt,0*t, lty=1, col="black",lwd=2)
for (i in 1:nsim){points(0.999, X[i,N], pch=1, lwd = 5, col = i)}
x <- seq(-2,2, length=100); px <- dnorm(x);par(mar = c(2,2,2,2))
plot(NULL , xlab="", ylab="", xlim = c(0, max(px,H$density)), ylim = c(-2,2),axes=F)
rect(0, H$breaks[1:(length(H$breaks) - 1)], col=rainbow(20,start=0.08,end=0.6), H$density,
    H$breaks[2:length(H$breaks)]); lines(px,x, lty=1, col="black",lwd=2)
```

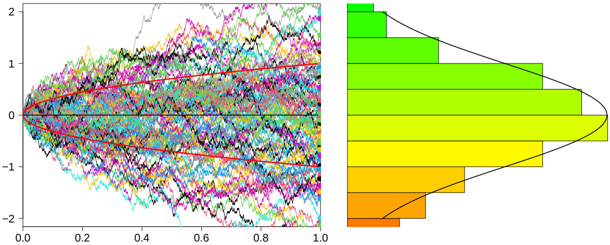


Fig. 1.5: Statistics of one-dimensional Brownian paths *vs.* Gaussian distribution.

The following R code* plots the 72 yearly return graphs of the S&P 500 index from 1950 to 2022, see Figure 1.6. The histogram of year-end returns is then fitted to a normalized Gaussian probability density function.

* Download the corresponding **IPython notebook** that can be run **here** or **here**.

```
library(quantmod); getSymbols("^GSPC",from="1950-01-01",to="2022-12-31",src="yahoo")
stock<-Cl(`GSPC`); s=0;y=0;j=0;count=0;N=240;nsim=72; X = matrix(0, nsim, N)
for (i in 1:nrow(GSPC)){if (s==0 && grepl('-01-0',index(stock[i]))) {if (count==0 || X[y,N]>0)
        {y=y+1;j=1;s=1;count=count+1;}}
if (j<=N) {X[y,j]=as.numeric(stock[i]);};if (grepl('-02-0',index(stock[i]))) {s=0;};j=j+1;}
t <- 0:(N-1); dt <- 1.0/N; dev.new(width=16,height=7);
layout(matrix(c(1,2), nrow =1, byrow = TRUE));par(mar=c(2,2,2,0), oma = c(2, 2, 2, 2))
m=mean(X[,N]/X[,1]-1);sigma=sd(X[,N]/X[,1]-1)
plot(t*dt, X[1,]/X[1,1]-1-m*t*dt, xlab = "", ylab = "", type = "l", ylim = c(-0.5, 0.5), col = 0,
        xaxs='i',las=1, cex.axis=1.6)
for (i in 1:nsim){lines(t*dt, X[i,]/X[i,1]-1-m*t*dt, type = "l", col = i)}
lines(t*dt,sigma*sqrt(t*dt),lty=1,col="red",lwd=3);lines(t*dt,-sigma*sqrt(t*dt), lty=1,
        col="red",lwd=3)
lines(t*dt,0*t, lty=1, col="black",lwd=2)
for (i in 1:nsim){points(0.999, X[i,N]/X[i,1]-1-m*N*dt, pch=1, lwd = 5, col = i)}
x <- seq(-0.5,0.5, length=100); px <- dnorm(x,0,sigma);par(mar = c(2,2,2,2))
H<-hist(X[,N]/X[,1]-1-m*N*dt,plot=FALSE);
plot(NULL , xlab="", ylab="", xlim = c(0, max(px,H$density)), ylim = c(-0.5,0.5),axes=F)
rect(0, H$breaks[1:(length(H$breaks) - 1)], col=rainbow(20,start=0.08,end=0.6), H$density,
H$breaks[2:length(H$breaks)]); lines(px,x, lty=1, col="black",lwd=2)
```

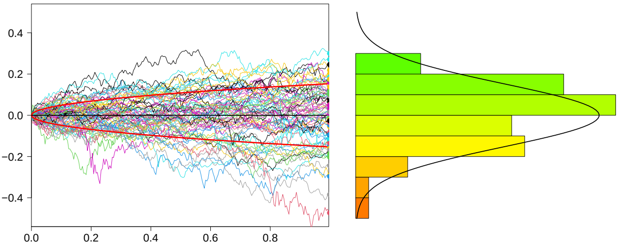


Fig. 1.6: Statistics of 72 S&P 500 yearly return graphs from 1950 to 2022.

1.2 Geometric Brownian Motion

The market returns of an asset price process $(S_t)_{t\in\mathbb{R}_+}$ over time can be estimated in various ways.

Definition 1.5. *1. Standard returns are defined as*

$$\frac{\Delta S_t}{S_t} := \frac{S_{t+\Delta t} - S_t}{S_t}. \tag{1.4}$$

2. Log-returns are defined as

$$\Delta \log S_t = \log S_{t+\Delta t} - \log S_t = \log \frac{S_{t+\Delta t}}{S_t} := \log\left(1 + \frac{\Delta S_t}{S_t}\right), \quad t \geqslant 0. \tag{1.5}$$

Estimating market returns with R

The R package quantmod can be used to fetch financial data from various sources such as Yahoo! Finance or the Federal Reserve Bank of St. Louis (FRED). It can be installed and run via the following command.

```r
install.packages("quantmod")
library(quantmod)
getSymbols("DEXJPUS",src="FRED") # Japan/U.S. Foreign Exchange Rate
getSymbols("CPIAUCNS",src='FRED') # Consumer Price Index
getSymbols("GOOG",src="yahoo") # Google Stock Price
```

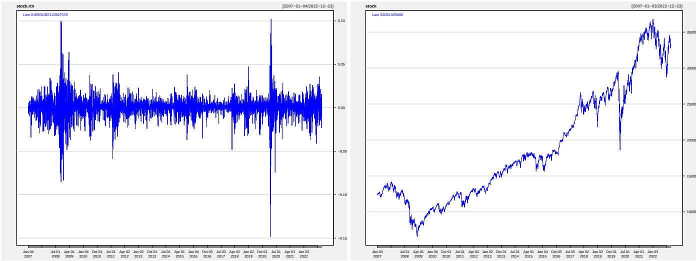
The package yfinance can be similarly used in Python.

```python
import pandas as pd
import yfinance as yf
# Google Stock Price
GOOG = yf.download("GOOG", start="2022-01-01", end="2023-12-31")
```

The following R script allows us to fetch market price data using the quantmod package. Market returns can be estimated using either the standard returns $\Delta S_t / S_t$ or the log-returns $\Delta \log S_t$, $t \geqslant 0$, which can be computed by the command diff(log(stock)), here with $dt = 1/365$.

```r
getSymbols("^DJI",from="2007-01-03",to=Sys.Date(),src="yahoo")
stock=Ad(`DJI`)
chartSeries(stock,up.col="blue",theme="white")
stock.logrtn=diff(log(stock)); # log returns
stock.rtn=(stock-lag(stock))/lag(stock); # standard returns
chartSeries(stock.rtn,up.col="blue",theme="white")
n = length(!is.na(stock.rtn))
```

The adjusted close price Ad() is the closing price computed after adjustments for applicable splits and dividend distributions.



(a) Stock returns. (b) Cumulative stock returns.

Fig. 1.7: Returns *vs.* cumulative returns.

The package yfinance can be similarly used in Python.

```
import numpy as np; import pandas as pd; import yfinance as yf
import matplotlib.pyplot as plt # Get stock data
stock_data = yf.download("^DJI", start="2007-01-03", end=pd.to_datetime("today"))
stock = stock_data["Adj Close"] # Extract adjusted close prices
stock_logrtn = stock.apply(lambda x: np.log(x)).diff() # Calculate log returns
stock_rtn = (stock - stock.shift(1)) / stock.shift(1) # Calculate standard returns
plt.subplot(1, 2, 1); stock_rtn.plot(color="blue");plt.title("Returns Chart");
plt.subplot(1, 2, 2); stock.plot(color="blue"); plt.title("Stock Chart"); # Plot stock chart
plt.tight_layout(); plt.show() # Plot returns chart
n = stock_rtn.count() # Calculate the number of non-NaN returns
```

Proposition 1.6. *The sequence* $(S_{t_k})_{k=1,\ldots,N}$ *of market prices can be recovered from the sequence of standard returns*

$$\frac{\Delta S_{t_0}}{S_{t_0}} = \frac{S_{t_1} - S_{t_0}}{S_{t_0}}, \ldots, \frac{\Delta S_{t_{N-1}}}{S_{t_{N-1}}} = \frac{S_{t_N} - S_{t_{N-1}}}{S_{t_{N-1}}}$$

from the telescoping product identity

$$S_{t_n} = S_0 \prod_{k=1}^{n} \frac{S_{t_k}}{S_{t_{k-1}}} = S_0 \prod_{k=1}^{n} \left(1 + \frac{\Delta S_{t_{k-1}}}{S_{t_{k-1}}}\right), \quad 1 \leqslant n \leqslant N. \tag{1.6}$$

Using log-returns, we have similarly

$$S_{t_n} = S_0 \exp\left(\sum_{k=1}^{n} \Delta \log S_{t_k}\right) = S_0 \prod_{k=1}^{n} \mathrm{e}^{\Delta \log S_{t_k}}, \qquad 1 \leqslant n \leqslant N.$$

The next R code recovers cumulative returns from standard market returns, according to (1.6).

```
stock<-stock[!is.na(stock.rtn)];stock.rtn<-stock.rtn[!is.na(stock.rtn)]
times=index(stock);dev.new(width=16,height=7);par(mfrow=c(1,2))
plot(times,stock.rtn,pch=19,xaxs="i",yaxs="i",cex=0.03,col="blue", ylab="", xlab="", main =
      'Asset returns', las=1, cex.lab=1.8, cex.axis=1.8, lwd=3)
segments(x0 = times, x1 = times, y0 = 0, y1 = stock.rtn,col="blue")
plot(times,100 * cumprod(1 + as.numeric(stock.rtn)),type='l',col='black',main = "Asset
      prices",ylab="", cex=0.1,cex.axis=1,las=1)
```

The package yfinance can be similarly used in Python.

```
# Remove NaN values from stock and stock_rtn
stock = stock[~np.isnan(stock_rtn)]; stock_rtn = stock_rtn[~np.isnan(stock_rtn)]
times = stock.index; plt.subplot(1, 2, 1); plt.plot(times, stock_rtn, color='blue')
plt.xlabel('Time'); plt.ylabel('Asset Returns')
plt.title('Asset Returns'); plt.xticks(rotation=90)
plt.grid(True) # Plot asset returns
plt.subplot(1, 2, 2); cumulative_returns = (1 + stock_rtn.astype(float)).cumprod() * 100
plt.plot(times, cumulative_returns, color='black'); plt.xlabel('Time')
plt.ylabel('Asset Prices'); plt.title('Asset Prices'); plt.xticks(rotation=90); plt.grid(True)
plt.tight_layout(); plt.show(); # Plot asset prices
```

Geometric Brownian Motion

Samuelson (1965) rediscovered Bachelier's ideas and proposed geometric Brownian motion as a model for stock prices. In an interview he stated "In the early 1950s I was able to locate by chance this unknown Bachelier (1900) book, rotting in the library of the University of Paris, and when I opened it up it was as if a whole new world was laid out before me." We refer to "Rational theory of warrant pricing" by Paul Samuelson, *Industrial Management Review*, p. 13-32, 1965.

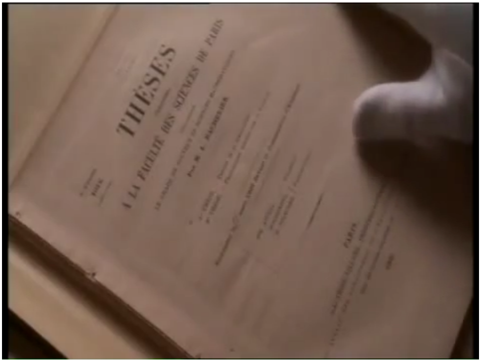


Fig. 1.8: Clark (2000) "As if a whole new world was laid out before me."*

The evolution of a riskless bank account value $(A_t)_{t\in\mathbb{R}_+}$ is constructed from standard returns, defined as follows:

$$\frac{A_{t+\Delta t}-A_t}{A_t}=r\Delta t, \quad i.e \quad \frac{\Delta A_t}{A_t}=r\Delta t.$$

This equation can be regarded as a discretization of the differential equation

$$A'_t=\frac{dA_t}{dt}=rA_t, \quad t\geqslant 0,$$

which has the solution

$$A_t=A_0\mathrm{e}^{rt}, \qquad t\geqslant 0, \tag{1.7}$$

where $r>0$ is the risk-free interest rate.†

In what follows, we will model the risky asset price process $(S_t)_{t\in\mathbb{R}_+}$ using standard returns, from the equation

$$\frac{S_{t+\Delta t}-S_t}{S_t}=\frac{\Delta S_t}{S_t}=\mu\Delta t+\sigma\Delta B_t, \qquad t\geqslant 0, \tag{1.8}$$

* Click on the figure to play the video (works in Acrobat Reader on the entire pdf file).
† "Anyone who believes exponential growth can go on forever in a finite world is either a madman or an economist", K. E. Boulding (1973), page 248.

which can be solved numerically according to the following R and Python codes.

```r
N=2000; t <- 0:N; dt <- 1.0/N;mu=0.5; sigma=0.2; nsim <- 10; S <- matrix(0, nsim, N+1)
Z <- matrix(rnorm(nsim*N,mean=0,sd=sqrt(dt)), nsim, N+1)
for (i in 1:nsim){S[i,1]=1.0;
for (j in 1:N+1){S[i,j]=S[i,j-1]*(1 + mu*dt+sigma*Z[i,j])}}
plot(t*dt, rep(0, N+1), xlab = "Time", ylab = "Geometric Brownian motion", lwd=2, ylim =
      c(min(S),max(S)), type = "l", col = 0,las=1, cex.axis=1.5,cex.lab=1.5, xaxs='i', yaxs='i')
for (i in 1:nsim){lines(t*dt, S[i, ], lwd=2, type = "l", col = i)}
```

```python
import numpy as np; import matplotlib.pyplot as plt
%matplotlib
N = 2000; t = np.arange(0, N+1); dt = 1.0 / N; mu = 0.5; sigma = 0.2; nsim = 10
X = np.zeros((nsim, N+1)); Z = np.random.normal(0, np.sqrt(dt), (nsim, N+1))
for i in range(nsim):
    X[i, 0] = 1.0
    for j in range(1, N+1):
        X[i, j] = X[i, j-1] + mu * X[i, j-1] * dt + sigma * X[i, j-1] * Z[i, j]
plt.plot(t*dt, np.zeros(N+1), color='black', linewidth=2)
plt.xlabel('Time'); plt.ylabel('Geometric Brownian motion'); plt.xlim(0, N*dt)
plt.ylim(np.min(X), np.max(X)); plt.xticks(fontsize=12); plt.yticks(fontsize=12)
for i in range(nsim):
    plt.plot(t*dt, X[i, :], linewidth=2)
plt.show()
```

The following proposition gives the explicit solution to (1.8) regarded as a discretization of the stochastic differential Equation (1.9) below.

Proposition 1.7 (Geometric Brownian motion). *The solution of the stochastic differential equation*

$$dS_t = \mu S_t dt + \sigma S_t dB_t \tag{1.9}$$

is given by

$$S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right), \qquad t \geqslant 0. \tag{1.10}$$

Proof. Using (1.8), the log-returns (1.5) of an asset priced $(S_t)_{t\in\mathbb{R}_+}$ satisfy

$$\frac{dS_t}{S_t} = \frac{S_{t+dt} - S_t}{S_t} = \mu dt + \sigma dB_t.$$

Hence, using the second order approximation $\log(1+x) \simeq x - x^2/2$ as x tends to zero, we have

$$\begin{aligned} d\log S_t &\simeq \log S_{t+dt} - \log S_t \\ &= \log \frac{S_{t+dt}}{S_t} \end{aligned}$$

$$
\begin{aligned}
&= \log\left(1 + \frac{S_{t+dt} - S_t}{S_t}\right) \\
&= \log\left(1 + \frac{dS_t}{S_t}\right) \\
&= \log(1 + \mu dt + \sigma dB_t) \\
&= \mu dt + \sigma dB_t - \frac{1}{2}(\mu dt + \sigma dB_t)^2 \\
&\simeq \mu dt + \sigma dB_t - \frac{\mu^2}{2}(dt)^2 - \mu\sigma dB_t \bullet dt - \frac{\sigma^2}{2}(dB_t)^2 \\
&\simeq \mu dt - \frac{\sigma^2}{2} dt + \sigma dB_t, \qquad t \geqslant 0.
\end{aligned}
$$

By integration over the time interval $[0, t]$ we find

$$
\begin{aligned}
\log S_t &= \log S_0 + \int_0^t d\log S_s \\
&= \int_0^t \mu ds - \frac{\sigma^2}{2}\int_0^t ds + \sigma \int_0^t dB_s \\
&= \mu t - \frac{\sigma^2 t}{2} + \sigma B_t,
\end{aligned}
$$

which yields (1.10) after exponentiation. □

The next Figure 1.9 presents an illustration of the geometric Brownian process of Proposition 1.7 according to the following R and Python codes.

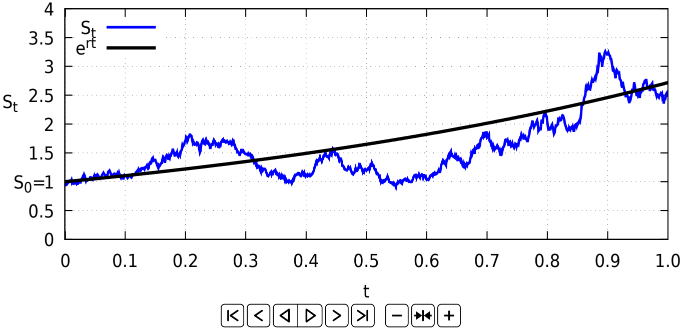


Fig. 1.9: Geometric Brownian motion started at $S_0 = 1$, with $\mu = r = 1$ and $\sigma^2 = 0.5$.*

* The animation works in Acrobat Reader on the entire pdf file.

```r
N=1000; t <- 0:N; dt <- 1.0/N; sigma=0.2; mu=0.5
Z <- rnorm(N,mean=0,sd=sqrt(dt));
plot(t*dt, exp(mu*t*dt), xlab = "time", ylab = "Geometric Brownian motion", type = "l", ylim
    = c(0.75, 2), col = 1,lwd=3)
lines(t*dt, exp(sigma*c(0,cumsum(Z))+mu*t*dt-sigma*sigma*t*dt/2),xlab = "time",type =
    "l",ylim = c(0, 4), col = 'blue', xaxs='i', yaxs='i')
```

```python
import numpy as np; import matplotlib.pyplot as plt
%matplotlib
N = 1000; t = np.arange(0, N+1); dt = 1.0 / N; sigma = 0.2; mu = 0.5
Z = np.random.normal(0, np.sqrt(dt), N)
GBM = np.exp(sigma * np.concatenate(([0], np.cumsum(Z))) + mu * t * dt - (sigma**2) * t *
    dt / 2)
plt.plot(t*dt, np.exp(mu * t * dt), color='blue', linewidth=3, label='Expected GBM')
plt.plot(t*dt, GBM, color='red', linewidth=1.5, label='Simulated GBM')
plt.xlabel('Time'); plt.ylabel('Geometric Brownian motion')
plt.xlim(0, N*dt); plt.ylim(0.75, 2); plt.legend(); plt.grid(True); plt.show()
```

Multiple sample paths of the solution (1.10) to (1.9) can also be simulated by the following R and Python codes.

```r
N=2000; t <- 0:N; dt <- 1.0/N; mu=0.5;sigma=0.2; nsim <- 10; par(oma=c(0,1,0,0))
X <- matrix(rnorm(nsim*N,mean=0,sd=sqrt(dt)), nsim, N)
X <- cbind(rep(0, nsim), t(apply(X, 1, cumsum)))
for (i in 1:nsim){X[i,] <- exp(mu*t*dt+sigma*X[i,]-sigma*sigma*t*dt/2)}
plot(t*dt, rep(0, N+1), xlab = "Time", ylab = "Geometric Brownian motion", lwd=2, ylim =
    c(min(X),max(X)), type = "l", col = 0,las=1,cex.axis=1.5,cex.lab=1.6, xaxs='i', yaxs='i')
for (i in 1:nsim){lines(t*dt, X[i, ], lwd=2, type = "l", col = i)}
```

```python
import numpy as np; import matplotlib.pyplot as plt
%matplotlib
N = 2000; t = np.arange(0, N+1); dt = 1.0 / N; mu = 0.5; sigma = 0.2; nsim = 10
X = np.random.normal(0, np.sqrt(dt), (nsim, N))
X = np.concatenate((np.zeros((nsim, 1)), np.cumsum(X, axis=1)), axis=1)
for i in range(nsim): X[i, :] = np.exp(mu * t * dt + sigma * X[i, :] - (sigma**2) * t * dt / 2)
plt.plot(t*dt, np.zeros(N+1), color='black', linewidth=1)
plt.xlabel('Time'); plt.ylabel('Geometric Brownian motion'); plt.xlim(0, N*dt)
plt.ylim(np.min(X), np.max(X)); plt.xticks(fontsize=12); plt.yticks(fontsize=12)
for i in range(nsim): plt.plot(t*dt, X[i, :], linewidth=1)
plt.show()
```

Figure 1.10 presents sample paths of geometric Brownian motion.

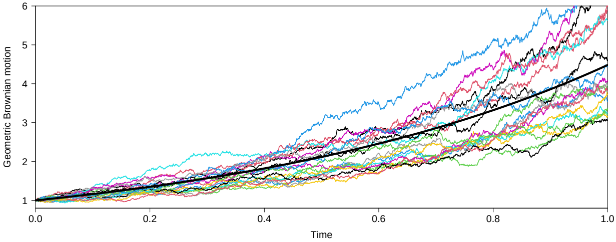


Fig. 1.10: Ten sample paths of geometric Brownian motion $(S_t)_{t\in\mathbb{R}_+}$.

The next R and Python codes compare geometric Brownian motion simulations to asset price data.

```
library(quantmod); getSymbols("0005.HK",from="2016-02-15",to="2017-05-11",src="yahoo")
Marketprices<-Ad(`0005.HK`);
returns = (Marketprices-lag(Marketprices)) / Marketprices
sigma=sd(as.numeric(returns[-1])); r=mean(as.numeric(returns[-1]))
N=length(Marketprices); t <- 0:N; a=(1+r)*(1-sigma)-1;b=(1+r)*(1+sigma)-1
X <- matrix((a+b)/2+(b-a)*rnorm( N-1, 0, 1)/2, 1, N-1)
X <- as.numeric(Marketprices[1])*cbind(0,t(apply((1+X),1,cumprod))); X[,1]=Marketprices[1];
x=seq(100,100+N-1); dates <- index(Marketprices)
GBM<-xts(x =X[1,], order.by = dates); myPars <- chart_pars();myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"; myTheme$rylab <- FALSE;
chart_Series(Marketprices,pars=myPars, theme = myTheme);
dexp<-as.numeric(Marketprices[1])*exp(r*seq(1,305)); ddexp<-xts(x =dexp, order.by = dates)
dev.new(width=16,height=8); par(mfrow=c(1,2));
add_TA(exp(log(ddexp)), on=1, col="black",layout=NULL, lwd=4 ,legend=NULL)
graph <- chart_Series(GBM,theme=myTheme,pars=myPars); myylim <- graph$get_ylim()
graph <- add_TA(exp(log(ddexp)), on=1, col="black",layout=NULL, lwd=4 ,legend=NULL)
myylim[[2]] <- structure(c(min(Marketprices),max(Marketprices)), fixed=TRUE)
graph$set_ylim(myylim); graph
```

```python
import numpy as np; import pandas as pd; import yfinance as yf
from matplotlib import pyplot as plt
data = yf.download("0005.HK", start="2016-02-15", end="2017-05-11", progress=False)
Marketprices = data['Adj Close'];
returns = (Marketprices - Marketprices.shift(1)) / Marketprices
sigma = np.std(returns[1:]); r = np.mean(returns[1:]); N = len(Marketprices);
t = np.arange(0, N); a = (1 + r) * (1 - sigma) - 1; b = (1 + r) * (1 + sigma) - 1
X = np.array([(a + b) / 2 + (b - a) * np.random.normal(0, 1, N-1) / 2])
X = np.concatenate(([Marketprices[0]], Marketprices[0]*np.cumprod(1 + X)))
x = np.arange(100, 100 + N); dates = Marketprices.index; GBM = pd.Series(X, index=dates)
Y = Marketprices[0]*pd.Series(np.exp(r * np.arange(1, N+1)), index=dates)
fig = plt.figure(figsize=(16, 8)); ax1 = fig.add_subplot(121)
ax1.plot(Marketprices, color='blue', linewidth=2)
ax1.set_ylabel('Price'); ax1.set_xlabel('Date'); ylim1 = ax1.get_ylim()
ax2 = fig.add_subplot(122); ax2.plot(GBM, color='blue', linewidth=2)
ax2.plot(Y, color='black', linewidth=2)
ax2.set_ylabel('GBM'); ax2.set_xlabel('Date'); ax2.set_ylim(ylim1)
plt.tight_layout(); plt.show()
```

Figure 1.11 presents a graph of underlying asset price market data, which is compared to the geometric Brownian motion simulation of Figure 1.10 in Figure 1.12.

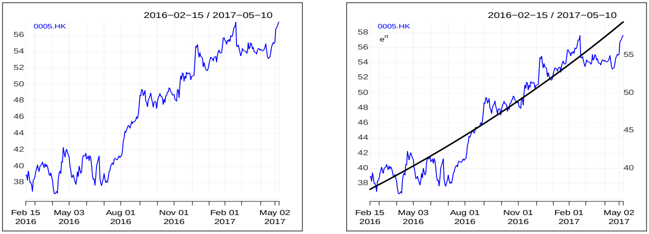


Fig. 1.11: Graph of underlying market prices.

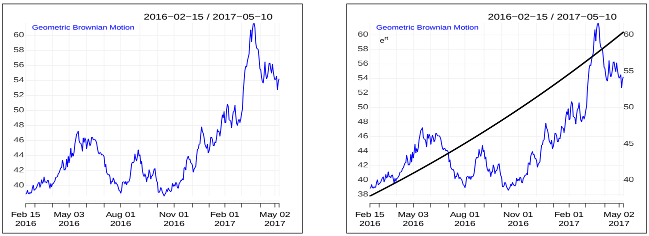


Fig. 1.12: Graph of simulated geometric Brownian motion.

The R package Sim.DiffProc can be used to estimate the coefficients of a geometric Brownian motion fitting observed market data.

```
library("Sim.DiffProc")
fx <- expression( theta[1]*x ); gx <- expression( theta[2]*x )
fitsde(data = as.ts(Marketprices), drift = fx, diffusion = gx, start = list(theta1=0.01,
      theta2=0.01),pmle="euler")
```

In the next proposition, we compute the probability distribution of geometric Brownian motion at any given time.

Proposition 1.8. *At any time $T > 0$, the random variable*

$$S_T := S_0 \mathrm{e}^{\sigma B_T + (\mu - \sigma^2/2)T}$$

has the lognormal distribution *with probability density function*

$$\boxed{x \longmapsto f(x) = \frac{1}{x\sigma\sqrt{2\pi T}} \mathrm{e}^{-(-(\mu-\sigma^2/2)T + \log(x/S_0))^2/(2\sigma^2 T)}, \qquad x > 0, \qquad (1.11)}$$

with log-variance σ^2 and log-mean $(\mu - \sigma^2/2)T + \log S_0$, see Figure 1.14.

Proof. For all $x \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}(S_T \leqslant x) &= \mathbb{P}\big(S_0 \mathrm{e}^{\sigma B_T + (\mu-\sigma^2/2)T} \leqslant x\big) \\
&= \mathbb{P}\left(\sigma B_T + \left(\mu - \frac{\sigma^2}{2}\right) T \leqslant \log \frac{x}{S_0}\right) \\
&= \mathbb{P}\left(B_T \leqslant \frac{1}{\sigma}\left(\log \frac{x}{S_0} - \left(\mu - \frac{\sigma^2}{2}\right) T\right)\right) \\
&= \int_{-\infty}^{(\log(x/S_0)-(\mu-\sigma^2/2)T)/\sigma} \mathrm{e}^{-y^2/(2T)} \frac{dy}{\sqrt{2\pi T}} \\
&= \int_{-\infty}^{(\log(x/S_0)-(\mu-\sigma^2/2)T)/(\sigma\sqrt{T})} \mathrm{e}^{-z^2/2} \frac{dz}{\sqrt{2\pi}} \\
&= \Phi\left(\frac{1}{\sigma\sqrt{T}}\left(\log \frac{x}{S_0} - \left(\mu - \frac{\sigma^2}{2}\right) T\right)\right),
\end{aligned}$$

where

$$\Phi(x) := \int_{-\infty}^{x} \mathrm{e}^{-y^2/2} \frac{dy}{\sqrt{2\pi T}}, \qquad x \in \mathbb{R}, \qquad (1.12)$$

denotes the standard Gaussian Cumulative Distribution Function (CDF) of a standard normal random variable $X \simeq \mathcal{N}(0,1)$, with the relation

$$\Phi(-x) = 1 - \Phi(x), \qquad x \in \mathbb{R}.$$

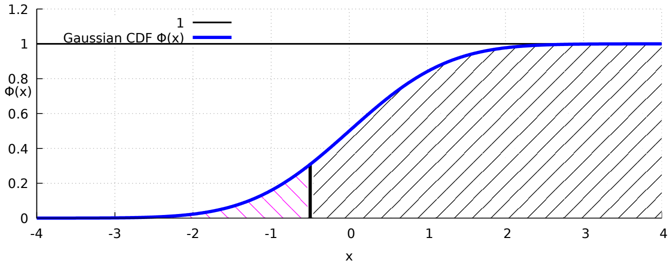


Fig. 1.13: Graph of the Gaussian Cumulative Distribution Function (CDF).

After differentiation with respect to x, we find the lognormal probability density function

$$\begin{aligned} f(x) &= \frac{d\mathbb{P}(S_T \leqslant x)}{dx} \\ &= \frac{\partial}{\partial x} \int_{-\infty}^{(\log(x/S_0)-(\mu-\sigma^2/2)T)/\sigma} e^{-y^2/(2T)} \frac{dy}{\sqrt{2\pi T}} \\ &= \frac{\partial}{\partial x} \Phi\left(\frac{1}{\sigma\sqrt{T}}\left(\log\frac{x}{S_0} - \left(\mu - \frac{\sigma^2}{2}\right)T\right)\right) \\ &= \frac{1}{x\sigma\sqrt{T}} \varphi\left(\frac{1}{\sigma\sqrt{T}}\left(\log\frac{x}{S_0} - \left(\mu - \frac{\sigma^2}{2}\right)T\right)\right) \\ &= \frac{1}{x\sigma\sqrt{2\pi T}} e^{-(-(\mu-\sigma^2/2)T+\log(x/S_0))^2/(2\sigma^2 T)}, \qquad x > 0, \end{aligned}$$

where

$$\varphi(y) = \Phi'(y) := \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \qquad y \in \mathbb{R},$$

denotes the standard Gaussian probability density function. □

The next R code is generating sample paths of geometric Brownian motion together with the histogram of terminal values fitted to the lognormal probability density function (1.11), see Figure 1.14.

```
N=1000; t <- 0:N; dt <- 1.0/N; nsim <- 100 # using Bernoulli samples
sigma=0.2;r=0.5;a=(1+r*dt)*(1-sigma*sqrt(dt))-1;b=(1+r*dt)*(1+sigma*sqrt(dt))-1
X <- matrix(a+(b-a)*rbinom( nsim * N, 1, 0.5), nsim, N)
X<-cbind(rep(0,nsim),t(apply((1+X),1,cumprod))); X[,1]=1;H<-hist(X[,N],plot=FALSE);
      dev.new(width=16,height=7);
layout(matrix(c(1,2), nrow =1, byrow = TRUE)); par(mar=c(2,2,2,0), oma = c(2, 2, 2, 2))
plot(t*dt,X[1,],xlab="time",ylab="",type="l",ylim=c(0.8,3), col = 0,xaxs='i',las=1,
      cex.axis=1.6)
for (i in 1:nsim){lines(t*dt, X[i, ], xlab = "time", type = "l", col = i)}
lines((1+r*dt)^t, type="l", lty=1, col="black",lwd=3,xlab="",ylab="", main="")
for (i in 1:nsim){points(0.999, X[i,N], pch=1, lwd = 5, col = i)}; x <- seq(0.01,3, length=100);
px <- exp(-(-(r-sigma^2/2)+log(x))^2/2/sigma^2)/x/sigma/sqrt(2*pi); par(mar = c(2,2,2,2))
plot(NULL , xlab="", ylab="", xlim = c(0, max(px,H$density)),ylim=c(0.8,3),axes=F, las=1)
rect(0, H$breaks[1:(length(H$breaks) - 1)], col=rainbow(20,start=0.08,end=0.6), H$density,
      H$breaks[2:length(H$breaks)])
lines(px,x, type="l", lty=1, col="black",lwd=3,xlab="",ylab="", main="")
```

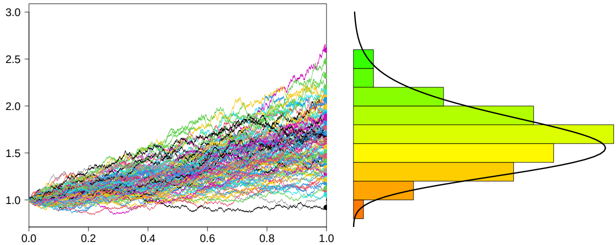


Fig. 1.14: Statistics of geometric Brownian paths *vs.* lognormal distribution.

Time-dependent coefficients

The above construction can be extended to time-dependent interest rate, drift and volatility processes $(r(t))_{t\in\mathbb{R}_+}$, $(\mu(t))_{t\in\mathbb{R}_+}$ and $(\sigma(t))_{t\in\mathbb{R}_+}$. In this case, we let $(A_t)_{t\in\mathbb{R}_+}$ be the risk-free asset with price given by

$$\frac{dA_t}{A_t} = r(t)dt, \qquad A_0 = 1, \qquad t \geqslant 0,$$

i.e.

$$A_t = A_0 \mathrm{e}^{\int_0^t r(s)ds}, \qquad t \geqslant 0,$$

and the price process $(S_t)_{t\in[0,T]}$ is defined by the stochastic differential equation

$$dS_t = \mu(t)S_t dt + \sigma(t)S_t dB_t, \qquad t \geqslant 0,$$

i.e., in integral form,

$$S_t = S_0 + \int_0^t \mu(u) S_u du + \int_0^t \sigma(u) S_u dB_u, \qquad t \geqslant 0,$$

with solution

$$S_t = S_0 \exp\left(\int_0^t \sigma(s) dB_s + \int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s) \right) ds \right),$$

$t \in \mathbb{R}_+$.

Moments and cumulants of random variables

In the sequel we will use the sequence $(\kappa_n(X))_{n\geqslant 1}$ of cumulants of a random variable X, defined from its Moment Generating Function (MGF)

$$\mathcal{M}_X(t) := \mathbb{E}\left[e^{tX}\right] = 1 + \sum_{n\geqslant 1} \frac{t^n}{n!} \mathbb{E}[X^n], \tag{1.13}$$

for t in a neighborhood of 0, see Thiele (1899).

Definition 1.9. *The* cumulants *of a random variable X are the coefficients $(\kappa_n(X))_{n\geqslant 1}$ appearing in the series expansion*

$$\log \mathbb{E}\left[e^{tX}\right] = \sum_{n\geqslant 1} \kappa_n(X) \frac{t^n}{n!}, \tag{1.14}$$

of the logarithmic moment generating function (log-MGF) of X.

The first cumulants of X can be identified as follows.

a) First moment and cumulant. We have $\kappa_1(X) = \mathbb{E}[X]$.

b) Variance and second cumulant. We have

$$\kappa_2(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right],$$

and $\sqrt{\kappa_2(X)}$ is the *standard deviation* of X.

c) The third cumulant of X is given as the third central moment

$$\kappa_3(X) = \mathbb{E}[(X - \mathbb{E}[X])^3].$$

d) Similarly, the fourth cumulant of X satisfies

$$\begin{aligned} \kappa_4(X) &= \mathbb{E}[(X - \mathbb{E}[X])^4] - 3(\kappa_2(X))^2 \\ &= \mathbb{E}[(X - \mathbb{E}[X])^4] - 3\left(\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]\right)^2. \end{aligned}$$

Example: Gaussian moments and cumulants

When X is centered we have $\kappa_1^X = 0$ and $\kappa_2^X = \mathbb{E}[X^2] = \mathrm{Var}[X]$, and X becomes Gaussian if and only if $\kappa_n^X = 0$, $n \geqslant 3$, *i.e.*

$$\kappa_n^X = \mathbb{1}_{\{n=2\}}\sigma^2, \qquad n \geqslant 1,$$

or

$$\left(\kappa_1^X, \kappa_2^X, \kappa_3^X, \kappa_4^X, \ldots\right) = (0, \sigma^2, 0, 0, \ldots).$$

Example: Poisson moments and cumulants

In the particular case of a Poisson random variable $Z \simeq \mathcal{P}(\lambda)$ with intensity $\lambda > 0$, we have

$$\begin{aligned}
\mathbb{E}_\lambda\left[\mathrm{e}^{tZ}\right] &= \sum_{n\geqslant 0} \mathrm{e}^{nt}\mathbb{P}(Z = n) \\
&= \mathrm{e}^{-\lambda} \sum_{n\geqslant 0} \frac{(\lambda \mathrm{e}^t)^n}{n!} \\
&= \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}, \qquad t \in \mathbb{R}_+,
\end{aligned} \tag{1.15}$$

hence $\kappa_n^Z = \lambda$, $n \geqslant 1$, or

$$\left(\kappa_1^Z, \kappa_2^Z, \kappa_3^Z, \kappa_4^Z, \ldots\right) = (\lambda, \lambda, \lambda, \lambda, \ldots),$$

Definition 1.10. *i) The* skewness *of X is defined as*

$$\mathrm{Sk}_X := \frac{\kappa_3(X)}{(\kappa_2(X))^{3/2}} = \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^3\right]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^{3/2}}.$$

ii) The excess kurtosis *of X is defined as*

$$\mathrm{EK}_X := \frac{\kappa_4(X)}{(\kappa_2(X))^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^2} - 3.$$

The cumulants of X were originally called "semi-invariants" due to the property $\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y)$, $n \geqslant 1$, when X and Y are independent random variables. Indeed, in this case we have

$$\begin{aligned}
\sum_{n\geqslant 1} \kappa_n(X + Y)\frac{t^n}{n!} &= \log\left(\mathbb{E}\left[\mathrm{e}^{t(X+Y)}\right]\right) \\
&= \log\left(\mathbb{E}\left[\mathrm{e}^{tX}\right]\mathbb{E}\left[\mathrm{e}^{tY}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= \log \mathbb{E}\left[e^{tX}\right] + \log \mathbb{E}\left[e^{tY}\right] \\
&= \sum_{n \geqslant 1} \kappa_n(X) \frac{t^n}{n!} + \sum_{n \geqslant 1} \kappa_n(Y) \frac{t^n}{n!} \\
&= \sum_{n \geqslant 1} \left(\kappa_n(X) + \kappa_n(Y)\right) \frac{t^n}{n!},
\end{aligned}
$$

showing that $\kappa_n(X+Y) = \kappa_n(X) + \kappa_n(Y)$, $n \geqslant 1$.

1.3 Distribution of Market Returns

Market returns *vs.* Gaussian and power tails

Consider for example the market returns data obtained from fetching DJI and STI index data using the R package quantmod and the following scripts.

```
library(quantmod)
getSymbols("^STI",from="1990-01-03",to="2015-01-03",src="yahoo");stock=Ad(`STI`);
getSymbols("^DJI",from="1990-01-03",to=Sys.Date(),src="yahoo");stock=Ad(`DJI`);
stock.rtn=diff(log(stock));returns <- as.vector(stock.rtn)
m=mean(returns,na.rm=TRUE);s=sd(returns,na.rm=TRUE);times=index(stock.rtn)
n = sum(is.na(returns))+sum(!is.na(returns));x=seq(1,n);y=rnorm(n,mean=m,sd=s)
dev.new(width=16,height=8)
plot(times,returns,pch=19,xaxs="i",yaxs="i",cex=0.03,col="blue", ylab="", xlab="", main = '',
      las=1, cex.lab=1.8, cex.axis=1.8, lwd=3)
segments(x0 = times, x1 = times, y0 = 0, y1 = returns,col="blue")
points(times,y,pch=19,cex=0.3,col="red")
abline(h = m+3*s, col="black", lwd =1);abline(h = m, col="black", lwd =1);abline(h =
      m+-3*s, col="black", lwd =1)
length(returns[abs(returns-m)>3*s])/length(stock.rtn)
length(y[abs(y-m)>3*s])/length(y);2*(1-pnorm(3*s,0,s))
```

Figure 1.15 shows the mismatch between the distributional properties of market log-returns *vs.* standardized Gaussian returns, which tend to underestimate the probabilities of extreme events. Note that when $X \simeq \mathcal{N}(0, \sigma^2)$, 99.73% of samples of X are falling within the interval $[-3\sigma, +3\sigma]$, *i.e.* $\mathbb{P}(|X| \leqslant 3\sigma) = 0.9973002$.

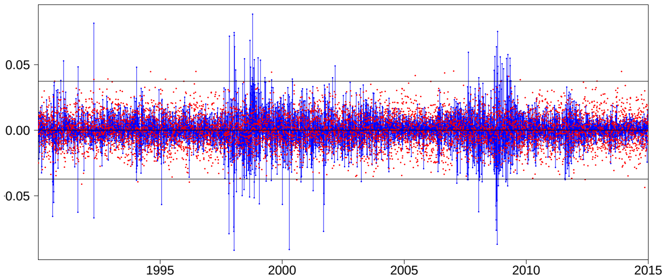


Fig. 1.15: Market returns (blue) *vs.* normalized Gaussian returns (red).

```
stock.ecdf=ecdf(as.vector(stock.rtn));x <- seq(-0.15, 0.15, length=200);px <- pnorm((x-m)/s)
dev.new(width=16,height=8)
plot(stock.ecdf, xlab = '', col="red",ylab = '', ylim=c(-0.002,1.002), main = '', xaxs="i",
      yaxs="i", las=1, cex.lab=1.8, cex.axis=1.8, lwd=4)
lines(x, px, type="l", lty=2, col="blue",xlab="",ylab="", main="", lwd=4)
legend("topleft", legend=c("Empirical CDF", "Gaussian CDF"),col=c("red", "blue"), lty=1:2,
      cex=2, lwd = 4);grid(lwd = 3)
```

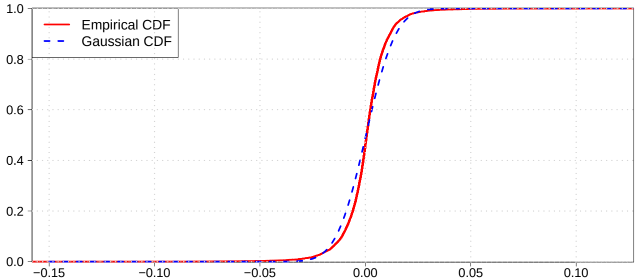


Fig. 1.16: Empirical *vs.* Gaussian CDF.

The following Quantile-Quantile plot is plotting the normalized empirical quantiles against the standard Gaussian quantiles, and is obtained with the `qqnorm(returns)` command.

```
dev.new(width=16,height=8)
qqnorm(returns, col = "blue", xaxs="i", yaxs="i", las=1, cex.lab=1.4, cex.axis=1, lwd=3)
grid(lwd = 2)
```

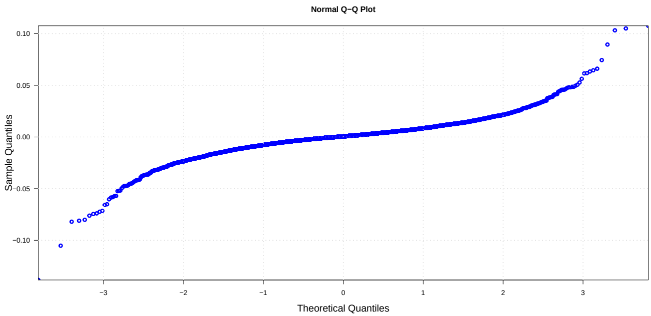


Fig. 1.17: Quantile-Quantile plot.

The following Kolmogorov-Smirnov test clearly rejects the null (normality) hypothesis of market returns.

```
n = sum(is.na(returns))+sum(!is.na(returns));x=seq(1,n);y=rnorm(n,mean=m,sd=s)
ks.test(y,"pnorm",mean=m,sd=s)
ks.test(returns,"pnorm",mean=m,sd=s)
```

```
One-sample Kolmogorov-Smirnov test

data:  returns
D = 0.075577, p-value < 2.2e-16
alternative hypothesis: two-sided
```

The mismatch in distributions observed in Figures 1.15-1.17 can be further illustrated by the empirical probability density plot in Figure 1.18, which is obtained from the following R code.

```
dev.new(width=16,height=8)
x <- seq(-0.25, 0.25, length=100);qx <- dnorm(x,mean=m,sd=s)
stock.dens=density(stock.rtn,na.rm=TRUE)
plot(stock.dens, xlab = 'x', lwd=4, col="red",ylab = '', main = '', xlim =c(-0.1,0.1),
      ylim=c(0,65), xaxs="i", yaxs="i", las=1, cex.lab=1.8, cex.axis=1.8)
lines(x, qx, type="l", lty=2, lwd=4, col="blue",xlab="x value",ylab="Density", main="")
legend("topleft", legend=c("Empirical density", "Gaussian density"),col=c("red", "blue"),
      lty=1:2, cex=1.5);grid(lwd = 2)
```

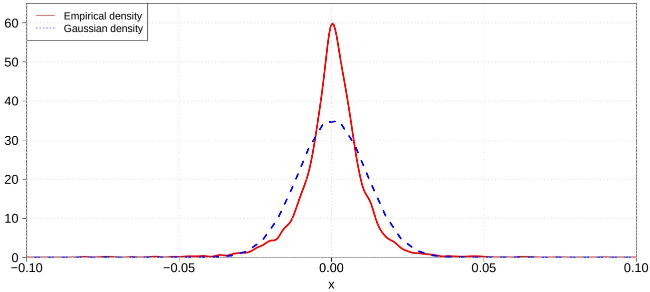


Fig. 1.18: Empirical density *vs.* normalized Gaussian density.

Form the above figure, we note that market returns have kurtosis higher than that of the normal distribution, i.e. their distribution is *leptokurtic*, in addition to showing some negative skewness.

The next R code and graph present a comparison of market prices to a calibrated lognormal distribution.

```
x <- seq(0, max(stock), length=100);qx <- dlnorm(x,mean=mean(log(stock)), sd=sd(log(stock)))
stock.dens=density(stock,na.rm=TRUE);dev.new(width=10, height=5)
plot(stock.dens, xlab = 'x', lwd=3, col="red",ylab = '', main = '', panel.first = abline(h = 0,
      col='grey', lwd =0.2), las=1, cex.axis=1, cex.lab=1, xaxs='i', yaxs='i')
lines(x, qx, type="l", lty=2, lwd=3, col="blue",xlab="x value",ylab="Density", main="")
legend("topright", legend=c("Empirical density", "Lognormal density"),col=c("red", "blue"),
      lty=1:2, cex=1.2)
```

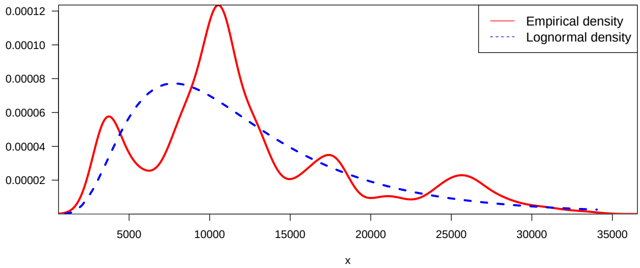


Fig. 1.19: Empirical density *vs.* normalized lognormal density.

Power tail distributions

We note that the empirical density has significantly higher kurtosis and non zero skewness in comparison with the Gaussian probability density. On the other hand, power tail probability densities of the form $\varphi(x) \simeq C_\alpha/|x|^\alpha$, $|x| \to \infty$, can provide a better fit of empirical probability density functions, as shown in Figure 1.20.

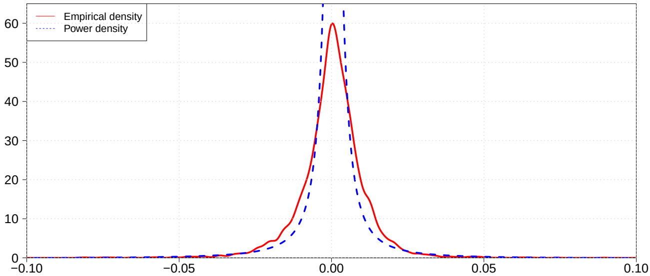


Fig. 1.20: Empirical density *vs.* power density.

The above fitting of empirical probability density function is using a power probability density function defined by a rational fraction obtained by the following R script.

```
install.packages("pracma")
library(pracma); x <- seq(-0.25, 0.25, length=1000)
stock.dens=density(returns,na.rm=TRUE, from = -0.1, to = 0.1, n = 1000)
a<-rationalfit(stock.dens$x, stock.dens$y, d1=2, d2=2)
dev.new(width=16,height=8)
plot(stock.dens$x,stock.dens$y, lwd=4, type = "l",xlab = '', col="red",ylab = '', main = '', xlim
      =c(-0.1,0.1), ylim=c(0,65), xaxs="i", yaxs="i", las=1, cex.lab=1.8, cex.axis=1.8)
lines(x,(a$p1[3]+a$p1[2]*x+a$p1[1]*x^2)/(a$p2[3]+a$p2[2]*x+a$p2[1]*x^2),
      type="l",lty=2,col="blue",xlab="x value",lwd=4, ylab="Density",main="")
legend("topleft", legend=c("Empirical density", "Power density"),col=c("red", "blue"), lty=1:2,
      cex=1.5);grid(lwd = 2)
```

The output of the **rationalfit** command is

```
$p1
[1] -0.184717249 -0.001591433  0.001385017

$p2
[1]  1.000000e+00 -6.460948e-04  1.314672e-05
```

which yields a rational fraction of the form

$$\begin{aligned} x \longmapsto & \frac{0.001385017 - 0.001591433 \times x - 0.184717249 \times x^2}{1.314672\ 10^{-5} - 6.460948\ 10^{-4} \times x + x^2} \\ \simeq & \ -0.184717249 - \frac{0.001591433}{x} + \frac{0.001385017}{x^2}, \end{aligned}$$

which approximates the empirical probability density function of DJI returns in the least squares sense.

A solution to this tail problem is to use stochastic processes with jumps, that will account for sudden variations of the asset prices. On the other hand, such jump models are generally based on the Poisson distribution which has a slower tail decay than the Gaussian distribution. This allows one to assign higher probabilities to extreme events, resulting in a more realistic modeling of asset prices. *Stable distributions* with parameter $\alpha \in (0, 2)$ provide typical examples of probability laws with power tails, as their probability density functions behave asymptotically as $x \mapsto C_\alpha / |x|^{1+\alpha}$ when $x \to \pm\infty$.

1.4 Gram-Charlier Expansions

In this section, we search for a better fit of market return distributions using the additional information provided by higher order moments. Let now

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \qquad x \in \mathbb{R},$$

denote the standard normal density function, and let

$$\Phi(x) := \int_{-\infty}^{x} \varphi(y) dy, \qquad x \in \mathbb{R},$$

denote the standard normal cumulative distribution function. Let also

$$H_n(x) := \frac{(-1)^n}{\varphi(x)} \frac{\partial^n \varphi}{\partial x^n}(x), \qquad x \in \mathbb{R},$$

denote the Hermite polynomial of degree n, with $H_0(x) = 1$. The next proposition summarizes the Gram-Charlier expansion method to obtain series expansion of a probability density function, see Gram (1883), Charlier (1914) and § 17.6 of Cramér (1946).

Proposition 1.11. *(Proposition 2.1 in Tanaka et al. (2010)) The Gram-Charlier expansion of the continuous probability density function $\phi_X(x)$ of a random variable X is given by*

$$\phi_X(x) =$$
$$\frac{1}{\sqrt{\kappa_2(X)}} \varphi\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right) + \frac{1}{\sqrt{\kappa_2(X)}} \varphi\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right) \sum_{n=3}^{\infty} c_n H_n\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right),$$

where $c_0 = 1$, $c_1 = c_2 = 0$, and the sequence $(c_n)_{n \geqslant 3}$ is given from the cumulants $(\kappa_n(X))_{n \geqslant 1}$ of X as

$$c_n = \frac{1}{(\kappa_2(X))^{n/2}} \sum_{m=1}^{[n/3]} \sum_{\substack{l_1 + \cdots + l_m = n \\ l_1, \ldots, l_m \geqslant 3}} \frac{\kappa_{l_1}(X) \cdots \kappa_{l_m}(X)}{m! l_1! \cdots l_m!}, \qquad n \geqslant 3.$$

The coefficients c_3 and c_4 can be expressed from the skewness $\kappa_3(X)/(\kappa_2(X))^{3/2}$ and the excess kurtosis $\kappa_4(X)/(\kappa_2(X))^2$ as

$$c_3 = \frac{\kappa_3(X)}{3!(\kappa_2(X))^{3/2}} \quad \text{and} \quad c_4 = \frac{\kappa_4(X)}{4!(\kappa_2(X))^2}.$$

a) The second-order expansion

$$\phi_X^{(1)}(x) = \frac{1}{\sqrt{\kappa_2(X)}} \varphi\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right)$$

corresponds to normal moment matching approximation.

b) The third-order expansion is given by

$$\phi_X^{(3)}(x) = \frac{1}{\sqrt{\kappa_2(X)}} \left(1 + c_3 H_3\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right)\right) \varphi\left(\frac{x - \kappa_1(X)}{\sqrt{\kappa_2(X)}}\right).$$

c) The fourth-order expansion is given by

$$\phi_X^{(4)}(x)$$
$$= \frac{1}{\sqrt{\kappa_2(X)}}\left(1 + c_3 H_3\left(\frac{x-\kappa_1(X)}{\sqrt{\kappa_2(X)}}\right) + c_4 H_4\left(\frac{x-\kappa_1(X)}{\sqrt{\kappa_2(X)}}\right)\right)\varphi\left(\frac{x-\kappa_1(X)}{\sqrt{\kappa_2(X)}}\right).$$

```
install.packages("SimMultiCorrData");install.packages("PDQutils")
library(SimMultiCorrData);library(PDQutils)
dev.new(width=16,height=8);m<-calc_moments(returns[!is.na(returns)]);
x <- stock.dens$x; qx <- dnorm(x,mean=m[1],sd=m[2])
plot(x,stock.dens$y, xlab = 'x', type = 'l', lwd=4, col="red",ylab = '', main = '', xlim
      =c(-0.1,0.1), ylim=c(0,65), xaxs="i", yaxs="i", las=1, cex.lab=1.8, cex.axis=1.8)
grid(lwd = 2); lines(x, qx, type="l", lty=2, lwd=4, col="blue")
cumulants<-c(m[1],m[2]**2);d2 <- dapx_edgeworth(x, cumulants)
lines(x, d2, type="l", lty=2, lwd=4, col="blue")
cumulants<-c(m[1],m[2]**2,m[3]*m[2]**3);d3 <- dapx_edgeworth(x, cumulants)
lines(x, d3, type="l", lty=2, lwd=4, col="green")
cumulants<-c(m[1],m[2]**2,0.5*m[3]*m[2]**3,0.2*m[4]*m[2]**4)
d4 <- dapx_edgeworth(x, cumulants);lines(x, d4, type="l", lty=2, lwd=4, col="purple")
legend("topleft", legend=c("Empirical density", "Gaussian density", "Third order
      Gram-Charlier", "Fourth order Gram-Charlier"),col=c("red", "blue", "green", "purple"),
      lty=1:2,cex=1.5)
```

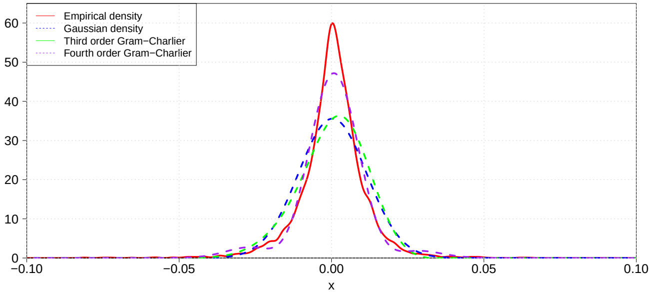


Fig. 1.21: Gram-Charlier expansions

Exercises

Exercise 1.1 Let $c > 0$. Using the definition of Brownian motion $(B_t)_{t\in\mathbb{R}_+}$, show that:

a) $(B_{c+t} - B_c)_{t\in\mathbb{R}_+}$ is a Brownian motion.
b) $(cB_{t/c^2})_{t\in\mathbb{R}_+}$ is a Brownian motion.

Exercise 1.2 Solve the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \tag{1.16}$$

defining geometric Brownian motion $(S_t)_{t\in\mathbb{R}_+}$, where $\mu, \sigma \in \mathbb{R}$.

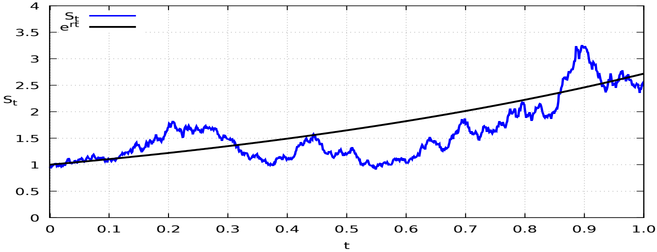


Fig. 1.22: Sample path of (1.16) with $r = 1$ and $\sigma^2 = 0.5$.

Exercise 1.3

a) Using the Gaussian moment generating function (MGF) formula (A.41), compute the *n-th* order moment $\mathbb{E}[S_t^n]$ for all $n \geqslant 1$.
b) Compute the lognormal mean and variance

$$\mathbb{E}[S_t] = S_0 e^{rt} \quad \text{and} \quad \text{Var}[S_t] = S_0^2 e^{2rt}\big(e^{\sigma^2 t} - 1\big), \quad t \geqslant 0.$$

Exercise 1.4 Consider two assets whose prices $S_t^{(1)}$, $S_t^{(2)}$ at time $t \in [0, T]$ follow the geometric Brownian dynamics

$$dS_t^{(1)} = \mu S_t^{(1)} dt + \sigma_1 S_t^{(1)} dW_t^{(1)} \quad \text{and} \quad dS_t^{(2)} = \mu S_t^{(2)} dt + \sigma_2 S_t^{(2)} dW_t^{(2)},$$

$t \in [0, T]$, where $\big(W_t^{(1)}\big)_{t\in[0,T]}$, $\big(W_t^{(2)}\big)_{t\in[0,T]}$ are two Brownian motions with correlation $\rho \in [-1, 1]$, *i.e.* we have $\mathbb{E}\big[W_t^{(1)} W_t^{(2)}\big] = \rho t$. Compute $\text{Var}\big[S_t^{(2)} - S_t^{(1)}\big]$, $t \in [0, T]$.

Exercise 1.5 We consider an economy in which individual income is modeled over time by a geometric Brownian motion

$$S_t = S_0 \mathrm{e}^{\sigma B_t + \mu t - \sigma^2 t/2}, \qquad t \geqslant 0.$$

Compute the value

$$\mathrm{Theil}_t := \log \mathbb{E}[S_t] - \mathbb{E}[\log S_t]$$

of the Theil (1967) inequality index at time $t \geqslant 0$ in this economy.

Hint: You may use the moment generating function (13.59) of the normal distribution.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Chapter 2
Time Series

Time series provide another family of models for sequences of data points indexed by discrete time. This chapter reviews the main time series models (moving average, autoregressive, integrated) and their properties, such as stationarity which is considered using autocovariance and unit root testing. Several examples are presented via the fitting of time series models to financial data in R. We conclude with an application to a pair trading algorithm on a financial market using the Dickey-Fuller stationarity test.

2.1 Autoregressive Moving Average

We use a step-by-step approach to the construction of time series, starting with the most basic setting of independent sequences, and building progressively towards more interaction. In what follows, we let

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

denote the set of all (signed) integers.

White noise

A white noise sequence is a sequence $(Z_n)_{n\in\mathbb{Z}}$ of independent, centered and identically distributed random variables with unit variance, with

$$\mathbb{E}[Z_n] = 0, \quad \text{and} \quad \mathrm{Cov}(Z_n, Z_m) = \mathbb{1}_{\{n=m\}}, \qquad n, m \in \mathbb{Z}.$$

```
Zn<-rnorm(100,0,1)
Zn
```

Moving Average (MA) model

Definition 2.1. *In the MA(q) model of order $q \geqslant 1$, the current state X_n of the system is expressed as the linear combination*

$$\begin{aligned} X_n &:= Z_n + \beta_1 Z_{n-1} + \cdots + \beta_q Z_{n-q} \\ &= Z_n + \sum_{k=1}^{q} \beta_k Z_{n-k}, \qquad n \geqslant 0, \end{aligned} \tag{2.1}$$

of the q previous values $Z_{n-1}, \ldots, Z_{n-q}$. Here, $\beta_1, \ldots, \beta_q$ is a sequence of deterministic coefficients such that $\beta_q \neq 0$.

We will use the "lag operator" or "backward time shift operator" L defined as

$$LZ_n := Z_{n-1}, \qquad n \geqslant 1. \tag{2.2}$$

```
library(zoo)
N=5; Zn<-zoo(rnorm(N,0,1))
Zn
# Lag operator
lag(Zn,-1, na.pad = TRUE)
```

The lag operator L can be iterated as

$$L^k Z_n = Z_{n-k}, \qquad n \in \mathbb{Z},\ k \geqslant 0,$$

and can be used to rewrite (2.1) as

$$X_n = Z_n + \beta_1 L Z_n + \cdots + \beta_q L^q Z_n$$

$$
\begin{aligned}
&= Z_n + \sum_{k=1}^{q} \beta_k L^k Z_n \\
&= Z_n + \psi(L) Z_n, \qquad n \geqslant q,
\end{aligned}
$$

where

$$\psi(L) := \beta_1 L + \cdots + \beta_q L^q = \sum_{k=1}^{q} \beta_k L^k$$

is the *moving average operator* given by the function

$$\psi(x) := \beta_1 x + \cdots + \beta_q x^q = \sum_{k=1}^{q} \beta_k x^k.$$

Example: generating MA(2) samples in R

The next R code generates samples of the MA(2) times series

$$X_n := Z_n - 0.7 \times Z_{n-1} + 0.1 \times Z_{n-2},$$

with $\beta_1 = -0.7$ and $\beta_2 = 0.1$.

```
n=41
ma.sim<-arima.sim(model=list(ma=c(-.7,.1)),n.start=100,n)
x=seq(100,100+n-1);dev.new(width=12, height=6)
plot(x,ma.sim,pch=19, ylab="", xlab="n",main = 'MA(2) Samples',col='blue',cex.axis=1.8,
      cex.lab=1.5,las=1)
lines(x,ma.sim,col='blue');grid()
```

The ARIMA command uses a parameter "n.start", here taken equal to 100, which creates a "burn-in" initial time interval which ensures sufficient randomness in the behavior of X_n.

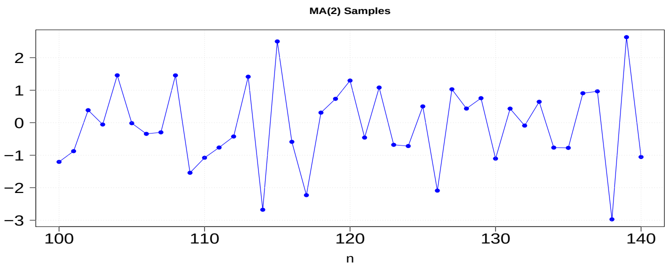


Fig. 2.1: MA(2) Samples.

Autoregressive (AR) model

In the simplest AR(1) model, the current state X_n of the system is expressed in feedback form as

$$X_n := Z_n + \alpha_1 X_{n-1}, \qquad n \geqslant 1, \tag{2.3}$$

Definition 2.2. *In the AR(p) model, $p \geqslant 1$, the state X_n of the system is expressed as the linear feedback combination*

$$\begin{aligned} X_n &:= Z_n + \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p} \\ &= Z_n + \sum_{k=1}^{p} \alpha_k X_{n-k}, \qquad n \geqslant p, \end{aligned} \tag{2.4}$$

of the p previous values $X_{n-1}, \ldots, X_{n-p}$ of the time series, where $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ is a sequence of deterministic coefficients such that $\alpha_p \neq 0$.

Using again the lag operator L defined in (2.2), we can rewrite (2.4) as

$$\begin{aligned} X_n &= Z_n + \alpha_1 L X_n + \cdots + \alpha_p L^p X_n \\ &= Z_n + \sum_{k=1}^{p} \alpha_k L^k X_n \\ &= Z_n + \phi(L) X_n, \qquad n \geqslant p, \end{aligned}$$

where

$$\phi(L) := \alpha_1 L + \cdots + \alpha_p L^p = \sum_{k=1}^{p} \alpha_k L^k$$

is the operator given by the function

$$\phi(x) := \alpha_1 x + \cdots + \alpha_p x^p = \sum_{k=1}^{p} \alpha_k x^k.$$

Proposition 2.3. *The equation*

$$X_n := Z_n + \alpha_1 X_{n-1}, \qquad n \in \mathbb{Z}, \tag{2.5}$$

defines an AR(1) time series $(X_n)_{n\in\mathbb{Z}}$, and can be solved recursively in the following cases:

a) When $|\alpha_1| < 1$, (2.5) admits the converging causal *moving average solution*

$$X_n = \sum_{k\geqslant 0} \alpha_1^k Z_{n-k}, \qquad n \in \mathbb{Z}, \tag{2.6}$$

with

$$\mathrm{Var}[X_n] = \sum_{k\geqslant 0} |\alpha_1|^{2k} = \frac{1}{1-|\alpha_1|^2} \geqslant 1, \qquad n \in \mathbb{Z}. \tag{2.7}$$

b) When $|\alpha_1| > 1$, (2.5) *admits the converging* non-causal *moving average solution*

$$X_n = -\sum_{k\geqslant 1} \frac{1}{\alpha_1^k} Z_{n+k}, \qquad n \in \mathbb{Z}, \tag{2.8}$$

with

$$\mathrm{Var}[X_n] = \sum_{k\geqslant 1} |\alpha_1|^{-2k} = \frac{1}{|\alpha_1|^2 - 1}, \qquad n \in \mathbb{Z}. \tag{2.9}$$

No such converging solutions exist when $|\alpha_1| = 1$.

Proof. a) When $|\alpha_1| < 1$ we may write, using backward induction,

$$\begin{aligned}
X_n &= Z_n + \alpha_1 X_{n-1} \\
&= Z_n + \alpha_1(Z_{n-1} + \alpha_1 X_{n-2}) \\
&= Z_n + \alpha_1(Z_{n-1} + \alpha_1(Z_{n-2} + \alpha_1 X_{n-3})) \\
&= Z_n + \alpha_1(Z_{n-1} + \alpha_1(Z_{n-2} + \alpha_1(Z_{n-3} + \alpha_1 X_{n-4}))) \\
&= Z_n + \alpha_1 Z_{n-1} + \alpha_1^2 Z_{n-2} + \alpha_1^3 Z_{n-3} + \alpha_1^4 X_{n-4} \\
&= \cdots \\
&= \sum_{k\geqslant 0} \alpha_1^k Z_{n-k}, \quad n \in \mathbb{Z},
\end{aligned}$$

which converges when the solution $z = 1/\alpha_1$ of the equation $\phi(z) = \alpha_1 z = 1$ satisfies $|\alpha_1| < 1$, *i.e.* $|z| > 1$.

b) When $|\alpha_1| > 1$, we write

$$X_n = -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1} X_{n+1}, \qquad n \geqslant 0,$$

which can be solved by forward induction as

$$\begin{aligned}
X_n &= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1} X_{n+1} \\
&= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+2} + \alpha_1^{-1} X_{n+2}) \\
&= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+2} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+3} + \alpha_1^{-1} X_{n+3})) \\
&= -\alpha_1^{-1} Z_{n+1} - \alpha_1^{-2} Z_{n+2} - \alpha_1^{-3} Z_{n+3} + \alpha_1^{-4} X_{n+3} \\
&= \cdots
\end{aligned}$$

$$= -\sum_{k \geqslant 1} \frac{1}{\alpha_1^k} Z_{n+k}, \quad n \in \mathbb{Z},$$

when the solution $z = 1/\alpha_1$ of the equation $\phi(z) = \alpha_1 z = 1$ satisfies $|z| < 1$. $\square$

Example: generating AR(2) samples in R

The next R code generates samples of the AR(2) times series

$$X_n := Z_n + 0.9 \times X_{n-1} - 0.2 \times X_{n-2}, \tag{2.10}$$

with $\alpha_1 = 0.9$ and $\alpha_2 = -0.1$.

```
n=41; ar.sim<-arima.sim(model=list(ar=c(.9,-.2)),n.start=100,n)
x=seq(100,100+n-1); dev.new(width=12, height=6)
plot(x,ar.sim,pch=19, ylab="", xlab="n", main = 'AR(2) Samples',col='blue',cex.axis=1.8,
      cex.lab=1.5,las=1)
lines(x,ar.sim,col='blue');grid()
```

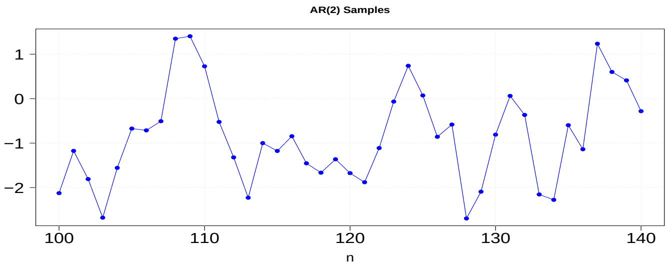


Fig. 2.2: AR(2) Samples.

Autoregressive Moving Average (ARMA) model

Definition 2.4. *In the ARMA(p, q) model with orders $p \geqslant 1$ and $q \geqslant 1$, the current state X_n of the system is expressed as the linear combination*

$$\begin{aligned} X_n &:= Z_n + \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p} + \beta_1 Z_{n-1} + \cdots + \beta_p Z_{n-q} \\ &= Z_n + \sum_{k=1}^{p} \alpha_k X_{n-k} + \sum_{k=1}^{q} \beta_k Z_{n-k}, \end{aligned} \tag{2.11}$$

of the p previous values $X_{n-1}, \ldots, X_{n-p}$ *and* $Z_{n-1}, \ldots, Z_{n-p}$, $n \geqslant \mathrm{Max}(p, q)$, *and* $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ *are sequences of nonnegative deterministic coefficients such that* $\alpha_p \neq 0$ *and* $\beta_q \neq 0$.

Using again the lag operator L defined in (2.2), we can rewrite (2.11) as

$$\begin{aligned} X_n &= Z_n + \sum_{k=1}^{p} \alpha_k L^k X_n + \sum_{k=1}^{q} \beta_k L^k Z_n \\ &= Z_n + \phi(L) X_n + \psi(L) Z_n, \qquad n \geqslant 1. \end{aligned}$$

Example: generating ARMA$(2,2)$ samples in R

The next R code generates samples of the ARMA$(2,2)$ times series

$$X_n := Z_n + 0.9 \times X_{n-1} - 0.2 \times X_{n-2} - 0.7 \times Z_{n-1} + 0.1 \times Z_{n-2},$$

with $\alpha_1 = 0.9$, $\alpha_2 = -0.1$, and $\beta_1 = -0.7$, $\beta_2 = 0.1$.

```r
n=41; arma.sim<-arima.sim(model=list(ar=c(.9,-.2),ma=c(-.7,.1)),n.start=100,n)
x=seq(100,100+n-1); dev.new(width=12, height=6)
plot(x,arma.sim,pch=19, ylab="", xlab="n", main = 'ARMA(2,2)
      Samples',col='blue',cex.axis=1.8, cex.lab=1.5,las=1)
lines(x,arma.sim,col='blue');grid()
```

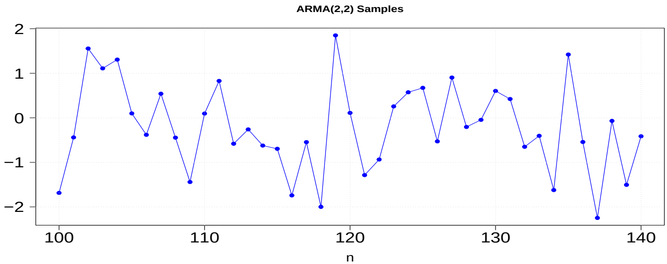


Fig. 2.3: ARMA$(2,2)$ Samples.

Autoregressive Integrated Moving Average

Consider the difference operator ∇ defined as

$$\nabla := \mathrm{I} - L$$

where I is the identity operator and L is the lag operator, so that

$$\nabla X_n := X_n - X_{n-1}, \qquad n \geqslant 1.$$

In addition, the operation $\nabla X_n = X_n - X_{n-1}$ can be iterated as follows:

$$\begin{aligned}\nabla^2 X_n &= \nabla\nabla X_n \\ &= \nabla(X_n - X_{n-1}) \\ &= \nabla X_n - \nabla X_{n-1} \\ &= X_n - X_{n-1} - (X_{n-1} - X_{n-2}) \\ &= X_n - 2X_{n-1} + X_{n-2},\end{aligned}$$

and

$$\begin{aligned}\nabla^3 X_n &= \nabla\nabla^2 X_n \\ &= \nabla X_n - 2\nabla X_{n-1} + \nabla X_{n-2} \\ &= X_n - X_{n-1} - 2(X_{n-1} - X_{n-2}) + X_{n-2} - X_{n-3} \\ &= X_n - 3X_{n-1} + 3X_{n-2} - X_{n-3}.\end{aligned}$$

The time series $(X_n)_{k\geqslant 1}$ can be recovered by integrating $(\nabla X_k)_{k\geqslant 1}$ using the telescoping identity

$$X_n = X_0 + \sum_{k=1}^n (X_k - X_{k-1}) = X_0 + \sum_{k=1}^n \nabla X_k, \qquad n \geqslant 1. \tag{2.12}$$

More generally, the time series $(X_n)_{n\geqslant 0}$ can be recovered by successive applications of the discrete integration formula (2.12) as in the next proposition, where we use the convention $\nabla^0 = \mathrm{I}$.

Proposition 2.5. *a) The iterated operator ∇^d satisfies*

$$\nabla^d X_n = \sum_{k=0}^d \binom{d}{k} (-1)^k X_{n-k}, \qquad n \geqslant d \geqslant 0.$$

b) The time series $(X_n)_{n\geqslant d}$ can be recovered from $\nabla X_n, \nabla^2 X_n, \ldots, \nabla^d X_n$ as

$$X_n = X_{n-d} + \sum_{k=1}^d \binom{d}{k} \nabla^k X_{n+k-d}, \qquad n \geqslant d \geqslant 0.$$

Proof. (a) This is a consequence of the binomial operator identity

$$\nabla^d = (\mathrm{I} - L)^d$$

$$
\begin{aligned}
&= \sum_{k=0}^{d} \binom{d}{k} (\mathrm{I})^{n-k} (-L)^k \\
&= \sum_{k=0}^{d} \binom{d}{k} (-1)^k L^k.
\end{aligned}
$$

(b) Apply the binomial operator identity

$$
\mathrm{I} = (\mathrm{I} - L + L)^d = (L + \nabla)^d = \sum_{k=0}^{d} \binom{d}{k} L^{d-k} \nabla^k.
$$

□

Definition 2.6. *In the ARIMA(p, d, q) model, the iterated difference process $(\nabla^d X_n)_{n \geqslant 0}$ is modeled as the ARMA(p, q) time series*

$$
\begin{aligned}
\nabla^d X_n &:= Z_n + \alpha_1 \nabla^d X_{n-1} + \cdots + \alpha_p \nabla^d X_{n-p} \\
&\quad + \beta_1 Z_{n-1} + \cdots + \beta_q Z_{n-q} \\
&= Z_n + \sum_{k=1}^{p} \alpha_k \nabla^d X_{n-k} + \sum_{k=1}^{q} \beta_k Z_{n-k}, \qquad (2.13)
\end{aligned}
$$

$n \geqslant \mathrm{Max}(p + d, q + d)$, *based on the p previous values $\nabla^d X_{n-1}, \ldots, \nabla^d X_{n-p}$ and $Z_{n-1}, \ldots, Z_{n-q}$, where and $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ are sequences of non-negative deterministic coefficients such that $\alpha_p \neq 0$ and $\beta_q \neq 0$.*

Using again the lag operator L defined in (2.2), we can rewrite (2.13) as

$$
\nabla^d X_n = Z_n + \phi(L) \nabla^d X_n + \psi(L) Z_n,
$$

$n \geqslant \mathrm{Max}(p + d, q + d)$, where the functions $\phi(z)$ and $\psi(z)$ are given by

$$
\phi(z) = \sum_{k=1}^{p} \alpha_k z^k \quad \text{and} \quad \psi(z) = \sum_{k=1}^{p} \beta_k z^k.
$$

In other words, we have

$$
(\mathrm{I} - \phi(L)) \nabla^d X_n = Z_n + \psi(L) Z_n,
$$

$n \geqslant \mathrm{Max}(p + d, q + d)$. The time series $(X_n)_{n \geqslant 0}$ can then be recovered by successive applications of the discrete integration formula (2.12) as in Proposition 2.5-(b). Alternatively, we can start by recovering $\nabla^{d-1} X_n$ from $\nabla^d X_n$

as

$$\begin{aligned}\nabla^{d-1} X_n &= \nabla^{d-1} X_0 + \sum_{k=1}^{n} \left(\nabla^{d-1} X_k - \nabla^{d-1} X_{k-1} \right) \\ &= \nabla^{d-1} X_0 + \sum_{k=1}^{n} \nabla^{d} X_k,\end{aligned}$$

followed by

$$\nabla^{d-2} X_n, \nabla^{d-3} X_n, \ldots, \nabla^2 X_n, \nabla X_n, X_n,$$

by induction on $n \geqslant d$.

Example: generating $\mathrm{ARIMA}(1,2,3)$ samples in R

The next R code generates samples of the $\mathrm{ARIMA}(1,2,3)$ times series $(X_n)_{n \geqslant 5}$ defined from (2.12) and

$$\nabla X_n := Z_n + 0.5 \times X_{n-1} + 0.5 \times Z_{n-1} + 0.5 \times Z_{n-2} - 0.5 \times Z_{n-3},$$

with $\alpha_1 = 0.5$ and $\beta_1 = 0.5$, $\beta_2 = 0.5$, $\beta_3 = -0.5$.

```
n=41;d=2
arima.sim<-arima.sim(model=list(ar=c(0.5),ma=c(0.5, 0.5, -0.5), order=c(1,d,3)), n.start=100,n)
x=seq(100,100+n+d-1); dev.new(width=12, height=6)
plot(x,arima.sim,pch=19, ylab="", xlab="n", main = paste("ARIMA(1,",d,",3) Samples", sep=""),cex.axis=1.8, cex.lab=1.5,las=1)
lines(x,arima.sim,col='blue');grid()
```

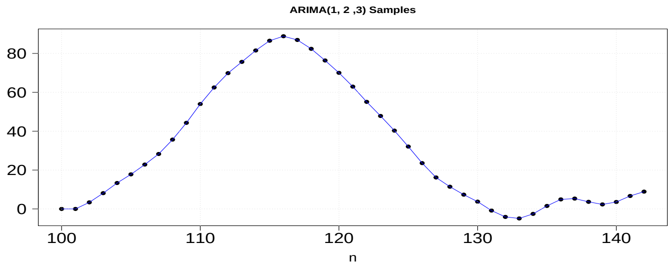


Fig. 2.4: $\mathrm{ARIMA}(1,2,3)$ Samples.

Note that the ARIMA graph of Figure 2.4 has more potential for prediction than the ARMA graph of Figure 2.3 due to increased dependence on past samples in the considered model, or longer memory.

2.2 Autoregressive Heteroskedasticity

As above, we consider a white noise sequence $(Z_n)_{n\in\mathbb{Z}}$ of independent, centered and identically distributed with unit variance.

Autoregressive Conditional Heteroskedasticity (ARCH) model

Heteroskedasticity refers to time-dependent variance in a time series.

Definition 2.7. *In the ARCH(p) model of order $p \geqslant 1$, the current state X_n of the system is expressed by the equation*

$$X_n := \sigma_n Z_n, \qquad n \geqslant 0, \tag{2.14}$$

where $\sigma_n > 0$ is given by

$$\begin{aligned}\sigma_n^2 &= \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{n-k}^2 \\ &= \alpha_0 + \sum_{k=1}^{p} \alpha_k \sigma_{n-k}^2 Z_{n-k}^2, \qquad n \geqslant p,\end{aligned} \tag{2.15}$$

and $\alpha_0, \ldots, \alpha_p \geqslant 0$ is a sequence of nonnegative deterministic coefficients such that $\alpha_p \neq 0$.

Using the lag operator L defined in (2.2), we can rewrite (2.15) as

$$\begin{aligned}\sigma_n^2 &= \alpha_0 + \sum_{k=1}^{p} \alpha_k L^k X_n^2 \\ &= \alpha_0 + \psi(L) X_n^2, \qquad n \geqslant p,\end{aligned}$$

where

$$\psi(L) := \alpha_1 L + \cdots + \alpha_p L^p = \sum_{k=1}^{p} \alpha_k L^k.$$

In particular, noting that σ_n is independent of Z_n since σ_n depends only on $Z_{n-1}, Z_{n-2}, \ldots$, we have $\mathbb{E}[X_n] = 0$ and

$$\mathbb{E}[X_n^2] = \mathbb{E}[\sigma_n^2 Z_n^2] = \mathbb{E}[Z_n^2]\mathbb{E}[\sigma_n^2] = \mathbb{E}[\sigma_n^2],$$

hence the recursion

$$\mathbb{E}[\sigma_n^2] = \alpha_0 + \sum_{k=1}^{p} \alpha_k \mathbb{E}[X_{n-k}^2] = \alpha_0 + \sum_{k=1}^{p} \alpha_k \mathbb{E}[\sigma_{n-k}^2].$$

Example: generating ARCH(2) samples in R

The next R code generates samples of the ARCH(2) times series $(X_n)_{n\geqslant 5}$ with variance process given by

$$\sigma_n^2 = \alpha_0 + 0.2 \times X_{n-1}^2 + 0.4 \times X_{n-2}^2, \qquad n \geqslant 2, \tag{2.16}$$

with $\alpha_1 = 0.2$ and $\alpha_2 = 0.4$.

```
library(fGarch)
arch.spec<-garchSpec(model=list(alpha0 = 10E-6,alpha=c(0.2,0.4),beta = 0))
n=100; arch.sim<-garchSim(arch.spec,n); x=seq(1,n); dev.new(width=12, height=6)
plot(x,arch.sim,pch=19, ylab="", xlab="n", main = 'ARCH(2) Samples', col='blue',
      cex.axis=1.5, cex.lab=1.7, las=1)
lines(x,arch.sim,col='blue');grid()
```

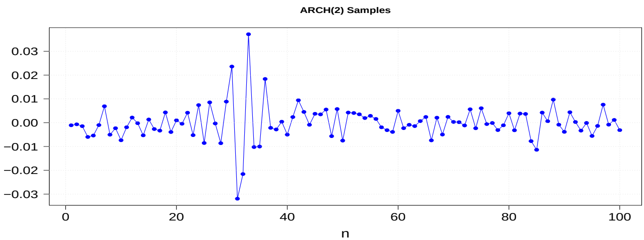


Fig. 2.5: ARCH(2) Samples.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model

Definition 2.8. *In the GARCH(p, q) model with orders $p \geqslant 1$ and $q \geqslant 1$, the current state X_n of the system is expressed by the equation (2.14), where $\sigma_n > 0$ is given by*

$$\begin{aligned} \sigma_n^2 &= \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{n-k}^2 + \sum_{k=1}^{q} \beta_k \sigma_{n-k}^2 \qquad (2.17) \\ &= \alpha_0 + \sum_{k=1}^{p} \alpha_k \sigma_{n-k}^2 Z_{n-k}^2 + \sum_{k=1}^{q} \beta_k \sigma_{n-k}^2, \quad n \geqslant \mathrm{Max}(p, q), \end{aligned}$$

and $\alpha_0, \ldots, \alpha_p \geqslant 0$, $\beta_1, \ldots, \beta_q \geqslant 0$ are sequences of nonnegative deterministic coefficients such that $\alpha_p \neq 0$ and $\beta_q \neq 0$.

Using the lag operator L, we can rewrite (2.17) as

$$\begin{aligned}\sigma_n^2 &= \alpha_0 + \sum_{k=1}^{p} \alpha_k L^k X_n^2 + \sum_{k=1}^{q} \beta_k L^k \sigma_n^2 \\ &= \alpha_0 + \phi(L) X_n^2 + \psi(L) \sigma_n^2, \qquad n \geqslant \mathrm{Max}(p,q),\end{aligned}$$

where

$$\phi(L) := \alpha_1 L + \cdots + \alpha_q L^q = \sum_{k=1}^{q} \alpha_k L^k$$

and

$$\psi(L) := \beta_1 L + \cdots + \beta_q L^q = \sum_{k=1}^{q} \beta_k L^k.$$

As above, we have

$$\mathbb{E}[X_n^2] = \mathbb{E}[\sigma_n^2 Z_n^2] = \mathbb{E}[Z_n^2]\mathbb{E}[\sigma_n^2] = \mathbb{E}[\sigma_n^2],$$

hence

$$\begin{aligned}\mathbb{E}[X_n^2] &= \alpha_0 + \sum_{k=1}^{p} \alpha_k \mathbb{E}[X_{n-k}^2] + \sum_{k=1}^{q} \beta_k \mathbb{E}[\sigma_{n-k}^2] \\ &= \alpha_0 + \sum_{k=1}^{p} \alpha_k \mathbb{E}[\sigma_{n-k}^2] + \sum_{k=1}^{q} \beta_k \mathbb{E}[\sigma_{n-k}^2], \quad n \geqslant \mathrm{Max}(p,q).\end{aligned}$$

Example: generating GARCH$(2,1)$ samples in R

The next R code generates samples of the GARCH$(2,1)$ times series times series $(X_n)_{n\geqslant 5}$ with variance process given by

$$\sigma_n^2 = \alpha_0 + 0.2 \times X_{n-1}^2 + 0.4 \times X_{n-2}^2 + 0.3 \times \sigma_{n-1}^2, \quad n \geqslant 2, \qquad (2.18)$$

where $\alpha_1 = 0.2$, $\alpha_2 = 0.4$, and $\beta_1 = 0.3$.

```
library(fGarch)
garch.spec<-garchSpec(model=list(alpha0 = 10E-6,alpha=c(0.2,0.4),beta = c(0.3)))
n=100; garch.sim<-garchSim(garch.spec,n); x=seq(1,n); dev.new(width=12, height=6)
plot(x,garch.sim,pch=19, ylab="", xlab="n", main = 'GARCH(2,1) Samples', col='blue',
        cex.axis=1.5, cex.lab=1.7, las=1)
lines(x,garch.sim,col='blue');grid()
```

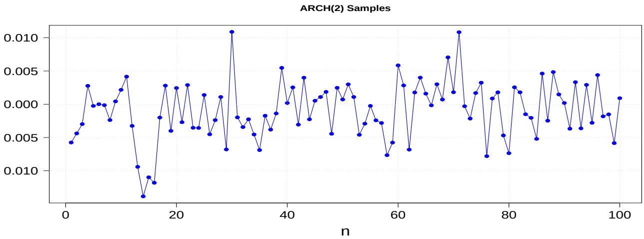


Fig. 2.6: GARCH$(2,1)$ Samples.

Similarly to Proposition 2.3, we have the following result.

Proposition 2.9. *The equations* $X_n = \sigma_n Z_n$ *and*

$$\sigma_n^2 := \alpha_0 + \alpha_1 X_{n-1}^2 + \beta_1 \sigma_{n-1}^2, \quad n \in \mathbb{Z},$$

with $\alpha_1, \beta_1 \geqslant 0$ *and* $\alpha_1 + \beta_1 < 1$*, define a GARCH*$(1,1)$ *time series* $(X_n)_{n\in\mathbb{Z}}$ *which admits the causal solution*

$$\sigma_n^2 = \alpha_0 \sum_{k\leqslant n} \prod_{l=k}^{n-1} \left(\alpha_1 Z_l^2 + \beta_1\right), \qquad n \in \mathbb{Z}, \tag{2.19}$$

with

$$\mathbb{E}[\sigma_n^2] = \alpha_0 \sum_{k\geqslant 0} (\alpha_1 + \beta_1)^k = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \qquad n \in \mathbb{Z}. \tag{2.20}$$

No such converging solution exists when $\alpha_1 + \beta_1 \geqslant 1$*.*

2.3 Time Series Stationarity

Strict stationarity

Definition 2.10. *A time series* $(X_n)_{n\in\mathbb{Z}}$ *is* strictly stationary *with order* $p \geqslant 1$ *if the equality in distribution*

$$(X_n, X_{n-1}, \ldots, X_{n-p}) \overset{d}{\simeq} (X_{n+m}, X_{n+m-1}, \ldots, X_{n+m-p}),$$

holds for all $n \in \mathbb{Z}$ *and* $m, p \geqslant 0$*.*

In other words, Definition 2.10 states that the random vectors

$$(X_n, X_{n-1}, \ldots, X_{n-p}) \quad \text{and} \quad (X_{n+m}, X_{n+m-1}, \ldots, X_{n+m-p})$$

have same distribution for all $m \in \mathbb{Z}$ and $p \geqslant 1$.

Example. The MA(q) time series

$$\begin{aligned} X_n &= Z_n + \beta_1 Z_{n-1} + \cdots + \beta_q Z_{n-q} \\ &= Z_n + \sum_{k=1}^{q} \beta_k Z_{n-k}, \qquad n \geqslant q, \end{aligned}$$

is *strictly stationary*, due to the equality in distribution

$$\begin{aligned} &\left(Z_n + \sum_{k=1}^{q} \beta_k Z_{n-k}, \ldots, Z_{n-p} + \sum_{k=1}^{q} \beta_k Z_{n-p-k} \right) \\ &\quad \overset{d}{\simeq} \left(Z_{n+m} + \sum_{k=1}^{q} \beta_k Z_{n+m-k}, \ldots, Z_{n+m-p} + \sum_{k=1}^{q} \beta_k Z_{n+m-p-k} \right), \end{aligned}$$

as $(Z_n)_{n \geqslant 0}$ is an *i.i.d.* sequence.

Weak stationarity

Definition 2.11. *A time series $(X_n)_{n \geqslant 0}$ is* weakly stationary *if*

i) $\mathbb{E}[X_n] = \mathbb{E}[X_0]$, $n \geqslant 0$, *and*

ii) the autocovariance [*]

$$(n, m) \longmapsto \mathrm{Cov}(X_n, X_m)$$

depends only on the absolute difference $|n - m|$, $n, m \geqslant 0$.

The autocovariances $\mathrm{Cov}(X_n, X_{n+l})$ and cross-covariances $\mathrm{Cov}(Y_n, X_{n+l})$ of time series with lag parameter $l \in \mathbb{Z}$ can be respectively estimated as follows:

```
n=1000; ar.sim<-arima.sim(model=list(ar=c(.9,-.2)),n.start=100,n); dev.new(width=12,
    height=6)
ar.acf<-acf(ar.sim,type="covariance",plot=T,col='blue',lwd=4); dev.new(width=12, height=6)
ar.ccf<-ccf(ar.sim,ar.sim,type="covariance",plot=T,lwd=4,col='blue',main='',cex.axis=1.8,
    cex.lab=1.5,las=1);grid()
```

[*] The covariance $\mathrm{Cov}(X, Y)$ is defined as $\mathrm{Cov}(X, Y) := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

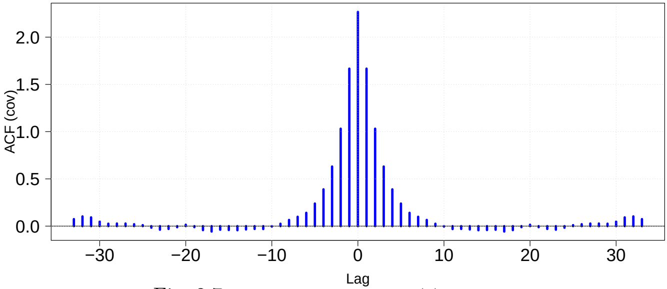


Fig. 2.7: Autocovariances of AR(2) Samples.

By representing an AR(q) series as a vector-valued AR(1) series we can obtain the following result, see *e.g.* Theorem 3.1.1 page 89 of Brockwell and Davis (1991) and Theorem 4.4 page 119 of Pourahmadi (2001).

Theorem 2.12. *Consider the AR(q) time series $(X_n)_{n\geqslant q}$ solution of*

$$X_n := Z_n + \phi(L)X_n = Z_n + \alpha_1 X_{n-1} + \cdots + \alpha_q X_{n-q},$$

with

$$\phi(z) = \alpha_1 z + \cdots + \alpha_q z^q, \qquad z \in \mathbb{C}.$$

1) Unit root test. *The AR(q) time series $(X_n)_{n\geqslant 0}$ is* weakly stationary *with lag order q if and only if* no *(complex) solution of the equation $\phi(z) = 1$ lies* on *the complex unit circle $\{z \in \mathbb{C} : |z| = 1\}$ in the complex plane $\mathbb{C}$.**

2) Causality. *The AR(q) time series $(X_n)_{n\geqslant q}$ admits a* causal *expression if and only if the equation $\phi(z) = 1$ has no solution inside the complex unit disk $\{z \in \mathbb{C} : |z| \leqslant 1\}$.*

Examples

i) In the AR(1) example

$$X_n := Z_n + \alpha_1 X_{n-1} = Z_n + \phi(L)X_n, \qquad n \geqslant 1,$$

with $\phi(z) = \alpha_1 z$, the unique solution $z = 1/\alpha_1$ of the equation $\phi(z) = \alpha_1 z = 1$ lies on the complex unit circle if and only if $\alpha_1 \neq \pm 1$, *i.e.* $|\alpha_1| \neq 1$. Hence, by Theorem 2.12 the time series $(X_n)_{n\geqslant 2}$ is (weakly) stationarity if and only if $|\alpha_1| \neq 1$.

* See (MOE and UCLES 2022, page 15) and (MOE and UCLES 2020, page 20).

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

In this case, by Proposition 2.3 we have the series representations

$$\begin{cases} X_n = \sum_{k\geqslant 0} \alpha_1^k Z_{n-k}, \qquad |\alpha_1| < 1, & (2.21a) \\ X_n = -\sum_{k\geqslant 1} \frac{1}{\alpha_1^k} Z_{n+k}, \qquad |\alpha_1| > 1, & (2.21b) \end{cases}$$

which converge when $|\alpha_1| \neq 1$, with

$$\mathrm{Var}[X_n] = \mathbb{E}\left[X_n^2\right] = \frac{1}{|1-|\alpha_1|^2|}, \qquad n \geqslant 1,$$

see (2.7) and (2.9).

a) In the case of (2.21a) with $|\alpha_1| < 1$, the causal solution

$$X_n = \sum_{k\geqslant 0} \alpha_1^k Z_{n-k}, \qquad n \geqslant 0,$$

satisfies

$$\begin{aligned} \mathrm{Cov}(X_n, X_m) &= \mathbb{E}\left[\sum_{k\geqslant 0} \alpha_1^k Z_{n-k} \sum_{l\geqslant 0} \overline{\alpha}_1^l Z_{m-l}\right] \\ &= \alpha_1^{n-m} \sum_{k\geqslant 0} |\alpha_1|^{2k} \\ &= \frac{\alpha_1^{n-m}}{1-|\alpha_1|^2}, \qquad n \geqslant m \geqslant 0. \end{aligned}$$

b) In the case of (2.21b) with $|\alpha_1| > 1$, the non-causal solution

$$X_n = -\sum_{k\geqslant 1} \frac{1}{\alpha_1^k} Z_{n+k}, \qquad n \geqslant 0,$$

satisfies

$$\begin{aligned} \mathrm{Cov}(X_n, X_m) &= \mathbb{E}\left[\sum_{k\geqslant 1} \frac{1}{\alpha_1^k} Z_{n+k} \sum_{l\geqslant 1} \frac{1}{\overline{\alpha}_1^l} Z_{m+l}\right] \\ &= \frac{\alpha_1^{m-n}}{|\alpha_1|^2 - 1}, \qquad n \geqslant m \geqslant 0. \end{aligned}$$

We note that the expressions (2.6)-(2.7) in the case $|\alpha_1| < 1$ correspond to strictly stationary times series, while (2.8)-(2.9) in the case $|\alpha_1| > 1$ correspond to weakly, but not strictly, stationary times series.

ii) In the AR(2) example

$$X_n := Z_n + 0.9 \times X_{n-1} - 0.2 \times X_{n-2}$$

of Figure 2.2 with $\phi(z) = 0.9z - 0.2z^2$, the solutions

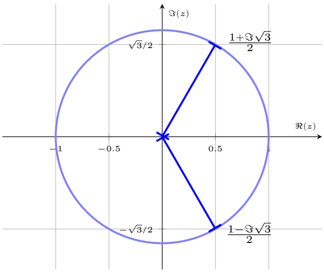
$$z_+ = \frac{0.9 + \sqrt{0.9^2 - 4 \times 0.2}}{2 \times 0.2} = \frac{5}{2}, \quad z_- = \frac{0.9 - \sqrt{0.9^2 - 4 \times 0.2}}{2 \times 0.2} = 2$$

of the equation $\phi(z) = 1$ do not lie on the complex unit circle, hence by Theorem 2.12 the time series $(X_n)_{n \geqslant 2}$ is (weakly) stationarity.

iii) Consider the AR(2) time series

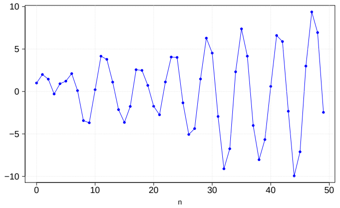
$$X_n := Z_n + X_{n-1} - X_{n-2} = Z_n + \phi(L)X_n, \qquad n \geqslant 2,$$

with $\phi(z) = z - z^2$.

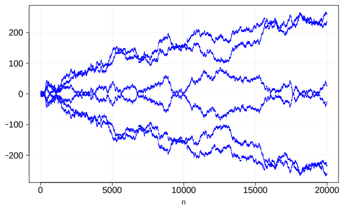


The solutions $z = (1 \pm i\sqrt{3})/2$ of the equation $\phi(z) = 1$ lie on the unit circle, hence by Theorem 2.12 the time series $(X_n)_{n \geqslant 2}$ is *not* (weakly) stationarity. The next Figure 2.8 presents a simulation of non-stationary time series according to the attached **R code**.*

* Right-click to save as attachment (may not work on 🍎).



(a) 50 time steps.



(b) $20,000$ time steps.

Fig. 2.8: Nonstationarity of AR(2) time series with $a_1 = 1$ and $a_2 = -1$.

Stationarity test

The Dickey-Fuller test allows us to test the null (nonstationarity) hypothesis H_0, *i.e.* "$|\alpha_1| = 1$", against the alternative (stationarity) hypothesis "$|\alpha_1| \neq 1$".

a) Under the alternative (stationarity) hypothesis $|\alpha_1| \neq 1$, consider the residual

$$\sum_{t=1}^{n} (X_t - \alpha_1 X_{t-1})^2$$

representing the quadratic distance between $(X_t)_{1 \leqslant t \leqslant n}$ and $(\alpha_1 X_{t-1})_{1 \leqslant t \leqslant n}$. By Ordinary Linear Regression (OLS), the value of α_1 that minimizes this distance is given by

$$\widehat{\alpha}_1^{(n)} := \frac{\sum_{t=1}^{n} X_{t-1} X_t}{\sum_{t=1}^{n} X_{t-1}^2}, \qquad n \geqslant 1,$$

which can be rewritten as

$$\widehat{\alpha}_1^{(n)} = \frac{\sum_{t=1}^{n} X_{t-1}(Z_t + \alpha_1 X_{t-1})}{\sum_{t=1}^{n} X_{t-1}^2} = \alpha_1 + \frac{\sum_{t=1}^{n} X_{t-1} Z_t}{\sum_{t=1}^{n} X_{t-1}^2}.$$

When $|\alpha_1| \neq 1$, by (2.7), (2.9) and the Central Limit Theorem the renormalized test statistics

$$\sqrt{n}\big(\widehat{\alpha}_1^{(n)} - \alpha_1\big)$$

converges in distribution to $\mathcal{N}(0, 1 - |\alpha_1|^2)$, see Chapters 8 and 17 of Hamilton (1994), since

$$\begin{aligned}
\sqrt{n}\big(\widehat{\alpha}_1^{(n)} - \alpha_1\big) &\simeq \sqrt{n}\frac{\displaystyle\sum_{t=1}^{n} X_{t-1}Z_t}{\displaystyle\sum_{t=1}^{n} X_{t-1}^2} \\
&\simeq \frac{|1-|\alpha_1|^2|}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}Z_t \\
&\simeq \sqrt{\frac{|1-|\alpha_1|^2|}{n}} \sum_{t=1}^{n} Z_t \\
&\simeq \sqrt{\frac{|1-|\alpha_1|^2|}{n}} \mathcal{N}(0, n) \\
&\simeq \mathcal{N}(0, |1-|\alpha_1|^2|^2),
\end{aligned}$$

as n tends to infinity.

b) Under the null (nonstationarity) hypothesis H_0, *i.e.* $|\alpha_1| = 1$, the test statistic

$$\hat{t}_n := \frac{\displaystyle\sum_{t=1}^{n} X_{t-1}Z_t}{\sqrt{\displaystyle\sum_{t=1}^{n} X_{t-1}^2 \sum_{k=1}^{n} \big(X_t - \widehat{\alpha}_1^{(n)} X_{k-1}\big)^2/(n-1)}}$$

converges in distribution of as n tends to infinity to the random variable

$$\frac{\displaystyle\int_0^1 B_s dB_s}{\displaystyle\int_0^1 B_s^2 ds},$$

where $(B_s)_{s\in[0,1]}$ is a standard Brownian motion, see Equation (17.4.11) in § 17.4 of Hamilton (1989). This asymptotic distribution can be used to test the null hypothesis H_0, *i.e.* "$|\alpha_1| = 1$", against the alternative stationarity hypothesis "$|\alpha_1| \neq 1$".

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

The (Augmented) Dickey-Fuller stationarity test uses an additional lag parameter and is implemented in the 'tseries' package, as follows:

```
install.packages('tseries')
library('tseries')
adf.test(ar.sim)
adf.test(arima.sim)
```

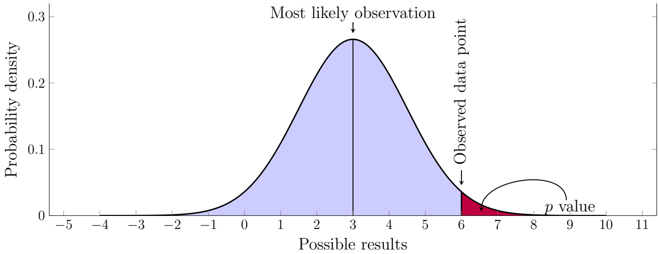


Fig. 2.9: Hypothesis testing.

The output of the adf.test command leads us to reject the nonstationarity (null) hypothesis H_0 at the level 1% for the AR(2) time series (2.10) of Figure 2.2:

```
          Augmented Dickey-Fuller Test
data: ar.sim
Dickey-Fuller = -13.765, Lag order = 16, p-value = 0.01
alternative hypothesis: stationary
Warning message:
In adf.test(ar.sim) : p-value smaller than printed p-value
```

Applying the Augmented Dickey-Fuller Test to the ARIMA time series of Figure 2.4 would not allow us to reject the nonstationarity (null) hypothesis H_0. Other stationarity tests for time series include the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which relies on linear regression.

2.4 Fitting Time Series to Financial Data

Market data can be fitted to an ARIMA(p, d, q) model using the auto.arima command in .

Fitting market returns

```
arima(data,order=c(p,d,q))
```

For an example based on market returns, we can use the following data set.

```
library(quantmod)
getSymbols("1800.HK",from="2013-01-01",to="2014-11-30",src="yahoo")
stock=Ad(`1800.HK`); stock=stock[!is.na(stock)];
stock.rtn=(stock-lag(stock))/lag(stock)[-1]; dev.new(width=12, height=6)
stock.rtn=stock.rtn[!is.na(stock.rtn)];
chartSeries(stock.rtn,up.col="blue",theme="white")
n = length(stock.rtn)
```

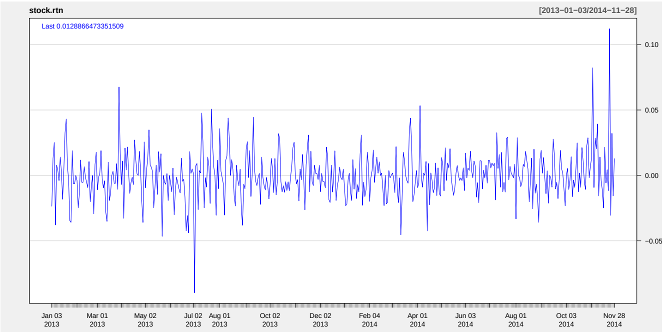


Fig. 2.10: Graph of stock returns.

```
library(forecast)
auto.arima(stock.rtn)
acf(stock.rtn,type="covariance",plot=T,col='blue',lwd=4);
```

The output of the auto.arima command identifies these data to a white noise, as ARIMA(0,0,0).

Series: stock.rtn
ARIMA(0,0,0) with zero mean
sigma^2 estimated as 0.0003266: log likelihood=1219.37
AIC=-2436.74 AICc=-2436.73 BIC=-2432.58

We can also fit these data to an MA(3) time series using the command

```
arima(stock.rtn,order=c(0,0,3))
```

which yields the output:

```
Coefficients:
       ma1    ma2     ma3
    0.0029 0.0470 -0.0416
s.e. 0.0452 0.0467 0.0465
```

and AIC=-2430.19. Sample data from this time series can be generated (up to rescaling) from

```
n=length(stock.rtn);
arima.sim<-arima.sim(model=list(ma=c(0.0029, 0.0470, -0.0416),order=c(0,0,3)),n.start=100,n)
x=seq(100,100+n-1); dates <- index(stock.rtn)
ma<-xts(x =arima.sim, order.by = dates)*sd(stock.rtn);
myPars <- chart_pars();myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"
myTheme$rylab <- FALSE; dev.new(width=16,height=8)
par(mfrow=c(1,2));chart_Series(stock.rtn,theme=myTheme,pars=myPars)
graph <-chart_Series(ma,theme=myTheme,pars=myPars); myylim <- graph$get_ylim()
myylim[[2]] <- structure(c(min(stock.rtn),max(stock.rtn)), fixed=TRUE)
graph$set_ylim(myylim); graph
```

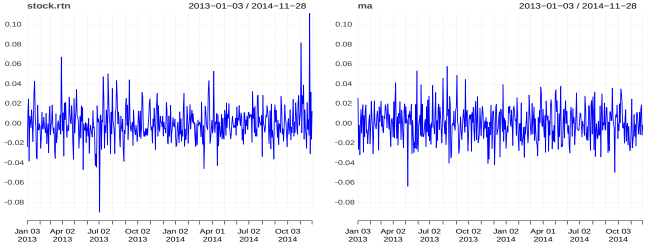


Fig. 2.11: ARIMA$(0, 0, 3)$ samples.

Next, we fit these data to an ARMA(2,2) time series using the command

```
arima(stock.rtn,order=c(2,0,2))
```

with the following output:

Coefficients:

	ar1	ar2	ma1	ma2
	-1.0593	-0.9048	1.0509	0.9508
s.e.	0.0679	0.0444	0.0474	0.0273

and AIC=-2432.57. Sample data from this time series can be generated (up to rescaling) from

```
n=length(stock.rtn);
arima.sim<-arima.sim(model=list(ar=c(-1.0593,-0.9048),ma=c(1.0509,0.9508),order=c(2,0,2)),
      n.start=100,n)
x=seq(100,100+n-1); dates <- index(stock.rtn)
ar<-xts(x =arima.sim, order.by = dates)*sd(stock.rtn);
myPars <- chart_pars();myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"
myTheme$rylab <- FALSE; dev.new(width=16,height=8); par(mfrow=c(1,2));
chart_Series(stock.rtn,theme=myTheme,pars=myPars)
graph <-chart_Series(ar,theme=myTheme,pars=myPars); myylim <- graph$get_ylim()
myylim[[2]] <- structure(c(min(stock.rtn),max(stock.rtn)), fixed=TRUE)
graph$set_ylim(myylim); graph
```

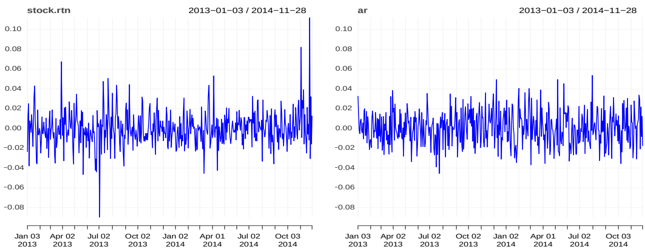


Fig. 2.12: $\mathrm{ARIMA}(2,0,2)$ samples.

These data can also be fit to a GARCH(1,1) time series using the following command.

```
library(fGarch)
garchFit(~ garch(1,1), data = stock.rtn, trace = FALSE)
```

with the following output.

Coefficients:

mu	omega	alpha1	beta1
-1.593e-04	9.124e-06	4.522e-02	9.308e-01

Sample data from this time series can be generated (up to rescaling) from

```
n=length(stock.rtn);
garch.spec<-garchSpec(model=list(alpha0 = 9.124e-06, alpha=c(4.522e-02),beta = c(9.308e-01)))
garch.sim<-garchSim(garch.spec,n);
x=seq(100,100+n-1); dates <- index(stock.rtn)
garch<-xts(x =garch.sim, order.by = dates)*sd(stock.rtn)/sd(garch.sim);
myPars <- chart_pars(); myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"
myTheme$rylab <- FALSE; dev.new(width=16,height=8)
par(mfrow=c(1,2));chart_Series(stock.rtn,theme=myTheme,pars=myPars)
graph <-chart_Series(garch,theme=myTheme,pars=myPars); myylim <- graph$get_ylim()
myylim[[2]] <- structure(c(min(stock.rtn),max(stock.rtn)), fixed=TRUE)
graph$set_ylim(myylim); graph
```

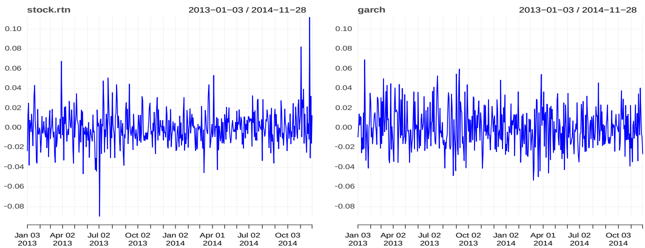


Fig. 2.13: GARCH samples.

Fitting market prices

Next, we fit market price data to an ARIMA time series.

```
library(quantmod)
getSymbols("1800.HK",from="2007-01-03",to="2011-12-02",src="yahoo")
stock=Ad(`1800.HK`)
chartSeries(stock,up.col="blue",theme="white")
n = length(stock)
arima(stock,order=c(2,1,2))
```

Fig. 2.14: Cumulative stock returns.

The output of the auto.arima(stock) command identifies these data to an ARIMA(2,1,0) time series of integrated order one.

Series: stock
ARIMA(2,1,0)
Coefficients: ar1 ar2 0.0605 -0.0006 s.e. 0.0288 0.0288
sigma^2 = 0.05082: log likelihood = 84.47
AIC=-162.94 AICc=-162.92 BIC=-147.63

We may also fit these data to an ARIMA(2,1,2) time series using the command arima(stock,order=c(2,1,2)).

Coefficients:

	ar1	ar2	ma1	ma2
	-0.3073	-0.9626	0.3452	0.9783
s.e.	0.0137	0.0178	0.0092	0.0155

sigma^2 estimated as 0.04987: log likelihood = 94.49, aic = -178.98

```
n=length(stock)-1
arima.sim<-arima.sim(model=list(ar=c(-0.3133,-0.9464),ma=c(0.3535,0.9637),order=c(2,1,2)),
        n.start=100,n)
x=seq(100,100+n)
dates <- index(stock); ar<-xts(x =arima.sim, order.by = dates)
myPars <- chart_pars();myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"
myTheme$rylab <- FALSE; dev.new(width=16,height=8)
par(mfrow=c(1,2));chart_Series(stock,theme=myTheme,pars=myPars)
chart_Series(as.vector(stock[1])+ar,theme=myTheme,pars=myPars)
```

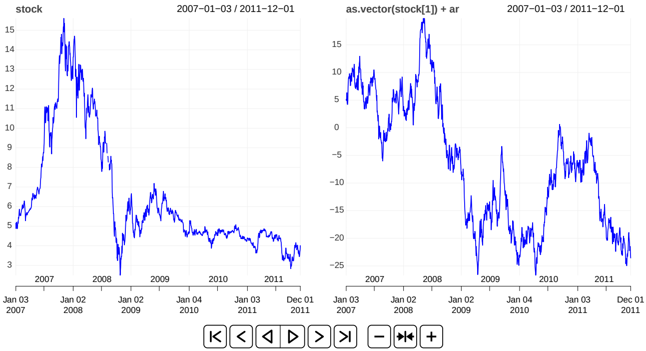


Fig. 2.15: ARIMA$(2, 1, 2)$ Samples.*

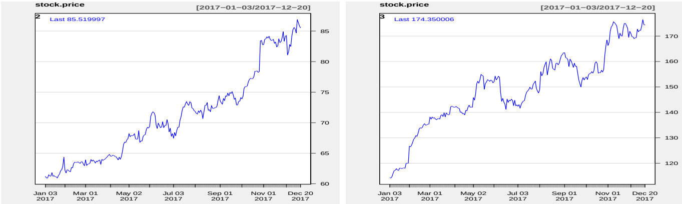
2.5 Application: Pair Trading

Pair trading data

We consider two assets that can be traded in pairs.

```
install.packages("quantmod");library(quantmod)
symbols = c("1800.HK","KO","PEP");symbols = c("1800.HK","MSFT","AAPL")
getSymbols(symbols, from=Sys.Date()-365, to=Sys.Date());ClosePrices <- lapply(symbols,
      function(x) Ad(get(x)))
getSymbols(symbols, from="2017-01-01", to="2018-01-01");ClosePrices <- lapply(symbols,
      function(x) Ad(get(x)))
stock<-do.call(merge, ClosePrices);stock.price<-stock[rowSums(is.na(stock[ , 1:3])) == 0, ];
chartSeries(stock.price[,2],up.col="purple",theme="white",name = symbols[2])
chartSeries(stock.price[,3],up.col="blue",theme="white",name = symbols[3])
myPars <- chart_pars();myPars$cex<-1.4;myTheme <- chart_theme();
      myTheme$col$line.col<-"blue"
dev.new(width=16,height=8);par(mfrow=c(1,3))
chart_Series(stock.price[,2],up.col="purple",theme=myTheme,name = symbols[2],pars=myPars)
chart_Series(stock.price[,3],up.col="blue",theme=myTheme,name = symbols[3],pars=myPars)
add_TA(stock.price[,2], col='purple', lw =2, on = 1)
```

* The animation works in Acrobat Reader on the entire pdf file.



(a) MSFT graph. (b) AAPL graph.

Fig. 2.16: MSFT *vs.* AAPL graphs.

Linear regression

As the two assets may evolve within different price ranges, we use a linear regression to put them both on the scale of the second asset.

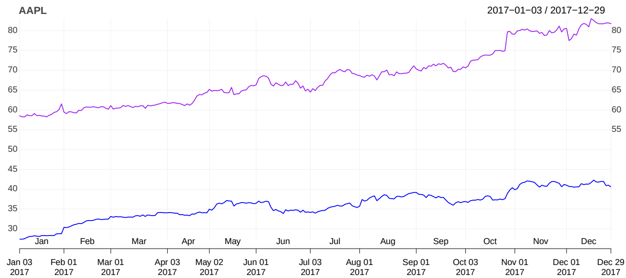


Fig. 2.17: Comparison graph before linear regression.

Letting

$$r_t^{(1)} := \log(S_t^{(1)}) \quad \text{and} \quad r_t^{(2)} := \log(S_t^{(2)}), \qquad t \geqslant 1,$$

by an Ordinary Least Square (OLS) regression using the R command *lm* (linear model), we derive a linear relationship of the form

$$r_t^{(2)} = a + b r_t^{(1)} + X_t, \quad t \geqslant 1, \tag{2.22}$$

between $\big(r_k^{(1)}\big)_{k\geqslant 1}$ and $\big(r_k^{(2)}\big)_{k\geqslant 1}$, where X_k is a random remainder term, by minimization of the quadratic residual distance

$$\sum_{t=1}^{n}\left(r_t^{(2)} - a - b r_t^{(1)}\right)^2 \tag{2.23}$$

between $\left(r_t^{(2)}\right)_{t=1,2,\ldots,n}$ and $\left(a + b r_t^{(1)}\right)_{t=1,2,\ldots,n}$, *i.e.*

$$\begin{cases} \widehat{a} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\left(r_k^{(2)} - \widehat{b} r_k^{(1)}\right), \\ \\ \text{and} \\ \\ \widehat{b} = \dfrac{\displaystyle\sum_{k=1}^{n} r_k^{(1)} r_k^{(2)} - \frac{1}{n}\sum_{k,l=1}^{n} r_k^{(1)} \tilde{r}_l^{(2)}}{\displaystyle\sum_{k=1}^{n}(r_k^{(1)})^2 - \frac{1}{n}\sum_{k,l=1}^{n} r_k^{(1)} r_l^{(1)}} = \dfrac{\displaystyle\sum_{k=1}^{n}\left(r_k^{(1)} - \frac{1}{n}\sum_{l=1}^{n} r_l^{(1)}\right)\left(r_k^{(2)} - \frac{1}{n}\sum_{l=1}^{n}\tilde{r}_l^{(2)}\right)}{\displaystyle\sum_{k=1}^{n}\left(r_k^{(1)} - \frac{1}{n}\sum_{k=1}^{n} r_k^{(1)}\right)^2}. \end{cases}$$

see Exercise 2.6. The coefficient a in (2.22) is called the *premium*.

```r
price.pair <- stock.price[,2:3]["2017-02-01::"]
reg <- lm(log(price.pair[,2]) ~ log(price.pair[,1]))
hedge.ratio <- as.numeric(reg$coef[2]);premium <- as.numeric(reg$coef[1])
myPars <- chart_pars();myPars$cex<-1.4
myTheme <- chart_theme();myTheme$col$line.col <- "blue"; dev.new(width=16,height=8)
chart_Series(price.pair[,2],theme=myTheme,pars=myPars,name = symbols[3])
add_TA(exp(premium+hedge.ratio*log(price.pair[,1])), col='purple', lw =2,on = 1)
```

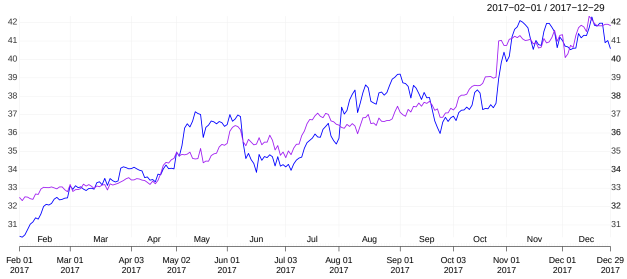


Fig. 2.18: Comparison graph after linear regression.

This allows us to define the *spread* $(X_t)_{t\geqslant 1}$ via the linear relationship

$$X_t := \log(S_t^{(2)}) - (a + b\log(S_t^{(1)})), \qquad t \geqslant 0.$$

```
spread <- log(price.pair[,2]) - ( hedge.ratio * log(price.pair[,1]) + premium )
list(premium = premium, hedge.ratio = hedge.ratio)
dev.new(width=16,height=7)
plot(spread,col='blue', main = "Spread",cex.axis=1.3)
```

Fig. 2.19: Spread graph.

Next, we model the spread X_t using an AR(1) time series and check for its stationarity, in which case the log-price processes $\log\big(S_t^{(1)}\big)_{t\geqslant 0}$ and $\log\big(S_t^{(2)}\big)_{t\geqslant 0}$ are said to be *cointegrated*.

This will be interpreted as the existence of a statistically significant connection between $\big(S_t^{(1)}\big)_{t\geqslant 0}$ and $\big(S_t^{(2)}\big)_{t\geqslant 0}$, also named *cointegration*. See Engle and Granger (1987) and Chapter 6 of Enders (2009) for more information on *cointegration*.

Dickey-Fuller test

Consider an AR(1) time series $(X_n)_{n\geqslant 0}$ given by

$$X_n := Z_n + \alpha_1 X_{n-1}.$$

The Dickey-Fuller test allows us to test the null hypothesis H_0, *i.e.* "$|\alpha_1| = 1$", against the alternative stationarity hypothesis "$|\alpha_1| \neq 1$".

```
install.packages('tseries'); library('tseries')
adf.test(spread)
```

Its output would lead us to reject the nonstationarity (null) hypothesis H_0 at the confidence level 5% when the p-value is lower than 0.05.

Augmented Dickey-Fuller Test
data: spread
Dickey-Fuller = -2.8771, Lag order = 6, p-value = 0.2077
alternative hypothesis: stationary

Pair trading

The trading signal is $\{-1, 1\}$-valued and determined by the alternating crossing times of a threshold level by the spread.

```
signal<-spread;threshold <- 0.02
signal[1] = sign(as.numeric(spread[1]));i=1;threshold=-as.numeric(signal[1])*threshold
while (i<length(spread)){i=i+1;
while (i<length(spread) &&
        sign(as.numeric(spread[i+1])-threshold)==sign(as.numeric(spread[i])-threshold))
        {signal[i]=sign(as.numeric(spread[i-1])-threshold);i=i+1;}
signal[i]=sign(as.numeric(spread[i-1])-threshold);threshold=-threshold;}
signal[i]=sign(as.numeric(spread[i-1])-threshold);threshold <- abs(threshold)
ratio1=range(spread)[1]/threshold;ratio2=range(spread)[2]/threshold
tblue <- rgb(0, 0, 1, alpha=0.8);tred <- rgb(1, 0, 0, alpha=0.5)
dev.new(width=16,height=7)
barplot(spread,col = tblue,lwd = 3, main = "",cex.axis=1.4,cex=1.6,las=1);par(new=TRUE);
barplot(-signal,offset=(range(spread)[1]+range(spread)[2])/threshold,ylim=c(ratio2,ratio1), xpd
        = FALSE, col=tred,space = 0, border ="blue",xaxt="n",yaxt="n",xlab="",ylab="")
```

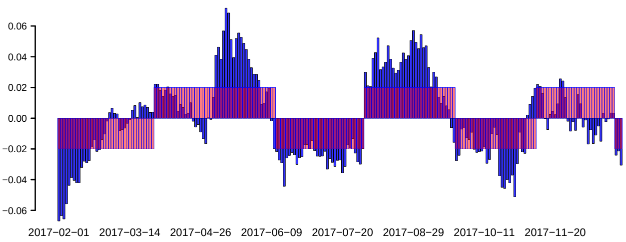


Fig. 2.20: Pair trading signals.

We construct a discrete-time self-financing portfolio strategy $\big(\xi_t^{(1)},\xi_t^{(2)}\big)_{t\geqslant 1}$ where $\xi_t^{(k)}$ denotes the (possibly fractional) quantity of asset n^o k held in the portfolio over the time interval $(t-1,t]$, $t=1,2$. Note that the portfolio allocation

$$\bar{\xi}_t = \big(\xi_t^{(1)},\xi_t^{(2)}\big)$$

is decided at time $t-1$ and remains constant over the interval $(t-1,t]$ while the stock price changes from $S_{t-1}^{(k)}$ to $S_t^{(k)}$ over this time interval. In other words, the quantity

$$\xi_t^{(k)}S_{t-1}^{(k)}$$

represents the amount invested in asset n^o k at the beginning of the time interval $(t-1,t]$, and

$$\xi_t^{(k)}S_t^{(k)}$$

represents the value of this investment at the end of the time interval $(t-1,t]$, $t=1,2,\ldots,N$.

Self-financing portfolio strategies

The opening price of the portfolio at the beginning of the trading period $(t-1,t]$ is

$$\xi_t^{(1)}S_{t-1}^{(1)}+\xi_t^{(2)}S_{t-1}^{(2)}.$$

At the end of the time interval $(t-1,t]$, it takes the closing value

$$\xi_t^{(1)}S_t^{(1)}+\xi_t^{(2)}S_t^{(2)}, \tag{2.24}$$

$t=1,2,\ldots,N$. After the new portfolio allocation $\big(\xi_{t+1}^{(1)},\xi_{t+1}^{(2)}\big)$ is designed we get the new portfolio opening price

$$\xi_{t+1}^{(1)}S_t^{(1)}+\xi_{t+1}^{(2)}S_t^{(2)}, \tag{2.25}$$

at the beginning of the next trading session $(t,t+1]$, $t=0,1,\ldots,N-1$.

Definition 2.13. *The portfolio prices at times $t=0,1,\ldots,N-1$ are given by*

$$V_t := \xi_{t+1}^{(1)}S_t^{(1)}+\xi_{t+1}^{(2)}S_t^{(2)}, \quad t=0,1,\ldots,N-1.$$

We say that the portfolio strategy $\big(\xi_t^{(1)},\xi_t^{(2)}\big)_{t=1,2,\ldots,N}$ is *self-financing* when (2.24) coincides with (2.25) for $t=0,1,\ldots,N-1$.

Definition 2.14. *A portfolio strategy $\big(\xi_t^{(1)},\xi_t^{(2)}\big)_{t=1,2,\ldots,N}$ is said to be* self-financing *if*

$$\underbrace{\xi_t^{(1)} S_t^{(1)} + \xi_t^{(2)} S_t^{(2)}}_{\text{Closing value}} = \underbrace{\xi_{t+1}^{(1)} S_t^{(2)} + \xi_{t+1}^{(2)} S_t^{(2)}}_{\text{Opening price}}, \qquad t = 1, 2, \ldots, N-1. \quad (2.26)$$

The next figure is an illustration of the self-financing condition.

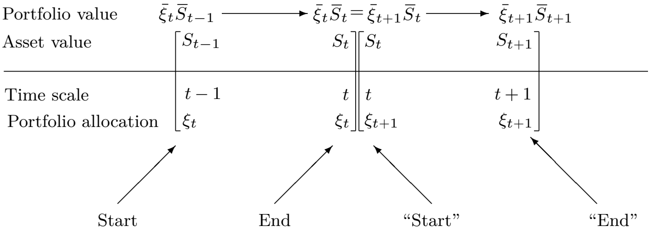


Fig. 2.21: Illustration of the self-financing condition (2.26).

By (2.24) and (2.25) the self-financing condition (2.26) can be rewritten as

$$\xi_t^{(1)} S_t^{(1)} + \xi_t^{(2)} S_t^{(2)} = \xi_{t+1}^{(1)} S_t^{(1)} + \xi_{t+1}^{(2)} S_t^{(2)}, \qquad t = 0, 1, \ldots, N-1,$$

or

$$\left(\xi_{t+1}^{(1)} - \xi_t^{(1)}\right) S_t^{(1)} + \left(\xi_{t+1}^{(2)} - \xi_t^{(2)}\right) S_t^{(2)} = 0, \qquad t = 0, 1, \ldots, N-1.$$

Note that any portfolio strategy $\left(\xi_t^{(1)}, \xi_t^{(2)}\right)_{t=1,2,\ldots,N}$ which is constant over time is self-financing by construction.

Under the self-financing condition (2.26), the portfolio closing values V_t at times $t = 1, 2, \ldots, N$ rewrite as

$$V_t = \bar{\xi}_t \cdot \overline{S}_t = \xi_t^{(1)} S_t^{(1)} + \xi_t^{(2)} S_t^{(2)}, \quad t = 1, 2, \ldots, N. \qquad (2.27)$$

Letting

$$\text{hedge.ratio}_t := \frac{S_t^{(2)}}{S_t^{(1)}}, \qquad t = 0, 1, \ldots, N-1,$$

we define the portfolio strategy as

$$\begin{cases} \xi_1^{(1)} := \text{signal}_1 \times \dfrac{\text{hedge.ratio}_0}{1 + \text{hedge.ratio}_0} \\[2ex] \xi_1^{(2)} := (-\text{signal}_1) \times \dfrac{1}{1 + \text{hedge.ratio}_0}, \end{cases}$$

and subsequently

$$\begin{cases} \xi_{t+1}^{(1)} := \xi_t^{(1)} + \dfrac{\text{signal}_{t+1} - \text{signal}_t}{2} \times \dfrac{\text{hedge.ratio}_t}{1 + \text{hedge.ratio}_t} \\ \\ \xi_{t+1}^{(2)} := \xi_t^{(2)} - \dfrac{\text{signal}_{t+1} - \text{signal}_t}{2} \times \dfrac{1}{1 + \text{hedge.ratio}_t}, \end{cases}$$

$t = 1, 2, \ldots, N = 1$, so that

$$\xi_1^{(1)} S_0^{(1)} + \xi_1^{(2)} S_0^{(2)} = 0,$$

and

$$\begin{aligned} & \big(\xi_{t+1}^{(1)} - \xi_t^{(1)}\big) S_t^{(1)} + \big(\xi_{t+1}^{(2)} - \xi_t^{(2)}\big) S_t^{(2)} \\ & = \frac{\text{signal}_{t+1} - \text{signal}_t}{2} \times \frac{\text{hedge.ratio}_t}{1 + \text{hedge.ratio}_t} S_t^{(1)} \\ & \quad - \frac{\text{signal}_{t+1} - \text{signal}_t}{2} \times \frac{1}{1 + \text{hedge.ratio}_t} S_t^{(2)} \\ & = 0, \quad t = 1, 2, \ldots, N, \end{aligned}$$

i.e. $\big(\xi_t^{(1)}, \xi_t^{(2)}\big)_{t=1,2,\ldots,N}$ is self-financing.

Backtesting

The performance of the pair trading algorithm can be estimated by the following code.

```
hedge.ratio=price.pair[,1]/price.pair[,2]; diff.xi1=diff(signal)[-1]/(1+hedge.ratio)/2
diff.xi2=-diff(signal)[-1]*hedge.ratio/(1+hedge.ratio)/2;
xi1=cumsum(c(signal[1]/(1+hedge.ratio[1]),diff.xi1))
xi2=cumsum(c(-signal[1]*hedge.ratio[1]/(1+hedge.ratio[1]),diff.xi2))
dev.new(width=16,height=7); portfolio=xi1*price.pair[,1]+xi2*price.pair[,2]
benchmark=as.numeric(xi1[1])*price.pair[,1]+as.numeric(xi2[1])*price.pair[,2]
plot(benchmark,col='orange',main = "",lwd = 3, cex.axis=1,cex=1,las=1)
lines(xi1,col='purple',main = "Xi1",lwd = 3, cex.axis=1,cex=2,las=1)
lines(xi2,col='blue',main = "Xi2",lwd = 3, cex.axis=1,cex=2,las=1)
lines(portfolio,col='red',main = "Portfolio performance" ,lwd = 4, cex.axis=1,cex=2,las=1)
legend("topleft", legend=c("Pair trading", "Benchmark", "Xi1", "Xi2"), col=c("red", "orange",
      "purple", "blue"), lty=1:2, cex=0.8,xpd=TRUE)
```

The performance of the pair trading portfolio return is plotted in Figure 2.22.

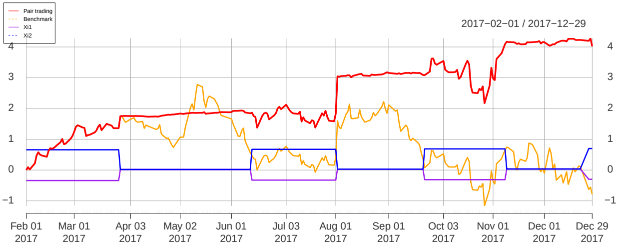


Fig. 2.22: Pair trading performance.

The portfolio strategy used as benchmark is built by investing a constant allocation in Stock 1 and Stock 2, and its return from time 0 to time n, given by

$$B_t := \xi_1^{(1)} S_t^{(1)} + \xi_2^{(2)} S_t^{(1)}, \qquad t = 1, \ldots, N,$$

is compared to the pair trading portfolio performance. The ADF test allows us to reject the non-stationarity hypothesis at the level $p = 8.5\%$.

```
> source("pairtrading.R")
Examples of pairs: 005930.KS vs AAPL, 2600.HK vs 1919.HK
Enter Stock 1 (Ex: GOOG):1800.HK
Enter Stock 2 (Ex: AAPL):1919.HK

          Augmented Dickey-Fuller Test
data: spread
Dickey-Fuller = -3.2232, Lag order = 6, p-value = 0.08465
alternative hypothesis: stationary
```

Figure 2.23 presents another example of pair trading backtesting.*

* Download the corresponding code.

Fig. 2.23: Pair trading performance.

See also the PairTrading package in Takayanagi and Ishikawa (2017).

Exercises

Exercise 2.1 Consider the MA(1) time series $(X_n)_{n\in\mathbb{Z}}$ defined as

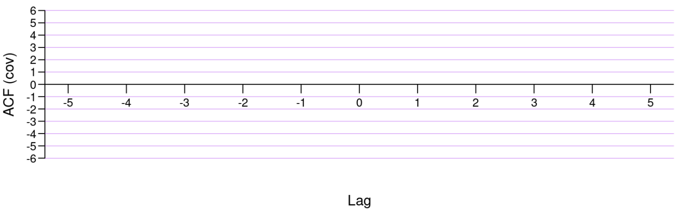
$$X_n := Z_n + aZ_{n-1}, \quad n \in \mathbb{Z},$$

where $(Z_n)_{ninZ}$ is a white noise sequence and $a \in \mathbb{R}$.

a) Compute the autocovariance function

$$\rho(k) = \mathrm{Cov}(X_n, X_{n+k})$$

for all $k \in \mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, and plot it on the graph below when $a = 2$.



b) Is the time series $(X_n)_{n\in\mathbb{Z}}$ weakly stationary? Strictly stationary?

Exercise 2.2 Let $(Z_n)_{n\geqslant 1}$ denote a discrete-time white noise.

a) Check the (weak) stationarity of the AR(1) time series $(X_n)_{n\geqslant 1}$ given by

$$X_n = Z_n + X_{n-1}, \qquad n \geqslant 1.$$

b) Check the (weak) stationarity of the AR(2) time series $(Y_n)_{n\geqslant 1}$ given by

$$Y_n = Z_n + \frac{3}{4} \times Y_{n-1} - \frac{1}{8} \times Y_{n-2}, \qquad n \geqslant 2.$$

Hint: Consider the roots of $\varphi(z) = 1$ where $\varphi(z)$ is the polynomial defined by $X_n = Z_n + \varphi(L)X_n$ and L is the lag operator $LX_n = X_{n-1}$.

Exercise 2.3 Let $\alpha \in \mathbb{R}$. Consider an *i.i.d.* white noise sequence $(Z_n)_{n\geqslant 0}$ with mean $\mathbb{E}[Z_n] = 0$ and variance $\mathrm{Var}[Z_n] = 1$, $n \geqslant 1$, and the AR(1) time series $(X_n)_{n\geqslant 0}$ given by $X_0 := 0$ and

$$X_n := Z_n + \alpha X_{n-1}, \qquad n \geqslant 1. \tag{2.28}$$

a) Find a recurrence relation for the mean $\mathbb{E}[X_n]$ in the parameter $n \geqslant 0$, and deduce the value of $\mathbb{E}[X_n]$ for all $n \geqslant 0$.
b) For fixed $n \geqslant 1$, find a recurrence relation in the parameter $k \geqslant 0$ for the covariance

$$\mathrm{Cov}(X_{n+k}, X_n) = \mathbb{E}[X_{n+k}X_n] - \mathbb{E}[X_{n+k}]\mathbb{E}[X_n], \qquad k \geqslant 0.$$

c) Find a recurrence relation in the parameter $n \geqslant 1$ for the variance

$$\mathrm{Var}[X_n] = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2, \qquad n \geqslant 0.$$

d) When is the time series $(X_n)_{n\geqslant 0}$ weakly stationary?

Exercise 2.4 Consider an *i.i.d.* white noise sequence $(Z_n)_{n\geqslant 0}$ with mean $\mathbb{E}[Z_n] = 0$ and variance $\mathrm{Var}[Z_n] = 1$, $n \geqslant 1$, and the AR(3) time series $(X_n)_{n\geqslant 3}$ given by

$$X_n := Z_{n-1} - Z_{n-2} + \alpha Z_{n-3}, \qquad n \geqslant 3.$$

a) Find the autocovariances

$$\begin{cases} \mathrm{Cov}(X_n, X_n) = \mathrm{Var}[X_n], \\[2ex] \mathrm{Cov}(X_{n+1}, X_n), \\[2ex] \mathrm{Cov}(X_{n+2}, X_n), \\[2ex] \mathrm{Cov}(X_{n+k}, X_n), \qquad k \geqslant 3. \end{cases}$$

b) Show that $(X_n)_{n\geqslant 3}$ has same distribution as an MA(q) time series $(Y_n)_{n\geqslant 3}$ of the form

$$Y_n = Z_n + \sum_{k=1}^{q} \beta_k Z_{n-k},$$

whose order q and coefficients $(\beta_k)_{1\leqslant k\leqslant q}$ will be determined.

Exercise 2.5 Consider an AR(1) time series $(X_n)_{n\geqslant 0}$ given by $X_0 = 0$ and

$$X_n := Z_n + \alpha_1 X_{n-1}, \qquad n \geqslant 1,$$

and the difference operator

$$\nabla X_n := X_n - X_{n-1}, \qquad n \geqslant 1,$$

also written $\nabla = I - L$, which can be integrated by the telescoping identity

$$X_n = X_0 + \sum_{k=1}^{n} (X_k - X_{k-1}) = \sum_{k=1}^{n} \nabla X_k, \quad n \geqslant 0.$$

a) Show that the first-order difference process $(\nabla X_n)_{n\geqslant 1} = (X_n - X_{n-1})_{n\geqslant 1}$. forms an ARMA$(1,2)$ time series.
b) Show that the second-order difference process

$$(\nabla^2 X_n)_{n\geqslant 2} = (\nabla X_n - \nabla X_{n-1})_{n\geqslant 2}$$

forms an ARMA$(1,3)$ time series.

Exercise 2.6 Consider two sequences $\big(r_k^{(1)}\big)_{k\geqslant 1}$ and $\big(r_k^{(2)}\big)_{k\geqslant 1}$ of market returns. We aim at deriving a linear relationship of the form

$$r_k^{(2)} = a + b r_k^{(1)} + X_k, \quad k \geqslant 0,$$

between $\big(r_k^{(1)}\big)_{k\geqslant 1}$ and $\big(r_k^{(2)}\big)_{k\geqslant 1}$, where X_k is a random remainder term, by minimization of the quadratic residual distance

$$\sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big)^2 \tag{2.29}$$

between $\big(r_k^{(2)}\big)_{k=1,2,\ldots,n}$ and $\big(a + b r_k^{(1)}\big)_{k=1,2,\ldots,n}$.

a) Compute the partial derivatives of (2.29) with respect to the parameters a and b.

b) By equating the derivatives to zero, find the least square estimates $\widehat{a}$ and $\widehat{b}$ of the parameters a and b based on the sequences $\big(r_k^{(1)}\big)_{k\geqslant 1}$ and $\big(r_k^{(2)}\big)_{k\geqslant 1}$.

Exercise 2.7 Consider the following ADF test output on a time series:

```
        Augmented Dickey-Fuller Test
data: series
Dickey-Fuller = -2.8771, Lag order = 6, p-value = 0.02377
alternative hypothesis: stationary
```

Does this test result allow us to reject the nonstationarity (null) hypothesis H_0 at a 5% confidence level?

Exercise 2.8 Let $(Z_n)_{n\in\mathbb{Z}}$ denote a white noise sequence with zero mean and variance σ^2, and consider the AR(2) time series $(X_n)_{n\in\mathbb{Z}}$ given by

$$X_n = Z_n + \alpha_1 X_{n-1} + \alpha_2 X_{n-2}, \qquad n \in \mathbb{Z}.$$

a) Assume that $\alpha_1 := a$ and $\alpha_2 := 2a^2$ for some $a \geqslant 0$. For which values of the parameter a is the time series $(X_n)_{n\in\mathbb{Z}}$ stationary?
   Hint: Consider the solutions of the equation $\varphi(z) = 1$, where $\varphi(z)$ is the polynomial defined by $X_n = Z_n + \varphi(L)X_n$ and L is the lag operator defined by $LX_n = X_{n-1}$.
   In the sequel, we assume that the time series $(X_n)_{n\in\mathbb{Z}}$ is stationary.
b) Show that $\mathbb{E}[X_n] = 0$, $n \in \mathbb{Z}$, if $\alpha_1 + \alpha_2 \neq 1$.
c) We assume that $(X_n)_{n\in\mathbb{Z}}$ is *causal*, *i.e.* for any $n \in \mathbb{Z}$, X_n depends only on $(Z_k)_{k<n}$. Show that $\mathrm{Cov}(X_n, Z_n) = \sigma^2$ for all $n \in \mathbb{Z}$.
d) Taking $\alpha_1 := 1/4$ and $\alpha_2 := 1/2$, compute the autocovariance $\mathrm{Cov}(X_{n+1}, X_n)$ given that $\mathrm{Cov}(X_n, X_n) = 16$, $n \in \mathbb{Z}$.

Chapter 3
Processes with Jumps

Modeling insurance risk requires to use continuous-time stochastic processes that allow for jumps in addition to a continuous component. This chapter presents the construction of Poisson and compound Poisson processes which are used for the modeling of insurance claim and reserve processes. Applications will be given to the closed form computation of ruin probabilities in the Cramér-Lundberg model.

3.1 The Poisson Process


The most elementary and useful jump process is the *standard Poisson process* $(N_t)_{t\in\mathbb{R}_+}$ which is a *counting process*, *i.e.* $(N_t)_{t\in\mathbb{R}_+}$ has jumps of size $+1$ only and its paths are constant in between two jumps, with $N_0 := 0$.

The counting process $(N_t)_{t\in\mathbb{R}_+}$ that can be used to model discrete arrival times such as claim dates in insurance, or connection logs.

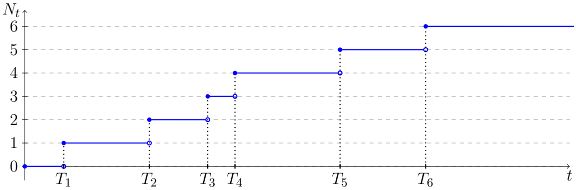


Fig. 3.1: Sample path of a counting process $(N_t)_{t\in\mathbb{R}_+}$.

Using the indicator functions

$$\mathbb{1}_{[T_k,\infty)}(t) = \begin{cases} 1 \text{ if } t \geqslant T_k, \\ 0 \text{ if } 0 \leqslant t < T_k, \quad k \geqslant 1, \end{cases}$$

the value of N_t at time t can be written as

$$N_t = \sum_{k\geqslant 1} \mathbb{1}_{[T_k,\infty)}(t), \qquad t \geqslant 0, \tag{3.1}$$

where and $(T_k)_{k\geqslant 1}$ is the increasing family of jump times of $(N_t)_{t\in\mathbb{R}_+}$ such that

$$\lim_{k\to\infty} T_k = +\infty.$$

The operation defined in (3.1) can be implemented in R using the following code.

```
T=10; Tn=c(1,3,4,7,9); dev.new(width=T, height=5)
plot(stepfun(Tn,c(0,1,2,3,4,5)),xlim =c(0,T),xlab="t",ylab=expression('N'[t]),pch=1, cex=0.8,
      col='blue', lwd=2, main="", cex.axis=1.2, cex.lab=1.4,xaxs='i'); grid()
```

In order for the counting process $(N_t)_{t\in\mathbb{R}_+}$ to be a Poisson process, it has to satisfy the following conditions:

1. Independence of increments: for all $0 \leqslant t_0 < t_1 < \cdots < t_n$ and $n \geqslant 1$ the increments
$$N_{t_1} - N_{t_0}, \ldots, \; N_{t_n} - N_{t_{n-1}},$$
are mutually independent random variables.

2. Stationarity of increments: $N_{t+h} - N_{s+h}$ has the same distribution as $N_t - N_s$ for all $h > 0$ and $0 \leqslant s \leqslant t$.

The meaning of the above stationarity condition is that for all fixed $k \geqslant 0$ we have

$$\mathbb{P}(N_{t+h} - N_{s+h} = k) = \mathbb{P}(N_t - N_s = k),$$

for all $h > 0$, *i.e.*, the value of the probability

$$\mathbb{P}(N_{t+h} - N_{s+h} = k)$$

does not depend on $h > 0$, for all fixed $0 \leqslant s \leqslant t$ and $k \geqslant 0$.

Based on the above assumption, given $T > 0$ a time value, a natural question arises:

what is the probability distribution of the random variable N_T?

We already know that N_t takes values in $\mathbb{N}$ and therefore it has a discrete distribution for all $t \in \mathbb{R}_+$.

It is a remarkable fact that the distribution of the increments of $(N_t)_{t\in\mathbb{R}_+}$, can be completely determined from the above conditions, as shown in the following theorem.

As seen in the next result, cf. Theorem 4.1 in Bosq and Nguyen (1996), the Poisson increment $N_t - N_s$ has the Poisson distribution with parameter $(t-s)\lambda$.

Theorem 3.1. *Assume that the counting process $(N_t)_{t\in\mathbb{R}_+}$ satisfies the above independence and stationarity Conditions 1 and 2 on page 76. Then, for all fixed $0 \leqslant s \leqslant t$ the increment $N_t - N_s$ follows the Poisson distribution with parameter $(t-s)\lambda$, i.e. we have*

$$\boxed{\mathbb{P}(N_t - N_s = k) = \mathrm{e}^{-(t-s)\lambda}\frac{((t-s)\lambda)^k}{k!}, \qquad k \geqslant 0,} \tag{3.2}$$

for some constant $\lambda > 0$.

The parameter $\lambda > 0$ is called the <u>intensity</u> of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$ and it is given by

$$\lambda := \lim_{h\to 0} \frac{1}{h}\mathbb{P}(N_h = 1). \tag{3.3}$$

The proof of the above Theorem 3.1 is technical and not included here, cf. *e.g.* Bosq and Nguyen (1996) for details, and we could in fact take this distribution property (3.2) as one of the hypotheses that define the Poisson process.

Precisely, we could restate the definition of the standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$ with intensity $\lambda > 0$ as being a stochastic process defined by (3.1), which is assumed to have independent increments distributed according to the Poisson distribution, in the sense that for all $0 \leqslant t_0 \leqslant t_1 < \cdots < t_n$,

$$(N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}})$$

is a vector of independent Poisson random variables with respective parameters

$$((t_1 - t_0)\lambda, \ldots, (t_n - t_{n-1})\lambda).$$

In particular, N_t has the Poisson distribution with parameter λt, *i.e.*,

$$\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad t > 0.$$

The *expected value* $\mathbb{E}[N_t]$ and the variance of N_t can be computed as

$$\mathbb{E}[N_t] = \mathrm{Var}[N_t] = \lambda t, \tag{3.4}$$

see Exercise A.1. As a consequence, the *dispersion index* of the Poisson process is

$$\frac{\mathrm{Var}[N_t]}{\mathbb{E}[N_t]} = 1, \qquad t \geqslant 0. \tag{3.5}$$

Short time behaviour

From (3.3) above we deduce the *short time asymptotics**

$$\begin{cases} \mathbb{P}(N_h = 0) = e^{-\lambda h} = 1 - \lambda h + o(h), & h \to 0, \\ \mathbb{P}(N_h = 1) = \lambda h e^{-\lambda h} \simeq \lambda h, & h \to 0. \end{cases}$$

By stationarity of the Poisson process we also find more generally that

$$\begin{cases} \mathbb{P}(N_{t+h} - N_t = 0) = e^{-\lambda h} = 1 - \lambda h + o(h), & h \to 0, \\ \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h e^{-\lambda h} \simeq \lambda h, & h \to 0, \\ \mathbb{P}(N_{t+h} - N_t = 2) \simeq h^2 \dfrac{\lambda^2}{2} = o(h), & h \to 0, \qquad t > 0, \end{cases} \tag{3.6}$$

for all $t > 0$. This means that within a "short" time interval $[t, t+h]$ of length h, the increment $N_{t+h} - N_t$ behaves like a Bernoulli random variable with parameter λh. This fact can be used for the random simulation of Poisson process paths.

* The notation $f(h) = o(h^k)$ means $\lim_{h\to 0} f(h)/h^k = 0$, and $f(h) \simeq h^k$ means $\lim_{h\to 0} f(h)/h^k = 1$.

The next R code and Figure 3.2 present a simulation of the standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$ according to its short time behavior (3.6).

```
lambda = 0.6;T=10;N=1000*lambda;h=T*1.0/N
t=0;s=c();for (k in 1:N) {if (runif(1)<lambda*h) {s=c(s,t)};t=t+h}
dev.new(width=T, height=5)
plot(stepfun(s,cumsum(c(0,rep(1,length(s))))),xlim
      =c(0,T),xlab="t",ylab=expression('N'[t]),pch=1, cex=0.8, col='blue', lwd=2, main="",
      cex.axis=1.2, cex.lab=1.4,xaxs='i'); grid()
```

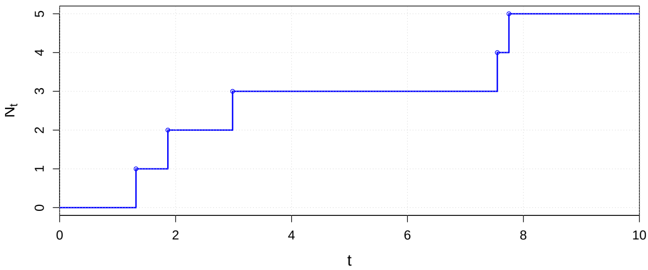


Fig. 3.2: Sample path of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$.

More generally, for $k \geqslant 1$ we have

$$\mathbb{P}(N_{t+h} - N_t = k) \simeq h^k \frac{\lambda^k}{k!}, \qquad h \to 0, \qquad t > 0.$$

Time-dependent intensity

The intensity of the Poisson process can in fact be made time-dependent (*e.g.* by a time change), in which case we have

$$\mathbb{P}(N_t - N_s = k) = \exp\left(-\int_s^t \lambda(u)du\right) \frac{\left(\int_s^t \lambda(u)du\right)^k}{k!}, \quad k = 0,1,2,\ldots.$$

Assuming that $\lambda(t)$ is a continuous function of time t we have in particular, as h tends to zero,

$$\mathbb{P}(N_{t+h} - N_t = k)$$

$$= \begin{cases} \exp\left(-\int_t^{t+h}\lambda(u)du\right) = 1-\lambda(t)h+o(h), & k=0, \\ \exp\left(-\int_t^{t+h}\lambda(u)du\right)\int_t^{t+h}\lambda(u)du = \lambda(t)h+o(h), & k=1, \\ o(h), & k\geqslant 2. \end{cases}$$

The intensity process $(\lambda(t))_{t\in\mathbb{R}_+}$ can also be made random, as in the case of Cox processes.

Poisson process jump times

In order to determine the distribution of the first jump time T_1 we note that we have the equivalence

$$\{T_1 > t\} \Longleftrightarrow \{N_t = 0\},$$

which implies

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0,$$

i.e., T_1 has an exponential distribution with parameter $\lambda > 0$.

In order to prove the next proposition we note that more generally, we have the equivalence

$$\{T_n > t\} \Longleftrightarrow \{N_t \leqslant n-1\},$$

for all $n \geqslant 1$. This allows us to compute the distribution of the random jump time T_n with its probability density function. It coincides with the *gamma* distribution with integer parameter $n \geqslant 1$, also known as the Erlang distribution in queueing theory.

Proposition 3.2. *For all $n \geqslant 1$, the probability distribution of T_n has the gamma probability density function*

$$t \longmapsto \lambda^n \mathrm{e}^{-\lambda t}\frac{t^{n-1}}{(n-1)!}$$

with shape parameter $n \geqslant 1$ and scaling parameter $\lambda > 0$ on $\mathbb{R}_+$, i.e., for all $t > 0$ the probability $\mathbb{P}(T_n \geqslant t)$ is given by

$$\mathbb{P}(T_n \geqslant t) = \lambda^n \int_t^\infty \mathrm{e}^{-\lambda s}\frac{s^{n-1}}{(n-1)!}ds.$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Proof. We have

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0,$$

and by induction, assuming that

$$\mathbb{P}(T_{n-1} > t) = \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-2}}{(n-2)!} ds, \qquad n \geqslant 2,$$

we obtain

$$\begin{aligned}
\mathbb{P}(T_n > t) &= \mathbb{P}(T_n > t \geqslant T_{n-1}) + \mathbb{P}(T_{n-1} > t) \\
&= \mathbb{P}(N_t = n-1) + \mathbb{P}(T_{n-1} > t) \\
&= \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} + \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-2}}{(n-2)!} ds \\
&= \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} ds, \qquad t \geqslant 0,
\end{aligned}$$

where we applied an integration by parts to derive the last line. □

In particular, for all $n \in \mathbb{Z}$ and $t \in \mathbb{R}_+$, we have

$$\mathbb{P}(N_t = n) = p_n(t) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

i.e., $p_{n-1} : \mathbb{R}_+ \to \mathbb{R}_+$, $n \geqslant 1$, is the probability density function of the random jump time T_n.

In addition to Proposition 3.2 we could show the following proposition which relies on the *strong Markov property*, see *e.g.* Theorem 6.5.4 of Norris (1998).

Proposition 3.3. *The (random) interjump times*

$$\tau_k := T_{k+1} - T_k$$

spent at state $k \geqslant 0$, with $T_0 = 0$, form a sequence of independent identically distributed random variables having the exponential distribution with parameter $\lambda > 0$, i.e.,

$$\mathbb{P}(\tau_0 > t_0, \ldots, \tau_n > t_n) = \mathrm{e}^{-(t_0 + t_1 + \cdots + t_n)\lambda}, \qquad t_0, t_1, \ldots, t_n \geqslant 0.$$

As the expectation of the exponentially distributed random variable τ_k with parameter $\lambda > 0$ is given by

$$\mathbb{E}[\tau_k] = \lambda \int_0^\infty x e^{-\lambda x} dx = \frac{1}{\lambda},$$

we can check that the nth jump time $T_n = \tau_0 + \cdots + \tau_{n-1}$ has the mean

$$\mathbb{E}[T_n] = \frac{n}{\lambda}, \qquad n \geqslant 1.$$

Consequently, the higher the intensity $\lambda > 0$ is (*i.e.*, the higher the probability of having a jump within a small interval), the smaller the time spent in each state $k \geqslant 0$ is on average.

As a consequence of Proposition 3.2, random samples of Poisson process jump times can be generated from Poisson jump times using the following R code according to Proposition 3.3.

```
lambda = 0.6;T=10;Tn=c();n=0;
S=0; while (S<T) {S=S+rexp(1,rate=lambda); Tn=c(Tn,S); n=n+1}
Z<-cumsum(c(0,rep(1,n))); dev.new(width=T, height=5)
plot(stepfun(Tn,Z),xlim =c(0,T),ylim=c(0,8),xlab="t",ylab=expression('N'[t]),pch=1, cex=1,
     col="blue", lwd=2, main="", las = 1, cex.axis=1.2, cex.lab=1.4,xaxs='i',yaxs='i'); grid()
```

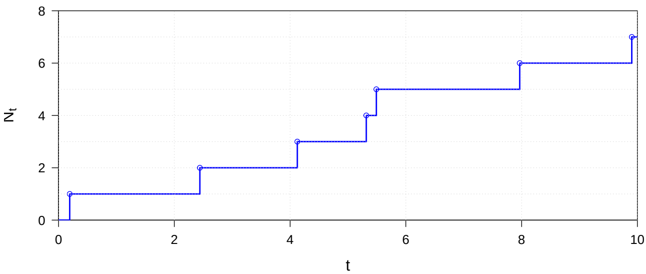


Fig. 3.3: Sample path of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$.

In addition, conditionally to $\{N_T = n\}$, the n jump times on $[0,T]$ of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$ are independent uniformly distributed random variables on $[0,T]^n$, cf. *e.g.* § 11.1 in Privault (2018). This fact can also be useful for the random simulation of Poisson process paths.

```
lambda = 0.6;T=10;n = rpois(1,lambda*T);Tn <- sort(runif(n,0,T)); Z<-cumsum(c(0,rep(1,n)));
    dev.new(width=T, height=5)
plot(stepfun(Tn,Z),xlim =c(0,T),ylim=c(0,8),xlab="t",ylab=expression('N'[t]),pch=1, cex=1,
    col="blue", lwd=2, main="", las = 1, cex.axis=1.2, cex.lab=1.4,xaxs='i',tick.ratio = 0.5);
    grid()
```

Compensated Poisson martingale

From (3.4) above we deduce that

$$\mathbb{E}[N_t - \lambda t] = 0, \tag{3.7}$$

i.e., the compensated Poisson process $(N_t - \lambda t)_{t\in\mathbb{R}_+}$ has *centered increments*.

```
lambda = 0.6;T=10;Tn=c();S=0;n=0;
while (S<T) {S=S+rexp(1,rate=lambda); Tn=c(Tn,S); n=n+1}
Z<-cumsum(c(0,rep(1,n)));
N <- function(t) {return(stepfun(Tn,Z)(t))};t <- seq(0,10,0.01)
dev.new(width=T, height=5)
plot(t,N(t)-lambda*t,xlim = c(0,10),ylim =
    c(-2,2),xlab="t",ylab=expression(paste('N'[t],'-t')),type="l",lwd=2,col="blue",main="",
    xaxs = "i", yaxs = "i", xaxs = "i", yaxs = "i", las = 1, cex.axis=1.2, cex.lab=1.4)
abline(h = 0, col="black", lwd =2)
points(Tn,N(Tn)-lambda*Tn,pch=1,cex=0.8,col="blue",lwd=2)
```

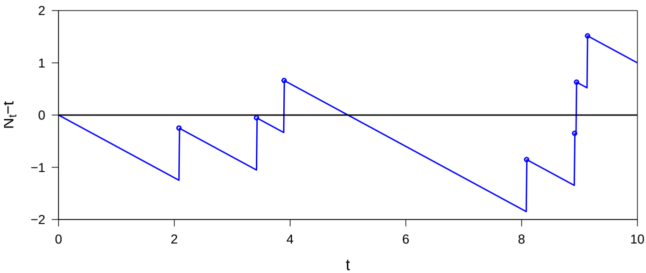


Fig. 3.4: Sample path of the compensated Poisson process $(N_t - \lambda t)_{t\in\mathbb{R}_+}$.

Since in addition $(N_t - \lambda t)_{t\in\mathbb{R}_+}$ also has independent increments, we get the following proposition. We let

$$\mathcal{F}_t := \sigma\big(N_s \ : \ s \in [0,t]\big), \qquad t \geqslant 0,$$

denote the *filtration* generated by the Poisson process $(N_t)_{t\in\mathbb{R}_+}$.

Proposition 3.4. *The compensated Poisson process*

$$(N_t - \lambda t)_{t\in\mathbb{R}_+}$$

is a martingale *with respect* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

The Poisson process belong to the family of *renewal processes*, which are counting processes of the form

$$N_t = \sum_{n\geqslant 1} \mathbb{1}_{[T_n,\infty)}(t), \qquad t \geqslant 0,$$

for which $\tau_k := T_{k+1} - T_k$, $k \geqslant 0$, is a sequence of independent identically distributed random variables.

3.2 Compound Poisson Process

The Poisson process itself appears to be too limited to develop realistic price models as its jumps are of constant size. Therefore there is some interest in considering jump processes that can have random jump sizes.

Let $(Z_k)_{k\geqslant 1}$ denote a sequence of independent, identically distributed (*i.i.d.*) square-integrable random variables, distributed as a common random variable Z with probability distribution $\nu(dy)$ on $\mathbb{R}$, independent of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$. We have

$$\mathbb{P}(Z \in [a,b]) = \nu([a,b]) = \int_a^b \nu(dy), \quad -\infty < a \leqslant b < \infty, \quad k \geqslant 1,$$

and when the distribution $\nu(dy)$ admits a probability density $\varphi(y)$ on $\mathbb{R}$, we write $\nu(dy) = \varphi(y)dy$ and

$$\mathbb{P}(Z \in [a,b]) = \int_a^b \varphi(y)dy, \quad -\infty < a \leqslant b < \infty, \quad k \geqslant 1.$$

Figure 3.5 shows an example of Gaussian jump size distribution.

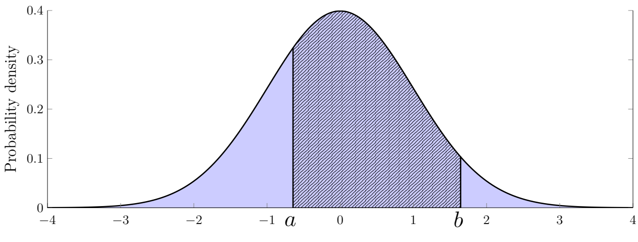


Fig. 3.5: Probability density function φ.

Definition 3.5. *The process* $(Y_t)_{t\in\mathbb{R}_+}$ *given by the random sum*

$$Y_t := Z_1 + Z_2 + \cdots + Z_{N_t} = \sum_{k=1}^{N_t} Z_k, \qquad t \geqslant 0, \tag{3.8}$$

is called a compound Poisson process.[*]

Letting Y_{t^-} denote the left limit

$$Y_{t^-} := \lim_{s \nearrow t} Y_s, \qquad t > 0,$$

we note that the jump size

$$\Delta Y_t := Y_t - Y_{t^-}, \qquad t \geqslant 0,$$

of $(Y_t)_{t\in\mathbb{R}_+}$ at time t is given by the relation

$$\Delta Y_t = Z_{N_t} \Delta N_t, \qquad t \geqslant 0, \tag{3.9}$$

where

$$\Delta N_t := N_t - N_{t^-} \in \{0, 1\}, \qquad t \geqslant 0,$$

denotes the jump size of the standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$, and N_{t^-} is the left limit

$$N_{t^-} := \lim_{s \nearrow t} N_s, \qquad t > 0,$$

The next Figure 3.6 represents a sample path of a compound Poisson process, with here $Z_1 = 0.9$, $Z_2 = -0.7$, $Z_3 = 1.4$, $Z_4 = 0.6$, $Z_5 = -2.5$, $Z_6 = 1.5$, $Z_7 = -0.5$, with the relation

$$Y_{T_k} = Y_{T_k^-} + Z_k, \qquad k \geqslant 1.$$

[*] We use the convention $\sum_{k=1}^{n} Z_k = 0$ if $n = 0$, so that $Y_0 = 0$.

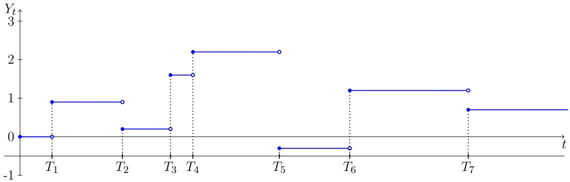


Fig. 3.6: Sample path of a compound Poisson process $(Y_t)_{t\in\mathbb{R}_+}$.

Example. Assume that the jump sizes Z are Gaussian distributed with mean δ and variance η^2, with

$$\nu(dy) = \frac{1}{\sqrt{2\pi\eta^2}} \mathrm{e}^{-(y-\delta)^2/(2\eta^2)} dy.$$

```
N<-50;Tk<-cumsum(rexp(N,rate=0.5)); Zk<-rexp(N,rate=0.5); Yk<-cumsum(c(0,Zk))
plot(stepfun(Tk,Yk),xlim = c(0,10),lwd=2,do.points = F,main="L=0.5",col="blue")
Zk<-rnorm(N,mean=0,sd=1); Yk<-cumsum(c(0,Zk))
plot(stepfun(Tk,Yk),xlim = c(0,10),lwd=2,do.points = F,main="L=0.5",col="blue")
```

Given that $\{N_T = n\}$, the n jump sizes of $(Y_t)_{t\in\mathbb{R}_+}$ on $[0,T]$ are independent random variables which are distributed on $\mathbb{R}$ according to $\nu(dx)$. Based on this fact, the next proposition allows us to compute the *Moment Generating Function* (MGF) of the increment $Y_T - Y_t$.

Proposition 3.6. *For any $t \in [0,T]$ and $\alpha \in \mathbb{R}$ we have*

$$\mathbb{E}\big[\mathrm{e}^{(Y_T-Y_t)\alpha}\big] = \exp\big((T-t)\lambda\big(\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big]-1\big)\big). \tag{3.10}$$

Proof. Since N_t has a Poisson distribution with parameter $t > 0$ and is independent of $(Z_k)_{k\geqslant 1}$, for all $\alpha \in \mathbb{R}$ we have, by conditioning on the value of $N_T - N_t = n$,

$$\begin{aligned}
\mathbb{E}\big[\mathrm{e}^{(Y_T-Y_t)\alpha}\big] &= \mathbb{E}\left[\exp\left(\alpha \sum_{k=N_t+1}^{N_T} Z_k\right)\right] = \mathbb{E}\left[\exp\left(\alpha \sum_{k=1}^{N_T-N_t} Z_{k+N_t}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\alpha \sum_{k=1}^{N_T-N_t} Z_k\right)\right] \\
&= \sum_{n\geqslant 0} \mathbb{E}\left[\exp\left(\alpha \sum_{k=1}^{N_T-N_t} Z_k\right) \,\middle|\, N_T - N_t = n\right] \mathbb{P}(N_T - N_t = n)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n\geqslant 0} \mathbb{E}\left[\exp\left(\alpha \sum_{k=1}^{n} Z_k\right)\right] \mathbb{P}(N_T - N_t = n) \\
&= \mathrm{e}^{-(T-t)\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!}(T-t)^n \mathbb{E}\left[\exp\left(\alpha \sum_{k=1}^{n} Z_k\right)\right] \\
&= \mathrm{e}^{-(T-t)\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!}(T-t)^n \prod_{k=1}^{n} \mathbb{E}\big[\mathrm{e}^{\alpha Z_k}\big] \\
&= \mathrm{e}^{-(T-t)\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!}(T-t)^n \big(\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big]\big)^n \\
&= \exp\Big((T-t)\lambda\big(\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big]-1\big)\Big).
\end{aligned}
$$

□

As a consequence of Proposition 3.6, we can derive the following version of the Lévy-Khintchine formula, after approximating $f : [0,T] \longrightarrow \mathbb{R}$ a bounded deterministic function of time by indicator functions:

$$
\mathbb{E}\left[\exp\left(\int_0^T f(t)dY_t\right)\right] = \exp\left(\lambda \int_0^T \int_{-\infty}^{\infty} \big(\mathrm{e}^{yf(t)}-1\big)\nu(dy)dt\right). \quad (3.11)
$$

We note that we can also write

$$
\begin{aligned}
\mathbb{E}\big[\mathrm{e}^{(Y_T-Y_t)\alpha}\big] &= \exp\left((T-t)\lambda \int_{-\infty}^{\infty} (\mathrm{e}^{\alpha y}-1)\nu(dy)\right) \\
&= \exp\left((T-t)\lambda \int_{-\infty}^{\infty} \mathrm{e}^{\alpha y}\nu(dy) - (T-t)\lambda \int_{-\infty}^{\infty} \nu(dy)\right),
\end{aligned}
$$

since the probability distribution $\nu(dy)$ of Z satisfies

$$
\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big] = \int_{-\infty}^{\infty} \mathrm{e}^{\alpha y}\nu(dy) \quad \text{and} \quad \int_{-\infty}^{\infty} \nu(dy) = 1.
$$

From the moment generating function (3.10) we can compute the expectation and variance of Y_t for fixed t. Note that the proofs of those identities require to exchange the differentiation and expectation operators, which is possible when the moment generating function (3.10) takes finite values for all α in a certain neighborhood $(-\varepsilon, \varepsilon)$ of 0.

Proposition 3.7. *i) The expectation of Y_t is given as the product of the mean number of jump times $\mathbb{E}[N_t] = \lambda t$ and the mean jump size $\mathbb{E}[Z]$, i.e.,*

$$\mathbb{E}[Y_t] = \mathbb{E}[N_t]\mathbb{E}[Z] = \lambda t \mathbb{E}[Z]. \tag{3.12}$$

ii) Regarding the variance, we have

$$\text{Var}\,[Y_t] = \mathbb{E}[N_t]\mathbb{E}\big[|Z|^2\big] = \lambda t \mathbb{E}\big[|Z|^2\big]. \tag{3.13}$$

Proof. (i) We use the relation

$$\mathbb{E}[Y_t] = \frac{\partial}{\partial \alpha}\mathbb{E}[\mathrm{e}^{\alpha Y_t}]_{|\alpha=0} = \lambda t \int_{-\infty}^{\infty} y\nu(dy) = \lambda t \mathbb{E}[Z].$$

(ii) By (3.10), we have

$$\begin{aligned}
\mathbb{E}\big[Y_t^2\big] &= \frac{\partial^2}{\partial \alpha^2}\mathbb{E}[\mathrm{e}^{\alpha Y_t}]_{|\alpha=0} \\
&= \frac{\partial^2}{\partial \alpha^2}\exp\big(\lambda t(\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big]-1)\big)_{|\alpha=0} \\
&= \frac{\partial}{\partial \alpha}\Big(\lambda t \mathbb{E}\big[Z\mathrm{e}^{\alpha Z}\big]\exp\big(\lambda t(\mathbb{E}\big[\mathrm{e}^{\alpha Z}\big]-1)\big)\Big)_{|\alpha=0} \\
&= \lambda t \mathbb{E}\big[Z^2\big] + (\lambda t \mathbb{E}[Z])^2 \\
&= \lambda t \int_{-\infty}^{\infty} y^2\nu(dy) + (\lambda t)^2\left(\int_{-\infty}^{\infty} y\nu(dy)\right)^2 \\
&= \lambda t \mathbb{E}[Z^2] + (\lambda t \mathbb{E}[Z])^2.
\end{aligned}$$

□

Relation (3.12) can be directly recovered using series summations, as

$$\begin{aligned}
\mathbb{E}[Y_t] &= \mathbb{E}\left[\sum_{k=1}^{N_t} Z_k\right] \\
&= \sum_{n\geqslant 1}\mathbb{E}\left[\sum_{k=1}^{N_t} Z_k \,\middle|\, N_t = n\right]\mathbb{P}(N_t = n) \\
&= \mathrm{e}^{-\lambda t}\sum_{n\geqslant 1}\frac{\lambda^n t^n}{n!}\mathbb{E}\left[\sum_{k=1}^{n} Z_k \,\middle|\, N_t = n\right] \\
&= \mathrm{e}^{-\lambda t}\sum_{n\geqslant 1}\frac{\lambda^n t^n}{n!}\mathbb{E}\left[\sum_{k=1}^{n} Z_k\right]
\end{aligned}$$

$$
\begin{aligned}
&= \lambda t \mathrm{e}^{-\lambda t} \mathbb{E}[Z] \sum_{n \geqslant 1} \frac{(\lambda t)^{n-1}}{(n-1)!} \\
&= \lambda t \mathbb{E}[Z] \\
&= \mathbb{E}[N_t] \mathbb{E}[Z].
\end{aligned}
$$

As a consequence, the *dispersion index* of the compound Poisson process

$$
\frac{\operatorname{Var}[Y_t]}{\mathbb{E}[Y_t]} = \frac{\mathbb{E}\left[|Z|^2\right]}{\mathbb{E}[Z]}, \qquad t \geqslant 0.
$$

coincides with the dispersion index of the random jump size Z. By a multivariate version of Theorem A.19, Proposition 3.6 can be used to show the next result.

Proposition 3.8. (i) *The compound Poisson process*

$$
Y_t = \sum_{k=1}^{N_t} Z_k, \qquad t \geqslant 0,
$$

has independent increments, i.e. for any finite sequence of times $t_0 < t_1 < \cdots < t_n$*, the increments*

$$
Y_{t_1} - Y_{t_0},\ Y_{t_2} - Y_{t_1}, \ldots,\ Y_{t_n} - Y_{t_{n-1}}
$$

are mutually independent random variables.

(ii) *In addition, the increment* $Y_t - Y_s$ *is stationary,* $0 \leqslant s \leqslant t$*, i.e. the distribution of* $Y_{t+h} - Y_{s+h}$ *does not depend of* $h \geqslant 0$.

Proof. This result relies on the fact that the result of Proposition 3.6 can be extended to sequences $0 \leqslant t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$, as

$$
\begin{aligned}
\mathbb{E}\left[\prod_{k=1}^{n} \mathrm{e}^{i\alpha_k (Y_{t_k} - Y_{t_{k-1}})}\right] &= \mathbb{E}\left[\exp\left(i \sum_{k=1}^{n} \alpha_k (Y_{t_k} - Y_{t_{k-1}})\right)\right] \\
&= \exp\left(\lambda \sum_{k=1}^{n} (t_k - t_{k-1}) \int_{-\infty}^{\infty} (\mathrm{e}^{i\alpha_k y} - 1) \nu(dy)\right) \qquad (3.14) \\
&= \prod_{k=1}^{n} \exp\left((t_k - t_{k-1}) \lambda \int_{-\infty}^{\infty} (\mathrm{e}^{i\alpha_k y} - 1) \nu(dy)\right) \\
&= \prod_{k=1}^{n} \mathbb{E}\left[\mathrm{e}^{i\alpha_k (Y_{t_k} - Y_{t_{k-1}})}\right],
\end{aligned}
$$

which also shows the stationarity in distribution of $Y_{t+h} - Y_{s+h}$ in $h \geqslant 0$, for $0 \leqslant s \leqslant t$. $\square$

Since the compensated compound Poisson process also has independent and centered increments by (3.7) we have the following counterpart of Proposition 3.4.

Proposition 3.9. *The compensated compound Poisson process*

$$M_t := Y_t - \lambda t \mathbb{E}[Z], \qquad t \geqslant 0,$$

is a martingale*.*

```
lambda = 0.6;T=10;Tn=c();S=0;n=0;
while (S<T) {S=S+rexp(1,rate=lambda); Tn=c(Tn,S); n=n+1}
Z<-cumsum(c(0,rep(1,n))); Zn<-cumsum(c(0,rexp(n,rate=2)));
Y <- function(t) {return(stepfun(Tn,Zn)(t))};t <- seq(0,10,0.01)
par(oma=c(0,0.1,0,0))
plot(t,Y(t)-0.5*lambda*t,xlim = c(0,10),ylim =
c(-2,2),xlab="t",ylab=expression(paste('Y'[t],'-t')),type="l",lwd=2,col="blue",main="", xaxs =
      "i", yaxs = "i", xaxs = "i", yaxs = "i", las = 1, cex.axis=1.2, cex.lab=1.4)
abline(h = 0, col="black", lwd =2)
points(Tn,Y(Tn)-0.5*lambda*Tn,pch=1,cex=0.8,col="blue",lwd=2);grid()
```

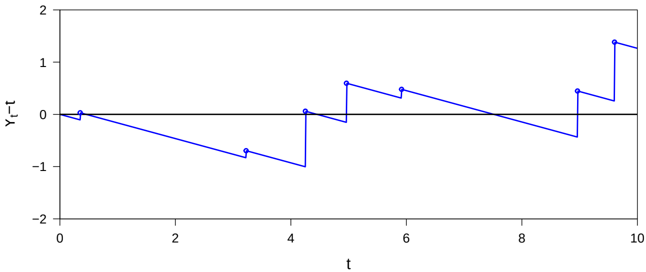


Fig. 3.7: Sample path of a compensated compound Poisson process $(Y_t - \lambda t \mathbb{E}[Z])_{t\in\mathbb{R}_+}$.

3.3 Claim and Reserve Processes

We consider

- a number N_t of claims made until $t \geqslant 0$, which is modeled by a Poisson process $(N_t)_{t\geqslant 0}$ with intensity $\lambda > 0$,
- a sequence $(Z_k)_{k\geqslant 1}$ of nonnegative independent, identically-distributed random variables, which represent the claim amounts.

We assume that the claim amounts $(Z_k)_{k\geqslant 1}$ and the process of arrivals $(N_t)_{t\geqslant 0}$ are independent. In the next definition we use the convention $S(t) = 0$ if $N_t = 0$.

Definition 3.10. *The* aggregate claim amount *up to time t is defined as the compound Poisson process*

$$S(t) = \sum_{k=1}^{N_t} Z_k.$$

The aggregate claim amount $(S(t))_{t\in\mathbb{R}_+}$ can also be written as

$$S(t) = Y_{N_t}, \qquad t \in \mathbb{R}_+,$$

where $(Y_k)_{k\geqslant 1}$ is the sequence of random variables independent of $(N_t)_{t\in\mathbb{R}_+}$ given by

$$Y_k = \sum_{j=1}^{k} Z_j = Z_1 + \cdots + Z_k, \qquad k \geqslant 1,$$

with $Y_0 := 0$. In the next definition, $f : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function mapping $t > 0$ to the premium income $f(t)$ received between time 0 and time t, with $f(0) = 0$.

Definition 3.11. Standard compound Poisson risk model. *The surplus (or reserve) process $(R_x(t))_{t\geqslant 0}$ is defined as*

$$R_x(t) = x + f(t) - S(t), \qquad t \geqslant 0,$$

where $x \geqslant 0$ is the amount of initial reserves and $f(t)$ is the premium income received between time 0 and time $t > 0$.

In the next Figures 3.8 and 3.9 we take $f(t) := ct$ with $c = 0.5$.

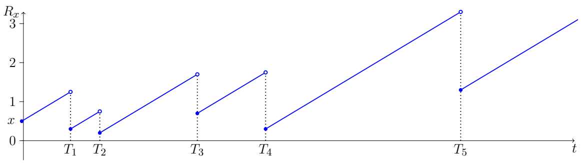


Fig. 3.8: Sample path (without ruin) of a reserve process $(R_x(t))_{t\in\mathbb{R}_+}$.

Unlike the above figure, the next Figure 3.9 contains a ruin event.

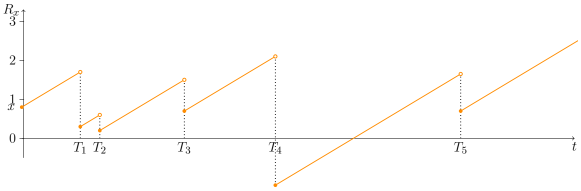


Fig. 3.9: Sample path (with ruin) of a reserve process $(R_x(t))_{t\in\mathbb{R}_+}$.

3.4 Ruin Probabilities

We will consider the infinite time ruin probability

$$\Psi(x) = \mathbb{P}\big(\exists\, t \geqslant 0 \ : \ R_x(t) < 0\big),$$

with $\Psi(x) = 1$ for $x < 0$, and the finite-time ruin probability defined as

$$\Psi_T(x) = \mathbb{P}\big(\exists\, t \in [0,T] \ : \ R_x(t) < 0\big),$$

given $T > 0$ a finite time horizon, with $\Psi_T(x) = 1$ for $x < 0$.

Denoting by m_0^T the infimum

$$m_0^T := \min_{0\leqslant t\leqslant T} (f(t) - S(t)),$$

the ruin probability $\Psi_T(x)$ can also be written as

$$\Psi_T(x) = \mathbb{P}\big(m_0^T < -x\big), \qquad x \geqslant 0.$$

Cramér-Lundberg Model

In Proposition 3.12 we compute the ruin probability in infinite time starting from an initial reserve $x \geqslant 0$.

Proposition 3.12. *Assume that the premium income function satisfies $f(t) = ct$ with premium rate $c > 0$.*

a) The ruin probability in infinite time starting from the initial reserve $x = 0$ is given by

$$\Psi(0) = \mathbb{P}\big(\exists\, t \geqslant 0 \ : \ R_0(t) < 0\big) = \frac{\lambda\mu}{c},$$

provided that $c \geqslant \lambda\mu$, where $\mu = \mathbb{E}[Z]$, and $\Psi(0) = 1$ if $c < \lambda\mu$.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

b) Assume that the claim sizes $(Z_k)_{k\geqslant 1}$ form a sequence of independent, exponentially distributed random variables with mean $\mu > 0$, i.e. with parameter $1/\mu$. Then, the ruin probability in infinite time starting from the initial reserve $x \geqslant 0$ is given by

$$\Psi(x) = \frac{\lambda\mu}{c}\mathrm{e}^{(\lambda/c-1/\mu)x}, \qquad x \geqslant 0, \tag{3.15}$$

provided that $c \geqslant \lambda\mu$, with $\Psi(x) = 1$ if $c < \lambda\mu$.

Proof. Let

$$\Phi(x) := 1 - \Psi(x) = \mathbb{P}\left(R_x(t) \geqslant 0, \ \forall t \geqslant 0\right)$$

denote the probability of non-ruin starting from an initial reserve $x \geqslant 0$. Since $c \geqslant 0$, letting

$$F(z) := \mathbb{P}(Z_1 \leqslant z), \qquad z \geqslant 0,$$

denote the cumulative distribution function of the claim size Z_1, for all $y \geqslant 0$ we have

$$\begin{aligned}
\Phi(y) &= \mathbb{P}(R_y(t) \geqslant 0, \ \forall \ t \geqslant 0) \\
&= \mathbb{P}\left(y + ct - \sum_{k=1}^{N_t} Z_k \geqslant 0, \ \forall \ t \geqslant 0\right) \\
&= \mathbb{P}\left(y + ct - \sum_{k=1}^{N_t} Z_k \geqslant 0, \ \forall \ t \geqslant T_1\right) \\
&= \mathbb{E}\left[\mathbb{1}_{\left\{y+cT_1-Z_1+c(t-T_1)-\sum_{k=2}^{N_t} Z_k \geqslant 0, \ \forall \ t \geqslant T_1\right\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\left\{y+cT_1-Z_1+c(t-T_1)-\sum_{k=2}^{N_t} Z_k \geqslant 0, \ \forall \ t \geqslant T_1\right\}} \,\middle|\, T_1\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(y + cT_1 - Z_1 + c(t-T_1) - \sum_{k=2}^{N_t} Z_k \geqslant 0, \ \forall \ t \geqslant T_1 \,\middle|\, T_1\right)\right] \\
&= \mathbb{E}\left[\Phi(y + cT_1 - Z_1)\right] \\
&= \lambda \int_0^\infty \mathrm{e}^{-\lambda s} \int_0^\infty \Phi(y + cs - z) dF(z) ds \\
&= \lambda \int_0^\infty \mathrm{e}^{-\lambda s} \int_0^{y+cs} \Phi(y + cs - z) dF(z) ds \\
&= \frac{\lambda}{c} \int_0^\infty \mathrm{e}^{-\lambda u/c} \int_0^{y+u} \Phi(y + u - z) dF(z) du
\end{aligned}$$

$$= \frac{\lambda}{c} e^{\lambda y/c} \int_y^\infty e^{-\lambda u/c} \int_0^u \Phi(u-z) dF(z) du. \qquad (3.16)$$

By differentiating (3.16) with respect to y, we find

$$\Phi'(y) = \frac{\lambda}{c} \left(\Phi(y) - \int_0^y \Phi(y-z) dF(z) ds \right), \qquad (3.17)$$

hence by integration by parts with respect to $z \in [0, y]$, we get

$$\begin{aligned}
\Phi(y) &= \Phi(0) + \int_0^y \Phi'(u) du \\
&= \Phi(0) + \frac{\lambda}{c} \int_0^y \Phi(u) du - \frac{\lambda}{c} \int_0^y \int_0^u \Phi(u-z) dF(z) du \\
&= \Phi(0) + \frac{\lambda}{c} \int_0^y \Phi(y-z)(1 - F(z)) dz,
\end{aligned}$$

a) Case $x = 0$. Letting y tend to infinity in the above inequality, we deduce

$$\begin{aligned}
\Phi(\infty) &= \Phi(0) + \frac{\lambda}{c} \int_0^\infty \Phi(\infty - z)(1 - F(z)) dz \\
&= \Phi(0) + \frac{\lambda}{c} \Phi(\infty) \int_0^\infty (1 - F(z)) dz \\
&= \Phi(0) + \Phi(\infty) \frac{\lambda}{c} \int_0^\infty \mathbb{P}(Z > z) dz \\
&= \Phi(0) + \Phi(\infty) \frac{\lambda}{c} \mathbb{E}[Z] \\
&= \Phi(0) + \Phi(\infty) \frac{\lambda \mu}{c}, \qquad (3.18)
\end{aligned}$$

since

$$\begin{aligned}
\int_0^\infty \mathbb{P}(Z_1 > z) dz &= \int_0^\infty \mathbb{E}[\mathbb{1}_{\{Z_1 > z\}}] dz \\
&= \mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{Z_1 > z\}} dz \right] \\
&= \mathbb{E}\left[\int_0^{Z_1} dz \right] \\
&= \mathbb{E}[Z_1] \\
&= \mu
\end{aligned}$$

is the average claim size. From (3.18) we have

$$\Phi(\infty) = \Phi(0) + \Phi(\infty) \frac{\lambda \mu}{c}.$$

When $\lambda\mu > c$ we find $\Phi(0) = \Phi(\infty) = 0$, whereas when $\lambda\mu \leqslant c$ we have $\Phi(\infty) = 1$ and we obtain $\Phi(0) = 1 - \lambda\mu/c$. In particular, the infinite time ruin probability $\Psi(0)$ starting from the initial reserve $x = 0$ is given by

$$\begin{aligned}\Psi(0) &= 1 - \Phi(0) \\ &= \mathbb{P}\big(\exists\, t \geqslant 0 \ : \ R_x(t) < 0\big) \\ &= \frac{\lambda\mu}{c},\end{aligned} \tag{3.19}$$

provided that $\lambda\mu \leqslant c$.

b) Case $x > 0$. We refer to Exercise 3.2 for the computation of the ruin probability $\Psi(x)$ starting from any $x > 0$. when the claim sizes $(Z_k)_{k\geqslant 1}$ are exponentially distributed. □

Analytic expressions for finite time ruin probabilities have also been obtained when $(Y_k)_{k\geqslant 1}$ are independent, exponentially distributed random variables with parameter $\mu > 0$ and $f(t) = ct$ is linear, $c \geqslant 0$, Theorem 4.1 and Relation (4.6) of Dozzi and Vallois (1997) show that

$$\begin{aligned}\Psi_T(x) &= \mathbb{P}\big(m_0^T < -x\big) \\ &= \lambda \int_0^T \left(x \sum_{n\geqslant 0} \frac{(\lambda\mu t(x+ct))^n}{(n!)^2} + ct \sum_{n\geqslant 0} \frac{(\lambda\mu t(x+ct))^n}{n!(n+1)!} \right) \frac{\mathrm{e}^{-\mu(x+ct)-\lambda t}}{x+ct} dt,\end{aligned}$$

see also Theorem 3.1 of León and Villa (2009) for other related expressions.

R simulation[*]

The following R code provides an approximation of the infinite time ruin probability (3.15) of Proposition 3.12 by Monte Carlo simulation when T is sufficiently large, see also (3.21) in Exercise 3.2.

[*] See Kaas et al. (2009), Example 4.3.7.

```
T=20; # Use T>=500 to approximate infinite time dev.new(width=T, height=5)
nSim = 50; lambda = 0.1; x = 7.5; mu = 10; c = 3; N <- rep(Inf, nSim)
for (k in 1:nSim){tauk <- rexp(10*T*lambda,lambda);Ti <- cumsum(tauk)
n=length(Ti[Ti<T]);if (n>=1) {Zk <- rexp(n,1/mu);Si <- x + Ti*c
Ri <- Si - cumsum(Zk);RRi <- Si - c(0,cumsum(Zk)[1:n-1]);
Si<-Si[Ti<T];Ri<-Ri[Ti<T];RRi<-RRi[Ti<T]
Ri <- c(Ri,Ri[n]+c*(T-Ti[n]));RRi <- c(RRi,Ri[n+1]);Ti <- c(Ti[Ti<T],T);
ruin <- !all(Ri[1:n]>=0);}
else {ruin<-FALSE;Ti=c(T);RRi=x+c*T;Ri=x+c*T;};color="blue";
if (ruin) {N[k] <- min(which(Ri<0));color="orange"}
par(mgp=c(0.8,1,1));par(mar=c(2,2,2,2))
plot(c(0,rbind(Ti,Ti)),c(x,rbind(RRi,Ri)),xlab="Time
      t",xlim=c(0,T*0.99),ylim=c(-c*T/3,x+c*T),lwd=3,ylab="R(t)",type="l",col=color,
      main=paste(length(N[N<Inf]),"/",k,"=",format(length(N[N<Inf])/k,digits=4)),
      axes=FALSE, cex.lab=1.4)
axis(1, pos=0, las = 1, cex.axis=1.2);axis(2, pos=0, las = 1, cex.axis=1.2);Sys.sleep(0.2)}
N <- N[N<Inf];length(N);mean(N);sd(N);max(N)
cat('Theoretical value:',lambda*mu*exp(-x*(1/mu-lambda/c))/c,'\n')
cat('Simulation:',length(N)/nSim,'\n')
```

Figure 3.10 computes an estimate of the infinite time ruin probability $\Psi(x)$ by generating the sample paths of the reserve process $(R_x(t))_{t\in\mathbb{R}_+}$.

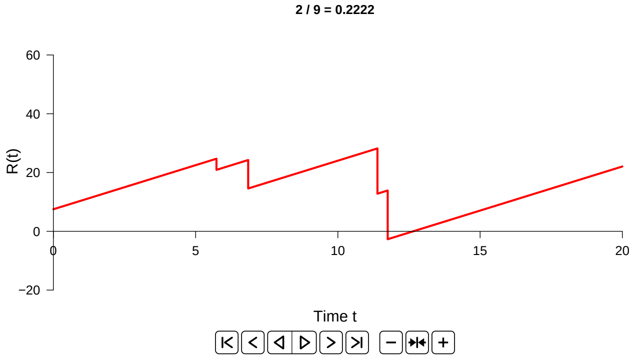


Fig. 3.10: Sample paths of a reserve process $(R_x(t))_{t\in\mathbb{R}_+}$.*

Probability density function

The probability density function of m_0^T at $-x<0$ can be computed as

$$-\frac{\partial \Psi_T}{\partial x}(x).$$

* The animation works in Acrobat Reader on the entire pdf file.

An important practical problem is to obtain numerical values of the sensitivity of the finite-time ruin probability with respect to the initial reserve

$$\frac{\partial \Psi_T}{\partial x}(x),$$

in particular due to new solvency regulations in Europe. The problem of computing the corresponding sensitivity for the finite-time ruin probability $\Psi_T(x)$ has been covered in Loisel and Privault (2009) based on multiple integration. Formulas for the finite-time ruin probability

$$\Psi_T(x) = \mathbb{P}\big(\exists\, t \in [0,T] \ : \ R_x(t) < 0\big)$$

have been proposed in Picard and Lefèvre (1997), see also De Vylder (1999) and Ignatova et al. (2001), Rullière and Loisel (2004). In Privault and Wei (2004; 2007), the Malliavin calculus has been used to provide a way to compute the sensitivity of the probability

$$\mathbb{P}\left(R_x(T) < 0\right)$$

that the terminal surplus is negative with respect to parameters such as the initial reserve or the interest rate of the model.

Non-constant rate of income

When the company income is an arbitrary function $f(t)$ of time such that $f(0) := 0$ we clearly we have $m_0^T \leqslant 0 = f(0)$, hence the distribution of m_0^T is supported on $(-\infty, 0]$. On the other hand, we have $m_0^T = 0$ if and only if $N_T = 0$ or $f(T_k) - Y_k > 0$ for all $k \geqslant 1$ such that $T_k \leqslant T$, hence the distribution of m_0^T has a Dirac mass at 0 with weight

$$\begin{aligned}\mathbb{P}\big(m_0^T = 0\big) &= \mathbb{P}(N_T = 0) + \mathbb{P}\big(\{m_0^T \geqslant 0\} \cap \{N_T \geqslant 1\}\big) \\ &= \mathrm{e}^{-\lambda T} + \mathrm{e}^{-\lambda T}\mathbb{E}\left[\sum_{k\geqslant 1} \lambda^k \int_0^T \int_0^{t_k} \cdots \int_0^{t_2} \mathbb{1}_{\{f(t_1) > Y_1\}} \cdots \mathbb{1}_{\{f(t_k) > Y_k\}} dt_1 \cdots dt_k\right],\end{aligned}$$

where we used the fact that Poisson jump times are independent uniformly distributed on the square $[0,T]^n$ given that $\{N_T = n\}$.

On the other hand, since f is increasing we have

$$m_0^T = \inf_{T_k \leqslant T,\ k \geqslant 0} (f(T_k) - Y_k) = \mathbb{1}_{\{N_T \geqslant 1\}} \inf_{T_k \leqslant T,\ k \geqslant 1} (f(T_k) - Y_k),$$

with $T_0 = 0$. Hence we have the integral expression

$$
\begin{aligned}
&P\big(\{m_0^T \geqslant y\} \cap \{N_T \geqslant 1\}\big) \qquad (3.20)\\
&= \mathrm{e}^{-\lambda T}\mathbb{E}\left[\sum_{k\geqslant 1}\lambda^k \int_0^T \int_0^{t_k} \cdots \int_0^{t_2} \mathbb{1}_{\{y<\inf_{1\leqslant l\leqslant k}(f(t_l)-Y_l)\}} dt_1 \cdots dt_k\right]\\
&= \lambda \mathrm{e}^{-\lambda T}\mathbb{E}\left[\sum_{k\geqslant 0}\lambda^k \int_0^T \int_0^{t_{k+1}} \cdots \int_0^{t_2} \mathbb{1}_{\{f(t_1)>Y_1+y\}} \cdots \mathbb{1}_{\{f(t_{k+1})>Y_{k+1}+y\}} dt_1 \cdots dt_{k+1}\right]
\end{aligned}
$$

Random rate of income

Here, we consider the infimum

$$
m_0^T = \inf_{0\leqslant t\leqslant T}(X_t - S(t))
$$

where $(X_t)_{t\in\mathbb{R}_+}$ is a stochastic process with independent increments and $X_0 = 0$, independent of $(S(t))_{t\in\mathbb{R}_+}$, and such that

$$
\inf_{t\in[a,b]} X_t, \qquad 0 \leqslant a < b,
$$

has a probability density function denoted by $\phi_{a,b}(x)$. For example, if $(X_t)_{t\in\mathbb{R}_+}$ is a standard Brownian motion then $\phi_{a,b}(x)$ is given by

$$
\begin{aligned}
&\int_x^\infty \phi_{a,b}(z)dz = \mathbb{P}\left(\inf_{t\in[a,b]} X_t \geqslant x\right)\\
&= \mathbb{E}\left[\mathbb{1}_{\{X_a<x\}}\mathbb{P}\left(\inf_{t\in[a,b]} X_t \geqslant x \;\middle|\; X_a\right)\right] + \mathbb{E}\left[\mathbb{1}_{\{X_a\geqslant x\}}\mathbb{P}\left(\inf_{t\in[a,b]} X_t \geqslant x \;\middle|\; X_a\right)\right]\\
&= \mathbb{E}\left[\mathbb{1}_{\{X_a<x\}}\mathbb{P}\left(\inf_{t\in[0,b-a]} B_t \geqslant x - X_a \;\middle|\; X_a\right)\right] + \mathbb{P}(X_a \geqslant x)\\
&= 2\mathbb{E}\big[\mathbb{1}_{\{X_a<x\}}\mathbb{P}\left(B_{b-a} \geqslant x - X_a \mid X_a\right)\big] + \mathbb{P}(X_a \geqslant x)\\
&= \frac{1}{\pi\sqrt{a(b-a)}}\int_0^\infty \mathrm{e}^{-(x-y)^2/(2a)}\int_y^\infty \mathrm{e}^{-z^2/(2(b-a))}dzdy + \frac{1}{\sqrt{2\pi a}}\int_x^\infty \mathrm{e}^{-z^2/(2a)}dz.
\end{aligned}
$$

We have $m_0^T \leqslant X_0 = 0$ *a.s.*, hence the distribution of m_0^T is carried by $(-\infty, 0]$.

Guaranteed Maturity Benefits

Variable annuity benefits offered by insurance companies are usually protected via different mechanisms such as Guaranteed Minimum Maturity Benefits (GMMBs) or Guaranteed Minimum Death Benefits (GMDBs). The computation of the corresponding risk measures is an important issue for the practitioner in risk management.

Given a fund value process $(F_t)_{t\in\mathbb{R}_+}$, an insurer is continuously charging annualized mortality and expense fees at the rate m from the account of variable annuities, resulting into a margin offset income M_t given by

$$M_t := mF_t \qquad t \in \mathbb{R}_+.$$

Denoting by τ_x the future lifetime of a policyholder at the age x, the future payment made by the insurer at maturity T is

$$(G - F_T)^+ \mathbb{1}_{\{\tau_x > T\}}$$

where G is the guarantee level expressed as a percentage of the initial fund value F_0, δ is a roll-up rate according to which the guarantee increases up to the payment time. In this case, the random variable X is taken equal to

$$X := \mathrm{e}^{-rT}(G - F_T)^+ \mathbb{1}_{\{\tau_x > T\}} - \int_0^{\min(T,\tau_x)} \mathrm{e}^{-rs} M_s \mathrm{d}s.$$

Exercises

Exercise 3.1 Consider N a Poisson random variable with distribution

$$\mathbb{P}(N = n) = \mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}, \qquad n = 0, 1, 2, \ldots$$

where $\lambda > 0$, and let $Y := \sum_{k=1}^{N} Z_k$, where $(Z_k)_{k\geqslant 1}$ is a sequence of independent centered $\mathcal{N}(0, \sigma^2)$ Gaussian random variables with variance σ^2 and cumulative distribution function

$$\mathbb{P}(Z_k \leqslant x) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x} \mathrm{e}^{-y^2/(2\sigma^2)}dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x/\sigma} \mathrm{e}^{-y^2/2}dy = \Phi\left(\frac{x}{\sigma}\right),$$

$x \in \mathbb{R}_+$.

a) Compute $\mathbb{P}(Y \geqslant y)$ using the conditioning

$$\mathbb{P}(Y \geqslant y) = \sum_{n\geqslant 1} \mathbb{P}\left(\sum_{k=1}^{N} Z_k \geqslant y \;\middle|\; N = n\right)\mathbb{P}(N = n) = \cdots$$

b) Find $\mathbb{E}[Y]$.

Exercise 3.2 Show that when the claim size distribution is exponential with mean $\mu > 0$, *i.e.* when $F(z) = 1 - \mathrm{e}^{-z/\mu}$, $z \geqslant 0$, the ruin probability is given by

$$\Psi(x) = \mathbb{P}\big(\exists\, t \in \mathbb{R}_+ \ : \ R_x(t) < 0\big) = \frac{\lambda\mu}{c}\mathrm{e}^{(\lambda/c - 1/\mu)x}, \quad x \geqslant 0, \qquad (3.21)$$

provided that $c \geqslant \lambda\mu$.

Exercise 3.3 An insurance company receives continuous-time premium income at the rate $\$\mu$ per year. Claim payments are filed by subscribers according to a Poisson process $(N_t)_{t\in\mathbb{R}_+}$ of intensity $\lambda > 0$ claims per year. All claims have same constant amount $\$C > 0$.

a) Compute the expected value $\mathbb{E}[R_T]$ and variance $\mathbb{E}[(R_T - \mathbb{E}[R_T])^2]$ of the company's reserve $R_T := R_0 + \mu T - CN_T$ at time $T > 0$, with constant initial reserve R_0.
b) Express the probability $\mathbb{P}(R_T < 0)$ of ruin at time T using the Poisson probability mass function $\mathbb{P}(N_T = k) = \mathrm{e}^{-\lambda T}(\lambda T)^k/k!$, $k \geqslant 0$.

Exercise 3.4 Consider

- a number N_t of claims made until $t \geqslant 0$, which is modeled by a Poisson process $(N_t)_{t\geqslant 0}$ with intensity $\lambda > 0$,
- a sequence $(Z_k)_{k\geqslant 1}$ of nonnegative independent, identically-distributed random variables, which represent the claim amounts.

We assume that the claim amounts $(Z_k)_{k\geqslant 1}$ and the process $(N_t)_{t\geqslant 0}$ of arrivals are independent. The *aggregate claim amount* made up to time t to an insurance company is defined as the compound Poisson process

$$S(t) := \sum_{k=1}^{N_t} Z_k = Z_1 + Z_2 + \cdots + Z_{N_t}, \qquad t \in [0, T].$$

The initial reserve of the company is denoted by $x \geqslant 0$ and the premium income received up to time $t \geqslant 0$ is denoted by $f(t)$.

a) Give the mean and variance of $S(T)$.
Hint: Use the mean $\mathbb{E}[N_T] = \lambda T$ and the moments $\mathbb{E}[Z_1]$ and $\mathbb{E}[Z_1^2]$.
b) Using the Chebyshev inequality (3.22), provide an upper bound for the ruin probability $\mathbb{P}(x + f(T) - S(T) < 0)$ at time $T > 0$, provided that $x + f(T) - \lambda T\mathbb{E}[Z_1] > 0$.

Hint: By the Chebyshev inequality inequality, for any random variable X with mean $\mu > 0$ and variance σ^2 we have

$$\mathbb{P}(X \leqslant 0) = \mathbb{P}(X - \mu \leqslant -\mu) \leqslant \mathbb{P}(|X - \mu| \geqslant \mu) \leqslant \frac{\sigma^2}{\mu^2}. \tag{3.22}$$

Chapter 4
Correlation and Dependence

Correlation and dependence play a capital role in risk management, in particular when assessing or preventing any potential "domino effect" arising from interactions between different entities exposed to uncertainties.* This chapter presents several standard models for the statistical interactions that arise in the modeling of correlated risk. For this, we use the concept of copulas, which can model the uncertainty and dependence properties observed between random variables or data samples.

4.1 Joint Bernoulli Distribution

Our study of dependence structures starts with the simplest case of two correlated random variables X and Y, each of them taking only two possible values. For this, let X and Y by two Bernoulli random variables, with

$$p_X = \mathbb{P}(X = 1) = \mathbb{E}[\mathbb{1}_{\{X=1\}}] \quad \text{and} \quad p_Y = \mathbb{P}(Y = 1) = \mathbb{E}[\mathbb{1}_{\{Y=1\}}]$$

and correlation coefficient

$$\rho := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}\,[X]\mathrm{Var}\,[Y]}}$$

* Correlation does not imply causation. Try "Spurious Correlations".

$$= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathrm{Var}\,[X]\mathrm{Var}\,[Y]}}$$
$$= \frac{\mathbb{P}(X = 1 \text{ and } Y = 1) - p_X p_Y}{\sqrt{p_X(1-p_X)p_Y(1-p_Y)}},$$

with $\rho \in [-1, 1]$ from the Cauchy-Schwarz inequality. We note that in this case, the joint distribution $\mathbb{P}(X = i \text{ and } Y = j)$, $i, j = 0, 1$, is fully determined by the data of $p_X = \mathbb{P}(X = 1)$, $p_Y = \mathbb{P}(Y = 1)$ and the correlation coefficient $\rho \in [-1, 1]$, as

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = \mathbb{E}[XY] \\ \qquad\qquad = p_X p_Y + \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = \mathbb{E}[(1-X)Y] = \mathbb{P}(Y = 1) - \mathbb{P}(X = 1 \text{ and } Y = 1) \\ \qquad\qquad = (1-p_X)p_Y - \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = \mathbb{E}[X(1-Y)] = \mathbb{P}(X = 1) - \mathbb{P}(X = 1 \text{ and } Y = 1) \\ \qquad\qquad = p_X(1-p_Y) - \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = \mathbb{E}[(1-X)(1-Y)] \\ \qquad\qquad = (1-p_X)(1-p_Y) + \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \end{cases}$$

see Exercise 4.2.

4.2 Joint Gaussian Distribution

Consider now two *centered* Gaussian random variables $X \simeq \mathcal{N}(0, \sigma^2)$ and $Y \simeq \mathcal{N}(0, \eta^2)$ with probability density functions

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-x^2/(2\sigma^2)} \quad \text{and} \quad f_Y(x) = \frac{1}{\sqrt{2\pi\eta^2}} \mathrm{e}^{-x^2/(2\eta^2)}, \qquad x \in \mathbb{R}.$$

Let

$$\rho = \operatorname{corr}(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}\,[X]\mathrm{Var}\,[Y]}} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathrm{Var}\,[X]\mathrm{Var}\,[Y]}}.$$

When the covariance matrix

$$\Sigma := \begin{bmatrix} \mathbb{E}[X^2] & \mathbb{E}[XY] \\ \mathbb{E}[XY] & \mathbb{E}[Y^2] \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma\eta \\ \rho\sigma\eta & \eta^2 \end{bmatrix} \tag{4.1}$$

with determinant

$$\begin{aligned}\det \Sigma &= \mathbb{E}\big[X^2\big]\mathbb{E}\big[Y^2\big] - \big(\mathbb{E}[XY]\big)^2 \\ &= \mathbb{E}\big[X^2\big]\mathbb{E}\big[Y^2\big]\big(1 - \big(\operatorname{corr}(X,Y)\big)^2\big) \\ &\geqslant 0,\end{aligned}$$

is invertible, there exists a probability density function

$$\begin{aligned} f_\Sigma(x,y) &= \frac{1}{\sqrt{2\pi \det \Sigma}} \exp\left(-\frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^\top \Sigma^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \right) \qquad (4.2) \\ &= \frac{1}{\sqrt{2\pi \det \Sigma}} \exp\left(-\frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^\top \begin{bmatrix} \mathbb{E}\big[X^2\big] & \mathbb{E}[XY] \\ \mathbb{E}[XY] & \mathbb{E}\big[Y^2\big] \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \right), \end{aligned}$$

with respective marginals $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(0, \eta^2)$.

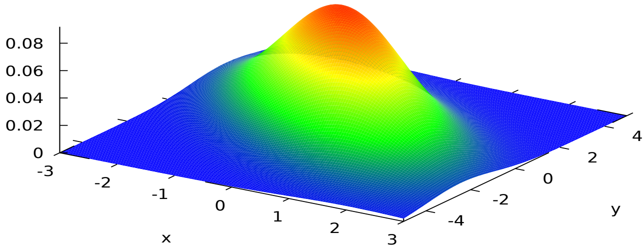


Fig. 4.1: Joint Gaussian probability density.

The probability density function (4.2) is called the centered joint (bivariate) Gaussian probability density with covariance matrix Σ.

Note that when $\rho = \operatorname{corr}(X,Y) = \pm 1$ we have $\det \Sigma = 0$ and the joint probability density function $f_\Sigma(x,y)$ is *not defined*.

Definition 4.1. *A random vector $(X_1, \ldots, X_n)$ is said to have a multivariate centered Gaussian distribution if every linear combination*

$$a_1 X_1 + \cdots + a_n X_n, \quad a_1, \ldots, a_n \in \mathbb{R}, \ n \geqslant 1,$$

has a centered Gaussian distribution.

Recall that if $(X_1, \ldots, X_n)$ has a multivariate centered Gaussian distribution, then its probability density function takes the form

$$f_\Sigma(x_1,\ldots,x_n) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x_1,\ldots,x_n)^T \Sigma^{-1}(x_1,\ldots,x_n)\right),$$

$(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where Σ is the covariance matrix

$$\Sigma = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_{n-1}) & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}[X_2] & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \mathrm{Var}[X_{n-1}] & \mathrm{Cov}(X_{n-1},X_n) \\ \mathrm{Cov}(X_1,X_n) & \mathrm{Cov}(X_2,X_n) & \cdots & \mathrm{Cov}(X_{n-1},X_n) & \mathrm{Var}[X_n] \end{bmatrix}.$$

The next remark plays an important role in the modeling of joint default probabilities, see **here** for a detailed discussion.

Remark 4.2. *There exist couples (X, Y) of random variables with Gaussian marginals $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(0, \eta^2)$, such that*

i) *(X, Y) does not have the bivariate Gaussian distribution with probability density function $f_\Sigma(x, y)$, where Σ is the covariance matrix (4.1) of (X, Y).*
ii) *the random variable $X + Y$ is not even Gaussian.*

Proof. See Exercise 4.5. □

4.3 Copulas and Dependence Structures

The word copula derives from the Latin noun for a "link" or "tie" that connects two different objects or concepts.

Definition 4.3. *A two-dimensional copula is any joint cumulative distribution function*

$$\begin{aligned} C : [0,1] \times [0,1] &\longrightarrow [0,1] \\ (u,v) &\longmapsto C(u,v) \end{aligned}$$

with uniform $[0,1]$-valued marginals.

In other words, any copula function $C(u, v)$ can be written as

$$C(u,v) = \mathbb{P}(U \leqslant u \text{ and } V \leqslant v), \qquad 0 \leqslant u, v \leqslant 1,$$

where U and V are uniform $[0,1]$-valued random variables.

Examples.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

i) The copula corresponding to independent uniform random variables (U, V) is given by

$$\begin{aligned} C(u, v) &= \mathbb{P}(U \leqslant u \text{ and } V \leqslant v) \\ &= \mathbb{P}(U \leqslant u)\mathbb{P}(V \leqslant v) \\ &= uv, \qquad 0 \leqslant u, v \leqslant 1. \end{aligned}$$

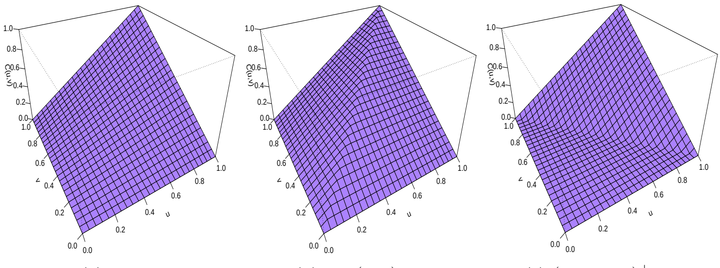
ii) The copula corresponding to the fully correlated case $U = V$ is given by

$$\begin{aligned} C(u, v) &= \mathbb{P}(U \leqslant u \text{ and } V \leqslant v) \\ &= \mathbb{P}(U \leqslant \min(u, v)) \\ &= \min(u, v), \qquad 0 \leqslant u, v \leqslant 1. \end{aligned}$$

iii) The copula corresponding to the fully anticorrelated case $U = 1 - V$ is given by

$$\begin{aligned} C(u, v) &:= \mathbb{P}(U \leqslant u \text{ and } V \leqslant v) \\ &= \mathbb{P}(U \leqslant u \text{ and } 1 - U \leqslant v) \\ &= \mathbb{P}(1 - v \leqslant U \leqslant u) \\ &= (u + v - 1)^+, \qquad 0 \leqslant u, v \leqslant 1. \end{aligned}$$

The above copulas are plotted in Figure 4.3a.



(a) uv. (b) $\min(u, v)$. (c) $(u + v - 1)^+$.

Fig. 4.2: Copula graphs $C(u, v) = uv$, $C(u, v) = \min(u, v)$, $C(u, v) = (u + v - 1)^+$.

In what follows, F_X^{-1} denotes the inverse of the *Cumulative Distribution Function* F_X of X.

Lemma 4.4. *Assume that the random variable X has a* continuous and strictly increasing *cumulative distribution function* $F_X(x) := \mathbb{P}(X \leqslant x)$, $x \in \mathbb{R}$. *Then, $U := F_X(X)$ is uniformly distributed on* $[0, 1]$.

Proof. We have

$$\begin{aligned} F_U(u) &= \mathbb{P}(U \leqslant u) \\ &= \mathbb{P}(F_X(X) \leqslant u) \\ &= \mathbb{P}(X \leqslant F_X^{-1}(u)) \\ &= F_X(F_X^{-1}(u)) \\ &= u, \qquad 0 \leqslant u \leqslant 1. \end{aligned}$$

□

As in Lemma 4.4, given (X, Y) a couple of random variables with joint cumulative distribution function

$$F_{(X,Y)}(x, y) := \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y), \qquad x, y \in \mathbb{R},$$

and continuous strictly increasing marginal cumulative distribution functions

$$F_X(x) = F_{(X,Y)}(x, \infty) = \mathbb{P}(X \leqslant x) \text{ and } F_Y(y) = F_{(X,Y)}(\infty, y) = \mathbb{P}(Y \leqslant y),$$

we note the following points.

i) The random variables

$$U := F_X(X) \quad \text{and} \quad V := F_Y(Y)$$

are uniformly distributed on $[0, 1]$.

ii) The copula function

$$C_{(X,Y)}(u, v) := \mathbb{P}(U \leqslant u \text{ and } V \leqslant v), \qquad 0 \leqslant u, v \leqslant 1,$$

satisfies

$$\begin{aligned} C_{(X,Y)}(u, v) &:= \mathbb{P}(U \leqslant u \text{ and } V \leqslant v) \\ &= \mathbb{P}\big(F_X(X) \leqslant u \text{ and } F_Y(Y) \leqslant v\big) \\ &= \mathbb{P}\big(X \leqslant F_X^{-1}(u) \text{ and } Y \leqslant F_Y^{-1}(v)\big) \\ &= F_{(X,Y)}\big(F_X^{-1}(u), F_Y^{-1}(v)\big), \qquad 0 \leqslant u, v \leqslant 1. \end{aligned}$$

iii) The joint cumulative distribution function of (X, Y) can be recovered from the copula $C_{(X,Y)}$ and the marginal cumulative distribution functions F_X, F_Y as

$$\begin{aligned} F_{(X,Y)}(x, y) &= \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y) \\ &= \mathbb{P}\big(F_X(X) \leqslant F_X(x) \text{ and } F_Y(Y) \leqslant F_Y(v)\big) \end{aligned}$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$= \mathbb{P}\big(U \leqslant F_X(x) \text{ and } V \leqslant F_Y(v)\big)$$
$$= C_{(X,Y)}\big(F_X(x), F_Y(y)\big), \qquad x, y \in \mathbb{R}.$$

Higher dimensional copulas

Definition 4.5. *An n-dimensional copula is any joint cumulative distribution function*

$$\begin{array}{rcl} C : [0,1] \times \cdots \times [0,1] & \longrightarrow & [0,1] \\ (u_1, \ldots, u_n) & \longmapsto & C(u_1, \ldots, u_n) \end{array}$$

of n uniform $[0,1]$-valued random variables.

Consider the joint cumulative distribution function

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$$

of a family $(X_1, \ldots, X_n)$ of random variables with marginal cumulative distribution functions

$$F_{X_i}(x) = F_{(X_1,\ldots,X_n)}(\infty, \ldots, +infty, x, \infty, \ldots, \infty), \quad x \in \mathbb{R},$$

$i = 1, 2, \ldots, n$. The copula defined in the next Sklar's Theorem 4.6 encodes the dependence structure of the vector $(X_1, \ldots, X_n)$.

Theorem 4.6. *[Sklar's theorem* (Sklar (1959; 2010))] Given a joint cumulative distribution function $F_{(X_1,\ldots,X_n)}$, there exists an n-dimensional copula $C(u_1, \ldots, u_n)$ such that*

$$F_{(X_1,\ldots,X_n)}(x_1, x_2, \ldots, x_n) = C\big(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\big),$$

$x_1, x_2, \ldots, x_n \in \mathbb{R}$.

The following corollary is a consequence of Sklar's Theorem 4.6.

Corollary 4.7. *Assume that the marginal distribution functions F_{X_i} are continuous and strictly increasing. Then the joint cumulative distribution function $F_{(X_1,\ldots,X_n)}$ defines a n-dimensional copula*

$$C(u_1, \ldots, u_n) := F_{(X_1,\ldots,X_n)}\big(F_{X_1}^{-1}(u_1), \ldots, F_{X_n}^{-1}(u_n)\big), \tag{4.3}$$

* "The author considers continuous non-decreasing functions C_n on the n-dimensional cube $[0,1]^n$ with $C_n(0, \ldots, 0) = 0$, $C_n(1, \ldots, 1, \alpha, 1, \ldots, 1) = \alpha$. Several theorems are stated relating n-dimensional distribution functions and their marginals in terms of functions C_n. No proofs are given." M. Loève, Math. Reviews MR0125600.

$u_1, u_2, \ldots, u_n \in [0,1]$, *which encodes the dependence structure of the vector* $(X_1, \ldots, X_n)$.

It can be checked as in Lemma 4.4 that $C(u_1, \ldots, u_n)$ defined in (4.3) has uniform marginal distributions on $[0,1]$, as

$$\begin{aligned}
&C(1, \ldots, 1, u, 1, \ldots, 1) \\
&\quad = F_{(X_1,\ldots,X_n)}\big(F_{X_1}^{-1}(1), \ldots, F_{X_{i-1}}^{-1}(1), F_{X_i}^{-1}(u), F_{X_{i+1}}^{-1}(1), \ldots, F_{X_n}^{-1}(1)\big) \\
&\quad = F_{(X_1,\ldots,X_n)}\big(\infty, \ldots, \infty, F_{X_i}^{-1}(u), \infty, \ldots, \infty\big) \\
&\quad = F_{X_i}\big(F_{X_i}^{-1}(u)\big) \\
&\quad = u, \qquad 0 \leqslant u \leqslant 1.
\end{aligned}$$

In the following proposition, we construct a vector of random variables from the data of a copula and a family of marginal distributions.

Proposition 4.8. *Given a family of continuous strictly increasing cumulative distribution functions* $F_1, \ldots, F_n$ *and a multidimensional copula* $C(u_1, \ldots, u_n)$, *the function*

$$F^C_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) := C(F_1(x_1), \ldots, F_n(x_n)), \quad x_1, x_2, \ldots, x_n \in \mathbb{R}, \tag{4.4}$$

defines a joint cumulative distribution function with marginals $F_1, \ldots, F_n$.

Proof. Given $(U_1, \ldots, U_n)$ a vector of n uniform random variables having the copula $C(u_1, \ldots, u_n)$ for cumulative distribution function, we let

$$X_1 := F_1^{-1}(U_1), \ldots, X_n := F_n^{-1}(U_n).$$

Then, we have

$$\begin{aligned}
\mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) &= \mathbb{P}\big(F_1^{-1}(U_1) \leqslant x_1, \ldots, F_n(U_n) \leqslant x_n\big) \\
&= \mathbb{P}(U_1 \leqslant F_1(x_1), \ldots, U_n \leqslant F_n(x_n)) \\
&= C(F_1(x_1), \ldots, F_n(x_n)) \\
&= F^C_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n), \quad x_1, x_2, \ldots, x_n \in \mathbb{R}.
\end{aligned}$$

We can also check that the marginal distributions generated by $F^C_{(X_1,\ldots,X_n)}$ coincide with the respective marginals of $(X_1, \ldots, X_n)$, as we have

$$\begin{aligned}
&F^C_{(X_1,\ldots,X_n)}(\infty, \ldots, \infty, u, \infty, \ldots, \infty) \\
&\quad = C\big(F_1(\infty), \ldots, F_{i-1}(\infty), F_i(u), F_{i+1}(\infty), \ldots, F_n(\infty)\big) \\
&\quad = C\big(1, \ldots, 1, F_i(u), 1, \ldots, 1\big) \\
&\quad = F_i(u), \qquad 0 \leqslant u \leqslant 1.
\end{aligned}$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

□

4.4 Examples of Copulas

Gaussian copulas

The choice of (4.2) above as joint probability density function, see Figure 4.1, actually induces a particular dependence structure between the Gaussian random variables X and Y, and corresponding to the joint cumulative distribution function

$$\begin{aligned}\Phi_\Sigma(x,y) &:= \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y)\\ &= \frac{1}{\sqrt{2\pi \det \Sigma}} \int_{-\infty}^{x}\int_{-\infty}^{y} \exp\left(-\frac{1}{2}\left\langle \begin{bmatrix} u \\ v \end{bmatrix}, \Sigma^{-1}\begin{bmatrix} u \\ v \end{bmatrix}\right\rangle\right) dudv,\end{aligned}$$

$x, y \in \mathbb{R}$. In case the random variables X, Y are normalized centered Gaussian random variables with unit variance, Σ is given by

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

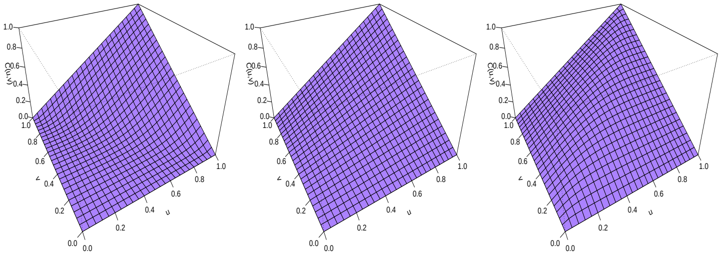
with correlation parameter $\rho \in (-1,1)$. Letting

$$F_X(x) := \mathbb{P}(X \leqslant x) \quad \text{and} \quad F_Y(y) := \mathbb{P}(Y \leqslant y),$$

denote the cumulative distribution functions of X and Y, the random variables $F_X(X)$ and $F_Y(Y)$ are known to be uniformly distributed on $[0,1]$, and $(F_X(X), F_Y(Y))$ is a $[0,1]\times[0,1]$-valued random variable with joint cumulative distribution function

$$\begin{aligned}C_\Sigma(u,v) &:= \mathbb{P}\big(F_X(X) \leqslant u \text{ and } F_Y(Y) \leqslant v\big)\\ &= \mathbb{P}\big(X \leqslant F_X^{-1}(u) \text{ and } Y \leqslant F_Y^{-1}(v)\big)\\ &= \Phi_\Sigma\big(F_X^{-1}(u), F_Y^{-1}(v)\big), \qquad 0 \leqslant u, v \leqslant 1.\end{aligned} \tag{4.5}$$

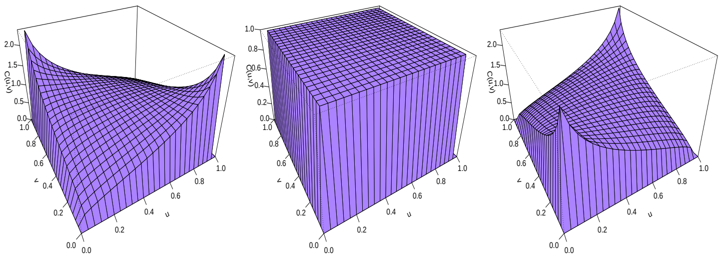
The function $C_\Sigma(u,v)$, which is the joint cumulative distribution function of a couple of uniformly distributed $[0,1]$-valued random variables, is called the *Gaussian copula* generated by the jointly Gaussian distribution of (X,Y) with covariance matrix Σ.



(a) $\rho = -0.9$. (b) $\rho = 0$. (c) $\rho = 0.9$.

Fig. 4.3: Gaussian copula graphs for $\rho = -0.9$, $\rho = 0$, and $\rho = 0.9$.

The graphs of Figures 4.3-(a) and 4.3-(c) correspond to intermediate dependence levels given by Gaussian copulas, cf. (4.5).

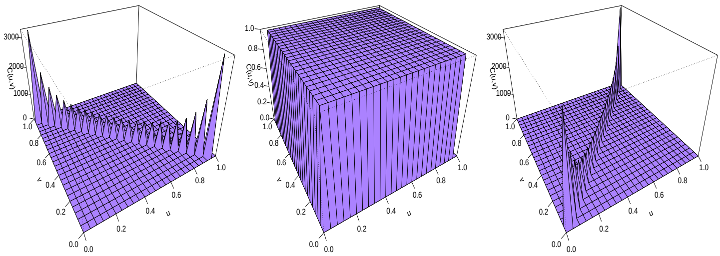


(a) $\rho = -0.35$. (b) $\rho = 0$. (c) $\rho = 0.35$.

Fig. 4.4: Different Gaussian copula *density* graphs for $\rho = -0.35$, $\rho = 0$ and $\rho = 0.35$.

Figures 4.4 and 4.5 present the corresponding copula density graphs.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html



(a) $\rho = -0.999$. (b) $\rho = 0$. (c) $\rho = 0.999$.

Fig. 4.5: Different Gaussian copula *density* graphs for $\rho = -0.999$, $\rho = 0$ and $\rho = 0.999$.

Figure 4.4-(a) represents a uniform (product) probability density function on the square $[0,1] \times [0,1]$, which corresponds to two independent uniformly distributed $[0,1]$-valued random variables U, V. Figure 4.4-(c) shows the probability distribution of the fully correlated couple (U,U), which does not admit a probability density on the square $[0,1] \times [0,1]$.

The Gaussian copula $C_\Sigma(u,u)$ admits a probability density function on $[0,1] \times [0,1]$ given by

$$\begin{aligned}
c_\Sigma(u,v) &= \frac{\partial^2 C_\Sigma}{\partial u \partial v}(u,v) \\
&= \frac{\partial^2}{\partial u \partial v} \Phi_\Sigma\big(F_X^{-1}(u), F_Y^{-1}(v)\big) \\
&= \frac{\partial}{\partial u}\left(\frac{1}{F_Y'(F_Y^{-1}(v))}\frac{\partial \Phi_\Sigma}{\partial y}\big(F_X^{-1}(u), F_Y^{-1}(v)\big)\right) \\
&= \frac{\partial}{\partial u}\left(\frac{1}{f_Y(F_Y^{-1}(v))}\frac{\partial \Phi_\Sigma}{\partial y}\big(F_X^{-1}(u), F_Y^{-1}(v)\big)\right) \\
&= \frac{1}{f_X(F_X^{-1}(u)) f_Y(F_Y^{-1}(v))}\frac{\partial^2 \Phi_\Sigma}{\partial x \partial y}\big(F_X^{-1}(u), F_Y^{-1}(v)\big) \\
&= \frac{f_\Sigma\big(F_X^{-1}(u), F_Y^{-1}(v)\big)}{f_X(F_X^{-1}(u)) f_Y(F_Y^{-1}(v))},
\end{aligned}$$

hence the Gaussian copula $C_\Sigma(u,v)$ can be computed as

$$C_\Sigma(u,v) = \int_0^u \int_0^v c_\Sigma(a,b)\,da\,db$$

$$= \int_0^u \int_0^v \frac{f_\Sigma\big(F_X^{-1}(a), F_Y^{-1}(b)\big)}{f_X(F_X^{-1}(a))f_Y(F_Y^{-1}(b))} dadb, \qquad 0 \leqslant u, v \leqslant 1.$$

The joint cumulative distribution function $F_{(X,Y)}(x, y)$ of (X, Y) can be recovered from Corollary 4.7 as

$$F_{(X,Y)}(x, y) = C_\Sigma(F_X(x), F_Y(y)), \qquad x, y \in \mathbb{R}. \tag{4.6}$$

from the Gaussian copula $C_\Sigma(x, y)$ and the respective cumulative distribution functions $F_X(x)$, $F_Y(y)$ of X and Y.

In that sense, the Gaussian copula $C_\Sigma(x, y)$ encodes the Gaussian dependence structure of the covariance matrix Σ. Moreover, the Gaussian copula $C_\Sigma(x, y)$ can be used to generate a joint distribution function $F^C_{(X,Y)}(x, y)$ by letting

$$F^C_{(X,Y)}(x, y) := C_\Sigma(F_X(x), F_Y(y)), \qquad x, y \in \mathbb{R}, \tag{4.7}$$

based on other, *possibly non-Gaussian* cumulative distribution functions $F_X(x)$, $F_Y(y)$ of two random variables X and Y. In this case we note that the marginals of the joint cumulative distribution function $F^C_{(X,Y)}(x, y)$ are $F_X(x)$ and $F_Y(y)$ because $C_\Sigma(x, y)$ has uniform marginals on $[0, 1]$.

Gumbel copula

The Gumbel copula is given by

$$C(u, v) = \exp\left(-\left((-\log u)^\theta + (-\log v)^\theta\right)^{1/\theta}\right), \qquad 0 \leqslant u, v \leqslant 1,$$

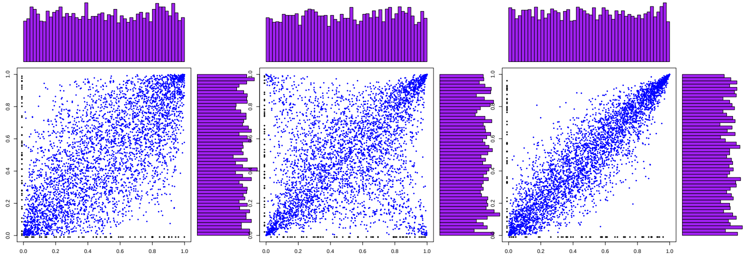
with $\theta \geqslant 1$, and $C(u, v) = uv$ when $\theta = 1$.

Uniform marginals with given copulas

The following R code generates random samples according to the Gaussian, Student, and Gumbel copulas with uniform marginals, as illustrated in Figure 4.6.

```
install.packages("copula"); install.packages("gumbel")
library(copula);library(gumbel)
norm.cop <- normalCopula(0.35); norm.cop
persp(norm.cop, pCopula, n.grid = 51, xlab="u", ylab="v", zlab="C(u,v)", main="", sub="",
      col='lightblue')
persp(norm.cop, dCopula, n.grid = 51, xlab="u", ylab="v", zlab="c(u,v)", main="", sub="",
      col='lightblue')
norm <- rCopula(4000,normalCopula(0.7))
plot(norm[,1],norm[,2],cex=3,pch='.',col='blue')
```

```
stud <- rCopula(4000,tCopula(0.5,dim=2,df=1))
points(stud[,1],stud[,2],cex=3,pch='.',col='red')
gumb <- rgumbel(4000,4)
points(gumb[,1],gumb[,2],cex=3,pch='.',col='green')
```



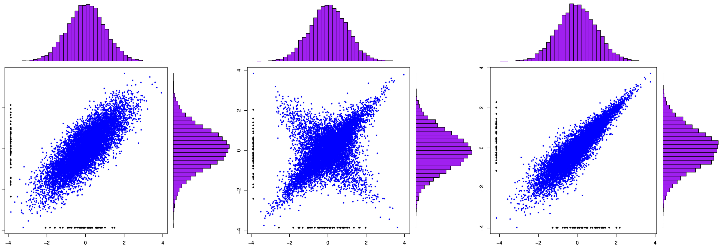
(a) Gaussian copula. (b) Student copula. (c) Gumbel copula.

Fig. 4.6: Samples with uniform marginals and given copulas.

The following R code is plotting the histograms of Figure 4.6.

```
joint_hist <- function(u) {x <- u[,1]; y <- u[,2]
xhist <- hist(x, breaks=40,plot=FALSE) ; yhist <- hist(y, breaks=40,plot=FALSE)
top <- max(c(xhist$counts, yhist$counts))
nf <- layout(matrix(c(2,0,1,3),2,2,byrow=TRUE), c(3,1), c(1,3), TRUE)
par(mar=c(3,3,1,1))
plot(x, y, xlab="", ylab="",col="blue",pch=19,cex=0.4)
points(x, -0.01+rep(min(y),length(x)), xlab="", ylab="",col="black",pch=18,cex=0.8)
points(-0.01+rep(min(x),length(y)), y, xlab="", ylab="",col="black",pch=18,cex=0.8)
par(mar=c(0,3,1,1))
barplot(xhist$counts, axes=FALSE, ylim=c(0, top), space=0,col="purple")
par(mar=c(3,0,1,1))
barplot(yhist$counts, axes=FALSE, xlim=c(0, top), space=0, horiz=TRUE,col="purple")}
joint_hist(norm);joint_hist(stud);joint_hist(gumb)
```

Gaussian marginals with given copulas

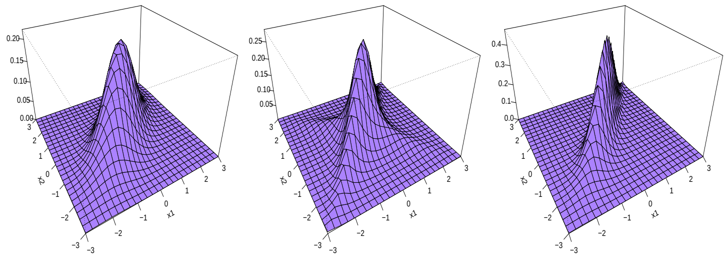


(a) Gaussian copula. (b) Student copula. (c) Gumbel copula.

Fig. 4.7: Samples with Gaussian marginals and given copulas.

The next R code generates random samples according to the Gaussian, Student, and Gumbel copulas with Gaussian marginals, as illustrated in Figure 4.7.

```
set.seed(100);N=10000
gaussMVD<-mvdc(normalCopula(0.8), margins=c("norm","norm"),
      paramMargins=list(list(mean=0,sd=1),list(mean=0,sd=1)))
norm <- rMvdc(N,gaussMVD)
studentMVD<-mvdc(tCopula(0.5,dim=2,df=1), margins=c("norm","norm"),
      paramMargins=list(list(mean=0,sd=1),list(mean=0,sd=1)))
stud <- rMvdc(N,studentMVD)
gumbelMVD<-mvdc(gumbelCopula(param=4, dim=2), margins=c("norm","norm"),
      paramMargins=list(list(mean=0,sd=1),list(mean=0,sd=1)))
gumb <- rMvdc(N,gumbelMVD)
plot(norm[,1],norm[,2],cex=3,pch='.',col='blue')
points(norm[,1], -0.01+rep(min(norm[,2]),N), xlab="", ylab="",col="black",pch=18,cex=0.8)
points(-0.01+rep(min(norm[,1]),N), norm[,2], xlab="", ylab="",col="black",pch=18,cex=0.8)
plot(stud[,1],stud[,2],cex=3,pch='.',col='blue')
points(stud[,1], -0.01+rep(min(stud[,2]),N), xlab="", ylab="",col="black",pch=18,cex=0.8)
points(-0.01+rep(min(stud[,1]),N), stud[,2], xlab="", ylab="",col="black",pch=18,cex=0.8)
plot(gumb[,1],gumb[,2],cex=3,pch='.',col='blue')
points(gumb[,1], -0.01+rep(min(gumb[,2]),N), xlab="", ylab="",col="black",pch=18,cex=0.8)
points(-0.01+rep(min(gumb[,1]),N), gumb[,2], xlab="", ylab="",col="black",pch=18,cex=0.8)
joint_hist(norm);joint_hist(stud);joint_hist(gumb)
```

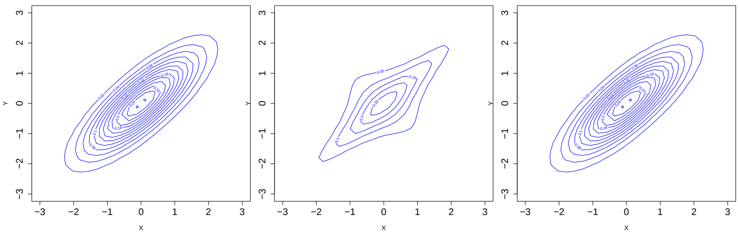


(a) Gaussian copula. (b) Student copula. (c) Gumbel copula.

Fig. 4.8: Joint densities with Gaussian marginals and given copulas.

The following R code is plotting joint densities with Gaussian marginals and given copulas, as illustrated in Figure 4.8.

```
persp(gaussMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
persp(studentMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
persp(gumbelMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
```



(a) Gaussian copula. (b) Student copula. (c) Gumbel copula.

Fig. 4.9: Joint density contour plots with Gaussian marginals and given copulas.

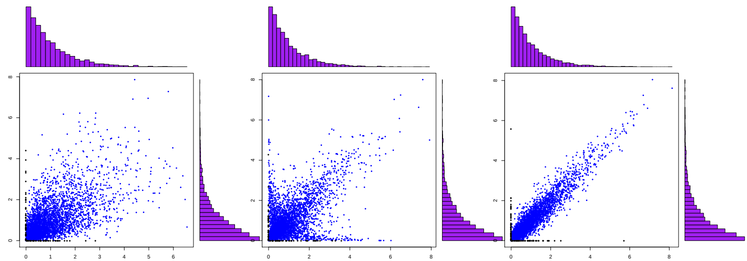
The following R code generates countour plots with Gaussian marginals and given copulas, as illustrated in Figure 4.9.

```
contour(gaussMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
contour(studentMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
contour(gaussMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
```

Exponential marginals with given copulas

The following R code generates random samples with exponential marginals according to the Gaussian, Student, and Gumbel copulas as illustrated in Figure 4.10.

```
library(copula);set.seed(100);N=4000
gaussMVD<-mvdc(normalCopula(0.7), margins=c("exp","exp"),
      paramMargins=list(list(rate=1),list(rate=1)))
norm <- rMvdc(N,gaussMVD)
studentMVD<-mvdc(tCopula(0.5,dim=2,df=1), margins=c("exp","exp"),
      paramMargins=list(list(rate=1),list(rate=1)))
stud <- rMvdc(N,studentMVD)
gumbelMVD<-mvdc(gumbelCopula(param=4, dim=2), margins=c("exp","exp"),
      paramMargins=list(list(rate=1),list(rate=1)))
gumb <- rMvdc(N,gumbelMVD)
plot(norm[,1],norm[,2],cex=3,pch='.',col='blue'); plot(stud[,1],stud[,2],cex=3,pch='.',col='blue');
      plot(gumb[,1],gumb[,2],cex=3,pch='.',col='blue')
persp(gaussMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
persp(studentMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
persp(gumbelMVD, dMvdc, xlim = c(0,1), ylim=c(0,1), col='lightblue')
contour(gaussMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
contour(studentMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
contour(gaussMVD,dMvdc,xlim=c(0,1),ylim=c(0,1),nlevels=10,xlab="X",ylab="Y",cex.axis=1.5)
joint_hist(norm); joint_hist(stud); joint_hist(gumb)
```



(a) Gaussian copula. (b) Student copula. (c) Gumbel copula.

Fig. 4.10: Samples with exponential marginals and given copulas.

Exercises

Exercise 4.1 Copulas. In what follows, U denotes a uniformly distributed $[0,1]$-valued random variable.

a) To which couple (U,V) of uniformly distributed $[0,1]$-valued random variables does the copula function

$$C_M(u,v) = \min(u,v), \qquad 0 \leqslant u,v \leqslant 1,$$

correspond?

b) Show that the function

$$C_m(u,v) := (u+v-1)^+, \qquad 0 \leqslant u,v \leqslant 1,$$

is the copula on $[0,1] \times [0,1]$ corresponding to $(U,V) = (U, 1-U)$.

c) Show that for any copula function $C(u,v)$ on $[0,1] \times [0,1]$ we have

$$C(u,v) \leqslant C_M(u,v), \qquad 0 \leqslant u,v \leqslant 1. \tag{4.8}$$

d) Show that for any copula function $C(u,v)$ on $[0,1] \times [0,1]$ we also have

$$C_m(u,v) \leqslant C(u,v), \qquad 0 \leqslant u,v \leqslant 1. \tag{4.9}$$

Hint: For fixed $v \in [0,1]$, let $h(u) := C(u,v) - (u+v-1)$ and show that $h(1) = 0$ and $h'(u) \leqslant 0$.

Exercise 4.2 Consider two Bernoulli random variables X and Y, with $p_X = \mathbb{P}(X=1)$, $p_Y = \mathbb{P}(Y=1)$, correlation coefficient $\rho \in [-1,1]$, and

$$\begin{cases} \mathbb{P}(X=1 \text{ and } Y=1) = p_X p_Y + \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X=0 \text{ and } Y=1) = (1-p_X)p_Y - \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X=1 \text{ and } Y=0) = p_X(1-p_Y) - \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X=0 \text{ and } Y=0) = (1-p_X)(1-p_Y) + \rho\sqrt{p_X p_Y (1-p_X)(1-p_Y)}. \end{cases}$$

a) Is it possible to have $\rho = 1$ *without* having $p_X = p_Y$ and

$$\begin{cases} \mathbb{P}(X=1 \text{ and } Y=1) = p_X = p_Y, \\ \\ \mathbb{P}(X=0 \text{ and } Y=1) = 0, \\ \\ \mathbb{P}(X=1 \text{ and } Y=0) = 0, \\ \\ \mathbb{P}(X=0 \text{ and } Y=0) = 1 - p_X = 1 - p_Y \text{ ?} \end{cases}$$

b) Similarly, is it possible to have $\rho = 1$ *without* having $p_X = 1 - p_Y$ and

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = p_Y, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = p_X, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = 0 \text{ ?} \end{cases}$$

Exercise 4.3 Exponential copulas. Consider the random vector (X, Y) of nonnegative random variables, whose joint distribution is given by the survival function

$$\mathbb{P}(X \geqslant x \text{ and } Y \geqslant y) := \mathrm{e}^{-\lambda x - \mu y - \nu \operatorname{Max}(x,y)}, \qquad x, y \in \mathbb{R}_+,$$

where $\lambda, \mu, \nu > 0$.

a) Find the marginal distributions of X and Y.
b) Find the joint cumulative distribution function $F(x, y) := \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y)$ of (X, Y).
c) Construct an "exponential copula" based on the joint cumulative distribution function of (X, Y).

Exercise 4.4 Gumbel bivariate logistic distribution. Consider the random vector (X, Y) of nonnegative random variables, whose joint distribution is given by the joint cumulative distribution function (CDF)

$$F_{(X,Y)}(x, y) := \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y) := \frac{1}{1 + \mathrm{e}^{-x} + \mathrm{e}^{-y}}, \qquad x, y \in \mathbb{R}.$$

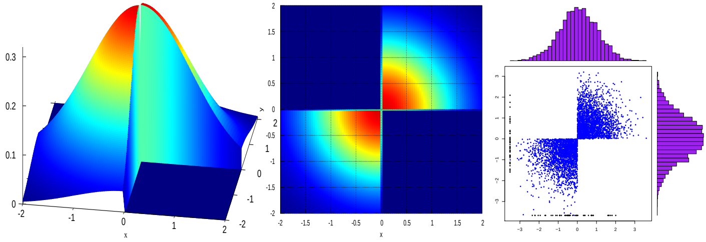
a) Find the marginal distributions of X and Y.
b) Construct the copula based on the joint CDF of (X, Y).

Exercise 4.5 Consider the random vector (X, Y) with the joint probability density function

$$\widetilde{f}(x, y) := \frac{1}{\pi\sigma\eta} \mathbb{1}_{\mathbb{R}^2_-}(x, y) \mathrm{e}^{-x^2/(2\sigma^2) - y^2/(2\eta^2)} + \frac{1}{\pi\sigma\eta} \mathbb{1}_{\mathbb{R}^2_+}(x, y) \mathrm{e}^{-x^2/(2\sigma^2) - y^2/(2\eta^2)},$$

plotted as a heat map in Figure 4.11b.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html



(a) 3D view. (b) Heat map. (c) Samples.

Fig. 4.11: Truncated two-dimensional Gaussian density.

```
library(MASS)
Sigma <- matrix(c(1,0,0,1),2,2);N=10000
u<-mvrnorm(N,rep(0,2),Sigma);j=1
for (i in 1:N){
if (u[i,1]>0 && u[i,2]>0) {j<-j+1;}
if (u[i,1]<0 && u[i,2]<0) {j<-j+1;}}
v<-matrix(nrow=j-1, ncol=2);j=1
for (i in 1:N){
if (u[i,1]>0 && u[i,2]>0) {v[j,]=u[i,];j<-j+1;}
if (u[i,1]<0 && u[i,2]<0) {v[j,]=u[i,];j<-j+1;}}
joint_hist(v) # Function defined the previous section
```

a) Show that (X, Y) has the Gaussian marginals $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(0, \eta^2)$.
b) Does the couple (X, Y) have the bivariate Gaussian distribution with probability density function $f_\Sigma(x, y)$, where Σ is the covariance matrix (4.1) of (X, Y)?
c) Show that the random variable $X + Y$ is not Gaussian (take $\sigma = \eta = 1$ for simplicity).
d) Show that under the rotation

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

of angle $\theta \in [0, 2\pi]$ the random vector $(X\cos\theta - Y\sin\theta, X\sin\theta + Y\cos\theta)$ can have an arbitrary covariance depending on the value of $\theta \in [0, 2\pi]$.

Exercise 4.6 Let τ_1, τ_2 and τ denote three independent exponentially distributed random times with respective parameters $\lambda_1, \lambda_2, \lambda > 0$. Consider two firms with respective default times $\tau_1 \wedge \tau = \min(\tau_1, \tau)$ and $\tau_2 \wedge \tau = \min(\tau_2, \tau)$, where τ represents the time of a macro-economic shock.

a) Find the tail (or survival) distribution functions of $\tau_1 \wedge \tau$ and $\tau_2 \wedge \tau$.

b) Compute the joint survival probability

$$\mathbb{P}(\tau_1 \wedge \tau > s \text{ and } \tau_2 \wedge \tau > t), \qquad s, t \geqslant 0.$$

Hint: Use the relation

$$\mathrm{Max}(s, t) = s + t - \min(s, t), \qquad s, t \geqslant 0.$$

c) Compute the joint cumulative distribution function

$$\mathbb{P}(\tau_1 \wedge \tau \leqslant s \text{ and } \tau_2 \wedge \tau \leqslant t), \qquad s, t \geqslant 0.$$

d) Compute the resulting copula

$$C(u, v) := F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)), \qquad 0 \leqslant u, v \leqslant 1.$$

e) Compute the resulting copula density function $\dfrac{\partial^2 C}{\partial u \partial v}(u, v)$, $u, v \in [0, 1]$.

Part II
Risk Measures

Chapter 5
Superhedging Risk Measure

In this chapter, we measure risk using the prices of financial derivatives such as options, that protect their holders against various kinds of market events. For this, we review some basic knowledge of call or put options and related financial derivatives, together with their pricing in the Black-Scholes framework. As a result, we introduce the superhedging risk measure, which can be defined from the price of a portfolio that hedges a given financial derivative.

5.1 Historical Sketch

Early accounts of option contracts can also be found in *The Politics* Aristotle (350 BCE) by Aristotle (384-322 BCE). Option credit contracts appear to have been used as early as the $10th$ century by traders in the Mediterranean.

Referring to the philosopher Thales of Miletus (c. 624 - c. 546 BCE), Aristotle writes:

> "He (Thales) knew by his skill in the stars while it was yet winter that there would be a great harvest of olives in the coming year; so, having a little money, he gave *deposits* for the use of all the olive-presses in Chios and Miletus, which he hired at a low price because no one bid against him. When the harvest-time came, and many were wanted all at once and of a sudden, he let them out at any rate which he pleased, and made a quantity of money".

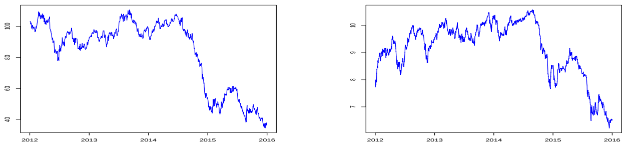
More recently, Robert Merton and Myron Scholes shared the 1997 Nobel Prize in economics: “In collaboration with Fisher Black, developed a pioneering formula for the valuation of stock options ... paved the way for economic valuations in many areas ... generated new types of financial instruments and facilitated more efficient risk management in society.”* See Black and Scholes (1973) “The Pricing of Options and Corporate Liabilities”. Journal of Political Economy 81 (3): 637-654.

The development of options pricing tools contributed greatly to the expansion of option markets and led to development several ventures such as the “Long Term Capital Management” (LTCM), founded in 1994. The fund yielded annualized returns of over 40% in its first years, but registered a loss of US$4.6 billion in less than four months in 1998, which resulted into its closure in early 2000.

As of year 2015, the size of the financial derivatives market is estimated at over one quadrillion (or one million billions, or 10^{15}) USD, which is more than 10 times the size of the total Gross World Product (GWP).

5.2 Financial Derivatives

The following graphs exhibit a correlation between commodity (oil) prices and an oil-related asset price.



(a) WTI price graph.

(b) Graph of Keppel Corp. stock price

Fig. 5.1: Comparison of WTI *vs.* Keppel price graphs.

The study of financial derivatives aims at finding functional relationships between the price of an underlying asset (a company stock price, a commodity price, etc.) and the price of a related financial contract (an option, a financial derivative, etc.).
In the above quote by Aristotle, olive oil can be regarded as the underlying asset, while the oil press stands for the financial derivative.

* This has to be put in relation with the modern development of Risk Societies; “societies increasingly preoccupied with the future (and also with safety), which generates the notion of risk” (Wikipedia).

Next, we move to a description of (European) call and put options, which are at the basis of risk management.

European put option contracts

As previously mentioned, an important concern for the buyer of a stock at time t is whether its price S_T can decline at some future date T. The buyer of the stock may seek protection from a market crash by purchasing a contract that allows him to sell his asset at time T at a guaranteed price K fixed at time t. This contract is called a put option with strike price K and exercise date T.

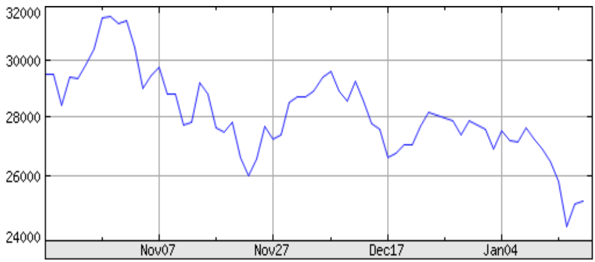


Fig. 5.2: Graph of the Hang Seng index - holding a put option might be useful here.

Definition 5.1. *A (European)* put *option is a contract that gives its holder the right (but not the obligation) to* sell *a quantity of assets at a predefined price K called the strike price (or exercise price) and at a predefined date T called the maturity.*

In case the price S_T falls down below the level K, exercising the contract will give the holder of the option a gain equal to $K - S_T$ in comparison to those who did not subscribe the option contract and have to sell the asset at the market price S_T. In turn, the issuer of the option contract will register a loss also equal to $K - S_T$ (in the absence of transaction costs and other fees).

If S_T is above K, then the holder of the option contract will not exercise the option as he may choose to sell at the price S_T. In this case, the profit derived from the option contract is 0.

Two possible scenarios (S_T finishing above K or below K) are illustrated in Figure 5.3.

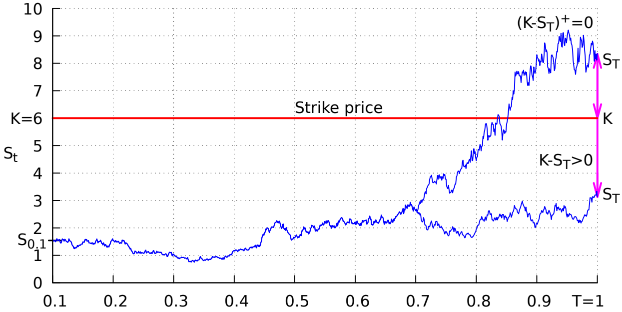


Fig. 5.3: Two put option scenarios.

In general, the payoff of a (so called *European*) put option contract can be written as

$$\phi(S_T) = (K - S_T)^+ := \begin{cases} K - S_T & \text{if } S_T \leqslant K, \\ 0, & \text{if } S_T \geqslant K. \end{cases}$$

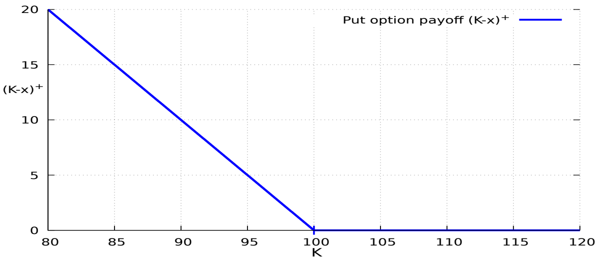


Fig. 5.4: Payoff function $x \mapsto (K - x)^+$ of a put option with strike price $K = 100$.

See *e.g.* https://optioncreator.com/stwwxvz.

Cash settlement *vs.* physical delivery

Physical delivery. In the case of physical delivery, the put option contract issuer will pay the strike price $\$K$ to the option contract holder in exchange for one unit of the risky asset priced S_T.

Cash settlement. In the case of a cash settlement, the put option issuer will satisfy the contract by transferring the amount $C = (K - S_T)^+$ to the option contract holder.

Examples of put options: The **buy back guarantee**[*] in currency exchange and the **price drop protection** in online ticket booking are common examples of European put options.

The derivatives market

As of year 2015, the size of the derivatives market was estimated at more that \$1.2 quadrillion,[†] or more than 10 times the Gross World Product (GWP). See **here** or **here** for up-to-date data on notional amounts outstanding and gross market value from the Bank for International Settlements (BIS).

European call option contracts

On the other hand, if the trader aims at buying some stock or commodity, his interest will be in prices not going up and he might want to purchase a call option, which is a contract allowing him to buy the considered asset at time T at a price not higher than a level K fixed at time t.

Definition 5.2. *A (European)* call *option is a contract that gives its holder the right (but not the obligation) to* purchase *a quantity of assets at a predefined price K called the strike price, and at a predefined date T called the maturity.*

Here, in the event that S_T goes above K, the buyer of the option contract will register a potential gain equal to $S_T - K$ in comparison to an agent who did not subscribe to the call option.

Two possible scenarios (S_T finishing above K or below K) are illustrated in Figure 5.5.

[*] Right-click to open or save the attachment.

[†] One thousand trillion, or one million billion, or 10^{15}.

Fig. 5.5: Two call option scenarios.

In general, the payoff of a (so called European) call option contract can be written as

$$\phi(S_T) = (S_T - K)^+ := \begin{cases} S_T - K & \text{if } S_T \geqslant K, \\ \\ 0, & \text{if } S_T \leqslant K. \end{cases}$$

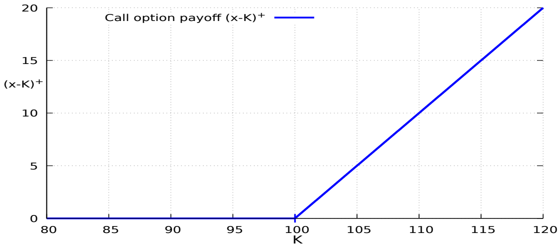


Fig. 5.6: Payoff function $x \mapsto (x - K)^+$ of a call option with strike price $K = 100$.

See *e.g.* https://optioncreator.com/stqhbgn.

Example of a call option: The **price lock guarantee*** is a common example of a European *call* option.

* Right-click to open or save the attachment.

According to market practice, options are often divided into a certain number n of *warrants*, the (possibly fractional) quantity n being called the *entitlement ratio*.

Cash settlement *vs.* physical delivery

Physical delivery. In the case of physical delivery, the call option contract issuer will transfer one unit of the risky asset priced S_T to the option contract holder in exchange for the strike price $\$K$. Physical delivery may include physical goods, commodities or assets such as coffee, airline fuel or live cattle.

Cash settlement. In the case of a cash settlement, the call option issuer will fulfill the contract by transferring the amount $C = (S_T - K)^+$ to the option contract holder.

Option pricing

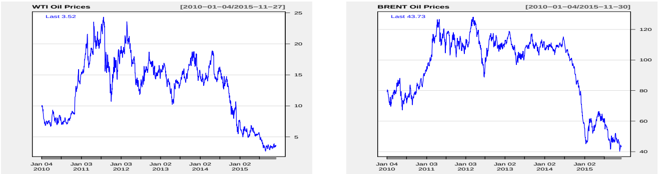
In order for an option contract to be fair, the buyer of the option contract should pay a fee (similar to an insurance fee) at the signature of the contract. The computation of this fee is an important issue, and is known as option *pricing*.

Option hedging

The second important issue is that of *hedging*, *i.e.* how to manage a given portfolio in such a way that it contains the required random payoff $(K - S_T)^+$ (for a put option) or $(S_T - K)^+$ (for a call option) at the maturity date T.

Example: Fuel hedging and the four-way zero-collar option

```
install.packages("Quandl"); library(Quandl); library(quantmod)
getSymbols("DCOILBRENTEU", src="FRED")
chartSeries(DCOILBRENTEU,up.col="blue",theme="white",name = "BRENT Oil
      Prices",lwd=5)
BRENT = Quandl("FRED/DCOILBRENTEU",start_date="2010-01-01",
      end_date="2015-11-30",type="xts")
chartSeries(BRENT,up.col="blue",theme="white",name = "BRENT Oil Prices",lwd=5)
getSymbols("WTI", from="2010-01-01", to="2015-11-30")
WTI <- Ad(`WTI`)
chartSeries(WTI,up.col="blue",theme="white",name = "WTI Oil Prices",lwd=5)
```



(a) WTI price graph. (b) Brent price graph

Fig. 5.7: Brent and WTI price graphs.

(April 2011)

Fuel hedge promises Kenya Airways smooth ride in volatile oil market.*

(November 2015)

A close look at the role of fuel hedging in Kenya Airways $259 million loss.*
The four-way call collar call option requires its holder to purchase the underlying asset (here, airline fuel) at a price specified by the blue curve in Figure 5.8, when the underlying asset price is represented by the red line.

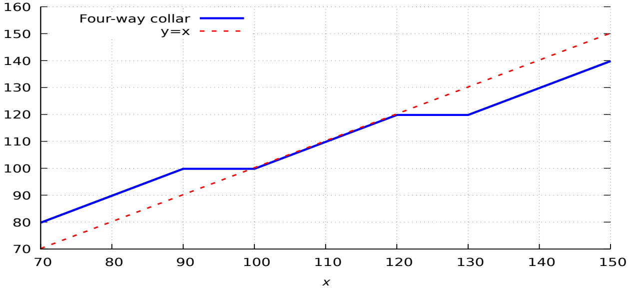


Fig. 5.8: Price map of a four-way collar option.

The four-way call collar option contract will result into a positive or negative payoff depending on current fuel prices, as illustrated in Figure 5.9.

* Right-click to open or save the attachment.

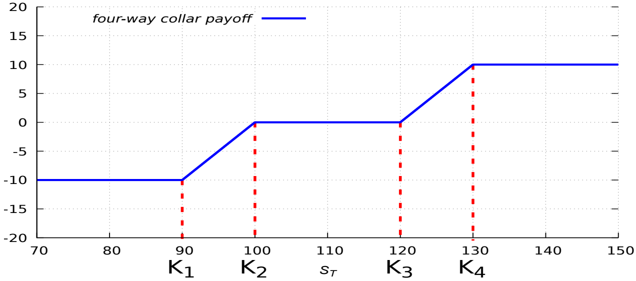


Fig. 5.9: Payoff function of a four-way call collar option.

The four-way call collar payoff can be written as a linear combination

$$\phi(S_T) = (K_1 - S_T)^+ - (K_2 - S_T)^+ + (S_T - K_3)^+ - (S_T - K_4)^+$$

of call and put option payoffs with respective strike prices

$$K_1 = 90, \quad K_2 = 100, \quad K_3 = 120, \quad K_4 = 130,$$

see *e.g.* https://optioncreator.com/st5rf51.

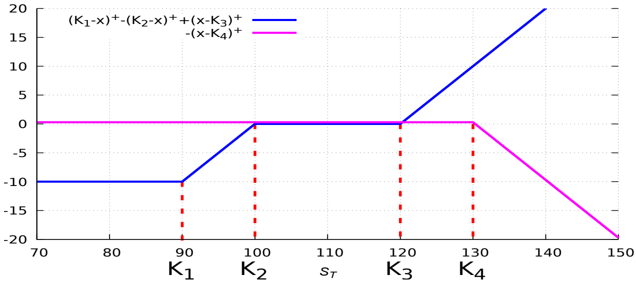


Fig. 5.10: Four-way call collar payoff as a combination of call and put options.*

Therefore, the four-way call collar option contract can be *synthesized* by:

* The animation works in Acrobat Reader on the entire pdf file.

1. purchasing a *put option* with strike price $K_1 = \$90$, and
2. selling (or issuing) a *put option* with strike price $K_2 = \$100$, and
3. purchasing a *call option* with strike price $K_3 = \$120$, and
4. selling (or issuing) a *call option* with strike price $K_4 = \$130$.

Moreover, the call collar option contract can be made *costless* by adjusting the boundaries K_1, K_2, K_3, K_4, in which case it becomes a *zero-collar* option.

Example - The one-step 4-5-2 model

We close this introduction with a simplified example of the pricing and hedging technique in a one-step binary model with two time instants $t = 0$ and $t = 1$. Consider:

i) A risky underlying stock valued $S_0 = \$4$ at time $t = 0$, and taking only two possible values

$$S_1 = \begin{cases} \$5 \\ \\ \$2 \end{cases}$$

at time $t = 1$.

ii) An option contract that promises a claim payoff C whose values are defined contingent to the market data of S_1 as:

$$C := \begin{cases} \$3 & \text{if } S_1 = \$5 \\ \\ \$0 & \text{if } S_1 = \$2. \end{cases}$$

Exercise: Does C represent the payoff of a put option contract? Of a call option contract? If yes, with which strike price K?

Quiz: Using this **online form**, input your own intuitive estimate for the price of the claim C.

At time $t = 0$ the option contract issuer (or writer) chooses to invest α units in the risky asset S, while keeping $\$\beta$ on our bank account, meaning that we invest a total amount

$$\alpha S_0 + \$\beta \qquad\qquad \text{at time} \quad t = 0.$$

Here, the amount $\$\beta$ may be positive or negative, depending on whether it is corresponds to savings or to debt, and is interpreted as a *liability*.

The following issues can be addressed:

a) *Hedging*: How to choose the portfolio allocation $(\alpha, \$\beta)$ so that the value

$$\alpha S_1 + \$\beta$$

of the portfolio matches the future payoff C at time $t = 1$?

b) *Pricing*: How to determine the amount $\alpha S_0 + \$\beta$ to be invested by the option contract issuer in such a portfolio at time $t = 0$?

$$\begin{tikzcd}
 & S_1 = 5 \text{ and } C = 3 \\
S_0 = 4 \arrow[ur] \arrow[dr] & \\
 & S_1 = 2 \text{ and } C = 0
\end{tikzcd}$$

$S_0 = 4$	$S_1 = 5$ and $C = 3$
	$S_1 = 2$ and $C = 0$

Hedging or *replicating* the contract means that at time $t = 1$ the portfolio value matches the future payoff C, *i.e.*

$$\alpha S_1 + \$\beta = C.$$

Hedge, then price. This condition can be rewritten as

$$C = \begin{cases} \$3 = \alpha \times \$5 + \$\beta & \text{if } S_1 = \$5, \\ \\ \$0 = \alpha \times \$2 + \$\beta & \text{if } S_1 = \$2, \end{cases}$$

i.e.

$$\begin{cases} 5\alpha + \beta = 3, \\ \\ 2\alpha + \beta = 0, \end{cases} \quad \text{which yields} \quad \begin{cases} \alpha = 1 \text{ stock}, \\ \\ \$\beta = -\$2. \end{cases}$$

In other words, the option contract issuer purchases 1 (one) unit of the stock S at the price $S_0 = \$4$, and borrows \$2 from the bank. The price of the option contract is then given by the portfolio value

$$\alpha S_0 + \$\beta = 1 \times \$4 - \$2 = \$2.$$

at time $t = 0$.

The above computation is implemented in the attached **IPython notebook***
that can be run **here** or **here**. This algorithm is scalable and can be extended to recombining binary trees over multiple time steps.

---

* Right-click to save as attachment (may not work on).

Definition 5.3. *The* arbitrage-free price *of the option contract is defined as the initial cost $\alpha S_0 + \$\beta$ of the portfolio hedging the claim payoff C.*

Conclusion: in order to deliver the random payoff $C = \begin{cases} \$3 & \text{if } S_1 = \$5 \\ \\ \$0 & \text{if } S_1 = \$2. \end{cases}$ to the option contract holder at time $t = 1$, the option contract issuer (or writer) will:

1. charge $\alpha S_0 + \$\beta = \$2$ (the option contract price) at time $t = 0$,
2. borrow $-\$\beta = \$2$ from the bank,
3. invest those $\$2 + \$2 = \$4$ into the purchase of $\alpha = 1$ unit of stock valued at $S_0 = \$4$ at time $t = 0$,
4. wait until time $t = 1$ to find that the portfolio value has evolved into

$$C = \begin{cases} \alpha \times \$5 + \$\beta = 1 \times \$5 - \$2 = \$3 & \text{if } S_1 = \$5, \\ \\ \alpha \times \$2 + \$\beta = 1 \times \$2 - \$2 = 0 & \text{if } S_1 = \$2, \end{cases}$$

   so that the option contract and the equality $C = \alpha S_1 + \$\beta$ can be fulfilled, allowing the option issuer to break even whatever the evolution of the risky asset price S.

   In a *cash settlement*, the stock is sold at the price $S_1 = \$5$ or $S_1 = \$2$, the payoff $C = (S_1 - K)^+ = \$3$ or $\$0$ is issued to the option contract holder, and the loan is refunded with the remaining $2.

   In the case of *physical delivery* , $\alpha = 1$ share of stock is handed in to the option holder in exchange for the strike price $K = \$2$ which is used to refund the initial $2 loan subscribed by the issuer.

Here, the option contract price $\alpha S_0 + \$\beta = \$2$ is interpreted as the cost of hedging the option. We will see that this model is scalable and extends to discrete time.

We note that the initial option contract price of $2 can be turned to $C = \$3$ (%50 profit) ... or into $C = \$0$ (total ruin).

Thinking further

1) The expected claim payoff at time $t = 1$ is

$$\begin{aligned} \mathbb{E}[C] &= \$3 \times \mathbb{P}(C = \$3) + \$0 \times \mathbb{P}(C = \$0) \\ &= \$3 \times \mathbb{P}(S_1 = \$5). \end{aligned}$$

In absence of arbitrage opportunities ("fair market"), this expected payoff $\mathbb{E}[C]$ should equal the initial amount \$2 invested in the option. In that case we should have

$$\begin{cases} \mathbb{E}[C] = \$3 \times \mathbb{P}(S_1 = \$5) = \$2 \\ \\ \mathbb{P}(S_1 = \$5) + \mathbb{P}(S_1 = \$2) = 1. \end{cases}$$

from which we can *infer* the probabilities

$$\begin{cases} \mathbb{P}(S_1 = \$5) = \dfrac{2}{3} \\ \\ \mathbb{P}(S_1 = \$2) = \dfrac{1}{3}, \end{cases} \tag{5.1}$$

which are called *risk-neutral* probabilities. We see that under the risk-neutral probabilities, the stock S has twice more chances to go up than to go down in a "fair" market.

2) Based on the probabilities (5.1) we can also compute the expected value $\mathbb{E}[S_1]$ of the stock at time $t = 1$. We find

$$\begin{aligned} \mathbb{E}[S_1] &= \$5 \times \mathbb{P}(S_1 = \$5) + \$2 \times \mathbb{P}(S_1 = \$2) \\ &= \$5 \times \frac{2}{3} + \$2 \times \frac{1}{3} \\ &= \$4 \\ &= S_0. \end{aligned}$$

Here this means that, on average, no extra profit or loss can be made from an investment on the risky stock, hence the term "risk-neutral". In a more realistic model we can assume that the riskless bank account yields an interest rate equal to r, in which case the above analysis is modified by letting $\$\beta$ become $\$(1+r)\beta$ at time $t = 1$, nevertheless the main conclusions remain unchanged.

Market-implied probabilities

By matching the theoretical price $\mathbb{E}[C]$ to an actual market price data $\$M$ as

$$\$M = \mathbb{E}[C] = \$3 \times \mathbb{P}(C = \$3) + \$0 \times \mathbb{P}(C = \$0) = \$3 \times \mathbb{P}(S_1 = \$5)$$

we can infer the probabilities

$$
\begin{cases}
\mathbb{P}(S_1 = \$5) = \dfrac{\$M}{3} \\
\\
\mathbb{P}(S_1 = \$2) = \dfrac{3 - \$M}{3},
\end{cases}
\tag{5.2}
$$

which are *implied probabilities* estimated from market data, as illustrated in Figure 5.11. We note that the conditions

$$
0 < \mathbb{P}(S_1 = \$5) < 1, \qquad 0 < \mathbb{P}(S_1 = \$2) < 1
$$

are equivalent to $0 < \$M < 3$, which is consistent with financial intuition in a non-deterministic market. Figure 5.11 shows the time evolution of probabilities $p(t)$, $q(t)$ of two opposite outcomes.

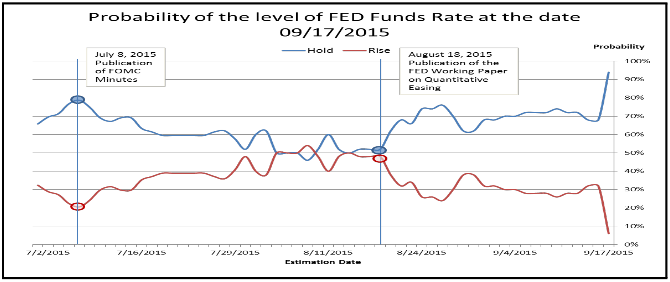


Fig. 5.11: Implied probabilities.

The *Practitioner* expects a good model to be:

- *Robust* with respect to missing, spurious or noisy data,
- *Fast* - prices have to be delivered daily in the morning,
- *Easy* to calibrate - parameter estimation,
- *Stable* with respect to re-calibration and the use of new data sets.

Typically, a medium size bank manages 5,000 options and 10,000 deals daily over 1,000 possible scenarios and dozens of time steps. This can mean a hundred million computations of $\mathbb{E}[C]$ daily, or close to a billion such computations for a large bank.

The *Mathematician* tends to focus on more theoretical features, such as:

- *Elegance*,
- *Sophistication*,

- *Existence* of analytical (closed-form) solutions / error bounds,
- *Significance* to mathematical finance.

This includes:

- *Creating* new payoff functions and structured products,
- *Defining* new models for underlying asset prices,
- *Finding* new ways to compute expectations $\mathbb{E}[C]$ and hedging strategies.

The methods involved include:

- Monte Carlo (60%),

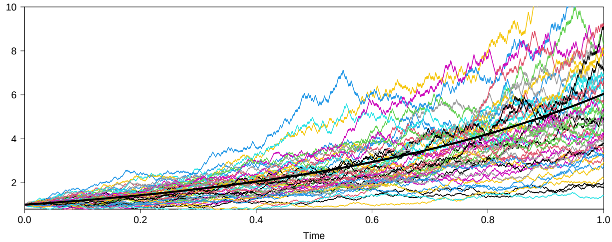


Fig. 5.12: One hundred sample price paths used for the Monte Carlo method.

- PDEs and finite differences (30%),
- Other analytic methods and approximations (10%),

+ AI and Machine Learning techniques.

5.3 Black-Scholes Analysis

In this section we consider an asset price process $(S_t)_{t\in\mathbb{R}_+}$ modeled as the geometric Brownian motion (1.10) of Proposition 1.7. Recall that the evolution of the riskless bank account value $(A_t)_{t\in\mathbb{R}_+}$ is constructed from standard returns, defined from the differential equation

$$\frac{dA_t}{dt} = rA_t, \quad t \geqslant 0,$$

with solution

$$A_t = A_0 \mathrm{e}^{rt}, \qquad t \geqslant 0,$$

where $r > 0$ is the risk-free interest rate. The risky asset price process $(S_t)_{t\in\mathbb{R}_+}$ is modeled from the equation

$$dS_t = rS_t dt + \sigma S_t dB_t$$

with solution

$$S_t = S_0 \exp\left(\sigma B_t + \left(r - \frac{1}{2}\sigma^2\right)t\right), \qquad t \geqslant 0.$$

Let α_t and β_t be the numbers of units invested at time $t \geqslant 0$, respectively in the assets priced $(S_t)_{t\in\mathbb{R}_+}$ and $(A_t)_{t\in\mathbb{R}_+}$.

In the sequel, we will consider a portfolio whose value V_t at time $t \geqslant 0$ is given by

$$V_t = \beta_t A_t + \alpha_t S_t, \qquad t \geqslant 0.$$

Black-Scholes formula for European call options

In the case of European call options with payoff function $\phi(x) = (x - K)^+$ we have the following Black-Scholes formula.

Proposition 5.4. *The price at time $t \in [0, T]$ of the European call option with strike price K and maturity T is given by*

$$\begin{aligned}\mathrm{BS}_c(S_t, K, r, T-t, \sigma) &= \mathrm{e}^{-(T-t)r}\mathbb{E}[(S_T - K)^+ \mid \mathcal{F}_t] \\ &= S_t \Phi\big(d_+(T-t)\big) - K\mathrm{e}^{-(T-t)r}\Phi\big(d_-(T-t)\big),\end{aligned} \tag{5.3}$$

$0 \leqslant t \leqslant T$, *with*

$$\begin{cases} d_+(T-t) := \dfrac{\log(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, & \text{(5.4a)} \\[2ex] d_-(T-t) := \dfrac{\log(S_t/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad 0 \leqslant t < T, & \text{(5.4b)} \end{cases}$$

where "log" denotes the natural logarithm "ln" *and*

$$\Phi(x) := \mathbb{P}(\mathcal{N} \leqslant x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-y^2/2} dy, \qquad x \in \mathbb{R},$$

is the standard Gaussian Cumulative Distribution Function (CDF).

We note the relation

$$d_+(T-t) = d_-(T-t) + |\sigma|\sqrt{T-t}, \qquad 0 \leqslant t < T. \tag{5.5}$$

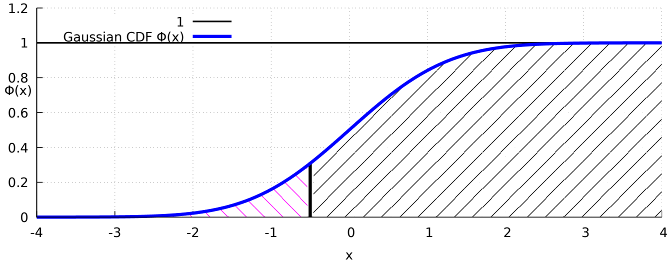


Fig. 5.13: Graph of the Gaussian Cumulative Distribution Function (CDF).

In other words, the European call option with strike price K and maturity T is priced at time $t \in [0, T]$ as

$$\mathrm{BS}_c(S_t, K, r, T-t, \sigma) = S_t \Phi\big(d_+(T-t)\big) - K\mathrm{e}^{-(T-t)r}\Phi\big(d_-(T-t)\big),$$

$0 \leqslant t \leqslant T$. The following R script implements the Black-Scholes formula for European call options in R.*

```
BSCall <- function(S, K, r, T, sigma)
{d1 <- (log(S/K)+(r+sigma^2/2)*T)/(sigma*sqrt(T));d2 <- d1 - sigma * sqrt(T)
BSCall = S*pnorm(d1) - K*exp(-r*T)*pnorm(d2)
BSCall}
```

In comparison with the discrete-time Cox-Ross-Rubinstein (CRR) model, the interest in the Black-Scholes formula is to provide an analytical solution that can be evaluated in a single step, which is computationally much more efficient.

---

* Download the corresponding **IPython notebook** that can be run **here** or **here**.

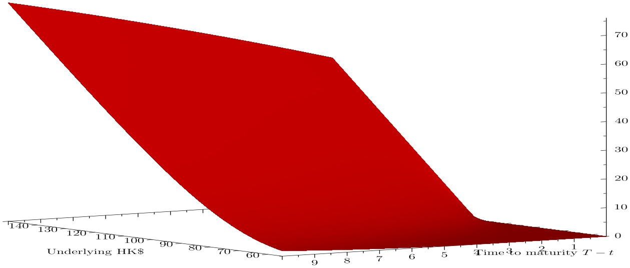


Fig. 5.14: Graph of the Black-Scholes call price map with strike price $K = 100$.*

Figure 5.14 presents an interactive graph of the Black-Scholes call price map, *i.e.* of the function

$$(t, x) \longmapsto \mathrm{BS}_c(x, K, r, T-t, \sigma) = x\Phi\big(d_+(T-t)\big) - K\mathrm{e}^{-(T-t)r}\Phi\big(d_-(T-t)\big).$$

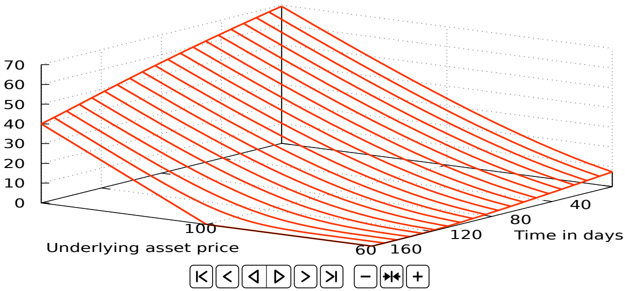


Fig. 5.15: Time-dependent solution of the Black-Scholes PDE (call option).†

The next proposition is proved by a direct differentiation of the Black-Scholes function.

Proposition 5.5. *The Black-Scholes* Delta *of the European call option is given by*

$$\alpha_t = \alpha_t(S_t) = \frac{\partial}{\partial x}\mathrm{BS}_c(x, K, r, T-t, \sigma)_{|x=S_t} = \Phi\big(d_+(T-t)\big) \in [0,1], \quad (5.6)$$

* Right-click on the figure for interaction and "Full Screen Multimedia" view.

† The animation works in Acrobat Reader on the entire pdf file.

where $d_+(T-t)$ *is defined in* (5.4a).

We note that the Black-Scholes call price splits into a risky component $S_t\Phi\big(d_+(T-t)\big)$ and a riskless component $-K\mathrm{e}^{-(T-t)r}\Phi\big(d_-(T-t)\big)$, as follows:

$$\mathrm{BS}_c(S_t,K,r,T-t,\sigma) = \underbrace{S_t\Phi\big(d_+(T-t)\big)}_{\text{Risky investment (held)}} - \underbrace{K\mathrm{e}^{-(T-t)r}\Phi\big(d_-(T-t)\big)}_{\text{Risk free investment (borrowed)}}, \tag{5.7}$$

$0 \leqslant t \leqslant T$, *i.e.* $\alpha_t = \Phi\big(d_+(T-t)\big)$ represents the quantity of assets invested in the risky asset priced at S_t. The following R script implements the Black-Scholes Delta for European call options.

```
DeltaCall <- function(S, K, r, T, sigma)
{d1 <- (log(S/K)+(r+sigma^2/2)*T)/(sigma*sqrt(T))
DeltaCall = pnorm(d1);DeltaCall}
```

In Figure 5.16 we plot the Delta of the European call option as a function of the underlying asset price and of the time remaining until maturity.

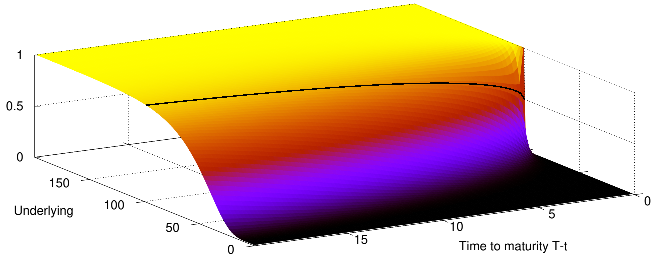


Fig. 5.16: Delta of a European call option with strike price $K = 100$, $r = 3\%$, $\sigma = 10\%$.

Black-Scholes formula for European Put Options

The European put option price is computed in the next proposition.

Proposition 5.6. *The price at time* $t \in [0,T]$ *of the European put option with strike price* K *and maturity* T *is given by*

$$\begin{aligned}\mathrm{BS}_p(S_t,K,r,T-t,\sigma) &= \mathrm{e}^{-(T-t)r}\mathbb{E}[(K-S_T)^+ \mid \mathcal{F}_t] \\ &= K\mathrm{e}^{-(T-t)r}\Phi\big(-d_-(T-t)\big) - S_t\Phi\big(-d_+(T-t)\big),\end{aligned}$$

$0 \leqslant t \leqslant T$, *where* $d_+(T-t)$ *and* $d_-(T-t)$ *are defined in* (5.4a)-(5.4b).

The Black-Scholes formula for European Put Options is plotted in illustrated in Figure 5.17.

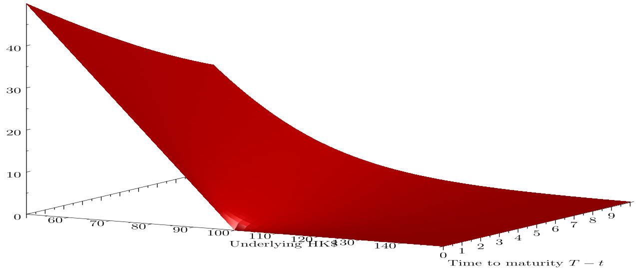


Fig. 5.17: Graph of the Black-Scholes put price function with strike price $K = 100$.*

In other words, the European put option with strike price K and maturity T is priced at time $t \in [0, T]$ as

$$Bl_p(S_t, K, r, T-t, \sigma) = K\mathrm{e}^{-(T-t)r}\Phi\big(-d_-(T-t)\big) - S_t\Phi\big(-d_+(T-t)\big),$$

$0 \leqslant t \leqslant T$.

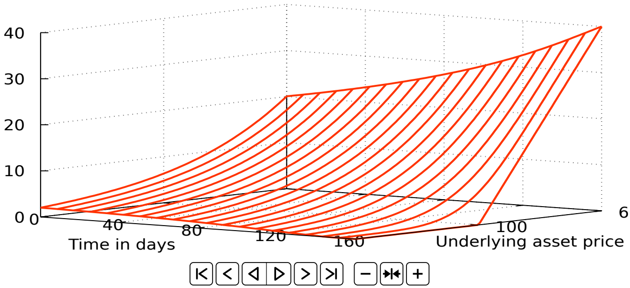


Fig. 5.18: Time-dependent solution of the Black-Scholes PDE (put option).†

The following R script implements the Black-Scholes formula for European put options in R.

* Right-click on the figure for interaction and "Full Screen Multimedia" view.

† The animation works in Acrobat Reader on the entire pdf file.

```
BSPut <- function(S, K, r, T, sigma)
{d1 = (log(S/K)+(r+sigma^2/2)*T)/(sigma*sqrt(T));d2 = d1 - sigma * sqrt(T);
BSPut = K*exp(-r*T) * pnorm(-d2) - S*pnorm(-d1);BSPut}
```

The Black-Scholes *Delta* of the European put option is computed in the following proposition.

Proposition 5.7. *The Black-Scholes* Delta *of the European put option is given by*

$$\boxed{\begin{aligned}\alpha_t = \alpha_t(S_t) &= \frac{\partial}{\partial x}\mathrm{BS}_p(x,K,r,T-t,\sigma)_{|x=S_t}\\ &= -\big(1-\Phi\big(d_+(T-t)\big)\big) = -\Phi\big(-d_+(T-t)\big)\in[-1,0],\quad 0\leqslant t\leqslant T,\end{aligned}}$$

where $d_+(T-t)$ is defined in (5.4a).

We note that the Black-Scholes put price splits into a risky component $-S_t\Phi\big(-d_+(T-t)\big)$ and a riskless component $K\mathrm{e}^{-(T-t)r}\Phi\big(-d_-(T-t)\big)$, as follows:

$$\mathrm{BS}_p(S_t,K,r,T-t,\sigma) = \underbrace{K\mathrm{e}^{-(T-t)r}\Phi\big(-d_-(T-t)\big)}_{\text{Risk}-\text{free investment (savings)}} - \underbrace{S_t\Phi\big(-d_+(T-t)\big)}_{\text{Risky investment (short)}}, \tag{5.8}$$

$0\leqslant t\leqslant T$, *i.e.* $-\Phi\big(-d_+(T-t)\big)$ represents the quantity of assets invested in the risky asset priced at S_t.

```
DeltaPut <- function(S, K, r, T, sigma)
{d1 <- (log(S/K)+(r+sigma^2/2)*T)/(sigma*sqrt(T)); DeltaPut = -pnorm(-d1);DeltaPut}
```

In Figure 5.19 we plot the Delta of the European put option as a function of the underlying asset price and of the time remaining until maturity.

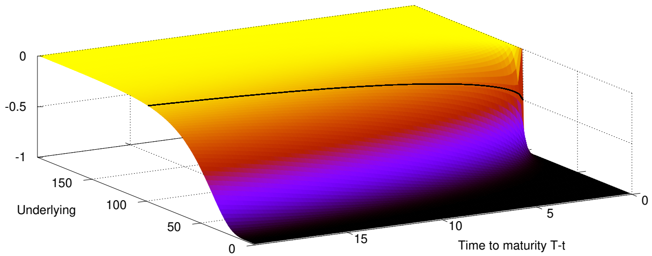


Fig. 5.19: Delta of a European put option with strike price $K = 100$, $r = 3\%$, $\sigma = 10\%$.

Exercises

Exercise 5.1 Consider a risky asset valued $S_0 = \$3$ at time $t = 0$ and taking only two possible values $S_1 \in \{\$1, \$5\}$ at time $t = 1$, and a financial claim given at time $t = 1$ by

$$C := \begin{cases} \$0 & \text{if } S_1 = \$5 \\ \\ \$2 & \text{if } S_1 = \$1. \end{cases}$$

Is C the payoff of a call option or of a put option? Give the strike price of the option.

Exercise 5.2 Consider a risky asset valued $S_0 = \$4$ at time $t = 0$, and taking only two possible values $S_1 \in \{\$2, \$5\}$ at time $t = 1$. Compute the initial value $V_0 = \alpha S_0 + \$\beta$ of the portfolio hedging the claim payoff

$$C = \begin{cases} \$0 & \text{if } S_1 = \$5 \\ \\ \$6 & \text{if } S_1 = \$2 \end{cases}$$

at time $t = 1$, and find the corresponding risk-neutral probability measure $\mathbb{P}^*$.

Exercise 5.3 Consider a risky asset valued $S_0 = \$4$ at time $t = 0$, and taking only two possible values $S_1 \in \{\$5, \$2\}$ at time $t = 1$, and the claim payoff

$$C = \begin{cases} \$3 & \text{if } S_1 = \$5 \\ \\ \$0 & \text{if } S_1 = \$2. \end{cases} \qquad \text{at time } t = 1.$$

We assume that the issuer charges \$1 for the option contract at time $t = 0$.

a) Compute the portfolio allocation (α, β) made of α stocks and $\$\beta$ in cash, so that:

   i) the full \$1 option price is invested into the portfolio at time $t = 0$,

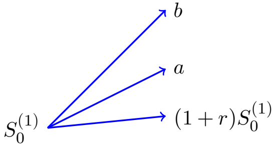
   and

   ii) the portfolio reaches the $C = \$3$ target if $S_1 = \$5$ at time $t = 1$.

b) Compute the loss incurred by the option issuer if $S_1 = \$2$ at time $t = 1$.

Exercise 5.4 Recall that an *arbitrage opportunity* consist of a portfolio allocation with zero or negative cost that can yield a nonnegative and possibly strictly positive payoff at maturity.

a) Consider the following market model:



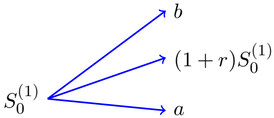
i) Does this model allow for arbitrage opportunities?

Yes | No |

ii) If this model allows for arbitrage opportunities, how can they be realized?

By shortselling | By borrowing on savings | N.A. |

b) Consider the following market model:



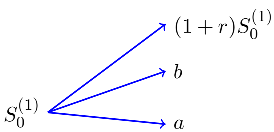
i) Does this model allow for arbitrage opportunities?

Yes | No |

ii) If this model allows for arbitrage opportunities, how can they be realized?

By shortselling | By borrowing on savings | N.A. |

c) Consider the following market model:



i) Does this model allow for arbitrage opportunities?

Yes |	No |

ii) If this model allows for arbitrage opportunities, how can they be realized?

By shortselling |	By borrowing on savings |	N.A. |

Exercise 5.5 In a market model with two time instants $t = 0$ and $t = 1$ and risk-free interest rate r, consider:

- a riskless asset valued $S_0^{(0)}$ at time $t = 0$, and value $S_1^{(0)} = (1 + r)S_0^{(0)}$ at time $t = 1$.

- a risky asset with price $S_0^{(1)}$ at time $t = 0$, and three possible values at time $t = 1$, with $a < b < c$, *i.e.*:

$$S_1^{(1)} = \begin{cases} S_0^{(1)}(1+a), \\ \\ S_0^{(1)}(1+b), \\ \\ S_0^{(1)}(1+c). \end{cases}$$

In general, is it possible to hedge (or replicate) a claim with three distinct claim payoff values C_a, C_b, C_c in this market?

Exercise 5.6 Superhedging risk measure. Consider a stock valued S_0 at time $t = 0$, and taking only two possible values $S_1 = \underline{S}_1$ or $S_1 = \overline{S}_1$ at time $t = 1$, with $\underline{S}_1 < \overline{S}_1$.

a) Compute the initial portfolio allocation (α, β) of a portfolio made of α units of stock and $\$\beta$ in cash, hedging the call option with strike price $K \in [\underline{S}_1, \overline{S}_1]$ and claim payoff

$$C = (S_1 - K)^+ = \begin{cases} \overline{S}_1 - K \text{ if } S_1 = \overline{S}_1 \\ \\ \$0 \qquad\quad \text{if } S_1 = \underline{S}_1, \end{cases} \quad \text{at time } t = 1.$$

b) Show that the risky asset allocation α satisfies the condition $\alpha \in [0, 1]$.

c) Compute the Superhedging Risk Measure SRM_C* of the claim $C = (S_1 - K)^+$.

Exercise 5.7 Given two strike prices $K_1 < K_2$, we consider a long box spread option, realized as the combination of four legs with same maturity date:

- One *long* call with strike price K_1 and payoff function $(x - K_1)^+$,
- One *short* put with strike price K_1 and payoff function $-(K_1 - x)^+$,
- One *short* call with strike price K_2 and payoff function $-(x - K_2)^+$,
- One *long* put with strike price K_2 and payoff function $(K_2 - x)^+$.

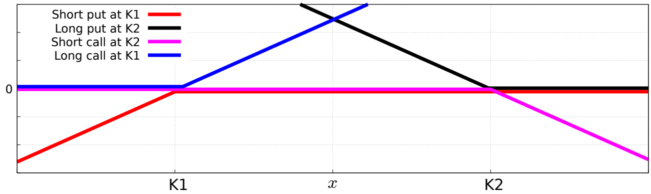


Fig. 5.20: Graphs of call/put payoff functions.

a) Find the payoff of the long box spread option in terms of K_1 and K_2.
b) From Table 5.1, find a choice of strike prices $K_1 < K_2$ that can be used to build a long box spread option on the Hang Seng Index (HSI).
c) Using Table 5.1, price the option built in part (b) in index points, and then in HK$.

   Hints.

   i) The closing prices in Table 5.1 are warrant prices quoted in index points.
   ii) Warrant prices are converted to option prices by multiplication by the number given in the "Entitlement Ratio" column.
   iii) The conversion rate from index points to HK$ is HK$50 per index point.

d) Would you buy the option priced in part (c) ?

* "*The smallest amount necessary to be paid for a portfolio at time $t = 0$ so that the value of this portfolio at time $t = 1$ is at least as great as C*".

DERIVATIVE WARRANT SEARCH

Link to Relevant Exchange Traded Options

Updated: 19 December 2022

Basic Data										Market Data								
DW Code	Issuer	UL	Call/ Put	DW Type	Listing (D-M-Y)	Maturity (D-M-Y)	Strike Currency	Strike	Entitle-ment Ratio^	Total Issue Size	O/S (%)	Delta (%)	IV. (%)	Trading Currency	Day High	Day Low	Closing Price	T/O ('000)
18606	SG	HSI	Put	Standard	22-11-2022	28-03-2023	HKD	25088	10000	300,000,000	8.01	(0.002)	30.968	HKD	0.054	0.042	0.053	459
19399	HT	HSI	Put	Standard	01-12-2021	28-03-2023	HKD	25200	10000	400,000,000	0.06	(0.002)	32.190	HKD	0.000	0.000	0.061	0
19485	BI	HSI	Put	Standard	02-12-2021	28-03-2023	HKD	25200	10000	150,000,000	21.41	(0.002)	28.154	HKD	0.044	0.037	0.044	59
22857	VT	HSI	Put	Standard	26-02-2021	28-03-2023	HKD	25000	8000	80,000,000	22.45	(0.002)	30.905	HKD	0.065	0.043	0.064	1,165
26601	BI	HSI	Call	Standard	27-12-2021	28-03-2023	HKD	25200	11000	150,000,000	0.00	0.018	25.347	HKD	0.390	0.360	0.370	84
27489	BP	HSI	Call	Standard	17-09-2021	28-03-2023	HKD	25000	7500	80,000,000	2.95	0.009	28.392	HKD	0.590	0.540	0.540	6
28231	HS	HSI	Call	Standard	29-09-2021	28-03-2023	HKD	25118	7500	200,000,000	0.00	0.012	24.897	HKD	0.000	0.000	0.570	0

^ The entitlement ratio in general represents the number of derivative warrants required to be exercised into one share or one unit of the underlying asset (subject to any adjustments as may be necessary to reflect any capitalization, rights issue, distribution or the like).

Delayed data on Delta and Implied Volatility of Derivative Warrants are provided by Reuters.
Users should not use such data provided by Reuters for commercial purposes without its prior written consent.

For underlying stock price, please refer to Securities Prices of Market Data.

Table 5.1: Call and put options on the Hang Seng Index (HSI).

Chapter 6
Value at Risk

Value at risk (VaR) is probably the most basic and widely used measure of risk. It relies on estimating the amount that can potentially be lost on a given investment within a certain time range. This chapter starts with a review the concept of risk measure in general, including quantile risk measures, before providing a mathematical treatment of Value at Risk, together with experiments based on actual financial data sets.

6.1 Risk Measures

Risk measures have two objectives:

i) to provide a measure for risk, and

ii) to determine an adequate level of capital reserves that matches the current level of risk.

In what follows, the potential losses associated to a given risk will be modeled by the values of a random variable X.

Definition 6.1. *A* risk measure *is a mapping that assigns a value V_X to a given loss random variable X.*

For insurance companies, which need to hold a capital in order to meet future liabilities, the capital C_X required to face the risk induced by a potential loss $X \geqslant 0$ can be defined as

$$C_X := V_X - L_X, \tag{6.1}$$

where

a) V_X stands for an upper "reasonable" estimate of the potential loss associated to X.

b) L_X represents the *liabilities* of the company.

   In other words, managing risk means here determining a level V_X of provision or capital requirement that will not be "too much" exceeded by X. When $L_X < 0$ the amount $-L_x > 0$ corresponds to a debt owed by the company, while $L_X > 0$ corresponds to positive liabilities such as deferred revenue or to a debt owed to the company.

Some examples of risk measures (Hardy (2006))

a) The *expected value premium principle* is the risk measure defined by

$$V_X := \mathbb{E}[X] + \alpha\mathbb{E}[X]$$

   for some $\alpha \geqslant 0$. For $\alpha = 0$, $V_X := \mathbb{E}[X]$ it is called the *pure premium* risk measure.

b) The *standard deviation premium principle* is the risk measure defined by

$$V_X := \mathbb{E}[X] + \alpha\sqrt{\mathrm{Var}[X]}$$

   for some $\alpha \geqslant 0$, where $\mathrm{Var}[X]$ denotes the variance of X.

In order to proceed with more examples of risk measures, we will need to use conditional expectations, see *e.g.* Lemma A.15 for the following proposition. The what follows, we let $\mathbb{1}_A$ denote the *indicator function* of any event A subset of Ω, defined as

$$\mathbb{1}_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A, \\ \\ 0 \text{ if } \omega \notin A. \end{cases}$$

Lemma 6.2. *Let A be an event such that $\mathbb{P}(A) > 0$. The* conditional expectation *of $X : \Omega \longrightarrow \mathbb{N}$ given the event A satisfies the relation*

$$\mathbb{E}[X \mid A] := \frac{1}{\mathbb{P}(A)}\mathbb{E}[X\mathbb{1}_A].$$

For example, consider the sample space $\Omega = \{1, 3, -1, -2, 5, 7\}$ with the non-uniform probability measure given by

$$\mathbb{P}(\{-1\}) = \mathbb{P}(\{-2\}) = \mathbb{P}(\{1\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{7\}) = \frac{1}{7}, \ \mathbb{P}(\{5\}) = \frac{2}{7},$$

and the random variable

$$X : \Omega \longrightarrow \mathbb{Z}$$

given by

$$X(k) = k, \qquad k = 1, 3, -1, -2, 5, 7.$$

Here, $\mathbb{E}[X \mid X > 1]$ denotes the expected value of X given the event

$$A := \{X > 1\} = \{3, 5, 7\} \subset \Omega,$$

i.e. the mean value of X given that X is strictly greater than one. This conditional expectation can be computed as

$$\begin{aligned}
&\mathbb{E}[X \mid X > 1] \\
&= 3 \times \mathbb{P}(X = 3 \mid X > 1) + 5 \times \mathbb{P}(X = 5 \mid X > 1) + 7 \times \mathbb{P}(X = 7 \mid X > 1) \\
&= 3 \times \frac{1}{4} + 5 \times \frac{2}{4} + 7 \times \frac{1}{4} \\
&= \frac{3 + 2 \times 5 + 7}{4}.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\frac{1}{\mathbb{P}(X > 1)} \mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big] \\
&= \frac{1}{\mathbb{P}(X > 1)} \left(3 \times \mathbb{P}(X = 3) + 5 \times \mathbb{P}(X = 5) + 7 \times \mathbb{P}(X = 7)\right) \\
&= \frac{1}{4/7} \left(3 \times \frac{1}{7} + 5 \times \frac{2}{7} + 7 \times \frac{1}{7}\right) \\
&= \frac{3 + 5 \times 2 + 7}{4},
\end{aligned}$$

where $\mathbb{P}(X > 1) = 4/7$ and the truncated expectation $\mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big]$ is given by

$$\mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big] = \frac{3 + 2 \times 5 + 7}{7}.$$

c) The *Conditional Tail Expectation* (CTE) of X given that $X > 0$ is the risk measure defined as the conditional mean

$$V^X := \mathbb{E}[X \mid X > 0] = \frac{\mathbb{E}[X \mathbb{1}_{\{X>0\}}]}{\mathbb{P}(X > 0)}. \qquad (6.2)$$

Next, we consider the following market returns data.

```
library(quantmod)
getSymbols("^HSI",from="2013-06-01",to="2014-10-01",src="yahoo")
stock<-Ad(`HSI`);returns <- as.vector((stock-lag(stock))/lag(stock));
times=index(stock);m=mean(returns[returns<0],na.rm=TRUE)
dev.new(width=16,height=7);par(oma=c(0,1,0,0))
plot(times,returns,pch=19,cex=0.4,col="blue", ylab="", xlab="", main = '', las=1, cex.lab=1.8,
      cex.axis=1.8, lwd=3)
segments(x0 = times, x1 = times, y0 = 0, y1 = returns,col="blue")
abline(h=m,col="red",lwd=3); length(returns)
```

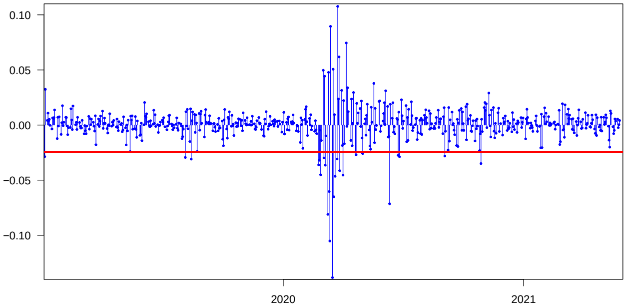


Fig. 6.1: Estimating liabilities by the conditional mean $\mathbb{E}[X \mid X < 0]$ over 346 market returns.

The conditional tail expectation (CTE) (6.2) estimated in Figure 6.1 can also be computed using the next R code, which also implements the statement of Lemma 6.2.

```
returns <- returns[!is.na(returns)]
condmean<-mean(returns[returns<0])
n <- length(returns); sum<-sum(returns[returns<0])
proportion<-length(returns[returns<0])/length(returns)
condmean; sum/proportion/n
condmean<-mean(returns[returns<(-0.025)])
n <- length(returns); sum<-sum(returns[returns<(-0.025)])
proportion<-length(returns[returns<(-0.025)])/length(returns)
condmean; sum/proportion/n
```

Coherent risk measures

Definition 6.3. *A risk measure V is said to be* coherent *if it satisfies the following four properties, for any two random variables X, Y:*

i) Monotonicity:

$$X \leqslant Y \Longrightarrow V_X \leqslant V_Y,$$

ii) (Positive) homogeneity:

$$V_{\lambda X} = \lambda V_X, \qquad \textit{for constant } \lambda > 0,$$

iii) Translation invariance:

$$V_{\mu+X} = \mu + V_X, \qquad \textit{for constant } \mu > 0,$$

iv) Subadditivity:

$$V_{X+Y} \leqslant V_X + V_Y.$$

Subadditivity means that the combined risk of several portfolios is lower than the sum of risks of those portfolios, as happens usually through *portfolio diversification*. For example, one person traveling might insure the unlikely loss of her phone for $V_X = \$100$. However, two people traveling together might want to insure the phone loss event at a level V_{X+Y} lower than $V_X + V_Y = \$100 + \$100$ as the simultaneous loss of both phones during a same trip seems even more unlikely.

The concept of subadditivity is common in most pricing engines, as shown in the following example:

Price(🍟🍔🥤) ⩽ Price(🍟) + Price(🍔) + Price(🥤).

The *expectation* of random variables

$$V_X := \mathbb{E}[X],$$

or *pure premium* risk measure, is an example of a coherent (and additive) risk measure satisfying the above conditions (i)-(iv).

Definition 6.4. *A* distortion risk measure *is a risk measure of the form*

$$M_X = \mathbb{E}[X f_X(X)],$$

where f_X is a distortion function, *i.e. a nonnegative, non-decreasing function such that*

i) $f_{\mu+X}(\mu + x) = f_X(x)$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$,

ii) $f_{\lambda X}(\lambda x) = f_X(x)$, $x \in \mathbb{R}$, $\lambda > 0$,

iii) $\mathbb{E}[f_X(X)] = 1$.

We note that distortion risk measures are positive homogeneous and translation invariant.

i) Positive homogeneity. For any $\lambda > 0$, we have

$$\begin{aligned} M_{\lambda X} = \mathbb{E}[\lambda X f_{\lambda X}(\lambda X)] &= \mathbb{E}[\lambda X f_X(X)] \\ &= \lambda \mathbb{E}[X f_X(X)] \\ &= \lambda M_X. \end{aligned}$$

ii) Translation invariance. For any $\mu \in \mathbb{R}$, we have

$$\begin{aligned} M_{\mu+X} &= \mathbb{E}[(\mu+X) f_{\mu+X}(\mu+X)] \\ &= \mathbb{E}[(\mu+X) f_X(X)] \\ &= \mathbb{E}[X f_X(X)] + \mu \mathbb{E}[f_X(X)] \\ &= \mu + \mathbb{E}[X f_X(X)] \\ &= \mu + M_X. \end{aligned}$$

See (7.2) and (7.7) below for examples of distortion risk measures.

6.2 Quantile Risk Measures

Definition 6.5. *The* Cumulative Distribution Function *(CDF) of a random variable X is the function*

$$F_X : \mathbb{R} \longrightarrow [0,1]$$

defined by

$$F_X(x) := \mathbb{P}(X \leqslant x), \qquad x \geqslant 0.$$

Any cumulative distribution function F_X satisfies the following properties:

i) $x \mapsto F_X(x)$ is non-decreasing,

ii) $x \mapsto F_X(x)$ is right-continuous,

iii) $\lim_{x\to\infty} F_X(x) = 1$,

iv) $\lim_{x\to-\infty} F_X(x) = 0$.

Cumulative distribution functions can be discontinuous functions, as illustrated in Figure 6.2 with

$$\mathbb{P}(X = 0) = \mathbb{P}(X \leqslant 0) - \mathbb{P}(X < 0) = 0.25 > 0.$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

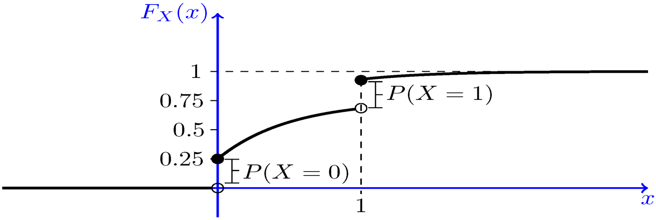


Fig. 6.2: Cumulative distribution function with discontinuities.*

Proposition 6.6 shows in particular that cumulative distribution functions admit left limits.

Proposition 6.6. *For any non-decreasing sequence* $(x_n)_{n\geqslant 1}$ *converging to* $x \in \mathbb{R}$, *we have*

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leqslant x_n) = \mathbb{P}(X < x). \tag{6.3}$$

Proof. By (A.7), we have

$$\begin{aligned}
\mathbb{P}(X < x) &= \mathbb{P}(X \in (-\infty, x)) \\
&= \mathbb{P}\left(X \in \bigcup_{n\geqslant 1} (-\infty, x_n]\right) \\
&= \lim_{n\to\infty} \mathbb{P}(X \in (-\infty, x_n]) \\
&= \lim_{n\to\infty} F_X(x_n).
\end{aligned}$$

□

As a consequence of Proposition 6.6, the gap generated by a discontinuity of a CDF at a point $x \in \mathbb{R}$ is given by

$$F_X(x) - \lim_{y\nearrow x} F_X(y) = \mathbb{P}(X \leqslant x) - \mathbb{P}(X < x) = \mathbb{P}(X = x).$$

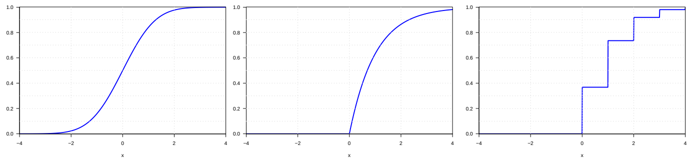
* Picture taken from https://www.probabilitycourse.com/.

```
x <- seq(-4, 4, length=1000)
plot(x, pnorm(x, mean=0, sd=1), type="l", lwd=3, xlab = 'x', ylab = '', main = '', col='blue',
      ylim=c(-0.001,1.002), las=1, cex.lab=1, cex.axis=1, xaxs='i', yaxs='i'); grid(4, 10, lwd =
      2)
plot(x, pexp(x, 1), type="l", lwd=3, xlab = 'x', ylab = '', main = '', col='blue',
      ylim=c(-0.001,1.002), las=1, cex.lab=1, cex.axis=1, xaxs='i', yaxs='i'); grid(4, 10, lwd =
      2)
plot(x, ppois(x, 1), type="l", lwd=3, xlab = 'x', ylab = '', main = '', col='blue',
      ylim=c(-0.001,1.002), las=1, cex.lab=1, cex.axis=1, xaxs='i', yaxs='i'); grid(4, 10, lwd =
      2)
```

Figure 6.3-(a) shows the continuous Cumulative Distribution Function

$$F_X(x) := \mathbb{P}(X \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2} dy, \qquad x \geqslant 0,$$

of a Gaussian random variable $X \simeq \mathcal{N}(0,1)$.



(a) Normal CDF. (b) Exponential CDF. (c) Poisson CDF.

Fig. 6.3: Cumulative distribution functions.

On the other hand, if $F_X(x)$ is differentiable in $x \in \mathbb{R}$ then the distribution of the random variable X is said to admit a *probability density function* (PDF) $f_X(x)$ given as the derivative

$$f_X(x) = F_X'(x), \qquad x \geqslant 0.$$

Definition 6.7. *Given X a random variable with cumulative distribution function $F_X : \mathbb{R} \longrightarrow [0,1]$ and a level $p \in (0,1)$, the p-quantile q_X^p of X is defined by*

$$q_X^p := \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\}. \tag{6.4}$$

We note that by (6.4), the function $p \mapsto q_X^p$ is the *generalized inverse* $F_X^{-1}(x)$ of the *Cumulative Distribution Function*

$$x \mapsto F_X(x) := \mathbb{P}(X \leqslant x), \qquad x \geqslant 0.$$

of X, see Definition 1 in Embrechts and Hofert (2013). As a consequence, we have the following.

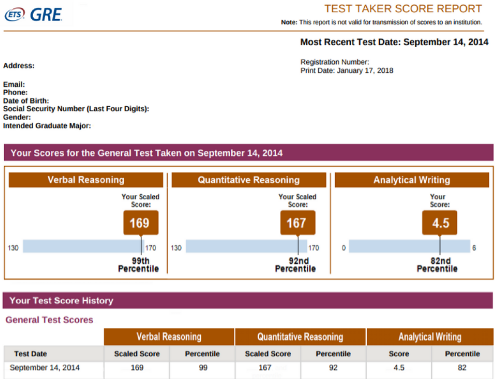
Proposition 6.8.

i) The function $p \mapsto q_X^p$ *is a non-decreasing, left-continuous function of* $p \in [0,1]$*, and it admits limits on the right.*

ii) For all $p \in [0,1]$ *and* $x \in \mathbb{R}$*, we have*

$$p \leqslant F_X(x) \iff q_X^p \leqslant x.$$

Proof. (i) follows from Proposition 1-(2) in Embrechts and Hofert (2013), since $F_X(x)$ is non-decreasing in $x \in \mathbb{R}$, and (ii) follows from Proposition 1-(5) therein, since $F_X(x)$ is right-continuous in $x \in \mathbb{R}$. $\square$



ETS GRE

TEST TAKER SCORE REPORT

Note: This report is not valid for transmission of scores to an institution.

Most Recent Test Date: September 14, 2014

Registration Number:
Print Date: January 17, 2018

Address:

Email:
Phone:
Date of Birth:
Social Security Number (Last Four Digits):
Gender:
Intended Graduate Major:

Your Scores for the General Test Taken on September 14, 2014

Verbal Reasoning	Quantitative Reasoning	Analytical Writing
Your Scaled Score: 169	Your Scaled Score: 167	Your Score: 4.5
130 – 170	130 – 170	0 – 6
99th Percentile	92nd Percentile	82nd Percentile

Your Test Score History

General Test Scores

	Verbal Reasoning		Quantitative Reasoning		Analytical Writing	
Test Date	Scaled Score	Percentile	Scaled Score	Percentile	Score	Percentile
September 14, 2014	169	99	167	92	4.5	82

Fig. 6.4: Example of quantiles given as percentiles.

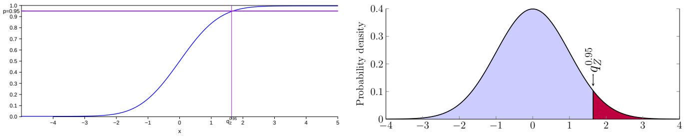
Quantiles of common distributions

The quantiles of various distributions can be obtained in R.

- *Gaussian distribution.* The command

```
qnorm(.95, mean=0, sd=1)
```

shows that the 95%-quantile of a $\mathcal{N}(0,1)$ Gaussian random variable is 1.644854.



(a) Gaussian quantile and CDF. (b) Gaussian quantile and CDF.

Fig. 6.5: Gaussian quantile $q_Z^p = 1.644854$ at $p = 0.95$.

- *Exponential distribution.* The command

```r
qexp(.95, 1)
```

displays the 95%-quantile of an exponentially distributed random variable with CDF

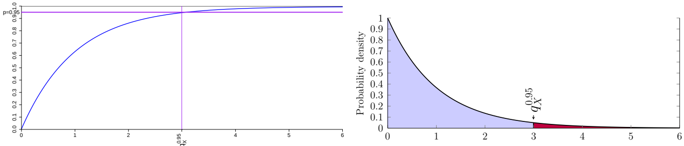
$$\mathbb{P}(X \leqslant x) = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geqslant 0.$$

By equating $\mathbb{P}(X \leqslant q_X^p) = p$, we find

$$\begin{aligned} q_X^p &= \inf\left\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\right\} \\ &= -\frac{1}{\lambda}\log(1-p) \\ &= \mathbb{E}[X]\log\frac{1}{1-p}, \end{aligned}$$

and when $p = 95\%$ and $\lambda = 1$ this yields

$$q_X^p = 2.995732 \simeq 2.996\mathbb{E}[X].$$



(a) Exponential quantile and CDF. (b) Exponential quantile and CDF.

Fig. 6.6: Exponential quantile $q_X^p = 2.995732$ at $p = 0.95$.

- *Student distribution.* The command

```
qt(.90, df=5)
```

displays the 90%-quantile of a Student t-distributed random variable with 5 degrees of freedom, which is 1.475884.

- *Bernoulli distribution.* Consider the Bernoulli random variable $X \in \{0,1\}$ with the distribution

$$\mathbb{P}(X = 1) = 2\%, \qquad \mathbb{P}(X = 0) = 98\%.$$

In this case, we check from Figure 6.7 that $q_X^{0.99} = 1$.

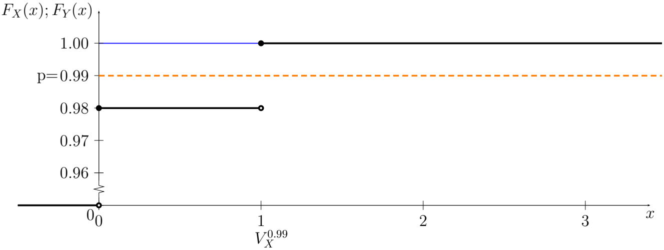


Fig. 6.7: Cumulative distribution function of X.

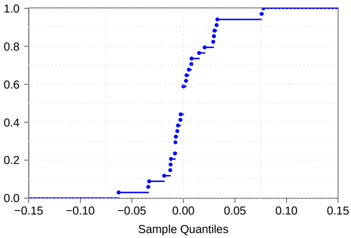
Empirical Cumulative Distribution Function

Definition 6.9. *The* empirical Cumulative Distribution Function *(CDF) of an N-point data set $\{x_1, x_2, x_3, \ldots, x_N\}$ is estimated as*

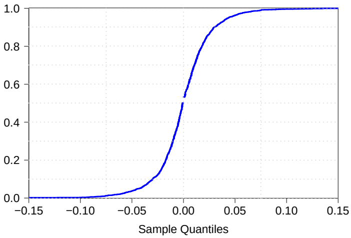
$$F_N(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{x_i \leqslant x\}}, \qquad x \geqslant 0.$$

```
getSymbols("^STI",from="1990-01-03",to="2015-01-03",src="yahoo")
getSymbols("1800.HK",from=Sys.Date()-50,to=Sys.Date(),src="yahoo")
stock=Ad(`1800.HK`);stock.rtn=(stock-lag(stock))/lag(stock);
stock.rtn <- stock.rtn[!is.na(stock.rtn)]
stock.ecdf=ecdf(as.vector(stock.rtn))
plot(stock.ecdf, xlab = 'Sample Quantiles', ylim=c(-0.001,1.002), xlim=c(-0.15,0.15), ylab = '',
     lwd = 3, main = '',col='blue', las=1, cex.lab=1.5, cex.axis=1.5, xaxs='i', yaxs='i');
     grid(4, 10, lwd = 2)
```

```
getSymbols("1800.HK",from=Sys.Date()-3650,to=Sys.Date(),src="yahoo")
stock=Ad(`1800.HK`);stock.rtn=(stock-lag(stock))/lag(stock);
stock.ecdf=ecdf(as.vector(stock.rtn))
plot(stock.ecdf, xlab = 'Sample Quantiles', ylim=c(-0.001,1.002), xlim=c(-0.15,0.15), ylab = '',
     lwd = 2, main = '',col='blue', cex=1, las=1, cex.lab=1.5, cex.axis=1.5, xaxs='i', yaxs='i')
grid(4, 10, lwd = 2)
```



(a) Empirical CDF on 50 samples.



(b) Empirical CDF on 2463 samples.

Fig. 6.8: Empirical cumulative distribution functions.

Note that the empirical distribution function in Figure 6.8-a) has a visible discontinuity (or gap) at $x = 0$, whose height 0.05483347 is given by

```
length(stock.rtn[stock.rtn==0])/length(stock.rtn)
```

6.3 Value at Risk (VaR)

Consider a random variable X used to model the potential losses associated to a given risk. The probability $\mathbb{P}(X > V)$ that X exceeds the level V is of a capital importance. Choosing the value of V such that for example

$$\mathbb{P}(X \leqslant V) \geqslant 0.95, \quad i.e. \quad \mathbb{P}(X > V) \leqslant 0.05,$$

means that insolvency will occur with probability less that 5%. In this setting, the 95%-quantile risk measure is the smallest value of $x \in \mathbb{R}$ such that

$$\mathbb{P}(X \leqslant x) \geqslant 0.95, \quad i.e. \quad \mathbb{P}(X > x) \leqslant 0.05.$$

More precisely, we have the following definition.

Definition 6.10. *The Value at Risk V_X^p of a random variable X at the level $p \in (0, 1)$ is the p-quantile of X defined by*

$$V_X^p := \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\}. \tag{6.5}$$

In other words, for some decreasing sequence $(x_n)_{n\geqslant 1}$ such that

$$\mathbb{P}(X \leqslant x_n) \geqslant p \quad \text{for all} \quad n \geqslant 1,$$

we have

$$V_X^p := \lim_{n\to\infty} x_n. \tag{6.6}$$

Similarly to the above, the function $p \mapsto V_X^p$ is the *generalized inverse* $F_X^{-1}(x)$ of the *Cumulative Distribution Function* $\mapsto F_X$ of X, and from Proposition 6.8-(i) we have the following result.

Proposition 6.11. *The function $p \mapsto V_X^p$ is a* non-decreasing, left-continuous *function of $p \in [0,1]$, and it admits limits on the right.*

In particular, if F_X is continuous and strictly increasing it admits an inverse F_X^{-1}, and in this case V_X^p is given by

$$V_X^p = F_X^{-1}(p), \qquad p \in (0,1).$$

Proposition 6.12. *The Value at Risk V_X^p of X at the level $p \in (0,1)$ satisfies the properties*

$$\mathbb{P}\big(X < V_X^p\big) \leqslant p \leqslant \mathbb{P}\big(X \leqslant V_X^p\big), \tag{6.7}$$

and

$$\mathbb{P}\big(X > V_X^p\big) \leqslant 1 - p \leqslant \mathbb{P}\big(X \geqslant V_X^p\big). \tag{6.8}$$

In particular, if $\mathbb{P}(X = V_X^p) = 0$, then we have

$$p = \mathbb{P}\big(X < V_X^p\big) = \mathbb{P}\big(X \leqslant V_X^p\big). \tag{6.9}$$

Proof. Using the decreasing sequence $(x_n)_{n\geqslant 1}$ in (6.6) and the right continuity of the cumulative distribution function F_X, we have

$$\begin{aligned}
\mathbb{P}\big(X \leqslant V_X^p\big) &= \mathbb{P}(X \leqslant \lim_{n\to\infty} x_n) \\
&= F_X\big(\lim_{n\to\infty} x_n\big) \\
&= \lim_{n\to\infty} F_X(x_n) \\
&= \lim_{n\to\infty} \mathbb{P}(X \leqslant x_n) \\
&\geqslant p.
\end{aligned}$$

On the other hand, if $\mathbb{P}\big(X < V_X^p\big) > p$ then there is a strictly increasing sequence $(y_n)_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} y_n = V_X^p$$

and by (6.3) we have

$$\mathbb{P}\big(X < V_X^p\big) = \lim_{n\to\infty} \mathbb{P}(X \leqslant y_n) > p,$$

in which case there would exist $n \geqslant 1$ such that $y_n < V_X^p$ and $\mathbb{P}(X \leqslant y_n) > p$, which contradicts (6.5). Relations (6.8)-(6.9) are direct consequences of (6.7). $\square$

When $\mathbb{P}\big(X = V_X^p\big) > 0$ we may have $\mathbb{P}\big(X > V_X^p\big) = 0$, for example in the case of a Bernoulli random variable $X \in \{0, 1\}$ with the distribution

$$\mathbb{P}(X = 1) = 2\%, \qquad \mathbb{P}(X = 0) = 98\%,$$

see Figure 6.7. The next proposition also follows from the Definition 6.10 of V_X^p and Proposition 6.8-(ii).

Proposition 6.13. *For all $x \in \mathbb{R}$ we have*

$$V_X^p \leqslant x \quad \Longleftrightarrow \quad \mathbb{P}(X \leqslant x) \geqslant p. \tag{6.10}$$

Proof. $\Leftarrow$) If $\mathbb{P}(X \leqslant x) \geqslant p$ then we have

$$V_X^p = \inf\{y \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant y) \geqslant p\} \leqslant x.$$

$\Rightarrow$) On the other hand, choosing a strictly decreasing sequence $(x_n)_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} x_n = V_X^p \quad \text{and} \quad \mathbb{P}(X \leqslant x_n) \geqslant p, \qquad n \geqslant 1,$$

if $V_X^p \leqslant x$ we have

$$\mathbb{P}(X \leqslant x) \geqslant \mathbb{P}(X \leqslant V_X^p) = \lim_{n\to\infty} \mathbb{P}(X \leqslant x_n) \geqslant p$$

by the right continuity of the cumulative distribution function F_X of X. $\square$

On the other hand, the Value at Risk V_X^p does not reveal any information on *how large* losses can be beyond V_X^p, see Chapter 7 for details. The next proposition shows how to estimate Value at Risk when switching the sign of the data.

Proposition 6.14. *Assume that the cumulative distribution function F_X is continuous and strictly increasing. Then, we have*

$$V_{-X}^p = -V_X^{1-p}, \qquad p \in (0, 1). \tag{6.11}$$

Proof. Since F_X is continuous, we have

$$F_{-X}(x) = \mathbb{P}(-X \leqslant x)$$

$$
\begin{aligned}
&= \mathbb{P}(X \geqslant -x) \\
&= 1 - \mathbb{P}(X < -x) \\
&= 1 - \mathbb{P}(X \leqslant -x) \\
&= 1 - F_X(-x),
\end{aligned}
$$

hence, taking

$$x := F_{-X}^{-1}(p),$$

we have

$$p = F_{-X}\big(F_{-X}^{-1}(p)\big) = 1 - F_X\big(-F_{-X}^{-1}(p)\big),$$

or

$$F_X\big(-F_{-X}^{-1}(p)\big) = 1 - p$$

i.e.

$$F_{-X}^{-1}(p) = -F_X^{-1}(1-p),$$

which yields

$$V_{-X}^{p} = F_{-X}^{-1}(p) = -F_X^{-1}(1-p) = -V_X^{1-p}, \qquad p \in (0,1).$$

□

In Figure 6.9 we choose a continuous CDF F_X with

$$F_{-X}(x) = 1 - F_X(-x), \qquad x \in \mathbb{R}.$$

In this case, the continuity of F_X ensures the symmetry property

$$V_{-X}^{p} = -V_X^{1-p}$$

of Proposition 6.14.

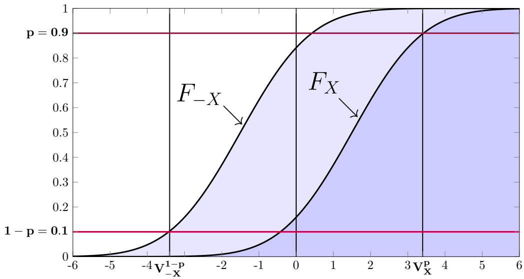


Fig. 6.9: Continuous CDF.

On the other hand, in Figure 6.10 we consider X with distribution

$$\mathbb{P}(X = 1) = 0.5, \quad \mathbb{P}(X = 2) = 0.3, \quad \mathbb{P}(X = 3.4) = 0.2,$$

hence

$$\mathbb{P}(-X = -3.4) = 0.2, \quad \mathbb{P}(-X = -2) = 0.3, \quad \mathbb{P}(-X = -1) = 0.5,$$

and we check that in this discontinuous case the relation $V_{-X}^q = -V_X^{1-q}$ fails for $p = 0.8$, although it still holds for $p' = 0.9$.

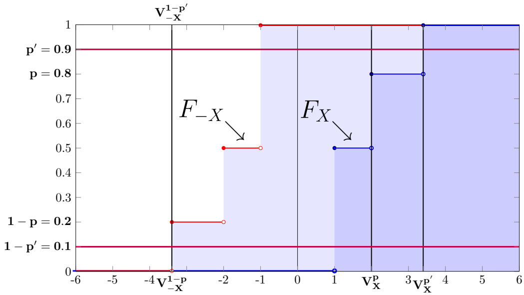


Fig. 6.10: Discontinuous CDF.

Next, we check the properties of Value at Risk. Although Value at Risk satisfies the monotonicity, positive homogeneity and translation invariance properties, it is *not* subadditive in general. Namely, the Value at Risk V^p_{X+Y} of $X+Y$ may be larger than the sum $V^p_X + V^p_Y$.

Proposition 6.15. *Value at Risk V^p_X is a monotone, positive homegeneous and translation invariant risk measure. However, it is not subadditive, and therefore it is not a coherent risk measure.*

Proof.

a) *Monotonicity.*

Value at Risk is a monotone risk measure. If $X \leqslant Y$ then

$$\mathbb{P}(Y \leqslant x) = \mathbb{P}(X \leqslant Y \leqslant x) \leqslant \mathbb{P}(X \leqslant x), \qquad x \geqslant 0,$$

hence

$$\mathbb{P}(Y \leqslant x) \geqslant p \quad \Longrightarrow \quad \mathbb{P}(X \leqslant x) \geqslant p, \qquad x \geqslant 0,$$

which shows that

$$V^p_X \leqslant V^p_Y$$

by (6.5).

b) *Positive homogeneity and translation invariance.*

Value at Risk satisfies the positive homogeneity and translation invariance properties. For any $\mu \in \mathbb{R}$ and $\lambda > 0$, we have

$$\begin{aligned} V^p_{\mu+\lambda X} &= \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(\mu + \lambda X \leqslant x) \geqslant p\} \\ &= \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant (x-\mu)/\lambda) \geqslant p\} \\ &= \inf\{\mu + \lambda y \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant y) \geqslant p\} \\ &= \mu + \lambda \inf\{y \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant y) \geqslant p\} \\ &= \mu + \lambda V^p_X. \end{aligned}$$

c) *Subadditivity and coherence.*

We show that Value at Risk is *not subadditive* by considering two *independent* Bernoulli random variables $X, Y \in \{0,1\}$ having the same distribution

$$\begin{cases} \mathbb{P}(X=1) = \mathbb{P}(Y=1) = 2\%, \\ \\ \mathbb{P}(X=0) = \mathbb{P}(Y=0) = 98\%, \end{cases}$$

hence $V^{0.975}_X = V^{0.975}_Y = 0$.

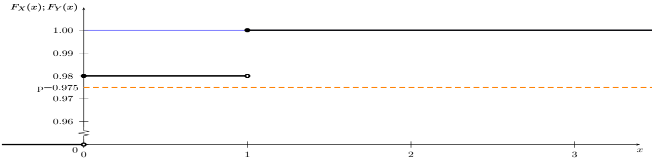


Fig. 6.11: Cumulative distribution function of X and Y.

On the other hand, we have

$$\begin{cases} \mathbb{P}(X+Y=2) = \mathbb{P}(X=1 \text{ and } Y=1) = (0.02)^2 = 0.04\%, \\ \\ \mathbb{P}(X+Y=1) = 2 \times 0.02 \times 0.98 = 3.92\%, \\ \\ \mathbb{P}(X+Y=0) = \mathbb{P}(X=0 \text{ and } Y=0) = (0.98)^2 = 96.04\%, \end{cases}$$

hence

$$V_{X+Y}^{0.975} = 1 > V_X^{0.975} + V_Y^{0.975} = 0.$$

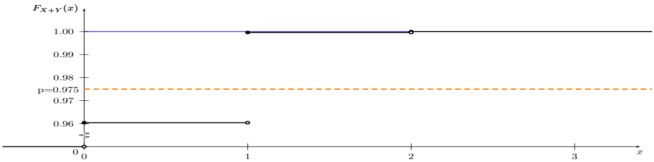


Fig. 6.12: Cumulative distribution function of $X+Y$.

$\square$

In the next proposition, we use the standard Gaussian Cumulative Distribution Function (CDF)

$$\Phi(x) := \int_{-\infty}^{x} \mathrm{e}^{-y^2/2} \frac{dy}{\sqrt{2\pi T}}, \qquad x \in \mathbb{R},$$

of a standard normal random variable $Z \simeq \mathcal{N}(0,1)$.

Proposition 6.16. Gaussian Value at Risk. *Given* $X \simeq \mathcal{N}(\mu_X, \sigma_X^2)$, *we have*

$$V_X^p = \mu_X + \sigma_X q_Z^p \tag{6.12}$$

where the normal quantile $q_Z^p = V_Z^p$ at the level p satisfies

$$\Phi(q_Z^p) = \mathbb{P}(Z \leqslant q_Z^p) = p \quad \textit{for} \quad Z \simeq \mathcal{N}(0,1),$$

i.e.

$$q_Z^p = \Phi^{-1}(p) \quad \textit{and} \quad V_X^p = \mu_X + \sigma_X \Phi^{-1}(p).$$

Proof. We represent the random variable $X \simeq \mathcal{N}(\mu_X, \sigma_X^2)$ as

$$X = \mu_X + \sigma_X Z,$$

where $Z \simeq \mathcal{N}(0,1)$ is a standard normal random variable, and use the relation

$$\begin{aligned} p &= \mathbb{P}(X \leqslant V_X^p) \\ &= \mathbb{P}(\mu_X + \sigma_X Z \leqslant V_X^p) \\ &= \mathbb{P}(Z \leqslant (V_X^p - \mu_X)/\sigma_X) \\ &= \mathbb{P}(Z \leqslant q_Z^p), \end{aligned}$$

which holds provided that $V_X^p = \mu_X + \sigma_X q_Z^p$. □

We also note that if $X \simeq \mathcal{N}(\mu_X, \sigma_X^2)$ then $-X \simeq \mathcal{N}(-\mu_X, \sigma_X^2)$, hence

$$\begin{aligned} V_{-X}^p &= -\mu_X + \sigma_X q_Z^p \\ &= -\mu_X - \sigma_X q_Z^{1-p} \\ &= -V_X^{1-p}, \end{aligned}$$

which is consistent with (6.11).

The next remark shows that, although Value at Risk is *not sub-additive* in general, it is sub-additive (and therefore coherent) on (not necessarily independent) Gaussian random variables.

Remark 6.17. *If X and Y are two Gaussian random variables, we have*

$$V_{X+Y}^p \leqslant V_X^p + V_Y^p.$$

Proof. By (6.12), for any two random variables X and Y, we have

$$\begin{aligned} &\sigma_{X+Y}^2 = \mathrm{Var}[X+Y] \\ &= \mathbb{E}\big[(X+Y)^2\big] - (\mathbb{E}[X+Y])^2 \\ &= \mathbb{E}\big[X^2\big] + \mathbb{E}\big[Y^2\big] + 2\mathbb{E}[XY] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}(X,Y) \qquad (6.13) \\
&\leqslant \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\sqrt{\mathbb{E}\big[(X - \mathbb{E}[X])^2\big]}\sqrt{\mathbb{E}\big[\big(Y - \mathbb{E}[Y]\big)^2\big]} \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]} \qquad (6.14) \\
&= \big(\sqrt{\mathrm{Var}[X]} + \sqrt{\mathrm{Var}[Y]}\big)^2,
\end{aligned}
$$

where, from (6.13) to (6.14) we applied the *Cauchy-Schwarz* inequality, hence $\sigma_{X+Y} \leqslant \sigma_X + \sigma_Y$. Assuming that X and Y are Gaussian, by (6.12) we find

$$
\begin{aligned}
V^p_{X+Y} &= \mu_{X+Y} + \sigma_{X+Y} q^p_Z \\
&= \mu_X + \mu_Y + \sigma_{X+Y} q^p_Z \\
&\leqslant \mu_X + \mu_Y + (\sigma_X + \sigma_Y) q^p_Z \\
&= V^p_X + V^p_Y.
\end{aligned}
$$

$\square$

6.4 Numerical estimates

In this section we are using the PerformanceAnalytics R package, see also § 6.1.1 of Mina and Xiao (2001). In case we care about negative return values, Definition 6.10 is replaced with

$$
\overline{V}^p_X := \mathrm{Sup}\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \leqslant 1 - p\}. \qquad (6.15)
$$

In case the CDF of X is continuous, we note the relation

$$
\begin{aligned}
\overline{V}^p_X &= \mathrm{Sup}\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \leqslant 1 - p\} \\
&= -\inf\{-x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \leqslant 1 - p\} \\
&= -\inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant -x) \leqslant 1 - p\} \\
&= -\inf\{x \in \mathbb{R} \ : \ \mathbb{P}(-X \geqslant x) \leqslant 1 - p\} \\
&= -\inf\{x \in \mathbb{R} \ : \ 1 - \mathbb{P}(-X \geqslant x) \geqslant p\} \\
&= -\inf\{x \in \mathbb{R} \ : \ \mathbb{P}(-X \leqslant x) \geqslant p\} \\
&= -V^p_{-X},
\end{aligned}
$$

hence the relation

$$
\overline{V}^p_X = -V^p_{-X} = V^{1-p}_X
$$

which is obtained from Proposition 6.14 when the cumulative distribution function F_X is continuous and strictly increasing.

```
install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
getSymbols("0700.HK",from="2010-01-03",to="2018-02-01",src="yahoo")
stock=Ad(`0700.HK`);chartSeries(stock,up.col="blue",theme="white")
stock.rtn=(stock-lag(stock))/lag(stock)[-1];stock.rtn <- stock.rtn[!is.na(stock.rtn)]
dev.new(width=16,height=7); chart.CumReturns(stock.rtn, main="Cumulative Returns")
var=VaR(stock.rtn, p=.95, method="historical");var
length(stock.rtn[stock.rtn<var[1]])/length(stock.rtn)
times=index(stock);chartSeries(stock.rtn,up.col="blue",theme="white")
abline(h=var,col="red",lwd=3)
```

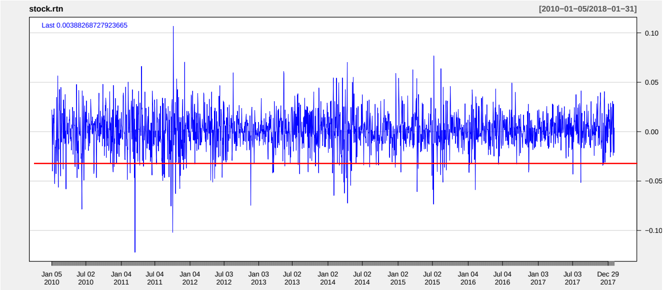


Fig. 6.13: Market returns *vs.* Value at Risk level in red.

The historical 95%-Value at Risk over N samples $(x_i)_{i=1,2,\ldots,N}$ can be estimated by inverting the *empirical cumulative distribution function* $F_N(x)$, and is found to be $\overline{V}_X^{95\%} = -0.03165963$.

```
VaR(stock.rtn, p=.95, method="gaussian",invert="FALSE")
VaR(stock.rtn, p=.95, method="gaussian",invert="TRUE")
```

The Gaussian 95%-Value at Risk is estimated from (6.12) with $p = 0.95$ as

$$\overline{V}_X^p = V_X^{1-p} = \mu + \sigma q_Z^{1-p} = \mu - \sigma q_Z^p,$$

where $-\mu = \mathbb{E}[-X]$ and $\sigma^2 = \mathrm{Var}[-X]$, and is found equal to

$$\overline{V}_X^{95\%} = -0.03115425.$$

It can be recovered up to approximation according to Proposition 6.16 from the following R code, which yields −0.0311592.

```
m=mean(stock.rtn,na.rm=TRUE); s=sd(stock.rtn,na.rm=TRUE)
q=qnorm(.95, mean=0, sd=1); m-s*q
```

Note that here we are concerned about large negative returns, which explains the negative sign in $m - s * q$.

The next lemma is useful for random simulation purposes, and it will also be used in the proof of Propositions 7.6 and 7.12 below.

Lemma 6.18. *Any random variable X can be represented as*

$$X = V_X^U = F_X^{-1}(U),$$

where U a uniformly distributed random variable on $[0,1]$.

Proof. It suffices to note that by (6.10) we have

$$\mathbb{P}\big(V_X^U \leqslant x\big) = \mathbb{P}(U \leqslant \mathbb{P}(X \leqslant x)) = \mathbb{P}(X \leqslant x) = F_X(x), \quad x \geqslant 0.$$

□

Exercises

Exercise 6.1 Consider a random variable X having the Pareto distribution with probability density function

$$f_X(x) = \frac{\gamma\theta^\gamma}{(\theta + x)^{\gamma+1}}, \qquad x \geqslant 0.$$

a) Compute the cumulative distribution function

$$F_X(x) := \int_0^x f_X(y)dy, \qquad x \geqslant 0.$$

b) Compute the value at risk V_X^p at the level p for any θ and γ, and then for $p = 99\%$, $\theta = 40$ and $\gamma = 2$.

Exercise 6.2 Consider a random variable X with the following cumulative distribution function:

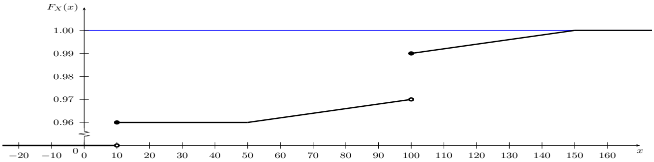


Fig. 6.14: Cumulative distribution function of X.

a) Give the value of $\mathbb{P}(X = 100)$.
b) Give the value of V_X^q for all q in the interval $[0.97, 0.99]$.
c) Compute the value of V_X^q for all q in the interval $[0.99, 1]$.

Hint: We have

$$F_X(x) = \mathbb{P}(X \leqslant x) = 0.99 + 0.01 \times \frac{x - 100}{50}, \qquad x \in [100, 150].$$

Exercise 6.3 Discrete distribution. Consider $X \in \{10, 100, 110\}$ with the distribution

$$\mathbb{P}(X = 10) = 90\%, \quad \mathbb{P}(X = 100) = 9.5\%, \quad \mathbb{P}(X = 110) = 0.5\%.$$

Compute the value at risk $V_X^{99\%}$.

Exercise 6.4 Exponential distribution. Assume that X has an exponential distribution with parameter $\lambda > 0$ and mean $1/\lambda$, *i.e.*

$$\mathbb{P}(X \leqslant x) = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geqslant 0.$$

a) Compute
$$V_X^p := \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\}$$
and $V_X^{95\%}$.
b) Assuming that the liabilities of a company are estimated by $\mathbb{E}[X]$, compute the amount of required capital C_X from (6.1).

Exercise 6.5 Given X a random variable having the geometric distribution with

$$\mathbb{P}(X = k) = (1 - p)^k p, \qquad k \geqslant 0,$$

compute the conditional expectation $\mathbb{E}[X \mid X \geqslant a]$ for $a > 0$.

Exercise 6.6 Estimating risk probabilities from moments.

a) Show that for every $r > 0$

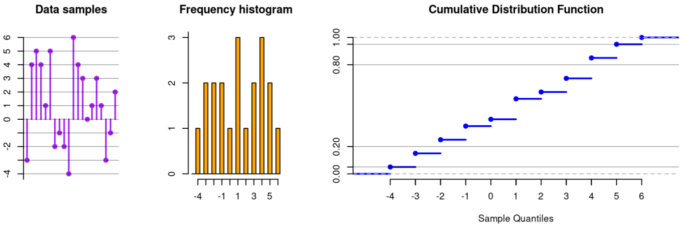
$$V_X^p \leqslant \left(\frac{\mathbb{E}[|X|^r]}{1-p}\right)^{1/r} = \frac{\|X\|_{L^r(\Omega)}}{(1-p)^{1/r}},$$

where $\|X\|_{L^r(\Omega)} := (\mathbb{E}[|X|^r])^{1/r}$.

Hint: Use the argument of the Markov inequality.

b) Give an upper bound for $V_X^{95\%}$ when $p = 95\%$ and $r = 1$.

Exercise 6.7 We consider a discrete random variable X having the following distribution.



a) Find the following quantities for the above data set, and mark their values on the graph.

i) Historical "Academic" Value at Risk at $p = 0.95$. $\mathrm{VaR}^{95}_{\mathrm{Ac\text{-}H}} =$ ________

ii) Historical "Academic" Value at Risk at $p = 0.80$. $\mathrm{VaR}^{80}_{\mathrm{Ac\text{-}H}} =$ ________

iii) Historical "Practitioner" Value at Risk at $p = 0.95$. $\overline{\mathrm{VaR}}^{95}_{\mathrm{Pr\text{-}H}} =$ ________

iv) Historical "Practitioner" Value at Risk at $p = 0.80$. $\overline{\mathrm{VaR}}^{80}_{\mathrm{Pr\text{-}H}} =$ ________

b) Knowing that mean=1.15, sd=3.048, qnorm(0.95)=1.645 and qnorm(0.80)=0.842, compute (from Proposition 6.16):

i) Gaussian "Academic" Value at Risk at $p = 0.95$. $\mathrm{VaR}^{95}_{\mathrm{Ac\text{-}G}} =$ ________

ii) Gaussian "Academic" Value at Risk at $p = 0.80$. $\mathrm{VaR}^{80}_{\mathrm{Ac\text{-}G}} =$ ________

iii) Gaussian "Practitioner" Value at Risk at $p = 0.95$. $\overline{\mathrm{VaR}}^{95}_{\mathrm{Pr\text{-}G}} =$ ________

iv) Gaussian "Practitioner" Value at Risk at $p = 0.80$. $\overline{\mathrm{VaR}}^{80}_{\mathrm{Pr\text{-}G}} =$ ________

Chapter 7
Expected Shortfall

This chapter presents the construction of Tail Value at Risk (TVaR) and the Expected Shortfall (ES), which, unlike Value at Risk, are coherent risk measures. The Tail Value at Risk at the confidence level $p \in (0,1)$ is defined as the average of losses suffered in the worst $(1-p)\%$ of events. Expected Shortfall provides an alternative computation of Tail Value at Risk (TVaR) by averaging potential losses above the VaR level. Experiments based on financial data sets are also included.

7.1 Tail Value at Risk (TVaR)

A basic shortcoming of Value at Risk is failing to provide information on the behavior of probability distribution tails beyond V_X^p. The next figure illustrates the limitations of Value at Risk, namely its inability to capture the properties of a probability distribution beyond V_X^p.[†]

[†] "Value at Risk is like an airbag that works all the time, except when you have a car accident". - D. Einhorn, hedge fund manager.

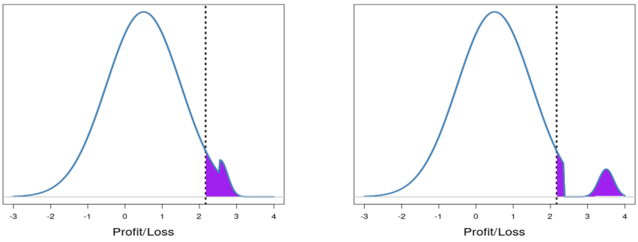


Fig. 7.1: Two distributions having the same Value at Risk $V_X^{95\%} = 2.145$.

The Tail Value at Risk (or Conditional Value at Risk) aims at providing a solution to the tail distribution problem observed with Value at Risk at the level $p \in (0,1)$ by averaging over confidence levels ranging from p to 1.

Definition 7.1. *The* Tail Value at Risk *(TVaR) of a random variable X at the level $p \in (0,1)$ is defined by the average*

$$\mathrm{TV}_X^p := \frac{1}{1-p} \int_p^1 V_X^q dq. \tag{7.1}$$

We note the following property.

Proposition 7.2. *The* Tail Value at Risk *(TVaR) and Value at risk (VaR) satisfy the following inequality:*

$$\mathrm{TV}_X^p \geqslant V_X^p, \qquad p \in (0,1).$$

Proof. Since the function $p \mapsto V_X^p$ is non-decreasing by Proposition 6.11, we have

$$\mathrm{TV}_X^p = \frac{1}{1-p} \int_p^1 V_X^q dq \geqslant \frac{1}{1-p} \int_p^1 V_X^p dq = V_X^p.$$

□

7.2 Conditional Tail Expectation (CTE)

Recall that by Lemma A.15, given an event A such that $\mathbb{P}(A) > 0$, the *conditional expectation* of $X : \Omega \longrightarrow \mathbb{N}$ given the event A satisfies

$$\mathbb{E}[X \mid A] = \frac{1}{\mathbb{P}(A)} \mathbb{E}\left[X \mathbb{1}_A\right],$$

see Section 6.1 for an example.

Definition 7.3. *Consider a random variable X such that $\mathbb{P}(X > V_X^p) > 0$. The* Conditional Tail Expectation *of X at the level $p \in (0,1)$ is the quantity*

$$\mathrm{CTE}_X^p := \mathbb{E}\big[X \mid X > V_X^p\big] = \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{E}\big[X \mathbb{1}_{\{X > V_X^p\}}\big].$$

The use of the strict inequality "$>$" in the definition of the Conditional Tail Expectation allows us to avoid any dependence on $\mathbb{P}(X = V_X^p)$, and to consider risky values strictly beyond V_X^p. The Conditional Tail Expectation is also called Conditional Value at Risk (CVaR).

Proposition 7.4. *The* Conditional Tail Expectation CTE_X^p *at the level $p \in (0,1)$ can be written as the* distortion risk measure

$$\mathrm{CTE}_X^p := \mathbb{E}[X f_X(X)], \tag{7.2}$$

where f_X is the function defined by

$$f_X(x) := \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{1}_{\{x > V_X^p\}}, \qquad x \in \mathbb{R}. \tag{7.3}$$

Proof. We check that the function f_X defined in (7.3) is a *distortion function* according to Definition 6.4. Indeed, by Proposition 6.15 we have

$$\begin{aligned} f_{\mu+X}(\mu + x) &= \frac{1}{\mathbb{P}(\mu + X > V_{\mu+X}^p)}\mathbb{1}_{\{\mu+x > V_{\mu+X}^p\}} \\ &= \frac{1}{\mathbb{P}(\mu + X > \mu + V_X^p)}\mathbb{1}_{\{\mu+x > \mu+V_X^p\}} \\ &= \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{1}_{\{x > V_X^p\}} \\ &= f_X(x), \qquad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \end{aligned}$$

and

$$\begin{aligned} f_{\lambda X}(\lambda x) &= \frac{1}{\mathbb{P}(\lambda X > V_{\lambda X}^p)}\mathbb{1}_{\{\lambda x > V_{\lambda X}^p\}} \\ &= \frac{1}{\mathbb{P}(\lambda X > \lambda V_X^p)}\mathbb{1}_{\{\lambda x > \lambda V_X^p\}} \\ &= \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{1}_{\{x > V_X^p\}} \\ &= f_X(x), \qquad x \in \mathbb{R}, \quad \lambda > 0. \end{aligned}$$

Finally, we note that

$$\begin{aligned}
\mathbb{E}[f_X(X)] &= \mathbb{E}\left[\frac{1}{\mathbb{P}\big(X > V_X^p\big)} \mathbb{1}_{\{x > V_X^p\}}\right] \\
&= \frac{1}{\mathbb{P}\big(X > V_X^p\big)} \mathbb{E}\left[\mathbb{1}_{\{x > V_X^p\}}\right] \\
&= \frac{\mathbb{P}\big(X > V_X^p\big)}{\mathbb{P}\big(X > V_X^p\big)} \\
&= 1.
\end{aligned}$$

□

Examples of Conditional Tail Expectations can be computed using the following R code in which computation is done on sign-changed data, *i.e.* according to the "practitioner" point of view.

```
library(quantmod); getSymbols("^HSI",from="2013-06-01",to="2014-10-01",src="yahoo")
returns <- as.vector(diff(log(Ad(`HSI`)))); library(PerformanceAnalytics)
var=VaR(returns, p=.95, method="historical")
cte=mean(returns[returns<as.numeric(var)],na.rm=TRUE)
```

The next proposition shows more precisely by which amount the Conditional Tail Expectation exceeds the Value at Risk.

Proposition 7.5. *Let X be a random variable X such that $\mathbb{P}\big(X > V_X^p\big) > 0$. For any $p \in (0, 1]$ we have $\mathrm{CTE}_X^p > \mathbb{E}[X]$ and $\mathrm{CTE}_X^p > V_X^p$ with, more precisely,*

$$\mathrm{CTE}_X^p = \mathbb{E}\big[X \mid X > V_X^p\big] = V_X^p + \mathbb{E}\big[\big(X - V_X^p\big)^+ \mid X > V_X^p\big].$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\big[X \mid X > V_X^p\big] &= \frac{1}{\mathbb{P}\big(X > V_X^p\big)} \mathbb{E}\big[X \mathbb{1}_{\{X > V_X^p\}}\big] \\
&= \frac{1}{\mathbb{P}\big(X > V_X^p\big)} \left(\mathbb{E}\big[\big(X - V_X^p\big)\mathbb{1}_{\{X > V_X^p\}}\big] + V_X^p \mathbb{E}\big[\mathbb{1}_{\{X > V_X^p\}}\big]\right) \\
&= \frac{1}{\mathbb{P}\big(X > V_X^p\big)} \left(\mathbb{E}\big[\big(X - V_X^p\big)^+\big] + V_X^p \mathbb{P}\big(X > V_X^p\big)\right) \\
&= V_X^p + \frac{1}{\mathbb{P}\big(X > V_X^p\big)} \mathbb{E}\big[\big(X - V_X^p\big)^+\big] \\
&= V_X^p + \mathbb{E}\big[\big(X - V_X^p\big)^+ \,\big|\, X > V_X^p\big].
\end{aligned}$$

See Exercise 7.2-(d) for a proof of $\mathrm{CTE}_X^p > \mathbb{E}[X]$. □

Next, we check that when $\mathbb{P}(X = V_X^p) = 0$, the Conditional Tail Expectation coincides with the Tail Value at Risk. Note that in this case, we have

$$\mathbb{P}(X > V_X^p) = 1 - p > 0$$

by (6.8) in Proposition 6.12.

Proposition 7.6. *Assume that* $\mathbb{P}(X = V_X^p) = 0$. *Then we have*

$$\mathrm{CTE}_X^p = \mathrm{TV}_X^p,$$

i.e.

$$\mathrm{CTE}_X^p = \mathbb{E}\left[X \mid X > V_X^p\right] = \mathbb{E}\left[X \mid X \geqslant V_X^p\right] = \frac{1}{1-p}\int_p^1 V_X^q dq = \mathrm{TV}_X^p. \tag{7.4}$$

Proof. By Lemma 6.18 we construct X as $X = V_X^U$ where U is uniformly distributed on $[0, 1]$, with

$$U \geqslant p \Longrightarrow V_X^U \geqslant V_X^p \Longleftrightarrow X \geqslant V_X^p,$$

and

$$U \leqslant p \Longrightarrow V_X^U \leqslant V_X^p \Longleftrightarrow X \leqslant V_X^p,$$

hence

$$X > V_X^p \Longleftrightarrow V_X^U > V_X^p \Longrightarrow U > p.$$

Since $\mathbb{P}(X = V_X^p) = 0$ we find that, with probability 1,

$$U \geqslant p \Longleftrightarrow U > p \Longleftrightarrow V_X^U \geqslant V_X^p \Longleftrightarrow X \geqslant V_X^p \Longleftrightarrow X > V_X^p,$$

hence

$$\begin{aligned}
\mathrm{CTE}_X^p &= \mathbb{E}\left[X \mid X > V_X^p\right] \\
&= \mathbb{E}\left[V_X^U \mid V_X^U > V_X^p\right] \\
&= \mathbb{E}\left[V_X^U \mid U \geqslant p\right] \\
&= \frac{1}{\mathbb{P}(U \geqslant p)}\mathbb{E}\left[V_X^U \mathbb{1}_{\{U \geqslant p\}}\right] \\
&= \frac{1}{1-p}\int_p^1 V_X^q dq.
\end{aligned}$$

$\square$

Figure 7.2 shows the locations of Value at Risk and Conditional Tail Expectation on a given data set. Note that here, the computation is done on sign-changed data according to Proposition 6.14, *i.e.* the output is computed according to the "practitioner" point of view.

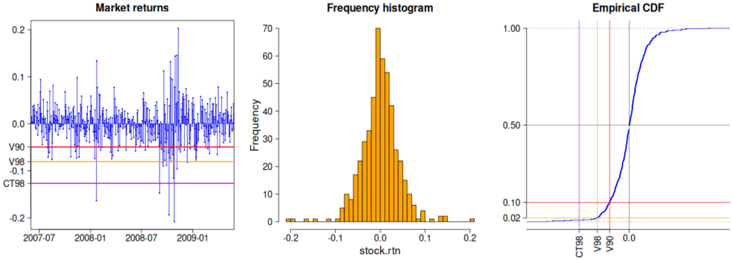


Fig. 7.2: Value at Risk and Conditional Tail Expectation.

The Conditional Tail Expectation of a Gaussian $\mathcal{N}(\mu,\sigma^2)$ random variable is computed in the next proposition, see also Proposition 6.16.

Proposition 7.7. Gaussian CTE. *Given* $X \simeq \mathcal{N}(\mu_X,\sigma_X^2)$ *we have*

$$\mathrm{CTE}_X^p = \mu_X + \frac{\sigma_X}{1-p}\phi(q_Z^p) = \mu_X + \frac{\sigma_X}{(1-p)\sqrt{2\pi}}\mathrm{e}^{-(q_Z^p)^2/2}, \qquad (7.5)$$

where $q_Z^p = \Phi^{-1}(p)$ *is the Gaussian quantile of* $Z \simeq \mathcal{N}(0,1)$ *at the level* $p \in (0,1)$ *and*

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}, \qquad z \in \mathbb{R},$$

is the standard normal probability density function.

Proof. Using the relation $\mathbb{P}(X \geqslant V_X^p) = \mathbb{P}(X > V_X^p) = 1-p$, cf. Proposition 6.13, by Proposition 7.6 we have

$$\begin{aligned}
\mathrm{CTE}_X^p &= \mathrm{TV}_X^p \\
&= \mathbb{E}\big[X \mid X > V_X^p\big] \\
&= \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{E}\big[X \mathbb{1}_{\{X > V_X^p\}}\big] \\
&= \frac{1}{1-p}\int_{V_X^p}^\infty x\mathrm{e}^{-(x-\mu_X)^2/(2\sigma_X^2)}\frac{dx}{\sqrt{2\pi\sigma_X^2}} \\
&= \frac{\mu_X}{1-p}\int_{V_X^p}^\infty \mathrm{e}^{-(x-\mu_X)^2/(2\sigma_X^2)}\frac{dx}{\sqrt{2\pi\sigma_X^2}} + \frac{1}{1-p}\int_{V_X^p}^\infty (x-\mu_X)\mathrm{e}^{-(x-\mu_X)^2/(2\sigma_X^2)}\frac{dx}{\sqrt{2\pi\sigma_X^2}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mu_X}{1-p}\mathbb{P}(X \geqslant V_X^p) + \frac{\sigma_X^2}{(1-p)\sqrt{2\pi\sigma_X^2}}\left[-\mathrm{e}^{-(x-\mu_X)^2/(2\sigma_X^2)}\right]_{V_X^p}^{\infty} \\
&= \mu_X + \frac{\sigma_X^2}{(1-p)\sqrt{2\pi\sigma_X^2}}\mathrm{e}^{-((V_X^p-\mu_X)/\sigma_X)^2/2} \\
&= \mu_X + \frac{\sigma_X}{(1-p)\sqrt{2\pi}}\mathrm{e}^{-(q_Z^p)^2/2} \\
&= \mu_X + \frac{\sigma_X}{1-p}\phi(q_Z^p),
\end{aligned}
$$

due to the rescaling relation $V_X^p = \mu_X + \sigma_X q_Z^p$, cf. (6.12). □

7.3 Expected Shortfall (ES)

There are several variants for the definition of the Expected Shortfall ES_X^p. Next is a frequently used definition.

Definition 7.8. *The Expected Shortfall* ES_X^p *of a random variable* X *at the level* $p \in (0,1)$ *is defined by*

$$
\mathrm{ES}_X^p := V_X^p + \frac{1}{1-p}\mathbb{E}[(X - V_X^p)^+]. \tag{7.6}
$$

The next proposition provides an alternative expression for ES_X^p.

Proposition 7.9. *The Expected Shortfall of* X *at the level* $p \in (0,1)$ *can be written as*

$$
\mathrm{ES}_X^p = \frac{1}{1-p}\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^p\}}\big] + \frac{V_X^p}{1-p}\big(1 - p - \mathbb{P}\big(X \geqslant V_X^p\big)\big).
$$

Proof. By Lemma 6.2, we have

$$
\begin{aligned}
\mathrm{ES}_X^p &= V_X^p + \frac{1}{1-p}\mathbb{E}[(X - V_X^p)^+] \\
&= V_X^p + \frac{1}{1-p}\mathbb{E}\big[\big(X - V_X^p\big)\mathbb{1}_{\{X \geqslant V_X^p\}}\big] \\
&= V_X^p + \frac{\mathbb{P}\big(X \geqslant V_X^p\big)}{1-p}\mathbb{E}\big[X - V_X^p \mid X \geqslant V_X^p\big] \\
&= V_X^p + \frac{1}{1-p}\big(\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^p\}}\big] - V_X^p \mathbb{E}\big[\mathbb{1}_{\{X \geqslant V_X^p\}}\big]\big) \\
&= V_X^p + \frac{1}{1-p}\big(\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^p\}}\big] - V_X^p \mathbb{P}\big(X \geqslant V_X^p\big)\big)
\end{aligned}
$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] + \frac{V_X^p}{1-p}\big(1-p-\mathbb{P}\big(X\geqslant V_X^p\big)\big).$$

□

When $\mathbb{P}\big(X = V_X^p\big) = 0$, Proposition 7.9 also yields

$$\mathrm{ES}_X^p = \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X>V_X^p\}}\big] + \frac{V_X^p}{1-p}\left(1-p-\mathbb{P}\big(X>V_X^p\big)\right).$$

Proposition 7.10. *When* $\mathbb{P}\big(X = V_X^p\big) = 0$ *the Expected Shortfall* ES_X^p *coincides with the* Conditional Tail Expectation CTE_X^p *and with the Tail Value at Risk* TV_X^p*, i.e., we have*

$$\mathrm{ES}_X^p = \mathbb{E}\big[X \mid X > V_X^p\big] = \mathbb{E}\big[X \mid X \geqslant V_X^p\big] = \mathrm{TV}_X^p.$$

Proof. By Relation (6.9) in Proposition 6.12, when $\mathbb{P}\big(X = V_X^p\big) = 0$ we have

$$p = \mathbb{P}(X \leqslant V_X^p) \quad \text{and} \quad 1-p = \mathbb{P}(X > V_X^p) = \mathbb{P}(X \geqslant V_X^p),$$

hence

$$\begin{aligned}
\mathrm{ES}_X^p &= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] \\
&= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X>V_X^p\}}\big] \\
&= \frac{1}{\mathbb{P}\big(X>V_X^p\big)}\mathbb{E}\big[X\mathbb{1}_{\{X>V_X^p\}}\big] \\
&= \mathbb{E}\big[X \mid X > V_X^p\big] \\
&= \mathrm{TV}_X^p,
\end{aligned}$$

by Proposition 7.6. □

From Propositions 7.9 and 7.10, we deduce that

$$\mathrm{ES}_X^p = \begin{cases} \dfrac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] = \mathbb{E}\big[X \mid X > V_X^p\big] = \mathrm{TV}_X^p & \text{if } \mathbb{P}\big(X = V_X^p\big) = 0, \\[2ex] \dfrac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] + \dfrac{V_X^p}{1-p}\big(1-p-\mathbb{P}\big(X\geqslant V_X^p\big)\big) & \text{if } \mathbb{P}\big(X = V_X^p\big) \geqslant 0. \end{cases}$$

In particular, by Propositions 7.7 and 7.10, the Gaussian Expected Shortfall of $X \simeq \mathcal{N}(\mu, \sigma^2)$ at the level $p \in (0,1)$ is also given by

$$\mathrm{ES}_X^p = \mathrm{CTE}_X^p = \mu + \frac{\sigma}{1-p}\phi(\Phi^{-1}(p)) = \mu + \frac{\sigma}{(1-p)\sqrt{2\pi}}\mathrm{e}^{-(\Phi^{-1}(p))^2/2}.$$

Proposition 7.11. *The Expected Shortfall* ES_X^p *at the level* $p \in (0,1)$ *can be written as the* distortion risk measure

$$\mathrm{ES}_X^p = \mathbb{E}[X f_X(X)], \tag{7.7}$$

where the function f_X *defined by*

$$f_X(x) := \frac{1}{1-p}\mathbb{1}_{\{x>V_X^p\}} + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{1-p-\mathbb{P}\big(X>V_X^p\big)}{(1-p)\mathbb{P}\big(X=V_X^p\big)}\mathbb{1}_{\{x=V_X^p\}},$$

$x \in \mathbb{R}$, *is a* distortion function.

Proof. By Proposition 7.9, we have

$$\begin{aligned}
\mathrm{ES}_X^p &= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] + \frac{V_X^p}{1-p}\big(1-p-\mathbb{P}\big(X\geqslant V_X^p\big)\big)\\
&= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{V_X^p}{1-p}\big(1-p-\mathbb{P}\big(X\geqslant V_X^p\big)\big)\\
&= \frac{1}{1-p}\mathbb{E}\left[\left(\mathbb{1}_{\{X\geqslant V_X^p\}} + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{1-p-\mathbb{P}\big(X\geqslant V_X^p\big)}{\mathbb{P}\big(X=V_X^p\big)}\mathbb{1}_{\{X=V_X^p\}}\right)X\right]\\
&= \frac{1}{1-p}\mathbb{E}\Bigg[\bigg(\mathbb{1}_{\{X\geqslant V_X^p\}} - \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\mathbb{1}_{\{X=V_X^p\}}\\
&\qquad + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{1-p-\mathbb{P}\big(X>V_X^p\big)}{\mathbb{P}\big(X=V_X^p\big)}\mathbb{1}_{\{X=V_X^p\}}\bigg)X\Bigg]\\
&= \frac{1}{1-p}\mathbb{E}\left[\left(\mathbb{1}_{\{X>V_X^p\}} + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{1-p-\mathbb{P}\big(X>V_X^p\big)}{\mathbb{P}\big(X=V_X^p\big)}\mathbb{1}_{\{X=V_X^p\}}\right)X\right].
\end{aligned}$$

In order to show that f_X is a distortion function according to Definition 6.4 we can proceed as in the proof of Proposition 7.4, and in particular we can check that

$$\begin{aligned}
\mathbb{E}[f_X(X)] &= \frac{1}{1-p}\mathbb{E}\left[\mathbb{1}_{\{X>V_X^p\}} + \mathbb{1}_{\{\mathbb{P}(X=V_X^p)>0\}}\frac{1-p-\mathbb{P}\big(X>V_X^p\big)}{\mathbb{P}\big(X=V_X^p\big)}\mathbb{1}_{\{X=V_X^p\}}\right]\\
&= \frac{1}{1-p}\Big(\mathbb{E}\big[\mathbb{1}_{\{X>V_X^p\}}\big] + 1-p-\mathbb{P}\big(X>V_X^p\big)\Big)\\
&= \frac{1}{1-p}\big(\mathbb{P}\big(X>V_X^p\big) + 1-p-\mathbb{P}\big(X>V_X^p\big)\big)\\
&= 1.
\end{aligned} \tag{7.8}$$

□

By Lemma 6.18 and Proposition 7.11, we also have

$$\mathrm{ES}_X^p = \int_0^1 V_X^q f_X(V_X^q) dq,$$

and we check that the distortion function f_X of Proposition 7.11 is a non-decreasing function that satisfies

$$f_X(x) \leqslant \frac{1}{1-p}, \qquad x \in \mathbb{R},$$

by (6.8). The following proposition, see Acerbi and Tasche (2001), shows that in general, the Expected Shortfall at the level $p \in (0,1)$ coincides with the Tail Value at Risk TV_X^p.

Theorem 7.12. *The Expected Shortfall* ES_X^p *coincides with the Tail Value at Risk* TV_X^p *for any* $p \in (0,1)$*, i.e. we have*

$$\mathrm{ES}_X^p = \mathrm{TV}_X^p = \frac{1}{1-p} \int_p^1 V_X^q dq.$$

Proof. Constructing X as $X = V_X^U$ where U is uniformly distributed on $[0,1]$ as in Lemma 6.18, by Proposition 6.11 we have

$$U \geqslant p \Longrightarrow V_X^U \geqslant V_X^p \Longrightarrow X \geqslant V_X^p$$

and

$$\begin{aligned}
\big(U < p \text{ and } X \geqslant V_X^p\big) &\Longrightarrow \big(V_X^U \leqslant V_X^p \text{ and } X \geqslant V_X^p\big) \\
&\Longrightarrow \big(X \leqslant V_X^p \text{ and } X \geqslant V_X^p\big) \\
&\Longrightarrow X = V_X^p.
\end{aligned}$$

Hence by (7.6) and the relations

$$1 - p = \mathbb{E}\big[\mathbb{1}_{\{U \geqslant p\}}\big] \quad \text{and} \quad \mathbb{P}\big(X \geqslant V_X^p\big) = \mathbb{E}\big[\mathbb{1}_{\{X \geqslant V_X^p\}}\big],$$

we have

$$\begin{aligned}
V_X^p\big(1 - p - \mathbb{P}\big(X \geqslant V_X^p\big)\big) &= -V_X^p \mathbb{E}\big[\mathbb{1}_{\{X \geqslant V_X^p\}} - \mathbb{1}_{\{U \geqslant p\}}\big] \\
&= -V_X^p \mathbb{E}\big[\mathbb{1}_{\{X \geqslant V_X^p\} \setminus \{U \geqslant p\}}\big] \\
&= -V_X^p \mathbb{E}\big[\mathbb{1}_{\{X \geqslant V_X^p\} \cap \{U < p\}}\big] \\
&= -\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^p\} \cap \{U < p\}}\big],
\end{aligned}$$

hence

$$\begin{aligned}
\mathrm{ES}_X^p &= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] + \frac{V_X^p}{1-p}\big(1-p-\mathbb{P}\big(X\geqslant V_X^p\big)\big) \\
&= \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}}\big] - \frac{1}{1-p}\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^p\}\cap\{U<p\}}\big] \\
&= \frac{1}{1-p}\mathbb{E}\big[V_X^U\mathbb{1}_{\{V_X^U\geqslant V_X^p\}}\big] - \frac{1}{1-p}\mathbb{E}\big[V_X^U\mathbb{1}_{\{V_X^U\geqslant V_X^p\}}\mathbb{1}_{\{U<p\}}\big] \\
&= \frac{1}{1-p}\mathbb{E}\big[V_X^U\mathbb{1}_{\{V_X^U\geqslant V_X^p\}}\mathbb{1}_{\{U\geqslant p\}}\big] \\
&= \frac{1}{1-p}\mathbb{E}\big[V_X^U\mathbb{1}_{\{U\geqslant p\}}\big] \\
&= \frac{1}{1-p}\int_p^1 V_X^q dq,
\end{aligned}$$

which is the Tail Value at Risk TV_X^p. □

Theorem 7.13. *Expected Shortfall* ES_X^p *and Tail Value at Risk* TV_X^p *are coherent risk measures.*

Proof. Since the Expected Shortfall ES_X^p is a *distortion risk measure* by Proposition 7.11, we can conclude to positive homogeneity and translation invariance as done after Definition 6.4.

Alternatively, as ES_X^p coincides with TV_X^p for all $p \in (0,1)$ from Theorem 7.12, we can use Relation (7.1) in Definition 7.1 or Relation (7.6) in Definition 7.8 to proceed as follows.

(i) Monotonicity. If $X \leqslant Y$, since Value at Risk is monotone by Proposition 6.15, we have

$$\begin{aligned}
\mathrm{ES}_X^p &= \mathrm{TV}_X^p \\
&= \frac{1}{1-p}\int_p^1 V_X^q dq \\
&\leqslant \frac{1}{1-p}\int_p^1 V_Y^q dq \\
&= \mathrm{TV}_Y^p \\
&= \mathrm{ES}_Y^p
\end{aligned}$$

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

(ii) Homogeneity and translation invariance. Similarly, since Value at Risk is satisfies the homogeneity and translation invariance properties, for all $\mu \in \mathbb{R}$ and $\lambda > 0$ we have

$$\begin{aligned}
\mathrm{ES}^p_{\mu+\lambda X} &= \mathrm{TV}^p_{\mu+\lambda X} \\
&= \frac{1}{1-p}\int_p^1 V^q_{\mu+\lambda X} dq \\
&= \frac{1}{1-p}\int_p^1 (\mu+\lambda V^q_X) dq \\
&= \mu+\lambda\frac{1}{1-p}\int_p^1 V^q_X dq \\
&= \mu+\lambda \mathrm{TV}^p_Y \\
&= \mu+\lambda \mathrm{ES}^p_Y
\end{aligned}$$

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

(iii) Sub-additivity. By Proposition 7.11, we have

$$\begin{aligned}
&(1-p)\left(\mathrm{ES}^p_{X+Y} - \mathrm{ES}^p_X - \mathrm{ES}^p_Y\right) \\
&= \mathbb{E}[(X+Y)f_{X+Y}(X+Y)] - \mathbb{E}[Xf_X(X)] - \mathbb{E}[Yf_Y(Y)] \\
&= \mathbb{E}[X(f_{X+Y}(X+Y) - f_X(X))] + \mathbb{E}[Y(f_{X+Y}(X+Y) - f_Y(Y))] \\
&= V^p_X\mathbb{E}[f_{X+Y}(X+Y) - f_X(X)] + \mathbb{E}[(X-V^p_X)(f_{X+Y}(X+Y) - f_X(X))] \\
&\quad + V^p_Y\mathbb{E}[f_{X+Y}(X+Y) - f_Y(Y)] + \mathbb{E}[(Y-V^p_Y)(f_{X+Y}(X+Y) - f_Y(Y))] \\
&= (1-1)V^p_X + \mathbb{E}[(X-V^p_X)(f_{X+Y}(X+Y) - f_X(X))] \\
&\quad + (1-1)V^p_Y + \mathbb{E}[(Y-V^p_Y)(f_{X+Y}(X+Y) - f_Y(Y))] \\
&\leqslant 0,
\end{aligned}$$

where we have used (7.8) and the following facts.

- When $x - V^p_X < 0$, we have

$$\begin{aligned}
(1-p)(f_{X+Y}(x+y) - f_X(x)) &= \mathbb{1}_{\{x+y>V^p_{X+Y}\}} - \mathbb{1}_{\{x>V^p_X\}} \\
&\quad + \mathbb{1}_{\{\mathbb{P}(X+Y=V^p_{X+Y})>0\}}\frac{1-p-\mathbb{P}(X+Y>V^p_{X+Y})}{\mathbb{P}(X+Y=V^p_{X+Y})}\mathbb{1}_{\{x+y=V^p_{X+Y}\}} \\
&\quad - \mathbb{1}_{\{\mathbb{P}(X=V^p_X)>0\}}\frac{1-p-\mathbb{P}\big(X>V^p_X\big)}{\mathbb{P}\big(X=V^p_X\big)}\mathbb{1}_{\{x=V^p_X\}} \\
&= \mathbb{1}_{\{x+y>V^p_{X+Y}\}} + \mathbb{1}_{\{\mathbb{P}(X+Y=V^p_{X+Y})>0\}}\frac{1-p-\mathbb{P}(X+Y>V^p_{X+Y})}{\mathbb{P}(X+Y=V^p_{X+Y})}\mathbb{1}_{\{x+y=V^p_{X+Y}\}} \\
&\geqslant 0, \qquad x < V^p_X,
\end{aligned}$$

  where we applied (6.8).
- When $x - V^p_X > 0$, we have

$$\begin{aligned}
(1-p)(f_{X+Y}(x+y)-f_X(x)) &= \mathbb{1}_{\{x+y>V^p_{X+Y}\}} - \mathbb{1}_{\{x>V^p_X\}} \\
&\quad + \mathbb{1}_{\{\mathbb{P}(X+Y=V^p_{X+Y})>0\}} \frac{1-p-\mathbb{P}(X+Y>V^p_{X+Y})}{\mathbb{P}(X+Y=V^p_{X+Y})} \mathbb{1}_{\{x+y=V^p_{X+Y}\}} \\
&\quad - \mathbb{1}_{\{\mathbb{P}(X=V^p_X)>0\}} \frac{1-p-\mathbb{P}\big(X>V^p_X\big)}{\mathbb{P}\big(X=V^p_X\big)} \mathbb{1}_{\{x=V^p_X\}} \\
&= \mathbb{1}_{\{x+y>V^p_{X+Y}\}} - \mathbb{1}_{\{x>V^p_X\}} \\
&\quad + \mathbb{1}_{\{\mathbb{P}(X+Y=V^p_{X+Y})>0\}} \frac{1-p-\mathbb{P}(X+Y>V^p_{X+Y})}{\mathbb{P}(X+Y=V^p_{X+Y})} \mathbb{1}_{\{x+y=V^p_{X+Y}\}} \\
&\leqslant \mathbb{1}_{\{x+y>V^p_{X+Y}\}} - \mathbb{1}_{\{x>V^p_X\}} + \mathbb{1}_{\{x+y=V^p_{X+Y}\}} \\
&= \mathbb{1}_{\{x+y\geqslant V^p_{X+Y}\}} - \mathbb{1}_{\{x>V^p_X\}} \\
&\leqslant 0, \qquad x > V^p_X,
\end{aligned}$$

where we applied (6.8).

□

Note that in general, the Conditional Tail Expectation is not a coherent risk measure when $\mathbb{P}\big(X = V^p_X\big) > 0$.

7.4 Numerical Estimates

We are using the PerformanceAnalytics R package, see also § 6.1.1 of Mina and Xiao (2001). In case we care about negative return values, Definitions 7.3 and 7.8 are replaced with

$$\overline{\mathrm{CTE}}^p_X := \mathbb{E}\big[X \mid X < \overline{V}^p_X\big] = \frac{1}{\mathbb{P}\big(X < \overline{V}^p_X\big)} \mathbb{E}\big[X \mathbb{1}_{\{X<\overline{V}^p_X\}}\big]$$

and

$$\overline{\mathrm{ES}}^p_X := \overline{V}^p_X + \frac{1}{1-p} \mathbb{E}\big[\big(X - \overline{V}^p_X\big)\mathbb{1}_{\{X\leqslant\overline{V}^p_X\}}\big].$$

From Proposition 6.14, when the cumulative distribution function F_X is continuous and strictly increasing we have

$$\begin{aligned}
\overline{\mathrm{CTE}}^p_X &= \overline{\mathrm{ES}}^p_X \\
&= \mathbb{E}\big[X \mid X < \overline{V}^p_X\big] \\
&= \mathbb{E}\big[X \mid X < -V^{1-p}_X\big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mathbb{P}\big(X < -V_X^{1-p}\big)} \mathbb{E}\big[X \mathbb{1}_{\{X < -V_X^{1-p}\}}\big] \\
&= -\frac{1}{\mathbb{P}\big(-X > V_X^{1-p}\big)} \mathbb{E}\big[-X \mathbb{1}_{\{-X > V_X^{1-p}\}}\big] \\
&= -\mathbb{E}\big[-X \mid -X > V_X^{1-p}\big] \\
&= -\mathrm{CTE}_{-X}^{1-p} \\
&= -\mathrm{ES}_{-X}^{1-p}.
\end{aligned}
$$

```
library(PerformanceAnalytics)
ES(returns, p=.95, method="historical",invert="TRUE")
ES(returns, p=.95, method="historical",invert="FALSE")
```

The 95% historical Expected Shortfall is $\mathrm{ES}_X^{95\%} = -0.02087832$, and can be exactly recovered by the empirical Conditional Tail Expectation (CTE) as

```
mean(returns[returns<(VaR(returns, p=.95, method="historical")[1])],na.rm=TRUE)
```

The Gaussian Expected Shortfall is given as -0.0191359 by

```
ES(returns, p=.95, method="gaussian",invert="FALSE")
ES(returns, p=.95, method="gaussian",invert="TRUE")
```

It can be recovered from (7.5) (after sign inversion) as

$$
\begin{aligned}
\mathrm{ES}_X^p &= -\mathrm{ES}_{-X}^{1-p} \\
&= -\left(-\mu + \frac{\sigma}{1-p}\phi(V_Z^{1-p})\right) \\
&= \mu - \frac{\sigma}{1-p}\phi(V_Z^p) \\
&= \mu - \frac{\sigma}{(1-p)\sqrt{2\pi}} \mathrm{e}^{-(V_Z^p)^2/2},
\end{aligned}
$$

i.e.

```
q=qnorm(.95, mean=0, sd=1)
mu=mean(returns,na.rm=TRUE)
sigma=sd(returns,na.rm=TRUE)
mu-sigma*dnorm(q)/0.05
```

with output -0.01916536.

The attached **code 1** and **code 2** compute the Expected Shortfall from the practitioner and academic points of views, and compare their outputs to the that of the PerformanceAnalytics package, as illustrated in the next Figure 7.3.

```
> source("comparison.R")
Number of samples= 265
VaR 95 = -0.03420879 , Threshold= 0.9433962
CTE 95 = -0.04646176
ES 95 = -0.04623058
Historical VaR 95 0= -0.03316604
Gaussian VaR 95 = -0.03209374
Historical ES 95 = -0.04552403
Gaussian ES 95 = -0.04043227
```

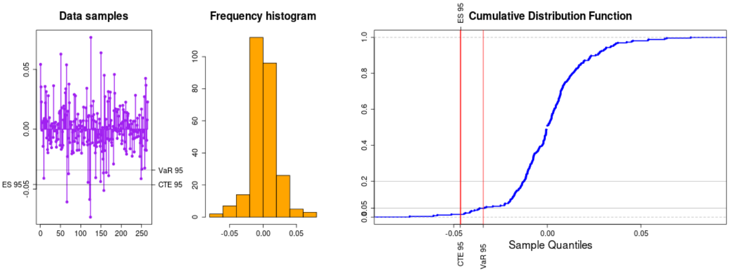


Fig. 7.3: Value at Risk and Expected Shortfall.

Value at Risk *vs.* Expected Shortfall

```
dev.new(width=16,height=8)
chart.VaRSensitivity(ts(returns),methods=c("HistoricalVaR","HistoricalES"),
        colorset=bluefocus, lwd=4)
```

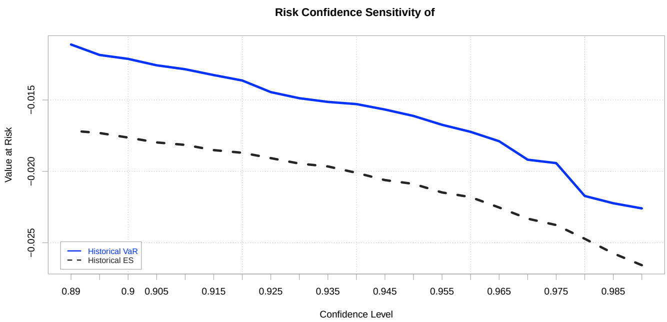


Fig. 7.4: Value at Risk *vs.* Expected Shortfall.

Historical *vs.* Gaussian risk measures

```
dev.new(width=16,height=8)
chart.VaRSensitivity(ts(returns),methods=c("HistoricalVaR","GaussianVaR"),
      colorset=bluefocus, lwd=4)
```

The next Figure 7.5 uses the above R code to compare the historical and Gaussian values at risk.

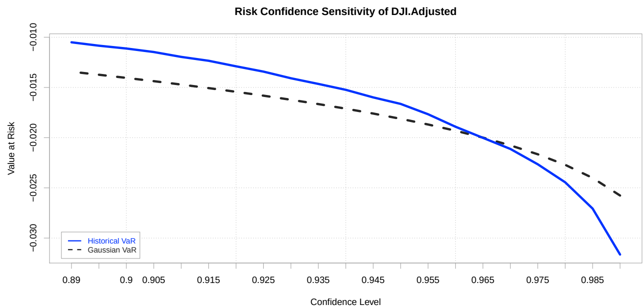


Fig. 7.5: Historical *vs.* Gaussian estimates of Value at Risk.

```
dev.new(width=16,height=8)
chart.VaRSensitivity(ts(returns),methods=c("HistoricalES","GaussianES"), colorset=bluefocus,
     lwd=4)
```

In the next Figure 7.6 we compare the Gaussian and historical estimates of Expected Shortfall.

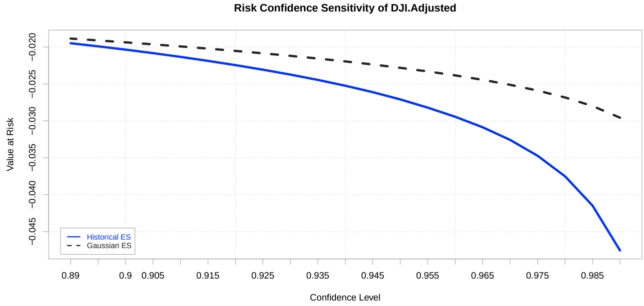


Fig. 7.6: Quantile function.

In Table 7.1 we summarize some properties of risk measures.

Risk Measure	Additivity	Homogeneity	Subadditivity	Coherence
V_X	✓	✓	✗	✗
CTE_X	✓	✓	✗	✗
TV_X	✓	✓	✓	✓
ES_X	✓	✓	✓	✓

Table 7.1: Summary of Risk Measures.

Note that Value at Risk V_X^p is *coherent* on Gaussian random variables according to Remark 6.17. Similarly, the Conditional Tail Expectation CTE_X^p

is *coherent* on random variables having a continuous CDF by Proposition 7.6 and Theorem 7.13.

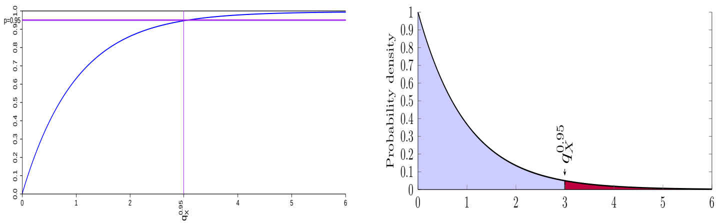
Exercises

Exercise 7.1 Let X denote an exponentially distributed random variable with parameter $\lambda > 0$, *i.e.* the distribution of X has the cumulative distribution function (CDF)

$$F_X(x) = \mathbb{P}(X \leqslant x) = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geqslant 0,$$

and the probability density function (PDF)

$$f_X(x) = F_X'(x) = \lambda \mathrm{e}^{-\lambda x}, \qquad x \geqslant 0.$$



(a) Exponential quantile and CDF.

(b) Exponential PDF.

a) Compute the conditional tail expectation

$$\mathbb{E}[X \mid X > \mathrm{VaR}_X^p] = \frac{1}{\mathbb{P}(X > \mathrm{VaR}_X^p)} \int_{\mathrm{VaR}_X^p}^{\infty} x f_X(x) dx.$$

b) Compute the tail value at risk

$$\mathrm{TV}_X^p = \frac{1}{1-p} \int_p^1 V_X^q dq.$$

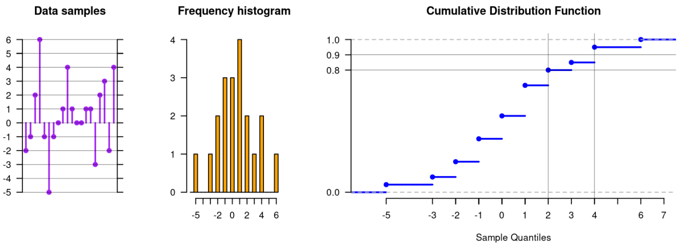
Exercise 7.2 Consider X an (integrable) random variable and $z \in \mathbb{R}$ such that $\mathbb{P}(X > z) > 0$.

a) Show that $\mathbb{E}[X \mid X > z] > z$.
b) Show that $\mathbb{E}[X \mid X > z] \geqslant \mathbb{E}[X]$.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

c) Show that $\mathbb{E}[X \mid X > z] > \mathbb{E}[X]$ if $\mathbb{P}(X \leqslant z) > 0$.
d) Show that $\mathrm{CTE}_X^p > \mathbb{E}[X]$.

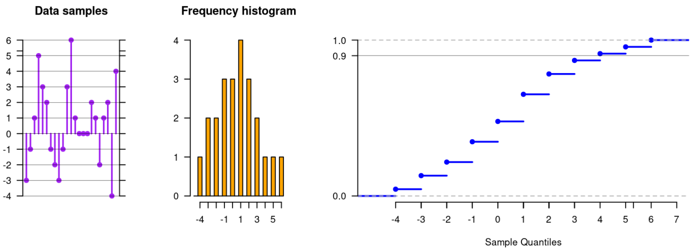
Exercise 7.3 Consider the following data set.



Find the Value at Risk VaR_X^p and the Conditional Tail Expectation $\mathrm{CTE}_X^p = \mathbb{E}\big[X \mid X > \mathrm{VaR}_X^p\big]$ and mark their values on the graph in the following cases.

a) $p = 0.9$.
b) $p = 0.8$.

Exercise 7.4



Let $p = 0.9$. For the above data set represented by the random variable X, compute the numerical values of the following quantities.

a) VaR_X^{90},
b) $\mathbb{E}\big[X \mathbb{1}_{\{X > V_X^{90}\}}\big]$,
c) $\mathbb{P}\big(X > V_X^{90}\big)$,
d) $\mathrm{CTE}_X^{90} = \mathbb{E}\big[X \mid X > V_X^{90}\big] = \mathbb{E}\big[X \mathbb{1}_{\{X > V_X^{90}\}}\big] / \mathbb{P}\big(X > V_X^{90}\big)$,
e) $\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^{90}\}}\big]$,

f) $\mathbb{P}\big(X \geqslant V_X^{90}\big)$,

g) $\mathrm{ES}_X^{90} = \dfrac{1}{1-p}\big(\mathbb{E}\big[X\mathbb{1}_{\{X\geqslant V_X^{90}\}}\big] + V_X^{90}\big(1-p-\mathbb{P}\big(X \geqslant V_X^{90}\big)\big)\big)$,

h) $\mathrm{TV}_X^{90} = \dfrac{1}{1-p}\displaystyle\int_p^1 V_X^q dq$,

and mark the values of $\mathrm{VaR}_X^{90\%}$, $\mathrm{CTE}_X^{90\%}$, $\mathrm{ES}_X^{90\%}$, $\mathrm{TV}_X^{90\%}$ on the above graph.

Exercise 7.5 Consider a random variable $X \in \{10, 100, 150\}$ with the distribution

$$\mathbb{P}(X = 10) = 96\%, \quad \mathbb{P}(X = 100) = 3\%, \quad \mathbb{P}(X = 150) = 1\%.$$

Compute

a) the Value at Risk $V_X^{98\%}$,

b) the Tail Value at Risk $\mathrm{TV}_X^{98\%}$,

c) the Conditional Tail Expectation $\mathbb{E}\big[X \mid X > V_X^{98\%}\big]$, and

d) the Expected Shortfall $E_X^{98\%}$.

Exercise 7.6 Consider two independent random variables X and Y with same distribution given by

$$\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = 90\% \quad \text{and} \quad \mathbb{P}(X = 100) = \mathbb{P}(Y = 100) = 10\%.$$

a) Plot the cumulative distribution function of X on the following graph:

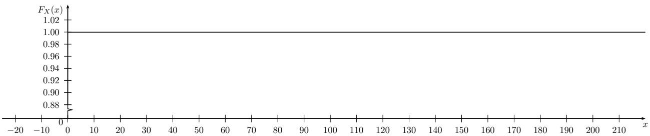


Fig. 7.8: Cumulative distribution function of X.

b) Plot the cumulative distribution function of $X + Y$ on the following graph:

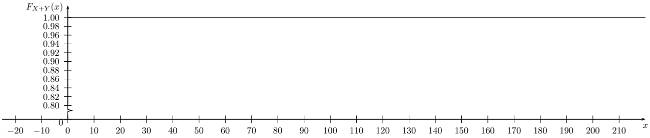


Fig. 7.9: Cumulative distribution function of $X+Y$.

c) Give the values at risk $V_{X+Y}^{99\%}$, $V_{X+Y}^{95\%}$, $V_{X+Y}^{90\%}$.
d) Compute the *Tail Value at Risk*

$$\mathrm{TV}_X^{90\%} := \frac{1}{1-p}\int_p^1 V_X^q dq$$

at the level $p = 90\%$.
e) Compute the *Tail Value at Risk*

$$\mathrm{TV}_{X+Y}^{p} := \frac{1}{1-p}\int_p^1 V_{X+Y}^q dq$$

at the levels $p = 90\%$ and $p = 80\%$.

Exercise 7.7 (Exercise 6.2 continued).

a) Compute the Tail Value at Risk

$$\mathrm{TV}_X^{p} := \frac{1}{1-p}\int_p^1 V_X^q dq$$

for all p in the interval $[0.99, 1]$, and give the value of $\mathrm{TV}_X^{99\%}$.
b) Taking $p = 0.98$, compute the Conditional Tail Expectation

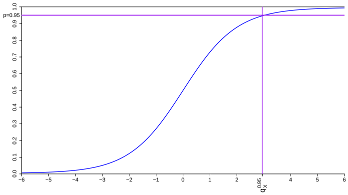
$$\mathrm{CTE}_X^{98\%} = \mathbb{E}\big[X \mid X > V_X^{98\%}\big] = \frac{1}{\mathbb{P}\big(X > V_X^p\big)}\mathbb{E}\left[X\mathbb{1}_{\{X>V_X^p\}}\right].$$

Exercise 7.8 We assume that the payoff X of a portfolio follows the standard logistic distribution with cumulative distribution function (CDF)

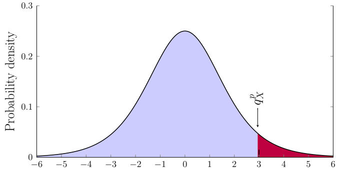
$$F_X(x) = \mathbb{P}(X \leqslant x) = \frac{1}{1+\mathrm{e}^{-x}}, \qquad x \in \mathbb{R},$$

and the probability density function (PDF)

$$f_X(x) = F_X'(x) = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2}, \qquad x \in \mathbb{R}.$$



(a) Logistic quantile and CDF.



(b) Logistic PDF.

a) Compute the quantile $q_X^p = \mathrm{VaR}_X^p$ of X at any level $p \in [0,1]$, defined by the relation

$$F_X(q_X^p) = \mathbb{P}(X \leqslant \mathrm{VaR}_X^p) = p.$$

b) Compute the conditional tail expectation

$$\mathbb{E}[X \mid X > \mathrm{VaR}_X^p] = \frac{1}{\mathbb{P}(X > \mathrm{VaR}_X^p)} \int_{\mathrm{VaR}_X^p}^\infty x f_X(x) dx.$$

Hint. We have

$$\int_a^\infty \frac{x\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2} dx = \log(1+\mathrm{e}^a) - \frac{a\mathrm{e}^a}{1+\mathrm{e}^a}, \qquad a \in \mathbb{R}.$$

c) Compute the tail value at risk

$$\mathrm{TV}_X^p = \frac{1}{1-p} \int_p^1 V_X^q dq.$$

Hint. We have $\int_p^1 \log q dq = p - 1 - p \log p$, $p \in (0,1)$.

Exercise 7.9

a) Show that for any random variable Z with probability density function $f_Z : \mathbb{R} \to \mathbb{R}_+$ we have

$$q\mathbb{P}(Z \geqslant q) \leqslant \mathbb{E}\big[Z \mathbb{1}_{\{Z \geqslant q\}}\big] = \int_q^\infty x f_Z(x) dx, \qquad q \geqslant 0. \tag{7.9}$$

b) Compute the left hand side and right hand side of the inequality (7.9) when $Z \simeq \mathcal{N}(0,1)$ has the standard normal distribution and q is the quantile q_Z^p of Z at the level $p \in [0,1]$.

c) Given $X \simeq \mathcal{N}(\mu_X, \sigma_X^2)$ a Gaussian random variable with mean μ_X and variance σ_X^2, show that the Gaussian Value at Risk

$$V_X^p = \mu_X + \sigma_X q_Z^p$$

is upper bounded by the Gaussian conditional tail expectation

$$\mathrm{CTE}_X^p = \mu_X + \frac{\sigma_X}{1-p}\phi(q_Z^p)$$

where

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}, \qquad z \in \mathbb{R},$$

is the standard normal probability density function.

Part III
Credit Risk

Chapter 8
Credit Scoring

Credit scoring provides a statistical assessment of a borrower's creditworthiness that helps financial institutions in making decisions on loan applications. In this chapter, we review the uses of discriminant analysis, and binomial logistic regression, with application to credit scoring. We also cover the properties of receiver operating characteristics curves.

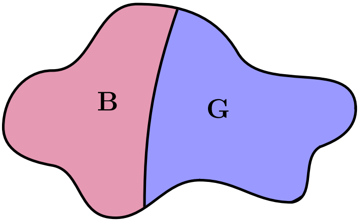
8.1 Discriminant Analysis

Consider a set Ω of credit applicants which is partitioned as

$$\Omega = B \cup G$$

into a subset G of "good" (or solvent) applicants, and a subset B of "bad" (or insolvent) applicants, with $B \cap G = \emptyset$. Applicants are selected at random within the set Ω according to a probability distribution $\mathbb{P}$, so that

$$\mathbb{P}(B) + \mathbb{P}(G) = 1.$$



Here, $\mathbb{P}(B)$ represents the probability that an applicant chosen at random may default, while $\mathbb{P}(G)$ represents the probability that a randomly selected applicant is solvent. In addition, each applicant $\omega \in \Omega$ is assigned a real-valued rating (or score) $X(\omega)$ via a *random variable**

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

with probability density function $f_X : \mathbb{R} \to \mathbb{R}_+$.

Definition 8.1. *1) The function*

$$\begin{aligned} \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto \mathbb{P}(B \mid X = x) \end{aligned}$$

is respectively called the probability default curve*.*

2) The function

$$\begin{aligned} \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto \mathbb{P}(G \mid X = x) \end{aligned}$$

is called the probability acceptance *curve.*

Denoting by $f_X(x \mid B)$, resp. $f_X(x \mid G)$, the probability density function of X given B, resp. X given G, we have

$$f_X(x) = f_X(x \mid G)\mathbb{P}(G) + f_X(x \mid B)\mathbb{P}(B), \quad x \in \mathbb{R},$$

and the Bayes formula yields

$$\mathbb{P}(B \mid X = x) = f_X(x \mid B)\frac{\mathbb{P}(B)}{f_X(x)}$$

* See (MOE and UCLES 2022, page 14 lines 4-5) and (MOE and UCLES 2020, page 19 lines 4-5).

$$= \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}, \qquad (8.1)$$

and

$$\mathbb{P}(G \mid X = x) = \frac{\mathbb{P}(G) f_X(x \mid G)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}.$$

Definition 8.2. *1) The True Positive Rate (TPR) is given by the tail distribution function*

$$\overline{F}_G(x) := \mathbb{P}(X > x \mid G) = \int_x^\infty f_X(y \mid G) dy, \qquad x \in \mathbb{R}.$$

2) The False Positive Rate (FPR) is given by the tail distribution function

$$\overline{F}_B(x) := \mathbb{P}(X > x \mid B) = \int_x^\infty f_X(y \mid B) dy, \qquad x \in \mathbb{R}.$$

Example. In case X is Gaussian distributed given $\{G, B\}$ with the conditional densities

$$f_X(x \mid G) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)} \qquad (8.2)$$

and

$$f_X(x \mid B) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}, \qquad (8.3)$$

with $\mu_B < \mu_G$, we have

$$f_X(x) = \frac{\mathbb{P}(G)}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)} + \frac{\mathbb{P}(B)}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}, \quad x \in \mathbb{R},$$

and by (8.1) we find the logistic probability default curve

$$\begin{aligned} \mathbb{P}(B \mid X = x) &= \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)} \\ &= \frac{\mathbb{P}(B) \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}}{\mathbb{P}(G) \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)} + \mathbb{P}(B) \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}} \\ &= \frac{1}{1 + \mathrm{e}^{\alpha + \beta x}}, \qquad x \in \mathbb{R}, \qquad (8.4) \end{aligned}$$

where we let

$$\beta := \frac{\mu_G - \mu_B}{\sigma^2} > 0$$

and

$$\alpha := -\beta \frac{\mu_G + \mu_B}{2} + \log \frac{\mathbb{P}(G)}{\mathbb{P}(B)}.$$

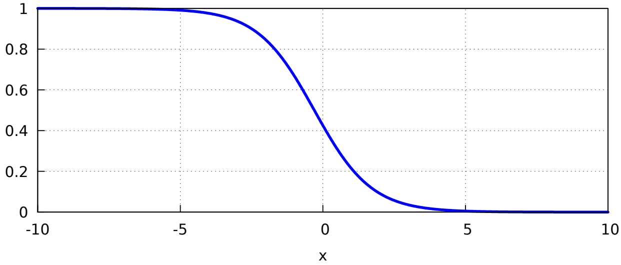


Fig. 8.1: Probability default curve $x \mapsto \mathbb{P}(B \mid X = x) = 1/(1 + e^{a+x})$.

Similarly, the probability acceptance curve is given by

$$\begin{aligned}\mathbb{P}(G \mid X = x) &= \frac{\mathbb{P}(G) f_X(x \mid G)}{\mathbb{P}(G) f_X(x \mid G) + (1 - \mathbb{P}(G)) f_X(x \mid B)} \\ &= \frac{\mathbb{P}(G) e^{-(x-\mu_G)^2/(2\sigma^2)}}{\mathbb{P}(G) e^{-(x-\mu_G)^2/(2\sigma^2)} + \mathbb{P}(B) e^{-(x-\mu_B)^2/(2\sigma^2)}} \\ &= \frac{1}{1 + e^{-\alpha-\beta x}} \\ &= \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}, \qquad x \in \mathbb{R}. \end{aligned} \tag{8.5}$$

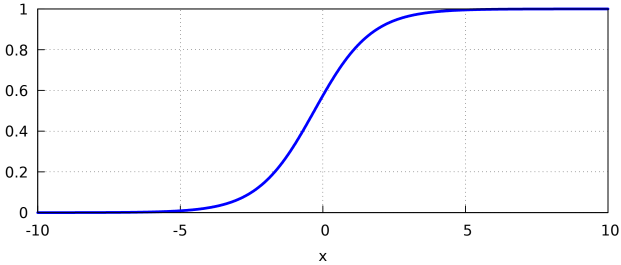


Fig. 8.2: Probability acceptance curve $x \mapsto \mathbb{P}(G \mid X = x) = e^{a+x}/(1 + e^{a+x})$.

8.2 Decision Rule

We decide to accept the applicants whose score $X(\omega)$ belongs to a decision (or acceptance) set $\mathscr{A} \subset \mathbb{R}$, and to reject those whose score $X(\omega)$ belongs to the rejection set $\mathscr{A}^c = \mathbb{R} \setminus \mathscr{A}$ which is the complement of $\mathscr{A}$ in $\mathbb{R}$.

Definition 8.3. *Let*

- $D(x)$ *represents the cost incurred by the default of an applicant with score* $x \in \mathscr{A}$, *and*
- $L(x)$ *represents the loss (or missed earnings) incurred by the rejection of an applicant with score* $x \in \mathscr{A}^c$.

The cost associated to this decision rule is defined as

$$\underbrace{D(X)\mathbb{1}_{\{X\in\mathscr{A}\}\cap B}}_{\text{Cost of accepting a "bad" applicant}} + \underbrace{L(X)\mathbb{1}_{\{X\in\mathscr{A}^c\}\cap G}}_{\text{Loss from rejecting a "good" applicant}}.$$

Theorem 8.4. *The optimal acceptance set* $\mathscr{A}^* \subset \mathbb{R}$ *that minimizes the expected cost*

$$\mathbb{E}\left[D(X)\mathbb{1}_{\{X\in\mathscr{A}\}\cap B} + L(X)\mathbb{1}_{\{X\in\mathscr{A}^c\}\cap G}\right]$$

is given by

$$\mathscr{A}^* = \left\{x \in \mathbb{R} \ : \ \lambda(x) \geqslant \frac{D(x)}{L(x)}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right\},$$

where $\lambda(x)$ *is the* likelihood ratio

$$\lambda(x) := \frac{f_X(x \mid G)}{f_X(x \mid B)} = \frac{\mathbb{P}(G \mid X = x)}{\mathbb{P}(B \mid X = x)}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}, \qquad x \in \mathbb{R}.$$

Proof. The expected cost corresponding to an acceptance set $\mathscr{A} \subset \mathbb{R}$ can be written as

$$\begin{aligned}
&\mathbb{E}\left[D(X)\mathbb{1}_{\{X\in\mathscr{A}\}\cap B} + L(X)\mathbb{1}_{\{X\in\mathscr{A}^c\}\cap G}\right]\\
&= \int_{\mathscr{A}} D(x)\mathbb{P}(B \cap \{X \in dx\}) + \int_{\mathscr{A}^c} L(x)\mathbb{P}(G \cap \{X \in dx\})\\
&= \int_{\mathscr{A}} D(x)\mathbb{P}(B \mid X = x) f_X(x)dx + \int_{\mathscr{A}^c} L(x)\mathbb{P}(G \mid X = x) f_X(x)dx\\
&= \int_{-\infty}^{\infty} \big(\mathbb{1}_{\mathscr{A}}(x)D(x)\mathbb{P}(B \mid X = x) + \mathbb{1}_{\mathscr{A}^c}(x)L(x)\mathbb{P}(G \mid X = x)\big) f_X(x)dx.
\end{aligned}$$

The expected cost can be minimized pointwise by finding the set $\mathscr{A}$ that minimizes the conditional expected cost

$$x \longmapsto \mathbb{E}\left[D(X)\mathbb{1}_{\{X\in\mathscr{A}\}\cap B} + L(X)\mathbb{1}_{\{X\in\mathscr{A}^c\}\cap G} \,\middle|\, X = x\right]$$

$$= \mathbb{1}_{\mathscr{A}}(x)D(x)\mathbb{P}(B \mid X = x) + \mathbb{1}_{\mathscr{A}^c}(x)L(x)\mathbb{P}(G \mid X = x)$$

given that $X = x$. For this, we note that

$$\begin{aligned}
&\mathbb{1}_{\mathscr{A}}(x)D(x)\mathbb{P}(B \mid X = x) + \mathbb{1}_{\mathscr{A}^c}(x)L(x)\mathbb{P}(G \mid X = x)\\
&\quad \geqslant \min(D(x)\mathbb{P}(B \mid X = x), L(x)\mathbb{P}(G \mid X = x))\\
&\quad = D(x)\mathbb{P}(B \mid X = x)\mathbb{1}_{\{D(x)\mathbb{P}(B|X=x)\leqslant L(x)\mathbb{P}(G|X=x)\}}\\
&\qquad + L(x)\mathbb{P}(G \mid X = x)\mathbb{1}_{\{L(x)\mathbb{P}(G|X=x)<D(x)\mathbb{P}(B|X=x)\}}\\
&\quad = D(x)\mathbb{P}(B \mid X = x)\mathbb{1}_{\mathscr{A}^*}(x) + L(x)\mathbb{P}(G \mid X = x)\mathbb{1}_{(\mathscr{A}^*)^c}(x),
\end{aligned}$$

where the set $\mathscr{A}^*$ which achieves equality in the above inequality is given by

$$\begin{aligned}
\mathscr{A}^* &:= \{D(x)\mathbb{P}(B \mid X = x) \leqslant L(x)\mathbb{P}(G \mid X = x)\}\\
&= \left\{x \in \mathbb{R} \ : \ \frac{D(x)}{L(x)} \leqslant \frac{\mathbb{P}(G \mid X = x)}{\mathbb{P}(B \mid X = x)}\right\}.
\end{aligned}$$

In addition, the optimal acceptance set $\mathscr{A}^*$ can be rewritten in terms of the *likelihood ratio* function

$$\lambda(x) := \frac{f_X(x \mid G)}{f_X(x \mid B)} = \frac{\mathbb{P}(G \mid X = x)}{\mathbb{P}(B \mid X = x)}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}, \qquad x \in \mathbb{R},$$

as

$$\mathscr{A}^* = \left\{x \in \mathbb{R} \ : \ \lambda(x) \geqslant \frac{D(x)}{L(x)}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right\}.$$

□

For simplicity, in the sequel we assume that $D = D(x)$ and $L = L(x)$ are constant in $x \in \mathbb{R}$, in which case we have

$$\mathscr{A}^* = \left\{x \in \mathbb{R} \ : \ \lambda(x) \geqslant \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right\}.$$

Proposition 8.5. *In the Gaussian example* (8.2)-(8.3) *with* $\mu_B < \mu_G$, *the optimal acceptance set* $\mathscr{A}^*$ *is given by*

$$\mathscr{A}^* = \left[\frac{\mu_G + \mu_B}{2} + \frac{1}{\beta}\log\left(\frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right), \infty\right), \tag{8.6}$$

where

$$\beta := \frac{\mu_G - \mu_B}{\sigma^2} > 0,$$

under the condition

$$\mathbb{E}[X \mid B] = \mu_B < \mu_G = \mathbb{E}[X \mid G].$$

Proof. In the Gaussian example (8.2)-(8.3), the likelihood ratio is given by

$$\begin{aligned}\lambda(x) &= \frac{f_X(x \mid G)}{f_X(x \mid B)} \\ &= \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)+(x-\mu_B)^2/(2\sigma^2)} \\ &= \mathrm{e}^{-(\mu_G^2-\mu_B^2-2x(\mu_G-\mu_B))/(2\sigma^2)} \\ &= \mathrm{e}^{\beta x-(\mu_G^2-\mu_B^2)/(2\sigma^2)}, \qquad x \in \mathbb{R},\end{aligned}$$

hence the condition

$$\lambda(x) = \mathrm{e}^{\beta x-(\mu_G^2-\mu_B^2)/(2\sigma^2)} \geqslant \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}$$

is equivalent to

$$\begin{aligned}x &\geqslant \frac{\mu_G^2-\mu_B^2}{2\beta\sigma^2} + \frac{1}{\beta}\log\left(\frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right) \\ &\geqslant \frac{\mu_G^2-\mu_B^2}{2(\mu_G-\mu_B)} + \frac{1}{\beta}\log\left(\frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right) \\ &\geqslant \frac{\mu_G+\mu_B}{2} + \frac{1}{\beta}\log\left(\frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right) \\ &=: x^*,\end{aligned}$$

hence $\mathscr{A}^* = [x^*, \infty)$, provided that

$$\beta := \frac{\mu_G-\mu_B}{\sigma^2} > 0,$$

which yields (8.6). □

We note that the optimal boundary point x^* satisfies the relation

$$\lambda(x^*) = \frac{\mathbb{P}(X = x^* \mid G)}{\mathbb{P}(X = x^* \mid B)} = \frac{\mathbb{P}(G \mid X = x^*)}{\mathbb{P}(B \mid X = x^*)}\frac{\mathbb{P}(B)}{\mathbb{P}(G)} = \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)},$$

i.e.

$$\frac{\mathbb{P}(G \mid X = x^*)}{\mathbb{P}(B \mid X = x^*)} = \frac{D}{L}. \tag{8.7}$$

Figure 8.3 illustrates the optimal decision rule by taking $L = D = \$1$ and using the default and acceptance curves (8.4)-(8.5).

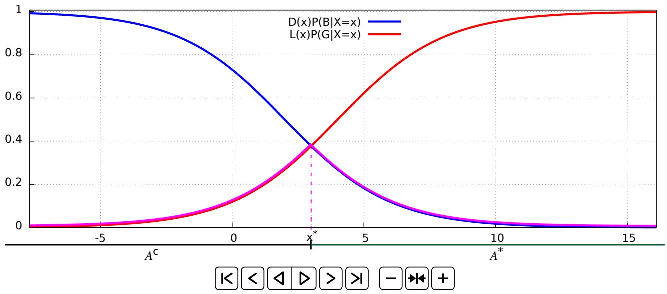


Fig. 8.3: Animated graph of optimal decision rule.*

In Figure 8.3, the conditional expected cost function

$$x \longmapsto \mathbb{1}_{\mathscr{A}}(x) D(x) \mathbb{P}(B \mid X = x) + \mathbb{1}_{\mathscr{A}^c}(x) L(x) \mathbb{P}(G \mid X = x)$$

is represented by the purple curve for a given set $\mathscr{A}$. Its uniform minimum over different threshold values is obtained for $\mathscr{A}^*$ of the form $\mathscr{A}^* = [x^*, \infty)$ where x^* lies at the intersection of the curves $x \mapsto D(x)\mathbb{P}(B \mid X = x)$ and $x \mapsto L(x)\mathbb{P}(G \mid X = x)$ as in (8.7).

The acceptance rate, or probability that an applicant is accepted according to the rule $\mathscr{A}$, is given by

$$\begin{aligned}\mathbb{P}(X \in \mathscr{A}) &= \mathbb{P}(\{X \in \mathscr{A}\} \cap G) + \mathbb{P}(\{X \in \mathscr{A}\} \cap B) \\ &= \mathbb{P}(X \in \mathscr{A} \mid G)\mathbb{P}(G) + \mathbb{P}(X \in \mathscr{A} \mid B)\mathbb{P}(B),\end{aligned}$$

where

$$\mathbb{P}(\{X \in \mathscr{A}\} \cap B) = \mathbb{P}(X \in \mathscr{A} \mid B)\mathbb{P}(B)$$

is the default rate, or probability that an applicant accepted according to the rule $\mathscr{A}$ will default.

We can also minimize the default rate $\mathbb{P}(\{X \in \mathscr{A}\} \cap B)$ subject to a given acceptance rate $\mathbb{P}(X \in \mathscr{A}) = a$.

8.3 Logistic Regression

In this section, we address the problem of constructing the random score variable $X : \Omega \to \mathbb{R}$ in a concrete setting. For this, consider a set of m

* The animation works in Acrobat Reader on the entire pdf file.

financial criteria or indicators $(x_{i,j})_{j=1,2,\ldots,m}$ applying to each of n credit applicants $i = 1, 2, \ldots, n$, with $\mathbb{P}(G) = 0.7$ and $\mathbb{P}(B) = 0.3$ in this data set.

```
install.packages("caret")
library(caret); data(GermanCredit); head(GermanCredit)
ggplot(GermanCredit, aes(x = Class)) + geom_bar(aes(y = (..count..)/sum(..count..)),fill=c(
      "red","darkgreen")) + labs(y = "prob.") + theme_bw()
```

j	Age	ForeignWorker	Property.RealEstate	Housing.Own	CreditHistory	Class
$x_{i,j}$	$x_{i,1} = 67$	$x_{i,2} = 1$	$x_{i,3} = 0$	$x_{i,4} = 1$	$x_{i,5} = 0$	$c_i = 1$

The credit scoring class ("good" or "bad") of applicant n°i is denoted by $c_i \in \{0, 1\}$ depending on his status, *i.e.* $c_i = 1$ for "good" applicants and $c_i = 0$ for "bad" applicants.

Linear regression

The score z_i of a given credit applicant in row n^o i is modeled as

$$z_i = F\left(\gamma_0 + \sum_{j=1}^{m} \gamma_j x_{i,j}\right), \qquad i = 1, 2, \ldots, n.$$

where $(\gamma_j)_{j=0,1,\ldots,m}$ is a family of linear coefficients, where $p_i := 1 - z_i$ represents the probability that applicant n^o i may default. In a linear regression model we would take $F(z) := z$, hence the system of equations

$$c_i = \gamma_0 + \sum_{j=1}^{m} \gamma_j x_{i,j}, \qquad i = 1, 2, \ldots, n,$$

would be used to estimate the coefficients $(\gamma_j)_{j=0,1,\ldots,m}$.

Binomial logistic regression

The shortcoming of linear models is that $F(z_i)$, which is assumed to represent a probability value, may exit the interval $[0, 1]$. Logistic regression models address this issue by replacing $F(x) = x$ with the logistic CDF F_L defined as

$$F_L(x) := \frac{\mathrm{e}^x}{1 + \mathrm{e}^x}, \qquad x \in \mathbb{R},$$

see also the Gaussian probability acceptance curve (8.5).

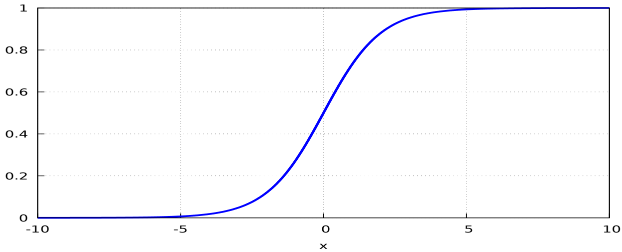


Fig. 8.4: Logistic CDF $x \mapsto F_L(x) = \mathrm{e}^x/(1+\mathrm{e}^x)$.

We model the probability $p_i = \mathbb{P}(i \in G)$ that applicant n^o i is rated "good" as

$$p_i = \mathbb{P}(i \in G) = \mathbb{P}\left(G \,\middle|\, X = \gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j} \right) = F_L\left(\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j} \right)$$

as in (8.5), $1 \leqslant i \leqslant n$. The probability of sampling a given $\{0,1\}$-valued sequence $(c_i)_{1\leqslant i \leqslant n}$ of applicant classifications is

$$\prod_{\substack{1\leqslant i\leqslant n \\ c_i=0}} \mathbb{P}(i \in B) \prod_{\substack{1\leqslant i\leqslant n \\ c_i=1}} \mathbb{P}(i \in G) = \prod_{i=1}^n \left((\mathbb{P}(i \in B))^{1-c_i} (\mathbb{P}(i \in G))^{c_i} \right).$$

In order to estimate the sequence of coefficients $(\gamma_j)_{j=0,1,\ldots,m}$, we aim at maximizing the log-likelihood ratio

$$\begin{aligned}
\log L(\beta|x) &:= \log \prod_{i=1}^n \left((\mathbb{P}(i \in B))^{1-c_i} (\mathbb{P}(i \in G))^{c_i} \right) \\
&= \log \prod_{i=1}^n \left(\left(\overline{F}_L \left(\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j} \right) \right)^{c_i} \left(F_L \left(\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j} \right) \right)^{1-c_i} \right) \\
&= \log \prod_{i=1}^n \left(\left(\frac{\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}}{1+\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}} \right)^{c_i} \left(\frac{1}{1+\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}} \right)^{1-c_i} \right) \\
&= \sum_{i=1}^n c_i \log \frac{\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}}{1+\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}} + \sum_{i=1}^n (1-c_i) \log \frac{1}{1+\mathrm{e}^{\gamma_0 + \sum_{j=1}^m \gamma_j x_{i,j}}}
\end{aligned}$$

over $\beta = (\gamma_j)_{j=0,1,\ldots,m}$ on the training set. The default probabilities are then estimated on the testing set from

$$1 - p_i = \overline{F}_L\left(\gamma_0 + \sum_{j=1}^{m} \gamma_j x_{i,j}\right), \qquad i = 1, \ldots, n,$$

and the *logit*

$$\overline{F}_L^{-1}(p_i) = -\log \frac{p_i}{1 - p_i} = \gamma_0 + \sum_{j=1}^{m} \gamma_j x_{i,j}, \qquad i = 1, \ldots, n,$$

or log-odds, represents probabilities on a logit scale. An implementation example of logistic regression is presented below using synthetic data.

```
y <- rnorm(15,1)
y<-(y-min(y))/(max(y)-min(y))
x <- rbinom(15,1,prob=y*y)
y<-data.frame(y)
glmreg<-glm(x ~ y,data=y,family="binomial")
xx<-seq(0,1,0.01)
xx<-data.frame(xx)
colnames(xx) <- c('y')
pred<-predict(glmreg,newdata=xx,type="response")
plot(y$y,x,col="blue")
lines(xx$y,pred,col="purple",lw=3)
points(y$y,((y$y-min(y$y))/(max(y$y)-min(y$y)))^2,col="red",lw=2)
```

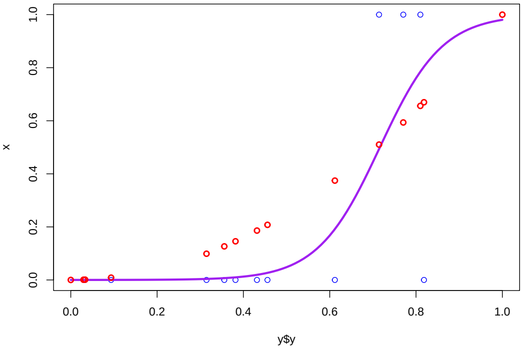


Fig. 8.5: GLM Regression.

```
Train <- createDataPartition(GermanCredit$Class, p=0.6, list=FALSE)
training <- GermanCredit[ Train, ];testing <- GermanCredit[ -Train, ]
```

The data is randomly split into a training set and a testing set using the createDataPartition command in R. The training set is used to fit the data in a generalized linear model using the glm() command. The testing set is then used to estimate the corresponding default probabilities.

```
mod.glm<-glm(Class ~ Age + ForeignWorker + Property.RealEstate + Housing.Own +
       CreditHistory.Critical, data=training, family="binomial")
head(testing$Class); head(predict(mod.glm, newdata=testing, type="response"))
testing <- cbind(rownames(testing), testing); colnames(testing)[1] <- "ID"
testing$Score5<-predict(mod.glm, newdata=testing, type="response")
testsamp <- head(testing,100);colnames(testsamp) <- make.unique(names(testsamp))
ggplot(testsamp, aes(x=ID, y=Score5, fill=Class)) + geom_bar(stat="identity") +
       scale_fill_manual(values=c( "red","darkgreen")) +
       scale_y_continuous(limits=c(0,1),expand = c(0,0)) + theme_bw(base_size = 18) +
       xlab(NULL) + theme(axis.text.x = element_blank(),aspect.ratio=0.5) +
       geom_hline(yintercept = 0.75,lwd=1.6)
```

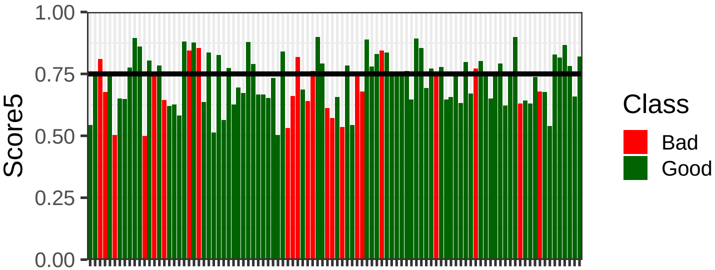


Fig. 8.6: Logistic regression output on 5 criteria.

Based on the above 100 samples, the next R code identifies the True Positive Rate (TPR) $= 34/77 = 44.16\%$, and the False Positive Rate (FPR) $= 8/23 = 34.78\%$, at the threshold $p = 0.75$.

```
cat('TPR5=',length(testsamp$Score5[testsamp$Score5>0.75 & testsamp$Class=="Good"])
       /length(testsamp$Score5[testsamp$Class=="Good"]),'\n')
cat('FPR5=',length(testsamp$Score5[testsamp$Score5>0.75 & testsamp$Class=="Bad"])
       /length(testsamp$Score5[testsamp$Class=="Bad"]),'\n')
pred <- ifelse(testsamp$Score5 > 0.75, "Good", "Bad")
confusionMatrix(factor(noquote(pred)),factor(testsamp$Class))
```

```
mod.glm<-glm(Class ~ Duration + Amount + InstallmentRatePercentage + ResidenceDuration
      + Age + NumberExistingCredits + NumberPeopleMaintenance + Telephone +
      ForeignWorker + CheckingAccountStatus.lt.0 + CheckingAccountStatus.0.to.200 +
      CheckingAccountStatus.gt.200 + CheckingAccountStatus.none +
      CreditHistory.NoCredit.AllPaid + CreditHistory.ThisBank.AllPaid +
      CreditHistory.PaidDuly + CreditHistory.Delay + CreditHistory.Critical +
      Purpose.NewCar + Purpose.UsedCar + Purpose.Furniture.Equipment +
      Purpose.Radio.Television + Purpose.DomesticAppliance + Purpose.Repairs +
      Purpose.Education + Purpose.Vacation + Purpose.Retraining + Purpose.Business +
      Purpose.Other + SavingsAccountBonds.lt.100 + SavingsAccountBonds.100.to.500 +
      SavingsAccountBonds.500.to.1000 + SavingsAccountBonds.gt.1000 +
      SavingsAccountBonds.Unknown + EmploymentDuration.lt.1 +
      EmploymentDuration.1.to.4 + EmploymentDuration.4.to.7 + EmploymentDuration.gt.7
      + EmploymentDuration.Unemployed + Personal.Male.Divorced.Seperated +
      Personal.Female.NotSingle + Personal.Male.Single + Personal.Male.Married.Widowed +
      Personal.Female.Single + OtherDebtorsGuarantors.None +
      OtherDebtorsGuarantors.CoApplicant + OtherDebtorsGuarantors.Guarantor +
      Property.RealEstate + Property.Insurance + Property.CarOther + Property.Unknown +
      OtherInstallmentPlans.Bank + OtherInstallmentPlans.Stores +
      OtherInstallmentPlans.None + Housing.Rent + Housing.Own + Housing.ForFree +
      Job.UnemployedUnskilled + Job.UnskilledResident + Job.SkilledEmployee +
      Job.Management.SelfEmp.HighlyQualified, data=training, family="binomial")
testing$Score61<-predict(mod.glm, newdata=testing, type="response")
testsamp <- head(testing,100);colnames(testsamp) <- make.unique(names(testsamp))
ggplot(testsamp, aes(x=ID, y=Score61, fill=Class)) + geom_bar(stat="identity") +
      scale_fill_manual(values=c( "red","darkgreen")) +
      scale_y_continuous(limits=c(0,1),expand = c(0,0)) + theme_bw(base_size = 18) +
      xlab(NULL) + theme(axis.text.x = element_blank(),aspect.ratio=0.5) +
      geom_hline(yintercept = 0.75,lwd=1.6)
```

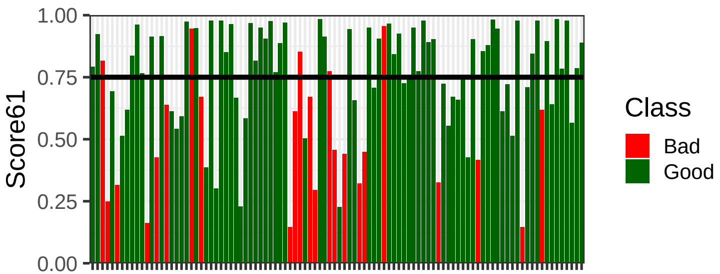


Fig. 8.7: Logistic regression output on 61 criteria.

In comparison with Figure 8.6, the 100 samples in Figure 8.7 above yield a higher True Positive Rate (TPR) $= 48/77 = 62.34\%$ and a lower False Positive Rate (FPR) $= 5/23 = 21.74\%$ at the level $p = 0.75$. In other words, the count of true positive samples has increased from 34 to 48, and the count of false positive samples has decreased from 6 to 5.

```
cat('TPR61=',length(testsamp$Score61[testsamp$Score61>0.75 &
      testsamp$Class=="Good"])/length(testsamp$Score61[testsamp$Class=="Good"]),'\n')
cat('FPR61=',length(testsamp$Score61[testsamp$Score61>0.75 &
      testsamp$Class=="Bad"])/length(testsamp$Score61[testsamp$Class=="Bad"]),'\n')
pred <- ifelse(testsamp$Score61 > 0.75, "Good", "Bad")
confusionMatrix(factor(noquote(pred)),factor(testsamp$Class))
```

8.4 ROC Curve

The ROC curve is a plot of the True Positive Rate values

$$x \longmapsto \overline{F}_G(x)$$

against the False Positive Rate function

$$x \longmapsto \overline{F}_B(x).$$

Definition 8.6. *The Receiver Operating Characteristic (ROC) curve is the function of the threshold $p \in [0,1]$ defined as*

$$\begin{aligned} [0,1] &\longrightarrow [0,1] \\ p \quad &\longmapsto \mathrm{ROC}(p) := \overline{F}_G\big(\overline{F}_B^{-1}(p)\big), \end{aligned}$$

where $\overline{F}_B^{-1}$ denotes the inverse of the tail distribution function $\overline{F}_B$.

The construction of Definition 8.6 proceeds in two steps.

a) Starting from a given FPR level p, compute the associated threshold x^* as $x^* := \overline{F}_B^{-1}(p)$, *i.e.*

$$p = \overline{F}_B(x^*) = \mathbb{P}(X > x^* \mid B).$$

b) From the threshold level, x^* estimate the corresponding TPR as

$$\mathbb{P}(X > x^* \mid G) = \overline{F}_G(x^*) = \overline{F}_G\big(\overline{F}_B^{-1}(p)\big) = \mathrm{ROC}(p).$$

Proposition 8.7. *The ROC function can be rewritten as the integral*

$$\mathrm{ROC}(p) = \int_0^p \lambda\big(\overline{F}_B^{-1}(q)\big)dq, \qquad 0 \leqslant p \leqslant 1,$$

where the likelihood ratio $\lambda(x)$ is given by

$$\lambda(x) = \frac{f_X(x \mid G)}{f_X(x \mid B)}, \qquad x \in \mathbb{R}.$$

Proof. The slope of the ROC curve at the point $p \in [0,1]$ is given by

$$\begin{aligned}\frac{d}{dp}\overline{F}_G\big(\overline{F}_B^{-1}(p)\big) &= \overline{F}'_G\big(\overline{F}_B^{-1}(p)\big)\frac{d}{dp}\overline{F}_B^{-1}(p) \\ &= \frac{\overline{F}'_G\big(\overline{F}_B^{-1}(p)\big)}{\overline{F}'_B\big(\overline{F}_B^{-1}(p)\big)} \\ &= \lambda\big(\overline{F}_B^{-1}(p)\big),\end{aligned}$$

hence we have

$$\overline{F}_G\big(\overline{F}_B^{-1}(p)\big) = \int_0^p \lambda\big(\overline{F}_B^{-1}(q)\big)dq, \qquad 0 \leqslant p \leqslant 1.$$

□

When X is Gaussian distributed given $\{G, B\}$ with the conditional densities

$$f_X(x \mid G) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)}$$

and

$$f_X(x \mid B) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)},$$

the likelihood ration is given by

$$\lambda(x) = \frac{f_X(x \mid G)}{f_X(x \mid B)} = \mathrm{e}^{\beta x-(\mu_G^2-\mu_B^2)/(2\sigma^2)}, \qquad x \in \mathbb{R},$$

with $\beta := (\mu_G - \mu_B)/\sigma^2$.

The ROC curve in this Gaussian setting can be computed using the following R code.

```
FG <- function(x, mug, sigma) {1 - pnorm((x - mug) / sigma)}
FBINV <- function(x, mub, sigma) {mub + sigma * qnorm(1 - x)}
ROC <- function(x, mub, mug, sigma) {
  sapply(x, function(x_val) ifelse(x_val == 0, 0, FG(FBINV(x_val, mub, sigma), mug, sigma)))}
```

Figure 8.8 presents three samples of ROC curves in the Gaussian example with successively $(\mu_B, \mu_G) = (1, 4)$, $(\mu_B, \mu_G) = (1, 2)$, and $(\mu_B, \mu_G) = (1, 1)$.

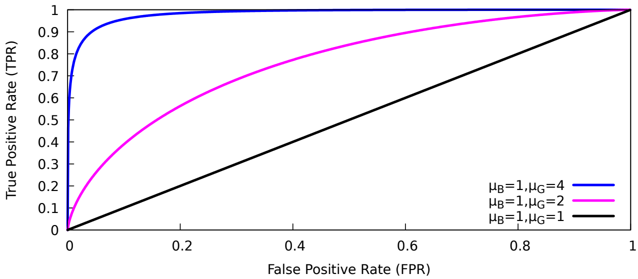


Fig. 8.8: Gaussian ROC curves.

We check that the classification is better when $\mu_B << \mu_G$. The ROC Curve shows the performance of the classification procedure, which is quantified by the Area Under The Curve (AUC). A perfect classification would correspond to a single point with coordinates $(0, 1)$ and AUC equal to 1, which corresponds to FPR=0% false negatives and TPR=100% true positives. The closer the AUC is to 1, the better the classification is performing.

On the other hand, a completely random guess would correspond to a point on the diagonal line. Points above the diagonal represent good classification results (better than random), while points below the line represent poor results (worse than random) (Wikipedia).

```
TPR5<-length(testing$Score5[testing$Score5>0.75 &
      testing$Class=="Good"])/length(testing$Score5[testing$Class=="Good"])
FPR5<-length(testing$Score5[testing$Score5>0.75 &
      testing$Class=="Bad"])/length(testing$Score5[testing$Class=="Bad"])
cat('TPR5=',TPR5,'\n'); cat('FPR5=',FPR5,'\n')
TPR61<-length(testing$Score61[testing$Score61>0.75 &
      testing$Class=="Good"])/length(testing$Score61[testing$Class=="Good"])
FPR61<-length(testing$Score61[testing$Score61>0.75 &
      testing$Class=="Bad"])/length(testing$Score61[testing$Class=="Bad"])
cat('TPR61=',TPR61,'\n'); cat('FPR61=',FPR61,'\n')
```

The above True Positive Rates (TPR) and False Positive Rates (FPR) based on 5 and 61 criteria at the level $p = 0.75$ are recomputed in the above R code on the whole 400 samples, and plotted on the next ROC graphs of Figure 8.9.

```
install.packages("ROCR"); library(ROCR)
pred5<-prediction(as.numeric(testing$Score5),as.numeric(testing$Class))
perf5 <- performance(pred5,"tpr","fpr")
dev.new(width=16,height=7); par(mar = c(4.5,4.5,2,2))
plot(perf5,col="purple",lwd=3, xaxs = "i", yaxs = "i",cex.lab=2,las=1)
segments(FPR5,0,FPR5,TPR5, col="purple", lwd =2)
segments(0,TPR5,FPR5,TPR5, col="purple", lwd =2)
pred61<-prediction(as.numeric(testing$Score61),as.numeric(testing$Class))
perf61 <- performance(pred61,"tpr","fpr"); par(new=TRUE)
plot(perf61,col="blue",lwd=3,main="", ann=FALSE, xaxs="i", yaxs="i")
legend("bottomright", legend=c("61 criteria","5 criteria"),col=c("blue","purple"), lwd=3, cex=3)
segments(0,0,1,1, col="black", lwd =3)
segments(FPR61,0,FPR61,TPR61, col="blue", lwd =2)
segments(0,TPR61,FPR61,TPR61, col="blue", lwd =2)
```

The ROC graphs in the next Figure 8.9 confirm the improvement in classification reached when switching from 5 to 61 criteria.

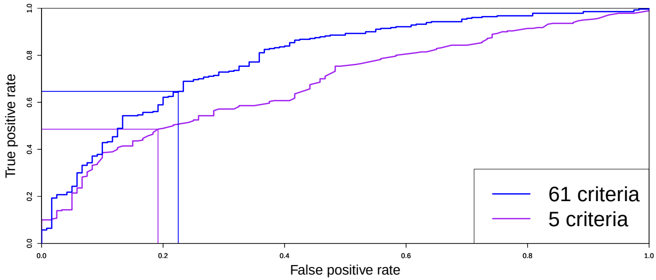


Fig. 8.9: ROC curves based on 5 criteria and 61 criteria.

Using a neural network in R

Note: Restarting R is *needed* before running the following code.

```
install.packages("neuralnet");
library(neuralnet);library(caret);data(GermanCredit);library(ROCR);
Train <- createDataPartition(GermanCredit$Class, p=0.6, list=FALSE)
training <- GermanCredit[ Train, ];testing <- GermanCredit[ -Train, ]
nn <- neuralnet(Class ~ Age + ForeignWorker + Property.RealEstate + Housing.Own +
        CreditHistory.Critical, data=training, hidden=c(3,1), linear.output=FALSE,
        threshold=0.05)
nn$result.matrix; plot(nn,col.intercept = "blue")
temp_test <- subset(testing, select = c("Age", "ForeignWorker", "Property.RealEstate",
        "Housing.Own", "CreditHistory.Critical"))
head(temp_test); nn.results <- compute(nn, temp_test)
results <- data.frame(actual = testing$Class, prediction = nn.results$net.result)
head(results); pred <- ifelse(results[,3] > 0.75, "Good", "Bad")
confusionMatrix(factor(results$actual),factor(noquote(pred)))
pred2<-prediction(as.numeric(results[,3]),as.numeric(results$actual))
perf <- performance(pred2,"tpr","fpr"); plot(perf,col="purple",lwd=3, xaxs = "i", yaxs = "i")
```

Improved performance may be achieved by rescaling the binary variables to $\{-1, 1\}$.

```
training$ForeignWorker <- ifelse(training$ForeignWorker == 0, -1, 1)
training$Property.RealEstate <- ifelse(training$Property.RealEstate  == 0, -1, 1)
training$Housing.Own <- ifelse(training$Housing.Own == 0, -1, 1)
training$CreditHistory.Critical <- ifelse(training$CreditHistory.Critical  == 0, -1, 1)
nn2 <- neuralnet(Class ~ Age + ForeignWorker + Property.RealEstate + Housing.Own +
        CreditHistory.Critical, data=training, hidden=c(3,1), linear.output=FALSE,
        threshold=0.05)
nn2.results <- compute(nn2, temp_test)
results2 <- data.frame(actual = testing$Class, prediction = nn2.results$net.result)
pred3 <- ifelse(results2[,3] > 0.75, "Good", "Bad")
confusionMatrix(factor(results2$actual),factor(noquote(pred3)))
pred4<-prediction(as.numeric(results2[,3]),as.numeric(results$actual))
dev.new(width=16,height=7); par(mar = c(4.5,4.5,2,2))
plot(perf,col="purple",lwd=3, xaxs = "i", yaxs = "i",cex.lab=2,las=1)
perf3 <- performance(pred4,"tpr","fpr"); par(new=TRUE)
plot(perf3,col="blue",lwd=3,main="", ann=FALSE, xaxs="i", yaxs="i")
legend("bottomright", legend=c("Rescaled","Non rescaled"),col=c("blue","purple"), lwd=3,
        cex=3)
```

Using a neural network in Python

Download the corresponding **IPython notebook*** that can be run **here** using this **data file**.

Using random forests in Python

Download the corresponding **IPython notebook** that can be run **here** using this **data file**.

* Right-click to save as attachment (may not work on).

Using XGBoost in R

Download the corresponding R **code**.

Exercises

Exercise 8.1 Consider a set Ω of applicants decomposed into the partition $\Omega = G \cup B$, where each applicant ω is assigned a score $X(\omega)$ which is exponentially distributed given $\{G, B\}$ with the conditional densities

$$f_X(x \mid G) = \lambda_G \mathrm{e}^{-\lambda_G x} \mathbb{1}_{[0,\infty)}(x)$$

and

$$f_X(x \mid B) = \lambda_B \mathrm{e}^{-\lambda_B x} \mathbb{1}_{[0,\infty)}(x),$$

where $\lambda_B > \lambda_G > 0$.

a) Compute the conditional expected values $\mathbb{E}[X \mid G]$ and $\mathbb{E}[X \mid B]$.
b) Compute the *probability default curve*

$$x \longmapsto \mathbb{P}(B \mid X = x) = \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}$$

in terms of the likelihood ratio $\lambda(x) := f_G(x)/f_B(x)$, $x \in \mathbb{R}$.
c) Determine the acceptance set

$$\mathscr{A} := \big\{ x \in \mathbb{R} \ : \ D\mathbb{P}(B \mid X = x) \leqslant L\mathbb{P}(G \mid X = x) \big\},$$

where

- L represents the loss incurred by the rejection of an applicant, and
- D represents the loss incurred by the default of an applicant.

Exercise 8.2 Consider a set Ω of applicants decomposed into the partition $\Omega = G \cup B$, where each applicant ω is assigned a uniformly distributed score $X(\omega)$ given $\{G, B\}$, with the conditional densities

$$f_X(x \mid G) = \frac{1}{\lambda_G} \mathbb{1}_{[0,\lambda_G]}(x) dx$$

and

$$f_X(x \mid B) = \frac{1}{\lambda_B} \mathbb{1}_{[0,\lambda_B]}(x) dx,$$

where $0 < \lambda_B < \lambda_G$.

a) Compute the *probability default curve*

$$x \longmapsto \mathbb{P}(B \mid X = x) = \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}.$$

b) Compute the conditional expected values $\mathbb{E}[X \mid G]$ and $\mathbb{E}[X \mid B]$.
c) Determine the acceptance set

$$A := \big\{x \in \mathbb{R}_+ \ : \ D\mathbb{P}(B \mid X = x) \leqslant L\mathbb{P}(G \mid X = x)\big\},$$

where

- L represents the missed earnings incurred by the rejection of applicant,
- D represents the loss incurred by the default of an applicant.

Exercise 8.3

a) Compute the ROC function $x \mapsto \overline{F}_G\big(\overline{F}_B^{-1}(x)\big)$ of Exercise 8.1, $x \in [0, 1]$, and draw its graph for sample values of λ_G, λ_B.
b) Compute the ROC function $x \mapsto \overline{F}_G\big(\overline{F}_B^{-1}(x)\big)$ of Exercise 8.2, $x \in [0, 1]$, and draw its graph for sample values of λ_G, λ_B.

Exercise 8.4 Consider a set Ω of customers decomposed as the partition $\Omega = G \cup B$, where each customer ω is assigned a uniformly distributed score $X(\omega)$ given $\{G, B\}$, with the conditional densities

$$f_X(x \mid G) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)}$$

and

$$f_X(x \mid B) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)},$$

with $\mu_B < \mu_G$.

a) Compute the *probability default curve*

$$x \longmapsto \mathbb{P}(B \mid X = x) := \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}.$$

b) Compute the likelihood ratio $\lambda(x)$ defined as $\lambda(x) := f_X(x \mid G)/f_X(x \mid B)$, $x \in \mathbb{R}$.
c) Letting

- $L(x) := \mathrm{e}^{-ax}$ represent the missed earnings incurred by rejecting an applicant with score $x \in \mathbb{R}$, with $a > 0$,
- $D(x) := \mathrm{e}^{bx}$ represent the loss incurred by the default of an applicant with score $x \in \mathbb{R}$, with $b > 0$,

determine the acceptance set

$$\mathcal{A} := \big\{ x \in \mathbb{R}_+ \ : \ D(x)\mathbb{P}(B \mid X = x) \leqslant L(x)\mathbb{P}(G \mid X = x) \big\}.$$

Chapter 9
Credit Risk - Structural Approach

Credit risk can be defined as the risk of default on the payment of a debt. In this chapter, credit risk is modeled using the value of a firm's assets, *i.e.* the default event is said to occur when the value of assets drops below a certain pre-defined level. This is in contrast to the reduced form approach to credit risk of Chapter 10, in which stochastic processes are used to model default probabilities. We also consider the modeling of correlation and dependence between multiple default times.

9.1 Merton Model

The Merton (1974) credit risk model reframes corporate debt as an option on a firm's underlying value. Precisely the value S_t of a firm's asset is modeled by a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

under the historical (or physical) measure $\mathbb{P}$. Recall that, using the standard Brownian motion

$$\widehat{B}_t = \frac{\mu - r}{\sigma} t + B_t, \qquad t \geqslant 0,$$

under the risk-neutral probability measure $\mathbb{P}^*$, the process $(S_t)_{t\in\mathbb{R}_+}$ is modeled as

$$dS_t = rS_t dt + \sigma S_t d\widehat{B}_t.$$

Assumption 9.1. *The company's debt is represented by an amount $K > 0$ in bonds to be paid at maturity T, see* e.g. *§ 4.1 of Grasselli and Hurd (2010).*

In this setting, default occurs if $S_T < K$ with probability $\mathbb{P}(S_T < K)$, the bond holder will receive the recovery value S_T. Otherwise, if $S_T \geqslant K$ the bond holder receives K and the equity holder is entitled to receive $S_T - K$, which can be represented as $(S_T - K)^+$ in general.

The discounted expected cash flow (or dividend) $\mathrm{e}^{-(T-t)r}\mathbb{E}^*[(S_T - K)^+ \mid \mathcal{F}_t]$ received by the equity holder can be estimated at time $t \in [0, T]$ as the price of a European call option, from the Black-Scholes formula

$$\mathrm{e}^{-(T-t)r}\mathbb{E}^*[(S_T - K)^+ \mid \mathcal{F}_t] = S_t \Phi\left(\frac{(r+\sigma^2/2)(T-t) + \log(S_t/K)}{\sigma\sqrt{T-t}}\right)$$
$$-K\mathrm{e}^{-(T-t)r}\Phi\left(\frac{(r-\sigma^2/2)(T-t) + \log(S_t/K)}{\sigma\sqrt{T-t}}\right), \quad 0 \leqslant t \leqslant T,$$

see Proposition 5.4.

Proposition 9.2. *The default probability $\mathbb{P}(S_T < K \mid \mathcal{F}_t)$ can be computed from the lognormal distribution of S_T as*

$$\mathbb{P}(S_T < K \mid \mathcal{F}_t) = \Phi\big(-d_-^\mu\big), \tag{9.1}$$

where Φ is the cumulative distribution function of the standard normal distribution, and

$$d_-^\mu := \frac{(\mu - \sigma^2/2)(T-t) + \log(S_t/K)}{\sigma\sqrt{T-t}}.$$

Proof. The default probability $\mathbb{P}(S_T < K \mid \mathcal{F}_t)$ can be computed from the lognormal distribution of S_T as

$$\begin{aligned}
\mathbb{P}(S_T < K \mid \mathcal{F}_t) &= \mathbb{P}\big(S_0 \mathrm{e}^{\sigma B_T + (\mu - \sigma^2/2)T} < K \mid \mathcal{F}_t\big) \\
&= \mathbb{P}\left(B_T < \frac{1}{\sigma}\left(-\left(\mu - \frac{\sigma^2}{2}\right)T + \log\frac{K}{S_0}\right) \,\Big|\, \mathcal{F}_t\right) \\
&= \mathbb{P}\left(B_T - B_t + y < \frac{1}{\sigma}\left(-\left(\mu - \frac{\sigma^2}{2}\right)T + \log\frac{K}{S_0}\right)\right)_{|y=B_t} \\
&= \mathbb{P}\left(B_T - B_t + \frac{1}{\sigma}\left(-\left(\mu - \frac{\sigma^2}{2}\right)t + \log\frac{K}{x}\right)\right. \\
&\qquad\qquad \left. < \frac{1}{\sigma}\left(-\left(\mu - \frac{\sigma^2}{2}\right)T + \log\frac{K}{S_0}\right)\right)_{|x=S_t}
\end{aligned}$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$
\begin{aligned}
&= \frac{1}{\sqrt{2(T-t)\pi}} \int_{-\infty}^{(-(\mu-\sigma^2/2)(T-t)+\log(K/S_t))/\sigma} \mathrm{e}^{-x^2/(2(T-t))} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(-(\mu-\sigma^2/2)(T-t)+\log(K/S_t))/(\sigma\sqrt{T-t})} \mathrm{e}^{-x^2/2} dx \\
&= 1-\Phi\left(\frac{(\mu-\sigma^2/2))(T-t)+\log(S_t/K))}{\sigma\sqrt{T-t}}\right) \\
&= 1-\Phi(d_-^\mu) \\
&= \Phi(-d_-^\mu).
\end{aligned}
$$

□

The formula (9.2) can be implemented as follows.

```
P <- function(S, K, mu, T, sigma)
{d1 <- (log(S/K)+(mu-sigma^2/2)*T)/(sigma*sqrt(T))
P = pnorm(d1);P}
```

Note that under the risk-neutral probability measure $\mathbb{P}^*$ we have, replacing μ with r,

$$
\mathbb{P}^*(S_T < K \mid \mathcal{F}_t) = \Phi(-d_-^r),
$$

with

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

which implies the relation

$$
d_-^r = d_-^\mu - \frac{\mu-r}{\sigma}\sqrt{T-t},
$$

or, denoting by Φ^{-1} the inverse function of Φ,

$$
\Phi^{-1}(\mathbb{P}(S_T < K \mid \mathcal{F}_t)) = -\frac{\mu-r}{\sigma}\sqrt{T-t} + \Phi^{-1}(\mathbb{P}^*(S_T < K \mid \mathcal{F}_t)).
$$

If the level of the firm's assets falls below the level K at time T, default may have occurred at a random time τ such that

$$
\mathbb{P}(\tau < T \mid \mathcal{F}_t) = \mathbb{P}(S_T < K \mid \mathcal{F}_t).
$$

In this case, the result of Proposition 9.2 can be reinterpreted in the next corollary.

Corollary 9.3. *The conditional distribution of the default time τ is given by*

$$\mathbb{P}(\tau < T \mid \mathcal{F}_t) = \mathbb{P}(S_T < K \mid \mathcal{F}_t) = \Phi\left(-\frac{(\mu - \sigma^2/2)(T-t) + \log(S_t/K))}{\sigma\sqrt{T-t}}\right), \tag{9.2}$$

$0 \leqslant t \leqslant T$.

We also have

$$\begin{aligned}
\mathbb{P}(\tau < T \mid \mathcal{F}_t) &= \mathbb{P}(S_T < K \mid \mathcal{F}_t) \\
&= \Phi\left(\Phi^{-1}(\mathbb{P}^*(S_T < K \mid \mathcal{F}_t)) - \frac{\mu - r}{\sigma}\sqrt{T-t}\right) \\
&= \Phi\left(\Phi^{-1}(\mathbb{P}^*(\tau < T \mid \mathcal{F}_t)) - \frac{\mu - r}{\sigma}\sqrt{T-t}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}^*(\tau < T \mid \mathcal{F}_t) &= \mathbb{P}^*(S_T < K \mid \mathcal{F}_t) \\
&= \Phi\left(-\frac{(r - \sigma^2/2))(T-t) + \log(S_t/K))}{\sigma\sqrt{T-t}}\right) \\
&= \Phi\left(\Phi^{-1}(\mathbb{P}(S_T < K \mid \mathcal{F}_t)) + \frac{\mu - r}{\sigma}\sqrt{T-t}\right) \\
&= \Phi\left(\Phi^{-1}(\mathbb{P}(\tau < T \mid \mathcal{F}_t)) + \frac{\mu - r}{\sigma}\sqrt{T-t}\right).
\end{aligned} \tag{9.3}$$

Note that when $\mu < r$, we have

$$\mathbb{P}(\tau < T \mid \mathcal{F}_t) > \mathbb{P}^*(\tau < T \mid \mathcal{F}_t),$$

whereas when $\mu > r$ we get

$$\mathbb{P}(\tau < T \mid \mathcal{F}_t) < \mathbb{P}^*(\tau < T \mid \mathcal{F}_t),$$

as illustrated in the next Figure 9.1.

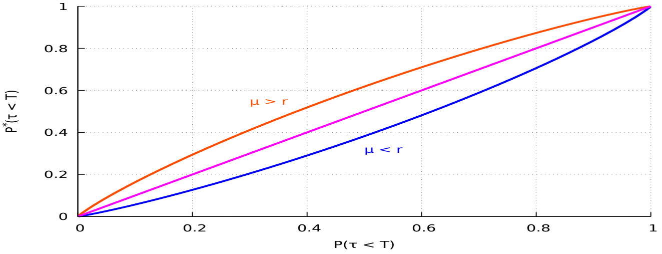


Fig. 9.1: Graph of the function $x \longmapsto \Phi\big(\Phi^{-1}(x) - (\mu - r)\sqrt{T}/\sigma\big)$ for $\mu > r$, $\mu = r$, and $\mu < r$.

9.2 Default Bonds

In the following proposition we price at time $t \in [0, T]$ the amount $\min(S_T, K)$ received by the bond holder (or junior creditor) at maturity, based on the recovery value S_T when $S_T < K$. This price can interpreted at the price $P(t, T)$ at time $t \in [0, T]$ of a default bond with face value \$1, maturity T and recovery value $\min(S_T/K, 1)$.

Proposition 9.4. *The amount received by the bond holder (or junior creditor) at maturity is priced at time $t \in [0, T]$ as*

$$\mathrm{e}^{-(T-t)r}\mathbb{E}^*\big[\min(S_T, K) \mid \mathcal{F}_t\big] = K\mathrm{e}^{-(T-t)r}\Phi(d_-^r) - S_t\Phi(-d_+^r), \quad 0 \leqslant t \leqslant T.$$

Proof. Using the Black-Scholes put option pricing formula and the identity

$$\min(x, K) = K - (K - x)^+, \qquad x \in \mathbb{R},$$

we have

$$\begin{aligned}
\mathrm{e}^{-(T-t)r}\mathbb{E}^*\big[\min(S_T, K) \mid \mathcal{F}_t\big] &= \mathrm{e}^{-(T-t)r}\mathbb{E}^*\big[K - (K - S_T)^+ \mid \mathcal{F}_t\big] \\
&= \mathrm{e}^{-(T-t)r}K - \mathrm{e}^{-(T-t)r}\mathbb{E}^*\big[(K - S_T)^+ \mid \mathcal{F}_t\big] \\
&= \mathrm{e}^{-(T-t)r}K + S_t\Phi(-d_+^r) - K\mathrm{e}^{-(T-t)r}\Phi(-d_-^r) \\
&= K\mathrm{e}^{-(T-t)r}\Phi(d_-^r) - S_t\Phi(-d_+^r).
\end{aligned}$$

□

Writing

$$
\begin{aligned}
P(t,T) &= \mathrm{e}^{-(T-t)y_{t,T}} \\
&= \frac{1}{K}\mathrm{e}^{-(T-t)r}\mathbb{E}^*[\min(S_T,K) \mid \mathcal{F}_t] \\
&= \mathrm{e}^{-(T-t)r}\Phi(d_-^r) - \frac{S_t}{K}\Phi(-d_+^r),
\end{aligned}
$$

gives the default bond yield

$$
\begin{aligned}
y_{t,T} &= -\frac{1}{T-t}\log(P(t,T)) \\
&= -\frac{1}{T-t}\log\left(\mathrm{e}^{-(T-t)r}\mathbb{E}^*\left[\min\left(1,\frac{S_T}{K}\right)\,\Big|\,\mathcal{F}_t\right]\right) \\
&= r - \frac{1}{T-t}\log\left(\mathbb{E}^*\left[\min\left(1,\frac{S_T}{K}\right)\,\Big|\,\mathcal{F}_t\right]\right) \\
&= r - \frac{1}{T-t}\log\left(\frac{1}{K}\mathbb{E}^*\left[\min\left(K,S_T\right)\,\Big|\,\mathcal{F}_t\right]\right) \\
&= r - \frac{1}{T-t}\log\left(\Phi(d_-^r) - \frac{S_t}{K}\mathrm{e}^{(T-t)r}\Phi(-d_+^r)\right),
\end{aligned}
$$

which is usually higher than the risk-free yield r.

9.3 Black-Cox Model

In the Black and Cox (1976) model the firm has to maintain an account balance above the level K throughout time, therefore default occurs at the first time the process S_t hits the level K, cf. § 4.2 of Grasselli and Hurd (2010). The default time τ_K is therefore the first hitting time

$$
\tau_K := \inf\left\{t \geqslant 0 \;:\; S_t := S_0\mathrm{e}^{\sigma B_t + (\mu-\sigma^2/2)t} \leqslant K\right\},
$$

of the level K by

$$
(S_t)_{t\in\mathbb{R}_+} = \left(S_0\mathrm{e}^{\sigma B_t+(\mu-\sigma^2/2)t}\right)_{t\in\mathbb{R}_+},
$$

after starting from $S_0 > K$.

Proposition 9.5. *The probability distribution function of the default time τ_K is given by*

$$
\mathbb{P}\big(\tau_K \leqslant T\big) = \mathbb{P}(S_T \leqslant K) + \left(\frac{S_0}{K}\right)^{1-2\mu/\sigma^2}\Phi\left(\frac{\log(K/S_0)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right), \tag{9.4}
$$

with $S_0 \geqslant K$.

Proof. By *e.g.* Corollary 7.2.2 and pages 297-299 of Shreve (2004), or from Relation (10.13) in Privault (2022), we have

$$\begin{aligned}
\mathbb{P}\big(\tau_K \leqslant T\big) &= \mathbb{P}\left(\min_{t\in[0,T]} S_t \leqslant K\right) \\
&= \mathbb{P}\left(\min_{t\in[0,T]} \mathrm{e}^{\sigma B_t+(\mu-\sigma^2/2)t} \leqslant \frac{K}{S_0}\right) \\
&= \mathbb{P}\left(\min_{t\in[0,T]} \left(B_t + \frac{(\mu-\sigma^2/2)t}{\sigma}\right) \leqslant \frac{1}{\sigma}\log\left(\frac{K}{S_0}\right)\right) \\
&= \Phi\left(\frac{\log(K/S_0)-(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right) \\
&\quad + \left(\frac{S_0}{K}\right)^{1-2\mu/\sigma^2} \Phi\left(\frac{\log(K/S_0)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right) \qquad (9.5) \\
&= \mathbb{P}(S_T \leqslant K) + \left(\frac{S_0}{K}\right)^{1-2\mu/\sigma^2} \Phi\left(\frac{\log(K/S_0)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right),
\end{aligned}$$

with $S_0 \geqslant K$. □

We check that when $S_0 = K$, using (9.1), Relation (9.4) reads

$$\begin{aligned}
\mathbb{P}\big(\tau_K \leqslant T\big) &= \Phi\left(-\frac{\log(K/S_0)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right) \\
&\quad + \Phi\left(\frac{\log(K/S_0)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}\right) \\
&= 1.
\end{aligned}$$

The cash flow

$$(S_T - K)^+ \mathbb{1}_{\{\tau_K > T\}} = (S_T - K) \mathbb{1}_{\left\{\min\limits_{t\in[0,T]} S_t > K\right\}}$$

received at maturity T by the equity holder can be priced at time $t \in [0,T]$ as a down-and-out barrier call option with strike price K and barrier level K is priced in the next proposition, in which BS_c denotes the Black-Scholes call pricing formula.

Proposition 9.6. *We have*

$$\mathrm{e}^{-(T-t)r}\mathbb{E}^*\left[(S_T-K)^+\mathbb{1}_{\left\{\min\limits_{0\leqslant t\leqslant T}S_t>K\right\}}\middle|\mathcal{F}_t\right]=\mathbb{1}_{\left\{\min\limits_{t\in[0,T]}S_t>B\right\}}g(t,S_t),$$

$t\in[0,T]$, *where*

$$g(t,S_t)=\mathrm{BS}_c(S_t,K,r,T-t,\sigma)-S_t\left(\frac{K}{S_t}\right)^{2r/\sigma^2}\mathrm{BS}_c(K/S_t,1,r,T-t,\sigma),$$

$0\leqslant t\leqslant T$.

Proof. By *e.g.* Relation (11.10) and Exercise 11.1 in Privault (2022), we have

$$\mathbb{E}^*\left[(S_T-K)^+\mathbb{1}_{\left\{\min\limits_{0\leqslant t\leqslant T}S_t>K\right\}}\middle|\mathcal{F}_t\right]=\mathbb{1}_{\left\{\min\limits_{t\in[0,T]}S_t>B\right\}}g(t,S_t),$$

$t\in[0,T]$, where

$$\begin{aligned}
&g(t,S_t)\\
&=S_t\Phi\left(\delta_+^{T-t}\left(\frac{S_t}{K}\right)\right)-\mathrm{e}^{-(T-t)r}K\Phi\left(\delta_-^{T-t}\left(\frac{S_t}{K}\right)\right)\\
&\quad -K\left(\frac{K}{S_t}\right)^{2r/\sigma^2}\Phi\left(\delta_+^{T-t}\left(\frac{K}{S_t}\right)\right)+\mathrm{e}^{-(T-t)r}K\left(\frac{S_t}{K}\right)^{1-2r/\sigma^2}\Phi\left(\delta_-^{T-t}\left(\frac{K}{S_t}\right)\right)\\
&=\mathrm{BS}_c(S_t,K,r,T-t,\sigma)\\
&\quad -K\left(\frac{S_t}{K}\right)^{-2r/\sigma^2}\Phi\left(\delta_+^{T-t}\left(\frac{K}{S_t}\right)\right)+\mathrm{e}^{-(T-t)r}S_t\left(\frac{S_t}{K}\right)^{-2r/\sigma^2}\Phi\left(\delta_-^{T-t}\left(\frac{K}{S_t}\right)\right)\\
&=\mathrm{BS}_c(S_t,K,r,T-t,\sigma)-S_t\left(\frac{K}{S_t}\right)^{2r/\sigma^2}\mathrm{BS}_c(K/S_t,1,r,T-t,\sigma),
\end{aligned}$$

$0\leqslant t\leqslant T$. □

For $t\geqslant 0$, taking now

$$\tau_K:=\inf\left\{u\in[t,\infty)\ :\ S_u:=S_0\mathrm{e}^{\sigma B_u+(\mu-\sigma^2/2)u}\leqslant K\right\},$$

the recovery value received by the bond holder at time $\min\left(\tau_K,T\right)$ is K, and it can be priced as in the next proposition.

Proposition 9.7. *After discounting from time* $\min\big(\tau_K, T\big)$ *to time* $t \in [0, T]$, *we have*

$$\mathbb{E}^*\big[K\mathrm{e}^{-(\min(\tau_K,T)-t)r} \mid \mathcal{F}_t\big]$$
$$= K\mathbb{1}_{\{\tau_K \geqslant t\}} \int_t^T \mathrm{e}^{-(u-t)r} \mathrm{d}\mathbb{P}^*\big(\tau_K \leqslant u \mid \mathcal{F}_t\big) + K\mathrm{e}^{-(T-t)r}\mathbb{P}^*\big(\tau_K > T \mid \mathcal{F}_t\big).$$

Proof. We have

$$\begin{aligned}
&\mathbb{E}^*\big[K\mathrm{e}^{-(\min(\tau_K,T)-t)r} \mid \mathcal{F}_t\big] \\
&= \mathbb{E}^*\big[K\mathrm{e}^{-(\tau_K-t)r}\mathbb{1}_{\{t\leqslant\tau_K\leqslant T\}} + K\mathrm{e}^{-(T-t)r}\mathbb{1}_{\{\tau_K>T\}} \mid \mathcal{F}_t\big] \\
&= K\mathbb{E}^*\big[\mathrm{e}^{-(\tau_K-t)r}\mathbb{1}_{\{t\leqslant\tau_K\leqslant T\}} \mid \mathcal{F}_t\big] + K\mathrm{e}^{-(T-t)r}\mathbb{P}^*\big(\tau_K > T \mid \mathcal{F}_t\big) \\
&= K\mathbb{1}_{\{\tau_K\geqslant t\}}\mathbb{E}^*\big[\mathrm{e}^{-(\tau_K-t)r}\mathbb{1}_{\{t\leqslant\tau_K\leqslant T\}} \mid \mathcal{F}_t\big] + K\mathrm{e}^{-(T-t)r}\mathbb{P}^*\big(\tau_K > T \mid \mathcal{F}_t\big) \\
&= K\mathbb{1}_{\{\tau_K\geqslant t\}} \int_t^T \mathrm{e}^{-(u-t)r}\mathrm{d}\mathbb{P}\big(\tau_K \leqslant u \mid \mathcal{F}_t\big) + K\mathrm{e}^{-(T-t)r}\mathbb{P}^*\big(\tau_K > T \mid \mathcal{F}_t\big),
\end{aligned}$$

$0 \leqslant t \leqslant T$. □

The above probabilities $\mathbb{P}^*\big(\tau_K \leqslant u \mid \mathcal{F}_t\big)$ and $\mathbb{P}^*\big(\tau_K > T \mid \mathcal{F}_t\big) = 1 - \mathbb{P}^*\big(\tau_K \leqslant T \mid \mathcal{F}_t\big)$ can be computed from (9.5) as

$$\begin{aligned}
\mathbb{P}^*\big(\tau_K \leqslant u \mid \mathcal{F}_t\big) &= \Phi\left(\frac{\log(K/S_t) - (r-\sigma^2/2)(u-t)}{\sigma\sqrt{u-t}}\right) \\
&\quad + \left(\frac{S_t}{K}\right)^{1-2r/\sigma^2}\Phi\left(\frac{\log(S_t/K) + (r-\sigma^2/2)(u-t)}{\sigma\sqrt{u-t}}\right) \\
&= \mathbb{P}(S_u \leqslant K \mid \mathcal{F}_t) + \left(\frac{S_t}{K}\right)^{1-2r/\sigma^2}\Phi\left(\frac{\log(S_t/K) + (r-\sigma^2/2)(u-t)}{\sigma\sqrt{u-t}}\right),
\end{aligned}$$

with $S_t \geqslant K$ and $u > t$, from which the probability density function of the hitting time τ_K can be estimated by differentiation with respect to $u > t$. Note also that we have

$$\begin{aligned}
\mathbb{P}^*\big(\tau_K < \infty \mid \mathcal{F}_t\big) &= \lim_{u\to\infty}\mathbb{P}^*\big(\tau_K \leqslant u \mid \mathcal{F}_t\big) \\
&= \begin{cases} \left(\dfrac{K}{S_t}\right)^{-1+2r/\sigma^2} & \text{if } r > \sigma^2/2 \\[2ex] 1 & \text{if } r \leqslant \sigma^2/2. \end{cases}
\end{aligned}$$

9.4 Correlated Default Times

In order to model correlated default and possible "domino effects", one can regard two given default times τ_1 and τ_2 as correlated random variables, with correlation coefficient

$$\rho := \frac{\mathrm{Cov}(\tau_1, \tau_2)}{\sqrt{\mathrm{Var}[\tau_1]\,\mathrm{Var}[\tau_2]}} \in [-1, 1].$$

Given two default events $\{\tau_1 \leqslant T\}$ and $\{\tau_2 \leqslant T\}$ with probabilities

$$\mathbb{P}(\tau_1 \leqslant T) = 1 - \exp\left(-\int_0^T \lambda_1(s)ds\right) \text{ and } \mathbb{P}(\tau_2 \leqslant T) = 1 - \exp\left(-\int_0^T \lambda_2(s)ds\right)$$

we can define the default correlation $\rho^D \in [-1, 1]$ as

$$\rho^D = \frac{C_\Sigma(\mathbb{P}(\tau_1 \leqslant T), \mathbb{P}(\tau_2 \leqslant T)) - \mathbb{P}(\tau_1 \leqslant T)\mathbb{P}(\tau_2 \leqslant T)}{\sqrt{\mathbb{P}(\tau_1 \leqslant T)(1 - \mathbb{P}(\tau_1 \leqslant T))}\sqrt{\mathbb{P}(\tau_2 \leqslant T)(1 - \mathbb{P}(\tau_2 \leqslant T))}}. \quad (9.6)$$

When the default probabilities are specified in the Merton model of credit risk as

$$\begin{aligned}
\mathbb{P}(\tau_i \leqslant T) &= \mathbb{P}(S_T < K) \\
&= \mathbb{P}\left(\mathrm{e}^{\sigma_i B_T + (\mu_i - \sigma_i^2/2)T} < \frac{K}{S_0}\right) \\
&= \mathbb{P}\left(B_T \leqslant -\frac{(\mu_i - \sigma_i^2/2)T}{\sigma_i} + \frac{1}{\sigma_i}\log\frac{K}{S_0}\right) \\
&= \Phi\left(\frac{\log(K/S_0) - (\mu_i - \sigma_i^2/2)T}{\sigma_i\sqrt{T}}\right), \qquad i = 1, 2,
\end{aligned}$$

where

$$\left(A_t^i\right)_{t\in\mathbb{R}_+} := \left(S_0 \mathrm{e}^{\sigma_i B_t + (\mu - \sigma_i^2/2)t}\right)_{t\in\mathbb{R}_+}, \qquad i = 1, 2,$$

the default correlation ρ^D becomes

$$\begin{aligned}
\rho^D &= \frac{\mathbb{P}(\tau_1 \leqslant T \text{ and } \tau_2 \leqslant T) - \mathbb{P}(\tau_1 \leqslant T)\mathbb{P}(\tau_2 \leqslant T)}{\sqrt{\mathbb{P}(\tau_1 \leqslant T)(1 - \mathbb{P}(\tau_1 \leqslant T))}\sqrt{\mathbb{P}(\tau_2 \leqslant T)(1 - \mathbb{P}(\tau_2 \leqslant T))}} \\
&= \frac{\Phi_\Sigma\left(\frac{\log(S_0/K) + (\mu_1 - \sigma_1^2/2)T}{\sigma_1\sqrt{T}}, \frac{\log(S_0/K) + (\mu_2 - \sigma_2^2/2)T}{\sigma_2\sqrt{T}}\right) - \mathbb{P}(\tau_1 \leqslant T)\mathbb{P}(\tau_2 \leqslant T)}{\sqrt{\mathbb{P}(\tau_1 \leqslant T)(1 - \mathbb{P}(\tau_1 \leqslant T))}\sqrt{\mathbb{P}(\tau_2 \leqslant T)(1 - \mathbb{P}(\tau_2 \leqslant T))}}.
\end{aligned}$$

When trying to build a dependence structure for the default times τ_1 and τ_2, the idea of Li (2000) is to use the normalized Gaussian copula $C_\Sigma(x, y)$, with

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

with correlation parameter $\rho \in [-1, 1]$, and to model the joint default probability $\mathbb{P}(\tau_1 \leqslant T \text{ and } \tau_2 \leqslant T)$ as

$$\mathbb{P}(\tau_1 \leqslant T \text{ and } \tau_2 \leqslant T) := C_\Sigma \left(\mathbb{P}(\tau_1 \leqslant T), \mathbb{P}(\tau_2 \leqslant T) \right),$$

where C_Σ is given by (4.5). Namely, it was suggested to use a single *average correlation* estimate, see (8.1) page 82 of the Credit Metrics™ Technical Document Gupton et al. (1997), and Appendix F therein.

It is worth noting that the outcomes of this methodology have been discussed in a number of magazine articles in recent years, to name a few:

"Recipe for disaster: the formula that killed Wall Street", *Wired Magazine*, by F. Salmon (2009);

"The formula that felled Wall Street", *Financial Times Magazine*, by S. Jones (2009);

"Formula from hell", *Forbes.com*, by S. Lee (2009),

see also **here**.

On the other hand, a more proper definition of the default correlation ρ^D should be

$$\rho^D := \frac{\mathbb{P}(\tau_1 \leqslant T \text{ and } \tau_2 \leqslant T) - \mathbb{P}(\tau_1 \leqslant T)\mathbb{P}(\tau_2 \leqslant T)}{\sqrt{\mathbb{P}(\tau_1 \leqslant T)(1 - \mathbb{P}(\tau_1 \leqslant T))}\sqrt{\mathbb{P}(\tau_2 \leqslant T)(1 - \mathbb{P}(\tau_2 \leqslant T))}},$$

which requires the actual computation of the joint default probability $\mathbb{P}(\tau_1 \leqslant T \text{ and } \tau_2 \leqslant T)$. An exact expression for this joint default probability in the first passage time Black-Cox model, and the associated correlation, have been recently obtained in Li and Krehbiel (2016).

Multiple default times

Consider now a sequence $(\tau_k)_{k=1,2,\ldots,n}$ of random default times and, for more flexibility, a standardized random variable M with probability density function $\phi(m)$ and variance $\operatorname{Var}[M] = 1$.

As in the Merton (1974) model, cf. § 9.1, a common practice, see Vašiček (1987), Gibson (2004), Hull and White (2004) is to parametrize the default probability associated to each τ_k by a conditioning of the form

$$\mathbb{P}(\tau_k \leqslant T \mid M = m) = \Phi\left(\frac{\Phi^{-1}(F_{\tau_k}(T)) - a_k m}{\sqrt{1-a_k^2}}\right), \tag{9.7}$$

see (9.3), where

$$F_{\tau_k}(T) := \mathbb{P}(\tau_k \leqslant T)$$

is the cumulative distribution function of τ_k, $k = 1, 2, \ldots, n$, and $a_1, \ldots, a_n \in (-1, 1)$. Note that we have

$$\begin{aligned}\mathbb{P}(\tau_k \leqslant T) &= \int_{-\infty}^{\infty} \mathbb{P}(\tau_k \leqslant T \mid M = m)\phi(m)dm \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi^{-1}(\mathbb{P}(\tau_k \leqslant T)) - a_k m}{\sqrt{1-a_k^2}}\right)\phi(m)dm, \end{aligned} \tag{9.8}$$

and $\phi(m)$ can be typically chosen as a standard normal Gaussian probability density function.

Next, we present a dependence structure which implements of the Gaussian copula correlation method of Li (2000) in the case of multiple default times.

Definition 9.8. *Given $X_1, X_2, \ldots, X_n$ Gaussian samples defined as*

$$X_k := a_k M + \sqrt{1-a_k^2} Z_k, \quad k = 1, 2, \ldots, n, \tag{9.9}$$

conditionally to M, where $Z_1, Z_2, \ldots, Z_n$ are normal random variables with same cumulative distribution function Φ, independent of M, we construct the correlated default times $(\tau_1, \ldots, \tau_n)$ as

$$\tau_k := F_{\tau_k}^{-1}(\Phi_{X_k}(X_k)), \tag{9.10}$$

where $F_{\tau_k}^{-1}$ denotes the inverse function of F_{τ_k} and Φ_{X_k} denotes the cumulative distribution function of X_k, $k = 1, 2, \ldots, n$.

In the next proposition we compute the joint distribution of the default times $(\tau_1, \ldots, \tau_n)$ according to the above dependence structure.

Proposition 9.9. *The default times $(\tau_k)_{k=1,2,\ldots,n}$ have the joint distribution*

$$\mathbb{P}(\tau_1 \leqslant y_1, \ldots, \tau_n \leqslant y_n) = C\big(\mathbb{P}(\tau_1 \leqslant y_1), \ldots, \mathbb{P}(\tau_n \leqslant y_n)\big), \tag{9.11}$$

where

$$C(x_1, \ldots, x_n) := \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi_{X_k}^{-1}(x_1) - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \Phi\left(\frac{\Phi_{X_k}^{-1}(x_n) - a_n m}{\sqrt{1-a_n^2}}\right)\phi(m)dm,$$

$x_1, x_2, \ldots, x_n \in [0,1]$.

Proof. We start by showing that Definition 9.8 recovers the conditional distribution (9.7), as follows:

$$\begin{aligned}
\mathbb{P}(\tau_k \leqslant T \mid M = m) &= \mathbb{P}\big(F_{\tau_k}^{-1}(\Phi_{X_k}(X_k)) \leqslant T \mid M = m\big) \\
&= \mathbb{P}\big(\Phi_{X_k}(X_k) \leqslant F_{\tau_k}(T) \mid M = m\big) \\
&= \mathbb{P}\big(X_k \leqslant \Phi_{X_k}^{-1}\big(F_{\tau_k}(T)\big) \mid M = m\big) \\
&= \mathbb{P}\left(a_k m + Z_k\sqrt{1-a_k^2} \leqslant \Phi_{X_k}^{-1}\big(F_{\tau_k}(T)\big)\right) \\
&= \mathbb{P}\left(Z_k\sqrt{1-a_k^2} \leqslant \Phi_{X_k}^{-1}\big(F_{\tau_k}(T)\big) - a_k m\right) \\
&= \mathbb{P}\left(Z_k \leqslant \frac{\Phi_{X_k}^{-1}\big(F_{\tau_k}(T)\big) - a_k m}{\sqrt{1-a_k^2}}\right) \\
&= \Phi\left(\frac{\Phi_{X_k}^{-1}\big(\mathbb{P}(\tau_k \leqslant T)\big) - a_k m}{\sqrt{1-a_k^2}}\right), \qquad k = 1, 2, \ldots, n.
\end{aligned}$$

Note that the above recovers the correct marginal distributions (9.8), *i.e.* we have

$$\begin{aligned}
\mathbb{P}(\tau_k \leqslant y_k) &= \mathbb{P}(\tau_1 \leqslant \infty, \ldots, \tau_{k-1} \leqslant \infty, \tau_k \leqslant y_k, \tau_{k+1} \leqslant \infty, \ldots, \tau_n \leqslant \infty) \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi_{X_k}^{-1}\big(\mathbb{P}(\tau_k \leqslant y_k)\big) - a_k m}{\sqrt{1-a_k^2}}\right)\phi(m)dm \\
&= \int_{-\infty}^{\infty} \mathbb{P}(\tau_k \leqslant T \mid M = m)\phi(m)dm, \quad k = 1, 2, \ldots, n.
\end{aligned}$$

Knowing that, given the sample $M = m$, the default times τ_k, $k = 1, 2, \ldots, n$, are independent random variables, we can compute the joint distribution

$$\begin{aligned}
&\mathbb{P}(\tau_1 \leqslant y_1, \ldots, \tau_n \leqslant y_n \mid M = m) \\
&\qquad = \mathbb{P}(\tau_1 \leqslant y_1 \mid M = m) \times \cdots \times \mathbb{P}(\tau_n \leqslant y_n \mid M = m),
\end{aligned}$$

conditionally to $M = m$. This yields

$$\begin{aligned}
&\mathbb{P}(\tau_1 \leqslant y_1, \ldots, \tau_n \leqslant y_n) = \int_{-\infty}^{\infty} \mathbb{P}(\tau_1 \leqslant y_1, \ldots, \tau_n \leqslant y_n \mid M = m)\phi(m)dm \\
&= \int_{-\infty}^{\infty} \mathbb{P}(\tau_1 \leqslant y_1 \mid M = m) \cdots \mathbb{P}(\tau_n \leqslant y_n \mid M = m)\phi(m)dm \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi_{X_1}^{-1}\big(\mathbb{P}(\tau_1 \leqslant y_1)\big) - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \Phi\left(\frac{\Phi_{X_n}^{-1}\big(\mathbb{P}(\tau_n \leqslant y_n)\big) - a_n m}{\sqrt{1-a_n^2}}\right)\phi(m)dm,
\end{aligned}$$

which is (9.11). □

The next corollary deals with the case where M is normally distributed.

Corollary 9.10. *Assume that M has the standard normal distribution with probability density function ϕ and is independent of $Z_1, \ldots, Z_n$. Then, the joint distribution of the default times $(\tau_k)_{k=1,2,\ldots,n}$ is given by*

$$\mathbb{P}(\tau_1 \leqslant y_1, \ldots, \tau_n \leqslant y_n) = C\big(\mathbb{P}(\tau_1 \leqslant y_1), \ldots, \mathbb{P}(\tau_n \leqslant y_n)\big),$$

where

$$C(x_1, \ldots, x_n) \\ := \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi^{-1}(x_1) - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \Phi\left(\frac{\Phi^{-1}(x_n) - a_n m}{\sqrt{1-a_n^2}}\right) \phi(m) dm,$$

$x_1, x_2, \ldots, x_n \in [0,1]$, *is the Gaussian copula with covariance matrix*

$$\Sigma = \begin{bmatrix} 1 & a_1 a_2 & \cdots & a_1 a_{n-1} & a_1 a_n \\ a_2 a_1 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & a_{n-1} a_n \\ a_n a_1 & a_n a_2 & \cdots & a_n a_{n-1} & 1 \end{bmatrix}. \tag{9.12}$$

Proof. When the random variable M is normally distributed and independent of $Z_1, \ldots, Z_n$, the random vector $(X_1, \ldots, X_n)$ has the covariance matrix (9.12), and the function

$$C(x_1, \ldots, x_n) \\ := \int_{-\infty}^{\infty} \Phi\left(\frac{\Phi^{-1}(x_1) - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \Phi\left(\frac{\Phi^{-1}(x_n) - a_n m}{\sqrt{1-a_n^2}}\right) \phi(m) dm,$$

$x_1, x_2, \ldots, x_n \in [0,1]$, is a Gaussian copula on $[0,1]^n$, built as

$$C(x_1, \ldots, x_n) = F\big(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_n)\big),$$

from the Gaussian cumulative distribution function

$$\begin{aligned} F(x_1, \ldots, x_n) &:= \int_{-\infty}^{\infty} \Phi\left(\frac{x_1 - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \Phi\left(\frac{x_n - a_n m}{\sqrt{1-a_n^2}}\right) \phi(m) dm \\ &= \int_{-\infty}^{\infty} \mathbb{P}\left(Z_1 \leqslant \frac{x_1 - a_1 m}{\sqrt{1-a_1^2}}\right) \cdots \mathbb{P}\left(Z_n \leqslant \frac{x_n - a_n m}{\sqrt{1-a_n^2}}\right) \phi(m) dm \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n \mid M = m) \phi(m) dm \end{aligned}$$

$$= \mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n), \qquad 0 \leqslant x_1, x_2, \ldots, x_n \leqslant 1,$$

of the vector $(X_1, \ldots, X_n)$, with covariance matrix given by (9.12). We conclude by Proposition 9.9. □

Exercises

Exercise 9.1 Compute the conditional probability density function of the default time τ defined in (9.2).

Exercise 9.2 Credit Default Contract. The assets of a company are modeled using a geometric Brownian motion $(S_t)_{t\in\mathbb{R}_+}$ with drift $r > 0$ under the risk-neutral probability measure $\mathbb{P}^*$. A Credit Default Contract pays \$1 as soon as the asset S_t hits a level $K > 0$. Price this contract at time $t > 0$ assuming that $S_t > K$.

Exercise 9.3

a) Check that the vector $(X_1, X_2, \ldots, X_n)$ defined in (9.9) has the covariance matrix given by (9.12).
b) Show that the vector $(X_1, X_2, \ldots, X_n)$, with covariance matrix (9.12) has standard Gaussian marginals.
c) By computing explicitly the probability density function of $(X_1, \ldots, X_n)$, recover the fact that it is a jointly Gaussian random vector with covariance matrix (9.12).

Exercise 9.4 Compute the inverse Σ^{-1} of the covariance matrix (9.12) in case $n = 2$.

Chapter 10
Credit Risk - Reduced-Form Approach

In this chapter, credit risk is estimated by modeling default probabilities using stochastic failure rate processes. In addition, information on default events is incorporated to the model by the use of exogeneous random variables and enlargement of filtrations. This is in contrast to the structural approach to credit risk of Chapter 9, in which bankruptcy is modeled from a firm's asset value. Applications are given to the pricing of default bonds.

10.1 Survival Probabilities

The reduced-form approach to credit risk relies on the concept of survival probability, defined as the probability $\mathbb{P}(\tau > t)$ that a random system with lifetime τ survives at least over t years, $t > 0$. Assuming that survival probabilities $\mathbb{P}(\tau > t)$ are strictly positive for all $t > 0$, we can compute the conditional probability for that system to survive up to time T, given that it was still functioning at time $t \in [0, T]$, as

$$\mathbb{P}(\tau > T \mid \tau > t) = \frac{\mathbb{P}(\tau > T \text{ and } \tau > t)}{\mathbb{P}(\tau > t)} = \frac{\mathbb{P}(\tau > T)}{\mathbb{P}(\tau > t)}, \qquad 0 \leqslant t \leqslant T,$$

with

$$\mathbb{P}(\tau \leqslant T \mid \tau > t) = 1 - \mathbb{P}(\tau > T \mid \tau > t)$$

$$\begin{aligned}
&= \frac{\mathbb{P}(\tau > t) - \mathbb{P}(\tau > T)}{\mathbb{P}(\tau > t)} \\
&= \frac{\mathbb{P}(\tau \leqslant T) - \mathbb{P}(\tau \leqslant t)}{\mathbb{P}(\tau > t)} \\
&= \frac{\mathbb{P}(t < \tau \leqslant T)}{\mathbb{P}(\tau > t)}, \qquad 0 \leqslant t \leqslant T. \qquad (10.1)
\end{aligned}$$

Such survival probabilities are typically found in life (or mortality) tables:

Age t	$\mathbb{P}(\tau \leqslant t+1 \mid \tau > t)$
20	0.0894%
30	0.1008%
40	0.2038%
50	0.4458%
60	0.9827%

Table 10.1: Mortality table.

The corresponding conditional survival probability distribution can be computed as follows:

$$\begin{aligned}
\mathbb{P}(\tau \in dx \mid \tau > t) &= \mathbb{P}(x < \tau \leqslant x + dx \mid \tau > t) \\
&= \mathbb{P}(\tau \leqslant x + dx \mid \tau > t) - \mathbb{P}(\tau \leqslant x \mid \tau > t) \\
&= \frac{\mathbb{P}(\tau \leqslant x + dx) - \mathbb{P}(\tau \leqslant x)}{\mathbb{P}(\tau > t)} \\
&= \frac{1}{\mathbb{P}(\tau > t)} d\mathbb{P}(\tau \leqslant x) \\
&= -\frac{1}{\mathbb{P}(\tau > t)} d\mathbb{P}(\tau > x), \qquad x > t.
\end{aligned}$$

Proposition 10.1. *The* failure rate *function, defined as*

$$\lambda(t) := \frac{\mathbb{P}(\tau \leqslant t + dt \mid \tau > t)}{dt},$$

satisfies

$$\mathbb{P}(\tau > t) = \exp\left(- \int_0^t \lambda(u) du \right), \qquad t \geqslant 0. \qquad (10.2)$$

Proof. By (10.1), we have

$$\lambda(t) := \frac{\mathbb{P}(\tau \leqslant t + dt \mid \tau > t)}{dt}$$

$$
\begin{aligned}
&= \frac{1}{\mathbb{P}(\tau > t)} \frac{\mathbb{P}(t < \tau \leqslant t + dt)}{dt} \\
&= \frac{1}{\mathbb{P}(\tau > t)} \frac{\mathbb{P}(\tau > t) - \mathbb{P}(\tau > t + dt)}{dt} \\
&= -\frac{d}{dt} \log \mathbb{P}(\tau > t) \\
&= -\frac{1}{\mathbb{P}(\tau > t)} \frac{d}{dt} \mathbb{P}(\tau > t), \qquad t > 0,
\end{aligned}
$$

and the differential equation

$$
\frac{d}{dt}\mathbb{P}(\tau > t) = -\lambda(t)\mathbb{P}(\tau > t),
$$

which can be solved as in (10.2) under the initial condition $\mathbb{P}(\tau > 0) = 1$.
□

Proposition 10.1 allows us to rewrite the (conditional) survival probability as

$$
\mathbb{P}(\tau > T \mid \tau > t) = \frac{\mathbb{P}(\tau > T)}{\mathbb{P}(\tau > t)} = \exp\left(- \int_t^T \lambda(u) du \right), \qquad 0 \leqslant t \leqslant T,
$$

with

$$
\mathbb{P}(\tau > t + h \mid \tau > t) = \mathrm{e}^{-\lambda(t)h} \simeq 1 - \lambda(t)h, \qquad [h \searrow 0],
$$

and

$$
\mathbb{P}(\tau \leqslant t + h \mid \tau > t) = 1 - \mathrm{e}^{-\lambda(t)h} \simeq \lambda(t)h, \qquad [h \searrow 0],
$$

as h tends to 0. When the failure rate $\lambda(t) = \lambda > 0$ is a constant function of time, Relation (10.2) shows that

$$
\mathbb{P}(\tau > T) = \mathrm{e}^{-\lambda T}, \qquad T \geqslant 0,
$$

i.e. τ has the exponential distribution with parameter λ. Note that given $(\tau_n)_{n \geqslant 1}$ a sequence of *i.i.d.* exponentially distributed random variables, letting

$$
T_n = \tau_1 + \tau_2 + \cdots + \tau_n, \qquad n \geqslant 1,
$$

defines the sequence of jump times of a standard Poisson process with intensity $\lambda > 0$, see Proposition 3.3.

10.2 Stochastic Default

When the random time τ is a *stopping time* with respect to $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ we have

$$\{\tau > t\} \in \mathcal{F}_t, \qquad t \geqslant 0,$$

i.e. the knowledge of whether default or bankruptcy has already occurred at time t is contained in $\mathcal{F}_t$, $t \in \mathbb{R}_+$, cf. *e.g.* Section 14.3 of Privault (2022). As a consequence, we can write

$$\mathbb{P}(\tau > t \mid \mathcal{F}_t) = \mathbb{E}\left[\mathbb{1}_{\{\tau>t\}} \mid \mathcal{F}_t\right] = \mathbb{1}_{\{\tau>t\}}, \qquad t \geqslant 0.$$

In what follows we will not assume that τ is an $\mathcal{F}_t$-stopping time, and by analogy with (10.2) we will write $\mathbb{P}(\tau > t \mid \mathcal{F}_t)$ as

$$\mathbb{P}(\tau > t \mid \mathcal{F}_t) = \exp\left(-\int_0^t \lambda_u du\right), \qquad t \geqslant 0, \tag{10.3}$$

where the failure rate function $(\lambda_t)_{t\in\mathbb{R}_+}$ is modeled as a random process adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

Remark 10.2. *The process $(\lambda_t)_{t\in\mathbb{R}_+}$ can also be chosen among the classical mean-reverting diffusion processes, including jump-diffusion processes. In Lando (1998), the process $(\lambda_t)_{t\in\mathbb{R}_+}$ is constructed as $\lambda_t := h(X_t)$, $t \in \mathbb{R}_+$, where h is a nonnegative function and $(X_t)_{t\in\mathbb{R}_+}$ is a stochastic process generating the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. The default time τ is then* defined *as*

$$\tau := \inf\left\{t \in \mathbb{R}_+ \ : \ \int_0^t h(X_u)du \geqslant L\right\},$$

where L is an exponentially distributed random variable with parameter $\mu > 0$ and distribution function $\mathbb{P}(L > x) = \mathrm{e}^{-\mu x}$, $x \geqslant 0$, independent of $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

In Remark 10.2, as τ is not an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time, we have

$$\begin{aligned}
\mathbb{P}(\tau > t \mid \mathcal{F}_t) &= \mathbb{P}\left(\int_0^t h(X_u)du < L \;\middle|\; \mathcal{F}_t\right) \\
&= \exp\left(-\mu \int_0^t h(X_u)du\right) \\
&= \exp\left(-\mu \int_0^t \lambda_u du\right), \qquad t \geqslant 0.
\end{aligned}$$

Definition 10.3. *Let $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ be the filtration defined by $\mathcal{G}_\infty := \mathcal{F}_\infty \vee \sigma(\tau)$ and*

$$\mathcal{G}_t := \big\{B \in \mathcal{G}_\infty \ : \ \exists A \in \mathcal{F}_t \text{ such that } A \cap \{\tau > t\} = B \cap \{\tau > t\}\big\}, \tag{10.4}$$

with $\mathcal{F}_t \subset \mathcal{G}_t$, $t \geqslant 0$.

In other words, $\mathcal{G}_t$ contains insider information on whether default at time τ has occurred or not before time t, and τ is a $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$-stopping time. Note that this information on τ may not be available to a generic user who has only access to the smaller filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. The next key Lemma 10.4, see Lando (1998), Guo et al. (2007), allows us to price a contingent claim given the information in the larger filtration $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$, by only using information in $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ and factoring in the default rate factor $\exp\left(-\int_t^T \lambda_u du\right)$.

Lemma 10.4. *(Guo et al. (2007), Theorem 1) For any $\mathcal{F}_T$-measurable integrable random variable F, we have*

$$\begin{aligned}\mathbb{E}\left[F\mathbb{1}_{\{\tau>T\}} \mid \mathcal{G}_t\right] &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\big[F\mathbb{P}(\tau>T\mid\tau>t)\mid\mathcal{F}_t\big]\\ &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[F\exp\left(-\int_t^T\lambda_u du\right)\ \middle|\ \mathcal{F}_t\right],\quad 0\leqslant t\leqslant T.\end{aligned}$$

Proof. By (10.3) we have

$$\frac{\mathbb{P}(\tau>T\mid\mathcal{F}_T)}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)} = \frac{\mathrm{e}^{-\int_0^T\lambda_u du}}{\mathrm{e}^{-\int_0^t\lambda_u du}} = \exp\left(-\int_t^T\lambda_u du\right),$$

hence, since F is $\mathcal{F}_T$-measurable,

$$\begin{aligned}\mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[F\exp\left(-\int_t^T\lambda_u du\right)\ \middle|\ \mathcal{F}_t\right] &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[F\frac{\mathbb{P}(\tau>T\mid\mathcal{F}_T)}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)}\ \middle|\ \mathcal{F}_t\right]\\ &= \frac{\mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)}\mathbb{E}\big[F\mathbb{E}[\mathbb{1}_{\{\tau>T\}}\mid\mathcal{F}_T]\mid\mathcal{F}_t\big]\\ &= \frac{\mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)}\mathbb{E}\big[\mathbb{E}[F\mathbb{1}_{\{\tau>T\}}\mid\mathcal{F}_T]\mid\mathcal{F}_t\big]\\ &= \mathbb{1}_{\{\tau>t\}}\frac{\mathbb{E}\big[F\mathbb{1}_{\{\tau>T\}}\,\big|\,\mathcal{F}_t\big]}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)}\\ &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\big[F\mathbb{1}_{\{\tau>T\}}\,\big|\,\mathcal{G}_t\big]\\ &= \mathbb{E}\big[F\mathbb{1}_{\{\tau>T\}}\,\big|\,\mathcal{G}_t\big],\qquad 0\leqslant t\leqslant T.\end{aligned}$$

In the last step of the above argument, we used the key relation

$$\mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[F\mathbb{1}_{\{\tau>T\}}\ \middle|\ \mathcal{G}_t\right] = \frac{\mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau>t\mid\mathcal{F}_t)}\mathbb{E}\big[F\mathbb{1}_{\{\tau>T\}}\mid\mathcal{F}_t\big],$$

cf. Relation (75.2) in § XX-75 page 186 of Dellacherie et al. (1992), Theorem VI-3-14 page 371 of Protter (2004), and Lemma 3.1 of Elliott et al.

(2000), under the conditional probability measure $\mathbb{P}_{|\mathcal{F}_t}$, $0 \leqslant t \leqslant T$. Indeed, according to (10.4), for any $B \in \mathcal{G}_t$ we have, for some event $A \in \mathcal{F}_t$,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_B \mathbb{1}_{\{\tau>t\}} F \mathbb{1}_{\{\tau>T\}}\right] &= \mathbb{E}\left[\mathbb{1}_{B\cap\{\tau>t\}} F \mathbb{1}_{\{\tau>T\}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{A\cap\{\tau>t\}} F \mathbb{1}_{\{\tau>T\}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_A \mathbb{1}_{\{\tau>t\}} F \mathbb{1}_{\{\tau>T\}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_A \mathbb{1}_{\{\tau>t\}} \frac{\mathbb{E}[\mathbb{1}_{\{\tau>t\}} \mid \mathcal{F}_t]}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} F \mathbb{1}_{\{\tau>T\}}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{\tau>t\}} \mid \mathcal{F}_t]}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} F \mathbb{1}_{\{\tau>T\}}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{\tau>t\}} \mid \mathcal{F}_t]}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{1}_A \mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{\tau>t\}} \mid \mathcal{F}_t]}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{1}_A \mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{1}_B \mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]\right],
\end{aligned}$$

hence by a standard characterization of conditional expectations, see *e.g.* Relation (A.26), we have

$$\mathbb{E}\big[\mathbb{1}_{\{\tau>t\}} F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{G}_t\big] = \frac{\mathbb{1}_{\{\tau>t\}}}{\mathbb{P}(\tau > t \mid \mathcal{F}_t)} \mathbb{E}\big[F \mathbb{1}_{\{\tau>T\}} \mid \mathcal{F}_t\big]$$

□

Taking $F = 1$ in Lemma 10.4 allows one to write the survival probability up to time T, given the information known up to time t, as

$$\begin{aligned}
\mathbb{P}(\tau > T \mid \mathcal{G}_t) &= \mathbb{E}\left[\mathbb{1}_{\{\tau>T\}} \mid \mathcal{G}_t\right] \qquad (10.5) \\
&= \mathbb{1}_{\{\tau>t\}} \mathbb{E}\left[\exp\left(-\int_t^T \lambda_u du\right) \,\middle|\, \mathcal{F}_t\right], \quad 0 \leqslant t \leqslant T.
\end{aligned}$$

In particular, applying Lemma 10.4 for $t = T$ and $F = 1$ shows that

$$\mathbb{E}\left[\mathbb{1}_{\{\tau>t\}} \mid \mathcal{G}_t\right] = \mathbb{1}_{\{\tau>t\}},$$

which shows that $\{\tau > t\} \in \mathcal{G}_t$ for all $t > 0$, and recovers the fact that τ is a $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$-stopping time, while in general, τ is not $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-stopping time.

The computation of $\mathbb{P}(\tau > T \mid \mathcal{G}_t)$ according to (10.5) is then similar to that of a bond price, by considering the failure rate $\lambda(t)$ as a "virtual" short-term interest rate. In particular the failure rate $\lambda(t,T)$ can be modeled in the HJM framework, cf. *e.g.* Chapter 18.3 of Privault (2022), and

$$\mathbb{P}(\tau > T \mid \mathcal{G}_t) = \mathbb{E}\left[\exp\left(-\int_t^T \lambda(t,u)du\right) \Bigm| \mathcal{F}_t\right]$$

can then be computed by applying HJM bond pricing techniques.

The computation of expectations given $\mathcal{G}_t$ as in Lemma 10.4 can be useful for pricing under insider trading, in which the insider has access to the augmented filtration $\mathcal{G}_t$ while the ordinary trader has only access to $\mathcal{F}_t$, therefore generating two different prices $\mathbb{E}^*[F \mid \mathcal{F}_t]$ and $\mathbb{E}^*[F \mid \mathcal{G}_t]$ for the same claim payoff F under the same risk-neutral probability measure $\mathbb{P}^*$. This leads to the issue of computing the dynamics of the underlying asset price by decomposing it using a $(\mathcal{F}_t)_{t\geqslant 0}$-martingale *vs.* a $(\mathcal{G}_t)_{t\geqslant 0}$-martingale instead of using different forward measures as in *e.g.* § 19.1 of Privault (2022). This can be obtained by the technique of enlargement of filtration, cf. Jeulin (1980), Jacod (1985), Yor (1985), Elliott and Jeanblanc (1999).

10.3 Defaultable Bonds

Bond pricing models are generally based on the terminal condition $P(T,T) =$ \$1 according to which the bond payoff at maturity is always equal to \$1, and default does not occurs. In this chapter we allow for the possibility of default at a random time τ, in which case the terminal payoff of a bond is allowed to vanish at maturity.

The price $P_d(t,T)$ at time t of a default bond with maturity T, (random) default time τ and (possibly random) recovery rate $\xi \in [0,1]$ is given by

$$\begin{aligned} P_d(t,T) = {} & \mathbb{E}^*\left[\mathbb{1}_{\{\tau>T\}} \exp\left(-\int_t^T r_u du\right) \Bigm| \mathcal{G}_t\right] \\ & + \mathbb{E}^*\left[\xi \mathbb{1}_{\{\tau\leqslant T\}} \exp\left(-\int_t^T r_u du\right) \Bigm| \mathcal{G}_t\right], \qquad 0 \leqslant t \leqslant T. \end{aligned}$$

Proposition 10.5. *The default bond with maturity T and default time τ can be priced at time $t \in [0,T]$ as*

$$P_d(t,T) = \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T(r_u+\lambda_u)du\right) \ \Big| \ \mathcal{F}_t\right]$$
$$+\mathbb{E}^*\left[\xi\mathbb{1}_{\{\tau\leqslant T\}}\exp\left(-\int_t^T r_u du\right) \ \Big| \ \mathcal{G}_t\right], \quad 0 \leqslant t \leqslant T.$$

Proof. We take $F = \exp\left(-\int_t^T r_u du\right)$ in Lemma 10.4, which shows that

$$\mathbb{E}^*\left[\mathbb{1}_{\{\tau>T\}}\exp\left(-\int_t^T r_u du\right) \ \Big| \ \mathcal{G}_t\right] = \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T(r_u+\lambda_u)du\right) \ \Big| \ \mathcal{F}_t\right],$$

cf. *e.g.* Lando (1998), Duffie and Singleton (2003), Guo et al. (2007). □

In the case of complete default (zero-recovery), we have $\xi = 0$ and

$$P_d(t,T) = \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T(r_s+\lambda_s)ds\right) \ \Big| \ \mathcal{F}_t\right], \quad 0 \leqslant t \leqslant T. \quad (10.6)$$

From the above expression (10.6) we note that the effect of the presence of a default time τ is to decrease the bond price, which can be viewed as an increase of the short rate by the amount λ_u. In a simple setting where the interest rate $r > 0$ and failure rate $\lambda > 0$ are constant, the default bond price becomes

$$P_d(t,T) = \mathbb{1}_{\{\tau>t\}}\mathrm{e}^{-(r+\lambda)(T-t)}, \qquad 0 \leqslant t \leqslant T.$$

In this case, the failure rate λ can be estimated at time $t \in [0,T]$ from a default bond price $P_d(t,T)$ and a non-default bond price $P(t,T) = \mathrm{e}^{-(T-t)r}$ as

$$\lambda = \frac{1}{T-t}\log\frac{P(t,T)}{P_d(t,T)}.$$

Finally, as in *e.g.* Proposition 19.1 in Privault (2022) the bond price (10.6) can also be expressed under the forward measure $\widehat{\mathbb{P}}$ with maturity T, as

$$\begin{aligned}
P_d(t,T) &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T(r_s+\lambda_s)ds\right) \ \Big| \ \mathcal{F}_t\right] \\
&= \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T r_s ds\right) \ \Big| \ \mathcal{F}_t\right]\widehat{\mathbb{E}}\left[\exp\left(-\int_t^T \lambda_s ds\right) \ \Big| \ \mathcal{F}_t\right] \\
&= \mathbb{1}_{\{\tau>t\}}N_t\widehat{\mathbb{P}}(\tau > T \mid \mathcal{G}_t),
\end{aligned}$$

where $(N_t)_{t\in\mathbb{R}_+}$ is the numéraire process

$$N_t := P(t,T) = \mathbb{E}^*\left[\exp\left(-\int_t^T r_s ds\right) \ \Big| \ \mathcal{F}_t\right], \qquad 0 \leqslant t \leqslant T,$$

and by (10.5),

$$\widehat{\mathbb{P}}(\tau > T \mid \mathcal{G}_t) = \mathbb{1}_{\{\tau>t\}} \widehat{\mathbb{E}}\left[\exp\left(- \int_t^T \lambda_s ds \right) \Bigm| \mathcal{F}_t \right]$$

is the survival probability under the forward measure $\widehat{\mathbb{P}}$ defined as

$$\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} := \frac{N_T}{N_0} e^{-\int_0^T r_t dt},$$

see Chen and Huang (2001) and Chen et al. (2008).

Estimating the default rates

Recall that the price of a default bond with maturity T, (random) default time τ and (possibly random) recovery rate $\xi \in [0,1]$ is given by

$$\begin{aligned} P_d(t,T) &= \mathbb{1}_{\{\tau>t\}} \mathbb{E}^*\left[\exp\left(- \int_t^T (r_u + \lambda_u) du \right) \Bigm| \mathcal{F}_t \right] \\ &\quad + \mathbb{E}^*\left[\xi \mathbb{1}_{\{\tau \leqslant T\}} \exp\left(- \int_t^T r_u du \right) \Bigm| \mathcal{G}_t \right], \qquad 0 \leqslant t \leqslant T, \end{aligned}$$

where ξ denotes the recovery rate. We consider a simplified deterministic step function model with zero recovery rate and tenor structure

$$\{t = T_0 < T_1 < \cdots < T_n = T\},$$

where

$$r(t) = \sum_{l=0}^{n-1} r_l \mathbb{1}_{(T_l, T_{l+1}]}(t) \quad \text{and} \quad \lambda(t) = \sum_{l=0}^{n-1} \lambda_l \mathbb{1}_{(T_l, T_{l+1}]}(t), \quad t \geqslant 0. \quad (10.7)$$

i) Estimating the default rates from default bond prices.

From Proposition 10.5, we have

$$\begin{aligned} P_d(t, T_k) &= \mathbb{1}_{\{\tau>t\}} \exp\left(- \int_t^{T_k} (r(u) + \lambda(u)) du \right) \\ &= \mathbb{1}_{\{\tau>t\}} \exp\left(- \sum_{l=0}^{k-1} (r_l + \lambda_l)(T_{l+1} - T_l) \right), \end{aligned}$$

$k = 1, 2, \ldots, n$, from which we can infer

$$\lambda_k = -r_k + \frac{1}{T_{k+1} - T_k} \log \frac{P_d(t, T_k)}{P_d(t, T_{k+1})} > 0, \quad k = 0, 1, \ldots, n-1.$$

ii) Estimating (implied) default probabilities $\mathbb{P}^*(\tau < T \mid \mathcal{G}_t)$ from default rates.

Based on the expression

$$\begin{aligned}\mathbb{P}^*(\tau > T \mid \mathcal{G}_t) &= \mathbb{E}^*\left[\mathbb{1}_{\{\tau>T\}} \mid \mathcal{G}_t\right] \\ &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T \lambda_u du\right) \,\Big|\, \mathcal{F}_t\right], \qquad 0 \leqslant t \leqslant T,\end{aligned} \tag{10.8}$$

of the survival probability up to time T, see (10.3), and given the information known up to time t, in terms of the hazard rate process $(\lambda_u)_{u\in\mathbb{R}_+}$ adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, we find

$$\begin{aligned}\mathbb{P}(\tau > T \mid \mathcal{G}_{T_k}) &= \mathbb{1}_{\{\tau>T_k\}} \exp\left(-\int_{T_k}^T \lambda_u du\right) \\ &= \mathbb{1}_{\{\tau>t\}} \exp\left(-\sum_{l=k}^{n-1} \lambda_l (T_{l+1} - T_l)\right), \quad k = 0, 1, \ldots, n-1,\end{aligned}$$

where

$$\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\{\tau \leqslant u\} : 0 \leqslant u \leqslant t), \quad t \geqslant 0,$$

i.e. $\mathcal{G}_t$ contains the additional information on whether default at time τ has occurred or not before time t.

In Table 10.2, bond ratings are determined according to hazard (or failure) rate thresholds.

Bond Credit	Moody's		S & P	
Ratings	Municipal	Corporate	Municipal	Corporate
Aaa/AAAs	0.00	0.52	0.00	0.60
Aa/AA	0.06	0.52	0.00	1.50
A/A	0.03	1.29	0.23	2.91
Baa/BBB	0.13	4.64	0.32	10.29
Ba/BB	2.65	19.12	1.74	29.93
B/B	11.86	43.34	8.48	53.72
Caa-C/CCC-C	16.58	69.18	44.81	69.19
Investment Grade	0.07	2.09	0.20	4.14
Non-Invest. Grade	4.29	31.37	7.37	42.35
All	0.10	9.70	0.29	12.98

Table 10.2: Cumulative historic default rates (in percentage).*

* Sources: Moody's, S&P.

Exercises

Exercise 10.1 Consider a standard zero-coupon bond with constant yield $r > 0$ and a defaultable (risky) bond with constant yield r_d and default probability $\alpha \in (0,1)$. Find a relation between r, r_d, α and the bond maturity T.

Exercise 10.2 A standard zero-coupon bond with constant yield $r > 0$ and maturity T is priced $P(t,T) = \mathrm{e}^{-(T-t)r}$ at time $t \in [0,T]$. Assume that the company can get bankrupt at a random time $t + \tau$, and default on its final \$1 payment if $\tau < T - t$.

a) Explain why the defaultable bond price $P_d(t,T)$ can be expressed as

$$P_d(t,T) = \mathrm{e}^{-(T-t)r}\mathbb{E}^*\big[\mathbb{1}_{\{\tau > T-t\}}\big]. \tag{10.9}$$

b) Assuming that the default time τ is exponentially distributed with parameter $\lambda > 0$, compute the default bond price $P_d(t,T)$ using (10.9).
c) Find a formula that can estimate the parameter λ from the risk-free rate r and the market data $P_M(t,T)$ of the defaultable bond price at time $t \in [0,T]$.

Exercise 10.3 Consider an interest rate process $(r_t)_{t\in\mathbb{R}_+}$ and a default rate process $(\lambda_t)_{t\in\mathbb{R}_+}$, modeled according to the Vasicek processes

$$\begin{cases} dr_t = -ar_t dt + \sigma dB_t^{(1)}, \\ d\lambda_t = -b\lambda_t dt + \eta dB_t^{(2)}, \end{cases}$$

where $\big(B_t^{(1)}\big)_{t\in\mathbb{R}_+}$ and $\big(B_t^{(2)}\big)_{t\in\mathbb{R}_+}$ are standard $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-Brownian motions with correlation $\rho \in [-1,1]$, and $dB_t^{(1)} \bullet dB_t^{(2)} = \rho dt$.

a) Taking $r_0 := 0$, show that we have

$$\int_t^T r_s ds = C(a,t,T)r_t + \sigma \int_t^T C(a,s,T)dB_s^{(1)},$$

and

$$\int_t^T \lambda_s ds = C(b,t,T)\lambda_t + \eta \int_t^T C(b,s,T)dB_s^{(2)},$$

where

$$C(a,t,T) = -\frac{1}{a}(\mathrm{e}^{-(T-t)a} - 1).$$

b) Show that the random variable

$$\int_t^T r_s ds + \int_t^T \lambda_s ds$$

is has a Gaussian distribution, and compute its conditional mean

$$\mathbb{E}^*\left[\int_t^T r_s ds + \int_t^T \lambda_s ds \,\Big|\, \mathcal{F}_t\right]$$

and variance

$$\mathrm{Var}\left[\int_t^T r_s ds + \int_t^T \lambda_s ds \,\Big|\, \mathcal{F}_t\right],$$

conditionally to $\mathcal{F}_t$.

Exercise 10.4 (Exercise 10.3 continued). Consider a (random) default time τ with cumulative distribution function

$$\mathbb{P}(\tau > t \mid \mathcal{F}_t) = \exp\left(-\int_0^t \lambda_u du\right), \qquad t \geqslant 0,$$

where λ_t is a (random) default rate process which is adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Recall that the probability of survival up to time T, given the information known up to time t, is given by

$$\mathbb{P}(\tau > T \mid \mathcal{G}_t) = \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T \lambda_u du\right) \,\Big|\, \mathcal{F}_t\right],$$

where $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\{\tau < u\} : 0 \leqslant u \leqslant t)$, $t \in \mathbb{R}_+$, is the filtration defined by adding the default time information to the history $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. In this framework, the price $P(t,T)$ of defaultable bond with maturity T, short-term interest rate r_t and (random) default time τ is given by

$$\begin{aligned} P(t,T) &= \mathbb{E}^*\left[\mathbb{1}_{\{\tau>T\}}\exp\left(-\int_t^T r_u du\right) \,\Big|\, \mathcal{G}_t\right] \\ &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}^*\left[\exp\left(-\int_t^T (r_u+\lambda_u) du\right) \,\Big|\, \mathcal{F}_t\right]. \end{aligned} \tag{10.10}$$

a) Give a justification for the fact that

$$\mathbb{E}^*\left[\exp\left(-\int_t^T (r_u+\lambda_u) du\right) \,\Big|\, \mathcal{F}_t\right]$$

can be written as a function $F(t, r_t, \lambda_t)$ of t, r_t and λ_t, $t \in [0,T]$.

b) Show that

$$t \longmapsto \exp\left(-\int_0^t (r_s + \lambda_s) ds\right) \mathbb{E}^*\left[\exp\left(-\int_t^T (r_u + \lambda_u) du\right) \Bigm| \mathcal{F}_t\right]$$

is an $(\mathcal{F}_t)_{t \geqslant 0}$-martingale under $\mathbb{P}$.

c) Use the Itô formula with two variables to derive a PDE on $\mathbb{R}^2$ for the function $F(t, x, y)$.

d) Compute $P(t, T)$ from its expression (10.10) as a conditional expectation.

e) Show that the solution $F(t, x, y)$ to the 2-dimensional PDE of Question (c) is

$$\begin{aligned} F(t, x, y) = & \exp\left(-C(a, t, T)x - C(b, t, T)y\right) \\ & \times \exp\left(\frac{\sigma^2}{2}\int_t^T C^2(a, s, T) ds + \frac{\eta^2}{2}\int_t^T C^2(b, s, T) ds\right) \\ & \times \exp\left(\rho\sigma\eta \int_t^T C(a, s, T) C(b, s, T) ds\right). \end{aligned}$$

f) Show that the defaultable bond price $P(t, T)$ can also be written as

$$P(t, T) = \mathrm{e}^{U(t,T)} \mathbb{P}(\tau > T \mid \mathcal{G}_t) \mathbb{E}^*\left[\exp\left(-\int_t^T r_s ds\right) \Bigm| \mathcal{F}_t\right],$$

where

$$U(t, T) = \rho\frac{\sigma\eta}{ab}\left(T - t - C(a, t, T) - C(b, t, T) + C(a + b, t, T)\right).$$

g) By partial differentiation of $\log P(t, T)$ with respect to T, compute the corresponding instantaneous short rate

$$f(t, T) = -\frac{\partial}{\partial T}\log P(t, T).$$

h) Show that $\mathbb{P}(\tau > T \mid \mathcal{G}_t)$ can be written using an HJM type default rate as

$$\mathbb{P}(\tau > T \mid \mathcal{G}_t) = \mathbb{1}_{\{\tau > t\}} \exp\left(-\int_t^T f_2(t, u) du\right),$$

where

$$f_2(t, u) = \lambda_t \mathrm{e}^{-(u-t)b} - \frac{\eta^2}{2} C^2(b, t, u).$$

i) Show how the result of Question (f) can be simplified when the processes $\big(B_t^{(1)}\big)_{t \in \mathbb{R}_+}$ and $\big(B_t^{(2)}\big)_{t \in \mathbb{R}_+}$ are independent.

Chapter 11
Credit Derivatives

Credit derivatives are option contracts that can be used as a protection against default risk in a creditor/debtor relationship, by transferring risk to a third party. This chapter reviews the construction and properties of several credit derivatives such as Collateralized Debt Obligations (CDOs) and Credit Default Swaps (CDSs). We also address the issue of counterparty default risk via the computation of Credit Valuation Adjustments (CVAs).

11.1 Credit Default Swaps (CDS)

Detailed information on the status of credit default swap (CDS) contracts can be obtained from the Bank for International Settlements. We note in particular that the outstanding notional amount of CDS contracts has decreased from its historical high of \$61.2 trillion at year-end 2007 to \$7.6 trillion at year-end 2019.

In this chapter, we work with a tenor structure $\{t = T_i < \cdots < T_j = T\}$ that represents a sequence of possible payment dates. We also let τ be a default time, and given a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, we consider the enlarged filtration $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ given by $\mathcal{G}_t := \mathcal{F}_t \vee \sigma(\tau)$, $t \geqslant 0$, which contains the additional information given by τ, see Definition 10.3.

Definition 11.1. *A Credit Default Swap (CDS) is a contract consisting in*

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

- A premium leg*: the buyer is purchasing protection at time t against default at time T_k, $k = i+1, \ldots, j$, and has to make a fixed spread payment $S_t^{i,j}$ at times $T_{i+1}, \ldots, T_j$ between t and T in compensation.*
- A protection leg*: the seller or issuer of the contract makes a compensation payment $1 - \xi_{k+1}$ to the buyer in case default occurs at time T_{k+1}, $k = i, \ldots, j-1$, where ξ_{k+1} is the* recovery rate*.*

In the sequel, we let

$$P(t, T_k) := \mathbb{1}_{\{\tau > t\}} \mathbb{E}\left[\exp\left(-\int_t^{T_k} (r_s + \lambda_s) ds\right) \Big| \mathcal{F}_t\right], \quad 0 \leqslant t \leqslant T_k,$$

denote the default bond price with maturity T_k, $k = i, \ldots, j-1$, see Lemma 10.4 and Proposition 10.5.

Proposition 11.2. *The discounted value at time t of the premium leg is given by*

$$V^p(t, T) = S_t^{i,j} P(t, T_i, T_j), \tag{11.1}$$

where $\delta_k := T_{k+1} - T_k$, $k = i, \ldots, j-1$, and

$$P(t, T_i, T_j) := \sum_{k=i}^{j-1} \delta_k P(t, T_{k+1})$$

is the (default) annuity numéraire, cf. e.g. *Relation (19.27) in Privault (2022).*

Proof. We have

$$\begin{aligned}
V^p(t, T) &= \mathbb{E}\left[\sum_{k=i}^{j-1} S_t^{i,j} \delta_k \mathbb{1}_{\{\tau > T_{k+1}\}} \exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Big| \mathcal{G}_t\right] \\
&= \sum_{k=i}^{j-1} S_t^{i,j} \delta_k \mathbb{E}\left[\mathbb{1}_{\{\tau > T_{k+1}\}} \exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Big| \mathcal{G}_t\right] \\
&= S_t^{i,j} \sum_{k=i}^{j-1} \delta_k P(t, T_{k+1}) \\
&= S_t^{i,j} P(t, T_i, T_j).
\end{aligned}$$

□

For simplicity, in the above proof we have ignored a possible accrual interest term over the time interval $[T_k, \tau]$ when $\tau \in [T_k, T_{k+1}]$ in the above value of the premium leg. Similarly, we have the following result.

Proposition 11.3. *The value at time t of the protection leg is given by*

$$V^d(t,T) := \mathbb{E}\left[\sum_{k=i}^{j-1} \mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Bigm| \mathcal{G}_t\right], \tag{11.2}$$

where ξ_{k+1} is the recovery rate associated with the maturity T_{k+1}, $k = i,\ldots,j-1$.

In the case of a non-random recovery rate ξ_k, the value of the protection leg becomes

$$\sum_{k=i}^{j-1}(1-\xi_{k+1})\mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Bigm| \mathcal{G}_t\right].$$

The spread $S_t^{i,j}$ is computed by equating the values of the premium (11.1) and protection (11.2) legs as $V^p(t,T) = V^d(t,T)$, *i.e.* from the relation

$$\begin{aligned} V^p(t,T) &= S_t^{i,j} P(t,T_i,T_j) \\ &= \mathbb{E}\left[\sum_{k=i}^{j-1} \mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Bigm| \mathcal{G}_t\right] \\ &= V^d(t,T), \end{aligned}$$

which yields

$$S_t^{i,j} = \frac{1}{P(t,T_i,T_j)}\sum_{k=i}^{j-1}\mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \Bigm| \mathcal{G}_t\right]. \tag{11.3}$$

The spread $S_t^{i,j}$, which is quoted in basis points per year and paid at regular time intervals, gives protection against defaults on payments of \$1. For a notional amount N the premium payment will become $N \times S_t^{i,j}$.

Fig. 11.1: CDS price evolution on Credit Suisse, 2023.

In the case of a constant recovery rate ξ, we find

$$S_t^{i,j} = \frac{1-\xi}{P(t,T_i,T_j)} \sum_{k=i}^{j-1} \mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau) \exp\left(-\int_t^{T_{k+1}} r_s ds \right) \Bigm| \mathcal{G}_t \right],$$

and if τ is constrained to take values in the tenor structure $\{t = T_i, \ldots, T_j\}$, we get

$$S_t^{i,j} = \frac{1-\xi}{P(t,T_i,T_j)} \mathbb{E}\left[\mathbb{1}_{(t,T]}(\tau) \exp\left(-\int_t^{\tau} r_s ds \right) \Bigm| \mathcal{G}_t \right].$$

The buyer of a Credit Default Swap (CDS) is purchasing protection at time t against default at time T_k, $k = i+1, \ldots, j$, by making a fixed payment $S_t^{i,j}$ (the premium leg) at times $T_{i+1}, \ldots, T_j$. On the other hand, the issuer of the contract makes a payment $1 - \xi_{k+1}$ to the buyer in case default occurs at time T_{k+1}, $k = i, \ldots, j-1$.

The contract is priced in terms of the swap rate $S_t^{i,j}$ (or spread) computed by equating the values $V^d(t,T)$ and $V^p(t,T)$ of the protection and premium legs, and acts as a compensation that makes the deal fair to both parties. Recall that from (11.3) and Lemma 10.4, we have

$$\begin{aligned}
S_t^{i,j} &= \frac{1}{P(t,T_i,T_j)} \sum_{k=i}^{j-1} \mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1}) \exp\left(-\int_t^{T_{k+1}} r_s ds \right) \Bigm| \mathcal{G}_t \right] \\
&= \frac{1}{P(t,T_i,T_j)} \sum_{k=i}^{j-1} \mathbb{E}\left[(\mathbb{1}_{\{T_k<\tau\}} - \mathbb{1}_{\{T_{k+1}<\tau\}})(1-\xi_{k+1}) \exp\left(-\int_t^{T_{k+1}} r_s ds \right) \Bigm| \mathcal{G}_t \right] \\
&= \frac{\mathbb{1}_{\{\tau>t\}}}{P(t,T_i,T_j)} \sum_{k=i}^{j-1} \mathbb{E}\left[(1-\xi_{k+1}) \left(\exp\left(-\int_t^{T_k} \lambda_s ds \right) - \exp\left(-\int_t^{T_{k+1}} \lambda_s ds \right) \right) \right. \\
&\qquad \left. \times \exp\left(-\int_t^{T_{k+1}} r_s ds \right) \Bigm| \mathcal{F}_t \right].
\end{aligned}$$

Estimating a deterministic failure rate

In case the rates $r(s)$, $\lambda(s)$ and the recovery rate ξ_{k+1} are deterministic, the above spread can be written as

$$\begin{aligned}S_t^{i,j} P(t, T_i, T_j) = \mathbb{1}_{\{\tau > t\}} \sum_{k=i}^{j-1} (1 - \xi_{k+1}) \exp\left(-\int_t^{T_{k+1}} r(s) ds\right) \\ \times \left(\exp\left(-\int_t^{T_k} \lambda(s) ds\right) - \exp\left(-\int_t^{T_{k+1}} \lambda(s) ds\right)\right).\end{aligned}$$

Given that

$$P(t, T_i, T_j) = \sum_{k=i}^{j-1} (T_{k+1} - T_k) P(t, T_{k+1}), \qquad T_i \leqslant t \leqslant T_{i+1},$$

we can write

$$\begin{aligned}&S_t^{i,j} \sum_{k=i}^{j-1} (T_{k+1} - T_k) \exp\left(-\int_t^{T_{k+1}} (r(s) + \lambda(s)) ds\right) \\ &= \mathbb{1}_{\{\tau > t\}} \sum_{k=i}^{j-1} (1 - \xi_k) \exp\left(-\int_t^{T_{k+1}} r(s) ds\right) \left(\exp\left(-\int_t^{T_k} \lambda(s) ds\right) - \exp\left(-\int_t^{T_{k+1}} \lambda(s) ds\right)\right).\end{aligned}$$

In particular, when $r(t)$ and $\lambda(t)$ are written as in (10.7) and assuming that $\xi_k = \xi$ is constant, $k = i, \ldots, j$, we get, with $t = T_i$ and writing $\delta_k = T_{k+1} - T_k$, $k = i, \ldots, j-1$,

$$\begin{aligned}&S_{T_i}^{i,j} \sum_{k=i}^{j-1} \delta_k \exp\left(-\sum_{p=i}^{k} \delta_p (r_p + \lambda_p)\right) \\ &\quad = \mathbb{1}_{\{\tau > t\}} (1 - \xi) \sum_{k=i}^{j-1} \exp\left(-\sum_{p=i}^{k} \delta_p (r_p + \lambda_p)\right) \left(\mathrm{e}^{\delta_k \lambda_k} - 1\right).\end{aligned}$$

Assuming further that $\lambda_k = \lambda$ is constant, $k = i, \ldots, j$, we have

$$\begin{aligned}&S_{T_i}^{i,j} \sum_{k=i}^{j-1} \delta_k \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right) \\ &\quad = (1 - \xi) \sum_{k=i}^{j-1} \left(\mathrm{e}^{\lambda \delta_k} - 1\right) \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right),\end{aligned} \tag{11.4}$$

which can be solved numerically for λ, cf. Sections 4 and 5 of Castellacci (2008) for the JP Morgan model, and Exercises 11.1-11.2.

11.2 Collateralized Debt Obligations (CDO)

The CDO market size was over \$200 billion before the 2008 financial crisis, and subsequently almost vanished. The market is currently rebounding and is expected to expand to \$40 billion by 2027. Consider a portfolio consisting of $N = j - i$ bonds with default times $\tau_k \in (T_k, T_{k+1}]$, $k = i, \ldots, j-1$, and recovery rates $\xi_k \in [0,1]$, $k = i+1, \ldots, j$.

A synthetic CDO is a structured investment product constructed by splitting the above portfolio into n ordered tranches numbered $i = 1, 2, \ldots, n$, where tranche n°i represents a percentage $p_i\%$ of the total portfolio value. We let

$$\alpha_l := p_1 + p_2 + \cdots + p_l, \qquad l = 1, 2, \ldots, n, \tag{11.5}$$

denote the corresponding cumulative percentages, with $\alpha_0 = 0$ and $\alpha_n = p_1 + p_2 + \cdots + p_n = 100\%$.

The tranches are ordered according to decreasing default risk, tranche n°1 being the riskiest one ("equity tranche"), and tranche n°n being the safest ("senior tranche"), while the intermediate tranches are referred to as "mezzanine tranches". In practice, losses occur first to the "equity" tranches, next to the "mezzanine" tranche holders, and finally to "senior" tranches.

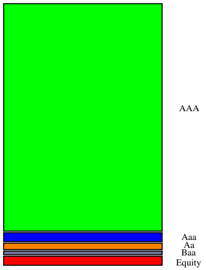


Fig. 11.2: A representation of CDO tranches.

CDOs can attract different types of investors.

- Unfunded investors (usually for the higher tranches) are receiving premiums and make payments in case of default.

- Funded (or short) investors, usually in the lower tranches, are investing in risky bonds to receive principal payments at maturity, and they are the first in line to incur losses.
- A CDO can also be used as a Credit Default Swap (CDS) for the "short investors" who make premium payments in exchange for credit protection in case of default.

The market for synthetic CDOs has been significantly reduced since the 2006-2008 subprime crisis.

Synthetic CDOs are based on $N = j - i$ bonds that can potentially generate a cumulative loss

$$L_t := \sum_{l=i}^{j-1} (1 - \xi_{l+1}) \mathbb{1}_{\{\tau_l \leqslant t\}} \in [0, N],$$

at time $t \in [T_i, T_j]$, based on the default time τ_l and recovery rate ξ_{l+1} of each involved bond, $k = i, \ldots, j-1$, with $N = j - i$.

When the first loss occurs, tranche n°1 is the first in line, and it loses the amount

$$L_t^1 = L_t \mathbb{1}_{\{L_t \leqslant p_1 N\}} + N p_1 \mathbb{1}_{\{L_t > p_1 N\}} = N \min(L_t / N, p_1).$$

In case $L_t > p_1 N$, then tranche n°2 takes the remaining loss up to the amount $N p_2$, that means the loss L_t^2 of tranche n°2 is

$$\begin{aligned}
L_t^2 &= (L_t - N p_1) \mathbb{1}_{\{p_1 N < L_t \leqslant (p_1 + p_2) N\}} + N p_2 \mathbb{1}_{\{L_t > (p_1 + p_2) N\}} \\
&= (L_t - N p_1) \mathbb{1}_{\{p_1 N < L_t \leqslant \alpha_2 N\}} + N p_2 \mathbb{1}_{\{L_t > \alpha_2 N\}} \\
&= (L_t - N p_1)^+ \mathbb{1}_{\{L_t \leqslant \alpha_2 N\}} + N p_2 \mathbb{1}_{\{L_t > \alpha_2 N\}} \\
&= \min((L_t - N p_1)^+, N p_2) \\
&= \mathrm{Max}(\min(L_t, N p_1 + N p_2) - N p_1, 0) \\
&= \mathrm{Max}(\min(L_t, N \alpha_2) - N p_1, 0).
\end{aligned}$$

By induction, the potential loss taken by tranche n°i is given by

$$\begin{aligned}
L_t^i &= (L_t - N \alpha_{i-1}) \mathbb{1}_{\{\alpha_{i-1} N < L_t \leqslant \alpha_i N\}} + N p_i \mathbb{1}_{\{L_t > \alpha_i N\}} \\
&= (L_t - N \alpha_{i-1})^+ \mathbb{1}_{\{L_t \leqslant \alpha_i N\}} + N p_i \mathbb{1}_{\{L_t > \alpha_i N\}} \\
&= \min((L_t - N \alpha_{i-1})^+, N p_i) \\
&= \mathrm{Max}(\min(L_t, N \alpha_i) - N \alpha_{i-1}, 0),
\end{aligned}$$

where $\alpha_i := p_1 + p_2 + \cdots + p_i$, $i = 1, 2, \ldots, n$.

In the end, tranche n°n will take the loss

$$L_t^n = (L_t - N\alpha_{n-1})\mathbb{1}_{\{\alpha_{n-1}N < L_t\}} = (L_t - N\alpha_{n-1})^+.$$

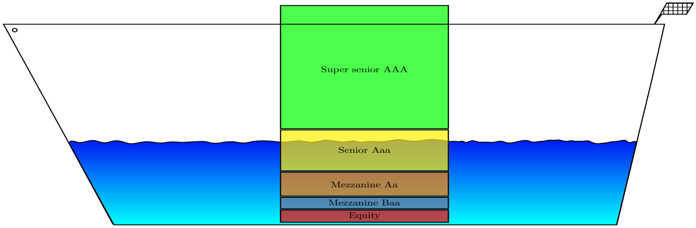


Fig. 11.3: A Titanic-style representation of cumulative tranche losses.

The CDO tranche n°l, $l = 1, 2, \ldots, n$, can be decomposed into:

- A premium leg: the short investor in tranche n°l is purchasing protection at time t against default at time T_k, $k = i+1, \ldots, j$, by making fixed payments S_t^l at times $T_{i+1}, \ldots, T_j$ between t and T in compensation. This premium can also be received by the unfunded investor.

  The discounted value at time t of the premium leg for the tranche n°l is

$$\begin{aligned} V_l^p(t,T) &= \mathbb{E}\left[\sum_{k=i}^{j-1} S_t^l \delta_k (Np_l - L_{T_{k+1}}^l) \exp\left(- \int_t^{T_{k+1}} r_s ds \right) \ \Big|\ \mathcal{G}_t \right] \\ &= S_t^l \sum_{k=i}^{j-1} \delta_k \mathbb{E}\left[(Np_l - L_{T_{k+1}}^l) \exp\left(- \int_t^{T_{k+1}} r_s ds \right) \ \Big|\ \mathcal{G}_t \right] \end{aligned} \quad (11.6)$$

  $l = 1, 2, \ldots, N$, where the premium spread S_t^l is quoted as a proportion of the compensation $Np_l - L_{T_{k+1}}^l$ and is paid at each time T_{k+1} until $k = j-1$ or $L_{T_{k+1}} = 100\%$, whichever comes first.

- A protection leg: the short investor receives protection against default from the premium leg, which can also be paid by the unfunded investors. Noting that at each default time $\tau_k \in (T_k, T_{k+1}]$, $k = i, \ldots, j-1$, the loss L_t^l taken by tranche n°l jumps by the amount $\Delta L_{\tau_k}^l = L_{\tau_k}^l - L_{\tau_k^-}^l$, the value at time t of the protection leg for tranche n°l can be written as

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$\begin{aligned}
V_l^d(t,T) &= \mathbb{E}\left[\sum_{k=i}^{j-1} \mathbb{1}_{[T_i,T_j]}(\tau_k)\Delta L^l_{\tau_k} \exp\left(-\int_t^{\tau_k} r_u du\right) \ \Big| \ \mathcal{G}_t\right] \\
&= \mathbb{E}\left[\int_{T_i}^{T_j} \exp\left(-\int_t^s r_u du\right) dL_s^l \,\Big|\, \mathcal{G}_t\right] \qquad (11.7) \\
&= \mathbb{E}\left[\exp\left(-\int_t^{T_j} r_u du\right) L^l_{T_j} - \exp\left(-\int_t^{T_i} r_u du\right) L^l_{T_i} \,\Big|\, \mathcal{G}_t\right] \\
&\quad + \mathbb{E}\left[\int_{T_i}^{T_j} r_s \exp\left(-\int_t^s r_u du\right) L_s^l ds \,\Big|\, \mathcal{G}_t\right] \\
&= \mathbb{E}\left[\exp\left(-\int_t^{T_j} r_u du\right) L^l_{T_j} \,\Big|\, \mathcal{G}_t\right] + \mathbb{E}\left[\int_{T_i}^{T_j} r_s \exp\left(-\int_t^s r_u du\right) L_s^l ds \,\Big|\, \mathcal{G}_t\right],
\end{aligned}$$

where we applied integration by parts on $[T_i, T_j]$ and used the fact that $L^l_{T_i} = 0$, $l = 1, 2, \ldots, n$.

The spread S_t^l paid by tranche n°l is computed by equating the values $V_l^p(t,T) = V_l^d(t,T)$ of the protection and premium legs in (11.6) and (11.7), which yields

$$\begin{aligned}
S_t^l &= \frac{\mathbb{E}\left[\int_{T_i}^{T_j} \exp\left(-\int_t^s r_u du\right) dL_s^l \,\Big|\, \mathcal{G}_t\right]}{\sum_{k=i}^{j-1} \delta_k \mathbb{E}\left[(Np_l - L^l_{T_{k+1}}) \exp\left(-\int_t^{T_{k+1}} r_s ds\right) \ \Big| \ \mathcal{G}_t\right]} \\
&= \frac{\mathbb{E}\left[\exp\left(-\int_t^{T_j} r_u du\right) L^l_{T_j} \,\Big|\, \mathcal{G}_t\right] + \mathbb{E}\left[\int_{T_i}^{T_j} r_s \exp\left(-\int_t^s r_u du\right) L_s^l ds \,\Big|\, \mathcal{G}_t\right]}{\sum_{k=i}^{j-1} \delta_k \mathbb{E}\left[(Np_l - L^l_{T_{k+1}}) \exp\left(-\int_t^{T_{k+1}} r_s ds\right) \ \Big| \ \mathcal{G}_t\right]} \\
&\geqslant 0,
\end{aligned}$$

$l = 1, 2, \ldots, n$.

Cumulative expected tranche loss

The expected cumulative loss given the value of the parameter M can be computed by linearity in the multiple default time model (9.7) of Chapter 9 as

$$\begin{aligned}
\mathbb{E}[L_t \mid M = m] &= \sum_{k=i}^{j-1} \mathbb{E}\left[(1-\xi_{k+1})\mathbb{1}_{\{\tau_k \leqslant t\}} \mid M = m\right] \\
&= \sum_{k=i}^{j-1} (1-\xi_{k+1}) \mathbb{P}\left(\tau_k \leqslant t \mid M = m\right)
\end{aligned}$$

$$= \sum_{k=i}^{j-1} (1 - \xi_{k+1}) \Phi \left(\frac{\Phi^{-1}(\mathbb{P}(\tau_k \leqslant t)) + a_k m}{\sqrt{1 - a_k^2}} \right),$$

by (9.7), and the expected cumulative loss can be written as

$$\mathbb{E}[L_t] = \int_{-\infty}^{\infty} \mathbb{E}\left[L_t \mid M = m\right] \phi(m) dm = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbb{E}\left[L_t \mid M = m\right] \mathrm{e}^{-m^2/2} dm.$$

The situation is different for the expected loss of tranche n°l is written as the expected value

$$\mathbb{E}[L_t^l] = \mathbb{E}\left[\min((L_t - N\alpha_{l-1})^+, Np_l)\right], \qquad l = 1, 2, \ldots, n,$$

of the *nonlinear* function $f_l(L_t) := \min((L_t - N\alpha_{l-1})^+, Np_l)$ of L_t, where α_{l-1} is defined in (11.5).

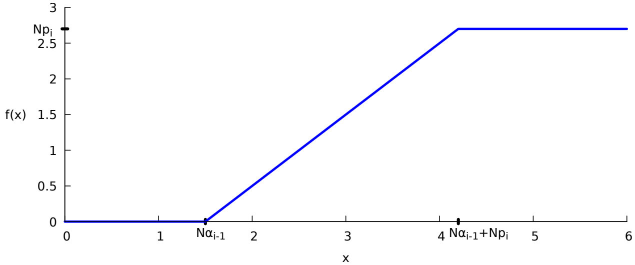


Fig. 11.4: Function $f_l(x) = \min((x - N\alpha_{l-1})^+, Np_l)$.

The expected tranche loss $\mathbb{E}[L_t^l]$ n°l can be estimated by the Monte Carlo method when the default times are generated according to (9.10).

In order to compute expected tranche losses we can model the cumulative loss L_t as a discrete random variable, with for example

$$\mathbb{P}\left(L_t = N - \sum_{k=i}^{j-1} \xi_{k+1} \right) = \mathbb{P}(\tau_i \leqslant t, \ldots, \tau_{j-1} \leqslant t),$$

and

$$\mathbb{P}(L_t = 0) = \mathbb{P}(\tau_i > t, \ldots, \tau_{j-1} > t),$$

which require the knowledge of the joint distribution of the default times $\tau_i, \ldots, \tau_{j-1}$.

If the $\tau_k's$ are independent and identically distributed with common cumulative distribution function F_τ and $a_k = a$, $\xi_k = \xi$, $k = i+1, \ldots, j$, then the cumulative loss L_t has a binomial distribution given M, given by

$$\mathbb{P}(L_t = (1-\xi)k \mid M) = \binom{N}{k}(1 - \mathbb{P}(\tau \leqslant T \mid M))^{N-k}(\mathbb{P}(\tau \leqslant T \mid M))^k$$
$$= \binom{N}{k}\left(1 - \Phi\left(\frac{\Phi^{-1}(F_\tau(T)) - aM}{\sqrt{1-a^2}}\right)\right)^{N-k}\left(\Phi\left(\frac{\Phi^{-1}(F_\tau(T)) - aM}{\sqrt{1-a^2}}\right)\right)^k,$$

$k = 0, 1, \ldots, N$. The expected loss of tranche n°l can then be expressed as

$$\begin{aligned}\mathbb{E}[L_t^l] &= \int_{-\infty}^{\infty} \mathbb{E}\left[f_l(L_t) \mid M = m\right]\phi(m)dm \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathbb{E}\left[f_l(L_t) \mid M = m\right]\mathrm{e}^{-m^2/2}dm,\end{aligned}$$

$l = 1, 2, \ldots, n$, where $\mathbb{E}\left[f_l(L_t) \mid M = m\right]$ is computed either by the Monte Carlo method, from the distribution of L_t.

In Vašiček (2002), the tranche loss has been approximated by a Gaussian random variable for very large portfolios with $N \to \infty$.

The α-percentile loss of the portfolio can be estimated as

$$\mathbb{E}[L_t \mid M = m] = \sum_{k=i}^{j-1}(1-\xi_{k+1})\Phi\left(\frac{\Phi^{-1}(\mathbb{P}(\tau_k \leqslant T)) - a_k\Phi^{-1}(\alpha)}{\sqrt{1-a_k^2}}\right),$$

where $m = \Phi^{-1}(\alpha)$.

Such (Gaussian) Merton (1974) and Vašiček (2002) type models have been implemented in the Basel II recommendations on Banking Supervision (2005). Namely in Basel II, banks are expected to hold capital in prevision of unexpected losses in a worst case scenario, according to the Internal Ratings-Based (IRB) formula

$$\sum_{k=i}^{j-1}(1-\xi_{k+1})\left(\Phi\left(\frac{\Phi^{-1}(\mathbb{P}(\tau_k \leqslant T)) - a_k\Phi^{-1}(\alpha)}{\sqrt{1-a_k^2}}\right) - \mathbb{P}(\tau_k \leqslant T)\right),$$

with confidence level set at $\alpha = 0.999$ *i.e.* $m = \Phi^{-1}(0.999) = 3.09$, cf. Relation (2.4) page 10 of Aas (2005). Recall that the function

$$x \longmapsto \Phi\left(\frac{\Phi^{-1}(x) + a_k \Phi^{-1}(\alpha)}{\sqrt{1-a_k^2}}\right)$$

always lies above the graph of x when $a_k < 0$, as in the next figure.

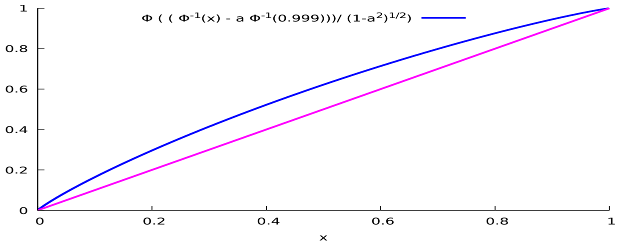


Fig. 11.5: Internal Ratings-Based (IRB) formula.

11.3 Credit Valuation Adjustment (CVA)

Credit Valuation Adjustments (CVA) aim at estimating the amount of capital required in the event of counterparty default, and are specially relevant to the Basel III regulatory framework. Other credit value adjustments (XVA) include the Funding Valuation Adjustments (FVA), Debit Valuation Adjustments (DVA), Capital Valuation Adjustments (KVA), and Margin Valuation Adjustments (MVA). The purpose of XVAs is also to take into account the future value of trades and their associated risks. The real-time estimation of XVA measures is generally highly demanding from a computational point of view.

Net Present Value (NPV) of a CDS

As above, we work with a tenor structure $\{t = T_i < \cdots < T_j = T\}$, and let

$$\begin{aligned}\Pi(T_l, T_j) &:= \text{protection\_leg} - \text{premium\_leg} \\ &= \sum_{k=l}^{j-1} \mathbb{1}_{(T_k, T_{k+1}]}(\tau)(1-\xi_{k+1}) \exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right) \\ &\quad - \sum_{k=l}^{j-1} S_{T_l}^{i,j} \delta_k \mathbb{1}_{\{\tau > T_{k+1}\}} \exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right)\end{aligned}$$

$$
\begin{aligned}
&= (1-\xi)\sum_{k=l}^{j-1} \mathbb{1}_{(T_k,T_{k+1}]}(\tau)\exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right) \\
&\quad -\sum_{k=l}^{j-1} S_{T_l}^{i,j}\delta_k \mathbb{1}_{\{\tau>T_{k+1}\}}\exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right) \\
&= \sum_{k=l}^{j-1}\left((1-\xi)\mathbb{1}_{(T_k,T_{k+1}]}(\tau)\exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right)\right. \\
&\quad \left. -\delta_k \mathbb{1}_{\{\tau>T_{k+1}\}}\exp\left(-\int_{T_l}^{T_{k+1}} r_s ds\right)\right) \qquad (11.8)
\end{aligned}
$$

denote the difference between the remaining protection and premium legs from time T_l until time T_j. Note that by definition of the spread $S_t^{i,j}$ we have $\Pi(t, T_j) = 0$, $0 \leqslant t \leqslant T_i$.

Definition 11.4. *The Net Present Value (NPV) at time T_l of the CDS is the conditional expected value*

$$
\mathrm{NPV}(T_l, T_j) := \mathbb{E}\big[\Pi(T_l, T_j) \mid \mathcal{G}_{T_l}\big]
$$

of the difference between the values at time T_l of the remaining protection and premium legs from time T_l until time T_j, where $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ is the filtration (10.4) enlarged as with the additional information on the default time τ.

The Net Present Value (NPV) at time T_l of the CDS satisfies

$$
\begin{aligned}
&\mathrm{NPV}(T_l, T_j) := \mathbb{E}\big[\Pi(T_l, T_j) \mid \mathcal{G}_{T_l}\big] \\
&= \mathbb{E}\left[\sum_{k=l}^{j-1} \mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \,\middle|\, \mathcal{G}_{T_l}\right] \qquad (11.9) \\
&\quad -\mathbb{E}\left[\sum_{k=l}^{j-1} S_t^{i,j}\delta_k \mathbb{1}_{\{\tau>T_{k+1}\}}\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \,\middle|\, \mathcal{G}_{T_l}\right] \\
&= (1-\xi)\sum_{k=l}^{j-1}\mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \,\middle|\, \mathcal{G}_t\right] - S_t^{i,j}\sum_{k=l}^{j-1}\delta_k P(t, T_{k+1}) \\
&= \sum_{k=l}^{j-1}\left((1-\xi)\mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)\exp\left(-\int_t^{T_{k+1}} r_s ds\right) \,\middle|\, \mathcal{G}_t\right] - S_t^{i,j}\delta_k P(t, T_{k+1})\right)
\end{aligned}
$$

which is the difference between the values at time T_l of the remaining protection and premium legs from time T_l until time T_j.

In addition to the credit default time τ we introduce a second $(\mathcal{G}_t)_{t\geqslant 0}$-stopping time $\nu \in [T_l, T_j]$ representing the possible default time of the party providing the protection leg. The Net Present Value $\mathrm{NPV}(\nu, T_j)$ is estimated when default occurs at time ν.

i) If $\mathrm{NPV}(\nu, T_j) > 0$ then a payment is due from the party providing the protection leg, and only a fraction $\eta\mathrm{NPV}(\nu, T_j)$ of this payment may be recovered, where $\eta \in [0, 1]$ is the recovery rate of the party providing protection in the CDS.

ii) On the other hand, if $\mathrm{NPV}(\nu, T_j) < 0$ then the original fee payment $-\mathrm{NPV}(\nu, T_j)$ is still due.

As a consequence, in the event of default at time $\nu \in [T_l, T_j]$, the net present value of the CDS at time ν is

$$\begin{aligned}
&\eta\mathrm{NPV}(\nu, T_j)\mathbb{1}_{\{\mathrm{NPV}(\nu,T_j)>0\}} + \mathrm{NPV}(\nu, T_j)\mathbb{1}_{\{\mathrm{NPV}(\nu,T_j)<0\}} \\
&\qquad = \eta\big(\mathrm{NPV}(\nu, T_j)\big)^+ - \big(\mathrm{NPV}(\nu, T_j)\big)^- \\
&\qquad = \eta\big(\mathrm{NPV}(\nu, T_j)\big)^+ - \big(-\mathrm{NPV}(\nu, T_j)\big)^+ \\
&\qquad = \eta\big(\mathrm{NPV}(\nu, T_j)\big)^+ + \big(\mathrm{NPV}(\nu, T_j) - \big(\mathrm{NPV}(\nu, T_j)\big)^+\big) \\
&\qquad = \mathrm{NPV}(\nu, T_j) - (1-\eta)\big(\mathrm{NPV}(\nu, T_j)\big)^+. \qquad (11.10)
\end{aligned}$$

Credit Valuation Adjustment (CVA)

Under the event of counterparty default at a time $\nu \in [T_l, T_j]$, we estimate the corresponding discounted payment estimated at time T_l as

$$\begin{aligned}
&\Pi^D(T_l, T_j) = \mathbb{1}_{\{T_j<\nu\}}\Pi(T_l, T_j) \\
&+ \mathbb{1}_{\{T_l<\nu\leqslant T_j\}}\left(\Pi(T_l,\nu) + \exp\left(-\int_{T_l}^{\nu} r_s ds\right)\left(\eta\big(\mathrm{NPV}(\nu,T_j)\big)^+ - \big(-\mathrm{NPV}(\nu,T_j)\big)^+\right)\right) \\
&= \Pi(T_l,\nu) + \exp\left(-\int_{T_l}^{\nu} r_s ds\right)\left(\eta\big(\mathrm{NPV}(\nu,T_j)\big)^+ - \big(-\mathrm{NPV}(\nu,T_j)\big)^+\right) \\
&= \Pi(T_l,\nu) + \exp\left(-\int_{T_l}^{\nu} r_s ds\right)\left(\mathrm{NPV}(\nu,T_j) - (1-\eta)\big(\mathrm{NPV}(\nu,T_j)\big)^+\right),
\end{aligned}$$

see Brigo and Masetti (2006), Brigo and Chourdakis (2009). As a consequence, we derive the following result.

Proposition 11.5. *The price at time T_l of the payoff $\Pi^D(T_l, T_j)$ under counterparty risk is given by*

$$\mathbb{E}\big[\Pi^D(T_l, T_j) \mid \mathcal{G}_{T_l}\big] = \mathrm{NPV}(T_l, T_j)$$

$$-(1-\eta)\mathbb{E}\left[\mathbb{1}_{\{T_l<\nu\leqslant T_j\}}\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\left(\mathrm{NPV}(\nu,T_j)\right)^+ \;\Big|\; \mathcal{G}_{T_l}\right].$$

Proof. From (11.8) and Definition 11.4, we note the relation

$$\begin{aligned}
\mathrm{NPV}(T_l,T_j) &= \mathbb{E}\big[\Pi(T_l,T_j) \mid \mathcal{G}_{T_l}\big] \\
&= \mathbb{E}\left[\Pi(T_l,\nu)+\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\Pi(\nu,T_j)\,\Big|\,\mathcal{G}_{T_l}\right] \\
&= \mathbb{E}\left[\Pi(T_l,\nu)+\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\mathrm{NPV}(\nu,T_j)\,\Big|\,\mathcal{G}_{T_l}\right].
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\mathbb{E}\big[\Pi^D(T_l,T_j) \mid \mathcal{G}_{T_l}\big] = \mathbb{E}\big[\Pi(T_l,T_j) \mid \mathcal{G}_{T_l}\big] \\
&= \mathbb{E}\left[\Pi(T_l,\nu)+\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\left(\mathrm{NPV}(\nu,T_j)-(1-\eta)\big(\mathrm{NPV}(\nu,T_j)\big)^+\right)\,\Big|\,\mathcal{G}_{T_l}\right] \\
&= \mathbb{E}\left[\Pi(T_l,T_j)-(1-\eta)\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\left(\mathrm{NPV}(\nu,T_j)\right)^+\,\Big|\,\mathcal{G}_{T_l}\right] \\
&= \mathrm{NPV}(T_l,T_j)-\mathbb{E}\left[(1-\eta)\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\left(\mathrm{NPV}(\nu,T_j)\right)^+\,\Big|\,\mathcal{G}_{T_l}\right].
\end{aligned}$$

□

The quantity

$$(1-\eta)\mathbb{E}\left[\mathbb{1}_{\{T_l<\nu\leqslant T_j\}}\exp\left(-\int_{T_l}^{\nu}r_s ds\right)\left(\mathrm{NPV}(\nu,T_j)\right)^+ \;\Big|\; \mathcal{G}_{T_l}\right]$$

is called the (positive) Counterparty Risk (CR) Credit Valuation Adjustment (CVA).

Exercises

Exercise 11.1 Show that the equation (11.4) admits a numerical solution $\lambda > 0$.

Exercise 11.2 Credit default swaps. From the CDS market data of Figure 11.6 on McDonald's Corp, estimate the first default rate λ_1 and the associated default probability in the framework of (11.4), cf. also Castellacci (2008).

Fig. 11.6: Cashflow data.

Exercise 11.3 Consider a tenor structure $\{t = T_i < \cdots < T_j = T\}$, a sequence

$$P(t,T_k) = \exp\left(-\int_t^{T_k} r(s)ds\right) = \mathrm{e}^{-(T_k-t)r_k}, \qquad k = i,\ldots,j,$$

of *deterministic* discount factors, and a family

$$Q(t,T_k) = \mathbb{E}\left[\exp\left(-\int_t^{T_k}\lambda_s ds\right) \;\Big|\; \mathcal{F}_t\right]$$

of survival probabilities.

a) Show that the discounted value at time t of the protection leg equals

$$\sum_{k=i}^{j-1}\mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r(s)ds\right) \;\Big|\; \mathcal{G}_t\right]$$
$$= \mathbb{1}_{\{\tau>t\}}(1-\xi)\sum_{k=i}^{j-1} P(t,T_{k+1})\left(Q(t,T_k)-Q(t,T_{k+1})\right).$$

b) Letting $\delta_k := T_{k+1}-T_k$, $k = i,\ldots,j-1$, show that the discounted value at time t of the premium leg, equals

$$V^p(t,T) = \mathbb{1}_{\{\tau>t\}} S_t^{i,j} \sum_{k=i}^{j-1} \delta_k P(t,T_{k+1}) Q(t,T_{k+1}).$$

c) By equating the protection and premium legs, find the value of $Q(t,T_{i+1})$ with $Q(t,T_i) = 1$, and derive a recurrence relation between $Q(t,T_{j+1})$ and $Q(t,T_i),\ldots,Q(t,T_j)$.

Exercise 11.4 (Exercise 11.3 continued). From the spread data and survival probabilities data of Figure 11.7 on the Coca-Cola Company, retrieve the corresponding CDS spreads $S_t^{i,j}$ and discount factors $P(t,T_i),\ldots,P(t,T_n)$, and estimate the corresponding survival probabilities $Q(t,T_i),\ldots,Q(t,T_n)$.

Fig. 11.7: CDS Market data.

Part IV
Appendix

Background on Probability Theory

In this appendix, we review a number of basic probabilistic tools that are needed in option pricing and hedging. We also refer to Jacod and Protter (2000), Devore (2003), Pitman (1999) for additional relevant probability background material.

A.1 Probability Sample Space and Events

We will need the following notation coming from set theory. Given A and B to abstract sets, "$A \subset B$" means that A is contained in B, and in this case, $B \setminus A$ denotes the set of elements of B which do not belong to A. The property that the element ω belongs to the set A is denoted by "$\omega \in A$", and given two sets A and Ω such that $A \subset \Omega$, we let $A^c = \Omega \setminus A$ denote the *complement* of A in Ω. The finite set made of n elements $\omega_1, \ldots, \omega_n$ is denoted by $\{\omega_1, \ldots, \omega_n\}$, and we will usually distinguish between the element ω and its associated singleton set $\{\omega\}$.

A probability sample space is an abstract set Ω that contains the possible outcomes of a random experiment.

Examples

i) Coin tossing: $\Omega = \{H, T\}$.

ii) Rolling one die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

iii) Picking one card at random in a pack of 52: $\Omega = \{1, 2, 3, \ldots, 52\}$.

iv) An integer-valued random outcome: $\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}$.

In this case the outcome $\omega \in \mathbb{N}$ can be the random number of trials needed until some event occurs.

v) A nonnegative, real-valued outcome: $\Omega = \mathbb{R}_+$.

In this case the outcome $\omega \in \mathbb{R}_+$ may represent the (nonnegative) value of a continuous random time.

vi) A random continuous parameter (such as time, weather, price or wealth, temperature, ...): $\Omega = \mathbb{R}$.

vii) Random choice of a continuous path in the space $\Omega = \mathcal{C}(\mathbb{R}_+)$ of all continuous functions on $\mathbb{R}_+$.

In this case, $\omega \in \Omega$ is a function $\omega : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and a typical example is the graph $t \longmapsto \omega(t)$ of a stock price over time.

Product spaces:

Probability sample spaces can be built as product spaces and used for the modeling of repeated random experiments.

i) Rolling two dice: $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.

In this case a typical element of Ω is written as $\omega = (k, l)$ with $k, l \in \{1, 2, 3, 4, 5, 6\}$.

ii) A finite number n of real-valued samples: $\Omega = \mathbb{R}^n$.

In this case the outcome ω is a vector $\omega = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with n components.

Note that to some extent, the more complex Ω is, the better it fits a practical and useful situation, *e.g.* $\Omega = \{H, T\}$ corresponds to a simple coin tossing experiment while $\Omega = \mathcal{C}(\mathbb{R}_+)$ the space of continuous functions on $\mathbb{R}_+$ can be applied to the modeling of stock markets. On the other hand, in many cases and especially in the most complex situations, we will *not* attempt to specify Ω explicitly.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Events

An event is a collection of outcomes, which is represented by a subset of Ω. In what follows we consider collections of events, called *σ-algebras* (or *σ-fields*), according to the following definition.

Definition A.6. *A collection $\mathcal{G}$ of events is a σ-*algebra *provided that it satisfies the following conditions:*

(i) $\emptyset \in \mathcal{G}$,
(ii) *For all countable sequences* $(A_n)_{n \geqslant 1}$ *such that* $A_n \in \mathcal{G}$, $n \geqslant 1$, *we have* $\bigcup_{n \geqslant 1} A_n \in \mathcal{G}$,
(iii) $A \in \mathcal{G} \Longrightarrow (\Omega \setminus A) \in \mathcal{G}$,

where $\Omega \setminus A := \{\omega \in \Omega \ : \ \omega \notin A\}$.

Note that Properties (ii) and (iii) above also imply the stability of σ-algebras under intersections, as

$$\bigcap_{n \geqslant 1} A_n = \left(\bigcup_{n \geqslant 1} A_n^c \right)^c \in \mathcal{G}, \tag{A.1}$$

for all countable sequences $A_n \in \mathcal{G}$, $n \geqslant 1$.

The collection of all events in Ω will often be denoted by $\mathcal{F}$. The empty set $\emptyset$ and the full space Ω are considered as events but they are of less importance because Ω corresponds to "any outcome may occur" while $\emptyset$ corresponds to an absence of outcome, or no experiment.

In the context of stochastic processes, two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ such that $\mathcal{F} \subset \mathcal{G}$ will refer to two different amounts of information, the amount of information associated to $\mathcal{F}$ being here lower than the one associated to $\mathcal{G}$.

The formalism of σ-algebras helps in describing events in a short and precise way.

Examples

i) Let $\Omega = \{1, 2, 3, 4, 5, 6\}$.

The event $A = \{2, 4, 6\}$ corresponds to

"the result of the experiment is an even number".

ii) Taking again $\Omega = \{1, 2, 3, 4, 5, 6\}$,

$$\mathcal{F} := \{\Omega, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$$

defines a σ-algebra on Ω which corresponds to the knowledge of parity of an integer picked at random from 1 to 6.

Note that in the set-theoretic notation, an event A is a subset of Ω, *i.e.* $A \subset \Omega$, while it is an element of $\mathcal{F}$, *i.e.* $A \in \mathcal{F}$. For example, we have $\Omega \supset \{2,4,6\} \in \mathcal{F}$, while $\{\{2,4,6\},\{1,3,5\}\} \subset \mathcal{F}$.

iii) Taking

$$\mathcal{G} := \{\Omega, \emptyset, \{2,4,6\}, \{2,4\}, \{6\}, \{1,2,3,4,5\}, \{1,3,5,6\}, \{1,3,5\}\} \supset \mathcal{F},$$

defines a σ-algebra on Ω which is bigger than $\mathcal{F}$ and includes the parity information contained in $\mathcal{F}$, in addition to information on whether the outcome of the experiment is equal to 6 or not.

iv) Take

$$\Omega = \{H,T\} \times \{H,T\} = \{(H,H),(H,T),(T,H),(T,T)\}.$$

In this case, the collection $\mathcal{F}$ of all possible events is given by

$$\begin{aligned}\mathcal{F} = \{\emptyset, &\{(H,H)\}, \{(T,T)\}, \{(H,T)\}, \{(T,H)\}, \\ &\{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \{(H,T),(T,T)\}, \\ &\{(T,H),(T,T)\}, \{(H,T),(H,H)\}, \{(T,H),(H,H)\}, \\ &\{(H,H),(T,T),(T,H)\}, \{(H,H),(T,T),(H,T)\}, \\ &\{(H,T),(T,H),(H,H)\}, \{(H,T),(T,H),(T,T)\}, \Omega\}.\end{aligned} \tag{A.2}$$

Note that the set $\mathcal{F}$ of all events considered in (A.2) above has altogether

$1 = \binom{n}{0}$ event of cardinality 0,

$4 = \binom{n}{1}$ events of cardinality 1,

$6 = \binom{n}{2}$ events of cardinality 2,

$4 = \binom{n}{3}$ events of cardinality 3,

$1 = \binom{n}{4}$ event of cardinality 4,

with $n = 4$, for a total of

$$16 = 2^n = \sum_{k=0}^{4} \binom{4}{k} = 1 + 4 + 6 + 4 + 1$$

events. The collection of events

$$\mathcal{G} := \{\emptyset, \{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \Omega\}$$

defines a sub σ-algebra of $\mathcal{F}$, which corresponds to the restricted information "the results of two coin tossings are different".

Exercise: Write down the set of all events on $\Omega = \{H, T\}$.

Note also that (H,T) is different from (T,H), whereas $\{(H,T),(T,H)\}$ is equal to $\{(T,H),(H,T)\}$.

In addition, we will distinguish between the *outcome* $\omega \in \Omega$ and its associated *event* $\{\omega\} \in \mathcal{F}$, which satisfies $\{\omega\} \subset \Omega$.

A.2 Probability Measures

Definition A.7. *A* probability measure *is a mapping* $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ *that assigns a probability* $\mathbb{P}(A) \in [0,1]$ *to any event* $A \in \mathcal{F}$*, with the properties*

a) $\mathbb{P}(\Omega) = 1$*, and*

b) $\mathbb{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \sum_{n\geqslant 1} \mathbb{P}(A_n)$*, whenever* $A_k \cap A_l = \emptyset$*,* $k \neq l$*.*

Property (b) above is named the *law of total probability*. It states in particular that we have

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$$

when the subsets $A_1, \ldots, A_n$ of Ω are disjoint, and

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{A.3}$$

if $A \cap B = \emptyset$. We also have the *complement rule*

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega \setminus A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A).$$

When A and B are not necessarily disjoint we can write

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$

which extends to arbitrary families of events $(A_i)_{i\in I}$ indexed by a finite set I as the inclusion-exclusion principle

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{J\subset I}(-1)^{|J|+1}\mathbb{P}\left(\bigcap_{j\in J} A_j\right), \tag{A.4}$$

and

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \sum_{I\subset J}(-1)^{|I|+1}\mathbb{P}\left(\bigcup_{i\in I} A_i\right). \tag{A.5}$$

The triple

$$(\Omega, \mathcal{F}, \mathbb{P}) \tag{A.6}$$

is called a *probability space*, and was introduced by A.N. Kolmogorov (1903-1987). This setting is generally referred to as the *Kolmogorov framework*.

A property or event is said to hold $\mathbb{P}$-*almost surely* (also written $\mathbb{P}$-*a.s.*) if it holds with probability equal to one.

Example

Take

$$\Omega = \big\{(T,T),(H,H),(H,T),(T,H)\big\}$$

and

$$\mathcal{F} = \{\emptyset, \{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \Omega\}.$$

The *uniform* probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is given by setting

$$\mathbb{P}(\{(T,T),(H,H)\}) := \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\{(H,T),(T,H)\}) := \frac{1}{2}.$$

In addition, we have the following convergence properties.

1. Let $(A_n)_{n\in\mathbb{N}}$ be a *non-decreasing* sequence of events, *i.e.* $A_n \subset A_{n+1}$, $n \geqslant 0$. Then we have

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n). \tag{A.7}$$

2. Let $(A_n)_{n\in\mathbb{N}}$ be a *non-increasing* sequence of events, *i.e.* $A_{n+1} \subset A_n$, $n \geqslant 0$. Then we have

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n). \tag{A.8}$$

Theorem A.8. *Borel-Cantelli Lemma. Let* $(A_n)_{n\geqslant 1}$ *denote a sequence of events on* $(\Omega, \mathcal{F}, \mathbb{P})$*, such that*

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$\sum_{n\geqslant 1} \mathbb{P}(A_n) < \infty.$$

Then we have

$$\mathbb{P}\left(\bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k\right) = 0,$$

i.e. the probability that A_n occurs infinitely many times occur is zero.

A.3 Conditional Probabilities and Independence

We start with examples.

Consider a population $\Omega = M \cup W$ made of a set M of men and a set W of women. Here the σ-algebra $\mathcal{F} = \{\Omega, \emptyset, W, M\}$ corresponds to the information given by gender. After polling the population, *e.g.* for a market survey, it turns out that a proportion $p \in [0,1]$ of the population declares to like apples, while a proportion $1-p$ declares to dislike apples. Let $A \subset \Omega$ denote the subset of individuals who like apples, while $A^c \subset \Omega$ denotes the subset individuals who dislike apples, with

$$p = \mathbb{P}(A) \quad \text{and} \quad 1-p = \mathbb{P}(A^c),$$

e.g. $p = 60\%$ of the population likes apples. It may be interesting to get a more precise information and to determine

- the relative proportion $\dfrac{\mathbb{P}(A \cap W)}{\mathbb{P}(W)}$ of women who like apples, and
- the relative proportion $\dfrac{\mathbb{P}(A \cap M)}{\mathbb{P}(M)}$ of men who like apples.

Here, $\mathbb{P}(A \cap W)/\mathbb{P}(W)$ represents the probability that a randomly chosen woman in W likes apples, and $\mathbb{P}(A \cap M)/\mathbb{P}(M)$ represents the probability that a randomly chosen man in M likes apples. Those two ratios are interpreted as *conditional probabilities*, for example $\mathbb{P}(A \cap M)/\mathbb{P}(M)$ denotes the probability that a given individual likes apples *given that* he is a man.

For another example, suppose that the population Ω is split as $\Omega = Y \cup O$ into a set Y of "young" people and another set O of "old" people, and denote by $A \subset \Omega$ the set of people who voted for candidate A in an election. Here it can be of interest to find out the relative proportion

$$\mathbb{P}(A \mid Y) = \frac{\mathbb{P}(Y \cap A)}{\mathbb{P}(Y)}$$

of young people who voted for candidate A.

Definition A.9. *Given any two events $A, B \subset \Omega$ with $\mathbb{P}(B) \neq 0$, we call*

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

the probability of A given B*, or* conditionally to B*.*

Remark A.10. *We note that if $\mathbb{P}(B) = 1$ we have $\mathbb{P}(A \cap B^c) \leqslant \mathbb{P}(B^c) = 0$, hence $\mathbb{P}(A \cap B^c) = 0$, which implies*

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \cap B),$$

and $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

We also recall the following property:

$$\begin{aligned}
\mathbb{P}\left(B \cap \bigcup_{n \geqslant 1} A_n \right) &= \mathbb{P}\left(\bigcup_{n \geqslant 1} (B \cap A_n) \right) \\
&= \sum_{n \geqslant 1} \mathbb{P}(B \cap A_n) \\
&= \sum_{n \geqslant 1} \mathbb{P}(B \mid A_n) \mathbb{P}(A_n) \\
&= \sum_{n \geqslant 1} \mathbb{P}(A_n \mid B) \mathbb{P}(B),
\end{aligned}$$

for any family of disjoint events $(A_n)_{n \geqslant 1}$ with $A_i \cap A_j = \emptyset$, $i \neq j$, and $\mathbb{P}(B) > 0$, $n \geqslant 1$. This also shows that conditional probability measures are probability measures, in the sense that whenever $\mathbb{P}(B) > 0$, we have

a) $\mathbb{P}(\Omega \mid B) = 1$, and

b) $\displaystyle \mathbb{P}\left(\bigcup_{n \geqslant 1} A_n \,\middle|\, B \right) = \sum_{n \geqslant 1} \mathbb{P}(A_n \mid B)$, whenever $A_k \cap A_l = \emptyset$, $k \neq l$.

In particular, if $\displaystyle \bigcup_{n \geqslant 1} A_n = \Omega$, $(A_n)_{n \geqslant 1}$ becomes a *partition* of Ω and we get the *law of total probability*

$$\mathbb{P}(B) = \sum_{n \geqslant 1} \mathbb{P}(B \cap A_n) = \sum_{n \geqslant 1} \mathbb{P}(A_n \mid B) \mathbb{P}(B) = \sum_{n \geqslant 1} \mathbb{P}(B \mid A_n) \mathbb{P}(A_n), \tag{A.9}$$

provided that $A_i \cap A_j = \emptyset$, $i \neq j$, and $\mathbb{P}(B) > 0$, $n \geqslant 1$.

Remark. In general we have

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$\mathbb{P}\left(A \,\middle|\, \bigcup_{n\geqslant 1} B_n\right) \neq \sum_{n\geqslant 1} \mathbb{P}(A \mid B_n),$$

even when $B_k \cap B_l = \emptyset$, $k \neq l$. Indeed, taking for example $A = \Omega = B_1 \cup B_2$ with $B_1 \cap B_2 = \emptyset$ and $\mathbb{P}(B_1) = \mathbb{P}(B_2) = 1/2$, we have

$$1 = \mathbb{P}(\Omega \mid B_1 \cup B_2) \neq \mathbb{P}(\Omega \mid B_1) + \mathbb{P}(\Omega \mid B_2) = 2.$$

Independent events

Definition A.11. *Two events A and B such that $\mathbb{P}(A), \mathbb{P}(B) > 0$ are said to be* independent *if*

$$\mathbb{P}(A \mid B) = \mathbb{P}(A). \tag{A.10}$$

We note that the independence condition (A.10) is equivalent to

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

A.4 Random Variables

A real-valued random variable is a mapping*

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

from a probability sample space Ω into the state space $\mathbb{R}$. Given

$$X : \Omega \longrightarrow \mathbb{R}$$

a random variable and a (measurable)† subset A of $\mathbb{R}$, we denote by $\{X \in A\}$ the event

$$\{X \in A\} := \{\omega \in \Omega \ : \ X(\omega) \in A\}.$$

Examples

i) Let $\Omega := \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, and consider the mapping

$$X : \Omega \longrightarrow \mathbb{R}$$

* See (MOE and UCLES 2022, page 14) lines 4-5 and (MOE and UCLES 2020, page 19) lines 4-5.

† Measurability of subsets of $\mathbb{R}$ refers to *Borel measurability*, a concept which will not be defined in this text.

$$(k, l) \longmapsto k + l.$$

Then X is a random variable giving the sum of the two numbers appearing on each die.

ii) the time needed everyday to travel from home to work or school is a random variable, as the precise value of this time may change from day to day under unexpected circumstances.

iii) the price of a risky asset can be modeled using a random variable.

In what follows, we will often use the notion of *indicator function* $\mathbb{1}_A$ of an event $A \subset \Omega$.

Definition A.12. *For any $A \subset \Omega$, the indicator function $\mathbb{1}_A$ is the random variable*

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ \omega &\longmapsto \mathbb{1}_A(\omega) \end{aligned}$$

defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Indicator functions satisfy the property

$$\mathbb{1}_{A \cap B}(\omega) = \mathbb{1}_A(\omega)\mathbb{1}_B(\omega), \tag{A.11}$$

since

$$\begin{aligned} \mathbb{1}_{A \cap B}(\omega) = 1 &\iff \omega \in A \cap B \\ &\iff \omega \in A \text{ and } \omega \in B \\ &\iff \mathbb{1}_A(\omega) = 1 \text{ and } \mathbb{1}_B(\omega) = 1 \\ &\iff \mathbb{1}_A(\omega)\mathbb{1}_B(\omega) = 1. \end{aligned}$$

We also have

$$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B,$$

and

$$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B, \tag{A.12}$$

if $A \cap B = \emptyset$.

For example, if $\Omega = \mathbb{N}$ and $A = \{k\}$, for all $l \geqslant 0$ we have

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

$$\mathbb{1}_{\{k\}}(l) = \begin{cases} 1 & \text{if } k = l, \\ \\ 0 & \text{if } k \neq l. \end{cases}$$

Given X a random variable, we also let

$$\mathbb{1}_{\{X=n\}} = \begin{cases} 1 & \text{if } X = n, \\ \\ 0 & \text{if } X \neq n, \end{cases}$$

and

$$\mathbb{1}_{\{X<n\}} = \begin{cases} 1 & \text{if } X < n, \\ \\ 0 & \text{if } X \geqslant n. \end{cases}$$

A.5 Probability Distributions

The *probability distribution* of a random variable $X : \Omega \longrightarrow \mathbb{R}$ is the collection

$$\{\mathbb{P}(X \in A) \ : \ A \text{ is a measurable subset of } \mathbb{R}\}.$$

As the collection of *measurable* subsets of $\mathbb{R}$ coincides with the σ-algebra generated by the intervals in $\mathbb{R}$, the distribution of X can be reduced to the knowledge of the probabilities

$$\{\mathbb{P}(a < X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a) \ : \ a < b \in \mathbb{R}\},$$

or of the cumulative distribution functions

$$\{\mathbb{P}(X \leqslant a) \ : \ a \in \mathbb{R}\}, \quad \text{or} \quad \{\mathbb{P}(X \geqslant a) \ : \ a \in \mathbb{R}\},$$

see *e.g.* Corollary 3.8 in Çınlar (2011).

Two random variables X and Y are said to be independent under the probability $\mathbb{P}$ if their probability distributions satisfy

$$\mathbb{P}(X \in A \ , \ Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all (measurable) subsets A and B of $\mathbb{R}$.

Distributions admitting a density

We say that the distribution of X admits a probability *density* distribution function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ if, for all $a \leqslant b$, the probability $\mathbb{P}(a \leqslant X \leqslant b)$ can be written as

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b \varphi_X(x) dx.$$

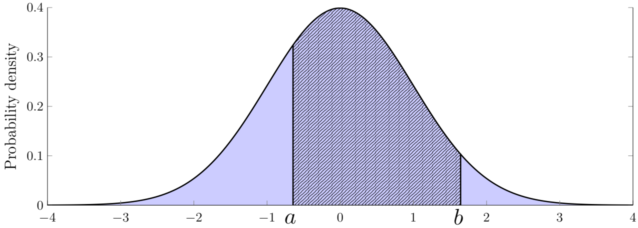


Fig. A.1: Probability density function φ_X.

We also say that the distribution of X is absolutely continuous, or that X is an absolutely continuous random variable. This, however, does *not* imply that the density function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ is continuous.

In particular, we always have

$$\int_{-\infty}^{\infty} \varphi_X(x) dx = \mathbb{P}(-\infty \leqslant X \leqslant \infty) = 1$$

for any probability density functions $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$.

Remark A.13. *Note that if the distribution of X admits a probability density function φ_X, then for all $a \in \mathbb{R}$ we have*

$$\mathbb{P}(X = a) = \int_a^a \varphi_X(x) dx = 0, \tag{A.13}$$

and this is not a contradiction.

In particular, Remark A.13 shows that

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(X = a) + \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a < X < b),$$

for $a \leqslant b$. Property (A.13) appears for example in the framework of lottery games with a large number of participants, in which a given number "a" selected in advance has a very low (almost zero) probability to be chosen.

The probability density function φ_X can be recovered from the Cumulative Distribution Functions (CDFs)

$$x \longmapsto F_X(x) := \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} \varphi_X(s) ds,$$

and

$$x \longmapsto 1 - F_X(x) = \mathbb{P}(X \geqslant x) = \int_x^\infty \varphi_X(s)ds,$$

as

$$\varphi_X(x) = \frac{\partial F_X}{\partial x}(x) = \frac{\partial}{\partial x}\int_{-\infty}^x \varphi_X(s)ds = -\frac{\partial}{\partial x}\int_x^\infty \varphi_X(s)ds, \qquad x \in \mathbb{R}.$$

Examples

i) The *uniform* distribution on an interval.

The probability density function of the uniform distribution on the interval $[a, b]$, $a < b$, is given by

$$\varphi(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x), \qquad x \in \mathbb{R}.$$

ii) The *Gaussian* distribution.

The probability density function of the standard normal distribution is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}, \qquad x \in \mathbb{R}.$$

More generally, the probability density function of the Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ is given by

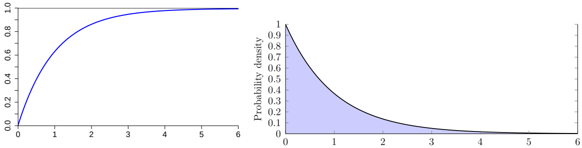
$$\varphi(x) := \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in \mathbb{R}.$$

In this case, we write $X \simeq \mathcal{N}(\mu, \sigma^2)$.

iii) The *exponential* distribution.

The probability density function of the exponential distribution with parameter $\lambda > 0$ is given by

$$\varphi(x) := \lambda \mathbb{1}_{[0,\infty)}(x)\mathrm{e}^{-\lambda x} = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x \geqslant 0 \\ \\ 0, & x < 0. \end{cases} \qquad \text{(A.14)}$$



(a) Exponential CDF.

(b) Exponential PDF.

Fig. A.2: Exponential CDF and PDF.

We also have

$$\mathbb{P}(X > t) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0. \tag{A.15}$$

iv) The *gamma* distribution.

The probability density function of the gamma distribution is given by

$$\varphi(x) := \frac{a^\lambda}{\Gamma(\lambda)} \mathbb{1}_{[0,\infty)}(x) x^{\lambda-1} \mathrm{e}^{-ax} = \begin{cases} \dfrac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \mathrm{e}^{-ax}, & x \geqslant 0 \\ \\ 0, & x < 0, \end{cases}$$

where $a > 0$ and $\lambda > 0$ are scale and shape parameters, and

$$\Gamma(\lambda) := \int_0^\infty x^{\lambda-1} \mathrm{e}^{-x} dx, \qquad \lambda > 0,$$

is the gamma function.

v) The *Cauchy* distribution.

The probability density function of the Cauchy distribution is given by

$$\varphi(x) := \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}.$$

vi) The *lognormal* distribution.

The probability density function of the lognormal distribution is given by

$$\varphi(x) := \mathbb{1}_{[0,\infty)}(x) \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-(\mu-\log x)^2/(2\sigma^2)} = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-(\mu-\log x)^2/(2\sigma^2)}, & x \geqslant 0 \\ \\ 0, & x < 0. \end{cases}$$

Exercise: For each of the above probability density functions φ, check that the condition

$$\int_{-\infty}^{\infty} \varphi(x) dx = 1$$

is satisfied.

Joint densities

Given two absolutely continuous random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$, we can form the $\mathbb{R}^2$-valued random variable (X, Y) defined by

$$\begin{aligned}(X, Y) : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (X(\omega), Y(\omega)).\end{aligned}$$

We say that (X, Y) admits a joint probability density

$$\varphi_{(X,Y)} : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$$

when

$$\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A \text{ and } Y \in B) = \int_B \int_A \varphi_{(X,Y)}(x, y) dx dy$$

for all *measurable* subsets A, B of $\mathbb{R}$, see Figure A.3.

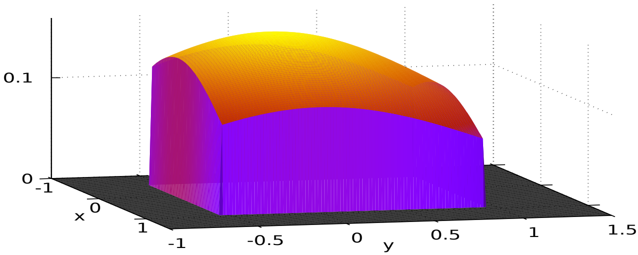


Fig. A.3: Probability $\mathbb{P}((X, Y) \in [-0.5, 1] \times [-0.5, 1])$ computed as a volume integral.

The probability density function $\varphi_{(X,Y)}$ can be recovered from the joint cumulative distribution function

$$(x, y) \longmapsto F_{(X,Y)}(x, y) := \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \varphi_{(X,Y)}(s, t) ds dt,$$

and

$$(x, y) \longmapsto \mathbb{P}(X \geqslant x \text{ and } Y \geqslant y) = \int_x^\infty \int_y^\infty \varphi_{(X,Y)}(s,t)dsdt,$$

as

$$\begin{aligned}
\varphi_{(X,Y)}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y) && \text{(A.16)} \\
&= \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^x \int_{-\infty}^y \varphi_{(X,Y)}(s,t)dsdt && \text{(A.17)} \\
&= \frac{\partial^2}{\partial x \partial y} \int_x^\infty \int_y^\infty \varphi_{(X,Y)}(s,t)dsdt,
\end{aligned}$$

$x, y \in \mathbb{R}$.

The probability densities $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ and $\varphi_Y : \mathbb{R} \longrightarrow \mathbb{R}_+$ of $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ are called the *marginal densities* of (X, Y), and are given by

$$\varphi_X(x) = \int_{-\infty}^\infty \varphi_{(X,Y)}(x,y)dy, \qquad x \in \mathbb{R}, \tag{A.18}$$

and

$$\varphi_Y(y) = \int_{-\infty}^\infty \varphi_{(X,Y)}(x,y)dx, \qquad y \in \mathbb{R}.$$

The conditional probability density $\varphi_{X|Y=y} : \mathbb{R} \longrightarrow \mathbb{R}_+$ of X given $Y = y$ is defined by

$$\varphi_{X|Y=y}(x) := \frac{\varphi_{(X,Y)}(x,y)}{\varphi_Y(y)}, \qquad x, y \in \mathbb{R}, \tag{A.19}$$

provided that $\varphi_Y(y) > 0$. In particular, X and Y are independent if and only if

$$\varphi_{X|Y=y}(x) = \varphi_X(x), \quad i.e., \quad \varphi_{(X,Y)}(x,y) = \varphi_X(x)\varphi_Y(y), \qquad x, y \in \mathbb{R}.$$

Example

If $X_1, \ldots, X_n$ are independent exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_n$ we have

$$\begin{aligned}
\mathbb{P}(\min(X_1, \ldots, X_n) > t) &= \mathbb{P}(X_1 > t, \ldots, X_n > t) \\
&= \mathbb{P}(X_1 > t) \cdots \mathbb{P}(X_n > t) \\
&= \mathrm{e}^{-(\lambda_1 + \cdots + \lambda_n)t}, \qquad t \geqslant 0, && \text{(A.20)}
\end{aligned}$$

hence $\min(X_1, \ldots, X_n)$ is an exponentially distributed random variable with parameter $\lambda_1 + \cdots + \lambda_n$.

From the joint probability density function of (X_1, X_2) given by

$$\varphi_{(X_1,X_2)}(x,y) = \varphi_{X_1}(x)\varphi_{X_2}(y) = \lambda_1\lambda_2 \mathrm{e}^{-\lambda_1 x - \lambda_2 y}, \qquad x, y \geqslant 0,$$

we can write

$$\begin{aligned}\mathbb{P}(X_1 < X_2) &= \mathbb{P}(X_1 \leqslant X_2)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} \varphi_{(X_1,X_2)}(x,y)dxdy\\ &= \lambda_1\lambda_2 \int_0^{\infty}\int_0^{y} \mathrm{e}^{-\lambda_1 x - \lambda_2 y}dxdy\\ &= \frac{\lambda_1}{\lambda_1+\lambda_2},\end{aligned} \tag{A.21}$$

and we note that

$$\mathbb{P}(X_1 = X_2) = \lambda_1\lambda_2 \int_{\{(x,y)\in\mathbb{R}_+^2 \ : \ x=y\}} \mathrm{e}^{-\lambda_1 x - \lambda_2 y}dxdy = 0.$$

Discrete distributions

We only consider integer-valued random variables, *i.e.* the distribution of X is given by the values of $\mathbb{P}(X = k)$, $k \geqslant 0$.

Examples

i) The *Bernoulli* distribution.

   We have

$$\mathbb{P}(X = 1) = p \qquad \text{and} \qquad \mathbb{P}(X = 0) = 1 - p, \tag{A.22}$$

   where $p \in [0, 1]$ is a parameter.

   Note that any Bernoulli random variable $X : \Omega \longrightarrow \{0, 1\}$ can be written as the indicator function

$$X = \mathbb{1}_A$$

   on Ω with $A = \{X = 1\} = \{\omega \in \Omega \ : \ X(\omega) = 1\}$.

ii) The *binomial* distribution.

   We have

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

   where $n \geqslant 1$ and $p \in [0, 1]$ are parameters and $\binom{n}{k} = n!/(k!(n-k)!)$, $0 \leqslant k \leqslant n$.

iii) The *geometric* distribution.

In this case, we have

$$\mathbb{P}(X = k) = (1-p)p^k, \qquad k \geqslant 0, \tag{A.23}$$

where $p \in (0,1)$ is a parameter. For example, if $(X_k)_{k\in\mathbb{N}}$ is a sequence of independent Bernoulli random variables with distribution (A.22), then the random variable,*

$$T_0 := \inf\{k \geqslant 0 \ : \ X_k = 0\}$$

can denote the duration of a game until the time that the wealth X_k of a player reaches 0. The random variable T_0 has the geometric distribution (A.23) with parameter $p \in (0,1)$.

iv) The *negative binomial* (or *Pascal*) distribution.

We have

$$\mathbb{P}(X = k) = \binom{k+r-1}{r-1}(1-p)^r p^k, \qquad k \geqslant 0, \tag{A.24}$$

where $p \in (0,1)$ and $r \geqslant 1$ are parameters. Note that the sum of $r \geqslant 1$ independent geometric random variables with parameter p has a negative binomial distribution with parameter (r,p). In particular, the negative binomial distribution recovers the geometric distribution when $r = 1$.

v) The *Poisson* distribution.

We have

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \qquad k \geqslant 0,$$

where $\lambda > 0$ is a parameter.

The probability that a discrete nonnegative random variable $X : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ is finite is given by

$$\mathbb{P}(X < \infty) = \sum_{k \geqslant 0} \mathbb{P}(X = k), \tag{A.25}$$

and we have

$$1 = \mathbb{P}(X = \infty) + \mathbb{P}(X < \infty) = \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} \mathbb{P}(X = k).$$

* The notation "inf" stands for "infimum", meaning the smallest $n \geqslant 0$ such that $X_n = 0$, if such an n exists.

Remark A.14. *The distribution of a discrete random variable cannot admit a probability density. If this were the case, by Remark A.13 we would have* $\mathbb{P}(X = k) = 0$ *for all* $k \geqslant 0$ *and*

$$1 = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(X \in \mathbb{N}) = \sum_{k \geqslant 0} \mathbb{P}(X = k) = 0,$$

which is a contradiction.

Given two discrete random variables X and Y, the conditional distribution of X given $Y = k$ is given by

$$\mathbb{P}(X = n \mid Y = k) = \frac{\mathbb{P}(X = n \text{ and } Y = k)}{\mathbb{P}(Y = k)}, \qquad n \geqslant 0,$$

provided that $\mathbb{P}(Y = k) > 0$, $k \geqslant 0$.

A.6 Expectation of Random Variables

The *expectation*, or *expected value*, of a random variable X is the mean, or average value, of X. In practice, expectations can be even more useful than probabilities. For example, knowing that a given equipment (such as a bridge) has a failure probability of 1.78493 out of a billion can be of less practical use than knowing the expected lifetime (*e.g.* 200000 years) of that equipment.

For example, the time $T(\omega)$ to travel from home to work/school can be a random variable with a new outcome and value every day, however we usually refer to its expectation $\mathbb{E}[T]$ rather than to its sample values that may change from day to day.

Expected value of a Bernoulli random variable

Any Bernoulli random variable $X : \Omega \longrightarrow \{0, 1\}$ can be written as the indicator function $X := \mathbb{1}_A$ where A is the event $A = \{X = 1\}$, and the parameter $p \in [0, 1]$ of X is given by

$$p = \mathbb{P}(X = 1) = \mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[X].$$

The expectation of a Bernoulli random variable with parameter p is defined as

$$\boxed{\mathbb{E}[\mathbb{1}_A] := 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A). \quad (A.26)}$$

Expected value of a discrete random variable

Next, let $X : \Omega \longrightarrow \mathbb{N}$ be a discrete random variable. The expectation $\mathbb{E}[X]$ of X is defined as the sum

$$\mathbb{E}[X] = \sum_{k\geqslant 0} k\mathbb{P}(X = k), \tag{A.27}$$

in which the possible values $k \geqslant 0$ of X are weighted by their probabilities. More generally we have

$$\mathbb{E}[\phi(X)] = \sum_{k\geqslant 0} \phi(k)\mathbb{P}(X = k),$$

for all sufficiently summable functions $\phi : \mathbb{N} \longrightarrow \mathbb{R}$.

The expectation of the indicator function $X = \mathbb{1}_A = \mathbb{1}_{\{X=1\}}$ can be recovered from (A.27) as

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_A] = 0 \times \mathbb{P}(\Omega \setminus A) + 1 \times \mathbb{P}(A) = 0 + \mathbb{P}(A) = \mathbb{P}(A).$$

Note that the expectation is a *linear* operation, *i.e.* we have

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \qquad a, b \in \mathbb{R}, \tag{A.28}$$

provided that

$$\mathbb{E}[|X|] + \mathbb{E}[|Y|] < \infty.$$

Examples

i) Expected value of a Poisson random variable with parameter $\lambda > 0$:

$$\mathbb{E}[X] = \sum_{k\geqslant 0} k\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \sum_{k\geqslant 1} k\frac{\lambda^k}{k!} = \lambda \mathrm{e}^{-\lambda} \sum_{k\geqslant 0} \frac{\lambda^k}{k!} = \lambda. \tag{A.29}$$

ii) Estimating the expected value of a Poisson random variable using R:

Taking $\lambda := 2$, we can use the following R code:

```
poisson_samples <- rpois(100000, lambda = 2)
poisson_samples
mean(poisson_samples)
```

Given $X : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ a discrete nonnegative random variable X, we have

$$\mathbb{P}(X < \infty) = \sum_{k\geqslant 0} \mathbb{P}(X = k),$$

and

$$1 = \mathbb{P}(X = \infty) + \mathbb{P}(X < \infty) = \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} \mathbb{P}(X = k),$$

and in general

$$\mathbb{E}[X] = +\infty \times \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} k \mathbb{P}(X = k).$$

In particular, $\mathbb{P}(X = \infty) > 0$ implies $\mathbb{E}[X] = \infty$, and the finiteness condition $\mathbb{E}[X] < \infty$ implies $\mathbb{P}(X < \infty) = 1$, however the converse is *not true*. For example, assume that X has the geometric distribution

$$\mathbb{P}(X = k) := \frac{1}{2^{k+1}}, \qquad k \geqslant 0, \tag{A.30}$$

with parameter $p = 1/2$, and

$$\mathbb{E}[X] = \sum_{k \geqslant 0} \frac{k}{2^{k+1}} = \frac{1}{4} \sum_{k \geqslant 1} \frac{k}{2^{k-1}} = \frac{1}{4} \frac{1}{(1 - 1/2)^2} = 1 < \infty.$$

Letting $\phi(X) := 2^X$, we have

$$\mathbb{P}(\phi(X) < \infty) = \mathbb{P}(X < \infty) = \sum_{k \geqslant 0} \frac{1}{2^{k+1}} = 1,$$

and

$$\mathbb{E}[\phi(X)] = \sum_{k \geqslant 0} \phi(k) \mathbb{P}(X = k) = \sum_{k \geqslant 0} \frac{2^k}{2^{k+1}} = \sum_{k \geqslant 0} \frac{1}{2} = +\infty,$$

hence the expectation $\mathbb{E}[\phi(X)]$ is *infinite* although $\phi(X)$ is *finite* with probability one.*

Conditional expectation

The notion of expectation takes its full meaning under conditioning. For example, the expected return of a random asset usually depends on information such as economic data, location, etc. In this case, replacing the expectation by a conditional expectation will provide a better estimate of the expected value.

* This is the St. Petersburg paradox.

For instance, life expectancy is a natural example of a conditional expectation since it typically depends on location, gender, and other parameters.

The *conditional expectation* of a finite discrete random variable $X : \Omega \longrightarrow \mathbb{N}$ given an event A is defined by

$$\mathbb{E}[X \mid A] = \sum_{k\geqslant 0} k\mathbb{P}(X = k \mid A) = \sum_{k\geqslant 1} k\frac{\mathbb{P}(X = k \text{ and } A)}{\mathbb{P}(A)}.$$

Lemma A.15. *Given an event A such that $\mathbb{P}(A) > 0$, we have*

$$\mathbb{E}[X \mid A] = \frac{1}{\mathbb{P}(A)}\mathbb{E}\big[X\mathbb{1}_A\big]. \tag{A.31}$$

Proof. The proof is done only for $X : \Omega \longrightarrow \mathbb{N}$ a discrete random variable, however (A.31) is valid for general real-valued random variables. By Relation (A.11) we have

$$\begin{aligned}
\mathbb{E}[X \mid A] &= \sum_{k\geqslant 0} k\mathbb{P}(X = k \mid A) \\
&= \frac{1}{\mathbb{P}(A)}\sum_{k\geqslant 0} k\mathbb{P}(\{X = k\} \cap A) = \frac{1}{\mathbb{P}(A)}\sum_{k\geqslant 0} k\mathbb{E}\big[\mathbb{1}_{\{X=k\}\cap A}\big] \\
&= \frac{1}{\mathbb{P}(A)}\sum_{k\geqslant 0} k\mathbb{E}\big[\mathbb{1}_{\{X=k\}}\mathbb{1}_A\big] = \frac{1}{\mathbb{P}(A)}\mathbb{E}\left[\mathbb{1}_A\sum_{k\geqslant 0} k\mathbb{1}_{\{X=k\}}\right] \\
&= \frac{1}{\mathbb{P}(A)}\mathbb{E}\big[\mathbb{1}_A X\big],
\end{aligned}$$

where we used the relation

$$X = \sum_{k\geqslant 0} k\mathbb{1}_{\{X=k\}}$$

which holds since X takes only integer values. □

Example

i) For example, consider $\Omega = \{1, 3, -1, -2, 5, 7\}$ with the non-uniform probability measure given by

$$\mathbb{P}(\{-1\}) = \mathbb{P}(\{-2\}) = \mathbb{P}(\{1\}) = \mathbb{P}(\{3\}) = \frac{1}{7},\ \mathbb{P}(\{5\}) = \frac{2}{7},\ \mathbb{P}(\{7\}) = \frac{1}{7},$$

and the random variable

$$X : \Omega \longrightarrow \mathbb{Z}$$

given by

$$X(k) = k, \qquad k = 1, 3, -1, -2, 5, 7.$$

Here, $\mathbb{E}[X \mid X > 1]$ denotes the expected value of X given

$$A = \{X > 1\} = \{3, 5, 7\} \subset \Omega,$$

i.e. the mean value of X given that X is strictly positive. This conditional expectation can be computed as

$$\begin{aligned}
&\mathbb{E}[X \mid X > 1] \\
&= 3 \times \mathbb{P}(X = 3 \mid X > 1) + 5 \times \mathbb{P}(X = 5 \mid X > 1) + 7 \times \mathbb{P}(X = 7 \mid X > 1) \\
&= \frac{3 + 2 \times 5 + 7}{4} \\
&= \frac{3 + 5 + 5 + 7}{7 \times 4/7} \\
&= \frac{1}{\mathbb{P}(X > 1)} \mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big],
\end{aligned}$$

where $\mathbb{P}(X > 1) = 4/7$ and the truncated expectation $\mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big]$ is given by $\mathbb{E}\big[X \mathbb{1}_{\{X>1\}}\big] = (3 + 2 \times 5 + 7)/7$.

ii) Estimating a conditional expectation using R:

```
geo_samples <- rgeom(100000, prob = 1/4)
mean(geo_samples)
mean(geo_samples[geo_samples<10])
```

Taking $p := 3/4$, we have

$$\mathbb{E}[X] = (1 - p) \sum_{k \geqslant 1} k p^k = \frac{p}{1 - p} = 3,$$

and

$$\begin{aligned}
\mathbb{E}[X \mid X < 10] &= \frac{1}{\mathbb{P}(X < 10)} \mathbb{E}\big[X \mathbb{1}_{\{X<10\}}\big] \\
&= \frac{1}{\mathbb{P}(X < 10)} \sum_{k=0}^{9} k \mathbb{P}(X = k) \\
&= \frac{1}{\sum_{k=0}^{9} p^k} \sum_{k=1}^{9} k p^k
\end{aligned}$$

$$
\begin{aligned}
&= \frac{p(1-p)}{1-p^{10}} \frac{\partial}{\partial p} \sum_{k=0}^{9} p^k \\
&= \frac{p(1-p)}{1-p^{10}} \frac{\partial}{\partial p} \left(\frac{1-p^{10}}{1-p} \right) \\
&= \frac{p(1-p^{10}-10(1-p)p^9)}{(1-p)(1-p^{10})} \\
&\simeq 2.4032603455.
\end{aligned}
$$

If the random variable $X : \Omega \longrightarrow \mathbb{N}$ is independent* of the event A, we have

$$
\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[X]\mathbb{P}(A),
$$

and we naturally find

$$
\mathbb{E}[X \mid A] = \mathbb{E}[X]. \tag{A.32}
$$

Taking $X = \mathbb{1}_A$ with

$$
\begin{aligned}
\mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\
\omega &\longmapsto \mathbb{1}_A := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}
\end{aligned}
$$

shows that, in particular,

$$
\begin{aligned}
\mathbb{E}[\mathbb{1}_A \mid A] &= 0 \times \mathbb{P}(X = 0 \mid A) + 1 \times \mathbb{P}(X = 1 \mid A) \\
&= \mathbb{P}(X = 1 \mid A) \\
&= \mathbb{P}(A \mid A) \\
&= 1.
\end{aligned}
$$

One can also define the conditional expectation of X given $A = \{Y = k\}$, as

$$
\mathbb{E}[X \mid Y = k] = \sum_{n \geqslant 0} n \mathbb{P}(X = n \mid Y = k),
$$

where $Y : \Omega \longrightarrow \mathbb{N}$ is a discrete random variable.

Proposition A.16. *Given X a discrete random variable such that $\mathbb{E}[|X|] < \infty$, we have the relation*

$$
\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]], \tag{A.33}
$$

which is sometimes referred to as the tower property.

* *i.e.*, $\mathbb{P}(\{X = k\} \cap A) = \mathbb{P}(\{X = k\})\mathbb{P}(A)$ for all $k \geqslant 0$.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Proof. We have

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X \mid Y]] &= \sum_{k\geqslant 0} \mathbb{E}[X \mid Y = k]\mathbb{P}(Y = k) \\ &= \sum_{k\geqslant 0}\sum_{n\geqslant 0} n\mathbb{P}(X = n \mid Y = k)\mathbb{P}(Y = k) \\ &= \sum_{n\geqslant 0} n \sum_{k\geqslant 0} \mathbb{P}(X = n \text{ and } Y = k) \\ &= \sum_{n\geqslant 0} n\mathbb{P}(X = n) = \mathbb{E}[X],\end{aligned}$$

where we used the marginal distribution

$$\mathbb{P}(X = n) = \sum_{k\geqslant 0} \mathbb{P}(X = n \text{ and } Y = k), \qquad n \geqslant 0,$$

that follows from the *law of total probability* (A.9) with $A_k = \{Y = k\}$, $k \geqslant 0$. □

Taking

$$Y = \sum_{k\geqslant 0} k\mathbb{1}_{A_k},$$

with $A_k := \{Y = k\}$, $k \geqslant 0$, from (A.33) we also get the *law of total expectation*

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X \mid Y]] \\ &= \sum_{k\geqslant 0} \mathbb{E}[X \mid Y = k]\mathbb{P}(Y = k) \\ &= \sum_{k\geqslant 0} \mathbb{E}[X \mid A_k]\mathbb{P}(A_k).\end{aligned} \tag{A.34}$$

Example

Life expectancy in Singapore is $\mathbb{E}[T] = 80$ years overall, where T denotes the lifetime of a given individual chosen at random. Let $G \in \{m, w\}$ denote the gender of that individual. The statistics show that

$$\mathbb{E}[T \mid G = m] = 78 \quad \text{and} \quad \mathbb{E}[T \mid G = w] = 81.9,$$

and we have

$$80 = \mathbb{E}[T]$$

$$
\begin{aligned}
&= \mathbb{E}[\mathbb{E}[T|G]] \\
&= \mathbb{P}(G = w)\mathbb{E}[T \mid G = w] + \mathbb{P}(G = m)\mathbb{E}[T \mid G = m] \\
&= 81.9 \times \mathbb{P}(G = w) + 78 \times \mathbb{P}(G = m) \\
&= 81.9 \times (1 - \mathbb{P}(G = m)) + 78 \times \mathbb{P}(G = m),
\end{aligned}
$$

showing that

$$
80 = 81.9 \times (1 - \mathbb{P}(G = m)) + 78 \times \mathbb{P}(G = m),
$$

i.e.

$$
\mathbb{P}(G = m) = \frac{81.9 - 80}{81.9 - 78} = \frac{1.9}{3.9} = 0.487.
$$

Variance

The *variance* of a random variable X is defined by

$$
\mathrm{Var}[X] := \mathbb{E}\big[X^2\big] - (\mathbb{E}[X])^2,
$$

provided that $\mathbb{E}\big[|X|^2\big] < \infty$. If $(X_k)_{k=1,\ldots,n}$ is a sequence of independent random variables, we have

$$
\begin{aligned}
&\mathrm{Var}\left[\sum_{k=1}^n X_k\right] = \mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] - \left(\mathbb{E}\left[\sum_{k=1}^n X_k\right]\right)^2 \\
&= \mathbb{E}\left[\sum_{k=1}^n X_k \sum_{l=1}^n X_l\right] - \mathbb{E}\left[\sum_{k=1}^n X_k\right]\mathbb{E}\left[\sum_{l=1}^n X_l\right] \\
&= \mathbb{E}\left[\sum_{k=1}^n \sum_{l=1}^n X_k X_l\right] - \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}[X_k]\mathbb{E}[X_l] \\
&= \sum_{k=1}^n \mathbb{E}\big[X_k^2\big] + \sum_{1 \leqslant k \neq l \leqslant n} \mathbb{E}[X_k X_l] - \sum_{k=1}^n (\mathbb{E}[X_k])^2 - \sum_{1 \leqslant k \neq l \leqslant n} \mathbb{E}[X_k]\mathbb{E}[X_l] \\
&= \sum_{k=1}^n \big(\mathbb{E}\big[X_k^2\big] - (\mathbb{E}[X_k])^2\big) \\
&= \sum_{k=1}^n \mathrm{Var}\,[X_k]. \qquad \text{(A.35)}
\end{aligned}
$$

Random sums

In what follows, we consider $Y : \Omega \longrightarrow \mathbb{N}$ an *a.s.* finite, integer-valued random variable, *i.e.* we have $\mathbb{P}(Y < \infty) = 1$ and $\mathbb{P}(Y = \infty) = 0$. Based on the tower property of conditional expectations (A.33) or ordinary conditioning, the expectation of a random sum $\sum_{k=1}^{Y} X_k$, where $(X_k)_{k\in\mathbb{N}}$ is a sequence of random variables, can be computed from the *tower property* (A.33) or from the *law of total expectation* (A.34) as

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^{Y} X_k\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{Y} X_k \,\Big|\, Y\right]\right] \\ &= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{Y} X_k \,\Big|\, Y = n\right] \mathbb{P}(Y = n) \\ &= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{n} X_k \,\Big|\, Y = n\right] \mathbb{P}(Y = n),\end{aligned}$$

and if the sequence $(X_k)_{k\in\mathbb{N}}$ is (mutually) independent of Y, this yields

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^{Y} X_k\right] &= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{n} X_k\right] \mathbb{P}(Y = n) \\ &= \sum_{n\geqslant 0} \mathbb{P}(Y = n) \sum_{k=1}^{n} \mathbb{E}[X_k].\end{aligned}$$

Random products

Similarly, for a random product we will have, using the independence of Y with $(X_k)_{k\in\mathbb{N}}$,

$$\begin{aligned}\mathbb{E}\left[\prod_{k=1}^{Y} X_k\right] &= \sum_{n\geqslant 0} \mathbb{E}\left[\prod_{k=1}^{n} X_k\right] \mathbb{P}(Y = n) \qquad \text{(A.36)} \\ &= \sum_{n\geqslant 0} \mathbb{P}(Y = n) \prod_{k=1}^{n} \mathbb{E}[X_k],\end{aligned}$$

where the last equality requires the (mutual) independence of the random variables in the sequence $(X_k)_{k\geqslant 1}$.

Distributions admitting a density

Given a random variable X whose distribution admits a probability density $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\varphi_X(x)dx,$$

and more generally,

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)\varphi_X(x)dx, \tag{A.37}$$

for all sufficiently integrable function ϕ on $\mathbb{R}$. For example, if X has a standard normal distribution we have

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

Examples

a) In case X has a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, we have

$$\mathbb{E}[\phi(X)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \phi(x)\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}dx. \tag{A.38}$$

b) The uniform random variable U on $[0,1]$ satisfies $\mathbb{E}[U] = 1/2 < \infty$ and

$$\mathbb{P}(1/U < \infty) = \mathbb{P}(U > 0) = \mathbb{P}(U \in (0,1]) = 1,$$

however we have

$$\mathbb{E}[1/U] = \int_0^1 \frac{dx}{x} = +\infty,$$

and $\mathbb{P}(1/U = +\infty) = \mathbb{P}(U = 0) = 0$.

c) If the random variable X has an exponential distribution with parameter $\mu > 0$ we have

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \mu\int_0^{\infty} \mathrm{e}^{\lambda x}\mathrm{e}^{-\mu x}dx = \begin{cases} \dfrac{\mu}{\mu-\lambda} < \infty & \text{if } \mu > \lambda, \\ \\ +\infty, & \text{if } \mu \leqslant \lambda. \end{cases}$$

Exercise: In case $X \simeq \mathcal{N}(\mu, \sigma^2)$ has a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, check that

$$\mu = \mathbb{E}[X] \qquad \text{and} \qquad \sigma^2 = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

When $(X,Y) : \Omega \longrightarrow \mathbb{R}^2$ is a $\mathbb{R}^2$-valued couple of random variables whose distribution admits a probability density $\varphi_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ we have

$$\mathbb{E}[\phi(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x,y)\varphi_{X,Y}(x,y)dxdy,$$

for all sufficiently integrable function ϕ on $\mathbb{R}^2$.

The expectation of an absolutely continuous random variable satisfies the same linearity property (A.28) as in the discrete case.

The conditional expectation of an absolutely continuous random variable can be defined as

$$\mathbb{E}[X \mid Y=y] = \int_{-\infty}^{\infty} x\varphi_{X|Y=y}(x)dx$$

where the conditional probability density $\varphi_{X|Y=y}(x)$ is defined in (A.19), with the relation

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] \tag{A.39}$$

which is called the *tower property* and holds as in the discrete case, since

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y=y]\varphi_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\varphi_{X|Y=y}(x)\varphi_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} \varphi_{(X,Y)}(x,y)dydx \\
&= \int_{-\infty}^{\infty} x\varphi_X(x)dx = \mathbb{E}[X],
\end{aligned}$$

where we used Relation (A.18) between the probability density of (X,Y) and its marginal X.

For example, an exponentially distributed random variable X with probability density function (A.14) has the expected value

$$\mathbb{E}[X] = \lambda\int_0^{\infty} x\mathrm{e}^{-\lambda x}dx = \frac{1}{\lambda}.$$

Proposition A.17. *(Fatou's lemma). Let* $(F_n)_{n\in\mathbb{N}}$ *be a sequence of* non-negative *random variable. Then we have*

$$\mathbb{E}\big[\liminf_{n\to\infty} F_n\big] \leqslant \liminf_{n\to\infty} \mathbb{E}[F_n].$$

In particular, Fatou's lemma shows that if in addition the sequence $(F_n)_{n\in\mathbb{N}}$ converges with probability one and the sequence $(\mathbb{E}[F_n])_{n\in\mathbb{N}}$ converges in $\mathbb{R}$

then we have

$$\mathbb{E}\Big[\lim_{n\to\infty} F_n\Big] \leqslant \lim_{n\to\infty} \mathbb{E}[F_n].$$

Moment Generating Functions

Characteristic functions

The *characteristic function* of a random variable X is the function

$$\Psi_X : \mathbb{R} \longrightarrow \mathbb{C}$$

defined by

$$\Psi_X(t) = \mathbb{E}\big[\mathrm{e}^{itX}\big], \qquad t \in \mathbb{R}.$$

The characteristic function Ψ_X of a random variable X with probability density function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ satisfies

$$\Psi_X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{ixt}\varphi(x)dx, \qquad t \in \mathbb{R}.$$

On the other hand, if $X : \Omega \longrightarrow \mathbb{N}$ is a discrete random variable we have

$$\Psi_X(t) = \sum_{n\geqslant 0} \mathrm{e}^{itn}\mathbb{P}(X = n), \qquad t \in \mathbb{R}.$$

One of the main applications of characteristic functions is to provide a characterization of probability distributions, as in the following theorem.

Theorem A.18. *Two random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ have same distribution if and only if*

$$\Psi_X(t) = \Psi_Y(t), \qquad t \in \mathbb{R}.$$

Theorem A.18 is used to identify or to determine the probability distribution of a random variable X, by comparison with the characteristic function Ψ_Y of a random variable Y whose distribution is known.

The characteristic function of a random vector (X, Y) is the function $\Psi_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{C}$ defined by

$$\Psi_{X,Y}(s,t) = \mathbb{E}\big[\mathrm{e}^{isX+itY}\big], \qquad s,t \in \mathbb{R}.$$

Theorem A.19. *The random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ are* independent *if and only if*

$$\Psi_{X,Y}(s,t) = \Psi_X(s)\Psi_Y(t), \qquad s,t \in \mathbb{R}.$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

A random variable X has a Gaussian distribution with mean μ and variance σ^2 if and only if its characteristic function satisfies

$$\mathbb{E}\left[e^{i\alpha X}\right] = e^{i\alpha\mu - \alpha^2\sigma^2/2}, \qquad \alpha \in \mathbb{R}. \tag{A.40}$$

From Theorems A.18 and A.19 we deduce the following proposition.

Proposition A.20. *Let $X \simeq \mathcal{N}(\mu, \sigma_X^2)$ and $Y \simeq \mathcal{N}(\nu, \sigma_Y^2)$ be independent Gaussian random variables. Then $X + Y$ also has a Gaussian distribution*

$$X + Y \simeq \mathcal{N}(\mu + \nu, \sigma_X^2 + \sigma_Y^2).$$

Proof. Since X and Y are independent, by Theorem A.19 the characteristic function Ψ_{X+Y} of $X + Y$ is given by

$$\begin{aligned}
\Phi_{X+Y}(t) &= \Phi_X(t)\Phi_Y(t) \\
&= e^{it\mu - t^2\sigma_X^2/2} e^{it\nu - t^2\sigma_Y^2/2} \\
&= e^{it(\mu+\nu) - t^2(\sigma_X^2 + \sigma_Y^2)/2}, \qquad t \in \mathbb{R},
\end{aligned}$$

where we used (A.40). Consequently, the characteristic function of $X + Y$ is that of a Gaussian random variable with mean $\mu + \nu$ and variance $\sigma_X^2 + \sigma_Y^2$ and we conclude by Theorem A.18. $\square$

Moment generating functions

The *moment generating function* of a random variable X is the function $\Phi_X : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$\Phi_X(t) := \mathbb{E}\left[e^{tX}\right],$$

for t in a neighborhood of 0. In particular, we have

$$\mathbb{E}[X^n] = \frac{\partial^n}{\partial t^n}\Phi_X(0), \qquad n \geqslant 1,$$

provided that $\mathbb{E}[|X|^n] < \infty$, and

$$\Phi_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{n \geqslant 0} \frac{t^n}{n!}\mathbb{E}[X^n],$$

provided that $\mathbb{E}\left[e^{t|X|}\right] < \infty$, $t \in \mathbb{R}$, and for this reason the moment generating function G_X characterizes the *moments* $\mathbb{E}[X^n]$ of $X : \Omega \longrightarrow \mathbb{N}$, $n \geqslant 0$.

The moment generating function Φ_X of a random variable X with probability density function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ satisfies

$$\Phi_X(t) = \int_{-\infty}^{\infty} e^{xt} \varphi(x) dx, \qquad t \in \mathbb{R}.$$

For example, the moment generating functions (MGF) of a Gaussian random variable X with mean μ and variance σ^2 is given by

$$\mathbb{E}\left[e^{\alpha X}\right] = e^{\alpha\mu + \alpha^2\sigma^2/2}, \qquad \alpha \in \mathbb{R}. \tag{A.41}$$

Note that in probability, the moment generating function is written as a *bilateral* transform defined using an integral from $-\infty$ to $+\infty$.

A.7 Conditional Expectation

The construction of conditional expectations of the form $\mathbb{E}[X \mid Y]$ given above for discrete and absolutely continuous random variables can be generalized to σ-algebras.

Definition A.21. *Given $\mathcal{F}$ a σ-algebra on Ω, a random variable $X : \Omega \longrightarrow \mathbb{R}$ is said to be $\mathcal{F}$-measurable if*

$$\{X \leqslant x\} := \{\omega \in \Omega \ : \ X(\omega) \leqslant x\} \in \mathcal{F},$$

for all $x \in \mathbb{R}$.

Intuitively, when X is $\mathcal{F}$-measurable, the knowledge of the values of X depends only on the information contained in $\mathcal{F}$. For example, when $\mathcal{F} = \sigma(A_1, \ldots, A_n)$ where $(A_n)_{n\geqslant 1}$ is a *partition* of Ω with $\bigcup_{n\geqslant 1} A_n = \Omega$, any $\mathcal{F}$-measurable random variable X can be written as

$$X(\omega) = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}(\omega), \qquad \omega \in \Omega,$$

for some $c_1, \ldots, c_n \in \mathbb{R}$.

Definition A.22. *Given $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space we let $L^2(\Omega, \mathcal{F})$ denote the space of $\mathcal{F}$-measurable and square-integrable random variables, i.e.*

$$L^2(\Omega, \mathcal{F}) := \left\{X : \Omega \longrightarrow \mathbb{R} \ : \ \mathbb{E}\left[|X|^2\right] < \infty\right\}.$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

More generally, for $p \geqslant 1$ one can define the space $L^p(\Omega, \mathcal{F})$ of $\mathcal{F}$-measurable and p-integrable random variables as

$$L^p(\Omega, \mathcal{F}) := \big\{ X : \Omega \longrightarrow \mathbb{R} \ : \ \mathbb{E}[|X|^p] < \infty \big\}.$$

We define a *inner product* $\langle \cdot, \cdot \rangle_{L^2(\Omega,\mathcal{F})}$ between elements of $L^2(\Omega, \mathcal{F})$, as

$$\langle X, Y \rangle_{L^2(\Omega,\mathcal{F})} := \mathbb{E}[XY], \quad X, Y \in L^2(\Omega, \mathcal{F}). \tag{A.42}$$

This inner product is associated to the norm $\| \cdot \|_{L^2(\Omega)}$ by the relation

$$\|X\|_{L^2(\Omega)} = \sqrt{\mathbb{E}\big[X^2\big]} = \sqrt{\langle X, X \rangle_{L^2(\Omega,\mathcal{F})}}, \qquad X \in L^2(\Omega, \mathcal{F}).$$

The norm $\| \cdot \|_{L^2(\Omega)}$ also defines the *mean-square* distance

$$\|X - Y\|_{L^2(\Omega)} = \sqrt{\mathbb{E}\big[(X - Y)^2\big]}$$

between random variables $X, Y \in L^2(\Omega, \mathcal{F})$, and it induces a notion of *orthogonality*, namely X is *orthogonal* to Y in $L^2(\Omega, \mathcal{F})$ if and only if

$$\langle X, Y \rangle_{L^2(\Omega,\mathcal{F})} = 0.$$

Proposition A.23. *The ordinary expectation* $\mathbb{E}[X]$ *achieves the minimum distance*

$$\big\|X - \mathbb{E}[X]\big\|^2_{L^2(\Omega)} = \min_{c \in \mathbb{R}} \|X - c\|^2_{L^2(\Omega)}. \tag{A.43}$$

Proof. It suffices to differentiate

$$\frac{\partial}{\partial c} \mathbb{E}\big[(X - c)^2\big] = -2\mathbb{E}[X - c] = 0,$$

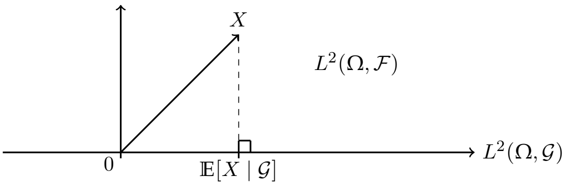
showing that the minimum in (A.43) is reached when $\mathbb{E}[X - c] = 0$, *i.e.* $c = \mathbb{E}[X]$. □

Similarly to Proposition A.23, the conditional expectation will be defined by a distance minimizing procedure.

Definition A.24. *Given* $\mathcal{G} \subset \mathcal{F}$ *a sub* σ*-algebra of* $\mathcal{F}$ *and* $X \in L^2(\Omega, \mathcal{F})$, *the conditional expectation of* X *given* $\mathcal{G}$, *and denoted*

$$\mathbb{E}[X \mid \mathcal{G}],$$

is defined as the orthogonal projection *of* X *onto* $L^2(\Omega, \mathcal{G})$.



As a consequence of the uniqueness of the orthogonal projection onto the subspace $L^2(\Omega,\mathcal{G})$ of $L^2(\Omega,\mathcal{F})$, the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is characterized by the relation

$$\langle Y, X - \mathbb{E}[X \mid \mathcal{G}]\rangle_{L^2(\Omega,\mathcal{F})} = 0,$$

which rewrites as

$$\mathbb{E}[Y(X - \mathbb{E}[X \mid \mathcal{G}])] = 0,$$

i.e.

$$\mathbb{E}[YX] = \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]],$$

for all bounded and $\mathcal{H}$-measurable random variables Y, where $\langle\cdot,\cdot\rangle_{L^2(\Omega,\mathcal{F})}$ denotes the inner product (A.42) in $L^2(\Omega,\mathcal{F})$. The next proposition extends Proposition A.23 as a consequence of Definition A.24. See Theorem 5.1.4 page 197 of Stroock (2011) for an extension of the construction of conditional expectation to the space $L^1(\Omega,\mathcal{F})$ of integrable random variable.

Proposition A.25. *The conditional expectation* $\mathbb{E}[X \mid \mathcal{G}]$ *realizes the minimum in mean-square distance between* $X \in L^2(\Omega,\mathcal{F})$ *and* $L^2(\Omega,\mathcal{G})$, *i.e. we have*

$$\|X - \mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\Omega)} = \min_{Y \in L^2(\Omega,\mathcal{G})} \|X - Y\|_{L^2(\Omega)}. \tag{A.44}$$

Proof. This is a consequence of the Pythagorean theorem written as

$$\|X - Y\|_{L^2(\Omega)} = \|X - \mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\Omega)} + \|\mathbb{E}[X \mid \mathcal{G}] - Y\|_{L^2(\Omega)},$$

for any $Y \in L^2(\Omega,\mathcal{G})$. □

The following proposition will often be used as a characterization of $\mathbb{E}[X \mid \mathcal{G}]$.

Proposition A.26. *Given* $X \in L^2(\Omega,\mathcal{F})$, $Z := \mathbb{E}[X \mid \mathcal{G}]$ *is the unique random variable* Z *in* $L^2(\Omega,\mathcal{G})$ *that satisfies the relation*

$$\mathbb{E}[YX] = \mathbb{E}[YZ] \tag{A.45}$$

for all bounded and $\mathcal{G}$*-measurable random variables* Y.

We note that taking $Y = \mathbf{1}$ in (A.45) yields

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]. \tag{A.46}$$

In particular, when $\mathcal{G} = \{\emptyset, \Omega\}$ we have $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid \{\emptyset, \Omega\}]$ and

$$\mathbb{E}[X \mid \{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X \mid \{\emptyset, \Omega\}]] = \mathbb{E}[X], \tag{A.47}$$

because $\mathbb{E}[X \mid \{\emptyset, \Omega\}]$ is in $L^2(\Omega, \{\emptyset, \Omega\})$ and is *a.s.* constant. In addition, the conditional expectation operator has the following properties.

i) $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$ if Y depends only on the information contained in $\mathcal{G}$.

   Proof. By the characterization (A.45) it suffices to show that

   $$\mathbb{E}[H(XY)] = \mathbb{E}[H(Y\mathbb{E}[X|\mathcal{G}])], \tag{A.48}$$

   for all bounded and $\mathcal{G}$-measurable random variables H, which implies $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$.

   Relation (A.48) holds from (A.45) because the product HY is $\mathcal{G}$-measurable hence Y in (A.45) can be replaced with HY.

ii) $\mathbb{E}[Y|\mathcal{G}] = Y$ when Y depends only on the information contained in $\mathcal{G}$.

   Proof. This is a consequence of point (i) above by taking $X := \mathbf{1}$.

iii) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ if $\mathcal{H} \subset \mathcal{G}$, called the *tower property*.

   Proof. First, we note that by (A.46), (iii) holds when $\mathcal{H} = \{\emptyset, \Omega\}$. Next, by the characterization (A.45) it suffices to show that

   $$\mathbb{E}[H\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[H\mathbb{E}[X|\mathcal{H}]], \tag{A.49}$$

   for all bounded and $\mathcal{H}$-measurable random variables H, which will imply (iii) from (A.45).

   In order to prove (A.49) we check that by point (i) above and (A.46) we have

   $$\begin{aligned}\mathbb{E}[H\mathbb{E}[X|\mathcal{G}]] &= \mathbb{E}[\mathbb{E}[HX|\mathcal{G}]] = \mathbb{E}[HX] \\ &= \mathbb{E}[\mathbb{E}[HX|\mathcal{H}]] = \mathbb{E}[H\mathbb{E}[X|\mathcal{H}]],\end{aligned}$$

   and we conclude by the characterization (A.45).

iv) $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ when X "does not depend" on the information contained in $\mathcal{G}$ or, more precisely stated, when the random variable X is *independent* of the σ-algebra $\mathcal{G}$.

Proof. It suffices to note that for all bounded $\mathcal{G}$-measurable Y we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[Y\mathbb{E}[X]],$$

and we conclude again by (A.45).

v) If Y depends only on $\mathcal{G}$ and X is independent of $\mathcal{G}$, then

$$\mathbb{E}[h(X,Y)|\mathcal{G}] = \mathbb{E}[h(X,x)]_{x=Y}. \tag{A.50}$$

Proof. This relation can be proved using the tower property, by noting that for any bounded $K \in L^2(\Omega, \mathcal{G})$ we have

$$\begin{aligned}\mathbb{E}[K\mathbb{E}[h(x,X)]_{x=Y}] &= \mathbb{E}[K\mathbb{E}[h(x,X) \mid \mathcal{G}]_{x=Y}] \\ &= \mathbb{E}[K\mathbb{E}[h(Y,X) \mid \mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[Kh(Y,X) \mid \mathcal{G}]] \\ &= \mathbb{E}[Kh(Y,X)],\end{aligned}$$

which yields (A.50) by the characterization (A.45).

The notion of conditional expectation can be extended from square-integrable random variables in $L^2(\Omega, \mathcal{F})$ to integrable random variables in $L^1(\Omega, \mathcal{F})$, cf. *e.g.* Theorem 5.1 in Kallenberg (2002).

Proposition A.27. *When the σ-algebra $\mathcal{G} := \sigma(A_1, A_2, \ldots, A_n)$ is generated by n disjoint events $A_1, A_2, \ldots, A_n \in \mathcal{F}$, we have*

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{k=1}^n \mathbb{1}_{A_k}\mathbb{E}[X \mid A_k] = \sum_{k=1}^n \mathbb{1}_{A_k}\frac{\mathbb{E}[X\mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}.$$

Proof. It suffices to note that the $\mathcal{G}$-measurable random variables can be generated by indicator functions of the form $\mathbb{1}_{A_l}$, and that

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{A_l}\sum_{k=1}^n \mathbb{1}_{A_k}\frac{\mathbb{E}[X\mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}\right] &= \mathbb{E}\left[\mathbb{1}_{A_l}\frac{\mathbb{E}[X\mathbb{1}_{A_l}]}{\mathbb{P}(A_l)}\right] \\ &= \frac{\mathbb{E}[X\mathbb{1}_{A_l}]}{\mathbb{P}(A_l)}\mathbb{E}[\mathbb{1}_{A_l}] \\ &= \mathbb{E}[X\mathbb{1}_{A_l}], \qquad l = 1, 2, \ldots, n,\end{aligned}$$

showing (A.45). The relation

$$\mathbb{E}[X \mid A_k] = \frac{\mathbb{E}[X\mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}, \qquad k = 1, 2, \ldots, n,$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

follows from Lemma A.15. □

For example, in case $\Omega = \{a, b, c, d\}$ and $\mathcal{G} = \big\{\emptyset, \Omega, \{a, b\}, \{c\}, \{d\}\big\}$, we have

$$\begin{aligned}\mathbb{E}[X \mid \mathcal{G}] &= \mathbb{1}_{\{a,b\}}\mathbb{E}[X \mid \{a, b\}] + \mathbb{1}_{\{c\}}\mathbb{E}[X \mid \{c\}] + \mathbb{1}_{\{d\}}\mathbb{E}[X \mid \{d\}] \\ &= \mathbb{1}_{\{a,b\}}\frac{\mathbb{E}\big[X\mathbb{1}_{\{a,b\}}\big]}{\mathbb{P}(\{a, b\})} + \mathbb{1}_{\{c\}}\frac{\mathbb{E}\big[X\mathbb{1}_{\{c\}}\big]}{\mathbb{P}(\{c\})} + \mathbb{1}_{\{d\}}\frac{\mathbb{E}\big[X\mathbb{1}_{\{d\}}\big]}{\mathbb{P}(\{d\})}.\end{aligned}$$

Regarding conditional probabilities we have similarly, for $A \subset \Omega = \{a, b, c, d\}$,

$$\begin{aligned}\mathbb{P}(A \mid \mathcal{G}) &= \mathbb{1}_{\{a,b\}}\frac{\mathbb{P}\big(A \cap \{a, b\}\big)}{\mathbb{P}(\{a, b\})} + \mathbb{1}_{\{c\}}\frac{\mathbb{P}\big(A \cap \{c\}\big)}{\mathbb{P}(\{c\})} + \mathbb{1}_{\{d\}}\frac{\mathbb{P}\big(A \cap \{d\}\big)}{\mathbb{P}(\{d\})} \\ &= \mathbb{1}_{\{a,b\}}\mathbb{P}(A \mid \{a, b\}) + \mathbb{1}_{\{c\}}\mathbb{P}(A \mid \{c\}) + \mathbb{1}_{\{d\}}\mathbb{P}(A \mid \{d\}).\end{aligned}$$

In particular, if $A = \{a\} \subset \Omega = \{a, b, c, d\}$ we find

$$\begin{aligned}\mathbb{P}(\{a\} \mid \mathcal{G}) &= \mathbb{1}_{\{a,b\}}\mathbb{P}(\{a\} \mid \{a, b\}) \\ &= \mathbb{1}_{\{a,b\}}\frac{\mathbb{P}\big(\{a\} \cap \{a, b\}\big)}{\mathbb{P}(\{a, b\})} \\ &= \mathbb{1}_{\{a,b\}}\frac{\mathbb{P}(\{a\})}{\mathbb{P}(\{a, b\})}.\end{aligned}$$

In other words, the probability of getting the outcome a is $\mathbb{P}(\{a\})/\mathbb{P}(\{a, b\})$ knowing that the outcome is either a or b, otherwise it is zero.

Exercises

Exercise A.1 Let X denote a Poisson random variable with parameter $\lambda > 0$.

a) Compute the expected value $\mathbb{E}[X]$ of X.
b) Compute the second moment $\mathbb{E}[X^2]$ and variance $\mathrm{Var}[X]$ of X.

Exercise A.2 Let X denote a centered Gaussian random variable with variance η^2, $\eta > 0$. Show that the probability $P(e^X > c)$ is given by

$$P\big(e^X > c\big) = \Phi(-(\log c)/\eta),$$

where $\log = \ln$ denotes the natural logarithm and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}dy, \qquad x \in \mathbb{R},$$

denotes the Gaussian cumulative distribution function.

Exercise A.3 Let $X \simeq \mathcal{N}(\mu, \sigma^2)$ be a Gaussian random variable with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, and probability density function

$$\varphi(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in \mathbb{R}.$$

a) Confirm that $\varphi \geqslant 0$ is indeed a probability density function, *i.e.*, show that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} dx = 1.$$

b) Write down $\mathbb{E}[X]$ as an integral, and show that

$$\mu = \mathbb{E}[X].$$

c) Write down $\mathbb{E}[X^2]$ as an integral, and show that

$$\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

d) Write down $\mathbb{E}[\mathrm{e}^X]$ as an integral and prove (A.41), *i.e.* show that

$$\mathbb{E}[\mathrm{e}^X] = \mathrm{e}^{\mu+\sigma^2/2}.$$

Exercise A.4 Let $X \simeq \mathcal{N}(0, \sigma^2)$ be a centered Gaussian random variable with variance $\sigma^2 > 0$ and probability density function

$$\varphi(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in \mathbb{R}.$$

a) Consider the function $x \mapsto x^+$ from $\mathbb{R}$ to $\mathbb{R}_+$, defined as

$$x^+ = \begin{cases} x & \text{if } x \geqslant 0, \\ \\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Compute $\mathbb{E}[X^+]$ as an integral.

b) Consider the function $x \mapsto (x-K)^+$ from $\mathbb{R}$ to $\mathbb{R}_+$, defined as

$$(x-K)^+ = \begin{cases} x - K & \text{if } x \geqslant K, \\ \\ 0 & \text{if } x \leqslant K, \end{cases}$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

where $K \in \mathbb{R}$. Compute $\mathbb{E}\big[(X-K)^+\big]$ as an integral using the cumulative distribution function of the standard normal distribution

$$\Phi(x) := \int_{-\infty}^{x} e^{-y^2/2} \frac{dy}{\sqrt{2\pi}}, \qquad x \in \mathbb{R}.$$

c) Consider the function $x \mapsto (K-x)^+$ from $\mathbb{R}$ to $\mathbb{R}_+$, defined as

$$(K-x)^+ = \begin{cases} K-x & \text{if } x \leqslant K, \\ \\ 0 & \text{if } x \geqslant K, \end{cases}$$

where $K \in \mathbb{R}$. Compute $\mathbb{E}\big[(K-X)^+\big]$ using the cumulative distribution function Φ.

Exercise A.5 Let $X \simeq \mathcal{N}(0, v^2)$ be a centered Gaussian random variable with variance $v^2 > 0$.

a) Compute

$$\mathbb{E}\big[e^{\sigma X} \mathbb{1}_{[K,\infty)}(x e^{\sigma X})\big] = \frac{1}{\sqrt{2\pi v^2}} \int_{\sigma^{-1}\log(K/x)}^{\infty} e^{\sigma y - y^2/(2v^2)} dy.$$

Hint. Use the completion of square identity

$$\sigma y - \frac{y^2}{v^2} = \frac{v^2\sigma^2}{4} - \left(\frac{y}{v} - \frac{v\sigma}{2}\right)^2.$$

b) Compute

$$\mathbb{E}[(e^{m+X} - K)^+] = \frac{1}{\sqrt{2\pi v^2}} \int_{-\infty}^{\infty} (e^{m+x} - K)^+ e^{-x^2/(2v^2)} dx.$$

Some Useful Identities

Here, we present a summary of algebraic identities that are used in this text.

Indicator functions

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ \\ 0 & \text{if } x \notin A. \end{cases} \qquad \mathbf{1}_{[a,b]}(x) = \begin{cases} 1 & \text{if } a \leqslant x \leqslant b, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Binomial coefficients

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}, \qquad 0 \leqslant k \leqslant n.$$

Exponential series

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \qquad x \in \mathbb{R}.$$

Geometric sum

$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}, \qquad r \neq 1. \tag{13.51}$$

Geometric series

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}, \qquad -1 < r < 1. \tag{13.52}$$

Differentiation of geometric series

$$\sum_{k=1}^{\infty} k r^{k-1} = \frac{\partial}{\partial r} \sum_{k=0}^{\infty} r^k = \frac{\partial}{\partial r} \frac{1}{1-r} = \frac{1}{(1-r)^2}, \qquad -1 < r < 1. \tag{13.53}$$

Binomial identities

$$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n.$$

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n. \tag{13.54}$$

$$\begin{aligned}\sum_{k=0}^{n} k \binom{n}{k} a^k b^{n-k} &= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} a^k b^{n-k} \\ &= \sum_{k=0}^{n-1} \frac{n!}{(n-1-k)!k!} a^{k+1} b^{n-1-k} \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} a^{k+1} b^{n-1-k} \\ &= na(a+b)^{n-1}, \qquad n \geqslant 1,\end{aligned}$$

$$\begin{aligned}\sum_{k=0}^{n} k \binom{n}{k} a^k b^{n-k} &= a \frac{\partial}{\partial a} \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \\ &= a \frac{\partial}{\partial a} (a+b)^n \\ &= na(a+b)^{n-1}, \qquad n \geqslant 1.\end{aligned}$$

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}.$$

Sums of integers

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}. \tag{13.55}$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}. \tag{13.56}$$

Taylor expansion

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \frac{x^k}{k!}\alpha(\alpha-1)\times\cdots\times(\alpha-(k-1)). \tag{13.57}$$

Differential equation

The solution of $f'(t) = cf(t)$ is given by $f(t) = f(0)\mathrm{e}^{ct}, \quad t \in \mathbb{R}_+.$ (13.58)

Gaussian MGF. Given Xa Gaussian random variable, we have

$$\mathbb{E}[\mathrm{e}^X] = \exp(\mathbb{E}[X] + \mathrm{Var}[X]/2). \tag{13.59}$$

Exercise Solutions

Chapter 1

Exercise 1.1 According to Definition 1.3, we need to check the following five properties of Brownian motion:

(i) starts at 0 at time 0,

(ii) independence of increments,

(iii) almost sure continuity of trajectories,

(iv) stationarity of the increments,

(v) Gaussianity of increments.

Checking conditions (i) to (iv) does not pose any particular problem since the time changes $t \mapsto c + t$ and $t \mapsto t/c^2$ are deterministic and continuous.

a) Let $X_t := B_{c+t} - B_t$, $t \in \mathbb{R}_+$. For any finite sequence of times $t_0 < t_1 < \cdots < t_n$, the sequence

$$(X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}) \\ = (B_{c+t_1} - B_{c+t_0}, B_{c+t_2} - B_{c+t_1}, \ldots, B_{c+t_n} - B_{c+t_{n-1}})$$

is made of independent random variables. Concerning (v), $X_t - X_s = B_{c+t} - B_{c+s}$ is normally distributed with mean zero and variance $t + c - (c+s) = t - s$ for any $0 \leqslant s < t$.

b) Let $X_t := cB_{t/c^2}-$, $t \in \mathbb{R}_+$. For any finite sequence of times $t_0 < t_1 < \cdots < t_n$, the sequence

$$(X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}) \\ = (cB_{t_1/c^2} - cB_{t_0/c^2}, cB_{t_2/c^2} - cB_{t_1/c^2}, \ldots, cB_{t_n/c^2} - cB_{t_{n-1/c^2}})$$

is made of independent random variables. Concerning (v), $X_t - X_s = cB_{t/c^2} - cB_{s/c^2}$ is normally distributed with mean zero and variance $c^2(t/c^2 - s/c^2) = t - s$ for any $0 \leqslant s < t$, since

$$\operatorname{Var}\left[cB_{t/c^2}\right] = c^2 \operatorname{Var}\left[B_{t/c^2}\right] = c^2 t/c^2 = t, \qquad t \geqslant 0.$$

Exercise 1.2 The solution to (1.16) is given by

$$S_t = S_0 \mathrm{e}^{\sigma B_t + (\mu - \sigma^2/2)t}, \qquad t \geqslant 0,$$

see the proof of Proposition 1.7 for details. The next R code can be used to generate Figure 1.22.

```
N=1000; t <- 0:N; dt <- 1.0/N; nsim <- 10; sigma=0.6; mu=0.001
Z <- c(rnorm(n = N, sd = sqrt(dt)));
plot(t*dt, exp(mu*t), xlab = "time", ylab = "Geometric Brownian motion", type = "l", ylim =
      c(0, 4), col = 1,lwd=3)
lines(t*dt, exp(sigma*c(0,cumsum(Z))+mu*t-sigma*sigma*t*dt/2),xlab = "time",type = "l",ylim
      = c(0, 4), col = 4)
```

Exercise 1.3

a) Those quantities can be computed from the expression of S_t^n as a function of the $\mathcal{N}(0,t)$ random variable B_t for $n \geqslant 1$. Namely, we have

$$\begin{aligned}
\mathbb{E}[S_t^n] &= \mathbb{E}\left[S_0^n \mathrm{e}^{n\sigma B_t - n\sigma^2 t/2 + nrt}\right] \\
&= S_0 \mathrm{e}^{-n\sigma^2 t/2 + nrt} \mathbb{E}\left[\mathrm{e}^{n\sigma B_t}\right] \\
&= S_0 \mathrm{e}^{-n\sigma^2 t/2 + nrt + n^2\sigma^2 t/2} \\
&= S_0^n \mathrm{e}^{nrt + (n-1)n\sigma^2 t/2},
\end{aligned}$$

where we used the Gaussian moment generating function (MGF) formula (A.41), *i.e.*

$$\mathbb{E}\left[\mathrm{e}^{n\sigma B_t}\right] = \mathrm{e}^{n^2\sigma^2 t/2}$$

for the normal random variable $B_t \simeq \mathcal{N}(0,t)$, $t > 0$.

b) By the result of Question (a)), we have $\mathbb{E}[S_t] = S_0 \mathrm{e}^{rt}$ and

$$\begin{aligned}
\mathbb{E}[S_t^2] &= \mathbb{E}\left[S_0^2 \mathrm{e}^{2\sigma B_t - \sigma^2 t + 2rt}\right] \\
&= S_0^2 \mathrm{e}^{-\sigma^2 t + 2rt} \mathbb{E}\left[\mathrm{e}^{2\sigma B_t}\right] \\
&= S_0^2 \mathrm{e}^{\sigma^2 t + 2rt}, \qquad t \geqslant 0.
\end{aligned}$$

Exercise 1.4 From the solution of Exercise 1.3, we have

$$\mathbb{E}\big[S_t^{(i)}\big] = S_0^{(i)}\mathrm{e}^{\mu t}, \qquad t \in [0,T], \quad i = 1,2,$$

and

$$\begin{aligned}\mathrm{Var}\left[S_t^{(i)}\right] &= \mathbb{E}\big[\big(S_t^{(i)}\big)^2\big] - \big(\mathbb{E}\big[S_t^{(i)}\big]\big)\\ &= \big(S_0^{(i)}\big)^2\mathrm{e}^{2\mu t+\sigma_i^2 t} - \big(S_0^{(i)}\big)^2\mathrm{e}^{2\mu t}\\ &= \big(S_0^{(i)}\big)^2\mathrm{e}^{2\mu t}\big(\mathrm{e}^{\sigma_i^2 t}-1\big), \quad t \in [0,T], \ \ i = 1,2.\end{aligned}$$

Hence, we have

$$\mathrm{Var}\left[S_t^{(2)} - S_t^{(1)}\right] = \mathrm{Var}\left[S_t^{(1)}\right] + \mathrm{Var}\left[S_t^{(2)}\right] - 2\,\mathrm{Cov}\left(S_t^{(1)}, S_t^{(2)}\right)$$

with

$$\begin{aligned}\mathbb{E}\big[S_t^{(1)}S_t^{(2)}\big] &= \mathbb{E}\big[S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t+\sigma_1 W_t^{(1)}-\sigma_1^2 t/2+\sigma_2 W_t^{(2)}-\sigma_2^2 t/2}\big]\\ &= S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t-\sigma_1^2 t/2-\sigma_2^2 t/2}\mathbb{E}\big[\mathrm{e}^{\sigma_1 W_t^{(1)}+\sigma_2 W_t^{(2)}}\big]\\ &= S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t-\sigma_1^2 t/2-\sigma_2^2 t/2}\exp\left(\frac{1}{2}\mathbb{E}\big[\big(\sigma_1 W_t^{(1)}+\sigma_2 W_t^{(2)}\big)^2\big]\right),\end{aligned}$$

with

$$\begin{aligned}\mathbb{E}\big[\big(\sigma_1 W_t^{(1)}+\sigma_2 W_t^{(2)}\big)^2\big] &= \mathbb{E}\big[\big(\sigma_1 W_t^{(1)}\big)^2\big] + 2\mathbb{E}\big[\sigma_1 W_t^{(1)}\sigma_2 W_t^{(2)}\big] + \mathbb{E}\big[\big(\sigma_2 W_t^{(2)}\big)^2\big]\\ &= \sigma_1^2 t + 2\rho\sigma_1\sigma_2 t + \sigma_2^2 t,\end{aligned}$$

hence

$$\mathbb{E}\big[S_t^{(1)}S_t^{(2)}\big] = S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t+\rho\sigma_1\sigma_2 t},$$

and

$$\mathrm{Cov}\left(S_t^{(1)}, S_t^{(2)}\right) = \mathbb{E}\big[S_t^{(1)}S_t^{(2)}\big] - \mathbb{E}\big[S_t^{(1)}\big]\mathbb{E}\big[S_t^{(2)}\big] = S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t}(\mathrm{e}^{\rho\sigma_1\sigma_2 t}-1),$$

and therefore

$$\begin{aligned}&\mathrm{Var}\left[S_t^{(2)} - S_t^{(1)}\right]\\ &= \big(S_0^{(1)}\big)^2\mathrm{e}^{2\mu t}(\mathrm{e}^{\sigma_1^2 t}-1) + \big(S_0^{(2)}\big)^2\mathrm{e}^{2\mu t}(\mathrm{e}^{\sigma_2^2 t}-1) - 2S_0^{(1)}S_0^{(2)}\mathrm{e}^{2\mu t}(\mathrm{e}^{\rho\sigma_1\sigma_2 t}-1)\\ &= \mathrm{e}^{2\mu t}\big(\big(S_0^{(1)}\big)^2\mathrm{e}^{\sigma_1^2 t} + \big(S_0^{(2)}\big)^2\mathrm{e}^{\sigma_2^2 t} - 2S_0^{(1)}S_0^{(2)}\mathrm{e}^{\rho\sigma_1\sigma_2 t} - \big(S_0^{(2)}-S_0^{(1)}\big)^2\big).\end{aligned}$$

Exercise 1.5 We have

$$\mathbb{E}\big[S_0\mathrm{e}^{\sigma B_t+\mu t-\sigma^2 t/2}\big] = S_0\mathrm{e}^{\mu t-\sigma^2 t/2}\mathbb{E}\big[\mathrm{e}^{\sigma B_t}\big] = S_0\mathrm{e}^{\mu t-\sigma^2 t/2}\mathrm{e}^{\sigma^2 t/2} = S_0\mathrm{e}^{\mu t}$$

and

$$\mathbb{E}[\log S_t] = \mathbb{E}\left[\log S_0 + \sigma B_t + \mu t - \frac{\sigma^2 t}{2}\right] = (\log S_0) + \mu t - \frac{\sigma^2 t}{2},$$

hence

$$\text{Theil}_t = \log \mathbb{E}[S_t] - \mathbb{E}[\log S_t] = \log S_0 + \mu t - \left((\log S_0) + \mu t - \frac{\sigma^2 t}{2}\right) = \frac{\sigma^2 t}{2}.$$

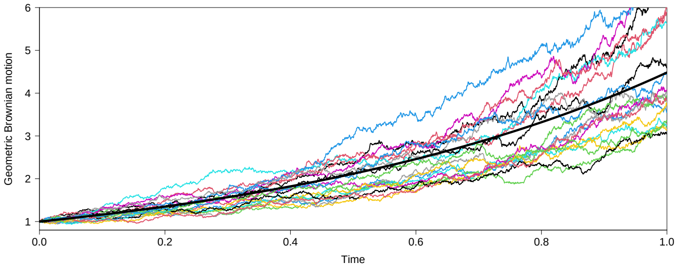


Fig. S.1: Twenty sample paths of geometric Brownian motion $(S_t)_{t\in\mathbb{R}_+}$.

Chapter 2

Exercise 2.1

a) i) By calculation (expected answer). We have

$$\begin{aligned}
\rho(0) &= \text{Cov}(X_n, X_n) \\
&= \text{Var}[X_n] \\
&= \mathbb{E}[X_n^2] \\
&= \mathbb{E}[(Z_n - aZ_{n-1})^2] \\
&= \mathbb{E}[Z_n^2 - 2aZ_{n-1}Z_n + a^2 Z_{n-1}^2] \\
&= \mathbb{E}[Z_n^2] - 2a\mathbb{E}[Z_{n-1}Z_n] + a^2\mathbb{E}[Z_{n-1}^2] \\
&= 1 - 2a\mathbb{E}[Z_{n-1}]\mathbb{E}[Z_n] + a^2 \\
&= 1 + a^2,
\end{aligned}$$

and

$$\begin{aligned}
&\rho(1) = \mathrm{Cov}(X_n, X_{n+1}) \\
&= \mathrm{Cov}(Z_n + aZ_{n-1}, Z_{n+1} + aZ_n) \\
&= \mathrm{Cov}(Z_n, Z_{n+1}) + a\,\mathrm{Cov}(Z_{n-1}, Z_{n+1}) + a\,\mathrm{Cov}(Z_{n-1}, Z_n) + a^2\,\mathrm{Cov}(Z_{n-1}, Z_n) \\
&= \mathrm{Cov}(Z_n, Z_{n+1}) + a\,\mathrm{Cov}(Z_{n-1}, Z_{n+1}) + a\,\mathrm{Cov}(Z_n, Z_n) + a^2\,\mathrm{Cov}(Z_{n-1}, Z_n) \\
&= a\,\mathrm{Var}[Z_n] \\
&= a,
\end{aligned}$$

and for $k \geqslant 2$,

$$\begin{aligned}
\rho(k) &= \mathrm{Cov}(X_n, X_{n+k}) \\
&= \mathrm{Cov}(Z_n + aZ_{n-1}, Z_{n+k} + aZ_{n+k-1}) \\
&= \mathrm{Cov}(Z_n, Z_{n+k}) + a\,\mathrm{Cov}(Z_{n-1}, Z_{n+k}) \\
&\quad + a\,\mathrm{Cov}(Z_n, Z_{n+k-1}) + a^2\,\mathrm{Cov}(Z_{n-1}, Z_{n+k-1}) \\
&= 0,
\end{aligned}$$

since $n < n + k - 1$. By a similar argument, we obtain $\rho(k) = \rho(-k)$ for $k \leqslant 0$.

ii) Confirmation by simulation with an MA(1) time series constructed by hand:

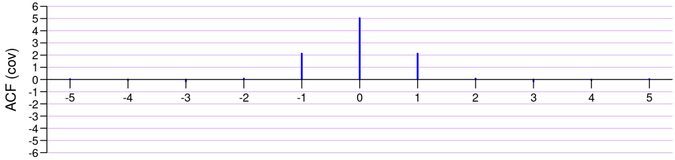
```
library(zoo)
N=10000;Zn<-zoo(rnorm(N,0,1))
Xn<-Zn+2*lag(Zn,-1, na.pad = TRUE);Xn<-Xn[-1]
k=0;cov(Xn[1:(length(Xn)-k)],lag(Xn,k))
```

or with an MA(1) time series constructed using arima.sim:

```
n=2000;a=2;
Xn<-arima.sim(model=list(ma=c(a)),n.start=100,n)
x=seq(100,100+n-1)
plot(x,Xn,pch=19, ylab="X", xlab="n", main = 'MA(1) Samples',col='blue')
lines(x,Xn,col='blue')
Xn<-zoo(Xn)
k=1;cov(Xn[1:(length(Xn)-k)],lag(Xn,k))
```

Using the command ccf to compute the autocovariance empirically, we find:

```
ccf(Xn,Xn,lag = 5, type="covariance", plot=T, lwd=2, col='blue', axes=FALSE,
      ylim=c(-1-a**2,1+a**2), main="")
axis(side = 1, at = seq(-5,5,1))
axis(side = 2, at = c(-1-a**2,-a,0,a,1+a**2), labels=c(expression(paste("-1-|a|"^"2")),
      "a",0,"a", expression(paste("1+|a|"^"2"))))
```



b) We note that $\mathbb{E}[X_n] = 0$, $n \in \mathbb{Z}$, and in addition the autocovariance $\mathrm{Cov}(X_n, X_{n+k})$ depends only of $|k|$ and not on $n \in \mathbb{Z}$. Therefore, the time series $(X_n)_{n\in\mathbb{Z}}$ is *weakly stationary* by Definition 2.11. In addition, by Theorem 2.12 this time series is strictly stationary only if $|a| \neq 1$.

Exercise 2.2

a) We rewrite the equation defining $(X_n)_{n\geqslant 1}$ as

$$X_n = Z_n + LX_n = Z_n + \phi(L)X_n, \qquad n \geqslant 1.$$

where L is the lag operator $LX_n = X_{n-1}$ and $\phi(L) = L$. Taking $\phi(z) := z$, by Theorem 2.12 we need to check whether the solutions of the equation $\phi(z) = 1$ lie on the complete unit circle. As $\phi(z) = 1$ admits the unique solution $z = 1$ which lies on the complete unit circle, we conclude that the AR(1) time series $(X_n)_{n\geqslant 1}$ is not weakly stationary.

b) As in part (a)), we rewrite the AR(2) equation for $(Y_n)_{n\geqslant 1}$ as

$$Y_n = Z_n + 0.75 \times LY_n - 0.125 \times L^2 Y_n = Z_n + \phi(L)Y_n, \qquad n \geqslant 2.$$

with $\phi(L) = 0.75 \times L - 0.125 \times L^2$. The equation $\phi(z) = 1$ with $\phi(z) = 0.75z - 0.125z^2$ reads $z^2 - 6z + 8 = (z-2)(z-4) = 0$. This equation has two solutions $z = 2, 4$ which lie outside the complex unit circle, hence by Theorem 2.12 the AR(2) time series $(Y_n)_{n\geqslant 2}$ is weakly stationary.

Exercise 2.3

a) We have

$$\begin{aligned}
\mathbb{E}[X_{n+1}] &= \mathbb{E}\big[Z_{n+1} + \alpha X_n\big] \\
&= \mathbb{E}\big[Z_{n+1}\big] + \alpha \mathbb{E}\big[X_n\big] \\
&= \alpha \mathbb{E}\big[X_n\big], \qquad n \geqslant 0,
\end{aligned}$$

hence by induction we have $\mathbb{E}[X_n] = \alpha^n \mathbb{E}[X_0] = 0$ for all $n \geqslant 0$.

b) We have

$$\begin{aligned}
\mathrm{Cov}(X_{n+k+1}, X_n) &= \mathbb{E}[X_{n+k+1}X_n] - \mathbb{E}[X_{n+k+1}]\mathbb{E}[X_n] \\
&= \mathbb{E}[X_{n+k+1}X_n] \\
&= \mathbb{E}[(Z_{n+k+1} + \alpha X_{n+k})X_n] \\
&= \mathbb{E}[Z_{n+k+1}X_n + \alpha X_{n+k}X_n] \\
&= \mathbb{E}[Z_{n+k+1}X_n] + \mathbb{E}[\alpha X_{n+k}X_n] \\
&= \mathbb{E}[Z_{n+k+1}]\mathbb{E}[X_n] + \alpha\mathbb{E}[X_{n+k}X_n] \\
&= \alpha\mathbb{E}[X_{n+k}X_n] \\
&= \alpha(\mathbb{E}[X_{n+k}X_n] - \mathbb{E}[X_{n+k}]\mathbb{E}[X_n]) \\
&= \alpha\,\mathrm{Cov}(X_{n+k}, X_n), \qquad n \geqslant 0,
\end{aligned}$$

hence, since $\mathbb{E}[X_n] = 0$, $n \geqslant 0$, we find

$$\mathrm{Cov}(X_{n+k}, X_n) = \alpha^k\,\mathrm{Cov}(X_n, X_n) = \alpha^k\,\mathrm{Var}[X_n], \qquad k, n \geqslant 0.$$

c) We have

$$\begin{aligned}
\mathrm{Var}[X_{n+1}] &= \mathbb{E}[X_{n+1}^2] \\
&= \mathbb{E}[(Z_{n+1} + \alpha X_n)^2] \\
&= \mathbb{E}[Z_{n+1}^2 + 2\alpha Z_{n+1}X_n + \alpha^2 X_n^2] \\
&= \mathbb{E}[Z_{n+1}^2] + 2\alpha\mathbb{E}[Z_{n+1}X_n] + \alpha^2\mathbb{E}[X_n^2] \\
&= 1 + 2\alpha\mathbb{E}[Z_{n+1}]\mathbb{E}[X_n] + \alpha^2\mathbb{E}[X_n^2] \\
&= 1 + \alpha^2\mathbb{E}[X_n^2] \\
&= 1 + \alpha^2\,\mathrm{Var}[X_n].
\end{aligned}$$

By applying the above relation recursively and using the geometric series identity (13.51), we obtain

$$\begin{aligned}
\mathrm{Var}[X_n] &= 1 + \alpha^2\,\mathrm{Var}[X_{n-1}] \\
&= 1 + \alpha^2(1 + \alpha^2\,\mathrm{Var}[X_{n-2}]) \\
&= 1 + \alpha^2(1 + \alpha^2(1 + \alpha^2\,\mathrm{Var}[X_{n-2}])) \\
&= 1 + \alpha^2 + \cdots + \alpha^{2n} \\
&= \sum_{k=0}^{n} \alpha^{2k} \\
&= \begin{cases} \dfrac{1 - \alpha^{2n+2}}{1 - \alpha^2}, & \alpha \neq \pm 1, \\[2ex] n + 1, & \alpha = \pm 1, \qquad n \geqslant 0. \end{cases}
\end{aligned}$$

d) We check that the solution of $\phi(z) := \alpha z$ is $z = 1/\alpha$, hence by Theorem 2.12 there exists an AR(1) solution of (2.28) which is weakly stationary when $\alpha \neq \pm 1$. However, the present time series $(X_n)_{n\geqslant 0}$ started at $X_0 = 0$ is *not* weakly stationary because $\mathrm{Cov}(X_n, X_n) = \mathbb{E}[X_n^2]$ is not constant in $n \geqslant 0$.

Exercise 2.4

a) We have

$$
\begin{aligned}
\mathrm{Var}[X_n] &= \mathrm{Var}[Z_{n-1} - Z_{n-2} + \alpha Z_{n-3}] \\
&= \mathrm{Var}[Z_{n-1}] + \mathrm{Var}[-Z_{n-2}] + \mathrm{Var}[\alpha Z_{n-3}] \\
&= \mathrm{Var}[Z_{n-1}] + \mathrm{Var}[Z_{n-2}] + \alpha^2 \mathrm{Var}[Z_{n-3}] \\
&= 2 + \alpha^2.
\end{aligned}
$$

Next, since using the linearity relation

$$
\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)
$$

and the fact that $\mathrm{Cov}(X, Z) = 0$ when X and Z are independent random variables, we have

$$
\begin{aligned}
\mathrm{Cov}(X_{n+1}, X_n) &= \mathrm{Cov}(Z_n - Z_{n-1} + \alpha Z_{n-2}, Z_{n-1} - Z_{n-2} + \alpha Z_{n-3}) \\
&= \mathrm{Cov}(Z_n - Z_{n-1} + \alpha Z_{n-2}, Z_{n-1}) \\
&\quad + \mathrm{Cov}(Z_n - Z_{n-1} + \alpha Z_{n-2}, -Z_{n-2}) \\
&\quad + \mathrm{Cov}(Z_n - Z_{n-1} + \alpha Z_{n-2}, \alpha Z_{n-3}) \\
&= \mathrm{Cov}(-Z_{n-1}, Z_{n-1}) + \mathrm{Cov}(\alpha Z_{n-2}, -Z_{n-2}) \\
&= -\mathrm{Cov}(Z_{n-1}, Z_{n-1}) - \alpha \mathrm{Cov}(Z_{n-2}, Z_{n-2}) \\
&= -\alpha - 1,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(X_{n+2}, X_n) &= \mathrm{Cov}(Z_{n+2} - Z_n + \alpha Z_{n-1}, Z_{n-1} - Z_{n-2} + \alpha Z_{n-3}) \\
&= \mathrm{Cov}(Z_{n+2} - Z_n + \alpha Z_{n-1}, Z_{n-1}) \\
&\quad + \mathrm{Cov}(Z_{n+2} - Z_n + \alpha Z_{n-1}, -Z_{n-2}) \\
&\quad + \mathrm{Cov}(Z_{n+2} - Z_n + \alpha Z_{n-1}, \alpha Z_{n-3}) \\
&= \mathrm{Cov}(\alpha Z_{n-1}, Z_{n-1}) \\
&= \alpha \mathrm{Cov}(Z_{n-1}, Z_{n-1}) \\
&= \alpha,
\end{aligned}
$$

and

$$\begin{aligned}\mathrm{Cov}(X_{n+k}, X_n) &= \mathrm{Cov}(Z_{n+k-1} - Z_{n+k-2} + \alpha Z_{n+k-3}, Z_{n-1} - Z_{n-2} + \alpha Z_{n-3}) \\ &= \mathrm{Cov}(Z_{n+k-1} - Z_{n+k-2} + \alpha Z_{n+k-3}, Z_{n-1}) \\ &\quad + \mathrm{Cov}(Z_{n+k-1} - Z_{n+k-2} + \alpha Z_{n+k-3}, -Z_{n-2}) \\ &\quad + \mathrm{Cov}(Z_{n+k-1} - Z_{n+k-2} + \alpha Z_{n+k-3}, \alpha Z_{n-3}) \\ &= 0\end{aligned}$$

for $k \geqslant 3$.

b) Since the white noise sequence $(Z_n)_{n\in\mathbb{Z}}$ is made of independent identically distributed random variables, we have the identity in distribution

$$X_n = Z_{n-1} - Z_{n-2} + \alpha Z_{n-3} \overset{d}{\simeq} Z_n - Z_{n-1} + \alpha Z_{n-2}, \qquad n \geqslant 2,$$

which shows that $(X_n)_{n\geqslant 3}$ has the same distribution as an MA(2) time series the form

$$Y_n = Z_n + \beta_1 Z_{n-1} + \beta_2 Z_{n-2},$$

with $\beta_1 = -1$ and $\beta_1 = \alpha$.

Exercise 2.5

a) We have

$$\begin{aligned}\nabla X_n &= X_n - X_{n-1} \\ &= Z_n + \alpha_1 X_{n-1} - Z_{n-1} - \alpha_1 X_{n-2} \\ &= Z_n - Z_{n-1} + \alpha_1 \nabla X_{n-1}, \qquad n \geqslant 2,\end{aligned}$$

hence $(\nabla X_n)_{n\geqslant 2}$ forms an ARMA$(1,1)$ time series.

b) We have

$$\begin{aligned}\nabla^2 X_n &= \nabla X_n - \nabla X_{n-1} \\ &= X_n - X_{n-1} - (X_{n-1} - X_{n-2}) \\ &= X_n - 2X_{n-1} + X_{n-2} \\ &= Z_n + \alpha_1 X_{n-1} - 2Z_{n-1} - 2\alpha_1 X_{n-2} + Z_{n-2} + \alpha_1 X_{n-3} \\ &= Z_n - 2Z_{n-1} + Z_{n-2} + \alpha_1 \nabla^2 X_{n-1}, \qquad n \geqslant 3,\end{aligned}$$

which forms an ARMA$(1,2)$ time series.

Exercise 2.6

a) We have

$$\frac{\partial}{\partial a} \sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big)^2 = -2 \left(\sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big) \right)$$
$$= 2an - 2 \sum_{k=1}^{n} r_k^{(2)} + 2b \sum_{k=1}^{n} r_k^{(1)},$$

and

$$\frac{\partial}{\partial b} \sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big)^2 = 2 \sum_{k=1}^{n} r_k^{(1)} \big(- a + r_k^{(2)} - b r_k^{(1)}\big)$$
$$= 2 \sum_{k=1}^{n} r_k^{(1)} \left(r_k^{(2)} - b r_k^{(1)} - \frac{1}{n} \sum_{l=1}^{n} \big(r_l^{(2)} - b r_l^{(1)}\big) \right)$$
$$= 2 \sum_{k=1}^{n} r_k^{(1)} r_k^{(2)} - \frac{2}{n} \sum_{k,l=1}^{n} r_k^{(1)} r_l^{(2)} - 2b \left(\sum_{k=1}^{n} (r_k^{(1)})^2 - \frac{1}{n} \sum_{k,l=1}^{n} r_k^{(1)} r_l^{(1)} \right).$$

b) In order to minimize the residual (2.29) over a and b we equate the above derivatives to zero, which yields the equations

$$\frac{\partial}{\partial a} \sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big)^2_{|a=\hat{a},\ b=\hat{b}} = 2\widehat{a}n - 2 \sum_{k=1}^{n} r_k^{(2)} + 2\widehat{b} \sum_{k=1}^{n} r_k^{(1)} = 0$$

and

$$\frac{\partial}{\partial b} \sum_{k=1}^{n} \big(r_k^{(2)} - a - b r_k^{(1)}\big)^2_{|a=\hat{a},\ b=\hat{b}}$$
$$= 2 \sum_{k=1}^{n} r_k^{(1)} r_k^{(2)} - \frac{2}{n} \sum_{k,l=1}^{n} r_k^{(1)} r_l^{(2)} - 2\widehat{b} \left(\sum_{k=1}^{n} (r_k^{(1)})^2 - \frac{1}{n} \sum_{k,l=1}^{n} r_k^{(1)} r_l^{(1)} \right)$$
$$= 0.$$

This leads to estimators $\widehat{a}$, $\widehat{b}$ of the parameters a and b respectively as the empirical mean and covariance of $(r_k^{(1)})_{k=1,2,\ldots,n}$, *i.e.*

$$\begin{cases} \widehat{a} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\left(r_k^{(2)} - \widehat{b}r_k^{(1)}\right), \\ \\ \text{and} \\ \\ \widehat{b} = \dfrac{\displaystyle\sum_{k=1}^{n} r_k^{(1)} r_k^{(2)} - \frac{1}{n}\sum_{k,l=0}^{n} r_k^{(1)} r_l^{(2)}}{\displaystyle\sum_{k=1}^{n} (r_k^{(1)})^2 - \frac{1}{n}\sum_{k,l=0}^{n} r_k^{(1)} r_l^{(1)}} = \dfrac{\displaystyle\sum_{k=1}^{n}\left(r_k^{(1)} - \frac{1}{n}\sum_{l=0}^{n} r_l^{(1)}\right)\left(r_k^{(2)} - \frac{1}{n}\sum_{l=0}^{n} r_l^{(2)}\right)}{\displaystyle\sum_{k=1}^{n}\left(r_k^{(1)} - \frac{1}{n}\sum_{k=1}^{n} r_k^{(1)}\right)^2}. \end{cases}$$

Exercise 2.7 Since the p-value $= 0.02377$ is lower than the 5% confidence level, we can reject the nonstationarity (null) hypothesis H_0 at that level.

Exercise 2.8

a) We consider the equation

$$\varphi(z) = \alpha_1 z + \alpha_2 z^2 = 1,$$

i.e.

$$\alpha_2 z^2 + \alpha_1 z - 1 = 0,$$

with solutions

$$z_\pm = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2\alpha_2} = \frac{-a \pm \sqrt{a^2 + 8a^2}}{4a^2} = \frac{-a \pm 3a}{4a^2} = \begin{cases} \dfrac{1}{2a} \\ -\dfrac{1}{a}, \end{cases}$$

hence by Theorem 2.12 the time series $(X_n)_{n\geqslant 1}$ is stationary for $a \notin \{-1, -1/2, 1/2, 1\}$.

b) We have

$$\begin{aligned} \mathbb{E}[X_n] &= \mathbb{E}[Z_n + \alpha_1 X_{n-1} + \alpha_2 X_{n-2}] \\ &= \mathbb{E}[Z_n] + \alpha_1 \mathbb{E}[X_{n-1}] + \alpha_2 \mathbb{E}[X_{n-2}] \\ &= \alpha_1 \mathbb{E}[X_{n-1}] + \alpha_2 \mathbb{E}[X_{n-2}] \\ &= \alpha_1 \mathbb{E}[X_n] + \alpha_2 \mathbb{E}[X_n], \end{aligned}$$

hence

$$(1 - \alpha_1 - \alpha_2)\mathbb{E}[X_n] = 0,$$

which implies $\mathbb{E}[X_n] = 0$, $n \in \mathbb{Z}$, since $1 - \alpha_1 - \alpha_2 \neq 0$.

c) We have

$$\begin{aligned}\operatorname{Cov}(X_n, Z_n) &= \operatorname{Cov}(Z_n + \alpha_1 X_{n-1} + \alpha_2 X_{n-2}, Z_n)\\ &= \operatorname{Cov}(Z_n, Z_n) + \alpha_1 \operatorname{Cov}(Z_n, X_{n-1}) + \alpha_2 \operatorname{Cov}(Z_n, X_{n-2})\\ &= \operatorname{Cov}(Z_n, Z_n)\\ &= \sigma^2.\end{aligned}$$

d) We have

$$\begin{aligned}\operatorname{Cov}(X_{n+1}, X_n) &= \operatorname{Cov}(Z_{n+1} + \alpha_1 X_n + \alpha_2 X_{n-1}, X_n)\\ &= \operatorname{Cov}(Z_{n+1}, X_n) + \alpha_1 \operatorname{Cov}(X_n, X_n) + \alpha_2 \operatorname{Cov}(X_{n-1}, X_n),\\ &= 16\alpha_1 + \alpha_2 \operatorname{Cov}(X_{n-1}, X_n)\\ &= 4 + \frac{1}{2} \operatorname{Cov}(X_{n-1}, X_n),\end{aligned}$$

hence

$$\operatorname{Cov}(X_{n+1}, X_n) = 8, \qquad n \in \mathbb{Z}.$$

Chapter 3

Exercise 3.1

a) Since $Z_1 + Z_2 + \cdots + Z_n$ has the centered Gaussian $\mathcal{N}(0, n\sigma^2)$ distribution with variance $n\sigma^2$, we have

$$\begin{aligned}\mathbb{P}(Y \geqslant y) &= \sum_{n\geqslant 1} \mathbb{P}\left(\sum_{k=1}^N Z_k \geqslant y \;\Big|\; N = n \right) \mathbb{P}(N = n)\\ &= \sum_{n\geqslant 1} \left(1 - \mathbb{P}\left(\sum_{k=1}^n Z_k < y \;\Big|\; N = n \right)\right) \mathbb{P}(N = n)\\ &= \sum_{n\geqslant 1} \left(1 - \mathbb{P}\left(\sum_{k=1}^n Z_k \leqslant y \;\Big|\; N = n \right)\right) \mathbb{P}(N = n)\\ &= \sum_{n\geqslant 1} \left(1 - \Phi\left(\frac{y}{\sqrt{n\sigma^2}} \right)\right) \mathbb{P}(N = n)\\ &= \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} \frac{\lambda^n}{n!} \Phi\left(-\frac{y}{\sqrt{n\sigma^2}} \right), \qquad y > 0.\end{aligned}$$

b) Since $\mathbb{E}[Z_k] = 0$ for all $k \geqslant 1$, we have

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{n\geqslant 1} \mathbb{E}\left[\sum_{k=1}^{N} Z_k \;\middle|\; N=n\right] \mathbb{P}(N=n) \\
&= \sum_{n\geqslant 1} \mathbb{E}\left[\sum_{k=1}^{n} Z_k \;\middle|\; N=n\right] \mathbb{P}(N=n) \\
&= \sum_{n\geqslant 1} \mathbb{P}(N=n) \sum_{k=1}^{n} \mathbb{E}[Z_k \mid N=n] \\
&= \sum_{n\geqslant 1} \mathbb{P}(N=n) \sum_{k=1}^{n} \mathbb{E}[Z_k] \\
&= 0,
\end{aligned}
$$

as in (3.12).

Exercise 3.2 By (3.17), we have

$$
\begin{aligned}
\Phi'(y) &= \frac{\lambda}{c}\Phi(y) - \frac{\lambda}{c}\int_0^y \Phi(y-z)dF(z) \\
&= \frac{\lambda}{c}\Phi(y) - \frac{\lambda}{\mu c}\int_0^y \Phi(y-z)\mathrm{e}^{-z/\mu}dz \\
&= \frac{\lambda}{c}\Phi(y) - \frac{\lambda}{\mu c}\int_0^y \Phi(z)\mathrm{e}^{-(y-z)/\mu}dz,
\end{aligned}
$$

hence the differential equation

$$
\begin{aligned}
\Phi''(y) &= \frac{\lambda}{c}\Phi'(y) - \frac{\lambda}{\mu c}\Phi(y) + \frac{\lambda}{\mu^2 c}\int_0^y \Phi(z)\mathrm{e}^{-(y-z)/\mu}dz \\
&= \frac{\lambda}{c}\Phi'(y) - \frac{\lambda}{\mu c}\Phi(y) + \frac{1}{\mu}\left(\frac{\lambda}{c}\Phi(y) - \Phi'(y)\right) \\
&= \left(\frac{\lambda}{c} - \frac{1}{\mu}\right)\Phi'(y),
\end{aligned}
$$

which can be solved as

$$
\Phi(y) = 1 - \frac{\lambda\mu}{c}\mathrm{e}^{(\lambda/c-1/\mu)y},
$$

given the boundary conditions $\Phi(\infty) = 1$ and $\Phi(0) = 1 - \lambda\mu/c$, cf. (3.19). We conclude that

$$
\Psi(y) = \frac{\lambda\mu}{c}\mathrm{e}^{(\lambda/c-1/\mu)y}, \qquad y \geqslant 0,
$$

provided that $c < \lambda\mu$.

Exercise 3.3

a) We have

$$\begin{aligned}\mathbb{E}[R_T] &= \mathbb{E}[R_0 + \mu T - CN_T] \\ &= \mathbb{E}[R_0] + \mathbb{E}[\mu T] - \mathbb{E}[CN_T] \\ &= R_0 + \mu T - C\mathbb{E}[N_T] \\ &= R_0 + \mu T - C\lambda T \\ &= R_0 + (\mu - \lambda C)T,\end{aligned}$$

and similarly

$$\mathrm{Var}[R_T] = \mathrm{Var}[\mu T - CN_T] = \mathrm{Var}[-CN_T] = C^2\,\mathrm{Var}[N_T] = \lambda C^2 T.$$

b) We find

$$\begin{aligned}\mathbb{P}(R_T < 0) &= \mathbb{P}(R_0 + \mu T - CN_T < 0) \\ &= \mathbb{P}(N_T > (R_0 + \mu T)/C) \\ &= \sum_{k > (R_0+\mu T)/C} \mathbb{P}(N_T = k) \\ &= \mathrm{e}^{-\lambda T} \sum_{k > (R_0+\mu T)/C} \frac{(\lambda T)^k}{k!}.\end{aligned}$$

Exercise 3.4

a) We have $\mathbb{E}[S(T)] = \lambda T \mathbb{E}[Z]$ and $\mathrm{Var}[S(T)] = \lambda T \mathbb{E}[Z^2]$.
b) We have

$$\begin{aligned}\mathbb{P}(x + f(T) - S(T) < 0) &\leqslant \frac{\mathrm{Var}[x + f(T) - S(T)]}{(\mathbb{E}[x + f(T) - S(T)])^2} \\ &= \frac{\mathrm{Var}[S(T)]}{(x + f(T) - \mathbb{E}[S(T)])^2} \\ &= \frac{\lambda T \mathbb{E}[Z_1^2]}{(x + f(T) - \lambda T \mathbb{E}[Z_1])^2}.\end{aligned}$$

Chapter 4

Exercise 4.1

a) Taking $(U,V)=(U,U)$, we have

$$\begin{aligned}\mathbb{P}(U\leqslant u \text{ and } V\leqslant v) &= \mathbb{P}(U\leqslant u \text{ and } U\leqslant v)\\ &= \mathbb{P}(U\leqslant \min(u,v))\\ &= \min(u,v)\\ &= C_M(u,v), \qquad u,v\in[0,1].\end{aligned}$$

b) Taking $(U,V)=(U,1-U)$, we have

$$\begin{aligned}\mathbb{P}(U\leqslant u \text{ and } V\leqslant v) &= \mathbb{P}(U\leqslant u \text{ and } 1-U\leqslant v)\\ &= \mathbb{P}(U\leqslant u \text{ and } U\geqslant 1-v)\\ &= \mathbb{P}(1-v\leqslant U\leqslant u)\\ &= \mathbb{1}_{\{0\leqslant 1-v\leqslant u\leqslant 1\}}\mathbb{P}(1-v\leqslant U\leqslant u)\\ &= \mathbb{1}_{\{0\leqslant u+v-1\leqslant 1\}}(u-(1-v))\\ &= (u+v-1)^+,\end{aligned}$$

$u,v\in[0,1]$.

c) We have

$$C(u,v)=\mathbb{P}(U\leqslant u \text{ and } V\leqslant v)\leqslant \mathbb{P}(U\leqslant u \text{ and } V\geqslant 1)\leqslant \mathbb{P}(U\leqslant u)=u,$$

$u,v\in[0,1]$, and similarly we find $C(u,v)\leqslant \mathbb{P}(U\leqslant v)=v$ for all $u,v\in[0,1]$, which yields (4.8).

d) For fixed $v\in[0,1]$ we have

$$\begin{aligned}\frac{\partial C}{\partial u}(u,v) &= \lim_{\varepsilon\to 0}\frac{C(u+\varepsilon,v)-C(u,v)}{\varepsilon}\\ &= \lim_{\varepsilon\to 0}\frac{\mathbb{P}(U\leqslant u+\varepsilon \text{ and } V\leqslant v)-\mathbb{P}(U\leqslant u \text{ and } V\leqslant v)}{\varepsilon}\\ &= \lim_{\varepsilon\to 0}\frac{\mathbb{P}(u\leqslant U\leqslant u+\varepsilon \text{ and } V\leqslant v)}{P(u\leqslant U\leqslant u+\varepsilon)}\\ &= \lim_{\varepsilon\to 0}\mathbb{P}(V\leqslant v \mid u\leqslant U\leqslant u+\varepsilon)\\ &= \mathbb{P}(V\leqslant v\mid U=u)\\ &\leqslant 1,\end{aligned}$$

$u,v\in[0,1]$, hence

$$h'(u)=\frac{\partial C}{\partial u}(u,v)-1=\mathbb{P}(V\leqslant v\mid U=u)-1\leqslant 0,$$

$u, v \in [0,1]$, and since $h(1) = C(1,v) - v = \mathbb{P}(V \leqslant v) - v = 0$, $v \in [0,1]$ we conclude that $h(u) \geqslant 0$, $u \in [0,1]$, which shows (4.9).

Exercise 4.2

a) When $\rho = 1$, we have

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = p_X p_Y + \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = (1-p_X) p_Y - \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = p_X (1-p_Y) - \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = (1-p_X)(1-p_Y) + \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \end{cases}$$

hence

$$\begin{cases} (1-p_X) p_Y \geqslant \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ p_X (1-p_Y) \geqslant \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \end{cases}$$

hence

$$(1-p_X) p_Y \geqslant p_X (1-p_Y) \quad \text{and} \quad p_X (1-p_Y) \geqslant p_Y (1-p_X),$$

showing that $(1-p_X)p_Y = p_X(1-p_Y)$, which implies $p_X = p_Y$, and

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = p_X^2 + p_X(1-p_X) = p_X = p_Y, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = 0, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = 1 - p_X = 1 - p_Y. \end{cases}$$

b) When $\rho = -1$, we have

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = p_X p_Y - \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = (1-p_X) p_Y + \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = p_X (1-p_Y) + \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = (1-p_X)(1-p_Y) - \sqrt{p_X p_Y (1-p_X)(1-p_Y)} \geqslant 0, \end{cases}$$

hence

$$\begin{cases} p_X p_Y \geqslant \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \\ \\ p_X p_Y \geqslant \sqrt{p_X p_Y (1-p_X)(1-p_Y)}, \end{cases}$$

hence

$$p_X p_Y \geqslant (1-p_X)(1-p_Y) \quad \text{and} \quad p_X p_Y \geqslant (1-p_X)(1-p_Y),$$

showing that $p_X p_Y = (1-p_X)(1-p_Y)$, which implies $p_X = 1 - p_Y$, and

$$\begin{cases} \mathbb{P}(X = 1 \text{ and } Y = 1) = 0, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 1) = 1, \\ \\ \mathbb{P}(X = 1 \text{ and } Y = 0) = 1, \\ \\ \mathbb{P}(X = 0 \text{ and } Y = 0) = 0. \end{cases}$$

Exercise 4.3

a) We have

$$\mathbb{P}(X \geqslant x) = \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0) = \mathrm{e}^{-(\lambda+\nu)x},$$

and

$$\mathbb{P}(Y \geqslant y) = \mathbb{P}(X \geqslant 0 \text{ and } Y \geqslant y) := \mathrm{e}^{-(\mu+\nu)y},$$

$x, y \geqslant 0$, *i.e.* X and Y are exponentially distributed with respective parameters $\lambda + \nu$ and $\mu + \nu$.

b) We have

$$\begin{aligned} &\mathbb{P}(X \leqslant x \text{ and } Y \leqslant 0) \\ &= \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0) - (\mathbb{P}(X \geqslant x) - \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0)) \\ &\quad -(\mathbb{P}(Y \geqslant x) - \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0)) \\ &= \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0) - \mathbb{P}(X \geqslant x) - \mathbb{P}(Y \geqslant x) + \mathbb{P}(X \geqslant x \text{ and } Y \geqslant 0)), \end{aligned}$$

$x, y \geqslant 0$, *i.e.* X and Y are exponentially distributed with respective parameters $\lambda + \nu$ and $\mu + \nu$.

c) Since $\mathrm{e}^{-(\lambda+\nu)X}$ and $\mathrm{e}^{-(\mu+\nu)Y}$ are uniformly distributed on $[0,1]$, a copula function $C(u,v)$ can be defined by

$$\begin{aligned} C(u,v) &:= \mathbb{P}(\mathrm{e}^{-(\lambda+\nu)X} \leqslant u \text{ and } \mathrm{e}^{-(\mu+\nu)Y} \leqslant v) \\ &= \mathbb{P}(X \leqslant -(\lambda+\nu)^{-1} \log u \text{ and } Y \leqslant -(\lambda+\nu)^{-1} \log v) \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{e}^{\lambda(\lambda+\nu)^{-1}\log u+\mu(\lambda+\nu)^{-1}\log v)y-\nu\,\mathrm{Max}(-(\lambda+\nu)^{-1}\log u,-(\lambda+\nu)^{-1}\log v))}\\
&= u^{\lambda/(\lambda+\nu)}v^{\mu/(\lambda+\nu)}\mathrm{e}^{-\nu\,\mathrm{Max}(-(\lambda+\nu)^{-1}\log u,-(\lambda+\nu)^{-1}\log v))}\\
&= u^{\lambda/(\lambda+\nu)}v^{\mu/(\lambda+\nu)}\mathrm{e}^{\nu\min(\log u^{(\lambda+\nu)^{-1}},\log v^{(\lambda+\nu)^{-1}}))}\\
&= u^{\lambda/(\lambda+\nu)}v^{\mu/(\lambda+\nu)}\mathrm{e}^{\log\min(u^{\nu/(\lambda+\nu)},v^{\nu/(\lambda+\nu)}))}\\
&= u^{\lambda/(\lambda+\nu)}v^{\mu/(\lambda+\nu)}\min(u^{\nu/(\lambda+\nu)},v^{\nu/(\lambda+\nu)}))\\
&= u^{\lambda/(\lambda+\nu)}v^{\mu/(\lambda+\nu)}\left(\min(u,v)\right)^{\nu/(\lambda+\nu)}, \qquad x,y\geqslant 0.
\end{aligned}
$$

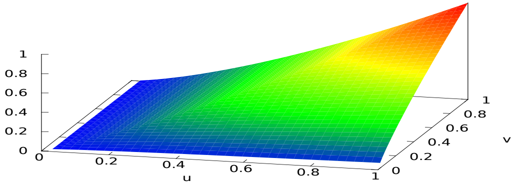


Fig. S.2: Exponential copula function $u,v\mapsto C(u,b)$ with $\lambda=1$, $\mu=2$, $\nu=4$.

Exercise 4.4

a) We have

$$
F_X(x)=\mathbb{P}(X\leqslant x)=\mathbb{P}(X\leqslant x \text{ and } Y\leqslant\infty)=\frac{1}{1+\mathrm{e}^{-x}}
$$

and

$$
F_Y(y)=\mathbb{P}(Y\leqslant y)=\mathbb{P}(X\leqslant\infty \text{ and } Y\leqslant y)=\frac{1}{1+\mathrm{e}^{-y}}, \qquad x,y\in\mathbb{R}.
$$

The probability densities are given by

$$
f_X(x)=f_Y(x)=F_X'(x)=F_Y'(x)=\frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2}, \qquad x\in\mathbb{R}.
$$

b) We have

$$
F_X^{-1}(u)=F_Y^{-1}(u)=-\log\frac{1-u}{u}, \qquad u\in(0,1),
$$

and the corresponding copula is given by

$$
\begin{aligned}
C(u,v)&=F_{(X,Y)}(F_X^{-1}(u),F_Y^{-1}(v))\\
&=F_{(X,Y)}\left(-\log\frac{1-u}{u},-\log\frac{1-v}{v}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{1+(1-u)/u+(1-v)/v} \\
&= \frac{1}{1+(1-u)/u+(1-v)/v} \\
&= \frac{uv}{u+v-uv}, \qquad u,v\in[0,1],
\end{aligned}
$$

which is a particular case of the Ali-Mikhail-Haq copula.

Exercise 4.5

a) We show that (X,Y) have Gaussian marginals $\mathcal{N}(0,\sigma^2)$ and $\mathcal{N}(0,\eta^2)$, according to the following computation:

$$
\begin{aligned}
\int_{-\infty}^{\infty}\widetilde{f}(x,y)dy &= \frac{1}{\pi\sigma\eta}\int_{-\infty}^{\infty}\mathbb{1}_{\mathbb{R}_-^2\cup\mathbb{R}_+^2}(x,y)\mathrm{e}^{-x^2/(2\sigma^2)-y^2/(2\eta^2)}dy \\
&= \frac{1}{\pi\sigma\eta}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_+}(x)\int_0^{\infty}\mathrm{e}^{-y^2/(2\eta^2)}dy + \\
&\quad \frac{1}{\pi\sigma\eta}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_-}(x)\int_{-\infty}^{0}\mathrm{e}^{-y^2/(2\eta^2)}dy \\
&= \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_+}(x) + \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_-}(x) \\
&= \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-x^2/(2\sigma^2)}, \qquad x\in\mathbb{R}.
\end{aligned}
$$

b) The couple (X,Y) does *not* have a joint Gaussian distribution, and its joint probability density function does *not* coincide with $f_\Sigma(x,y)$.

c) When $\sigma=\eta=1$, the random variable $X+Y$ has the probability density function

$$
\begin{aligned}
&\frac{\partial}{\partial a}\mathbb{P}(X+Y\leqslant a) = \frac{\partial}{\partial a}\int_{-\infty}^{\infty}\int_{-\infty}^{a-x}\widetilde{f}(x,y)dydx \\
&= \frac{1}{\pi}\frac{\partial}{\partial a}\int_0^a\int_0^{a-x}\mathrm{e}^{-x^2/2-y^2/2}dydx \\
&= \frac{1}{\pi}\frac{\partial}{\partial a}\int_0^a\mathrm{e}^{-(a-z)^2/2}\int_0^z\mathrm{e}^{-y^2/2}dydz \\
&= \frac{1}{\pi}\int_0^a\mathrm{e}^{-y^2/2}dy - \frac{1}{\pi}\int_0^a(a-z)\mathrm{e}^{-(a-z)^2/2}\int_0^z\mathrm{e}^{-y^2/2}dydz \\
&= \frac{1}{\pi}\int_0^a\mathrm{e}^{-y^2/2}dy - \frac{1}{\pi}\int_0^a(a-z)\mathrm{e}^{-(a-z)^2/2}dz\int_0^a\mathrm{e}^{-y^2/2}dy \\
&\quad +\frac{1}{\pi}\int_0^a\mathrm{e}^{-y^2/2}\int_0^y\mathrm{e}^{-(a-z)^2/2}dzdy
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\pi}\mathrm{e}^{-a^2/2}\int_0^a \mathrm{e}^{-y^2/2}dy + \frac{1}{\pi}\int_0^a \mathrm{e}^{-y^2/2}(\mathrm{e}^{-(a-y)^2/2} - \mathrm{e}^{-a^2/2})dy \\
&= \frac{1}{\pi}\mathrm{e}^{-a^2/2}\int_0^a \mathrm{e}^{-y^2/2-(a-y)^2/2}dy \\
&= \frac{1}{\pi}\int_0^a \mathrm{e}^{-((\sqrt{2}y-a/\sqrt{2})^2-a^2/2)/2}dy \\
&= \frac{\mathrm{e}^{-a^2/4}}{\pi\sqrt{2}}\int_0^{a\sqrt{2}} \mathrm{e}^{-((y-a/\sqrt{2})^2)/2}dy \\
&= \frac{\mathrm{e}^{-a^2/4}}{\pi\sqrt{2}}\int_{-a/\sqrt{2}}^{a(\sqrt{2}-1/\sqrt{2})} \mathrm{e}^{-y^2/2}dy \\
&= \frac{\mathrm{e}^{-a^2/4}}{\sqrt{\pi}\sqrt{2\pi}}\int_{-a/\sqrt{2}}^{a/\sqrt{2}} \mathrm{e}^{-y^2/2}dy \\
&= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-a^2/4}\frac{1}{\sqrt{\pi}}\big(2\Phi\big(a/\sqrt{2}\big)-1\big), \quad a \geqslant 0,
\end{aligned}
$$

which vanishes at $a = 0$.

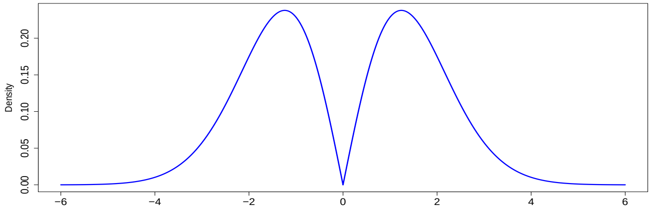


Fig. S.3: Density of $X + Y$.

d) The random variables X and Y are positively correlated, as

$$
\begin{aligned}
\int_{-\infty}^{\infty} y f_\Sigma(x,y)dy &= \frac{1}{\pi\sigma\eta}\int_{-\infty}^{\infty} \mathbb{1}_{\mathbb{R}_-^2\cup\mathbb{R}_+^2}(x,y)y\mathrm{e}^{-x^2/(2\sigma^2)-y^2/(2\eta^2)}dy \\
&= \frac{1}{\pi\sigma\eta}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_+}(x)\int_0^{\infty} y\mathrm{e}^{-y^2/(2\eta^2)}dy \\
&\quad + \frac{1}{\pi\sigma\eta}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_-}(x)\int_{-\infty}^{0} y\mathrm{e}^{-y^2/(2\eta^2)}dy \\
&= \frac{\eta}{\pi\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_+}(x) - \frac{\eta}{\pi\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}\mathbb{1}_{\mathbb{R}_-}(x),
\end{aligned}
$$

hence

$$
\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_\Sigma(x,y)dydx
$$

$$= \frac{\eta}{\pi\sigma}\int_0^\infty x\mathrm{e}^{-x^2/(2\sigma^2)}dx - \frac{\eta}{\pi\sigma}\int_{-\infty}^0 x\mathrm{e}^{-x^2/(2\sigma^2)}dx$$
$$= \frac{2\sigma\eta}{\pi},$$

and

$$\rho = \frac{\mathbb{E}[XY]}{\sigma\eta} = \frac{2}{\pi}.$$

Under a rotation

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

of angle $\theta \in [0, 2\pi]$ we would find

$$\begin{aligned}
&\mathbb{E}[(X\cos\theta - Y\sin\theta)(X\sin\theta + Y\cos\theta)] \\
&= \sin\theta\cos\theta\mathbb{E}\big[X^2\big] + (\cos^2\theta - \sin^2\theta)\mathbb{E}[XY] - \sin\theta\cos\theta\mathbb{E}\big[Y^2\big] \\
&= \sigma^2\sin\theta\cos\theta + (\cos^2\theta - \sin^2\theta)\frac{2\sigma\eta}{\pi} - \eta^2\sin\theta\cos\theta \\
&= \frac{\sigma^2}{2}\sin(2\theta) + \cos(2\theta)\frac{2\sigma\eta}{\pi} - \frac{\eta^2}{2}\sin(2\theta),
\end{aligned}$$

and

$$\rho = \frac{\sigma}{2\eta}\sin(2\theta) + \cos(2\theta)\frac{2}{\pi} - \frac{\eta}{2\sigma}\sin(2\theta),$$

i.e. $\theta = \pi/4$ and $\sigma = \eta$ would lead to uncorrelated random variables.

Exercise 4.6

a) We have

$$\begin{aligned}
\mathbb{P}(\tau_i \wedge \tau \geqslant s) &= \mathbb{P}(\tau_i \geqslant s \text{ and } \tau \geqslant s) \\
&= \mathbb{P}(\tau_i \geqslant s)\mathbb{P}(\tau \geqslant s) \\
&= \mathrm{e}^{-\lambda_i s}\mathrm{e}^{-\lambda s} \\
&= \mathrm{e}^{-(\lambda_i+\lambda)s}, \qquad s \geqslant 0,
\end{aligned}$$

hence $\tau_i \wedge \tau$ is an exponentially distributed random variable with parameter $\lambda_i + \lambda$, $i = 1, 2$.

b) Next, we have

$$\begin{aligned}
\mathbb{P}(\tau_1 \wedge \tau > s \text{ and } \tau_2 \wedge \tau > t) &= \mathbb{P}(\tau_1 > s \text{ and } \tau > s \text{ and } \tau_2 > t \text{ and } \tau > t) \\
&= \mathbb{P}(\tau_1 > s \text{ and } \tau_2 > t \text{ and } \tau > \mathrm{Max}(s,t)) \\
&= \mathbb{P}(\tau_1 > s)\mathbb{P}(\tau_2 > t)\mathbb{P}(\tau > \mathrm{Max}(s,t)) \\
&= \mathrm{e}^{-\lambda_1 s}\mathrm{e}^{-\lambda_2 t}\mathrm{e}^{-\lambda\,\mathrm{Max}(s,t)}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{e}^{-\lambda_1 s-\lambda_2 t-\lambda \operatorname{Max}(s,t)} \\
&= \mathrm{e}^{-(\lambda_1+\lambda)s-(\lambda_2+\lambda)t+\lambda \min(s,t)} \\
&= (1-F_X(s))(1-F_Y(t)) \min(\mathrm{e}^{\lambda s}, \mathrm{e}^{\lambda t}),
\end{aligned}
$$

$s,t \geqslant 0$.

c) We have

$$
\begin{aligned}
F_{X,Y}(s,t) &= \mathbb{P}(\tau_1 \wedge \tau \leqslant s \text{ and } \tau_2 \wedge \tau \leqslant t) \\
&= \mathbb{P}(\tau_1 \wedge \tau \leqslant s) - \mathbb{P}(\tau_1 \wedge \tau \leqslant s \text{ and } \tau_2 \wedge \tau > t) \\
&= \mathbb{P}(\tau_1 \wedge \tau \leqslant s) - \big(\mathbb{P}(\tau_2 \wedge \tau > t) - \mathbb{P}(\tau_1 \wedge \tau > s \text{ and } \tau_2 \wedge \tau > t)\big) \\
&= \mathbb{P}(\tau_1 \wedge \tau \leqslant s) + \mathbb{P}(\tau_2 \wedge \tau \leqslant t) + \mathbb{P}(\tau_1 \wedge \tau > s \text{ and } \tau_2 \wedge \tau > t) - 1 \\
&= F_X(s) + F_Y(t) + (1-F_X(s))(1-F_Y(t)) \min(\mathrm{e}^{\lambda s}, \mathrm{e}^{\lambda t}) - 1.
\end{aligned}
$$

d) We find

$$
\begin{aligned}
C(u,v) &= F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)) \\
&= F_X(F_X^{-1}(u)) + F_Y(F_X^{-1}(v)) \\
&\quad +(1-F_X(F_X^{-1}(u)))(1-F_Y(F_X^{-1}(v))) \min\big(\mathrm{e}^{\lambda F_X^{-1}(u)}, \mathrm{e}^{\lambda F_X^{-1}(v)}\big) - 1 \\
&= u+v-1+(1-u)(1-v) \min\big(\mathrm{e}^{\lambda F_X^{-1}(u)}, \mathrm{e}^{\lambda F_Y^{-1}(v)}\big) \\
&= u+v-1+(1-u)(1-v) \min\big(\mathrm{e}^{-\lambda \log(1-u)/(\lambda_1+\lambda)}, \mathrm{e}^{-\lambda \log(1-v)/(\lambda_2+\lambda)}\big) \\
&= u+v-1+\min\big((1-v)(1-u)^{1-\lambda/(\lambda_1+\lambda)}, (1-u)(1-v)^{1-\lambda/(\lambda_2+\lambda)}\big) \\
&= u+v-1+\min\big((1-v)(1-u)^{1-\theta_1}, (1-u)(1-v)^{1-\theta_2}\big), \quad u,v \in [0,1],
\end{aligned}
$$

with

$$
\theta_1 = \frac{\lambda}{\lambda_1+\lambda} \quad \text{and} \quad \theta_2 = \frac{\lambda}{\lambda_2+\lambda}.
$$

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

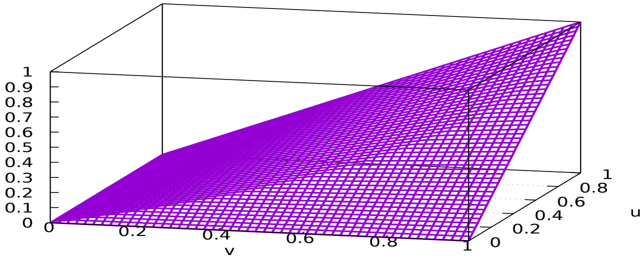


Fig. S.4: Survival copula graph with $\theta_1 = 0.3$ and $\theta_2 = 0.7$.

e) We have

$$\begin{aligned} C(u,v) = {} & u+v-1+(1-u)(1-v)^{1-\theta_2}\mathbb{1}_{\{(1-u)^{\theta_1}<(1-v)^{\theta_2}\}} \\ & +(1-v)(1-u)^{1-\theta_1}\mathbb{1}_{\{(1-u)^{\theta_1}>(1-v)^{\theta_2}\}}, \quad u,v\in[0,1], \end{aligned}$$

hence

$$\begin{aligned} \frac{\partial C}{\partial u}(u,v) = {} & -(1-v)^{1-\theta_2}\mathbb{1}_{\{(1-u)^{\theta_1}<(1-v)^{\theta_2}\}} \\ & -(1-\theta_1)(1-v)(1-u)^{-\theta_1}\mathbb{1}_{\{(1-u)^{\theta_1}>(1-v)^{\theta_2}\}} \end{aligned}$$

and the survival copula density is given by

$$\begin{aligned} \frac{\partial^2 C}{\partial u \partial v}(u,v) = {} & (1-\theta_2)(1-v)^{-\theta_2}\mathbb{1}_{\{(1-u)^{\theta_1}<(1-v)^{\theta_2}\}} \\ & +(1-\theta_1)(1-u)^{-\theta_1}\mathbb{1}_{\{(1-u)^{\theta_1}>(1-v)^{\theta_2}\}}, \quad u,v\in[0,1], \end{aligned}$$

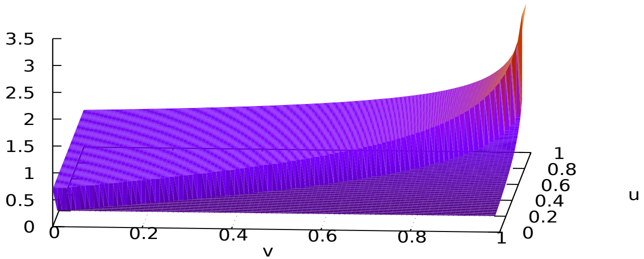


Fig. S.5: Survival copula density graph with $\theta_1 = 0.3$ and $\theta_2 = 0.7$.

Remark: When $\lambda = 0$ we have $\theta_1 = \theta_2 = 0$ and $\tau = +\infty$ *a.s.*, therefore we have

$$\min(\tau_1, \tau) = \tau_1 \quad \text{and} \quad \min(\tau_2, \tau) = \tau_2,$$

hence the copula $C(u, v)$ is given by

$$C(u, v) = u + v - 1 + (1 - v)(1 - u) = uv, \qquad u, v \in [0, 1],$$

which coincides with the copula of independence.

Chapter 5

Exercise 5.1 The payoff C is that of a *put* option with strike price $K = \$3$.

Exercise 5.2 Each of the two possible scenarios yields one equation:

$$\begin{cases} 5\xi + \eta = 0 \\ \\ 2\xi + \eta = 6, \end{cases} \qquad \text{with solution} \qquad \begin{cases} \xi = -2 \\ \\ \eta = +10. \end{cases}$$

The hedging strategy at $t = 0$ is to **shortsell** $-\xi = +2$ units of the asset S priced $S_0 = 4$, and to put $\eta = \$10$ on the savings account. The price $V_0 = \xi S_0 + \eta$ of the initial portfolio at time $t = 0$ is

$$V_0 = \xi S_0 + \eta = -2 \times 4 + 10 = \$2,$$

which yields the price of the claim at time $t = 0$. In order to hedge then option, one should:

i) At time $t = 0$,

a. Charge the \$2 option price.
b. Shortsell $-\xi = +2$ units of the stock priced $S_0 = 4$, which yields \$8.
c. Put $\eta = \$8 + \$2 = \$10$ on the savings account.

ii) At time $t = 1$,

a. If $S_1 = \$5$, spend \$10 from savings to buy back $-\xi = +2$ stocks.
b. If $S_1 = \$2$, spend \$4 from savings to buy back $-\xi = +2$ stocks, and deliver a \$10 - \$4 = \$6 payoff.

Pricing the option by the expected value $\mathbb{E}^*[C]$ yields the equality

$$\begin{aligned}
\$2 &= \mathbb{E}^*[C] \\
&= 0 \times \mathbb{P}^*(C = 0) + 6 \times \mathbb{P}^*(C = 6) \\
&= 0 \times \mathbb{P}^*(S_1 = 2) + 6 \times \mathbb{P}^*(S_1 = 5) \\
&= 6 \times q^*,
\end{aligned}$$

hence the risk-neutral probability measure $\mathbb{P}^*$ is given by

$$p^* = \mathbb{P}^*(S_1 = 5) = \frac{2}{3} \quad \text{and} \quad q^* = \mathbb{P}^*(S_1 = 2) = \frac{1}{3}.$$

Exercise 5.3

a) Each of the stated conditions yields one equation, *i.e.*

$$\begin{cases} 4\xi + \eta = 1 \\ \\ 5\xi + \eta = 3, \end{cases} \qquad \text{with solution} \qquad \begin{cases} \xi = 2 \\ \\ \eta = -7. \end{cases}$$

Therefore, the portfolio allocation at $t = 0$ consists to purchase $\xi = 2$ unit of the asset S priced $S_0 = 4$, and to borrow $-\eta = \$7$ in cash.

We can check that the price $V_0 = \xi S_0 + \eta$ of the initial portfolio at time $t = 0$ is

$$V_0 = \xi S_0 + \eta = 2 \times 4 - 7 = \$1.$$

b) This loss is expressed as

$$\xi \times \$2 + \eta = 2 \times 2 - 7 = -\$3.$$

Note that the \$1 received when selling the option is not counted here because it has already been fully invested into the portfolio.

Exercise 5.4

a) i) Does this model allow for arbitrage? | Yes | ✓ | | No | |

ii) If this model allows for arbitrage opportunities, how can they be realized? | By shortselling | | | By borrowing on savings | ✓ | | N.A. | |

b) i) Does this model allow for arbitrage? | Yes | | | No | ✓ |

ii) If this model allows for arbitrage opportunities, how can they be realized? | By shortselling | | | By borrowing on savings | | | N.A. | ✓ |

c) i) Does this model allow for arbitrage? | Yes | ✓ | | No | |

ii) If this model allows for arbitrage opportunities, how can they be realized? | By shortselling | ✓ | | By borrowing on savings | | | N.A. | |

Exercise 5.5 Hedging a claim with possible payoff values C_a, C_b, C_c would require to solve

$$\begin{cases} (1+a)\xi S_0^{(1)} + (1+r)\eta S_0^{(0)} = C_a \\ \\ (1+b)\xi S_0^{(1)} + (1+r)\eta S_0^{(0)} = C_b \\ \\ (1+c)\xi S_0^{(1)} + (1+r)\eta S_0^{(0)} = C_c, \end{cases}$$

for ξ and η, which is not possible in general due to the existence of three conditions with only two unknowns.

Exercise 5.6

a) Each of two possible scenarios yields one equation:

$$\begin{cases} \alpha \overline{S}_1 + \beta = \overline{S}_1 - K \\ \\ \alpha \underline{S}_1 + \beta = 0, \end{cases} \quad \text{with solution} \quad \begin{cases} \alpha = \dfrac{\overline{S}_1 - K}{\overline{S}_1 - \underline{S}_1} \\ \\ \beta = -\underline{S}_1 \dfrac{\overline{S}_1 - K}{\overline{S}_1 - \underline{S}_1}. \end{cases}$$

b) We have

$$0 \leqslant \alpha = \frac{\overline{S}_1 - K}{\overline{S}_1 - \underline{S}_1} \leqslant 1$$

since $K \in [\underline{S}_1, \overline{S}_1]$.

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

c) We find

$$\begin{aligned}\mathrm{SRM_C} &= \alpha S_0 + \beta \\ &= \alpha(S_0 - \underline{S}_1) \\ &= (S_0 - \underline{S}_1)\frac{\overline{S}_1 - K}{\overline{S}_1 - \underline{S}_1}.\end{aligned}$$

We note that when $S_0 < \underline{S}_1$ the value of $\mathrm{SRM_C}$ is negative because in this case, investing in the zero-cost portfolio $(\alpha, -\alpha S_0)$ that would yield a payoff at least equal to $\alpha(\underline{S}_1 - S_0) = -\alpha(\underline{S}_0 - S_1) > 0$, which represents an *arbitrage opportunity*.

Exercise 5.7

a) The payoff of the long box spread option is given in terms of K_1 and K_2 as

$$\begin{aligned}(x - K_1)^+ - (K_1 - x)^+ - (x - K_2)^+ + (K_2 - x)^+ &= x - K_1 - (x - K_2) \\ &= K_2 - K_1.\end{aligned}$$

b) From Table 5.1 we check that the strike prices suitable for a long box spread option on the Hang Seng Index (HSI) are $K_1 = 25,000$ and $K_2 = 25,200$.

c) Based on the data provided, we note that the long box spread can be realized in two ways.

   i) Using the put option issued by BI (BOCI Asia Ltd.) at 0.044.

   In this case, the box spread option represents a short position priced

$$\underbrace{0.540}_{\text{Long call}} \times 7,500 \underbrace{-0.064}_{\text{Short put}} \times 8,000 \underbrace{-0.370}_{\text{Short call}} \times 11,000 \underbrace{+0.044}_{\text{Long put}} \times 10,000 = -92$$

   index points, or $-92 \times \$50 = -\$4,600$.

   Note that option prices are quoted in index points (to be multiplied by the relevant option/warrant entitlement ratio), and every index point is worth \$50.

   ii) Using the put option issued by HT (Haitong Securities) at 0.061.

   In this case, the box spread option represents a long position priced

$$\underbrace{0.540}_{\text{Long call}} \times 7,500 \underbrace{-0.044}_{\text{Short put}} \times 8,000 \underbrace{-0.370}_{\text{Short call}} \times 11,000 \underbrace{+0.061}_{\text{Long put}} \times 10,000 = +78$$

   index points, or $78 \times \$50 = \$3,900$.

d) As the option built in i)) represents a short position paying \$4,600 today with an additional $\$50 \times (K_2 - K_1) = 200 = \$10,000$ payoff at maturity on March 28, I would definitely enter this position.

As for the option built in ii)), it is less profitable because it costs \$3,900, however it is still profitable taking into account the \$10,000 payoff at maturity on March 28.

Chapter 6

Exercise 6.1

a) We have

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leqslant x) \\ &= \int_0^x f_X(y)dy \\ &= \gamma\theta^\gamma \int_0^x \frac{1}{(\theta+y)^{\gamma+1}}dy \\ &= \left[-\left(\frac{\theta}{\theta+y}\right)^\gamma\right]_0^x \\ &= 1-\left(\frac{\theta}{\theta+x}\right)^\gamma, \qquad x \in \mathbb{R}_+. \end{aligned}$$

b) Since the distribution of X admits a probability density function, the cumulative distribution function $x \longmapsto F_X(x)$ is continuous in x and we have $\mathbb{P}(X = x) = 0$ for all $x > 0$. Hence the Value at Risk V_X^p at the level p is given by the relation $F_X(V_X^p) = p$, *i.e.*

$$\left(\frac{\theta}{\theta+V_X^p}\right)^\gamma = 1-p,$$

which gives

$$V_X^p = \theta\left(\frac{1}{(1-p)^{1/\gamma}} - 1\right).$$

In particular, with $p = 99\%$, $\theta = 40$ and $\gamma = 2$, we find

$$V_X^p = ((1-p)^{-1/\gamma} - 1)\theta = 40\big(\sqrt{100} - 1\big) = \$360.$$

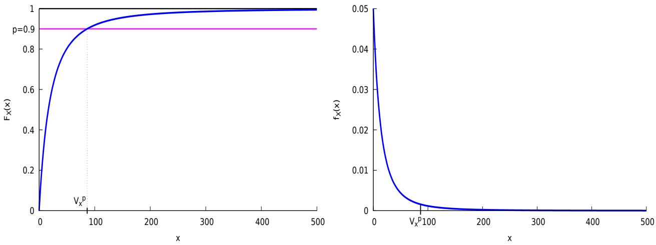


Fig. S.6: Pareto CDF $x \mapsto F_X(x)$ and PDF $x \mapsto f_X(x)$ with $V_X^{99\%} = \$86.49$.

Exercise 6.2

a) We have $\mathbb{P}(X = 100) = 0.02$.
b) We have $V_X^q = 100$ for all $q \in [0.97, 0.99]$.
c) The value at risk V_X^q at the level $q \in [0.99, 1]$ satisfies

$$F_X(V_X^q) = \mathbb{P}(X \leqslant V_X^q) = 0.99 + 0.01 \times (V_X^q - 100)/50 = q,$$

hence

$$V_X^q = 100 + 50(100q - 99) = 5000q - 4850, \qquad q \in [0.99, 1].$$

Exercise 6.3 We find $V_X^{99\%} = 100$ according to the following cumulative distribution function:

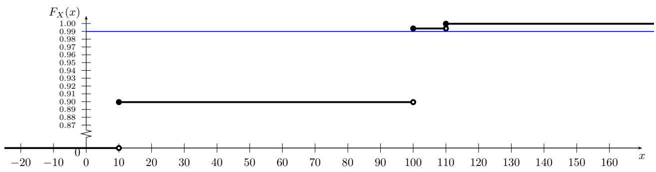


Fig. S.7: Cumulative distribution function of X and Y.

Exercise 6.4

a) We have

$$V_X^p := \inf\left\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\right\} = -\frac{1}{\lambda}\log(1-p) = \mathbb{E}[X]\log\frac{1}{1-p}.$$

When $p = 95\%$ this yields

$$V_X^p \simeq 2.996\mathbb{E}[X].$$

b) We find that the required capital C_X satisfies

$$C_X = V_X^p - \mathbb{E}[X] = \mathbb{E}[X] \log \frac{1}{1-p} - \mathbb{E}[X],$$

i.e.

$$C_X = V_X^{95\%} - \mathbb{E}[X] \simeq 1.996\mathbb{E}[X],$$

which means doubling the estimated amount of liabilities.

Exercise 6.5 By Proposition 6.2 and the geometric series identity (13.53), we have

$$\begin{aligned}
\mathbb{E}[X \mid X \geqslant a] &= \frac{1}{\mathbb{P}(X \geqslant a)} \mathbb{E}\big[X \mathbb{1}_{\{X \geqslant a\}}\big] \\
&= \frac{1}{\mathbb{P}(X \geqslant a)} \sum_{k \geqslant a} k \mathbb{P}(X = k) \\
&= \frac{1}{\displaystyle\sum_{k \geqslant a} (1-p)^k} \sum_{k \geqslant a} k (1-p)^k \\
&= \frac{(1-p)^a}{(1-p)^a \displaystyle\sum_{k \geqslant 0} (1-p)^k} \sum_{k \geqslant 0} (k+a)(1-p)^k \\
&= a + \frac{1}{\displaystyle\sum_{k \geqslant 0} (1-p)^k} \sum_{k \geqslant 0} k (1-p)^k \\
&= a + p \sum_{k \geqslant 0} k (1-p)^k \\
&= a + \frac{1}{p} \\
&= a + \mathbb{E}[X].
\end{aligned}$$

This can be recovered numerically for example with $a = 11$ using the R code below.

```
geo_samples <- rgeom(100000, prob = 1/4)
mean(geo_samples)
mean(geo_samples[geo_samples>=10])
```

Exercise 6.6

a) As in the proof of the Markov inequality, for every $x > 0$ and $r > 0$ we have

$$\begin{aligned} x^r \mathbb{P}(X \geqslant x) &= x^r \mathbb{E}\left[\mathbb{1}_{\{X \geqslant x\}}\right] \\ &\leqslant \mathbb{E}\left[X^r \mathbb{1}_{\{X \geqslant x\}}\right] \\ &\leqslant \mathbb{E}\left[|X|^r\right], \end{aligned}$$

hence

$$\mathbb{P}(X \leqslant x) \geqslant 1 - \frac{1}{x^r}\mathbb{E}[|X|^r], \qquad x > 0. \tag{A.1}$$

From the inequality (A.1), it follows that

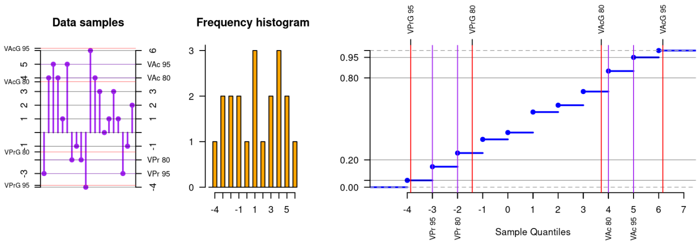
$$\begin{aligned} V_X^p &= \inf\left\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p\right\} \\ &\leqslant \inf\left\{x \in \mathbb{R} \ : \ 1 - \frac{1}{x^r}\mathbb{E}[|X|^r] \geqslant p\right\} \\ &= \inf\left\{x \in \mathbb{R} \ : \ x^r \geqslant \frac{1}{1-p}\mathbb{E}[|X|^r]\right\} \\ &= \left(\frac{\mathbb{E}[|X|^r]}{1-p}\right)^{1/r} \\ &= \frac{\|X\|_{L^r(\Omega)}}{(1-p)^{1/r}}. \end{aligned}$$

b) Taking $p = 95\%$ and $r = 1$ we get

$$V_X^{95\%} \leqslant \frac{1}{1-p}\mathbb{E}[|X|] = 20\mathbb{E}[|X|].$$

To summarize, a smaller L^r-norm of X tends to make the value at risk V_X smaller.

Exercise 6.7



See the attached **code*** for a solution using R.

a) i) $\mathrm{VaR}^{95}_{\mathrm{Ac\text{-}H}} = 5$.

ii) $\mathrm{VaR}^{80}_{\mathrm{Ac\text{-}H}} = 4$.

iii) $\mathrm{VaR}^{95}_{\mathrm{P}r\text{-}\mathrm{H}} = -3$.

iv) $\mathrm{VaR}^{80}_{\mathrm{P}r\text{-}\mathrm{H}} = -2$.

b) By Proposition 6.16, we have:

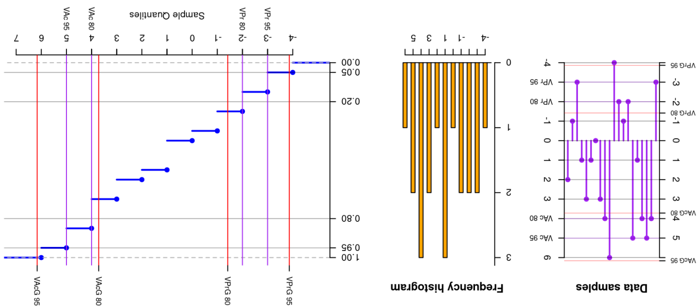
i) $\mathrm{VaR}^{95}_{\mathrm{Ac\text{-}G}} = 1.15+3.048\times$ qnorm(0.95)=6.164,

ii) $\mathrm{VaR}^{80}_{\mathrm{Ac\text{-}G}} = 1.15+3.048\times$ qnorm(0.80)=3.71551,

iii) $\mathrm{VaR}^{95}_{\mathrm{P}r\text{-}\mathrm{G}} = 1.15\text{-}3.048\times$ qnorm(0.95)=-3.864.

iv) $\mathrm{VaR}^{80}_{\mathrm{P}r\text{-}\mathrm{G}} = 1.15\text{-}3.048\times$ qnorm(0.80)=-1.41551,

Remark. The "Practitioner" Values at Risk can be better visualized after applying top-down and left-right symmetries (or a 180^o rotation) to the original CDF, as in the next figure.



* Right-click to save as attachment (may not work on 🍎).

Chapter 7

Exercise 7.1

a) Noting that $p = 1 - \mathrm{e}^{-\lambda \mathrm{VaR}_X^p}$ and using integration by parts on $\left[\mathrm{VaR}_X^p, \infty\right)$ with $u(x) = x$ and $v'(x) = \mathrm{e}^{-\lambda x}$, we have

$$\begin{aligned}
\mathbb{E}[X \mid X > \mathrm{VaR}_X^p] &= \frac{1}{\mathbb{P}(X > \mathrm{VaR}_X^p)} \int_{\mathrm{VaR}_X^p}^\infty x f_X(x) dx \\
&= \frac{\lambda}{1-p} \int_{\mathrm{VaR}_X^p}^\infty x \mathrm{e}^{-\lambda x} dx \\
&= \frac{\lambda}{1-p} \int_{\mathrm{VaR}_X^p}^\infty u(x) v'(x) dx \\
&= \frac{\lambda}{1-p} \left(\big[u(x) v(x) \big]_{\mathrm{VaR}_X^p}^\infty - \int_{\mathrm{VaR}_X^p}^\infty u'(x) v(x) dx \right) \\
&= \frac{\lambda}{1-p} \left(\left[-\frac{x}{\lambda} \mathrm{e}^{-\lambda x} \right]_{\mathrm{VaR}_X^p}^\infty + \frac{1}{\lambda} \int_{\mathrm{VaR}_X^p}^\infty \mathrm{e}^{-\lambda x} dx \right) \\
&= \frac{\lambda}{1-p} \left(\frac{\mathrm{VaR}_X^p}{\lambda} \mathrm{e}^{-\lambda \mathrm{VaR}_X^p} + \frac{1}{\lambda^2} \mathrm{e}^{-\lambda \mathrm{VaR}_X^p} \right) \\
&= \frac{\lambda}{1-p} \left(\frac{\mathrm{VaR}_X^p}{\lambda} (1-p) + \frac{1-p}{\lambda^2} \right) \\
&= \mathrm{VaR}_X^p + \frac{1}{\lambda} \\
&= \frac{1}{\lambda} - \frac{\log(1-p)}{\lambda}.
\end{aligned}$$

b) We have

$$\begin{aligned}
\mathrm{TV}_X^p &= \frac{1}{1-p} \int_p^1 V_X^q dq \\
&= -\frac{1}{\lambda(1-p)} \int_p^1 \log(1-q) dq \\
&= -\frac{1}{\lambda(1-p)} \int_0^{1-p} (\log q) dq \\
&= \frac{1 - p + (1-p) \log \dfrac{1}{1-p}}{\lambda(1-p)} \\
&= \frac{1}{\lambda} + \frac{1}{\lambda} \log \frac{1}{1-p} \\
&= \mathbb{E}[X] \left(1 + \log \frac{1}{1-p} \right)
\end{aligned}$$

$$= \mathbb{E}[X] + V_X^p.$$

Exercise 7.2

a) If $\mathbb{P}(X > z) > 0$ we have $\mathbb{E}\big[(X - z)\mathbb{1}_{\{X>z\}}\big] > 0$, hence

$$\mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big] > \mathbb{E}\big[z\mathbb{1}_{\{X>z\}}\big] = z\mathbb{P}(X > z),$$

and

$$\mathbb{E}[X \mid X > z] = \frac{\mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big]}{\mathbb{P}(X > z)} > z. \tag{A.2}$$

Recall that $\mathbb{E}[X \mid X > z]$ is not defined if $\mathbb{P}(X > z) = 0$.

b) We have

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\big[X\mathbb{1}_{\{X\leqslant z\}}\big] + \mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big] \\
&\leqslant z\mathbb{E}\big[\mathbb{1}_{\{X\leqslant z\}}\big] + \mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big] \\
&= z\mathbb{P}(X \leqslant z) + \mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big] \\
&\leqslant \mathbb{E}[X \mid X > z]\mathbb{P}(X \leqslant z) + \mathbb{E}[X \mid X > z]\mathbb{P}(X > z) \\
&= \mathbb{E}[X \mid X > z].
\end{aligned}$$

Note that $\mathbb{E}[X] = \mathbb{E}[X \mid X > z]$ when $\mathbb{P}(X \leqslant z) = 0$, *i.e.* $\mathbb{P}(X > z) = 1$.

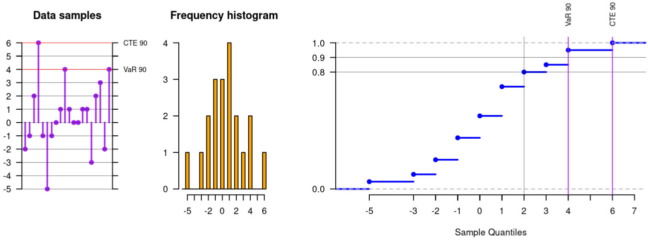
c) When $\mathbb{P}(X \leqslant z) > 0$, from (A.2) we find

$$\begin{aligned}
\mathbb{E}[X] &\leqslant z\mathbb{P}(X \leqslant z) + \mathbb{E}\big[X\mathbb{1}_{\{X>z\}}\big] \\
&< \mathbb{E}[X \mid X > z]\mathbb{P}(X \leqslant z) + \mathbb{E}[X \mid X > z]\mathbb{P}(X > z) \\
&= \mathbb{E}[X \mid X > z].
\end{aligned}$$

d) By (6.7) we have $\mathbb{P}\big(X \leqslant V_X^p\big) \geqslant p > 0$, hence by (c)) above we find $\mathrm{CTE}_X^p = \mathbb{E}\big[X \,\big|\, X > V_X^p\big] > \mathbb{E}[X]$.

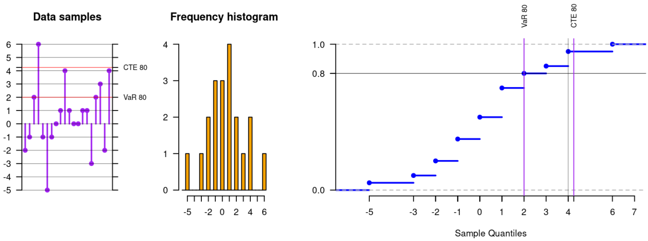
Exercise 7.3

a) We have $\mathrm{VaR}_X^{0.9} = 4$ and $\mathrm{CTE}_X^{0.9} = 6$.



b) We have $\mathrm{VaR}_X^{0.8} = 2$ and

$$\mathrm{CTE}_X^{0.8} = \frac{3 + 2 \times 4 + 6}{4} = \frac{17}{4} = 4.25.$$



Equivalently, we have

$$\begin{aligned}\mathrm{CTE}_X^{0.8} &= \frac{0.05 \times 3 + 0.1 \times 4 + 0.05 \times 6}{0.05 + 0.1 + 0.05} \\ &= \frac{0.05 \times 3 + 0.1 \times 4 + 0.05 \times 6}{0.2} \\ &= \frac{0.85}{0.2} = 4.25.\end{aligned}$$

Exercise 7.4

a) $\mathrm{VaR}_X^{90\%} = 4$.

b) $\mathbb{E}\big[X \mathbb{1}_{\{X > V_X^{90\%}\}}\big] = \dfrac{5+6}{23} = \dfrac{11}{23}$.

c) $\mathbb{P}\big(X > V_X^{90\%}\big) = \dfrac{2}{23}$.

d) $\mathrm{CTE}_X^{90\%} = \mathbb{E}\big[X \mid X > V_X^{90\%}\big] = \dfrac{\mathbb{E}\big[X \mathbb{1}_{\{X > V_X^{90\%}\}}\big]}{\mathbb{P}\big(X > V_X^{90\%}\big)} = \dfrac{5+6}{2} = \dfrac{11}{2} = 5.50.$

e) $\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^{90\%}\}}\big] = \dfrac{4+5+6}{23} = \dfrac{15}{23}.$

f) $\mathbb{P}\big(X \geqslant V_X^{90\%}\big) = \dfrac{3}{23}.$

g) $\mathrm{ES}_X^{90\%} = \dfrac{1}{1-p}\big(\mathbb{E}\big[X \mathbb{1}_{\{X \geqslant V_X^{90\%}\}}\big] + V_X^{90\%}\big(1-p-\mathbb{P}\big(X \geqslant V_X^{90\%}\big)\big)\big) = 10 \times \dfrac{4+5+6}{23} + 10 \times 4\left(0.1 - \dfrac{3}{23}\right) = \dfrac{150}{23} + 40 \times \dfrac{2.3-3}{23} = \dfrac{150 - 40 \times 0.7}{23} = \dfrac{122}{23} = 5.304.$

h)
$$\begin{aligned}
\mathrm{TV}_X^{90\%} &= \frac{1}{1-p}\int_p^1 V_X^q dq = \frac{1}{1-p}\left(\int_p^{21/23} V_X^q dq + \int_{21/23}^{22/23} V_X^q dq + \int_{22/23}^1 V_X^q dq\right)\\
&= \frac{1}{1-p}\left(\int_p^{21/23} 4dq + \int_{21/23}^{22/23} 5dq + \int_{22/23}^1 6dq\right)\\
&= \frac{1}{1-p}\left(4\left(\frac{21}{23}-p\right) + \frac{5}{23} + \frac{6}{23}\right) = \frac{84-92p+5+6}{23(1-p)} = \frac{122}{23} = 5.304.
\end{aligned}$$

We note that $\mathrm{ES}_X^{90\%} = \mathrm{TV}_X^{90\%}$ according to Proposition 7.12. The attached **code** computes the above risk measures, as illustrated in Figure S.8.

```
> source("var-cte_quiz.R")
VaR90= 4, Threshold= 0.9130435
CTE90= 5.5
ES90= 5.304348
```

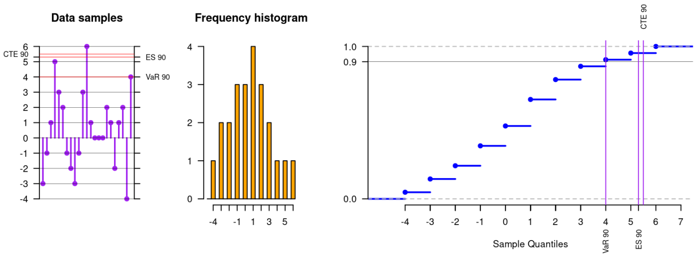


Fig. S.8: Value at Risk and Expected Shortfall for small data.

Exercise 7.5

a) The value at risk is $V_X^{98\%} = 100$.

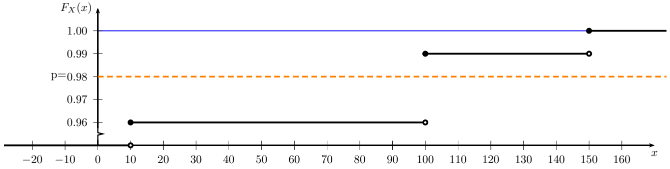


Fig. S.9: Cumulative distribution function of X.

b) Taking $p = 0.98$, we have

$$\begin{aligned}\mathrm{TV}_X^{98\%} &= \frac{1}{1-p}\int_p^1 V_X^q dq \\ &= \frac{1}{0.02}\left((0.99-0.98)\times 100 + (1-0.99)\times 150\right) = 125.\end{aligned}$$

c) We have

$$\begin{aligned}\mathrm{CTE}_X^{98\%} &= \frac{1}{\mathbb{P}(X > V_X^p)}\mathbb{E}\left[X\mathbb{1}_{\{X>V_X^p\}}\right] \\ &= \frac{1}{0.01}\times 150\times 0.01 = 150.\end{aligned}$$

d) We have

$$\begin{aligned}\mathrm{ES}_X^{98\%} &= \frac{1}{1-p}\mathbb{E}\left[X\mathbb{1}_{\{X\geqslant V_X^p\}}\right] + \frac{V_X^p}{1-p}\left(1-p-\mathbb{P}(X\geqslant V_X)\right) \\ &= \frac{1}{0.02}\left(100\times 0.03 + 150\times 0.01\right) + \frac{100}{0.02}\left(0.02-(0.03+0.01)\right) \\ &= \frac{4.5}{0.02} + \frac{100}{0.02}\left(0.02-(0.03+0.01)\right) = 125.\end{aligned}$$

Note that we also have

$$\begin{aligned}\mathrm{ES}_X^{98\%} &= \frac{1}{1-p}\mathbb{E}\left[X\mathbb{1}_{\{X> V_X^p\}}\right] + \frac{V_X^p}{1-p}\left(1-p-\mathbb{P}(X> V_X)\right) \\ &= \frac{1}{0.02}\left(150\times 0.01\right) + \frac{100}{0.02}\left(0.02-0.01\right) \\ &= 125,\end{aligned}$$

hence the Expected Shortfall $\mathrm{ES}_X^{98\%}$ does coincide with the tail value at risk $TV_X^{98\%}$.

Exercise 7.6

a) The cumulative distribution function of X is given by the following graph:

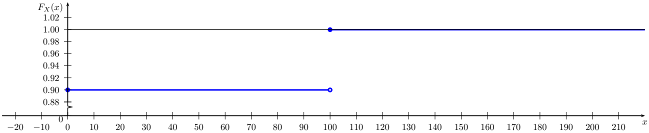


Fig. S.10: Cumulative distribution function of X.

b) The distribution of $X+Y$ is given by

$$\mathbb{P}(X+Y=0)=81\%, \quad \mathbb{P}(X+Y=100)=18\%, \quad \mathbb{P}(X+Y=200)=1\%.$$

The cumulative distribution function of $X+Y$ is given by the following graph:

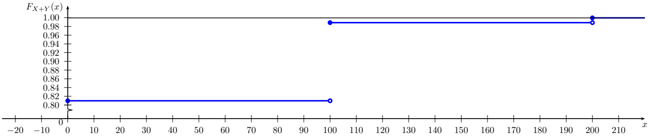


Fig. S.11: Cumulative distribution function of $X+Y$.

c) We have $V_{X+Y}^{99\%} = V_{X+Y}^{95\%} = V_{X+Y}^{90\%} = 100$.

Note that we have $V_{X+Y}^{99\%} = 100$ because

$$V_X^{99\%} = \inf\{x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant 0.99\} = 100.$$

d) We have

$$\mathrm{TV}_X^{90\%} = \frac{1}{1-0.9}\int_{0.9}^1 V_X^q dq = 100 \times \frac{1-0.9}{1-0.9} = 100.$$

e) We have

$$\begin{aligned}
\mathrm{TV}_{X+Y}^{99\%} &= \frac{1}{1-0.9}\int_{0.9}^{1} V_{X+Y}^{q} dq \\
&= \frac{1}{0.1}\left(\int_{0.9}^{0.99} 100 dq + \int_{0.99}^{1} 200 dq\right) \\
&= \frac{1}{0.1}(100\times 0.09 + 200\times 0.01) \\
&= 110,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{TV}_{X+Y}^{80\%} &= \frac{1}{1-0.8}\int_{0.9}^{1} V_{X+Y}^{q} dq \\
&= \frac{1}{0.2}\left(\int_{0.8}^{0.81} 0 dq + \int_{0.81}^{0.99} 100 dq + \int_{0.99}^{1} 200 dq\right) \\
&= \frac{1}{0.2}(100\times 0.18 + 200\times 0.01) = 100.
\end{aligned}$$

Exercise 7.7 (Exercise 6.2 continued).

a) For all $p \in [0.99, 1]$ we have

$$\begin{aligned}
\mathrm{TV}_{X}^{p} &= \frac{1}{1-p}\int_{p}^{1} V_{X}^{q} dq \\
&= \frac{1}{1-p}\int_{p}^{1} (5000q - 4850) dq \\
&= \frac{1}{1-p}\left(5000\frac{(1-p^2)}{2} - (1-p)4850\right) \\
&= 2500p - 2350.
\end{aligned}$$

In particular,

$$\mathrm{TV}_{X}^{99\%} = 2500\times 0.99 - 2350 = 125 \geqslant V_{X}^{99\%} = 100.$$

b) We have $V_X^{98\%} = 100$ and

$$\begin{aligned}
\mathrm{CTE}_{X}^{98\%} &= \mathbb{E}\big[X \mid X > V_X^{98\%}\big] \\
&= \frac{1}{\mathbb{P}\big(X > V_X^{98\%}\big)}\mathbb{E}\left[X \mathbb{1}_{\{X > V_X^{98\%}\}}\right] \\
&= \frac{1}{0.01}\int_{100}^{\infty} x f_X(x) dx \\
&= \frac{1}{0.01}\int_{100}^{\infty} x \frac{dF_X(x)}{dx} dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{0.01} \frac{0.01}{50} \int_{100}^{150} x dx \\
&= \frac{150^2 - 100^2}{2 \times 50} \\
&= 125.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathrm{TV}_X^{98\%} &= \frac{0.01}{0.02} \times 100 + \frac{1}{0.02} \left(5000 \frac{(1-0.99^2)}{2} - 0.01 \times 4850 \right) \\
&= 112.50 \\
&\geqslant V_X^{98\%} = 100,
\end{aligned}
$$

which differs from $\mathrm{CTE}_X^{98\%} = 125$ since

$$
\mathbb{P}\big(X = V_X^{98\%}\big) = \mathbb{P}(X = 100) = 0.02 > 0.
$$

Exercise 7.8

a) We have

$$
V_X^p := \inf \big\{ x \in \mathbb{R} \ : \ \mathbb{P}(X \leqslant x) \geqslant p \big\} = \log \frac{p}{1-p}.
$$

b) We have

$$
\begin{aligned}
&\mathbb{E}[X \mid X > \mathrm{VaR}_X^p] = \frac{1}{\mathbb{P}(X > \mathrm{VaR}_X^p)} \int_{\mathrm{VaR}_X^p}^\infty x f_X(x) dx \\
&= \frac{1}{1-p} \int_{\mathrm{VaR}_X^p}^\infty x \mathrm{e}^{-\lambda x} dx \\
&= \frac{1}{1-p} \int_{\mathrm{VaR}_X^p}^\infty \frac{x \mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2} dx \\
&= \frac{1}{1-p} \log(1 + \mathrm{e}^{\mathrm{VaR}_X^p}) - \frac{\mathrm{VaR}_X^p \mathrm{e}^{\mathrm{VaR}_X^p}}{1 + \mathrm{e}^{\mathrm{VaR}_X^p}} \\
&= \frac{1}{1-p} \log \left(1 + \frac{p}{1-p} \right) - \frac{1}{1-p} \frac{p}{1-p} \frac{1}{1 + \frac{p}{1-p}} \log \frac{p}{1-p} \\
&= \frac{1}{1-p} \log \frac{1}{1-p} - \frac{p}{1-p} \log \frac{p}{1-p} \\
&= -\frac{p}{1-p} \log p - \log(1-p).
\end{aligned}
$$

c) We have

$$\begin{aligned}
\mathrm{TV}_X^p &= \frac{1}{1-p}\int_p^1 V_X^q dq \\
&= \frac{1}{1-p}\int_p^1 \log\frac{q}{1-q}dq \\
&= \frac{1}{1-p}\int_p^1 \log q dq - \frac{1}{1-p}\int_p^1 \log(1-q)dq \\
&= \frac{1}{1-p}\int_p^1 \log q dq - \frac{1}{1-p}\int_0^{1-p} \log q dq \\
&= \frac{1}{1-p}\int_p^1 \log q dq - \frac{1}{1-p}\left(\int_0^1 \log q dq - \int_{1-p}^1 \log q dq\right) \\
&= \frac{p-1-p\log p}{1-p} - \frac{-1+p+(1-p)\log(1-p)}{1-p} \\
&= -\frac{p}{1-p}\log p - \log(1-p).
\end{aligned}$$

Exercise 7.9

a) We have

$$q\mathbb{P}(Z \geqslant q) = \mathbb{E}\big[q\mathbb{1}_{\{Z\geqslant q\}}\big] \leqslant \mathbb{E}\big[Z\mathbb{1}_{\{Z\geqslant q\}}\big] = \int_q^\infty x f_Z(x)dx, \qquad q \geqslant 0.$$

b) We have

$$\begin{aligned}
\int_q^\infty x f_Z(x)dx &= \int_q^\infty x\phi(x)dx \\
&= \frac{1}{\sqrt{2\pi}}\int_q^\infty x\mathrm{e}^{-x^2/2}dx \\
&= -\frac{1}{\sqrt{2\pi}}\left[\mathrm{e}^{-x^2/2}\right]_q^\infty \\
&= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-q^2/2} \\
&= \phi(q), \qquad q \geqslant 0,
\end{aligned}$$

and $1-p = \mathbb{P}(Z \geqslant q)$, hence

$$(1-p)q \leqslant \int_q^\infty x f_Z(x)dx = \phi(q), \qquad q \geqslant 0.$$

c) Taking $q := q_Z^p$ with $1-p = \mathbb{P}(Z \geqslant q_Z^p)$, we recover

$$V_X^p = \mu_X + \sigma_X q_Z^p \leqslant \mu_X + \frac{\sigma_X}{1-p}\phi(q_Z^p) = \mathrm{CTE}_X^p,$$

see Proposition 7.5.

Chapter 8

Exercise 8.1

a) We have

$$\mathbb{E}[X \mid G] = \lambda_G \int_0^\infty x e^{-\lambda_G x} dx = \frac{1}{\lambda_G}$$

and

$$\mathbb{E}[X \mid B] = \lambda_B \int_0^\infty x e^{-\lambda_B x} dx = \frac{1}{\lambda_B}.$$

b) We find

$$\begin{aligned}
\mathbb{P}(B \mid X = x) &= \frac{f_X(x \mid B)\mathbb{P}(B)}{f_X(x \mid G)\mathbb{P}(G) + f_X(x \mid B)\mathbb{P}(B)} \\
&= \frac{\lambda_B e^{-\lambda_B x}\mathbb{P}(B)}{\lambda_G e^{-\lambda_G x}\mathbb{P}(G) + \lambda_B e^{-\lambda_B x}\mathbb{P}(B)} \\
&= \frac{1}{1 + \dfrac{\lambda_G \mathbb{P}(G)}{\lambda_B \mathbb{P}(B)} e^{(\lambda_B - \lambda_G)x}} \\
&= \frac{1}{1 + \lambda(x)\dfrac{\mathbb{P}(G)}{\mathbb{P}(B)}},
\end{aligned}$$

where $\lambda(x)$ is the *likelihood ratio*

$$\lambda(x) = \frac{f_X(x \mid G)}{f_X(x \mid B)} = \frac{\lambda_G}{\lambda_B} e^{(\lambda_B - \lambda_G)x}, \qquad x > 0.$$

c) The condition

$$D\mathbb{P}(B \mid X = x) \leqslant L\mathbb{P}(G \mid X = x)$$

rewrites as

$$D\mathbb{P}(B \mid X = x) \leqslant L(1 - \mathbb{P}(B \mid X = x)),$$

i.e.

$$(L + D)\mathbb{P}(B \mid X = x) \leqslant L$$

or

$$\frac{L + D}{1 + \lambda(x)\dfrac{\mathbb{P}(G)}{\mathbb{P}(B)}} \leqslant L,$$

or

$$\lambda(x) = \frac{\lambda_G}{\lambda_B} \mathrm{e}^{(\lambda_B - \lambda_G)x} \geqslant \frac{D}{L} \frac{\mathbb{P}(B)}{\mathbb{P}(G)}.$$

This condition holds if and only if

$$x \geqslant \frac{1}{\lambda_B - \lambda_G} \log\left(\frac{D}{L} \frac{\lambda_B \mathbb{P}(B)}{\lambda_G \mathbb{P}(G)}\right),$$

provided that $\lambda_B > \lambda_G$. Therefore we have

$$\begin{aligned} \mathscr{A} &= \left\{ x \in \mathbb{R} \ : \ \lambda(x) \geqslant \frac{D}{L} \frac{\mathbb{P}(B)}{\mathbb{P}(G)} \right\} \\ &= \left[\frac{1}{\lambda_B - \lambda_G} \log\left(\frac{D}{L} \frac{\lambda_B \mathbb{P}(B)}{\lambda_G \mathbb{P}(G)}\right), \infty \right), \end{aligned}$$

under the condition

$$\mathbb{E}[X \mid B] = \frac{1}{\lambda_B} < \frac{1}{\lambda_G} = \mathbb{E}[X \mid G].$$

Exercise 8.2

a) We find

$$\begin{aligned} \mathbb{P}(B \mid X = x) &= \frac{f_X(x \mid B)\mathbb{P}(B)}{f_X(x \mid G)\mathbb{P}(G) + f_X(x \mid B)\mathbb{P}(B)} \\ &= \mathbb{1}_{[0,\lambda_B]}(x) \frac{\mathbb{P}(B)/\lambda_B}{\mathbb{P}(G)/\lambda_G + \mathbb{P}(B)/\lambda_B} \\ &= \mathbb{1}_{[0,\lambda_B]}(x) \frac{1}{1 + \dfrac{\lambda_B \mathbb{P}(G)}{\lambda_G \mathbb{P}(B)}}. \end{aligned}$$

b) We have

$$\mathbb{E}[X \mid G] = \int_{-\infty}^{\infty} y f_X(y \mid G) dy = \frac{1}{\lambda_G} \int_0^{\lambda_G} y dy = \frac{\lambda_G}{2},$$

and similarly

$$\mathbb{E}[X \mid B] = \int_{-\infty}^{\infty} y f_X(y \mid B) dy = \frac{1}{\lambda_B} \int_0^{\lambda_B} y dy = \frac{\lambda_B}{2}.$$

c) The condition

$$D\mathbb{P}(B \mid X = x) \leqslant L\mathbb{P}(G \mid X = x)$$

rewrites as

$$D\mathbb{P}(B \mid X = x) \leqslant L(1 - \mathbb{P}(B \mid X = x)),$$

i.e.

$$(L + D)\mathbb{P}(B \mid X = x) \leqslant L$$

or

$$D\mathbb{1}_{[0,\lambda_B]}(x) \leqslant L\mathbb{1}_{(\lambda_B,\infty)}(x) + L\frac{\lambda_B \mathbb{P}(G)}{\lambda_G \mathbb{P}(B)}.$$

This condition holds if and only if

$$\frac{\lambda_B}{\lambda_G} \geqslant \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)}.$$

when $x \in [0, \lambda_B]$, and is always satisfied when $x \in (\lambda_B, \infty)$. Therefore, we have $\mathscr{A} = \mathbb{R}$ if

$$\frac{\lambda_B}{\lambda_G} \geqslant \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)},$$

and $\mathscr{A} = (\lambda_B, \infty)$ if

$$\frac{\lambda_B}{\lambda_G} < \frac{D}{L}\frac{\mathbb{P}(B)}{\mathbb{P}(G)},$$

under the condition

$$\mathbb{E}[X \mid B] = \lambda_B < \lambda_G = \mathbb{E}[X \mid G].$$

Exercise 8.3

a) We have

$$\overline{F}_G(x) = \mathrm{e}^{-\lambda_G x} \quad \text{and} \quad \overline{F}_B(x) = \mathrm{e}^{-\lambda_B x}, \quad x \geqslant 0,$$

hence

$$\overline{F}_B^{-1}(y) = -\frac{\log y}{\lambda_B}, \quad y \in (0, 1],$$

and

$$\begin{aligned}\overline{F}_G(\overline{F}_B^{-1}(y)) &= \overline{F}_G\left(-\frac{\log y}{\lambda_B}\right) \\ &= \mathrm{e}^{\lambda_G (\log y)/\lambda_B} \\ &= y^{\lambda_G/\lambda_B}, \qquad y \in [0, 1].\end{aligned}$$

We check that, according to Proposition 8.7,

$$
\begin{aligned}
\frac{d}{dy}\overline{F}_G\big(\overline{F}_B^{-1}(y)\big) &= \frac{d}{dy}y^{\lambda_G/\lambda_B} \\
&= \frac{d}{dy}\mathrm{e}^{\lambda_G(\log y)/\lambda_B} \\
&= \frac{\lambda_G}{y\lambda_B}\mathrm{e}^{\lambda_G(\log y)/\lambda_B} \\
&= \frac{\lambda_G}{\lambda_B}\mathrm{e}^{(\lambda_G-\lambda_B)(\log y)/\lambda_B} \\
&= \frac{\lambda_G}{\lambda_B}\mathrm{e}^{(\lambda_B-\lambda_G)\overline{F}_B^{-1}(y)} \\
&= \lambda\big(\overline{F}_B^{-1}(y)\big), \qquad x \in [0,1].
\end{aligned}
$$

Figure S.12 presents three samples of exponential ROC curves, with successively $(\lambda_B, \lambda_G) = (10, 1)$, $(\lambda_B, \lambda_G) = (2, 1)$, and $(\lambda_B, \lambda_G) = (1, 1)$.

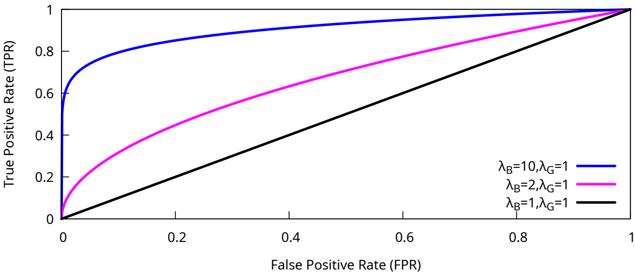


Fig. S.12: Exponential ROC curves.

b) We have

$$\overline{F}_G(x) := 1 - \frac{x}{\lambda_G}, \qquad x \in [0, \lambda_G],$$

and

$$\overline{F}_B(x) := 1 - \frac{x}{\lambda_B}, \qquad x \in [0, \lambda_B],$$

hence

$$\overline{F}_B^{-1}(y) := \lambda_B(1-y), \qquad y \in [0,1],$$

hence

$$\overline{F}_G\big(\overline{F}_B^{-1}(x)\big) = 1 - \frac{\lambda_B}{\lambda_G}(1-y) = \frac{\lambda_G - \lambda_B}{\lambda_G} + \frac{\lambda_B}{\lambda_G}y, \quad y \in [0,1].$$

Figure S.13 presents three samples of uniform ROC curves, with successively $(\lambda_B, \lambda_G) = (1, 8)$, $(\lambda_B, \lambda_G) = (1, 2)$, and $(\lambda_B, \lambda_G) = (1, 1)$.

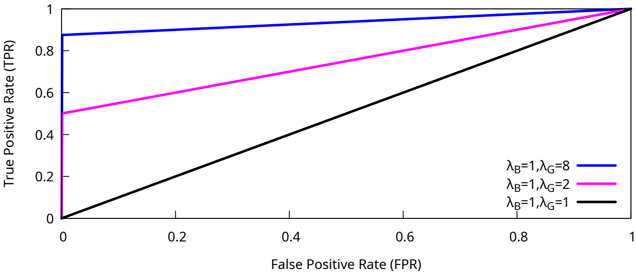


Fig. S.13: Uniform ROC curves.

Exercise 8.4

a) We have

$$\begin{aligned}\mathbb{P}(B \mid X = x) &= \frac{\mathbb{P}(B) f_X(x \mid B)}{\mathbb{P}(G) f_X(x \mid G) + \mathbb{P}(B) f_X(x \mid B)}\\ &= \frac{\mathbb{P}(B) \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}}{\mathbb{P}(G) \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)} + \mathbb{P}(B) \mathrm{e}^{-(x-\mu_B)^2/(2\sigma^2)}}\\ &= \frac{1}{1 + \mathrm{e}^{\alpha+\beta x}}, \qquad x \in \mathbb{R},\end{aligned}$$

with

$$\beta := \frac{\mu_G - \mu_B}{\sigma^2} > 0$$

and

$$\alpha := -\beta \frac{\mu_G + \mu_B}{2} + \log\left(\frac{\mathbb{P}(G)}{\mathbb{P}(B)}\right).$$

b) We have

$$\begin{aligned}\lambda(x) &= \frac{f_X(x \mid G)}{f_X(x \mid B)}\\ &= \mathrm{e}^{-(x-\mu_G)^2/(2\sigma^2)+(x-\mu_B)^2/(2\sigma^2)}\\ &= \mathrm{e}^{-(\mu_G^2-\mu_B^2-2x(\mu_G-\mu_B))/(2\sigma^2)}\\ &= \mathrm{e}^{\beta x-(\mu_G^2-\mu_B^2)/(2\sigma^2)}, \qquad x \in \mathbb{R}.\end{aligned}$$

c) The condition

$$\lambda(x) = \mathrm{e}^{\beta x - (\mu_G^2 - \mu_B^2)/(2\sigma^2)} \geqslant \frac{D(x)}{L(x)} \frac{\mathbb{P}(B)}{\mathbb{P}(G)}$$

is equivalent to

$$\begin{aligned}\beta x &\geqslant \frac{\mu_G^2 - \mu_B^2}{2\sigma^2} + \log\left(\frac{D(x)}{L(x)} \frac{\mathbb{P}(B)}{\mathbb{P}(G)}\right)\\ &\geqslant \frac{\mu_G^2 - \mu_B^2}{2\sigma^2} + \log \frac{\mathbb{P}(B)}{\mathbb{P}(G)} + \log \frac{D(x)}{L(x)}\\ &\geqslant \frac{\mu_G^2 - \mu_B^2}{2\sigma^2} + \log \frac{\mathbb{P}(B)}{\mathbb{P}(G)} + x(a+b),\end{aligned}$$

hence

$$x \geqslant \frac{\mu_G^2 - \mu_B^2}{2\sigma^2(\beta - a - b)} + \frac{1}{\beta - a - b} \log \frac{\mathbb{P}(B)}{\mathbb{P}(G)},$$

provided that

$$\beta := \frac{\mu_G - \mu_B}{\sigma^2} > a + b.$$

In this case, we have

$$\mathscr{A}^* = [x^*, \infty) = \left[\frac{\mu_G^2 - \mu_B^2}{2\sigma^2(\beta - a - b)} + \frac{1}{\beta - a - b} \log \frac{\mathbb{P}(B)}{\mathbb{P}(G)}, \infty\right),$$

where

$$x^* := \frac{\mu_G^2 - \mu_B^2}{2\sigma^2(\beta - a - b)} + \frac{1}{\beta - a - b} \log \frac{\mathbb{P}(B)}{\mathbb{P}(G)}.$$

Chapter 9

Exercise 9.1 By differentiation of (9.2), *i.e.*

$$\begin{aligned}\mathbb{P}(\tau < T \mid \mathcal{F}_t) &:= \mathbb{P}(S_T < K \mid \mathcal{F}_t)\\ &= \Phi\left(-\frac{(\mu - \sigma^2/2)(T-t) + \log(S_t/K))}{\sigma\sqrt{T-t}}\right), \qquad T \geqslant t,\end{aligned}$$

with respect to T, we find

$$d\mathbb{P}(\tau \leqslant T \mid \mathcal{F}_t) = \frac{dT}{2\sigma\sqrt{2\pi(T-t)}} \left(\frac{\sigma^2}{2} - \mu + \frac{\log(S_t/K)}{T-t}\right)$$

$$\times \exp\left(-\frac{\left((\mu-\sigma^2/2))(T-t)+\log(S_t/K))\right)^2}{2(T-t)\sigma^2}\right),$$

provided that $\mu < \sigma^2/2$.

Exercise 9.2 Consider the first hitting time

$$\tau_K := \inf\{u \geqslant t \ : \ S_u \leqslant K\}$$

of the level $K > 0$ starting from $S_t > K$. By Lemma 15.1 in Privault (2022), we have

$$\mathbb{E}^*\left[\mathrm{e}^{-(\tau_K - t)r} \mid \mathcal{F}_t\right] = \left(\frac{K}{S_t}\right)^{2r/\sigma^2},$$

provided that $S_t \geqslant K$.

Exercise 9.3

a) We have

$$\begin{aligned}
\mathbb{E}[X_k X_l] &= \mathbb{E}\left[(a_k M + \sqrt{1-a_k^2} Z_k)(a_l M + \sqrt{1-a_l^2} Z_l)\right] \\
&= \mathbb{E}\left[a_k a_l M^2 + a_k M\sqrt{1-a_l^2} Z_l + a_l M\sqrt{1-a_k^2} Z_k + \sqrt{1-a_k^2}\sqrt{1-a_l^2} Z_k Z_l\right] \\
&= a_k a_l \mathbb{E}\left[M^2\right] + a_k\sqrt{1-a_l^2}\mathbb{E}[Z_l M] + a_l\sqrt{1-a_k^2}\mathbb{E}[Z_k M] \\
&\quad + \sqrt{1-a_k^2}\sqrt{1-a_l^2}\mathbb{E}[Z_k Z_l] \\
&= a_k a_l \mathbb{E}\left[M^2\right] + a_k\sqrt{1-a_l^2}\mathbb{E}[Z_l]\mathbb{E}[M] + a_l\sqrt{1-a_k^2}\mathbb{E}[Z_k]\mathbb{E}[M] \\
&\quad + \sqrt{1-a_k^2}\sqrt{1-a_l^2}\mathbb{1}_{\{k=l\}} \\
&= a_k a_l + (1-a_k^2)\mathbb{1}_{\{k=l\}} \\
&= \mathbb{1}_{\{k=l\}} + a_k a_l \mathbb{1}_{\{k\neq l\}}, \qquad k,l = 1,2,\ldots,n,
\end{aligned}$$

b) We check that the vector $(X_1, \ldots, X_n)$, with covariance matrix (9.12) has the probability density function

$$\begin{aligned}
&\varphi(x_1, \ldots, x_n) \\
&\quad = \frac{1}{(2\pi)^{n/2}} \prod_{k=1}^n (1-a_k^2)^{-1/2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x_1-a_1 m)^2}{2(1-a_1^2)}} \cdots \mathrm{e}^{-\frac{(x_n-a_n m)^2}{2(1-a_n^2)}} \frac{\mathrm{e}^{-m^2/2}}{\sqrt{2\pi}} dm
\end{aligned}$$

which is jointly Gaussian, with marginals given by

$$
\begin{aligned}
x_k \longmapsto & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, \ldots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n \\
&= \frac{1}{\sqrt{2\pi(1-a_k^2)}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x_k - a_k m)^2}{2(1-a_k^2)}} \frac{\mathrm{e}^{-m^2/2}}{\sqrt{2\pi}} dm \\
&= \frac{1}{2\pi\sqrt{1-a_k^2}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x_k - a_k m)^2}{2(1-a_k^2)} - m^2/2} dm \\
&= \frac{1}{2\pi\sqrt{1-a_k^2}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x_k^2 - 2a_k x_k m + m^2}{2(1-a_k^2)}} dm \\
&= \frac{\mathrm{e}^{-x_k^2/2}}{2\pi\sqrt{1-a_k^2}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(m - a_k x_k)^2}{2(1-a_k^2)}} dm \\
&= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x_k^2/2}, \qquad x_k \in \mathbb{R}.
\end{aligned}
$$

c) We have

$$
\begin{aligned}
\varphi(x_1, \ldots, x_n) &= \frac{1}{(2\pi)^{n/2}} \prod_{k=1}^{n} (1-a_k^2)^{-1/2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x_1 - a_1 m)^2}{2(1-a_1^2)}} \cdots \mathrm{e}^{-\frac{(x_n - a_n m)^2}{2(1-a_n^2)}} \frac{\mathrm{e}^{-m^2/2}}{\sqrt{2\pi}} dm \\
&= \frac{1}{(2\pi)^{n/2}} \prod_{k=1}^{n} (1-a_k^2)^{-1/2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2 + a_1^2 m^2 - 2x_1 a_1 m}{1-a_1^2} + \cdots + \frac{x_n^2 + a_n^2 m^2 - 2x_n a_n m}{1-a_n^2} + m^2 \right)} \frac{dm}{\sqrt{2\pi}} \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2}{1-a_1^2} + \cdots + \frac{x_n^2}{1-a_n^2} \right)} \prod_{k=1}^{n} (1-a_k^2)^{-1/2} \\
&\quad \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{m^2}{2}\left(1 + \frac{a_1^2}{1-a_1^2} + \cdots + \frac{a_n^2}{1-a_n^2}\right) + 2m\left(\frac{x_1 a_1}{2(1-a_1^2)} + \cdots + \frac{x_n a_n}{2(1-a_n^2)} \right)} dm \\
&= \frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2}{1-a_1^2} + \cdots + \frac{x_n^2}{1-a_n^2} \right)}}{\sqrt{(2\pi)^n (1-a_1^2) \cdots (1-a_n^2)}} \exp\left(\frac{\frac{1}{2}\left(\frac{x_1 a_1}{1-a_1^2} + \cdots + \frac{x_n a_n}{1-a_n^2} \right)^2}{1 + \frac{a_1^2}{1-a_1^2} + \cdots + \frac{a_n^2}{1-a_n^2}} \right) \\
&\quad \times \left(1 + \frac{a_1^2}{1-a_1^2} + \cdots + \frac{a_n^2}{1-a_n^2} \right)^{-1/2} \\
&= \frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2}{1-a_1^2} + \cdots + \frac{x_n^2}{1-a_n^2} \right)}}{\sqrt{(2\pi)^n \alpha^2 (1-a_1^2) \cdots (1-a_n^2)}} \exp\left(\frac{1}{2\alpha^2} \left(\frac{x_1 a_1}{1-a_1^2} + \cdots + \frac{x_n a_n}{1-a_n^2} \right)^2 \right)
\end{aligned}
$$

$$= \frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2}{1-a_1^2}+\cdots+\frac{x_n^2}{1-a_n^2}\right)}}{\sqrt{(2\pi)^n\alpha^2(1-a_1^2)\cdots(1-a_n^2)}}\exp\left(\frac{1}{2\alpha^2}\left(\frac{x_1a_1}{1-a_1^2}+\cdots+\frac{x_na_n}{1-a_n^2}\right)^2\right)$$
$$= \frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{x_1^2}{1-a_1^2}\left(1-\frac{a_1^2}{\alpha^2(1-a_1^2)}\right)+\cdots+\frac{x_n^2}{1-a_n^2}\left(1-\frac{a_n^2}{\alpha^2(1-a_n^2)}\right)\right)}}{\sqrt{(2\pi)^n\alpha^2(1-a_1^2)\cdots(1-a_n^2)}}\exp\left(\frac{1}{2\alpha^2}\sum_{1\leqslant p\neq l\leqslant n}\frac{x_px_la_pa_l}{(1-a_p^2)(1-a_l^2)}\right)$$
$$= \frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\mathrm{e}^{-\frac{1}{2}\langle x,\Sigma^{-1}x\rangle},$$

where

$$\alpha^2 := 1+\frac{a_1^2}{1-a_1^2}+\cdots+\frac{a_n^2}{1-a_n^2},$$

and

$$\Sigma^{-1} = \frac{1}{\alpha^2}\begin{bmatrix}
\frac{\alpha^2(1-a_1^2)-a_1^2}{(1-a_1^2)^2} & \frac{-a_1a_2}{(1-a_1^2)(1-a_2^2)} & \cdots & \frac{-a_1a_n}{(1-a_1^2)(1-a_n^2)} \\
\frac{-a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2(1-a_2^2)-a_2^2}{(1-a_2^2)^2} & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \frac{\alpha^2(1-a_{n-1}^2)-a_{n-1}^2}{(1-a_{n-1}^4)} & \frac{-a_{n-1}a_n}{(1-a_{n-1}^2)(1-a_n^2)} \\
\frac{-a_na_1}{(1-a_n^2)(1-a_1^2)} & \ddots & \frac{-a_na_{n-1}}{(1-a_n^2)(1-a_{n-1}^2)} & \frac{\alpha^2(1-a_n^2)-a_n^2}{(1-a_n^2)^2}
\end{bmatrix}.$$

Exercise 9.4 We have

$$\Sigma = \begin{bmatrix} 1 & a_1a_2 \\ a_2a_1 & 1 \end{bmatrix},$$

and letting

$$\begin{aligned}
\alpha^2 &:= 1+\frac{a_1^2}{1-a_1^2}+\frac{a_2^2}{1-a_2^2} \\
&= \frac{(1-a_1^2)(1-a_2^2)+a_1^2(1-a_2^2)+a_2^2(1-a_1^2)}{(1-a_1^2)(1-a_2^2)} \\
&= \frac{1-a_2^2a_1^2}{(1-a_1^2)(1-a_2^2)},
\end{aligned}$$

we find

$$\begin{aligned}
\Sigma^{-1} &= \frac{1}{\alpha^2}\begin{bmatrix} \frac{\alpha^2(1-a_1^2)-a_1^2}{(1-a_1^2)^2} & -\frac{a_1a_2}{(1-a_1^2)(1-a_2^2)} \\ -\frac{a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2(1-a_2^2)-a_2^2}{(1-a_2^2)^2} \end{bmatrix} \\
&= \frac{1}{\alpha^2}\begin{bmatrix} \frac{\alpha^2}{1-a_1^2}\left(1-\frac{(1-a_2^2)a_1^2}{1-a_2^2a_1^2}\right) & -\frac{a_1a_2}{(1-a_1^2)(1-a_2^2)} \\ -\frac{a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2}{1-a_2^2}\left(1-\frac{(1-a_1^2)a_2^2}{1-a_2^2a_1^2}\right) \end{bmatrix} \\
&= \frac{1}{\alpha^2}\begin{bmatrix} \frac{\alpha^2}{1-a_2^2a_1^2} & -\frac{a_1a_2}{(1-a_1^2)(1-a_2^2)} \\ -\frac{a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2}{1-a_2^2a_1^2} \end{bmatrix} \\
&= \frac{(1-a_1^2)(1-a_2^2)}{1-a_2^2a_1^2}\begin{bmatrix} \frac{\alpha^2}{1-a_2^2a_1^2} & -\frac{a_1a_2}{(1-a_1^2)(1-a_2^2)} \\ -\frac{a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2}{1-a_2^2a_1^2} \end{bmatrix} \\
&= \frac{1}{\alpha^2}\begin{bmatrix} \frac{\alpha^2}{1-a_2^2a_1^2} & -\frac{a_1a_2}{(1-a_1^2)(1-a_2^2)} \\ -\frac{a_2a_1}{(1-a_2^2)(1-a_1^2)} & \frac{\alpha^2}{1-a_2^2a_1^2} \end{bmatrix} \\
&= \frac{1}{1-a_2^2a_1^2}\begin{bmatrix} 1 & -a_1a_2 \\ -a_1a_2 & 1 \end{bmatrix}.
\end{aligned}$$

In particular, the case $n = 2$ is able to recover all two-dimensional copulas by setting the correlation coefficient $\rho = a_1a_2$. In the general case, Σ is parametrized by n numbers, which offers less degrees of freedom compared with the joint Gaussian copula correlation method which relies on $n(n-1)/2$ coefficients, see also Exercise 9.3.

Chapter 10

Exercise 10.1 By absence of arbitrage we have $(1-\alpha)\mathrm{e}^{r_dT} = \mathrm{e}^{rT}$, hence $\alpha = 1 - \mathrm{e}^{(r-r_d)T}$.

Exercise 10.2

a) The bond payoff $\mathbb{1}_{\{\tau > T-t\}}$ is discounted according to the risk-free rate, before taking expectation.
b) We have $\mathbb{E}\big[\mathbb{1}_{\{\tau>T-t\}}\big] = \mathrm{e}^{-\lambda(T-t)}$, hence $P_d(t,T) = \mathrm{e}^{-(\lambda+r)(T-t)}$.
c) We have $P_M(t,T) = \mathrm{e}^{-(\lambda+r)(T-t)}$, hence $\lambda = -r - \frac{1}{T-t}\log P_M(t,T)$.

Exercise 10.3

a) We have

$$r_t = -a\int_0^t r_s ds + \sigma B_t^{(1)}, \quad t \geqslant 0,$$

hence

$$\begin{aligned}\int_0^t r_s ds &= \frac{1}{a}\big(\sigma B_t^{(1)} - r_t\big)\\ &= \frac{\sigma}{a}\left(B_t^{(1)} - \int_0^t \mathrm{e}^{-(t-s)a} dB_s^{(1)}\right)\\ &= \frac{\sigma}{a}\int_0^t (1-\mathrm{e}^{-(t-s)a}) dB_s^{(1)},\end{aligned}$$

and

$$\begin{aligned}&\int_t^T r_s ds = \int_0^T r_s ds - \int_0^t r_s ds\\ &= \frac{\sigma}{a}\int_0^T (1-\mathrm{e}^{-(T-s)a}) dB_s^{(1)} - \frac{\sigma}{a}\int_0^t (1-\mathrm{e}^{-(t-s)a}) dB_s^{(1)}\\ &= -\frac{\sigma}{a}\left(\int_0^t (\mathrm{e}^{-(T-s)a} - \mathrm{e}^{-(t-s)a}) dB_s^{(1)} + \int_t^T (\mathrm{e}^{-(T-s)a} - 1) dB_s^{(1)}\right)\\ &= -\frac{\sigma}{a}(\mathrm{e}^{-(T-t)a} - 1)\int_0^t \mathrm{e}^{-(t-s)a} dB_s^{(1)} - \frac{\sigma}{a}\int_t^T (\mathrm{e}^{-(T-s)a} - 1) dB_s^{(1)}\\ &= -\frac{1}{a}(\mathrm{e}^{-(T-t)a} - 1) r_t - \frac{\sigma}{a}\int_t^T (\mathrm{e}^{-(T-s)a} - 1) dB_s^{(1)}.\end{aligned}$$

The answer for λ_t is similar.

b) As a consequence of the answer to the previous question, we have

$$\mathbb{E}\left[\int_t^T r_s ds + \int_t^T \lambda_s ds \,\Big|\, \mathcal{F}_t\right] = C(a,t,T) r_t + C(b,t,T)\lambda_t,$$

and

$$\begin{aligned}&\mathrm{Var}\left[\int_t^T r_s ds + \int_t^T \lambda_s ds \,\Big|\, \mathcal{F}_t\right]\\ &= \mathrm{Var}\left[\int_t^T r_s ds \,\Big|\, \mathcal{F}_t\right] + \mathrm{Var}\left[\int_t^T \lambda_s ds \,\Big|\, \mathcal{F}_t\right]\\ &\quad +2\,\mathrm{Cov}\left(\int_t^T X_s ds, \int_t^T Y_s ds \,\Big|\, \mathcal{F}_t\right)\\ &= \frac{\sigma^2}{a^2}\int_t^T (\mathrm{e}^{-(T-s)a} - 1)^2 ds\\ &\quad +2\rho\frac{\sigma\eta}{ab}\int_t^T (\mathrm{e}^{-(T-s)a} - 1)(\mathrm{e}^{-(T-s)b} - 1) ds\\ &\quad +\frac{\eta^2}{b^2}\int_t^T (\mathrm{e}^{-(T-s)b} - 1)^2 ds\\ &= \sigma^2\int_t^T C^2(a,s,T) ds + 2\rho\sigma\eta\int_t^T C(a,s,T)C(b,s,T) ds\end{aligned}$$

$$+\eta^2 \int_t^T C^2(b, sT)ds,$$

from the Itô isometry.

Exercise 10.4 (Exercise 10.3 continued).

a) We use the fact that $(r_t, \lambda_t)_{t\in[0,T]}$ is a Markov process.
b) We use the tower property (A.33) of the conditional expectation given $\mathcal{F}_t$.

c) Writing $F(t, r_t, \lambda_t) = P(t, T)$, we have

$$\begin{aligned}
& d\left(\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} P(t,T)\right) \\
& = -(r_t+\lambda_t)\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} P(t,T)dt + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} dP(t,T) \\
& = -(r_t+\lambda_t)\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} P(t,T)dt + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} dF(t, r_t, \lambda_t) \\
& = -(r_t+\lambda_t)\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} P(t,T)dt + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial F}{\partial x}(t, r_t, \lambda_t) dr_t \\
& + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial F}{\partial y}(t, r_t, \lambda_t) d\lambda_t + \frac{1}{2}\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial^2 F}{\partial x^2}(t, r_t, \lambda_t)\sigma_1^2(t, r_t)dt \\
& + \frac{1}{2}\mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial^2 F}{\partial y^2}(t, r_t, \lambda_t)\sigma_2^2(t, \lambda_t)dt \\
& + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \rho \frac{\partial^2 F}{\partial x \partial y}(t, r_t, \lambda_t)\sigma_1(t, r_t)\sigma_2(t, \lambda_t)dt + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial F}{\partial t}(t, r_t, \lambda_t)dt \\
& = \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial F}{\partial x}(t, r_t, \lambda_t)\sigma_1(t, r_t)dB_t^{(1)} + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \frac{\partial F}{\partial y}(t, r_t, \lambda_t)\sigma_2(t, \lambda_t)dB_t^{(2)} \\
& + \mathrm{e}^{-\int_0^t (r_s+\lambda_s)ds} \left(-(r_t+\lambda_t)P(t,T) + \frac{\partial F}{\partial x}(t, r_t, \lambda_t)\mu_1(t, r_t) \right. \\
& + \frac{\partial F}{\partial y}(t, r_t, \lambda_t)\mu_2(t, \lambda_t) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, r_t, \lambda_t)\sigma_1^2(t, r_t) + \frac{1}{2}\frac{\partial^2 F}{\partial y^2}(t, r_t, \lambda_t)\sigma_2^2(t, \lambda_t) \\
& \left. + \rho \frac{\partial^2 F}{\partial x \partial y}(t, r_t, \lambda_t)\sigma_1(t, r_t)\sigma_2(t, \lambda_t) + \frac{\partial F}{\partial t}(t, r_t, \lambda_t) \right) dt,
\end{aligned}$$

hence the bond pricing PDE is

$$\begin{aligned}
& -(x+y)F(t,x,y) + \mu_1(t,x)\frac{\partial F}{\partial x}(t,x,y) \\
& + \mu_2(t,y)\frac{\partial F}{\partial y}(t,x,y) + \frac{1}{2}\sigma_1^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x,y) \\
& + \frac{1}{2}\sigma_2^2(t,y)\frac{\partial^2 F}{\partial y^2}(t,x,y) + \rho\sigma_1(t,x)\sigma_2(t,y)\frac{\partial^2 F}{\partial x \partial y}(t,x,y) + \frac{\partial F}{\partial t}(t, r_t, \lambda_t) = 0.
\end{aligned}$$

d) We have

$$\begin{aligned}
P(t,T) &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[\exp\left(-\int_t^T r_s ds - \int_t^T \lambda_s ds\right) \;\Big|\; \mathcal{F}_t\right] \\
&= \mathbb{1}_{\{\tau>t\}} \exp\left(-\mathbb{E}\left[\int_t^T r_s ds \;\Big|\; \mathcal{F}_t\right] - \mathbb{E}\left[\int_t^T \lambda_s ds \;\Big|\; \mathcal{F}_t\right]\right) \\
&\quad \times \exp\left(\frac{1}{2}\,\mathrm{Var}\left[\int_t^T r_s ds + \int_t^T \lambda_s ds \;\Big|\; \mathcal{F}_t\right]\right) \\
&= \mathbb{1}_{\{\tau>t\}} \exp\left(-C(a,t,T)r_t - C(b,t,T)\lambda_t\right) \\
&\quad \times \exp\left(\frac{\sigma^2}{2}\int_t^T C^2(a,s,T)ds + \frac{\eta^2}{2}\int_t^T C^2(b,s,T)\mathrm{e}^{-(T-s)b}ds\right) \\
&\quad \times \exp\left(\rho\sigma\eta\int_t^T C(a,s,T)C(b,s,T)ds\right).
\end{aligned}$$

e) This is a direct consequence of the answers to Questions (c)) and (d)).

f) The above analysis shows that

$$\begin{aligned}
\mathbb{P}(\tau > T \mid \mathcal{G}_t) &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[\exp\left(-\int_t^T \lambda_s ds\right) \;\Big|\; \mathcal{F}_t\right] \\
&= \mathbb{1}_{\{\tau>t\}} \exp\left(-C(b,t,T)\lambda_t + \frac{\eta^2}{2}\int_t^T C^2(b,s,T)ds\right),
\end{aligned}$$

for $a = 0$ and

$$\mathbb{E}\left[\exp\left(-\int_t^T r_s ds\right) \;\Big|\; \mathcal{F}_t\right] = \exp\left(-C(a,t,T)r_t + \frac{\sigma^2}{2}\int_t^T C^2(a,s,T)ds\right),$$

for $b = 0$, and this implies

$$\begin{aligned}
U_\rho(t,T) &= \exp\left(\rho\sigma\eta\int_t^T C(a,s,T)C(b,s,T)ds\right) \\
&= \exp\left(\rho\frac{\sigma\eta}{ab}\left(T - t - C(a,t,T) - C(b,t,T) + C(a+b,t,T)\right)\right).
\end{aligned}$$

g) We have

$$\begin{aligned}
f(t,T) &= -\mathbb{1}_{\{\tau>t\}}\frac{\partial}{\partial T}\log P(t,T) \\
&= \mathbb{1}_{\{\tau>t\}}\left(r_t\mathrm{e}^{-(T-t)a} - \frac{\sigma^2}{2}C^2(a,t,T) + \lambda_t\mathrm{e}^{-(T-t)b} - \frac{\eta^2}{2}C^2(b,t,T)\right) \\
&\quad -\mathbb{1}_{\{\tau>t\}}\rho\sigma\eta C(a,t,T)C(b,t,T).
\end{aligned}$$

h) We use the relation

$$\begin{aligned}\mathbb{P}(\tau > T \mid \mathcal{G}_t) &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}\left[\exp\left(-\int_t^T \lambda_s ds\right) \ \Big|\ \mathcal{F}_t\right] \\ &= \mathbb{1}_{\{\tau>t\}} \exp\left(-C(b,t,T)\lambda_t + \frac{\eta^2}{2}\int_t^T C^2(b,s,T)ds\right) \\ &= \mathbb{1}_{\{\tau>t\}} \mathrm{e}^{-\int_t^T f_2(t,u)du},\end{aligned}$$

where $f_2(t,T)$ is the Vasicek forward rate corresponding to λ_t, *i.e.*

$$f_2(t,u) = \lambda_t \mathrm{e}^{-(u-t)b} - \frac{\eta^2}{2} C^2(b,t,u).$$

i) In this case we have $\rho = 0$ and

$$P(t,T) = \mathbb{P}(\tau > T \mid \mathcal{G}_t)\mathbb{E}\left[\exp\left(-\int_t^T r_s ds\right) \ \Big|\ \mathcal{F}_t\right],$$

since $U_\rho(t,T) = 0$.

Chapter 11

Exercise 11.1 It suffices to check that as λ tends to ∞, the ratio

$$\frac{S_{T_i}^{i,j} \displaystyle\sum_{k=i}^{j-1} \delta_k \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right)}{(1-\xi)\displaystyle\sum_{k=i}^{j-1} \left(1 - \mathrm{e}^{-\lambda\delta_k}\right) \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right)}$$

converges to 0, while it tends to $+\infty$ as λ goes to 0. Therefore, the equation (11.4) admits a numerical solution.

Exercise 11.2 Equation (11.4) reads

$$\begin{aligned}&S_{T_i}^{i,j} \sum_{k=i}^{j-1} \delta_k \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right) \\ &\quad = (1-\xi)\sum_{k=i}^{j-1} \left(\mathrm{e}^{\lambda\delta_k} - 1\right) \exp\left(-\sum_{p=i}^{k} \delta_p r_p - \lambda \sum_{p=i}^{k} \delta_p\right),\end{aligned}$$

or

$$S_{T_i}^{i,j}\sum_{k=i}^{j-1}\delta_k P(0,T_{k+1})\exp\left(-\lambda\sum_{p=i}^{k}\delta_p\right)$$
$$=(1-\xi)\sum_{k=i}^{j-1}\big(\mathrm{e}^{\lambda\delta_k}-1\big)P(0,T_{k+1})\exp\left(-\lambda\sum_{p=i}^{k}\delta_p\right),$$

since

$$P(0,T_{k+1})=\exp\left(-\sum_{p=i}^{k}\delta_p r_p\right),\quad k=0,1,2.$$

From the terminal data of Figure 11.6 we infer with

$$\begin{cases} i=0,\\ j=3,\\ T_0=03/20/2015,\\ T_1=06/22/2015,\\ t=04/12/2015,\\ T_2=09/21/2015,\\ T_3=12/21/2015,\\ \delta_1=\delta_2=\delta_3=0.25,\\ \xi=0.4,\\ S_{T_1}^{1,3}=0.1079.\end{cases}$$

Hence, from the data of Figure S.14, Equation (11.4) rewrites as

$$\begin{aligned}&0.1079\times 0.25\\ &\times\big(0.99952277\times\mathrm{e}^{-\lambda\times 0.25}+0.99827639\times\mathrm{e}^{-\lambda\times 0.5}+0.99607821\times\mathrm{e}^{-\lambda\times 0.75}\big)\\ &=(1-0.4)\times\big(\mathrm{e}^{\lambda\times 0.25}-1\big)\\ &\times\big(0.99952277\times\mathrm{e}^{-\lambda\times 0.25}+0.99827639\times\mathrm{e}^{-\lambda\times 0.5}+0.99607821\times\mathrm{e}^{-\lambda\times 0.75}\big),\end{aligned}$$

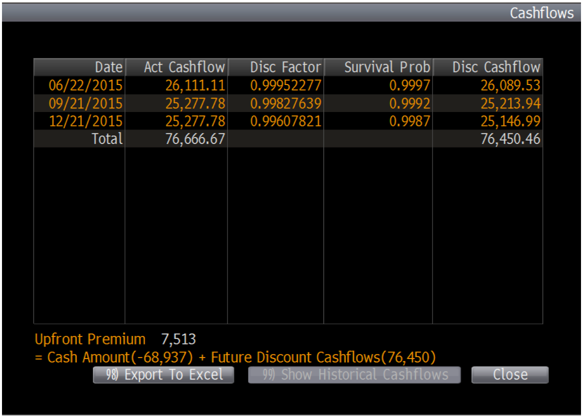
with solution $\lambda=0.0017987468$. The default probability is given by $p=1-\mathrm{e}^{-\lambda\times 0.75}=0.001348151$.

Next, from the discount factors of Figure S.14 we solve the Equation (11.4) numerically in Table S.1 below to find the default rate $\lambda_1=0.0017987468$ and default probability 0.0012460256, which is consistent with the value of 0.0013 in Figure 11.6, see also Castellacci (2008).

This version: May 3, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Date	Delta	Discount Factor	Premium Leg	Protection Leg
Jun 22, 2015	0.2611111	0.99952277	0.0002814722	0.0002814708
Sep 21, 2015	0.2527778	0.99827639	0.0002721533	0.000272154
Dec 21, 2015	0.2527778	0.99607821	0.0002715541	0.0002715548
		Sum	0.0008251796	0.0008251796

Table S.1: CDS Market data.


Cashflows

Date	Act Cashflow	Disc Factor	Survival Prob	Disc Cashflow
06/22/2015	26,111.11	0.99952277	0.9997	26,089.53
09/21/2015	25,277.78	0.99827639	0.9992	25,213.94
12/21/2015	25,277.78	0.99607821	0.9987	25,146.99
Total	76,666.67			76,450.46

Upfront Premium 7,513
= Cash Amount(-68,937) + Future Discount Cashflows(76,450)
98) Export To Excel
99) Show Historical Cashflows
Close

Fig. S.14: CDS Price data.

Exercise 11.3

a) We have

$$\begin{aligned}
&\sum_{k=i}^{j-1} \mathbb{E}\left[\mathbb{1}_{(T_k,T_{k+1}]}(\tau)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r(s)ds\right) \,\Big|\, \mathcal{G}_t\right] \\
&= \sum_{k=i}^{j-1} \mathbb{E}\left[\left(\mathbb{1}_{\{T_k<\tau\}} - \mathbb{1}_{\{T_{k+1}<\tau\}}\right)(1-\xi_{k+1})\exp\left(-\int_t^{T_{k+1}} r(s)ds\right) \,\Big|\, \mathcal{G}_t\right] \\
&= \mathbb{1}_{\{\tau>t\}} \sum_{k=i}^{j-1} \mathbb{E}\left[(1-\xi_{k+1})\left(\mathrm{e}^{-\int_t^{T_k}\lambda_s ds} - \mathrm{e}^{-\int_t^{T_{k+1}}\lambda_s ds}\right)\mathrm{e}^{-\int_t^{T_{k+1}} r(s)ds} \,\Big|\, \mathcal{F}_t\right]
\end{aligned}$$

$$= \mathbb{1}_{\{\tau>t\}}(1-\xi)\sum_{k=i}^{j-1} \mathrm{e}^{-\int_t^{T_{k+1}} r(s)ds}\mathbb{E}\left[\mathrm{e}^{-\int_t^{T_k}\lambda_s ds} - \mathrm{e}^{-\int_t^{T_{k+1}}\lambda_s ds} \ \Big|\ \mathcal{F}_t\right]$$
$$= \mathbb{1}_{\{\tau>t\}}(1-\xi)\sum_{k=i}^{j-1} P(t,T_{k+1})\left(Q(t,T_k)-Q(t,T_{k+1})\right).$$

b) We have

$$\begin{aligned}
V^p(t,T) &= S_t^{i,j}\sum_{k=i}^{j-1}\delta_k\mathbb{E}\left[\mathbb{1}_{\{\tau>T_{k+1}\}}\exp\left(-\int_t^{T_{k+1}} r(s)ds\right)\ \Big|\ \mathcal{G}_t\right]\\
&= S_t^{i,j}\sum_{k=i}^{j-1}\delta_k\mathbb{E}\left[\mathbb{1}_{\{T_{k+1}<\tau\}}\exp\left(-\int_t^{T_{k+1}} r(s)ds\right)\ \Big|\ \mathcal{G}_t\right]\\
&= \mathbb{1}_{\{\tau>t\}}S_t^{i,j}\sum_{k=i}^{j-1}\delta_k\mathbb{E}\left[\exp\left(-\int_t^{T_{k+1}}\lambda_s ds\right)\exp\left(-\int_t^{T_{k+1}} r(s)ds\right)\ \Big|\ \mathcal{F}_t\right]\\
&= S_t^{i,j}\mathbb{1}_{\{\tau>t\}}\sum_{k=i}^{j-1}\delta_k\exp\left(-\int_t^{T_{k+1}} r(s)ds\right)\mathbb{E}\left[\exp\left(-\int_t^{T_{k+1}}\lambda_s ds\right)\ \Big|\ \mathcal{F}_t\right]\\
&= \mathbb{1}_{\{\tau>t\}}S_t^{i,j}\sum_{k=i}^{j-1}\delta_k P(t,T_{k+1})Q(t,T_{k+1}).
\end{aligned}$$

c) By equating the protection and premium legs, we find

$$\begin{aligned}
(1-\xi)\sum_{k=i}^{j-1} P(t,T_{k+1})\left(Q(t,T_k)-Q(t,T_{k+1})\right)&\\
&= S_t^{i,j}\sum_{k=i}^{j-1}\delta_k P(t,T_{k+1})Q(t,T_{k+1}).
\end{aligned}$$

For $j=i+1$, this yields

$$(1-\xi)P(t,T_{i+1})\left(Q(t,T_i)-Q(t,T_{i+1})\right) = S_t^{i,i+1}\delta_i P(t,T_{i+1})Q(t,T_{i+1}),$$

hence

$$Q(t,T_{i+1}) = \frac{1-\xi}{S_t^{i,i+1}\delta_i+1-\xi},$$

with $Q(t,T_i)=1$, and the recurrence relation

$$(1-\xi)P(t,T_{j+1})\left(Q(t,T_j)-Q(t,T_{j+1})\right)$$

$$
\begin{aligned}
&+ (1-\xi) \sum_{k=i}^{j-1} P(t, T_{k+1}) \left(Q(t, T_k) - Q(t, T_{k+1}) \right) \\
&= S_t^{i,j} \delta_j P(t, T_{j+1}) Q(t, T_{j+1}) + S_t^{i,j} \sum_{k=i}^{j-1} \delta_k P(t, T_{k+1}) Q(t, T_{k+1}),
\end{aligned}
$$

i.e.

$$
\begin{aligned}
Q(t, T_{j+1}) &= \frac{(1-\xi) Q(t, T_j)}{1 - \xi + S_t^{i,j} \delta_j} \\
&+ \sum_{k=i}^{j-1} \frac{P(t, T_{k+1}) \big((1-\xi) Q(t, T_k) - Q(t, T_{k+1}) \big((1-\xi) + \delta_k S_t^{i,j} \big) \big)}{P(t, T_{j+1}) (1 - \xi + S_t^{i,j} \delta_j)}.
\end{aligned}
$$

Exercise 11.4 (Exercise 11.3 continued). From the terminal data of Figure 11.7, we find the following spread data and survival probabilities:

k	Maturity	T_k	$S_t^{1,k}$ (bp)	$Q(t, T_k)$
1	6M	0.5	10.97	0.999087
2	1Y	1	12.25	0.997961
3	2Y	2	14.32	0.995235
4	3Y	3	19.91	0.990037
5	4Y	4	26.48	0.982293
6	5Y	5	33.29	0.972122
7	7Y	7	52.91	0.937632
8	10Y	10	71.91	0.880602

Table S.2: Spread and survival probabilities.

Background on Probability Theory

Exercise A.1

a) We have

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k \geqslant 0} k \mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \sum_{k \geqslant 0} k \frac{\lambda^k}{k!} \\
&= \mathrm{e}^{-\lambda} \sum_{k \geqslant 1} \frac{\lambda^k}{(k-1)!} = \lambda \mathrm{e}^{-\lambda} \sum_{k \geqslant 0} \frac{\lambda^k}{k!} = \lambda.
\end{aligned}
$$

b) We have

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k\geqslant 0} k^2 \mathbb{P}(X=k) \\
&= \mathrm{e}^{-\lambda} \sum_{k\geqslant 1} k^2 \frac{\lambda^k}{k!} \\
&= \mathrm{e}^{-\lambda} \sum_{k\geqslant 1} k \frac{\lambda^k}{(k-1)!} \\
&= \mathrm{e}^{-\lambda} \sum_{k\geqslant 2} \frac{\lambda^k}{(k-2)!} + \mathrm{e}^{-\lambda} \sum_{k\geqslant 1} \frac{\lambda^k}{(k-1)!} \\
&= \lambda^2 \mathrm{e}^{-\lambda} \sum_{k\geqslant 0} \frac{\lambda^k}{k!} + \lambda \mathrm{e}^{-\lambda} \sum_{k\geqslant 0} \frac{\lambda^k}{k!} \\
&= \lambda^2 + \lambda,
\end{aligned}$$

and

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda = \mathbb{E}[X].$$

Exercise A.2 We have

$$\begin{aligned}
\mathbb{P}(\mathrm{e}^X > c) = \mathbb{P}(X > \log c) &= \int_{\log c}^\infty \mathrm{e}^{-y^2/(2\eta^2)} \frac{dy}{\sqrt{2\pi\eta^2}} \\
&= \int_{(\log c)/\eta}^\infty \mathrm{e}^{-y^2/2} \frac{dy}{\sqrt{2\pi}} = 1 - \Phi((\log c)/\eta) = \Phi(-(\log c)/\eta).
\end{aligned}$$

Exercise A.3

a) Using the change of variable $z = (x-\mu)/\sigma$, we have

$$\begin{aligned}
\int_{-\infty}^\infty \varphi(x) dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty \mathrm{e}^{-y^2/(2\sigma^2)} dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \mathrm{e}^{-z^2/2} dz.
\end{aligned}$$

Next, using the polar change of coordinates $dxdy = r dr d\theta$, we find[†]

[†] "In a discussion with Grothendieck, Messing mentioned the formula expressing the integral of e^{-x^2} in terms of π, which is proved in every calculus course. Not only did

$$\begin{aligned}
\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-z^2/2}dz\right)^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}dy\int_{-\infty}^{\infty}\mathrm{e}^{-z^2/2}dz\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-(y^2+z^2)/2}dydz\\
&= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty}r\mathrm{e}^{-r^2/2}drd\theta\\
&= \int_{0}^{\infty}r\mathrm{e}^{-r^2/2}dr\\
&= \lim_{R\to+\infty}\int_{0}^{R}r\mathrm{e}^{-r^2/2}dy\\
&= -\lim_{R\to+\infty}\left[\mathrm{e}^{-r^2/2}\right]_0^R\\
&= \lim_{R\to+\infty}(1-\mathrm{e}^{-R^2/2})\\
&= 1,
\end{aligned}$$

or

$$\int_{-\infty}^{\infty}\mathrm{e}^{-z^2/2}dz = \sqrt{2\pi}.$$

b) We have

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty}x\varphi(x)dx\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}x\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}dx\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(\mu+y)\mathrm{e}^{-y^2/(2\sigma^2)}dx\\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}dy+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y\mathrm{e}^{-y^2/2}dy\\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}dy+\frac{\sigma}{\sqrt{2\pi}}\lim_{A\to+\infty}\int_{-A}^{A}y\mathrm{e}^{-y^2/2}dy\\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}dy\\
&= \mu\int_{-\infty}^{\infty}\varphi(y)dy\\
&= \mu\mathbb{P}(X\in\mathbb{R})\\
&= \mu,
\end{aligned}$$

by symmetry of the function $y \longmapsto y\mathrm{e}^{-y^2/2}$ on $\mathbb{R}$.

c) Similarly, by integration by parts twice on $\mathbb{R}$, we find

Grothendieck not know the formula, but he thought that he had never seen it in his life". Milne (2005).

$$\begin{aligned}
\mathbb{E}[(X-\mathbb{E}[X])^2] &= \int_{-\infty}^{\infty}(x-\mu)^2\varphi(x)dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} y^2\mathrm{e}^{-(y-\mu)^2/(2\sigma^2)}dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y\times y\mathrm{e}^{-y^2/2}dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-y^2/2}dy \\
&= \sigma^2.
\end{aligned}$$

d) By a completion of squares argument, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{e}^X] &= \int_{-\infty}^{\infty}\mathrm{e}^x\varphi(x)dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{e}^{x-(x-\mu)^2/(2\sigma^2)}dx \\
&= \frac{\mathrm{e}^\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{e}^{y-y^2/(2\sigma^2)}dy \\
&= \frac{\mathrm{e}^\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{e}^{\sigma^2/2+(y-\sigma^2)^2/(2\sigma^2)}dy \\
&= \frac{\mathrm{e}^\mu+\sigma^2/2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{e}^{x^2/(2\sigma^2)}dy \\
&= \mathrm{e}^\mu+\frac{\sigma^2}{2}.
\end{aligned}$$

Exercise A.4

a) We have

$$\begin{aligned}
\mathbb{E}[X^+] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}x^+\mathrm{e}^{-x^2/(2/\sigma^2)}dx \\
&= \frac{\sigma}{\sqrt{2\pi}}\int_0^\infty x\mathrm{e}^{-x^2/2}dx \\
&= \frac{\sigma}{\sqrt{2\pi}}\left[-\mathrm{e}^{-x^2/2}\right]_{x=0}^{x=\infty} \\
&= \frac{\sigma}{\sqrt{2\pi}}.
\end{aligned}$$

b) We have

$$\mathbb{E}[(X-K)^+] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(x-K)^+\mathrm{e}^{-x^2/(2\sigma^2)}dx$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_K^\infty (x-K)\mathrm{e}^{-x^2/(2\sigma^2)}dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_K^\infty x\mathrm{e}^{-x^2/(2\sigma^2)}dx - \frac{K}{\sqrt{2\pi\sigma^2}} \int_K^\infty \mathrm{e}^{-x^2/(2\sigma^2)}dx \\
&= \frac{\sigma}{\sqrt{2\pi}} \left[-\mathrm{e}^{-x^2/(2\sigma^2)}\right]_{x=K}^\infty - \frac{K}{\sqrt{2\pi}} \int_{-\infty}^{-K/\sigma} \mathrm{e}^{-x^2/2}dx \\
&= \frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-K^2/(2\sigma)^2} - K\Phi\left(-\frac{K}{\sigma}\right).
\end{aligned}$$

c) Similarly, we have

$$\begin{aligned}
\mathbb{E}[(K-X)^+] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty (K-x)^+\mathrm{e}^{-x^2/(2\sigma^2)}dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^K (K-x)\mathrm{e}^{-x^2/(2\sigma^2)}dx \\
&= \frac{K}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^K \mathrm{e}^{-x^2/(2\sigma^2)}dx - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^K x\mathrm{e}^{-x^2/(2\sigma^2)}dx \\
&= \frac{K}{\sqrt{2\pi}} \int_{-\infty}^{K/\sigma} \mathrm{e}^{-x^2/2}dx - \frac{\sigma}{\sqrt{2\pi}} \left[-\mathrm{e}^{-x^2/(2\sigma^2)}\right]_{-\infty}^{x=K} \\
&= \frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-K^2/(2\sigma)^2} + K\Phi\left(\frac{K}{\sigma}\right).
\end{aligned}$$

Index

Author index

References

[1] K. Aas. The Basel II IRB approach for credit portfolios: A survey. Note No SAMBA/33/05, Norwegian Computing Center, 2005. (Cited on page 263).

[2] C. Acerbi and D. Tasche. On the coherence of expected shortfall. Preprint arXiv:cond-mat/0104295, 2001. (Cited on page 184).

[3] Aristotle. *Politics, Book one, Part XI*. The Internet Classics Archive, 350 BCE. http://classics.mit.edu/Aristotle/politics.1.one.html. (Cited on page 125).

[4] L. Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure, Série 3*, 17:21–86, 1900. (Cited on pages 5 and 13).

[5] F. Black and J.C. Cox. Valuing corporate securities. *Journal of Finance*, 31:351–357, 1976. (Cited on page 228).

[6] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. of Political Economy*, 81, 1973. (Cited on page 126).

[7] D. Bosq and H.T. Nguyen. *A Course in Stochastic Processes: Stochastic Models and Statistical Inference*. Mathematical and Statistical Methods. Kluwer, 1996. (Cited on page 77).

[8] K.E. Boulding. *In "Energy Reorganization Act of 1973. Hearings, Ninety-third Congress, first session, on H.R. 11510"*. U.S. Government Printing Office, Washington, 1973. (Cited on page 14).

[9] D. Brigo and K. Chourdakis. Counterparty risk for credit default swaps: impact of spread volatility and default correlation. *Int. J. Theor. Appl. Finance*, 12(7):1007–1026, 2009. (Cited on page 266).

[10] D. Brigo and M. Masetti. Risk neutral pricing of counterparty risk. In M. Pykhtin, editor, *In Counter-party Credit Risk Modeling: Risk Management, Pricing and Regulation*, London, 2006. Risk Books. (Cited on page 266).

[11] P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991. (Cited on page 50).

[12] R. Brown. A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine*, 4:161–173, 1828. (Cited on page 5).

[13] G. Castellacci. Bootstrapping credit curves from CDS spread curves. Available at https://dx.doi.org/10.2139/ssrn.2042177, 2008. (Cited on pages 258, 267, and 372).

[14] C.V.L. Charlier. Frequency curves of type A in heterograde statistics. *Ark. Mat. Astr. Fysik*, 9(25):1–17, 1914. (Cited on page 31).

[15] R.-R. Chen and J.-Z. Huang. Credit spread bounds and their implications for credit risk modeling. Working paper, Rutgers University and Penn State University, 2001. (Cited on page 247).

[16] R.-R. Chen, X. Cheng, F.J. Fabozzi, and B. Liu. An explicit, multifactor credit default swap pricing model with correlated factors. *Journal of Financial and Quantitative Analysis*, 43(1):123–160, 2008. (Cited on page 247).

[17] E. Çınlar. *Probability and stochastics*, volume 261 of *Graduate Texts in Mathematics*. Springer, New York, 2011. (Cited on page 283).

[18] M. Clark. Trillion Dollar Bet. PBS Nova Documentaries, 2000. (Cited on page 14).

[19] H. Cramér. *Mathematical methods of statistics*. Princeton University Press, Princeton, NJ, 1946. (Cited on page 31).

[20] F. De Vylder. Numerical finite-time ruin probabilities by the Picard-Lefèvre formula. *Scand. Act. J.*, 2:97–105, 1999. (Cited on page 97).

[21] C. Dellacherie, B. Maisonneuve, and P.A. Meyer. *Probabilités et Potentiel*, volume 5. Hermann, 1992. (Cited on page 243).

[22] J. L. Devore. *Probability and Statistics for Engineering and the Sciences*. Duxbury Press, sixth edition edition, 2003. (Cited on page 273).

[23] M. Dozzi and P. Vallois. Level crossing times for certain processes without positive jumps. *Bull. Sci. Math.*, 121(5):355–376, 1997. (Cited on page 95).

[24] D. Duffie and K.J. Singleton. *Credit risk*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2003. (Cited on page 246).

[25] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17:549–560, 1905. (Cited on page 5).

[26] R.J. Elliott and M. Jeanblanc. Incomplete markets with jumps and informed agents. *Math. Methods Oper. Res.*, 50(3):475–492, 1999. (Cited on page 245).

[27] R.J. Elliott, M. Jeanblanc, and M. Yor. On models of default risk. *Math. Finance*, 10(2):179–195, 2000. (Cited on page 243).

[28] P. Embrechts and M. Hofert. A note on generalized inverses. *Math. Methods Oper. Res.*, 77(3):423–432, 2013. (Cited on pages 158 and 159).

[29] W. Enders. *Applied Econometric Time Series*. Wiley, 2009. (Cited on page 64).

[30] R.F. Engle and C.W.J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987. (Cited on page 64).

[31] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, second edition, 1999. (Cited on page 6).

[32] M. Gibson. Understanding the risk of synthetic CDOs. Federal Reserve Board Working paper, 2004. (Cited on page 233).

[33] C. Gourieroux and J. Jasiak. Chapter 10 - value at risk. In Y. Aït-Sahalia and L.P. Hansen, editors, *Handbook of Financial Econometrics: Tools and Techniques*, volume 1 of *Handbooks in Finance*, pages 553–615, San Diego, 2010. North-Holland. (Cited on page v).

[34] J.P. Gram. Über die Entwicklung reeller Funktionen in Reihen mittelst der Methode der kleinsten Quadraten. *J. Reine Angew. Math*, 94:41–73, 1883. (Cited on page 31).

[35] M.R. Grasselli and T.-R. Hurd. Credit risk modeling. Lecture notes, McMaster University, 2010. (Cited on pages 224 and 228).

[36] X. Guo, R. Jarrow, and C. Menn. A note on Lando's formula and conditional independence. Preprint, 2007. (Cited on pages 243 and 246).

[37] G.M. Gupton, C.C. Finger, and M. Bhatia. Creditmetrics: technical document. JP Morgan & Co., 1997. (Cited on page 233).

[38] J.D. Hamilton. A new approach to the economic analysis of non-stationary time series. *Econometrica*, 57:357–384, 1989. (Cited on page 54).

[39] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994. (Cited on page 54).

[40] M.R. Hardy. An introduction to risk measures for actuarial applications. Study Notes, 2006. (Cited on page 152).

[41] J. Hull and A. White. Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation. *Journal of Derivatives*, 12:8–23, 2004. (Cited on page 233).

[42] G. Ignatova, V. Kaishev, and R. Krachunov. An improved finite-time ruin probabilities formula and its Mathematica implementation. *Insurance Math. Econom.*, 29:375–386, 2001. (Cited on page 97).

[43] K. Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944. (Cited on page 6).

[44] K. Itô. On stochastic differential equations. *Mem. Amer. Math. Soc.*, No. 4:51, 1951. (Cited on page 6).

[45] J. Jacod. Grossissement initial, hypothèse (H') et théorème de Girsanov. In Th. Jeulin and M. Yor, editors, *Grossissements de filtrations: exemples et applications*, volume 1118 of *Lecture Notes in Mathematics*, pages 6–14. Springer-Verlag, Berlin, 1985. (Cited on page 245).

[46] J. Jacod and P. Protter. *Probability essentials*. Springer-Verlag, Berlin, 2000. (Cited on page 273).

[47] Th. Jeulin. *Semi-martingales et grossissement d'une filtration*, volume 833 of *Lecture Notes in Mathematics*. Springer Verlag, 1980. (Cited on page 245).

[48] S. Jones. The formula that felled Wall Street. Financial Times Magazine, April 24, 2009. (Cited on page 233).

[49] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory: Using R*. Springer, 2009. (Cited on page 95).

[50] O. Kallenberg. *Foundations of Modern Probability*. Probability and its Applications. Springer-Verlag, New York, second edition, 2002. (Cited on page 308).

[51] D. Lando. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120, 1998. (Cited on pages 242, 243, and 246).

[52] S. Lee. Formula from hell. Forbes.com, August 8, 2009. (Cited on page 233).

[53] J.A. León and J. Villa. On the distributions of the sup and inf of the classical risk process with exponential claim. *Commun. Stoch. Anal.*, 3 (1):69–84, 2009. (Cited on page 95).

[54] D.X. Li. On default correlation - a copula function approach. *The Journal of Fixed Income*, 9(4):43–54, 2000. (Cited on pages 232 and 234).

[55] W. Li and T. Krehbiel. An improved approach to evaluate default probabilities and default correlations with consistency. *Int. J. Theor. Appl. Finance*, 19(5), 2016. (Cited on page 233).

[56] S. Loisel and N. Privault. Sensitivity analysis and density estimation for finite-time ruin probabilities. *J. Comput. Appl. Math.*, 230(1):107–120, 2009. (Cited on page 97).

[57] R. C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974. (Cited on pages 223, 233, and 263).

[58] J.S. Milne. Mathematical apocrypha. https://www.jmilne.org/math/apocrypha.html, 2005. Accessed: 2020-06-12. (Cited on page 377).

[59] J. Mina and J.Y. Xiao. Return to RiskMetrics: The Evolution of a Standard. RiskMetrics, 2001. Introduction by Christopher C. Finger. (Cited on pages 170 and 187).

[60] MOE and UCLES. Mathematics (Syllabus 9758), Singapore-Cambridge General Certificate of Education Advanced Level Higher 2 (2022). https:

//www.seab.gov.sg/docs/default-source/national-examinations/syllabus/alevel/2022syllabus/9758_y22_sy.pdf, 2020. Accessed: 2022-12-28. (Cited on pages 50, 202, and 281).

[61] MOE and UCLES. Mathematics (Syllabus 4048), Singapore-Cambridge General Certificate of Education Ordinary Level (2022). https://www.seab.gov.sg/docs/default-source/national-examinations/syllabus/olevel/2022syllabus/4048_y22_sy.pdf, 2022. Accessed: 2024-04-20. (Cited on pages 50, 202, and 281).

[62] J.R. Norris. *Markov Chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. (Cited on page 81).

[63] Basel Committee on Banking Supervision. An explanatory note on the Basel II IRB risk weight functions. July, 2005. (Cited on page 263).

[64] P. Picard and C. Lefèvre. The probability of ruin in finite time with discrete claim size distribution. *Scand. Act. J.*, 1:58–69, 1997. (Cited on page 97).

[65] J. Pitman. *Probability*. Springer, 1999. (Cited on page 273).

[66] M. Pourahmadi. *Foundations of time series analysis and prediction theory*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2001. (Cited on page 50).

[67] N. Privault. *Understanding Markov Chains*. Springer Undergraduate Mathematics Series. Springer, second edition, 2018. (Cited on page 82).

[68] N. Privault. *Introduction to Stochastic Finance with Market Examples (2nd edition)*. Financial Mathematics Series. Chapman & Hall/CRC, 2022. (Cited on pages 229, 230, 242, 245, 246, 254, and 364).

[69] N. Privault and X. Wei. A Malliavin calculus approach to sensitivity analysis in insurance. *Insurance Math. Econom.*, 35(3):679–690, 2004. (Cited on page 97).

[70] N. Privault and X. Wei. Integration by parts for point processes and Monte Carlo estimation. *J. Appl. Probab.*, 44:806–823, 2007. (Cited on page 97).

[71] P. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, second edition, 2004. (Cited on page 243).

[72] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, 1994. (Cited on page 6).

[73] D. Rullière and S. Loisel. Another look at the Picard-Lefèvre formula for finite-time ruin probabilities. *Insurance Math. Econom.*, 35(2):187–203, 2004. (Cited on page 97).

[74] F. Salmon. Recipe for disaster: the formula that killed Wall Street. Wired Magazine, 23 February, 2009. (Cited on page 233).

[75] P.A. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 6(2):13–39, 1965. (Cited on page 13).

[76] R.L. Schilling and L. Partzsch. *Brownian motion*. De Gruyter Graduate. De Gruyter, Berlin, second edition, 2014. (Cited on page 8).

[77] S.E. Shreve. *Stochastic calculus for finance. II*. Springer Finance. Springer-Verlag, New York, 2004. Continuous-time models. (Cited on page 229).

[78] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959. (Cited on page 109).

[79] M. Sklar. Fonctions de répartition a n dimensions et leurs marges [republication of MR0125600]. *Ann. I.S.U.P.*, 54(1-2):3–6, 2010. With an introduction by Denis Bosq. (Cited on page 109).

[80] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, second edition, 2011. (Cited on page 306).

[81] S. Takayanagi and K. Ishikawa. *PairTrading: classical pair trading based on cointegration in finance*, 2017. R package version 1.1. (Cited on page 70).

[82] K. Tanaka, T. Yamada, and T. Watanabe. Applications of Gram-Charlier expansion and bond moments for pricing of interest rates and credit risk. *Quant. Finance*, 10(6):645–662, 2010. (Cited on page 31).

[83] H. Theil. *Economics and Information Theory*. North-Holland, 1967. (Cited on page 34).

[84] T.N. Thiele. On semi invariants in the theory of observations (Om Iagttagelseslærens Halvinvarianter). *Kjöbenhavn Overs.*, pages 135–141, 1899. (Cited on page 23).

[85] O. Vašiček. Probability of loss on loan portfolio. KMV Corporation Technical Report, 1987. (Cited on page 233).

[86] O. Vašiček. Loan portfolio value. *Risk Magazine*, pages 160–162, 2002. (Cited on page 263).

[87] N. Wiener. Differential space. *Journal of Mathematics and Physics of the Massachusetts Institute of Technology*, 2:131–174, 1923. (Cited on page 6).

[88] M. Yor. Grossissements de filtration et absolue continuité de noyaux. In Th. Jeulin and M. Yor, editors, *Grossissements de filtrations: exemples et applications*, volume 1118 of *Lecture Notes in Mathematics*, pages 6–14. Springer-Verlag, Berlin, 1985. (Cited on page 245).

Notes on Financial Risk and Analytics

This document gives a presentation of mathematical tools used for financial risk modeling and related analytics, and is divided into three parts. The first part focuses on stochastic modeling using diffusion processes (Chapter 1), time series (Chapter 2), jump processes and insurance risk (Chapter 3), and random dependence structures (Chapter 4). The second part covers classical risk measures, starting with the superhedging risk measure, which is constructed in Chapter 5 from basic financial derivatives. Value at risk (VaR) is considered in Chapter 6, followed by tail value at risk (TVaR), conditional tail expectation (CTE) and expected shortfall (ES) in Chapter 7. The third part deals with credit risk, starting with Chapter 8 on credit scoring. Chapter 9 on credit risk builds on the geometric Brownian motion model of Chapter 1 for the pricing of default bonds. The remaining Chapters 10 and 11 consider credit default via defaultable bonds, credit default swaps (CDS) and collateralized debt obligations (CDOs), and involve more advanced knowledge of stochastic processes. The concepts presented are illustrated by 95 coding examples in R and Python, and accompanied with 159 figures and 70 exercises with complete solutions.