Selected Topics from Portfolio Optimization

Including some topics on

Stochastic Optimization

Lecture Notes

Winter 2023/ 24

Alois Pichler

TECHNISCHE UNIVERSITÄT CHEMNITZ
Faculty of Mathematics

DRAFT
Version as of November 29, 2023

Preface and Acknowledgment

Le silence éternel de ces espaces infinis m'effraie.

Blaise Pascal, *Pensées*,
Fragment Transition n^{o} 7 / 8

The purpose of these lecture notes is to facilitate the content of the lecture and the course. From experience it is helpful and recommended to attend and follow the lectures in addition. These lecture notes do not cover the lectures completely.

I am indebted to Prof. Georg Ch. Pflug for numerous discussions in the area and significant support over years.

Please report mistakes, errors, violations of copyright, improvements or necessary completions.

Content: https://www.tu-chemnitz.de/mathematik/studium/module/2013/M16.pdf

Contents

[1]By Ruben Schlotter
[2]Also: Newsboy, or Newsvendor problem

1 Historical Milestones in Portfolio Optimization

Probability is the foundation of banking.

Francis Ysidro Edgeworth, 1845–1926,
Anglo-Irish philosopher and political
economist. *Edgeworth [1888]*

1.1 IN BANKING

- 1938: Bond Duration, Edgeworth [1888]
- 1952: ***Markowitz*** mean-variance framework
- 1963: Sharp's capital asset pricing model
- 1966: Multiple factor models
- 1973: Black & Scholes option pricing model, the "Greeks"
- 1988: Risk weighted assets for banks
- 1993: ***Value-at-Risk***
- 1994: ***Risk Metrics***
- 1997: Credit Metrics
- 1998: Integration of credit and market risk
- 1998: Risk Budgeting, the Basel Rules
- 2007: Basel II
- 2017: Basel III

1.2 IN INSURANCE

The natural business of insurance companies is concerned with Risk.

- Pricing of individual Contracts
- Reserving in the Portfolio
- The Cramér-Lundberg model
- Solvability
- Solvency II
- US and Canada Insurance Supervisory: **Conditional Tail Expectation**

Bibliography

P. Artzner, F. Delbaen, and D. Heath. Thinking coherently. *Risk*, 10:68–71, 1997. 41

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9:203–228, 1999. doi:10.1111/1467-9965.00068. 41

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization*. SIAM Series on Optimization. 2001. 14

R. I. Boţ, S.-M. Grad, and G. Wanka. *Duality in Vector Optimization*. Springer, 2009. doi:10.1007/978-3-642-02886-1. 85

C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions*. Number 580 in Lecture Notes in Mathematics. Springer, 1977. doi:10.1007/BFb0087685. URL `https://books.google.com/books?id=Fev0CAAAQBAJ`. 99

G. Cornuejols and R. Tütüncü. *Optimization Methods in Finance*. Cambridge University Press (CUP), 2006. doi:10.1017/cbo9780511753886. 65

D. Denneberg. *Non-additive measure and integral*, volume 27. Springer Science & Business Media, 1994. doi:10.1007/978-94-017-2434-0. 80

D. Dentcheva and A. Ruszczyński. Portfolio optimization with risk control by stochastic dominance constraints. In G. Infanger, editor, *Stochastic Programming*, volume 150 of *International series in Operations Research & Management Science*, chapter 9, pages 189–211. Springer Science+Business Media, LLC, 2011. doi:10.1007/978-1-4419-1642-6. 62

F. Y. Edgeworth. The mathematical theory of banking. *Journal of the Royal Statistical Society*, 51(1): 113–127, 1888. 9

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter Studies in Mathematics 27. Berlin, Boston: De Gruyter, 2004. ISBN 978-3-11-046345-3. doi:10.1515/9783110218053. URL `http://books.google.com/books?id=cL-bZSOrqWoC`. 48

C. Hess. Set-valued integration and set-valued probability theory: An overview. In E. Pap, editor, *Handbook of Measure Theory*, volume I, II of *Handbook of Measure Theory*, chapter 14, pages 617–673. Elsevier, 2002. doi:10.1016/B978-044450263-6/50015-4. 98

A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley series in probability and statistics. Wiley, Chichester, 2002. ISBN 978-0-471-49446-1. URL `https://books.google.com/books?id=a8uPRWteCeUC`. 61

G. Ch. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2014. ISBN 978-3-319-08842-6. doi:10.1007/978-3-319-08843-3. URL `https://books.google.com/books?id=q_VWBQAAQBAJ`. 69, 94

G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, River Edge, NJ, 2007. doi:10.1142/9789812708724. 17, 37, 49, 70

S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems Volume I: Theory, Volume II: Applications*, volume XXV of *Probability and its applications*. Springer, New York, 1998. doi:10.1007/b98893. 94

R. T. Rockafellar. *Conjugate Duality and Optimization*, volume 16. CBMS-NSF Regional Conference Series in Applied Mathematics. 16. Philadelphia, Pa.: SIAM, Society for Industrial and Applied Mathematics. VI, 74 p., 1974. doi:10.1137/1.9781611970524. 98

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer Nature Switzerland AG, 1997. doi:10.1007/978-3-642-02431-3. URL `https://books.google.com/books?id=w-NdOE5fD8AC`. 98

A. Shapiro. Time consistency of dynamic risk measures. *Operations Research Letters*, 40(6):436–439, 2012. doi:10.1016/j.orl.2012.08.007. 49

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. MOS-SIAM Series on Optimization. SIAM, third edition, 2021. doi:10.1137/1.9781611976595. 49, 89

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi:10.1017/CBO9780511802256. URL `http://books.google.com/books?id=UEuQEM5RjWgC`. 37

A. E. van Heerwaarden and R. Kaas. The Dutch premium principle. *Insurance: Mathematics and Economics*, 11:223–230, 1992. doi:10.1016/0167-6687(92)90049-H. 49

Introduction and Classification of Stochastic Programs 2

Universitäten sind gefährlicher als Handgranaten.

Ruhollah Chomeini, 1902–1989

We employ the usual axioms in probability theory and denote a probability space by

$$(\Omega, \mathcal{F}, P).$$

Typically, we denote random variables mapping to a state space Ξ by

$$\xi\colon \Omega \to \Xi$$

(or sometimes also $Y\colon \Omega \to \mathbb{R}$).

2.1 RELATIONS AND CONNECTIONS TO PORTFOLIO OPTIMIZATION: MARKOWITZ

See Markowitz, Section 3 below for details.

Definition 2.1. A portfolio $x^* \in \mathbb{R}^J$ (with J indicating the number of stocks) is *efficient* if it solves

$$\begin{aligned} \text{minimize}_{\text{ in } x\in\mathbb{R}^J}\ & \operatorname{var} x^\top \xi \\ \text{subject to }\ & \mathbb{E}\, x^\top \xi \geq \mu, \\ & \mathbb{1}^\top x \leq 1, \\ & (x \geq 0). \end{aligned} \tag{2.1}$$

2.2 ALTERNATIVE FORMULATIONS OF THE MARKOWITZ PROBLEM

Instead of Markowitz (2.1) one might consider the problem

$$\begin{aligned} \text{maximize }\ & \mathbb{E}\, x^\top \xi \\ \text{subjet to }\ & \operatorname{var} x^\top \xi \leq q, \\ & \mathbb{1}^\top x \leq 1, \\ & (x \geq 0). \end{aligned} \tag{2.2}$$

2.3 INVOLVING RISK FUNCTIONALS

2.3.1 Risk Neutral

... is about the expectation, as

$$\begin{aligned} \text{minimize}_{\text{in x}}\ & \mathbb{E}\, Q(x, \xi) \\ \text{subject to }\ & \mathbb{E}\, G_i(x, \xi) \leq 0,\ i = 1, \dots, k \end{aligned}$$

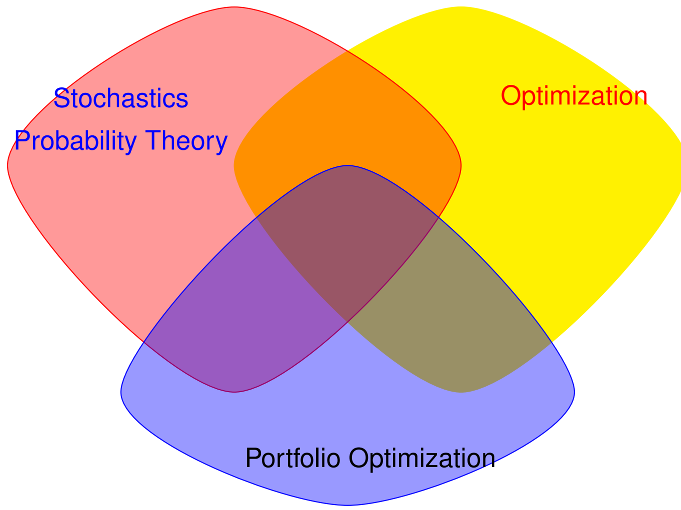


Figure 2.1: Wide intersections: in theory and (economic) practice

This may be reformulated by employing the risk functional

$$\mathcal{R}\big(Q(x,\cdot)\big) := \mathbb{E}_\xi\, Q(x,\xi).$$

2.3.2 Utility Functions

A function $u\colon \mathbb{R} \to \mathbb{R}$ is a utility function if it satisfies some model-design properties in addition. Optimization problems involving utility function generally read

$$\begin{aligned} \text{minimize}_{\text{in x}} \quad & \mathbb{E}\, u\big(Q(x,\xi)\big) \\ \text{subject to} \quad & \mathbb{E}\, G_i(x,\xi) \le 0,\ i = 1,\dots,k \end{aligned}$$

or

$$\begin{aligned} \text{minimize}_{\text{in x}} \quad & \mathcal{R}\big(Q(x,\xi)\big) \\ \text{subject to} \quad & \mathcal{R}\big(G_i(x,\xi)\big) \le 0,\ i = 1,\dots,k \end{aligned}$$

where we might want to put

$$\mathcal{R}\big(Q(x,\cdot)\big) := \mathbb{E}\, u\big(Q(x,\xi)\big).$$

2.3.3 Robust Optimization

Robust optimization considers the problem (Ben-Tal and Nemirovski [2001])

$$\begin{aligned} \text{minimize}_{\text{in x}} \quad & \max_{\xi\in\Xi} Q(x,\xi) \\ \text{subject to} \quad & \max_{\xi\in\Xi} G_i(x,\xi) \le 0,\ i = 1,\dots,k \end{aligned}$$

Note, that there is no probability measure

$$\mathcal{R}\big(Q(x,\cdot)\big) := \operatorname{ess\,sup} Q(x,\xi)$$

and the problem is basically about the *support* of the probability measure.

2.3.4 Distributionally Robust Optimization

Distributionally robust optimization involves the probability measure instead,

$$\begin{aligned} \text{minimize}_{\text{in x}}\ & \max_{P\in\mathscr{P}} \mathcal{R}_P\big(Q(x,\xi)\big),\\ \text{subject to } & \mathcal{R}_P\big(G_i(x,\xi)\big) \le 0,\ i=1,\dots,k, \end{aligned}$$

where we indicate the probability measure $P \in \mathscr{P}$ explicitly.

2.4 PROBABILISTIC CONSTRAINTS

This is about the problem

$$\begin{aligned} \text{minimize (in } x)\ & \mathcal{R}\big(Q(x,\xi)\big)\\ \text{subject to } & P\big(G_i(x,\xi) \le 0\big) \ge \alpha,\ i=1,\dots,k \end{aligned}$$

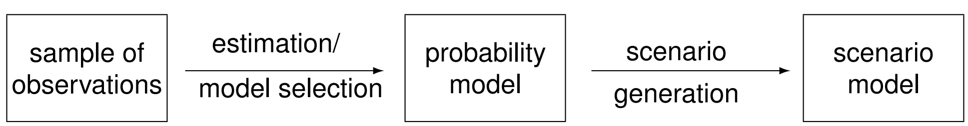
Example 2.2 (Economic example)**.** Call Center

2.5 STOCHASTIC DOMINANCE

$$\begin{aligned} \text{minimize}_{\text{in x}}\ & \mathbb{E}\, u\big(Q(x,\xi)\big)\\ \text{subject to } & G_i(x,\xi) \succcurlyeq Y,\ i=1,\dots,k \end{aligned} \tag{2.3}$$

for some (stochastic) order relation $\succcurlyeq$.

2.6 ON GENERAL DIFFICULTIES IN STOCHASTIC OPTIMIZATION



Problem 2.3 (Accuracy)**.** How large/ small do we need $\varepsilon > 0$?

The Markowitz Model

3

The trend is your friend.

Börsenweisheit

This section follows Pflug and Römisch [2007, Section 4].

The model of Markowitz[1] is historically the first model to determine the decomposition of an optimal portfolio.

3.1 INTRODUCTION

For portfolio optimization people often use the simple model provided by Harry Markowitz which involves the variance. Note that it is a *significant drawback* of the Markowitz model that positive deviations (profits – this is what the investor wants) and negative deviations (losses – this is what the investor tries to avoid) are treated exactly the same way. So the Markowitz model is of historical interest (it was the first model on asset allocation with the objective to reduce the variance) but it violates some natural objectives of an investor.

Extensions of the problem described at the end of this section avoid this downside.

3.2 EMPIRICAL PROBLEM FORMULATION AND VARIABLES

(i) J is the number of stocks considered ($J = 5$ in the example which Table 3.1 displays);

(ii) each stock $j \in \{1, \dots, J\}$ is observed at $n + 1$ consecutive times $t_0, \dots, t_n$ ($n = 12$ in Table 3.1);

(iii) the price observed of stock j at time t is S_t^j;

(iv) $\xi_i = \left(\xi_i^1, \xi_i^2, \dots \xi_i^J\right)^\top$ collects the annualized returns of all J stocks; note that $\xi_i^j = e_j^\top \xi_i$;

(v) x_j represents the fraction of cash invested in stock j, $j \in \{1, \dots, J\}$; we set $x := (x_1, x_2, \dots, x_J)^\top$, x is the allocation vector;

(vi) The budget constraint: the total amount of cash to be invested is not more than the budget available. 1€ is the default value (or $1m$€, say): the budget constraint thus reads $x^\top \mathbb{1} \leq 1$€, where $\mathbb{1} = (1, \dots, 1)^\top$;

(vii) Short-selling constraint: occasionally we do not allow short-selling (i.e., negative positions), then the constraints $x \geq 0$ has to be added. $x \geq 0$ is understood as $x_j \geq 0$, $j \in \{1, \dots, J\}$, for each stock.

To solve the problem one needs to specify the probability measure P which is used to compute the expectation $\mathbb{E}$ and the variance var.

[1]Harry Max Markowitz. 1927. Nobel Memorial Price in Economic Sciences in 1990

$S_t/€$		DAX	RWE	gold	oil	US-$/ €
t_0	January	9798.11	12.870	981.49	34.9	0.9228
t_1	February	9495.40	10.540	1030.49	33.08	0.9197
t_2	March	9965.51	11.375	1144.73	33.77	0.8787
t_3	April	10038.97	13.045	1137.17	37.04	0.8729
t_4	May	10262.74	11.765	1192.35	43.71	0.8983
t_5	June	9680.09	14.190	1121.85	45.81	0.9005
t_6	July	10337.50	15.905	1222.25	46.04	0.8949
t_7	August	10592.69	14.665	1244.67	39.78	0.8962
t_8	September	10511.02	15.335	1204.53	43.35	0.8896
t_9	October	10665.01	14.460	1212.10	46.15	0.9107
t_{10}	November	10640.30	11.860	1180.53	44.95	0.9444
t_{11}	December	11481.06	11.815	1082.17	47.72	0.9509
t_{12}	January	11599.01	11.800	1081.43	52.69	0.9494

Table 3.1: Prices $S_{t_i}^j$ observed in 2016. www.investing.com

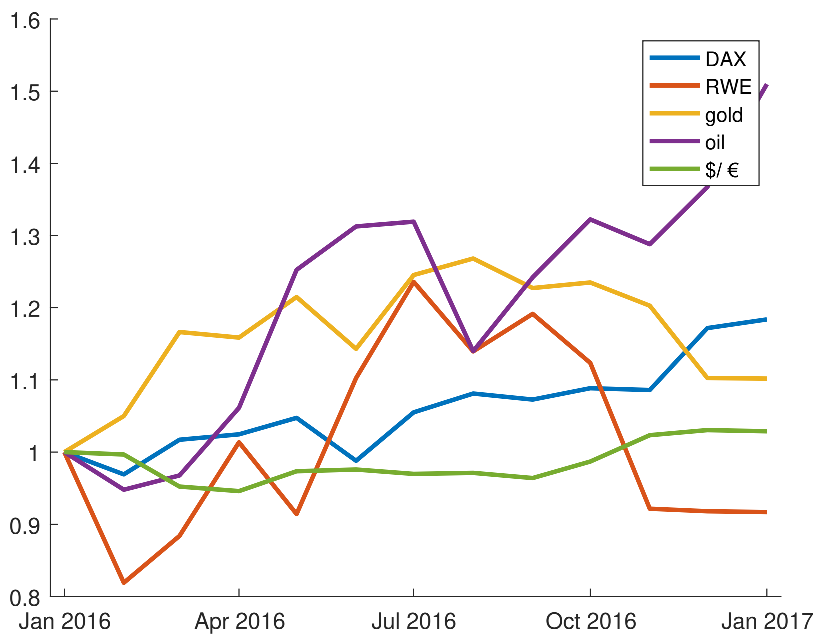


Figure 3.1: Prices of Table 3.1

3.3 THE EMPIRICAL/ DISCRETE MODEL

Empirical models extract the probability model from historic observations.

For this observe a stock at $n+1$ successive times $(t_i)_{i=0}^n$ and collect the prices S_{t_i}.

Definition 3.1. Its *annualized return* during the time-period $[t_{i-1}, t_i]$ is[2]

$$\xi_i \coloneqq \frac{1}{t_i - t_{i-1}} \ln \frac{S_{t_i}}{S_{t_{i-1}}}.$$

Define the weights (probabilities) $p_i \coloneqq \frac{t_i - t_{i-1}}{t_n - t_0}$, a random variable $\xi \colon \Omega \to \mathbb{R}$ and a probability measure with

$$P(\xi = \xi_i) \coloneqq p_i = \frac{t_i - t_{i-1}}{t_n - t_0}$$

and we set

$$p^\top \coloneqq (p_1, \dots, p_n) \coloneqq \left(\frac{t_1 - t_0}{t_n - t_0}, \frac{t_2 - t_1}{t_n - t_0}, \dots, \frac{t_n - t_{n-1}}{t_n - t_0} \right).$$

Remark 3.2. Note that $\sum_{i=1}^n p_i = p^\top \cdot \mathbb{1} = \sum_{i=1}^n \frac{t_i - t_{i-1}}{t_n - t_0} = 1$, thus

$$\underbrace{\frac{1}{t_n - t_0} \ln \frac{S_{t_n}}{S_{t_0}}}_{\text{annual return}} = \frac{1}{t_n - t_0} \sum_{i=1}^n \ln \frac{S_{t_i}}{S_{t_{i-1}}} \tag{3.1}$$

$$= \sum_{i=1}^n \underbrace{\frac{t_i - t_{i-1}}{t_n - t_0}}_{p_i} \cdot \underbrace{\frac{1}{t_i - t_{i-1}} \ln \frac{S_{t_i}}{S_{t_{i-1}}}}_{\text{annualized return per period}} = \underbrace{\sum_{i=1}^n p_i \cdot \xi_i}_{\text{average of returns}} = \mathbb{E}_P\, \xi. \tag{3.2}$$

Based on this observation it follows that the *annual return* for the entire period is the expected value of the *annualized returns of successive periods* (cf. Table 3.2).

Definition 3.3 (The first moment)**.** The expected return is $r \coloneqq \mathbb{E}\, \xi$.

Obviously, one may observe all J stocks in parallel, at the same time. So put

$$\xi_i^j \coloneqq \frac{1}{t_i - t_{i-1}} \ln \frac{S_{t_i}^j}{S_{t_{i-1}}^j}, \qquad i = 1, \dots, n,\ j = 1, \dots, J$$

and set $\xi_i \coloneqq (\xi_i^1, \dots, \xi_i^J)$. Collect all observations in the $n \times J$-matrix

$$\Xi \coloneqq \left(\xi_i^j\right)_{i=1:n}^{j=1:J} = \left(\frac{1}{t_i - t_{i-1}} \ln \frac{S_{t_i}^j}{S_{t_{i-1}}^j} \right)_{i=1:n}^{j=1:J}$$

(cf. Table 3.2). For the random return vector we have that

$$P\left(\xi = \left(\xi_i^1, \dots, \xi_i^J\right)\right) = P\left(\xi = \Xi_i\right) = p_i$$

where Ξ_i is the ith row in the matrix Ξ). Note that the return of stock j rewrites for the empirical probability measure

$$P(\cdot) = \sum_{i=1}^n p_i \cdot \delta_{\left(\xi_i^1, \dots, \xi_i^J\right)}(\cdot),$$

[2] Here and always: logarithmus naturalis with basis $e = 2.718\ldots$

annualized monthly returns ξ_t^j	DAX	RWE	gold	oil	US-\$/ €
$p_1 = 1/12$ or $p_1 = 31/365$	-37.7%	-239.7%	58.5%	-64.3%	-4.0%
$p_2 = 1/12$ or $p_2 = 28/365$	58.0%	91.5%	126.2%	24.8%	-54.7%
$p_3 = 1/12$ or $p_3 = 31/365$	8.8%	164.4%	-8.0%	110.9%	-7.9%
$p_4 = 1/12$ or $p_4 = 30/365$	26.5%	-123.9%	56.9%	198.7%	34.4%
$p_5 = 1/12$ or $p_5 = 31/365$	-70.1%	224.9%	-73.1%	56.3%	2.9%
$p_6 = 1/12$ or $p_6 = 30/365$	78.8%	136.9%	102.9%	6.0%	-7.5%
$p_7 = 1/12$ or $p_7 = 31/365$	29.3%	-97.4%	21.8%	-175.4%	1.7%
$p_8 = 1/12$ or $p_8 = 31/365$	-9.3%	53.6%	-39.3%	103.1%	-8.9%
$p_9 = 1/12$ or $p_9 = 30/365$	17.5%	-70.5%	7.5%	75.1%	28.1%
$p_{10} = 1/12$ or $p_{10} = 31/365$	-2.8%	-237.9%	-31.7%	-31.6%	43.6%
$p_{11} = 1/12$ or $p_{11} = 30/365$	91.3%	-4.6%	-104.4%	71.8%	8.2%
$p_{12} = 1/12$ or $p_{12} = 31/365$	12.3%	-1.5%	-0.8%	118.9%	-1.9%
average of monthly returns, (3.2)	16.9%	-8.7%	9.7%	41.2%	2.8%
annual return, (3.1)	16.9%	-8.7%	9.7%	41.2%	2.8%

Table 3.2: Matrix Ξ, collecting the returns ξ_i^j, cf. Table 3.1

i.e., each stock is a random variable with $P\left(\xi^j = \xi_i^j\right) = p_i$, independently of j. For every j thus

$$\mathbb{E}\,\xi^j = \sum_{i=1}^n p_i\,\xi_i^j = p^\top \xi^j,$$

where

$$p^\top = (p_1, \dots, p_n) = \left(\frac{t_1 - t_0}{t_n - t_0}, \frac{t_2 - t_1}{t_n - t_0}, \dots, \frac{t_n - t_{n-1}}{t_n - t_0}\right).$$

Remark 3.4. It follows from the Taylor series expansion

$$\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots$$

for small x that

$$\xi_i^j = \frac{1}{t_i - t_{i-1}} \ln \frac{S_{t_i}^j}{S_{t_{i-1}}^j} = \frac{1}{t_i - t_{i-1}} \ln\left(1 + \frac{S_{t_i}^j}{S_{t_{i-1}}^j} - 1\right) \approx \frac{1}{t_i - t_{i-1}} \left(\frac{S_{t_i}^j}{S_{t_{i-1}}^j} - 1\right).$$

3.4 THE FIRST MOMENT: RETURN

Suppose an amount of x_j is invested in the stock j. Then the total return of the investment is

$$x^\top \xi = \sum_{j=1}^J x_j \xi^j.$$

Note, that $e_j^\top \xi = \xi_j$, where $e_j^\top = (\underbrace{0, \dots, 0}_{i-1\text{ times}}, 1, \underbrace{0, \dots, 0}_{J-i\text{ times}})$ is the j-th vector in the canonical basis.

Lemma 3.5. *The expected return is $r := \mathbb{E}\,\xi = p^\top \Xi$.*

The return observed of the portfolio in period i is $\sum_{j=1}^J \xi_i^j x_j = (\Xi \cdot x)_i$ (the ith line in the matrix $\Xi \cdot x$). Markowith adds the constraint

$$\mathbb{E}\,x^\top \xi = \sum_{i=1}^n p_i \left(\sum_{j=1}^J \xi_i^j x_j\right) = p^\top \Xi x = r^\top x \ge \mu,$$

		DAX	RWE	gold	oil	US-\$/ €
return	$r_j = \mathbb{E}\, e_j^\top \xi$	16.9%	-8.7%	9.7%	41.2%	2.8%
variance	$\Sigma_{jj} = \operatorname{var}(e_j^\top \xi)$	19.2%	208.8%	42.8%	88.9%	5.9%

Table 3.3: Return and variance (cf. Table 3.1)

which means, that a minimum return μ is required.

Remark 3.6. Suppose the total cash invested in stock i is C^i (i.e., C^j/S_0^j is the number of shares of stock j) with total initial cash $C_0 = \sum_{j=1}^J C^j$, then it is natural to define the fraction $x_j := \frac{C^j}{\sum_{j=1}^J C^j}$ so that $\sum_{j=1}^J x_j = 1$. The total portfolio value at time t then is $C_t = C_0 \cdot \sum_{j=1}^J x_j \frac{S_t^j}{S_0^j}$. Note, however, that $\sum_{j=1}^J x_j \sum_{i=1}^I S_{t_i}^j = C_{t_I}$, but $\sum_{j=1}^J x_j \sum_{i=1}^I \xi_j^{t_i} \neq \frac{C_{t_i}}{C_{t_0}}$.

3.5 THE SECOND MOMENT: RISK

The covariance is

$$\begin{aligned}
\operatorname{var} x^\top \xi &= \mathbb{E}\left(x^\top \xi - \mathbb{E}\, x^\top \xi\right)^2 = \mathbb{E}\left(x^\top \xi\right)^2 - \left(\mathbb{E}\, x^\top \xi\right)^2 = \\
&= \sum_{i=1}^n p_i \left(\sum_{j=1}^J \xi_i^j x_j\right)^2 - \left(\sum_{i=1}^n p_i\, \xi_i^j x_j\right)^2 \\
&= \sum_{i=1}^n p_i \left(\sum_{j=1}^J \sum_{j'=1}^J \xi_i^j x_j \cdot \xi_i^{j'} x_{j'}\right) - \sum_{i=1}^n p_i \left(\sum_{j=1}^J \xi_i^j x_j\right) \cdot \sum_{i'=1}^n p_{i'} \left(\sum_{j'=1}^J \xi_{i'}^{j'} x_{j'}\right) \\
&= \sum_{j=1}^J x_j \sum_{j'=1}^J x_{j'} \cdot \sum_{i=1}^n p_i \left(\xi_i^j \xi_i^{j'}\right) - \sum_{j=1}^J x_j \sum_{j'=1}^J x_{j'} \cdot \sum_{i=1}^n p_i \xi_i^j \sum_{i'=1}^n p_{i'} \xi_{i'}^{j'} \\
&= \sum_{j=1}^J x_j \sum_{j'=1}^J x_{j'} \cdot \underbrace{\left(\sum_{i=1}^n p_i \cdot \xi_i^j \xi_i^{j'} - \sum_{i=1}^n p_i \xi_i^j \cdot \sum_{i'=1}^n p_{i'} \xi_{i'}^{j'}\right)}_{=:\Sigma_{j,j'}}
\end{aligned}$$

and thus

$$\operatorname{var} x^\top \xi = \sum_{j=1}^J x_j \sum_{j'=1}^J x_{j'} \Sigma_{jj'} = x^\top \Sigma x,$$

where Σ is the covariance matrix (aka. variance-covariance matrix) with entries

$$\begin{aligned}
\Sigma_{jj'} &= \sum_{i=1}^n p_i \xi_i^j \xi_i^{j'} - \sum_{i=1}^n p_i \xi_i^j \cdot \sum_{i'=1}^n p_{i'} \xi_{i'}^{j'} \\
&= \mathbb{E}\, \xi^j \xi^{j'} - \mathbb{E}\, \xi^j \cdot \mathbb{E}\, \xi^{j'} = \operatorname{cov}\left(\xi^j, \xi^{j'}\right).
\end{aligned}$$

Remark 3.7 (Bessel's correction). For the empirical measure $p_i = 1/n$, the entries of the variance–covariance matrix are

$$\Sigma_{jj'} = \operatorname{cov}\left(\xi^j, \xi^{j'}\right) = \frac{1}{n} \sum_{i=1}^n \left(\xi_i^j - \overline{\xi}^j\right)\left(\xi_i^{j'} - \overline{\xi}^{j'}\right), \text{ where } \overline{\xi}^j := \sum_{i=1}^n \xi_i^j.$$

Bessel's correction replaces this quantity by $\Sigma_{j,j'} = \frac{1}{n-1} \sum_{i=1}^n \left(\xi_i^j - \overline{\xi}^j\right)\left(\xi_i^{j'} - \overline{\xi}^{j'}\right)$.

Example 3.8. Cf. Table 3.4.

19.2%	4.1%	7.9%	1.3%	-1.9%
4.1%	208.8%	-7.4%	44.5%	-18.9%
7.9%	-7.4%	42.8%	-10.3%	-6.7%
1.3%	44.5%	-10.3%	88.9%	2.9%
-1.9%	-18.9%	-6.7%	2.9%	5.9%

(a) Covariance matrix Σ of returns in Table 3.1, cf. also Table 3.3

5.72	-0.04	-1.01	-0.20	0.69
-0.04	1.03	0.74	-0.57	4.41
-1.01	0.74	3.59	-0.14	6.22
-0.20	-0.57	-0.14	1.49	-2.78
0.69	4.41	6.22	-2.78	39.83

(b) The inverse Σ^{-1}

Table 3.4: Covariance matrix Σ of returns in Table 3.1 and its inverse

3.6 THE NON-EMPIRICAL FORMULATION

Here, the random variable is $\xi\colon \Omega \to \mathbb{R}^J$. We define $r := \mathbb{E}\, x^\top \xi$ and observe that

$$\begin{aligned}
\operatorname{var} x^\top \xi &= \mathbb{E}\left(x^\top \xi\right)^2 - \left(\mathbb{E}\, x^\top \xi\right)^2 \\
&= \mathbb{E}\left(\sum_{j=1}^J \xi^j x_j\right)^2 - \left(\sum_{j=1}^J x_j\, \mathbb{E}\,\xi^j\right)^2 \\
&= \sum_{j=1}^J \sum_{j'=1}^J x_j x_{j'}\, \mathbb{E}\left(\xi^j \xi^{j'}\right) - \sum_{j=1}^J \sum_{j'=1}^J x_j x_{j'} \left(\mathbb{E}\,\xi^j\right)\left(\mathbb{E}\,\xi^{j'}\right) \\
&= \sum_{j=1}^J \sum_{j'=1}^J x_j x_{j'} \underbrace{\left(\mathbb{E}\left(\xi^j \xi^{j'}\right) - \left(\mathbb{E}\,\xi^j\right)\left(\mathbb{E}\,\xi^{j'}\right)\right)}_{\operatorname{cov}\left(\xi^j, \xi^{j'}\right)} = x^\top \operatorname{cov}(\xi)\, x.
\end{aligned}$$

Definition 3.9. The covariance matrix[3] is

$$\Sigma := \operatorname{cov}(\xi) = \mathbb{E}\left(\xi \cdot \xi^\top\right) - (\mathbb{E}\,\xi) \cdot (\mathbb{E}\,\xi)^\top .$$

Remark 3.10. Note, that

$$\Sigma = \Xi^\top \cdot \operatorname{diag}(p) \cdot \Xi - \underbrace{p^\top \Xi}_{r} \cdot \underbrace{\Xi^\top p}_{r^\top}$$

and Σ is symmetric.

3.7 THE CAPITAL ASSET PRICING MODEL (CAPM)

Markowitz considers the problem

$$\begin{aligned}
\text{minimize (in } x \in \mathbb{R}^J) \quad & \operatorname{var} x^\top \xi \\
\text{subject to} \quad & \mathbb{E}\, x^\top \xi \ge \mu, \\
& \mathbb{1}^\top x \le 1€, \\
& (x \ge 0)
\end{aligned} \tag{3.3}$$

The Markowitz problem (3.3) is quadratic, with linear constraints: J varialbes, 2 constraints.

Definition 3.11. A portfolio $x^* \in \mathbb{R}^J$ is *efficient* if it solves (3.3).

[3]The covariance cov is a.k.a. *variance matrix*.

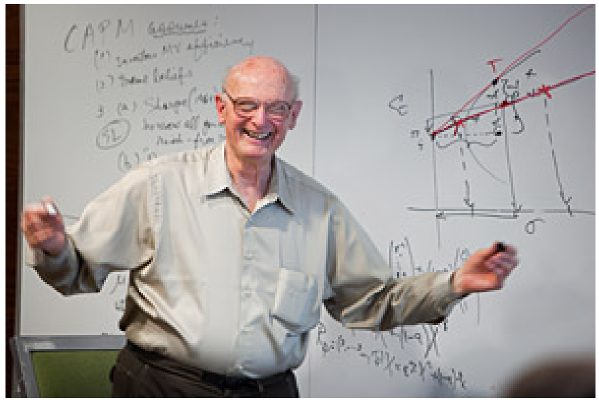


Figure 3.2: Harry Markowitz (1927) explains the CAPM and the mean-variance plot. Nobel Memorial Prize in Economic Sciences (1990)

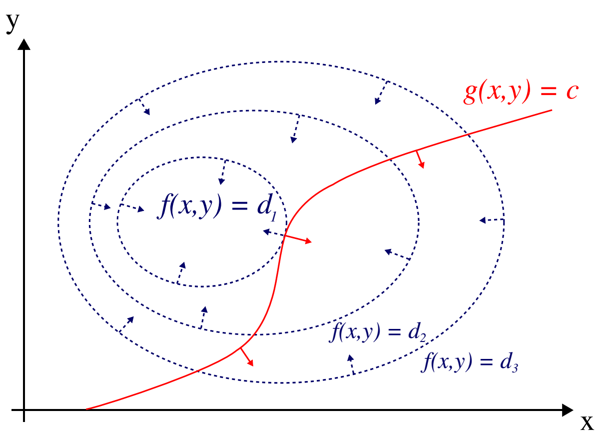


Figure 3.3: Illustration of Lagrange multipliers, contour lines of f

Expressed by matrices the Markowitz problem (3.3) is

$$\begin{aligned} \text{minimize }_{\text{in } x\in\mathbb{R}^J} \; & x^\top \Sigma x \\ \text{subject to } & x^\top r \geq \mu, \\ & x^\top \mathbb{1} \leq 1, \\ & (x \geq 0). \end{aligned}$$

Theorem 3.12. *The efficient Markowitz portfolio is given by*

$$x^*(\mu) = \mu\left(\frac{c}{d}\Sigma^{-1}r - \frac{b}{d}\Sigma^{-1}\mathbb{1}\right) - \frac{b}{d}\Sigma^{-1}r + \frac{a}{d}\Sigma^{-1}\mathbb{1}, \tag{3.4}$$

where $a := r^\top\Sigma^{-1}r$, $b := r^\top\Sigma^{-1}\mathbb{1}$, $c := \mathbb{1}^\top\Sigma^{-1}\mathbb{1}$ *and* $d := ac - b^2$ *are auxiliary quantities.*

Remark 3.13. The units of the auxiliary are $[a] = \frac{\text{interest}^2}{\text{variance}}$, $[b] = \frac{\text{interest}}{\text{variance}}$, $[c] = \frac{1}{\text{variance}}$ and $[\Sigma] =$ variance.

Proof. Differentiate the Lagrangian[4]

$$L(x;\lambda,\gamma) := \frac{1}{2}x^\top\Sigma x - \lambda(r^\top x - \mu) - \gamma(\mathbb{1}^\top x - 1)$$

[4]We could choose $x^\top\Sigma x$ or $\sqrt{x^\top\Sigma x}$ equally well in the Lagrangian function L.

to get the necessary conditions for optimality,

$$0 = \frac{\partial L}{\partial x} = \frac{1}{2}(\Sigma x)^\top + \frac{1}{2}x^\top \Sigma - \lambda r^\top - \gamma \mathbb{1}^\top, \tag{3.5}$$

$$0 = \frac{\partial L}{\partial \lambda} = -r^\top x + \mu, \tag{3.6}$$

$$0 = \frac{\partial L}{\partial \gamma} = -\mathbb{1}^\top x + 1. \tag{3.7}$$

It follows from (3.5) that

$$x^* = \lambda \Sigma^{-1} r + \gamma \Sigma^{-1} \mathbb{1}\,. \tag{3.8}$$

To determine the shadow prices λ and γ we employ (3.6) and (3.7), i.e.,

$$\begin{aligned} \mu &= r^\top x^* = \lambda r^\top \Sigma^{-1} r + \gamma r^\top \Sigma^{-1} \mathbb{1} \text{ and} \\ 1 &= \mathbb{1}^\top x^* = \lambda \mathbb{1}^\top \Sigma^{-1} r + \gamma \mathbb{1}^\top \Sigma^{-1} \mathbb{1}\,. \end{aligned} \tag{3.9}$$

We may rewrite these latter equations as a usual matrix equation,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} \mu \\ 1 \end{pmatrix} \tag{3.10}$$

with solutions $\lambda^* = \frac{\mu c - b}{ac - b^2}$ and $\gamma^* = \frac{a - \mu b}{ac - b^2}$. Substitute them in (3.8) to get the assertion (3.4) of the theorem, i.e., the efficient portfolio. □

Corollary 3.14. *Note from (3.9) that*

$$\mathbb{E}\, x^{*\top} \xi = x^{*\top} \xi = \mu$$

and

$$\begin{aligned} \text{var}\left(x^{*\top}\xi\right) &= x^{*\top}\Sigma x^* = \frac{\mu^2 c - 2\mu b + a}{ac - b^2}, \\ \text{cov}\left(x^{*\top}\xi, \xi_i\right) &= x^{*\top}\Sigma e_i = \frac{\mu c - b}{ac - b^2} r_i + \frac{a - \mu b}{ac - b^2}. \end{aligned} \tag{3.11}$$

Proof. We have

$$\begin{aligned} \text{var}\left(x^{*\top}\xi\right) &= x^{*\top}\Sigma x^* = \left(\lambda \Sigma^{-1} r + \gamma \Sigma^{-1} \mathbb{1}\right)^\top \Sigma x^* \\ &= \lambda r^\top x^* + \gamma \mathbb{1}^\top x^* = \lambda \mu + \gamma \\ &= \frac{\mu c - b}{ac - b^2}\mu + \frac{a - \mu b}{ac - b^2} \\ &= \mu^2 \frac{c}{ac - b^2} - 2\mu \frac{b}{ac - b^2} + \frac{a}{ac - b^2} \\ &=: \sigma^2(\mu). \end{aligned}$$

□

Corollary 3.15. *It holds that $ac > b^2$.*

Proof. The matrix Σ is positive definite as it is a covariance matrix, and so is its inverse. It thus holds that $c = \mathbb{1}^\top \Sigma^{-1} \mathbb{1} > 0$. The variance is also positive for every $\mu > 0$, so it follows from (3.11) that $ac - b^2 > 0$. □

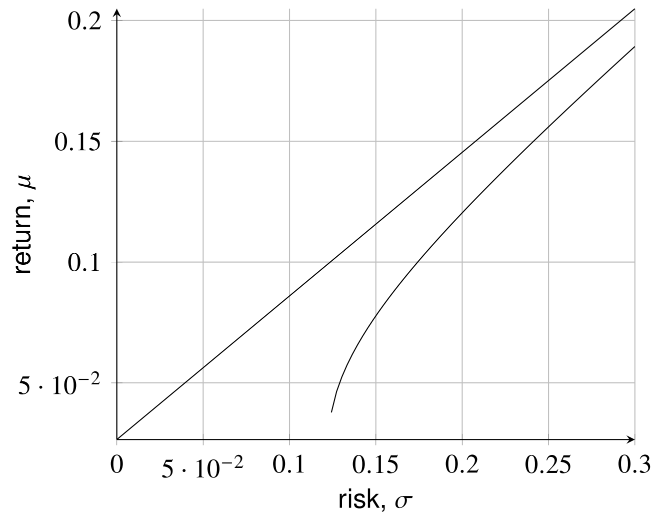


Figure 3.4: The mean-variance plot (3.12), the efficient frontier and asymptotic (3.13)

3.7.1 The Mean-Variance Plot

This section studies the Markowitz problem as a function of the parameter μ.

Corollary 3.16 (Mean-variance)**.** *Set* $\sigma^2 := \operatorname{var} Y_{x^*}$*, then we have (for* $\sigma > \frac{1}{\sqrt{c}}$*) that*

$$\mu(\sigma) = \frac{b + \sqrt{(ac - b^2)(c\sigma^2 - 1)}}{c} = \frac{b}{c} + \sqrt{\left(a - \frac{b^2}{c}\right)\left(\sigma^2 - \frac{1}{c}\right)}. \tag{3.12}$$

Proof. Solve (3.11) for $\operatorname{var} Y_x = \sigma^2$ using the quadratic formula. □

Figure 3.4 graphs the relation (3.12), i.e., the mean and the variance of efficient portfolios.

Corollary 3.17. *It holds that*

$$\mu(\sigma) \le \frac{b}{c} + \sigma\sqrt{a - \frac{b^2}{c}} \tag{3.13}$$

and every efficient portfolio satisfies $\sigma \ge \frac{1}{\sqrt{c}}$ *and* $\mu \ge \frac{b}{c}$*.*

Proof. This is immediate from (3.12); cf. also Figure 3.4. □

Remark 3.18*.* The Markowitz portfolio with *smallest variance* which does not include a risk free asset is given for $\mu = \frac{b}{c}$ (differentiate (3.11) with respect to μ) and this portfolio thus is of particular interest. In particular, note that its variance, by (3.11), is

$$\sigma^2_{\text{min}} = \operatorname{var} x^* \left(\frac{b}{c}\right)^\top \xi = \frac{\left(\frac{b}{c}\right)^2 c - 2\frac{b}{c} b + a}{a\,c - b^2} = \frac{1}{c}\frac{b^2 - 2b^2 + a\,c}{a\,c - b^2} = \frac{1}{c}.$$

Exercise 3.1. *The portfolio with smallest-variance in our data is* $\mu = \frac{b}{c} = \frac{1.76}{66.3} = 2.66\%$*, the corresponding standard deviation, which cannot be improved, is* $\sigma = \frac{1}{\sqrt{c}} = \frac{1}{\sqrt{\mathbb{1}^\top \Sigma^{-1}\mathbb{1}}} = 12.3\%$*; cf. Figure 3.4 and Figure 3.5.*

3.7.2 Tangency portfolio

For some fixed risk free rate r_0 we study the particular reward

$$\mu_m := \frac{a - r_0\, b}{b - r_0\, c}. \tag{3.14}$$

Lemma 3.19. *The variance corresponding to the reward μ_m is*

$$\sigma_m^2 := \sigma^2(\mu_m) = \frac{a - 2r_0\, b + r_0^2\, c}{(b - r_0\, c)^2}. \tag{3.15}$$

Proof. From (3.11) it follows that

$$\begin{aligned}
\sigma^2(\mu_m) &= \frac{c\mu_m^2 - 2b\mu_m + a}{ac - b^2}\\
&= \frac{1}{(b - r_0 c)^2} \frac{c(a - r_0 b)^2 + 2b(a - r_0 b)(b - r_0 c) + a(b - r_0 c)^2}{ac - b^2}\\
&= \cdots = \frac{a - 2r_0\, b + r_0^2\, c}{(b - r_0\, c)^2}
\end{aligned}$$

after some annoying, but elementary algebra. □

Definition 3.20 (Sharpe ratio)**.** The Sharpe ratio of the portfolio with reward μ_m is

$$s_m := \frac{\mu_m - r_0}{\sigma_m}. \tag{3.16}$$

Remark 3.21*.* Note first that $\frac{\mu_m - r_0}{b - r_0\, c} = \sigma_m^2$. It follows that

$$b - r_0\, c = \frac{\mu_m - r_0}{\sigma_m^2} \underset{(3.16)}{=} \frac{s_m}{\sigma_m} \tag{3.17}$$

and with (3.14) that

$$a - r_0\, b \underset{(3.14)}{=} \mu_m\, (b - r_0\, c) \underset{(3.17)}{=} \frac{\mu_m \cdot s_m}{\sigma_m}.$$

From (3.15) we deduce further that

$$a - 2r_0\, b + r_0^2\, c \underset{(3.15)}{=} \sigma_m^2\, (b - r_0\, c)^2 \underset{(3.17)}{=} s_m^2. \tag{3.18}$$

Definition 3.22 (Market portfolio, tangency portfolio)**.** The *market portfolio* is

$$x_m := x^*(\mu_m) = \frac{\sigma_m}{s_m} \cdot \Sigma^{-1}\big(r - r_0 \cdot \mathbb{1}\big). \tag{3.19}$$

Remark 3.23*.* It follows according (3.4) is

$$\begin{aligned}
x_m = x^*(\mu_m) &= \frac{\mu_m c - b}{ac - b^2}\Sigma^{-1} r - \frac{\mu_m b - a}{ac - b^2}\Sigma^{-1}\,\mathbb{1}\\
&= \frac{1}{b - r_0 c}\Sigma^{-1} r - \frac{r_0}{b - r_0 c}\Sigma^{-1}\,\mathbb{1}
\end{aligned}$$

and with (3.17) thus (3.19).

Remark 3.24*.* For the line $t(\sigma) := r_0 + \sigma \cdot \frac{\mu_m - r_0}{\sigma_m}$ it holds that

$$\begin{aligned}
t(\sigma_m) &= \mu(\sigma_m) \text{ and}\\
t'(\sigma_m) &= \mu'(\sigma_m) \qquad \text{(cf. (3.12))}.
\end{aligned}$$

The line t thus is the tangent drawn from the point of the risk-free asset $t(0) = r_0$ to the feasible region for risky assets. The tangent line is called *capital market line*. The portfolio with decomposition (3.19) is also called the *most efficient portfolio*, it has the highest *reward-to-volatility ratio*.

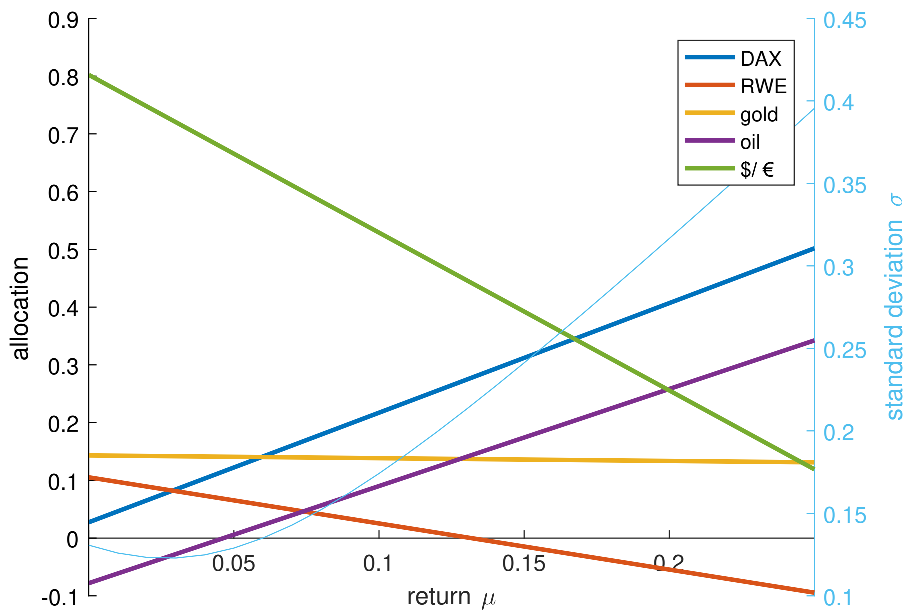


Figure 3.5: Asset allocation according to Markowitz for varying return μ

3.7.3 The Two Fund Theorem

Note that μ is a model-parameter in the Markowitz model (2.1). We now compare efficient portfolios for different returns μ.

Theorem 3.25 (Two fund theorem)**.** *If x_1^* and x_2^* are different efficient portfolios (for different μs), then every efficient portfolio can be obtained as an affine combination of these two.*

Proof. Recall from (3.4) that

$$\begin{aligned} x^*(\mu) &= \frac{\mu c - b}{ac - b^2}\Sigma^{-1} r + \frac{a - \mu b}{ac - b^2}\Sigma^{-1}\mathbb{1} \\ &= \left(-\frac{b}{ac - b^2}\Sigma^{-1} r + \frac{a}{ac - b^2}\Sigma^{-1}\mathbb{1}\right) \\ &\quad + \mu\left(\frac{c}{ac - b^2}\Sigma^{-1} r - \frac{b}{ac - b^2}\Sigma^{-1}\mathbb{1}\right). \end{aligned}$$

□

Exercise 3.2. *The result for our data is (cf. Figure 3.5)*

$$x^*(\mu) = \begin{pmatrix} 2.7\% \\ 10.5\% \\ 14.3\% \\ -7.8\% \\ 80.2\% \end{pmatrix} + \mu \begin{pmatrix} +1.90 \\ -0.80 \\ -0.05 \\ +1.68 \\ -2.73 \end{pmatrix};$$

the auxiliary quantities are $a = 0.40$, $b = 1.76$ and $c = 66.31$.

3.8 MARKOWITZ PORTFOLIO INCLUDING A RISK FREE ASSET

Set $r := \mathbb{E}\,\xi$, i.e., $r_j := \mathbb{E}\,\xi^j$, $j = 1, \dots, J$. Further, let r_0 be the return of the risk-free asset. We set

$$\tilde{r} = \begin{pmatrix} r_0 \\ r_1 \\ \vdots \\ r_J \end{pmatrix} = \begin{pmatrix} r_0 \\ \\ r \\ \end{pmatrix} \text{ with } r = \begin{pmatrix} r_1 \\ \vdots \\ r_J \end{pmatrix} \text{ and } \tilde{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_J \end{pmatrix} = \begin{pmatrix} x_0 \\ \\ x \\ \end{pmatrix} \text{ with } x = \begin{pmatrix} x_1 \\ \vdots \\ x_J \end{pmatrix}.$$

Note that $S_t^0 = S_0^0 \cdot e^{r_0 t}$ and the annualized return $\xi_i^0 = \frac{1}{t_{i+1}-t_i} \ln \frac{S_{t+1}^0}{S_{t+1}^0} = r_0$ is the *constant* risk-free interest rate. The risk free asset is not correlated with other assets. The covariance matrix

$$\tilde{\Sigma} := \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix}$$

thus is *not* invertible. Consequently, the results from the previous section do *not* apply.

Theorem 3.26. *The Markowitz portfolio is given by*

$$\tilde{x}^*(\mu) = \begin{pmatrix} x_0^*(\mu) \\ x^*(\mu) \end{pmatrix} = \begin{pmatrix} \frac{\mu_m - \mu}{\mu_m - r_0} \\ \frac{\mu - r_0}{\mu_m - r_0} \cdot x_m \end{pmatrix}, \tag{3.20}$$

where x_m (cf. (3.19)) is the tangency portfolio (market portfolio).

Proof. Differentiate the Lagrangian (cf. Figure 3.3 for illustration)

$$L(\tilde{x}; \lambda, \gamma) := \frac{1}{2} x^\top \Sigma x - \lambda(\tilde{r}^\top \tilde{x} - \mu) - \gamma(\mathbb{1}^\top \tilde{x} - 1)$$

to get the necessary conditions for optimality,

$$0 = \frac{\partial L}{\partial x} = \Sigma x - \lambda r^\top - \gamma\, \mathbb{1}^\top, \tag{3.21}$$

$$0 = \frac{\partial L}{\partial x_0} = -\lambda r_0 - \gamma, \tag{3.22}$$

$$0 = \frac{\partial L}{\partial \lambda} = \tilde{r}^\top \tilde{x} - \mu = r_0 x_0 + r^\top x - \mu, \tag{3.23}$$

$$0 = \frac{\partial L}{\partial \gamma} = \mathbb{1}^\top \tilde{x} - 1 = x_0 + \mathbb{1}^\top x - 1, \tag{3.24}$$

We get from (3.21) that $x^* = \lambda \Sigma^{-1} r + \gamma \Sigma^{-1}\, \mathbb{1}$. Substitute x^* in (3.23) and (3.24), and after collecting terms in (3.22)–(3.24) one finds (cf. (3.10))

$$\begin{pmatrix} 0 & r_0 & 1 \\ r_0 & a & b \\ 1 & b & c \end{pmatrix} \begin{pmatrix} x_0 \\ \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \\ 1 \end{pmatrix}.$$

This linear matrix equation has explicit solution

$$\begin{pmatrix} x_0^* \\ \lambda^* \\ \gamma^* \end{pmatrix} = \underbrace{\frac{1}{a - 2br_0 + cr_0^2}}_{=s_m^2,\ \text{cf. (3.18)}} \begin{pmatrix} a - br_0 + \mu(cr_0 - b) \\ -r_0 + \mu \\ r_0^2 - r_0 \mu \end{pmatrix}. \tag{3.25}$$

So we finally get

$$\tilde{x}^* = \begin{pmatrix} x_0^* \\ x^* \end{pmatrix} = \begin{pmatrix} \frac{1}{s_m^2}\left(a - br_0 + \mu(cr_0 - b)\right) \\ \lambda^* \Sigma^{-1} r + \gamma^* \Sigma^{-1}\, \mathbb{1} \end{pmatrix} = \frac{1}{s_m^2} \begin{pmatrix} s_m^2 - (b - r_0 c)(\mu - r_0) \\ (\mu - r_0)\left(\Sigma^{-1} r - r_0 \Sigma^{-1}\, \mathbb{1}\right) \end{pmatrix}$$

from (3.24); cf. Exercise 3.3. ◻

Corollary 3.27. *The variance (standard deviation, resp.) of the portfolio corresponding to μ is*

$$\operatorname{var}\left(\tilde{x}^*(\mu)^\top \xi\right) = \left(\frac{\mu - r_0}{s_m}\right)^2 \qquad \left(\sigma(\mu) = \frac{|\mu - r_0|}{s_m},\ \textit{resp.}\right).$$

Proof. Recall the special structure of $\tilde{\Sigma}$. It thus follows from (3.20) that

$$\begin{aligned}\operatorname{var}\left(\tilde{x}^*(\mu)^\top \xi\right) &= \tilde{x}^*(\mu)^\top \Sigma \tilde{x}^*(\mu)\\ &= \frac{(\mu - r_0)^2}{s_m^4}\,(r - r_0\,\mathbb{1})^\top\, \Sigma^{-1}\Sigma\Sigma^{-1}\,(r - r_0\,\mathbb{1})\\ &= \frac{(\mu - r_0)^2}{s_m^4}\left(r^\top\Sigma^{-1}r - 2r_0\,\mathbb{1}^\top\,\Sigma^{-1}r + r_0^2\,\mathbb{1}^\top\,\Sigma^{-1}\,\mathbb{1}\right)\\ &= \frac{(\mu - r_0)^2}{s_m^4}\left(a - 2r_0 b + r_0^2 c\right) = \left(\frac{\mu - r_0}{s_m}\right)^2,\end{aligned} \tag{3.26}$$

which is the assertion. □

Remark 3.28. Note, that the portfolio with smallest variance is attained here for $\tilde{\mu} = r_0$, the corresponding variance by (3.26) is *zero* , i.e, there is *no risk*. This is in significant contrast to Remark 3.18.

3.9 ONE FUND THEOREM

Theorem 3.29 (One fund theorem, Tobin[5]-separation, market portfolio)**.** *Every efficient portfolio is the affine combination of*

(i) a portfolio without a risk-free asset (the market portfolio), and

(ii) the risk free asset.

Definition 3.30. The portfolio allocation in (i) is called *market portfolio*.

Proof. Choose

(i) $\mu := r_0$, then the portfolio in (3.20) is $\tilde{x}^*(r_0) = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$; this portfolio does not involve stocks and thus is completely *free of risk*, i.e., it consists of the risk-free asset solely;

(ii) For the market portfolio, recall (cf. the tangency portfolio (3.14))

$$\mu_m := \mu_t = \frac{a - r_0\, b}{b - r_0\, c}$$

and $x_m := x_t$. Then, by (3.19) and (3.20),

$$\tilde{x}^*(\mu_m) = \begin{pmatrix}0\\ x^*(\mu_t)\end{pmatrix} = \begin{pmatrix}0\\ \frac{\sigma_m}{s_m}\,\Sigma^{-1}\,(r - r_0\cdot\mathbb{1})\end{pmatrix} = \begin{pmatrix}0\\ x_m\end{pmatrix},$$

which means that the portfolio $\tilde{x}^*(\mu_m)$ is free from risk-free assets, i.e., *does not contain the risk free asset*.

[5]Tobin-Separation, 1918–2002, American economist

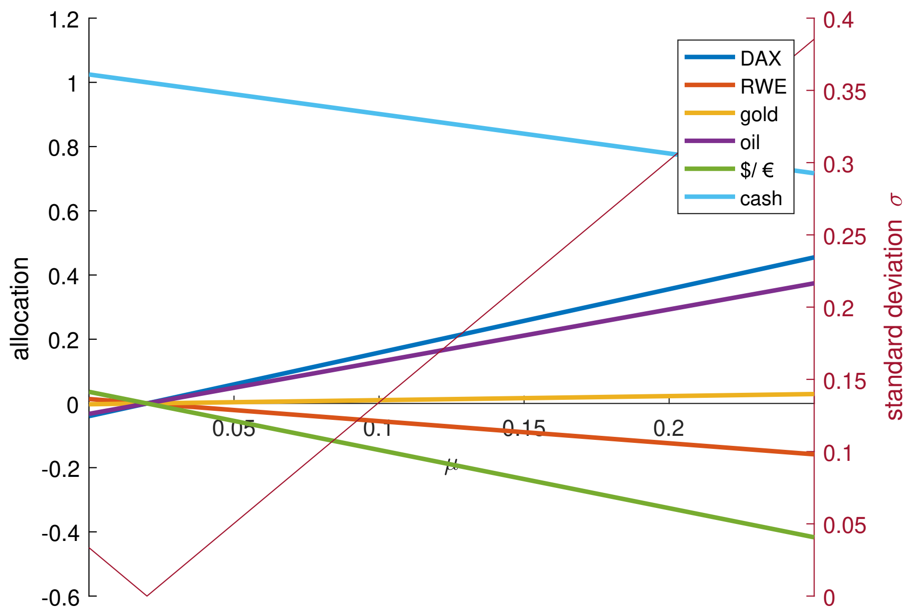


Figure 3.6: Markowitz portfolio including a risk free asset (cash) with return $r_0 = 2\%$ for varying return μ

The assertion follows, as every portfolio is a linear combination of both portfolios by the Two Fund Theorem, Theorem 3.25. Explicitly, the optimal portfolio (cf. (3.20)) is

$$\tilde{x}^*(\mu) = \frac{\mu_m - \mu}{\mu_m - r_0}\begin{pmatrix}1\\0\end{pmatrix} + \frac{\mu - r_0}{\mu_m - r_0}\begin{pmatrix}0\\ \frac{\sigma_m}{s_m}\Sigma^{-1}\left(r - r_0\,\mathbb{1}\right)\end{pmatrix} = \frac{\mu_m - \mu}{\mu_m - r_0}\begin{pmatrix}1\\0\end{pmatrix} + \frac{\mu - r_0}{\mu_m - r_0}\begin{pmatrix}0\\x_m\end{pmatrix}.$$

□

Remark 3.31. The tangency portfolio coincides with the market portfolio, $x_m = x_t$.

Example 3.32. For $r_0 := 2\%$, the optimal portfolios for our data are given according the one fund theorem as

$$\tilde{x}^*(\mu) = \frac{83.3\% - \mu}{81.3\%}\underbrace{\begin{pmatrix}100\%\\0\\0\\0\\0\\0\end{pmatrix}}_{\text{no stocks}} + \frac{\mu - 2\%}{81.3\%}\underbrace{\begin{pmatrix}0\\160.7\%\\-56.0\%\\10.3\%\\132.2\%\\-147.3\%\end{pmatrix}}_{\text{no cash}};$$

cf. Figure 3.6.

3.9.1 Capital Asset Pricing Model (CAPM)

Recall that the market (or tangency) portfolio

$$x_m = x_t = \frac{\sigma_m}{s_m}\Sigma^{-1}\left(r - r_0\cdot\mathbb{1}\right)$$

has expectation (cf. (3.14))

$$\mathbb{E}\, x_m^\top \xi = \mu_m$$

and variance (cf. (3.15))

$$\operatorname{var}\left(x_m^\top \xi\right) = \sigma_m^2.$$

Remark 3.33 (Covariance of the market portfolio). The covariance of the market portfolio with asset j is

$$\operatorname{cov}\left(e_j^\top \xi,\, x_m^\top \xi\right) = e_j^\top \Sigma x_m = e_j^\top \Sigma \cdot \frac{\sigma_m}{s_m}\, \Sigma^{-1}\left(r - r_0 \cdot \mathbb{1}\right) = \frac{\sigma_m}{s_m}\left(r_j - r_0\right),$$

where $r = \mathbb{E}\,\xi$ and $r_j = \mathbb{E}\,\xi_j$.

It follows from the definition of the Sharpe ratio (3.16) that

$$\beta_j := \frac{\operatorname{cov}\left(x_m^\top \xi,\, e_j^\top \xi\right)}{\operatorname{var}\left(x_m^\top \xi\right)} = \frac{\frac{\sigma_m}{s_m}\left(r_j - r_0\right)}{\sigma_m^2} = \frac{r_j - r_0}{s_m\, \sigma_m} = \frac{r_j - r_0}{\mu_m - r_0},$$

i.e.,

$$r_j = r_0 + \beta_j \cdot \left(\mu_m - r_0\right). \tag{3.27}$$

The quantity β_j is the sensitivity of the expected excess asset returns to the expected excess market returns

The relation (3.27) is the core of the capital asset pricing model (CAPM). The graph of (3.27),

$$\beta \mapsto r_0 + \beta\left(\mu_m - r_0\right)$$

is also called *security market line* (SML in the μ-β-diagram).

Remark 3.34. For the market portfolio x_m it holds that $x_m^\top \beta = \beta_m = 1$.

Indeed, with (3.27),

$$\mu_m = x_m^\top r = r_0 x_m^\top \mathbb{1} + x_m^\top \beta \cdot \left(\mu_m - r_0\right) = r_0 + x_m^\top \beta \cdot \left(\mu_m - r_0\right)$$

and thus the assertion.

3.9.2 On systematic and specific risk

Observe that the correlation of asset j with the market is defined as $\rho_{j,m} := \frac{\operatorname{cov}\left(e_j^\top \xi,\, \xi^\top x_m^\top \xi\right)}{\sqrt{\operatorname{var}\left(x_m^\top \xi\right) \cdot \operatorname{var}\left(e_j^\top \xi\right)}}$ so that

$$\beta_j = \rho_{j,m} \cdot \frac{\sigma_j}{\sigma_m}.$$

It follows that

$$\sigma_j = \rho_{j,m}\, \sigma_j + \left(1 - \rho_{j,m}\right) \sigma_j = \underbrace{\beta_j\, \sigma_m}_{\text{systematic}} + \underbrace{\left(1 - \beta_j \frac{\sigma_m}{\sigma_j}\right) \sigma_j}_{\text{specific}}.$$

- ▹ The *systematic risk*[6] is also called *aggregate* or *undiversifiable risk*;
- ▹ the *specific risk*[7] is also called *unsystematic*, *residual* or *idiosyncratic risk*.

[6] systematisches Risiko (dt.)
[7] unsystematisches, spezifisches, diversifizierbares Risiko (dt.)

3.9.3 Sharpe ratio

Note that $\mathbb{E}\, e_j^\top \xi = r_j$ and var $e_j^\top \xi = e_j^\top \Sigma e_j = \Sigma_{jj}$.

Definition 3.35. Then quantity

$$\frac{r_j - r_0}{\sqrt{\Sigma_{jj}}}$$

is the *Sharpe ratio* of asset j.[8]

It holds that

$$\beta_j = \frac{\operatorname{cov}\left(\xi^\top e_j,\, \xi^\top x_m^*\right)}{\operatorname{var}\left(\xi^\top x_m^*\right)} = \frac{\operatorname{corr}\left(\xi^\top e_j,\, \xi^\top x_m^*\right)\, \sigma_m \sqrt{\Sigma_{jj}}}{\sigma_m^2} = \frac{\sqrt{\Sigma_{jj}}}{\sigma_m} \rho_{j,m}.$$

The security market line (SML) is

$$\text{SML}\colon \beta \mapsto r_0 + \beta \cdot (\mu_m - r_0)\,.$$

Note, from (3.27), that

$$\begin{aligned} \text{SML}(\beta_j) &= r_j, \\ \text{SML}(0) &= r_0 \text{ and} \\ \text{SML}(1) &= r_m. \end{aligned}$$

3.10 ALTERNATIVE FORMULATIONS OF THE MARKOWITZ PROBLEM

Instead of Markowitz (3.3) one may consider the problem

$$\begin{aligned} \text{maximize } & r^\top x \\ \text{subjet to } & x^\top \Sigma x \le q, \\ & \mathbb{1}^\top x \le 1, \\ & (x \ge 0) \end{aligned}$$

Proposition 3.36 (Utility maximization)**.** *The explicit solution of ($\kappa > 0$)*

$$\begin{aligned} \textit{maximize } & \mathbb{E}\, x^\top \xi - \frac{\kappa}{2} \operatorname{var} x^\top \xi \\ \textit{subjet to } & \mathbb{1}^\top x \le 1, \\ & (x \ge 0) \end{aligned} \tag{3.28}$$

is

$$x = \frac{1}{\kappa} \Sigma^{-1} \left(r + \frac{\kappa - \mathbb{1}^\top \Sigma^{-1} r}{\mathbb{1}^\top \Sigma^{-1} \mathbb{1}}\; \mathbb{1} \right).$$

Proof. The first order conditions for the Lagrangian $L(x;\lambda) \coloneqq x^\top r - \frac{\kappa}{2} x^\top \Sigma x + \lambda\left(\mathbb{1}^\top x - 1\right)$ are

$$\begin{aligned} 0 &= r - \kappa\, \Sigma\, x + \lambda\, \mathbb{1} \text{ and} \\ 1 &= \mathbb{1}^\top x. \end{aligned}$$

from which follows that $x = \frac{1}{\kappa}\Sigma^{-1}(r + \lambda\, \mathbb{1})$. Further, $1 = \mathbb{1}^\top x = \frac{1}{\kappa}\, \mathbb{1}^\top \Sigma^{-1}(r + \lambda\, \mathbb{1})$, i.e., $\lambda = \frac{\kappa - \mathbb{1}^\top \Sigma^{-1} r}{\mathbb{1}^\top \Sigma^{-1} \mathbb{1}}$. Hence the result. □

Remark 3.37*.* The portfolios of (3.28) and (3.4) coincide for $\kappa = \frac{d}{c\,\mu - b}$. In this case, $\mu = \frac{d + b\,\kappa}{c\,\kappa}$ and $\sigma^2 = \frac{d+\kappa^2}{c\,\kappa^2} = \frac{1}{c} + \frac{d}{c\,\kappa^2}$.

[8] William Sharpe, 1934, Nobel memorial Price in Economic Sciences (1990)

Eigenvalue	2.254	0.773	0.440	0.165	0.024
Variance explained	61.7%	21.2%	12.0%	4.5%	0.6%

(a) Eigenvalues and percentages of explained variance

	PC1	PC2	PC3
DAX	-0.02	-0.34	0.31
RWE	-0.95	0.30	0.05
gold	0.05	-0.24	-0.91
oil	-0.31	0.91	0.26
FX	0.08	0.14	0.13

(b) The first 3 principal components explain 94.9%

Table 3.5: Principal component analysis

	Stocks:				
return μ	S_1	S_2	S_3	S_4	S_5
$0\,\%$	$2.8\,\%$	$10.5\,\%$	$14.3\,\%$	$-7.8\,\%$	$80.2\,\%$
$15\,\%$	$31.2\,\%$	$-1.5\,\%$	$13.6\,\%$	$17.4\,\%$	$39.3\,\%$
$5\,\%$					

(a) Markowitz portfolio

	Stocks:		
return μ	S_1	S_2	S_3
$2\,\%$	$-4\,\%$	$-2\,\%$	$106\,\%$
$14\,\%$	$12\,\%$	$6\,\%$	$82\,\%$

(b) Markowitz portfolio

Table 3.6: Markowitz portfolios for various μ

3.11 PRINCIPAL COMPONENTS

Table 3.5 collects the eigenvalues and principal components of the covariance matrix Σ for the three components according the Karhunen–Loève decomposition. The first three principal components explain 95 % of the data.

3.12 PROBLEMS

Exercise 3.3. *Verify (3.25) and (3.20).*

Exercise 3.4. *The following portfolios (asset allocations, Table 3.6a) are efficient (in the sense of Markowitz). Give the Markowitz portfolio for $\mu = 5\%$?*

Exercise 3.5. *Is there a risk free asset among $S_1, \dots S_5$ in Table 3.6a?*

Exercise 3.6. *Give two pros and two cons for Markowtz's model.*

Exercise 3.7. *The portfolios in Table 3.6b are efficient. What is the risk free rate?*

Exercise 3.8. *Give the portfolio in Table 3.6b which does not contain a risk free asset.*

Exercise 3.9. *Verify Remark 3.37.*

Value-at-Risk

4

Never catch a falling knife.

investment strategy

4.1 DEFINITIONS

Definition 4.1 (Cumulative distribution function, cdf)**.** Let $Y\colon \Omega \to \mathbb{R}$ be a real-valued random variable. The *cumulative distribution function* (cdf, or just distribution function) is[1]

$$F_Y(x) := P(Y \le x). \tag{4.1}$$

Definition 4.2. The Value-at-Risk at (confidence, or risk) level $\alpha \in [0,1]$ is[2]

$$\mathsf{V@R}_\alpha(Y) := F_Y^{-1}(\alpha) = \inf\{x\colon P(Y \le x) \ge \alpha\}. \tag{4.2}$$

The Value-at-Risk is also called the *quantile funciton* $q_\alpha(Y) := \mathsf{V@R}_\alpha(Y)$ or *generalized inverse*.

Example 4.3. Cf. Figur 4.1 and Table 4.1 or Figure 9.1.

4.2 HOW ABOUT ADDING RISK?

Fact. *Consider the random variables (cf. Table 4.2) for which*

$$\mathsf{V@R}_{40\%}(X+Y) = 9 \le \mathsf{V@R}_{40\%}(X) + \mathsf{V@R}_{40\%}(Y) = 4 + 6 = 10$$

but

$$\mathsf{V@R}_{20\%}(X+Y) = 4 > \mathsf{V@R}_{20\%}(X) + \mathsf{V@R}_{20\%}(Y) = 2 + 0.$$

Lemma 4.4 (Cf. Figur 4.1)**.** *It holds that*

[1]Note that $F_Y(\cdot)$ is càdlàg, i.e., continue à droite, limite à gauche: right continuous with left limits

[2]Note, that $\inf \emptyset = +\infty$.

i	1	2	3	4	5
y_i	3	7	−3	8	−5
$P(Y = y_i)$	1/5	1/5	1/5	1/5	1/5

(a) Observations

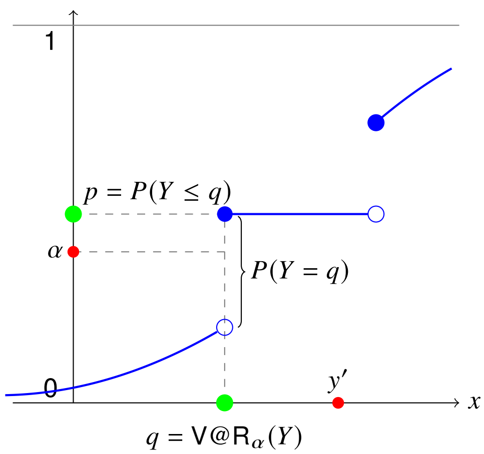
y	−5	−3	3	7	8	3.9
$F_Y(y)$	1/5	2/5	3/5	4/5	5/5	3/5

(b) Cumulative distribution function

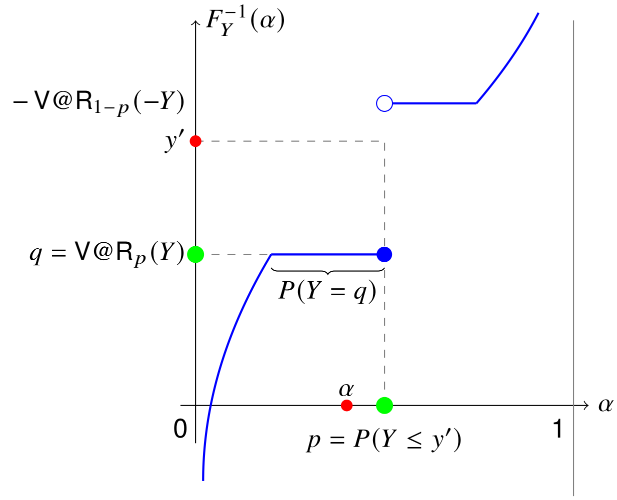
α	1/5	2/5	3/5	4/5	5/5
$\mathsf{V@R}_\alpha(Y)$	−5	−3	3	7	8

(c) Value-at-Risk

Table 4.1: Value-at-Risk



(a) Cumulative distribution function, cdf



(b) Quantile function, the generalized inverse of Figure 4.1a

Figure 4.1: Cumulative distribution and its corresponding quantile function

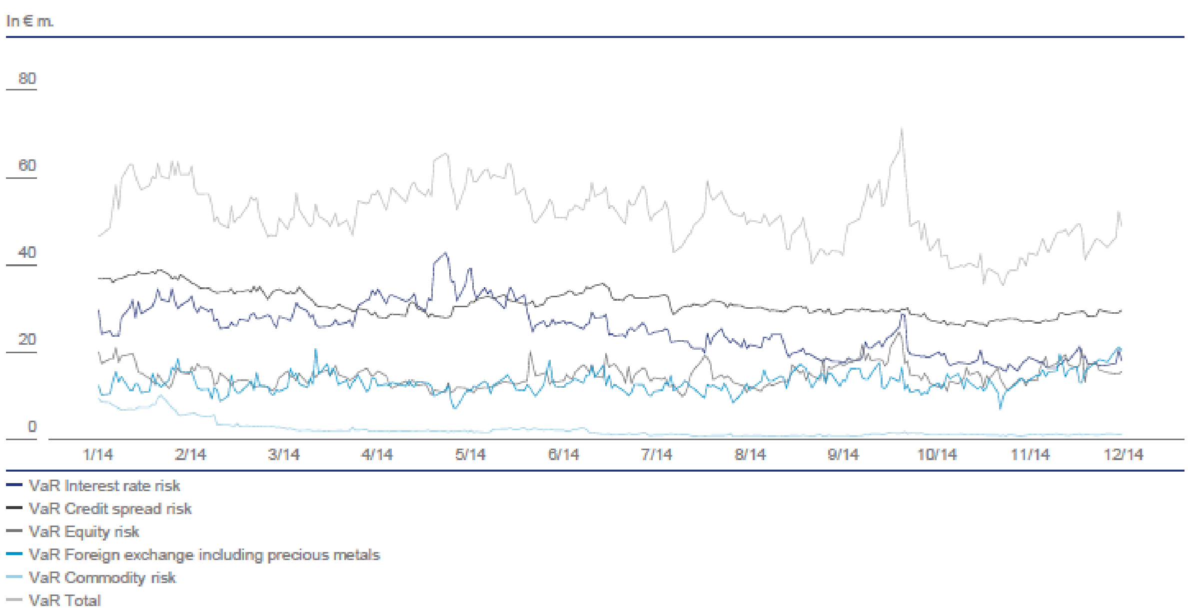


Figure 4.2: Deutsche Bank, annual report 2014, Values-at-Risk

$P(X = x_i, Y = y_i)$	1/3	1/3	1/3
X	2	4	5
Y	7	0	6
$X + Y$	9	4	11

Table 4.2: Counterexample

(i) $F_Y^{-1}(\alpha) \le x$ *if and only if* $\alpha \le F_Y(x)$ *(Galois connection, cf. van der Vaart [1998, Lemma 21.1]);*

(ii) F_Y *is continuous from right (upper semi-continuous);*

(iii) F_Y^{-1} *is continuous from left (lower semi-continuous);*

(iv) $F_Y^{-1}(F_Y(x)) \le x$ *for all* $x \in \mathbb{R}$ *and* $F_Y\left(F_Y^{-1}(\alpha)\right) \ge \alpha$ *for all* $\alpha \in (0,1)$*;*

(v) $F_Y^{-1}\left(F_Y\left(F_Y^{-1}(y)\right)\right) = F_Y^{-1}(y)$ *and* $F_Y\left(F_Y^{-1}(F_Y(y))\right) = F_Y(y)$*.*

Remark 4.5 (Quantile transform)*.* Let U be uniformly distributed, i.e., $P(U \le u) = u$ for every $u \in (0,1)$. The random variables Y and $F_Y^{-1}(U)$ share the same distribution.

Proof. $P\left(F_Y^{-1}(U) \le y\right) = P\left(U \le F_Y(y)\right) = F_Y(y)$, the assertion. □

The converse does not hold true, i.e., $F_Y(Y)$ is not necessarily uniformly distributed. However, we have the following:

Lemma 4.6 (The generalized quantile transform, Pflug and Römisch [2007, Proposition 1.3])**.** *Let* U *be uniform and independent from* Y*. Then*

$$F(Y,U) \coloneqq (1-U)\cdot F(Y-) + U \cdot F(Y) \tag{4.3}$$

is uniformly $[0,1]$ *and*

$$F_Y^{-1}\big(F(Y,U)\big) = Y \text{ almost surely,}$$

where $F(x-) \coloneqq \lim_{x' \nearrow x} F(x')$*.*

Proof. For $p \in (0,1)$ fixed let y_p satisfy $F_Y(y_p-) \le p \le F(y_p)$. Then

$$P\big(F(Y,U) \le p \,|\, Y\big) = \begin{cases} 1 & \text{if } Y < y_p \\ \frac{p-F(y_p-)}{F(y_p)-F(y_p-)} & \text{if } Y = y_p \\ 0 & \text{if } Y > y_p \end{cases}$$

and thus $P\big(F(Y,U) \le p\big) = F(y_p-) + \big(F(y_p) - F(y_p-)\big)\frac{p-F(y_p-)}{F(y_p)-F(y_p-)} = p$, i.e., $F(Y,U)$ is uniformly distributed.

Conditional on $\{Y = y\}$ it holds that $F(Y,U) \in \left[F^{-1}(y-), F^{-1}(y)\right]$. But $F^{-1}(u) = y$ for every $u \in \left[F^{-1}(y-), F^{-1}(y)\right]$ and thus the assertion. □

4.3 PROPERTIES OF THE VALUE-AT-RISK

Nice properties (cf. Lemma 4.4)

(i) Homogeneity: it holds that $\mathsf{V@R}_\alpha(\lambda Y) = \lambda \cdot \mathsf{V@R}_\alpha(Y)$ for $\lambda \ge 0$;[3]

$$F_{\lambda Y}(\cdot) = F_Y(\cdot/\lambda).$$

(ii) Cash-invariance: $\mathsf{V@R}_\alpha(Y+c) = c + \mathsf{V@R}_\alpha(Y)$, where $c \in \mathbb{R}$ is a constant. Note also that

$$F_{Y+c}(\cdot) = F_Y(\cdot - c).$$

(iii) Law-invariance: if X and Y share the same law, i.e., $P(X \le z) = P(Y \le z)$ for all $z \in \mathbb{R}$, then $\mathsf{V@R}_\alpha(X) = \mathsf{V@R}_\alpha(Y)$ (note X and Y may have the same law, even if $X(\omega) \ne Y(\omega)$ for all $\omega \in \Omega$ and even if $X\colon \Omega \to \mathbb{R}$ and $Y\colon \Omega' \to \mathbb{R}$);

[3] Throughout this lecture shall investigate *positively homogeneous* acceptability functionals.

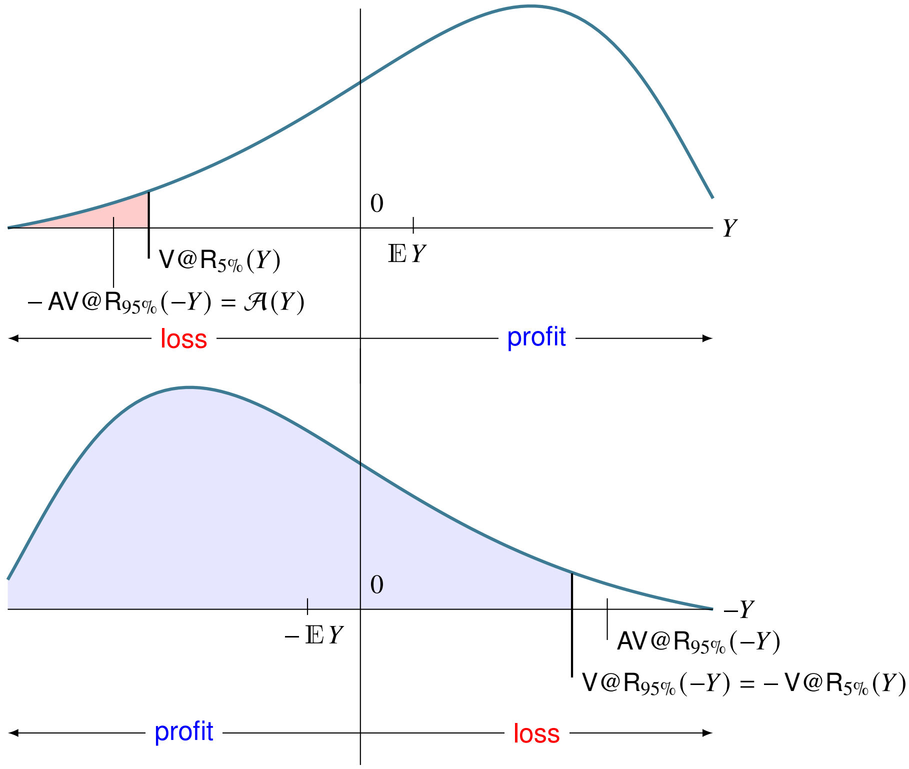


Figure 4.3: Profit versus loss

(iv) $F_Y\left(F_Y^{-1}(p)\right) \geq p$, with equality, if p is in the range of F_Y, equivalently, if $F_Y^{-1}(p)$ is a point of continuity of F_Y;

(v) $F_Y^{-1}\left(F_Y(u)\right) \leq u$, with equality, if u is in the range of F_Y^{-1}, equivalently, if $F_Y(u)$ is a point of continuity of F_Y^{-1};

(vi) comonotonic additive (cf. Section 14), i.e., $\mathsf{V@R}_\alpha(X+Y)=\mathsf{V@R}_\alpha(X)+\mathsf{V@R}_\alpha(Y)$, provided that X and Y are comonotonic.

4.4 PROFIT VERSUS LOSS

For an illustration see Figure 4.3.

Lemma 4.7 (Profit vs. loss, cf. Figure 4.3)**.** *It holds that*

$$\mathsf{V@R}_\alpha(Y) \leq -\mathsf{V@R}_{1-\alpha}(-Y) \tag{4.4}$$

with equality if $F_Y(y+h) > F_Y(y)$ for $h>0$ at $y=\mathsf{V@R}_\alpha(Y)$.

Proof. First,

$$\begin{aligned}
-\mathsf{V@R}_{1-\alpha}(-Y) &= -\inf\{y\colon P(-Y\leq y)\geq 1-\alpha\}\\
&= \sup\{-y\colon P(Y\geq -y)\geq 1-\alpha\}\\
&= \sup\{y\colon P(Y\geq y)\geq 1-\alpha\}\\
&= \sup\{y\colon 1-P(Y\geq y)\leq \alpha\}\\
&= \sup\{y\colon P(Y<y)\leq \alpha\} \qquad (4.5)
\end{aligned}$$

	S_1	S_2	S_3
Ξ:	5 %	−4 %	−2 %
	8 %	2 %	0 %
	4 %	1 %	0 %
	−9 %	0 %	−10 %

Table 4.3: Annualized, logarithmic returns

Now observe that

$$\{y\colon P(Y<y)\le\alpha\}\,\dot\cup\,\{y\colon P(Y<y)>\alpha\}$$

and disjoint intervals. It follows that

$$\begin{aligned}-\mathsf{V@R}_{1-\alpha}(-Y)&=\inf\{y\colon P(Y<y)>\alpha\}\\&\ge\inf\{y\colon P(Y\le y)>\alpha\}\\&\ge\inf\{y\colon P(Y\le y)\ge\alpha\},\end{aligned}$$

the assertion.

Now set $y:=\mathsf{V@R}_\alpha(Y)$. As the cdf $y\mapsto P(Y\le y)$ is right-continuous it follows that y is feasible for (4.2) and (4.5), i.e., $P(Y<y)\le\alpha\le P(Y\le y)$. Hence the result. ◻

Problem 4.8. A Markowitz-like formulation involving the Value-at-Risk is

$$\begin{array}{rll}\text{maximize}\\ \text{(in } x=(x_1,\dots x_S)\text{)} & \frac{1}{N}\sum_{n=1}^N x^\top\xi_n=x^\top\xi\\ \text{subject to} & \mathsf{V@R}_{5\,\%}\,(x^\top\xi)\ge-2\$ & 5\%\text{ worst profits}>-2\$\\ & x_1+\dots+x_S\le 1.000\$ & \text{budget constraint}\\ & (x\ge 0) & \text{shortselling allowed / not allowed}\end{array}$$

or, with (4.4),

$$\begin{array}{rll}\text{maximize}\\ \text{(in } x=(x_1,\dots x_S)\text{)} & \frac{1}{N}\sum_{n=1}^N x^\top\xi_n=x^\top\xi\\ \text{subject to} & \mathsf{V@R}_{95\,\%}\,(-x^\top\xi)\le 2\$ & 95\%\text{ of all losses}<2\$\\ & x_1+\dots+x_S\le 1.000\$ & \text{budget constraint}\\ & (x\ge 0) & \text{shortselling allowed / not allowed}\end{array}$$

4.5 PROBLEMS

Exercise 4.1. *Is Problem 4.8 always feasible? Where is the Risk? Which statistics are involved?*

Exercise 4.2. *Is Problem 4.8 clever? – Downsides? How can one obtain higher returns?*

Exercise 4.3. *The matrix Ξ in Table 4.3 contains logarithmic, annualized returns of 3 shares at the end of 4 quarters. You are invested with $x=(40\,\%,\,30\,\%,\,30\,\%)$. What is the Value-at-Risk at risk-level $\alpha=30\,\%$, $\alpha=70\,\%$ of your returns?*

5 Axiomatic Treatment of Risk

If in trouble, double.

Börsenweisheit

Definition 5.1 (Artzner et al. [1999, 1997])**.** A positively homogeneous *risk measure*, aka. *risk functional* or *coherent risk measure* is a mapping $\mathcal{R}\colon L^p \to \mathbb{R} \cup \{\infty\}$ with the following properties:

(i) MONOTONICITY: $\mathcal{R}(Y_1) \le \mathcal{R}(Y_2)$ whenever $Y_1 \le Y_2$ almost surely;

(ii) CONVEXITY: $\mathcal{R}\big((1-\lambda)Y_0 + \lambda Y_1\big) \le (1-\lambda)\mathcal{R}(Y_0) + \lambda\mathcal{R}(Y_1)$ for $0 \le \lambda \le 1$;

(iii) TRANSLATION EQUIVARIANCE:[1] $\mathcal{R}(Y+c) = \mathcal{R}(Y) + c$ if $c \in \mathbb{R}$;

(iv) POSITIVE HOMOGENEITY: $\mathcal{R}(\lambda Y) = \lambda\mathcal{R}(Y)$ whenever $\lambda > 0$.

Throughout this lecture shall investigate *positively homogeneous* risk functionals.

Remark 5.2*.* In the present context Y is associated with loss. In the literature the mapping

$$\rho\colon Y \mapsto \mathcal{R}(-Y)$$

is often called *coherent risk measure* instead, when Y is associated with a reward rather than a loss: Whereas $\mathcal{R}$ is more frequent in an actuarial (insurance) context, ρ is typically used in a banking context.

The term *acceptability functional* (or *Utility Function*, cf. Figure 4.3) is frequently used for the concave mapping

$$\mathcal{A}\colon Y \mapsto -\mathcal{R}(-Y). \tag{5.1}$$

Example 5.3 (Simple Examples of risk measures)**.** The functionals

$$\mathcal{R}(Y) := \mathbb{E}\,Y$$

and

$$\mathcal{R}(Y) := \operatorname{ess\,sup} Y$$

are risk measures.

Example 5.4. The functional $\mathcal{R}(Y) := \mathbb{E}\,YZ$ is a risk functional, provided that $Z \ge 0$ and $\mathbb{E}\,Z = 1$.

Proof. By translation equivariance we have that

$$\mathcal{R}(Y) + c = \mathcal{R}(Y + c\,\mathbb{1}) = \mathbb{E}(Y + c\,\mathbb{1})Z = \mathbb{E}\,Y + c\,\mathbb{E}\,Z,$$

hence $\mathbb{E}\,Z = 1$.

Further, we have for all $Y_1 \le Y_2$ that $\mathbb{E}\,Y_1 Z = \mathcal{R}(Y_1) \le \mathcal{R}(Y_1) = \mathbb{E}\,Y_2 Z$, i.e., $\mathbb{E}\,ZY \ge 0$ for all $Y \ge 0$. This can only hold true for $Z \ge 0$. □

Theorem 5.5. *Suppose a risk functional $\mathcal{R}(\cdot)$ is well defined on L^∞ and satisfies (i) and (iii) in Definition 5.1. Then $\mathcal{R}$ is Lipschitz-continuous with respect to $\|\cdot\|_\infty$; the Lipschitz constant is 1.*

[1] In an economic or monetary environment this is often called CASH INVARIANCE instead.

Proof. For X and $Y \in L^\infty$ it is true that $X - Y \le \|X - Y\|_\infty$ and hence $X \le \|Y - X\|_\infty + Y$. By monotonicity and translation equivariance thus

$$\mathcal{R}(X) \le \mathcal{R}(Y + \|Y - X\|_\infty) = \|Y - X\|_\infty + \mathcal{R}(Y)$$

and thus

$$\mathcal{R}(X) - \mathcal{R}(Y) \le \|Y - X\|_\infty\,;$$

interchanging the role of X and Y reveals the result. □

Lemma 5.6. *If $\mathcal{R}_1$ and $\mathcal{R}_2$ are risk measures, then so are*

- $\frac{1}{2}\mathcal{R}_1 + \frac{1}{2}\mathcal{R}_2$ *and*
- $\max\{\mathcal{R}_1, \mathcal{R}_2\}$.

By the Fenchel–Moreau theorem (see below), no other risk measures are possible.

6 Examples of Coherent Risk Functionals

Buy on rumors, sell on facts.

Börsenweisheit

6.1 MEAN SEMI-DEVIATION

The semi-deviation risk measure is[1]

$$\mathcal{R}(Y) := \mathbb{E}\, Y + \beta \cdot \|(Y - \mathbb{E}\, Y)_+\|_p\,, \tag{6.1}$$

where $\beta \in [0, 1]$ and $p \geq 1$.

Proposition 6.1. *The semi-deviation (6.1) is a risk measure.*

Proof. Convexity (ii), translation equivariance (iii) and homogeneity (iv) are evident. To show monotonicity (i) assume that $X \leq Y$. By Jensens inequality (i.e., $x \mapsto x_+$ is convex) we have that $(x + y)_+ \leq x_+ + y_+$ and it follows that

$$\begin{aligned}(X - \mathbb{E}\, X)_+ &= \big(Y - \mathbb{E}\, Y + (X - Y - \mathbb{E}(X - Y))\big)_+ \\ &\leq (Y - \mathbb{E}\, Y)_+ + \big(X - Y - \mathbb{E}(X - Y)\big)_+.\end{aligned}$$

Now we have that $x_+ = x + (-x)_+$ (cf. Footnote 1) and thus further

$$(X - \mathbb{E}\, X)_+ \leq (Y - \mathbb{E}\, Y)_+ + \big(X - Y - \mathbb{E}(X - Y)\big) + \big(Y - X - \mathbb{E}(Y - X)\big)_+.$$

Recall that $Y - X \geq 0$ and further that $\mathbb{E}\, X \leq \mathbb{E}\, Y$, and thus $\big(Y - X - \mathbb{E}(Y - X)\big)_+ \leq Y - X$. Consequently

$$\begin{aligned}(X - \mathbb{E}\, X)_+ &\leq (Y - \mathbb{E}\, Y)_+ + X - Y - \mathbb{E}(X - Y) + Y - X \\ &= (Y - \mathbb{E}\, Y)_+ + \mathbb{E}(Y - X).\end{aligned}$$

It follows that

$$\mathbb{E}\, X + (X - \mathbb{E}\, X)_+ \leq \mathbb{E}\, Y + (Y - \mathbb{E}\, Y)_+.$$

Now multiply the latter inequality with the density Z with $Z \geq 0$, $\mathbb{E}\, Z = 1$ and $\|Z\|_q \leq 1$ to obtain

$$\mathbb{E}\, X + \|(X - \mathbb{E}\, X)_+\|_p \leq \mathbb{E}\, Y + \|(Y - \mathbb{E}\, Y)_+\|_p$$

by Hölder's inequality. Multiplying this inequality with β and adding $(1 - \beta)$ times the inequality $\mathbb{E}\, X \leq \mathbb{E}\, Y$ finally gives monotonicity (i). ◻

6.2 AVERAGE VALUE-AT-RISK

The most important and prominent acceptability functional satisfying all axioms of the Definition is the *Average Value-at-Risk.*

[1] $x_+ := \max\{0,\ x\}$. Note, that $x_+ - (-x)_+ = x$.

Definition 6.2. The *Average Value-at-Risk*[2] at level $\alpha \in [0,1]$ is given by

$$\mathsf{AV@R}_\alpha(Y) := \frac{1}{1-\alpha}\int_\alpha^1 \mathsf{V@R}_p(Y)\,\mathrm{d}p = \frac{1}{1-\alpha}\int_\alpha^1 F_Y^{-1}(u)\,\mathrm{d}u \qquad (0 \le \alpha < 1)$$

and

$$\mathsf{AV@R}_1(Y) := \operatorname{ess\,sup} Y.$$

Proposition 6.3. *Representations of the Average Value-at-Risk include(cf. Footnote 1 on the preceding page)*

$$\begin{aligned}
\mathsf{AV@R}_\alpha(Y) &= \frac{1}{1-\alpha}\int_\alpha^1 F_Y^{-1}(p)\,\mathrm{d}p && (6.2)\\
&= \inf_{q\in\mathbb{R}} q + \frac{1}{1-\alpha}\mathbb{E}(Y-q)_+ && (6.3)\\
&= \sup\left\{\mathbb{E}\, YZ\colon 0 \le Z \le (1-\alpha)^{-1},\ \mathbb{E}\,Z = 1\right\} && (6.4)\\
&= \sup\left\{\mathbb{E}_Q\, Y\colon \frac{\mathrm{d}Q}{\mathrm{d}P} \le \frac{1}{1-\alpha}\right\}\\
&\underset{\text{cf. Remark 6.4}}{=} \mathbb{E}\big(Y \mid Y > \mathsf{V@R}_\alpha(Y)\big). && (6.5)
\end{aligned}$$

The α-quantile $q^* = F_Y^{-1}(\alpha)$ minimizes (6.3).

Remark 6.4. Equation (6.5) is correct, *provided that* $P(Y > \mathsf{V@R}_\alpha(Y)) = 1-\alpha$, i.e., $P(Y \le \mathsf{V@R}_\alpha(Y)) = \alpha$, or $F_Y\left(F_Y^{-1}(\alpha)\right) = \alpha$ (cf. Lemma 4.4 (iv)). In this case (6.5) follows from (6.2).

Proof. Differentiate the objective in (6.3) with respect to q to get the necessary condition of optimality

$$0 = 1 - \frac{1}{1-\alpha}\mathbb{E}\,\mathbb{1}_{\{q<Y\}} = 1 - \frac{1}{1-\alpha}\big(1 - P(Y \le q)\big),$$

i.e., $P(Y \le q) = \alpha$, so that $q^* = F_Y^{-1}(\alpha) = \mathsf{V@R}_\alpha(Y)$. The objective in (6.3) hence is

$$\begin{aligned}
q^* + \frac{1}{1-\alpha}\mathbb{E}(Y-q^*)_+ &= q^* + \frac{1}{1-\alpha}\int_0^1 \left(F_Y^{-1}(u) - q^*\right)_+ \mathrm{d}u\\
&= F_Y^{-1}(\alpha) + \frac{1}{1-\alpha}\int_\alpha^1 F_Y^{-1}(u) - F_Y^{-1}(\alpha)\,\mathrm{d}u\\
&= \frac{1}{1-\alpha}\int_\alpha^1 F_Y^{-1}(u)\,\mathrm{d}u = \mathsf{AV@R}_\alpha(Y).
\end{aligned}$$

See Lemma (6.7) below for the remaining assertion. □

Lemma 6.5. *It holds that*

(i) $\mathsf{AV@R}_0(Y) = \mathbb{E}\,Y$*;*

(ii) $\mathsf{V@R}_\alpha(Y) \le \mathsf{AV@R}_\alpha(Y)$*;*

---

[2]The

- *Average Value-at-Risk, or conditional Value-at-Risk,*

is sometimes also called

- *Conditional Tail Expectation (CTE)*
- *expected shortfall,*
- *tail value-at-risk* or newly
- *super-quantile.*

(iii) $\mathsf{AV@R}_{\alpha'}(Y) \le \mathsf{AV@R}_\alpha(Y)$, *provided that* $\alpha' \le \alpha$, *and particularly*

(iv) $\underbrace{\mathbb{E}\,Y}_{\textit{risk neutral}} \le \underbrace{\mathsf{AV@R}_\alpha(Y)}_{\textit{risk averse}} \le \underbrace{\operatorname{ess\,sup} Y}_{\textit{completely risk averse}}.$

Proof. Indeed, substitute $u \leftarrow F_Y(y)$ and we have $\mathsf{AV@R}_0(Y) = \int_0^1 F_Y^{-1}(u)\,\mathrm{d}u = \int_{\mathbb{R}} y\,\mathrm{d}F_Y(y) = \mathbb{E}\,Y$. In case a density is available then $\mathrm{d}F_Y(y) = f_Y(y)\,\mathrm{d}y$ and thus $\mathsf{AV@R}_0(Y) = \int_{\mathbb{R}} y f_Y(y)\,\mathrm{d}y$.

For the proof of (iii) see the more general Proposition 6.14 below. □

Proposition 6.6. *The Average Value-at-Risk is a risk functional according the axioms of Definition 5.1.*

Proof. Monotonicity, translation equivariance and positive homogeneity are evident by (6.2) and (6.3).

As for convexity let q_0^* (q_1^*, resp.) be optimal in (6.3) for the random variable Y_0 (Y_1, resp.). Set $Y_\lambda := (1-\lambda)Y_0 + \lambda Y_1$ and $q_\lambda := (1-\lambda)q_0^* + \lambda q_1^*$. Then

$$\begin{aligned} \mathsf{AV@R}_\alpha\big((1-\lambda)Y_0 + \lambda Y_1\big) &\le q_\lambda + \frac{1}{1-\alpha}\,\mathbb{E}\,(Y_\lambda - q_\lambda)_+ \\ &= (1-\lambda)q_0^* + \lambda q_1^* + \frac{1}{1-\alpha}\,\mathbb{E}\,\big((1-\lambda)\left(Y_0 - q_0^*\right) + \lambda\left(Y_1 - q_1^*\right)\big)_+ \\ &\le (1-\lambda)q_0^* + \lambda q_1^* + \frac{1-\lambda}{1-\alpha}\,\mathbb{E}\left(Y_0 - q_0^*\right)_+ + \frac{\lambda}{1-\alpha}\,\mathbb{E}\left(Y_1 - q_1^*\right)_+ \qquad (6.6) \\ &= (1-\lambda)\,\mathsf{AV@R}_\alpha(Y_0) + \lambda\,\mathsf{AV@R}_\alpha(Y_1), \end{aligned}$$

where we have used Jensen's inequality in (6.6) for the convex function $y \mapsto (y-q)_+$; thus the assertion. □

Lemma 6.7. *It holds that*

$$\mathsf{AV@R}_\alpha(Y) = \max\left\{\mathbb{E}\,YZ\colon 0 \le Z \le \frac{1}{1-\alpha},\ \mathbb{E}\,Z = 1\right\} = \min_{c\in\mathbb{R}}\left\{c + \frac{1}{1-\alpha}\,\mathbb{E}\,(Y-c)_+\right\}. \qquad (6.7)$$

Proof. We provide a prove of the statement for discrete random variables based on duality.

Recall first that the linear problems

$$\begin{array}{lll} \text{minimize (in } x) & c^\top x \\ \text{subject to} & A_1 x = b_1 \\ & A_2 x \ge b_2 \\ & x \ge 0 \end{array} \quad \text{and} \quad \begin{array}{ll} \text{maximize (in } \lambda,\mu) & \lambda^\top b_1 + \mu^\top b_2 \\ \text{subject to} & \lambda^\top A_1 + \mu^\top A_2 \le c^\top \\ & \mu \ge 0 \end{array}$$

are dual to each other. We rewrite the initial problem (6.7)

$$\begin{array}{ll} -\text{minimize (in } Z) & \sum_i -p_i Y_i Z_i \\ \text{subject to} & \sum_i p_i Z_i = 1 \\ & -p_i Z_i \ge -p_i \frac{1}{1-\alpha} \\ & Z_i \ge 0 \end{array}$$

with $c_i = -p_i Y_i$, $A_{1,i} = p_i$, $b_1 = 1$, $A_{2,i} = -p_i$ and $b_{2,i} = -\frac{p_i}{1-\alpha}$. Inserting in the dual gives

$$\begin{array}{ll} -\text{maximize (in } \lambda,\mu) & \lambda - \sum_i \frac{p_i}{1-\alpha}\mu_i \\ \text{subject to} & \lambda p_i - p_i \mu_i \le -p_i Y_i, \\ & \mu_i \ge 0. \end{array} \qquad (6.8)$$

Now note that the latter two inequalities are $\mu_i \ge 0$ and $\mu_i \ge \lambda + Y_i$. The maximum in (6.8) is attained for $\mu_i = \max\{0, \lambda + Y_i\}$. Hence (6.8) rewrites as

$$-\text{maximize}_\lambda = \lambda - \frac{1}{1-\alpha}\sum_i p_i\,(Y_i + \lambda)_+ = \text{minimize}_\lambda = -\lambda + \frac{1}{1-\alpha}\sum_i p_i\,(Y_i + \lambda)_+\,,$$

from which the assertion follows. □

6.3 ENTROPIC VALUE-AT-RISK

Definition 6.8. The *Entropic Value-at-Risk* at risk level $\alpha \in [0,1)$ is

$$\mathsf{EV@R}_\alpha(Y) = \inf_{t>0} \frac{1}{t} \log \frac{1}{1-\alpha} \mathbb{E}\, e^{tY}. \tag{6.9}$$

It is a risk measure satisfying all conditions (i)–(iv) in Definition 5.1.

Remark 6.9. For small values of t, it holds that

$$\frac{1}{t} \log \mathbb{E}\, e^{tY} \approx \mathbb{E}\, X + \frac{t}{2} \operatorname{var} Y + O(t^2)$$

(cf. (3.28)).

Proof. Indeed, $\mathbb{E}\, e^{tY} = 1 + t\,\mathbb{E}\, Y + \frac{1}{2} t^2\, \mathbb{E}\, Y^2$. Using $\log(1+x) \approx x - \frac{1}{2}x^2 + O(x^3)$ gives

$$\begin{aligned} \log \mathbb{E}\, e^{tY} &\approx \log\left(1 + t\,\mathbb{E}\, Y + \frac{1}{2} t^2\, \mathbb{E}\, Y^2\right) \\ &= t\,\mathbb{E}\, Y + \frac{1}{2} t^2\, \mathbb{E}\, Y^2 - \frac{1}{2}\big(t\,\mathbb{E}\, Y\big)^2 + O(t^3) \\ &= t\,\mathbb{E}\, Y + \frac{t^2}{2} \operatorname{var} Y + O(t^3) \end{aligned}$$

and thus the assertion. ◻

Remark 6.10. A good guess for the optimal t^* in (6.9) thus is $t^* \approx \sqrt{\frac{\operatorname{var} Y}{2 \log \frac{1}{1-\alpha}}}$.

6.4 SPECTRAL RISK MEASURES

Definition 6.11. Spectral risk measures[3] are

$$\mathcal{R}_\sigma(Y) := \int_0^1 \sigma(u) F_Y^{-1}(u)\, \mathrm{d}u$$

for some spectral function $\sigma\colon [0,1] \to \mathbb{R}$; occasionally, the function $\sigma(\cdot)$ is also called *spectrum*.

Proposition 6.12. *Spectral functions $\sigma\colon [0,1) \to \mathbb{R}_{\ge 0}$ necessarily satisfy*

(i) $\int_0^1 \sigma(u)\, \mathrm{d}u = 1$ (by translation equivariance),

(ii) $\sigma(\cdot) \ge 0$ (by monotonicity) and

(iii) $\sigma(\cdot)$ is nondecreasing (by convexity).

Proof. See Exercise 6.3. ◻

Remark 6.13. The Average Value-at-Risk is a spectral risk measure for the spectrum

$$\sigma(u) := \begin{cases} 0 & \text{if } u < \alpha, \\ \frac{1}{1-\alpha} & \text{if } u \ge \alpha. \end{cases}$$

Proposition 6.14. *Suppose that $\int_\alpha^1 \sigma_1(u)\, \mathrm{d}u \le \int_\alpha^1 \sigma_2(u)\, \mathrm{d}u$ for all $\alpha \in (0,1)$, then $\mathcal{R}_{\sigma_1}(Y) \le \mathcal{R}_{\sigma_2}(Y)$.*

[3] Also: distortion risk funtionals

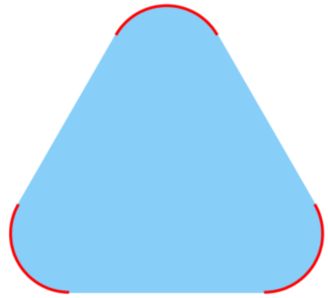


Figure 6.1: AV@R$_\alpha$'s are extreme points, so there is some Choquet-Bishop-de Leeuw Representation for $\mathcal{A}$ (Krein-Milman Theorem).

Proof. We verify the statement for Y bounded. Set $\Sigma_i(u) := \int_u^1 \sigma_i(p)\,\mathrm{d}p$. Then

$$\mathcal{R}_{\sigma_1}(Y) = \int_0^1 \sigma_1(u) F_Y^{-1}(u)\,\mathrm{d}u = -\int_0^1 F_Y^{-1}(u)\,\mathrm{d}\Sigma_1(u),$$

and by integration by parts

$$\mathcal{R}_{\sigma_1}(Y) = -F_Y^{-1}(u)\Sigma_1(u)\big|_{u=0}^1 + \int_0^1 \Sigma_1(u)\,\mathrm{d}F_Y^{-1}(u) = F_Y^{-1}(0) + \int_0^1 \Sigma_1(u)\,\mathrm{d}F_Y^{-1}(u).$$

Note, that $\Sigma_1(\cdot) \le \Sigma_2(\cdot)$ by assumption and $F_Y^{-1}(\cdot)$ is an increasing function, thus

$$\mathcal{R}_{\sigma_1}(Y) = F_Y^{-1}(0) + \int_0^1 \Sigma_1(u)\,\mathrm{d}F_Y^{-1}(u) \le F_Y^{-1}(0) + \int_0^1 \Sigma_2(u)\,\mathrm{d}F_Y^{-1}(u) = \mathcal{R}_{\sigma_2}(Y).$$

□

Lemma 6.15. $\mathcal{R}_\mu(Y) := \int_0^1 \mathrm{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha)$ *is a spectral risk measure, provided that*

- ▹ $\int_0^1 \mu(\mathrm{d}\alpha) = 1$ *(to ensure translation equivariance) and*
- ▹ $\mu(\cdot) \ge 0$ *(to ensure monotonicity).*

Proof. Indeed,

$$\mathcal{R}_\mu(Y) = \int_0^1 \mathrm{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha) = \int_0^1 \mathrm{V@R}_\alpha(Y)\,\sigma(\alpha)\,\mathrm{d}\alpha,$$

where $\sigma(p) = \int_0^p \frac{\mu(\mathrm{d}\alpha)}{1-\alpha}$ is the *spectrum*. □

Lemma 6.16. *For $Y \ge 0$ a.s. we have the representation*

$$\mathcal{R}_\sigma(Y) = \int_0^\infty \Sigma(F_Y(q))\,\mathrm{d}q \qquad \text{(if } Y \ge 0 \text{ a.s.)},$$

where $\Sigma(\alpha) = \int_\alpha^1 \sigma(p)\,\mathrm{d}p$ is the negative antiderivative.

6.5 KUSUOKA'S REPRESENTATION OF LAW INVARIANT RISK MEASURES

A supremum of Choquet representations.

Theorem 6.17. *Suppose $\mathcal{R}$ is a law invariant Risk measure. Then it has the Kusuoka-representation*

$$\mathcal{R}\left(Y\right)=\sup_{\mu\in\mathcal{M}}\int_0^1 \mathsf{AV@R}_\alpha\left(Y\right)\mu\left(\mathrm{d}\alpha\right)$$

($\mathcal{M}$ is a set of positive measures on $[0,1]$).

Proof. We have that $\mathcal{R}(Y)=\sup_Z \mathbb{E}\,YZ-\mathcal{R}^*(Z)=\sup_{Z\in\mathcal{Z}}\mathbb{E}\,YZ$ as $\mathcal{R}^*(Z)\in\{0,\infty\}$. For $Z\in\mathcal{Z}$ given, let Y' have the same distribution as Y so that Y' and Z are comonotone. Then

$$\mathcal{R}(Y)=\mathcal{R}(Y')=\sup_{Z\in\mathcal{Z}}\mathbb{E}\,YZ-\mathcal{R}^*(Z)=\sup_{Z\in\mathcal{Z}}\int_0^1 F_Y^{-1}(u)F_Z^{-1}(u)\,\mathrm{d}u=\sup_{\sigma\in\Sigma}\int_0^1\sigma(u)F_Y^{-1}(u)\,\mathrm{d}u,$$

where $\Sigma=\{F_Z(\cdot)\colon Z\in\mathcal{Z}\}$ collects all distribution functions of $\mathcal{Z}$. Hence the result. □

Proposition 6.18. *For any law invariant risk functional $\mathcal{R}$ it holds that*

$$\mathbb{E}\,Y\leq\mathcal{R}(Y).$$

Proof. Consider the functional $\mathcal{R}_\sigma(\cdot)$ first. Find $\tilde{u}$ such that $\sigma(u)\leq 1$ whenever $u\leq\tilde{u}$ and $\sigma(u)\geq 1$ for $u\geq\tilde{u}$. Note as well that $\int_0^{\tilde{u}}1-\sigma(u)\,\mathrm{d}u=\int_{\tilde{u}}^1\sigma(u)-1\,\mathrm{d}u$, as $\left(\int_0^{\tilde{u}}+\int_{\tilde{u}}^1\right)\sigma(u)\,\mathrm{d}u=\left(\int_0^{\tilde{u}}+\int_{\tilde{u}}^1\right)1\,\mathrm{d}u=1$. Then

$$\begin{aligned}\int_0^{\tilde{u}}(1-\sigma(u))F_Y^{-1}(u)\,\mathrm{d}u&\leq\int_0^{\tilde{u}}(1-\sigma(u))F_Y^{-1}(\tilde{u})\,\mathrm{d}u\\&=\int_{\tilde{u}}^1(\sigma(u)-1)F_Y^{-1}(\tilde{u})\,\mathrm{d}u\leq\int_{\tilde{u}}^1(\sigma(u)-1)F_Y^{-1}(u)\,\mathrm{d}u.\end{aligned}$$

The assertion follows from Kusuoka's representation. □

Proposition 6.19. *For any law invariant risk measure $\mathcal{R}$ and sub-sigma algebra $\mathcal{G}$ we have that*

$$\mathcal{R}\big(\mathbb{E}(Y|\,\mathcal{G})\big)\leq\mathcal{R}\left(Y\right).$$

Proof. Note that $x\mapsto(x-q)_+$ is convex. Thus, by the conditional Jensen inequality

$$\left(\mathbb{E}(Y|\,\mathcal{G})-q\right)_+\leq\mathbb{E}\left((Y-q)_+|\mathcal{G}\right).$$

It follows that

$$\begin{aligned}\mathsf{AV@R}_\alpha\big(\mathbb{E}(Y|\mathcal{G})\big)&=\min_{q\in\mathbb{R}}q+\frac{1}{1-\alpha}\mathbb{E}\left((\mathbb{E}\,Y|\,\mathcal{G})-q\right)_+\\&\leq\min_{q\in\mathbb{R}}q+\frac{1}{1-\alpha}\mathbb{E}\,\mathbb{E}\left((Y-q)_+|\,\mathcal{G}\right)\\&=\min_{q\in\mathbb{R}}q+\frac{1}{1-\alpha}\mathbb{E}\left((Y-q)_+\right)=\mathsf{AV@R}_\alpha(Y).\end{aligned}$$

The assertion follows form Kusuoka's representation. □

Remark 6.20. For the Average Value-at-Risk we have that $\mathsf{AV@R}_\alpha\left(\mathbb{E}(Y|\mathcal{F})\right)\leq\mathsf{AV@R}_\alpha\left(Y\right)$, where $\mathcal{F}$ is a sub-sigma algebra, and thus $\mathcal{R}\left(\mathbb{E}(Y|\mathcal{F})\right)\leq\mathcal{R}\left(Y\right)$ for law invariant risk functionals by Kusuoka's theorem.

Proposition 6.21 (Cf. Föllmer and Schied [2004, Theorem 4.67])**.** $\mathsf{AV@R}_\alpha(\cdot)$ *is the smallest* law invariant *coherent risk functional dominating* $\mathsf{V@R}_\alpha(\cdot)$ *(cf. Lemma 6.5 (ii) and Exercise 6.4).*

Proof. By translation equivariance and for $Y\in L^\infty$ we may assume that $Y>0$. Set $A:=\{Y>\mathsf{V@R}_\alpha(Y)\}$ and consider $X:=Y\cdot\mathbb{1}_{A^\complement}+\mathbb{E}(Y|A)\cdot\mathbb{1}_A$. Notice, that $X=\mathbb{E}\left(Y\,\middle|\,Y\cdot\mathbb{1}_{A^\complement}\right)$ and $\mathsf{V@R}_\alpha(X)=\mathbb{E}(Y|A)$. Suppose the coherent risk functional $\mathcal{R}(\cdot)$ dominates $\mathsf{V@R}_\alpha(\cdot)$, then

$$\mathcal{R}(Y)\geq\mathcal{R}(X)\geq\mathsf{V@R}_\alpha(X)=\mathbb{E}(Y|A)=\mathsf{AV@R}_\alpha(Y)$$

by (6.5). □

6.6 APPLICATION IN INSURANCE

The Dutch Premium Principle

A comprehensive list of Kusuoka representations for important risk functionals is provided in [Pflug and Römisch, 2007]. A compelling example is the absolute semi-deviation risk measure (the Dutch premium principle, introduced in [van Heerwaarden and Kaas, 1992]) for some fixed $\theta \in [0,1]$,

$$\mathcal{R}_\theta(Y) := \mathbb{E}\left[Y + \theta \cdot (Y - \mathbb{E} Y)_+\right],$$

assigning an additional loading of θ to any loss L exceeding the net-premium price $\mathbb{E} L$. Its Kusuoka representation is

$$\mathcal{R}_\theta(Y) = \sup_{0 \le \mu \le 1} (1 - \theta \cdot \mu)\, \mathbb{E} Y + \theta\, \mu \cdot \mathsf{AV@R}_{1-\mu}(Y)$$

(cf. Shapiro [2012], Shapiro et al. [2021]).

Theorem 6.22. *$\mathcal{R}_\theta(\cdot)$ is a convex risk functional.*

Proof. We verify that $\mathcal{R}_\theta(\cdot)$ is convex. Indeed, define $Y_\lambda := (1-\lambda)Y_0 + \lambda Y_1$. Then

$$(Y_\lambda - \mathbb{E} Y_\lambda)_+ = \left((1-\lambda)(Y_0 - \mathbb{E} Y_0) + \lambda (Y_1 - \mathbb{E} Y_1)\right)_+ \le (1-\lambda)(Y_0 - \mathbb{E} Y_0)_+ + \lambda (Y_1 - \mathbb{E} Y_1)_+$$

and hence

$$\begin{aligned}
\mathbb{E}\left[Y_\lambda + \theta \cdot (Y_\lambda - \mathbb{E} Y_\lambda)_+\right] &\le (1-\lambda)\, \mathbb{E} Y_0 + \lambda\, \mathbb{E} Y_1 + \theta(1-\lambda)(Y_0 - \mathbb{E} Y_0)_+ + \lambda (Y_1 - \mathbb{E} Y_1)_+ \\
&= (1-\lambda)\left(\mathbb{E} Y_0 + \theta (Y_0 - \mathbb{E} Y_0)_+\right) + \lambda \left(\mathbb{E} Y_1 + \theta (Y_1 - \mathbb{E} Y_1)_+\right) \\
&= (1-\lambda)\mathcal{R}_\theta(Y_0) + \lambda \mathcal{R}_\theta(Y_1),
\end{aligned}$$

i.e., $\mathcal{R}_\theta$ is convex. □

6.7 PROBLEMS

Exercise 6.1. *Compute the Average Value-at-Risk for the returns given in Table 9.2 for $\alpha = 20\,\%$, 40%, 60% and $\alpha = 80\%$.*

Exercise 6.2 (Cf. Exercise 4.3)**.** *The matrix Ξ in Table 4.3 contains logarithmic, annualized returns of 3 shares at the end of 4 quarters.*

(i) Compute the Average Value-at-Risk for the risk levels $\alpha = 20\,\%$, 40%, 60% and $\alpha = 80\%$ for each asset.

(ii) You are invested with $x = (40\,\%, 30\,\%, 30\,\%)$. What is the Value-at-Risk of your returns at the above risk-levels?

Exercise 6.3. *Verify Proposition 6.12. Hint: try the random variables $\mathbb{1}_{[\lambda,1]}(U)$ to show monotonicity, and $Y_0 := \mathbb{1}_{[u,1]}(U)$ and $Y_1 := \mathbb{1}_{[u-\Delta,1-\Delta]}(U)$ for convexity.*

Exercise 6.4. *Give a risk functional $\mathcal{R}$ and a random variable $Y \in L^\infty$ so that $\mathcal{R}(Y) > \mathbb{E} Y$.*

7 Portfolio Optimization Problems Involving Risk Measures

. . . : daß keines von ihnen verloren gehe.

Edith Stein, ESGA, Band 1

7.1 INTEGRATED RISK MANAGEMENT FORMULATION

The portfolio optimization problem we want to consider here for simplicity and introduction is (cf. Figure 4.3 and (iv) in Theorem 9.10)

$$\begin{array}{rl} \underset{\text{in } x}{\text{maximize}} & -\mathcal{R}\left(-x^\top \xi\right) = \mathcal{A}\left(x^\top \xi\right) \\ \text{subject to} & x^\top \mathbb{1} \le 1€, \\ & x \ge 0, \end{array}$$

where $\mathcal{A}(\cdot) := -\mathcal{R}(-\cdot)$ is an *acceptability functional*, cf. (5.1), Remark 5.2. The problem is notably unbounded without shortselling constraints. Equivalent is the formulation (cf. Figure 4.3, again)

$$\begin{array}{rl} \underset{\text{in } x}{\text{minimize}} & \mathcal{R}\left(-x^\top \xi\right) \\ \text{subject to} & x^\top \mathbb{1} \le 1€, \\ & x \ge 0, \end{array}$$

which is apparently a convex problem formulation.

Typical risk functionals are $\mathcal{R}(Y) := (1-\gamma)\,\mathbb{E}\,Y + \gamma\,\mathsf{AV@R}_\alpha(Y)$.

7.2 MARKOWITZ TYPE FORMULATION

Recall from Figure 4.3 that $\mathbb{E}\,Y + \mathcal{R}(-Y)$ (≥ 0) is a one-sided deviation from the mean, which can be interpreted as risk. The formulation

$$v(\mu) := \begin{array}{rl} \underset{\text{in } x}{\text{minimize}} & \mathbb{E}\,x^\top \xi + \mathcal{R}\left(-x^\top \xi\right) \\ \text{subject to} & \mathbb{E}\,x^\top \xi \ge \mu, \\ & x^\top \mathbb{1} \le 1€, \\ & (x \ge 0) \end{array} \tag{7.1}$$

specifies a convex problem in x, as the objective (i.e., $\mathcal{R}$) is convex, the constraints are even linear. The function v is nondecreasing and it holds that $0 \le \mathbb{E}\,Y + \mathcal{R}\,(-Y)$, i.e., $v(\mu) \ge 0$.

Let x_μ denote the optimal diversification in (7.1). For $\lambda \in (0,1)$ set $\mu_\lambda := (1-\lambda)\mu_0 + \lambda\mu_1$ and $x_{\mu_\lambda} := (1-\lambda)x_{\mu_0} + \lambda x_{\mu_1}$. By linearity, x_{μ_λ} is feasible for $v(\mu_\lambda)$ and we have from convexity of $\mathcal{R}$ that

$$v(\mu_\lambda) \le \mathbb{E}\,x_{\mu_\lambda}^\top \xi + \mathcal{R}\left(-x_{\mu_\lambda}^\top \xi\right) \le (1-\lambda)v(\mu_0) + \lambda v(\mu_1),$$

the function $v(\cdot)$ thus is convex. This gives rise to the efficient frontier $\mu \mapsto \begin{pmatrix} v(\mu) \\ \mu \end{pmatrix}$ (which is concave) and an accordant tangency portfolio.

Example 7.1. Consider $\mathcal{R}(Y) := (1-\gamma)\,\mathbb{E}\,Y + \gamma\,\mathsf{AV@R}_\alpha(Y)$, then the problem of integrated risk management is (cf. Figure 4.3, and again)

$$\begin{array}{rl} \text{minimize} & \\ \text{in } x & \gamma\cdot\mathbb{E}\,x^\top\xi + \gamma\cdot\mathsf{AV@R}_\alpha\left(-x^\top\xi\right) \\ \text{subject to} & \mathbb{E}\,x^\top\xi \ge \mu, \\ & x^\top\mathbb{1} \le 1€, \\ & (x\ge 0)\,, \end{array} \tag{7.2}$$

with parameters $\alpha,\gamma\in(0,1)$.

Recall that $\mathsf{AV@R}_\alpha(Y) = \min_{q\in\mathbb{R}}\left\{q + \frac{1}{1-\alpha}\,\mathbb{E}(Y-q)_+\right\}$. The discrete model problem (7.2) thus may be rewritten as

$$\begin{array}{rl} \text{minimize} & \\ \text{in } x,\ q & \gamma\cdot p^\top\Xi x + \gamma\cdot\left(q + \frac{1}{1-\alpha}p^\top\left(-\Xi x - q\right)_+\right) \\ \text{subject to} & p^\top\Xi x \ge \mu, \\ & x^\top\mathbb{1} \le 1€, \\ & (x\ge 0)\,. \end{array}$$

To eliminate the nonlinear expression $(\cdot)_+$ define the slack variable $z^i := (-q - x^\top\xi_i)_+$ $(i=1,\dots,n)$ and note that $z^i\ge 0$ and $-q - x^\top\xi_i \le z_i$. So we get the *linear program*

$$\begin{array}{rl} \text{minimize} & \\ \text{in } x,\ z,\ q & \gamma\cdot p^\top\Xi x + \gamma\cdot q + \frac{\gamma}{1-\alpha}p^\top z \\ \text{subject to} & -q - x^\top\xi_i \le z_i \quad (i=1,\dots n)\,, \\ & p^\top\Xi x \ge \mu, \\ & x^\top\mathbb{1} \le 1€, \\ & z\ge 0,\ (x\ge 0)\,, \end{array}$$

or re-written in matrix-form

$$\begin{array}{rl} \text{minimize} & \\ \text{in } x,\ z,\ q & -\gamma p^\top\Xi x + \gamma\cdot q + \frac{\gamma}{1-\alpha}p^\top z \\ \text{subject to} & \begin{pmatrix} -\Xi & -I_n & -\mathbb{1}_n \\ \mathbb{1}_S^\top & 0\dots 0 & 0 \\ -p^\top\Xi & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ z \\ q\end{pmatrix} \le \begin{pmatrix} 0 \\ 1 \\ -\mu\end{pmatrix}, \\ & z\ge 0,\ (x\ge 0)\,. \end{array}$$

Example 7.2. Consider $\mathcal{R}(Y) := \mathsf{EV@R}_\alpha(Y)$, then the problem of integrated risk management is

$$\begin{array}{rl} \text{minimize} & \\ \text{in } x,t & \mathbb{E}\,x^\top\xi + t\log\frac{1}{1-\alpha}\,\mathbb{E}\,e^{-x^\top\xi/t} \\ \text{subject to} & \mathbb{E}\,x^\top\xi \ge \mu, \\ & x^\top\mathbb{1} \le 1€, \\ & t>0,\ (x\ge 0)\,. \end{array} \tag{7.3}$$

7.3 ALTERNATIVE FORMULATION

The formulation

$$\begin{array}{rrl} \tilde v(c) := & \text{maximize} & \\ & \text{in } x & \mathbb{E}\,x^\top\xi \\ & \text{subject to} & -\mathcal{R}\left(-x^\top\xi\right) \ge c, \\ & & x^\top\mathbb{1} \le 1€, \\ & & (x\ge 0) \end{array}$$

specifies a convex problem in $x \in \mathbb{R}^S$ as well (the objective is linear, the constraints convex). It holds that $c \le -\mathcal{R}\left(-x^\top \xi\right) \le \mathbb{E}\, x^\top \xi$ and thus $\tilde{v}(c) \ge c$. The function $\tilde{v}(c)$ is concave and the frontier $c \mapsto \begin{pmatrix} c \\ \tilde{v}(c) \end{pmatrix}$ (or $c \mapsto \begin{pmatrix} c \\ \tilde{v}(c) - c \end{pmatrix}$) is a concave efficient frontier, which again gives rise for a tangency portfolio.

Expected Utility Theory

8

Buy on bad news, sell on good news.

Börsenweisheit

The concept of utility functions dates back to Oskar Morgenstern[1] and John von Neumann,[2] expected utilities to Kenneth Arrow[3] and John W. Pratt.[4]

Preference is given to Y over X, if

$$\mathbb{E}\, u(X) \le \mathbb{E}\, u(Y). \tag{8.1}$$

8.1 EXAMPLES OF UTILITY FUNCTIONS

The *exponential utility* for $\gamma \ge 0$ is defined as

$$u(x) = 1 - e^{-\gamma x}. \tag{8.2}$$

For $\alpha > 0$, $\alpha \ne 1$ the *polynomial utility* functions are defined as

$$u(x) = \begin{cases} \frac{x^{1-\alpha}}{1-\alpha} & x \ge 0, \\ -\infty & x < 0; \end{cases} \tag{8.3}$$

they are sometimes also termed power utility functions,

$$u(x) = \begin{cases} \frac{x^{\kappa}}{\kappa} & x \ge 0, \\ -\infty & x < 0; \end{cases} \tag{8.4}$$

and in case of $\alpha = 1$,

$$u(x) = \begin{cases} \log x & x \ge 0, \\ -\infty & x < 0. \end{cases}$$

Definition 8.1 (HARA utilities)**.** Hyperbolic risk aversion (HARA) utilities are $U(w) = \frac{1-\gamma}{\gamma}\left(\frac{aw}{1-\gamma} + b\right)^{\gamma}$.

8.2 ARROW–PRATT MEASURE OF ABSOLUTE RISK AVERSION

Definition 8.2. The *local risk aversion coefficient at* c (cf. Arrow–Pratt measure of absolute risk-aversion (ARA), also known as the coefficient of absolute risk), is

$$A(c) = -\frac{u''(c)}{u'(c)},$$

the coefficient for relative risk aversion is

$$R(c) = -\frac{c \cdot u''(c)}{u'(c)}.$$

[1] 1902 (in Görlitz) – 1977
[2] 1903 – 1958
[3] 1921–2017, Nobel memorial price in economic sciences in 1972
[4] 1931

For a motivation consider the Taylor-series expansion $u(y) \approx u(x) + (y-x)u'(x) + \frac{(y-x)^2}{2}u''(x)$. At $x = \mathbb{E}\,Y$, $y = Y$ and after taking expectations we obtain

$$\mathbb{E}\,u(Y) \approx u(\mathbb{E}\,Y) + \mathbb{E}(Y - \mathbb{E}\,Y) \cdot u'(\mathbb{E}\,Y) + \frac{\mathbb{E}(Y-\mathbb{E}\,Y)^2}{2}u''(\mathbb{E}\,Y) = u(\mathbb{E}\,Y) + \frac{\operatorname{var} Y}{2}u''(\mathbb{E}\,Y). \tag{8.5}$$

Now apply a Taylor-series expansion to the inverse $u^{-1}(x) \approx u^{-1}(y) + \frac{x-y}{u'(u^{-1}(y))}$ with $x = \mathbb{E}\,u(Y)$ and $y = u(\mathbb{E}\,Y)$ to get

$$u^{-1}(\mathbb{E}\,u(Y)) \approx \mathbb{E}\,Y + \frac{\mathbb{E}\,u(Y) - u(\mathbb{E}\,Y)}{u'(\mathbb{E}\,Y)} \approx \mathbb{E}\,Y + \underbrace{\frac{u''(\mathbb{E}\,Y)}{2u'(\mathbb{E}\,Y)}}_{\text{Arrow–Pratt at } \mathbb{E}\,Y} \cdot \operatorname{var} Y,$$

by (8.5).

Example 8.3. The risk aversion coefficient is $-\gamma$ (thus constant) for the utility function (8.2), while $A(c) = \frac{\alpha}{c}$ and $R(c) = \alpha$ for the utility given in (8.3).

Example 8.4. Consider $u(x) = \log x$, then $A(c) = -\frac{u''(c)}{u'(c)} = \frac{1}{c}$.

8.3 EXAMPLE: ST. PETERSBURG PARADOX[5]

Consider the following game. A fair coin is tossed until heads appears for the first time and suppose this happens at the Nth toss. The player will then get 2^{N-1} euros. What is the fair amount a player should pay in order to play the game?

The fee is given by the expected payout. By definition of the geometric distribution:

$$P(N = k) = 2^{-k} \qquad k = 1, 2, \dots$$

Therefore the expected payout is

$$\mathbb{E}\,2^{N-1} = \sum_{k=1}^{\infty} 2^{-k}2^{k-1} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty.$$

This result obviously does not make sense. Several approaches were developed by N. Bernoulli and G. Cramer. Insetad of calculating the expected payout $\mathbb{E}\left[2^{N-1}\right]$, $c = u^{-1}\left(\mathbb{E}\left[u(2^{N-1})\right]\right)$ with $u(x) = \log(x)$ or $u(x) = \sqrt{x}$ is calculated. In the case of $u(x) = \log(x)$, it follows that

$$\begin{aligned}\mathbb{E}\log 2^{N-1} &= \sum_{k=1}^{\infty} 2^{-k}(k-1)\log 2 = \log(2)\sum_{k=0}^{\infty} k \cdot 2^{-k} \\ &= \frac{\log(2)}{4}\sum_{k=0}^{\infty} k2^{k-1} = \frac{\log 2}{4}\frac{1}{\left(1-\frac{1}{2}\right)^2} = \log 2\end{aligned}$$

and $c = e^{\log 2} = 2$.

In case of $u(x) = \sqrt{x}$,

$$\mathbb{E}\sqrt{2^{N-1}} = \sum_{k=1}^{\infty} 2^{\frac{k-1}{2}}2^{-k} = \frac{1}{\sqrt{2}}\sum_{k=1}^{\infty}\left(\frac{\sqrt{2}}{2}\right)^k = \frac{1}{\sqrt{2}}\frac{\frac{\sqrt{2}}{2}}{1-\frac{\sqrt{2}}{2}} = \frac{1}{2-\sqrt{2}}$$

and therefore $c = \left(2-\sqrt{2}\right)^{-2} \approx 2.914$.

The expected payout was weighted with $u(\cdot)$ which yields a finite value. The weighting with u can be interpreted as giving less importance to very high payouts which have small probabilities of occurring.

Remark. Note the shape of both $\log(x)$ and $\sqrt{x}$. Such functions are called utility functions.

[5] by ruben schlotter

8.4 PREFERENCES AND UTILITY FUNCTIONS

The main aim is to

- model decisions under uncertainty
- compare random payouts (lotteries)

Definition 8.5. A function $F\colon \mathbb{R} \to [0,1]$ is called (cumulative) distribution function on $\mathbb{R}$, if

- F is monotone increasing and right continuous.
- $\lim_{\{x\to-\infty} F(x) = 0$ and $\lim_{\{x\to\infty} F(x) = 1$

Let $\mathcal{M}$ be the set of all distribution functions on$\mathbb{R}$. We define a relation on$\mathcal{M}$. A preference on$\mathcal{M}$ is a relation $\preccurlyeq$ such that

- $F \preccurlyeq F$ for all $F \in \mathcal{M}$
- $(F \preccurlyeq G) \wedge (G \preccurlyeq H)$ implies that $F \preccurlyeq H$ for all $F, G, H \in \mathcal{M}$
- Either $(F \preccurlyeq G)$ or $(G \preccurlyeq F)$

$F \preccurlyeq G$ is interpreted as G is preferred over F. F and G are called equivalent, denoted by$F \sim G$ if$F \preccurlyeq G$ and$G \preccurlyeq F$. The preference$\preccurlyeq$satisfies the continuity axiom if for all $F, G, H \in \mathcal{M}$ such that $F \preccurlyeq G \preccurlyeq H$ there exists an $\alpha \in [0,1]$ with

$$(1-\alpha)F + \alpha H \sim G.$$

The preference$\preccurlyeq$satisfies the independence of irrelevant alternatives axiom, if for all $F, G, H \in \mathcal{M}$ and all $\alpha \in [0,1]$

$$F \preccurlyeq G \Longleftrightarrow (1-\alpha)F + \alpha H \preccurlyeq (1-\alpha)G + \alpha H$$

Remark. The continuity axiom means that "good" and "bad" risks can be pooled into an average one.

A preference$\preccurlyeq$has a numerical representation if there is a mapping $U\colon \mathcal{M} \to [-\infty, \infty)$, such that

$$F \preccurlyeq G \Longleftrightarrow U(F) \le U(G)$$

This representation has a v*on Neumann-Morgenstern representation* if there exists another function $u\colon \mathbb{R} \to [-\infty, \infty)$ such that for all X with (cumulative) distribution function F

$$U(F) = \mathbb{E}\, u(X)$$

Theorem 8.6. *Let $\preccurlyeq$ be a preference on $\mathcal{M}$. Then the following are equivalent*

(i) *$\preccurlyeq$satisfies the continuity and the independence axiom*

(ii) *$\preccurlyeq$ has a vNM representation*

Definition 8.7. $u\colon \mathbb{R} \to [-\infty, \infty)$ is called *Bernoulli-utility function* if u is monotone increasing and strictly concave.

Theorem 8.8. *Let $\preccurlyeq$be a preference with a vNM-representation. Then u is a Bernoulli utility function if and only if*

(i) $c \preccurlyeq d \Longleftrightarrow c \le d$ *for all* $c, d \in \mathbb{R}$

(ii) $X \preccurlyeq \mathbb{E}\, X$

Let$u$ be a Bernoulli utility function (wrt. .: $(\preccurlyeq_u)$ and X have finite expectation then define the *certainty equivalent* of X by

$$c := c(X,u) \in \mathbb{R}, \qquad \text{such that } c \sim_u X$$

In the introductory example the certainty equivalent of the payout N with respect to 2 different Bernoulli utility functions was calculated.

9 Stochastic Orderings

> Ich bin so glücklich, ich habe meinen Posten verloren. Mein Chef ist nämlich in Konkurs gegangen. Mich bringt niemand mehr in ein Bankhaus.
>
> Arnold Schönberg, 1874–1951, an David Josef Bach

A particular utility function $u(\cdot)$ is occasionally considered as artifact which specifies a very particular and individual personal preference. Different investors might employ very different utility functions to express their individual preference.

Some concepts of stochastic orderings robustify decisions by replacing a single utility function by a class of functions, so that (8.1) holds for all of them.

9.1 STOCHASTIC DOMINANCE OF FIRST ORDER

Definition 9.1. For $\mathbb{R}$-valued random variables X and Y we say that X is dominated by Y in *first order stochastic dominance*,

$$X \preccurlyeq_{(1)} Y \text{ or } X \preccurlyeq_{FSD} Y,$$

if (8.1) holds for all $u \in \mathcal{U}_{FSD} := \{u\colon \mathbb{R} \to \mathbb{R} \text{ nondecreasing}\}$ for which the integrals exists.

Remark 9.2. Not that the function $u(x) := x$ in nondecreasing, so that $\mathbb{E}\, X \le \mathbb{E}\, Y$ whenever $X \preccurlyeq_{(1)} Y$.

Remark 9.3. The relation

$$X \le Y \qquad \text{almost surely}$$

(cf. Definition 5.1 (i)) is occasionally referred to as *stochastic dominance of order* 0 and denoted $X \preccurlyeq_{(0)} Y$.

Theorem 9.4. *The following are equivalent:*

(i) $X \preccurlyeq_{(1)} Y$*,*

(ii) $F_X(\cdot) \ge F_Y(\cdot)$*, i.e.,* $P(X \le z) \ge P(Y \le z)$ *for all* $z \in \mathbb{R}$ *and*

(iii) $F_X^{-1}(\cdot) \le F_Y^{-1}(\cdot)$*, i.e.,* $\mathsf{V@R}_\alpha(X) \le \mathsf{V@R}_\alpha(Y)$ *for all* $\alpha \in (0,1)$*.*

Proof. The function $u_z(x) := \mathbb{1}_{(z,\infty)}(x)$ is nondecreasing and thus $u_z \in \mathcal{U}_{FSD}$. Note that $\mathbb{E}\, u_z(X) = P(X > z)$, thus (8.1) is equivalent to $F_X(z) = P(X \le z) = 1 - P(X > z) = 1 - \mathbb{E}\, u_z(X) \ge 1 - \mathbb{E}\, u_z(Y) = 1 - P(Y > z) = P(Y \le z) = F_Y(z)$, so that (ii) follows from (i).

As for the converse note that every nondecreasing function $u(\cdot)$ may be approximated by a simple step function $u_n(\cdot) = \sum_{i=1}^n \alpha_i\, \mathbb{1}_{(z_i,\infty)}(\cdot)$ with $\alpha_i > 0$ so that $|\mathbb{E}\, u(X) - \mathbb{E}\, u_n(X)| < \varepsilon$ and $|\mathbb{E}\, u(Y) - \mathbb{E}\, u_n(Y)| < \varepsilon$. With (ii) it follows that $\mathbb{E}\, u_n(X) = \sum_{i=1}^n \alpha_i\, P(X > z_i) \le \sum_{i=1}^n \alpha_i\, P(Y > z_i) = \mathbb{E}\, u_n(Y)$, so that stochastic dominance in first order follows.

It is evident that (ii) and (iii) are equivalent. □

Remark 9.5. Exercise 9.5 (Table 9.1a) demonstrates that the order $\preccurlyeq_{(1)}$ is not convex, i.e., the sets $\{Y : X \preccurlyeq_{(1)} Y\}$ and $\{Y : Y \preccurlyeq_{(1)} X\}$ are not convex.

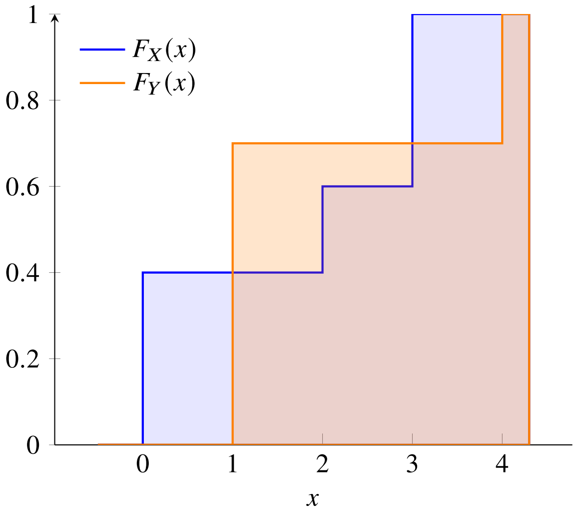
probabilities	40%	20%	40%
$Y_0 = X$	2	4	4
Y_1	4	4	2
$\frac{1}{2}(Y_0 + Y_1)$	3	4	3

(a) The relation $\preccurlyeq_{(1)}$ is not convex, cf. Remark 9.5. Note that $Y_0 \neq Y_1$, but $F_{Y_0} = F_{Y_1}$.

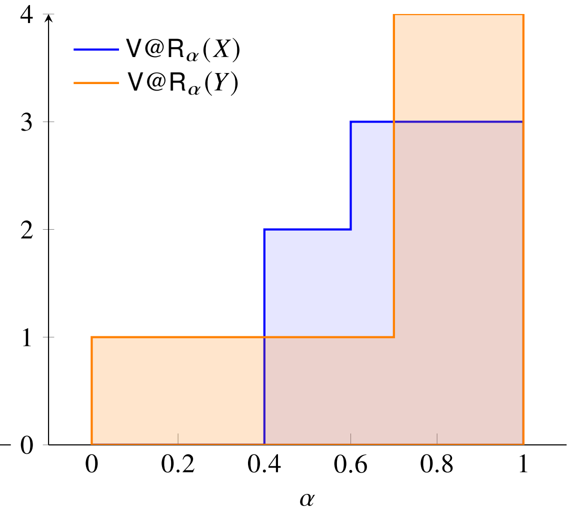
probabilities	40%	20%	10%	30%
X	0	2	3	3
Y	1	1	1	4

(b) $X \not\preccurlyeq_{(1)} Y$, but $X \preccurlyeq_{(2)} Y$, cf. Figure 9.1

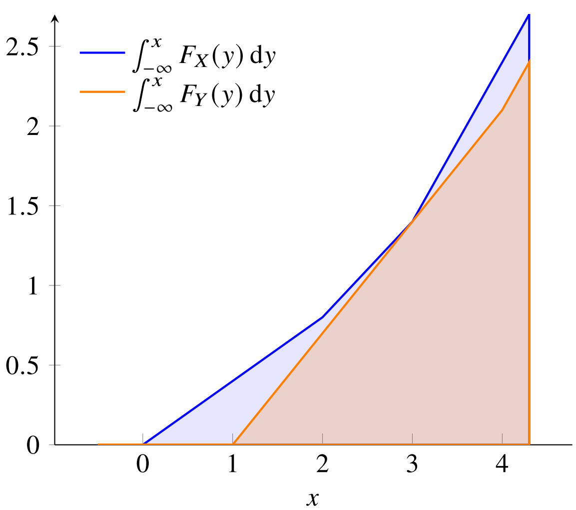
Table 9.1: Counterexamples



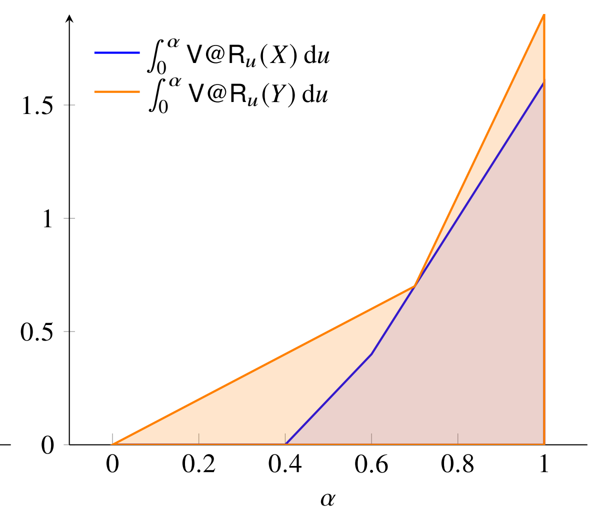
(a) Cumulative distribution function, Theorem 9.4(ii)



(b) Generalized inverse, V@R, Theorem 9.4(iii)



(c) Integrated cdf, Theorem 9.10(ii)



(d) Integrated V@R, Theorem 9.10(iii)

Figure 9.1: Random variables X (blue) and Y (orange) from Table 9.1b

9.2 STOCHASTIC DOMINANCE OF SECOND ORDER

Definition 9.6. For $\mathbb{R}$-valued random variables X and Y we say that X is dominated by Y in second order stochastic dominance,

$$X \preccurlyeq_{(2)} Y \text{ or } X \preccurlyeq_{SSD} Y,$$

if (8.1) holds for all $u \in \mathcal{U}_{SSD} := \{u \colon \mathbb{R} \to \mathbb{R} \text{ nondecreasing and concave}\}$, for which the integrals exists.

Remark 9.7. As in Remark 9.2 we have that that $\mathbb{E}\, X \le \mathbb{E}\, Y$ whenever $X \preccurlyeq_{(2)} Y$.

Remark 9.8. By Jensen's inequality $u(\mathbb{E}\, Y) \ge \mathbb{E}\, u(Y)$. This can be read as possessing (i.e., not investing) the amount $u(\mathbb{E}\, Y)$ is given preference to investing.

Lemma 9.9. *The set $\big\{Y \colon X \preccurlyeq_{(2)} Y\big\}$ is convex (cf. Remark 9.5).*

Proof. Let $X \preccurlyeq_{(2)} Y_0$, $X \preccurlyeq_{(2)} Y_1$ and $u \in \mathcal{U}_{SSD}$ be chosen. Define $Y_\lambda := (1-\lambda)\, Y_0 + \lambda\, Y_1$. By Jensen's inequality it holds that $u\big((1-\lambda)\, Y_0 + \lambda\, Y_1\big) \ge (1-\lambda) u(Y_0) + \lambda\, u(Y_1)$ and thus

$$\begin{aligned}\mathbb{E}\, u(Y_\lambda) = \mathbb{E}\, u\big((1-\lambda) Y_0 + \lambda\, Y_1\big) &\underset{\text{Jensen's inequality}}{\ge} (1-\lambda)\, \mathbb{E}\, u(Y_0) + \lambda\, \mathbb{E}\, u(Y_1)\\ &\ge (1-\lambda)\, \mathbb{E}\, u(X) + \lambda\, \mathbb{E}\, u(X) = \mathbb{E}\, u(X)\end{aligned}$$

by Jensen's inequality and hence $X \preccurlyeq_{(2)} Y_\lambda$. ◻

Theorem 9.10. *The following are equivalent:*

(i) $X \preccurlyeq_{(2)} Y$,

(ii) $\int_{-\infty}^{q} F_X(z)\, \mathrm{d}z \ge \int_{-\infty}^{q} F_Y(z)\, \mathrm{d}z$ *for all $q \in \mathbb{R}$ and*

(iii) $\int_0^\alpha F_X^{-1}(u)\, \mathrm{d}u \le \int_0^\alpha F_Y^{-1}(u)\, \mathrm{d}u$ *(this is what is called the* absolute Lorentz function*) for $\alpha \in (0,1)$,*

(iv) $-\mathrm{AV@R}_\alpha(-X) \le -\mathrm{AV@R}_\alpha(-Y)$ *for all $\alpha \in (0,1)$.*[1]

Proof. By Riemann–Stieltjes integration by parts we have that

$$\begin{aligned}\int_{-\infty}^{q} F_X(x)\, \mathrm{d}x &= x \cdot F_X(x)\big|_{x=-\infty}^{q} - \int_{-\infty}^{q} x\, \mathrm{d}F_X(x) = \int_{-\infty}^{q} q - x\, \mathrm{d}F_X(x) \\ &= \int_{-\infty}^{\infty} (q-x)_+\, \mathrm{d}F_X(x) = -\,\mathbb{E}\, u_q(X),\end{aligned} \tag{9.1}$$

where

$$u_q(x) := -(q-x)_+.$$

The function $u_q(\cdot)$ is nondecreasing and concave, hence it follows from (8.1) that $\mathbb{E}\, u_q(X) \le \mathbb{E}\, u_q(Y)$ and thus the assertion (ii).

A nondecreasing and concave function $u(\cdot) \in \mathcal{U}_{SSD}$ can be approximated by $u_n(x) := \sum_{i=1}^n \alpha_i \cdot u_{q_i}(x)$ where $\alpha_i > 0$ so that $|\mathbb{E}\, u(X) - \mathbb{E}\, u_n(X)| < \varepsilon$ and $|\mathbb{E}\, u(Y) - \mathbb{E}\, u_n(Y)| < \varepsilon$ (see Müller and Stoyan [2002] for details). Assertion (i) then follows by combining (9.1) and (ii).

Define

$$G_X(q) := \int_{-\infty}^{q} F_X(x)\, \mathrm{d}x \text{ and } G_X^{-1}(\alpha) := \int_0^\alpha F_X^{-1}(p)\, \mathrm{d}p.$$

Recall Young's inequality (15.7), i.e.,

$$q\, \alpha \le \underbrace{\int_0^q F_X(x)\, \mathrm{d}x}_{G_X(q)} + \underbrace{\int_0^\alpha F_X^{-1}(p)\, \mathrm{d}p}_{G_X^{-1}(\alpha)}.$$

[1] Recall from Remark 5.2 that $\mathcal{A}(Y) := -\mathrm{AV@R}_\alpha(-Y)$ is an acceptability functional.

From (ii) we deduce that $q\,\alpha - G_X(q) \le q\,\alpha - G_Y(q)$ and thus

$$G_X^{-1}(\alpha) = \sup_{q\in\mathbb{R}} \{q\,\alpha - G_X(q)\} \le \sup_{q\in\mathbb{R}} \{q\,\alpha - G_Y(q)\} = G_Y^{-1}(\alpha).$$

The converse follows by the same reasoning.

As for (iv) recall that $\mathbb{E}\,u_q(X) \le \mathbb{E}\,u_q(Y)$ is equivalent to $\mathbb{E}(q-X)_+ \ge \mathbb{E}(q-Y)_+$, which holds true for every $q \in \mathbb{R}$. Thus $q + \frac{1}{1-\alpha}\,\mathbb{E}(-X-q)_+ \ge q + \frac{1}{1-\alpha}\,\mathbb{E}(-Y-q)_+$, from which assertion (iv) follows after taking the infimum with respect to $q \in \mathbb{R}$.

As for the converse define $q_\alpha^* := -\mathsf{V@R}_\alpha(-Y)$ and recall that $\mathsf{AV@R}_\alpha(-Y) = -q_\alpha^* + \frac{1}{1-\alpha}\,\mathbb{E}(-Y+q_\alpha^*)_+$. It follows that

$$-q_\alpha^* + \frac{1}{1-\alpha}\,\mathbb{E}(-X+q_\alpha^*)_+ \ge \mathsf{AV@R}_\alpha(-X) \ge \mathsf{AV@R}_\alpha(-Y) = -q_\alpha^* + \frac{1}{1-\alpha}\,\mathbb{E}\left(-Y+q_\alpha^*\right)_+,$$

and thus $\mathbb{E}\,u_{q_\alpha^*}(X) \le \mathbb{E}\,u_{q_\alpha^*}(Y)$. The assertion follows now, as q_α^* can be adjusted for every $\alpha \in (0,1)$. □

Remark 9.11. It is evident that $X \preccurlyeq_{(1)} Y$ implies $X \preccurlyeq_{(2)} Y$, but the converse is not true: cf. Exercise 9.6 (Table 9.1b).

9.3 PORTFOLIO OPTIMIZATION

Let X be a random variable understood as a benchmark (cf. (2.3) in the introduction). By Lemma 9.9, the problem

$$\begin{array}{rl} \text{maximize} & \\ \text{in } x & \mathbb{E}\,x^\top\xi \\ \text{subject to} & X \preccurlyeq_{(2)} x^\top\xi \\ & x^\top \mathbb{1} \le 1€, \\ & (x \ge 0) \end{array} \tag{9.2}$$

is a convex optimization problem.

The relation $X \preccurlyeq_{SSD} x^\top\xi$ in (9.2) can be restated as

$$\begin{array}{rll} \text{maximize} & & \\ \text{in } x & \mathbb{E}\,x^\top\xi & \\ \text{subject to} & -\mathsf{AV@R}_\alpha(-X) \le -\mathsf{AV@R}_\alpha(-x^\top\xi) & \text{for all } \alpha \in (0,1), \\ & x^\top \mathbb{1} \le 1€, & \\ & (x \ge 0) & \end{array} \tag{9.3}$$

so that the problem has infinity many (uncountably many) constraints. We refer to Dentcheva and Ruszczyński [2011] for a discussion and numerical implementation schemes.

9.4 PROBLEMS

Exercise 9.1. *Show that*

$$(1-\alpha)\,\mathsf{AV@R}_\alpha(Y) - \alpha\,\mathsf{AV@R}_{1-\alpha}(-Y) = \mathbb{E}\,Y.$$

Exercise 9.2. *Compute the preferences* $\mathbb{E}\,u_\gamma(X) \le \mathbb{E}\,u_\gamma(Y)$ *for* $u_\gamma(x) := \frac{x^\gamma}{\gamma}$ *with* $\gamma = 0.5$ *and* $\gamma = 0.6$ *for the random variable specified in Table 4.2.*

Exercise 9.3. *Compute the preferences* $\mathbb{E}\,u_\lambda(X)$ *for* $u_\lambda(x) := 1 - e^{-\lambda x}$ *with selections of* λ *to get different preferences (again Table 4.2).*

Exercise 9.4. *Show by using Theorem 9.4 that we have* $X \not\preccurlyeq_{(1)} Y$ *and* $X \not\preccurlyeq_{(2)} Y$ *for the random variable in Table 4.2.*

probabilities	$30\,\%$	$70\,\%$
return Y_1	$10\,\%$	$25\,\%$
return Y_2	$25\,\%$	$10\,\%$

Table 9.2: Return of different portfolios Y_1 and Y_2

Exercise 9.5. *Consider the random variables in Table 9.1a and verify that the set* $\{Y : X \preccurlyeq_{(1)} Y\}$ *is not convex.*

Exercise 9.6. *Verify for the random variables in Table 9.1b that* $X \not\preccurlyeq_{(1)} Y$*, but* $X \preccurlyeq_{(2)} Y$*.*

Exercise 9.7. *Which portfolio is preferable in Table 9.2 if employing the utility function* $u(x) = x^\kappa$ *for* $\kappa \in (0,1)$*?*

10 Arbitrage

> On ne peut vivre de frigidaires, de politique, de bilans et de mots croisés, voyez-vous! On ne peut plus vivre sans poésie, couleur ni amour.
>
> Antoine de Saint-Exupéry,
> *Lettre au général X*, 30 juillet 1944

This section follows Cornuejols and Tütüncü [2006, Chapter 4].

Definition 10.1. Arbitrage is a trading strategy,

Type A: that has a positive cash flow and no risk of a later loss;

Type B: that requires no initial cash input, has no risk of a loss, and a positive probability of making profits in the future: $V_0 = 0$, $P(V_t \ge 0) = 1$ and $P(V_t > 0) > 0$, where V_t is the portfolio value at time t.

10.1 TYPE A

Consider the exchange rates in Table 10.1. Note, that converting any (!) currency forwards and backwards will result in a loss; for example

$$1 \text{ EUR} = 1.12 \text{ US\$} = 1.12 * 0.892 \text{ EUR} = 0.99904 \text{ EUR} < 1 \text{ EUR},$$

etc.

Exchanging a sequence of currencies is a loss as well (in general), e.g.,

$$\begin{aligned} 1 \text{ EUR} &= 121 \text{ JPY} \\ &= 121 * 0.00701 \text{ GBP} \\ &= 121 * 0.00701 * 1.286 \text{ US\$} \\ &= 121 * 0.00701 * 1.286 * 0.892 \text{ EUR} = 0.97 \text{ EUR} < 1 \text{ EUR}. \end{aligned} \tag{10.1}$$

However, the table allows for arbitrage (a free lunch of 1.76%), for example by converting

$$\begin{aligned} 1 \text{ EUR} &= 1.12 \text{ US\$} \\ &= 1.12 * 0.777 \text{ GBP} \\ &= 1.12 * 0.777 * 142.6 \text{ JPY} \\ &= 1.12 * 0.777 * 142.6 * 0.0082 \text{ EUR} = 1.0176 \text{ EUR} > 1 \text{ EUR}. \end{aligned} \tag{10.2}$$

	to:	EUR	US$	GBP	JPY
1 EUR	=		1.12	0.87	121.0
1 US$	=	0.892		0.777	110.7
1 GBP	=	1.149	1.286		142.6
1 JPY	=	0.0082	0.00900	0.00701	

Table 10.1: Exchange rates

Investing $1€$ thus will result in a profit of EUR 0.0176 without risk. Note that (10.2) is actually the reverse order of (10.1).

How can one detect an opportunity for arbitrage?

Define the variables

ED, etc.: quantity of EUR (i.e., number of EUR banknotes) changed to US$, etc.

A: quantity of EUR generated by arbitrage.

Then we may consider the optimization problem

$$\begin{aligned}
\text{maximize}\quad & A \\
\text{subject to}\quad & 0.892DE - ED + 1.149PE - EP + 0.0082YE - EY \geq A, && (10.3)\\
& 1.12ED - DE + 1.286PD - DP + 0.009YD - DY \geq 0, \\
& 0.87EP - PE + 0.777DP - PD + 0.00701YP - PY \geq 0, \\
& 121EY - YE + 110.7DY - YD + 142.6PY - YP \geq 0, && (10.4)\\
& ED + EY + EP \leq 1000 \text{ EUR}, && (10.5)\\
& ED \geq 0,\ DE \geq 0,\ \text{etc. and } A \geq 0.
\end{aligned}$$

The constraints (10.3)–(10.4) are balance equations (conservation equations[1]) for EUR, USD, GBP and JPY (resp.), while the budget constraint (10.5) limits the total, initial amount available. $(0, \dots, 0)$ is always a feasible solution with profit $A = 0$ (convert nothing, no arbitrage). We have found an arbitrage opportunity, if our solver returns a solution with objective $A > 0$.

Indeed, this is possible here as $ED = 1$, $DP = 1.12$, $PY = 1.12 * 0.777 = 0.87024$, $YE = 1.12 * 0.777 * 142.6 = 124.09$ (all other are $EP = EY = DE = \cdots = 0$) and $A = 0.0176 > 0$ (cf. (10.2)) is feasible and indeed the optimal solution (up to scaling with 1000, cf. (10.5)).

10.2 TYPE B

Consider as series of options $i = 1, \dots n$ with payoff $\Psi_i(\cdot)$, written on one single underlying with random terminal price S. By investing the amount of x_i in each option (we allow short-selling, i.e., $x_i \leq 0$ or $x_i \geq 0$), we obtain the random payoff

$$\Psi_x(S) = \sum_{i=1}^{n} x_i \cdot \Psi_i(S).$$

For call and put options with strike K_i, the function $\Psi_x(\cdot)$ is piecewise linear with kinks at the strikes K_i and thus everywhere nonnegative (thus generating arbitrage) in the range of the underlying $S \in [0, \infty)$, if

$$\Psi_x(0) \geq 0, \quad \Psi_x(K_j) \geq 0 \text{ for all } j = 1, \dots, n \text{ and } \Psi_x'(K^{max}) \geq 0,$$

where $K^{max} := \max_{j=1,\dots n} K_j$ is the largest of all strikes.

[1] Erhaltungsgleichungen, in German

Assume the price of option i is p_i and solve the linear problem

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^n} \quad & \sum_{i=1}^n x_i p_i \\ \text{subject to} \quad & \sum_{i=1}^n x_i \cdot \Psi_i(0) \geq 0, \\ & \sum_{i=1}^n x_i \cdot \Psi_i(K_j) \geq 0 \text{ for } j = 1, \dots, n \text{ and} \\ & \sum_{i=1}^n x_i \cdot \Big(\Psi_i\left(K^{max}+1\right) - \Psi_i\left(K^{max}\right)\Big) \geq 0. \end{aligned} \tag{10.6}$$

Then there is type B arbitrage, if the objective (10.6) ≤ 0 or unbounded; no arbitrage is possible, if (10.6) > 0.

Example 10.2. How should one modify problem (10.6) to incorporate interest, for example because the options are exercised in 1 year, e.g.?

11 The Flowergirl Problem[1]

Sell in May and go away.

investment strategy

11.1 THE FLOWERGIRL PROBLEM

Example 11.1 (The flowergirl problem, cf. [Pflug and Pichler, 2014])**.** A flowergirl has to decide how many flowers she orders from the wholesaler.

- She
  - buys for the price b per flower and
  - sells them for a price $s > b$.
  - The random demand is ξ.
  - If the demand is higher than the available stock, she may procure additional flowers for an extra price $e > b$.
  - Unsold flowers may be returned for a price of $r < b$
- What is the optimal order quantity x^*, if the expected profit should be maximized?

We formulate the profit as negative costs (expenses minus revenues) are

$$\begin{aligned} \text{total costs} \coloneqq\ & \text{initial purchase} && = b\,x \\ & - \ \text{revenue from sales} && - s\,\xi \\ & + \ \text{extra procurement costs} && + e\,[\xi - x]_+ \\ & - \ \text{revenue from returns.} && - r\,[\xi - x]_-; \end{aligned}$$

here, $[a]_+ = \max\{a,\, 0\}$ is the positive part of a and $[a]_- = \max\{-a, 0\}$ is the negative part of a. Since $a = [a]_+ - [a]_-$, the cost function may be rewritten as

$$\begin{aligned} Q(x,\xi) &= (b-r)\,x - (s-r)\,\xi + (e-r)\,[\xi - x]_+ \\ &= (b-r)\left\{x + \frac{1}{1-\frac{e-b}{e-r}}\,[\xi - x]_+\right\} - (s-r)\,\xi. \end{aligned}$$

Since $\mathsf{AV@R}_\alpha(Y) = \min\left\{x + \frac{1}{1-\alpha}\,\mathbb{E}[Y-x]_+ \colon x \in \mathbb{R}\right\}$ (see Average Value-at-Risk below) it follows that

$$\min_{x\in\mathbb{R}} \mathbb{E}\,Q(x,\xi) = (b-r)\ \mathsf{AV@R}_\alpha(\xi) - (s-r)\ \mathbb{E}\,\xi,$$

where $\alpha \coloneqq \frac{e-b}{e-r}$.

To determine the order quantity consider the function

$$x \mapsto (b-r)\left\{x + \frac{1}{1-\frac{e-b}{e-r}}\,\mathbb{E}[\xi - x]_+\right\} - (s-r)\,\mathbb{E}\,\xi \tag{11.1}$$

[1] Also: Newsboy, or Newsvendor problem

and its derivative

$$0 = (b-r)\left\{1 - \frac{1}{1-\alpha}\,\mathbb{E}\,\mathbb{1}_{\xi > x}\right\} = (b-r)\left\{1 - \frac{1}{1-\alpha} P\left(\xi > x\right)\right\}.$$

Note next that this is equivalent to $\alpha = P(\xi \le x)$. The optimal procurement quantity of the flowergirl thus has the explicit expression

$$x^* = \mathsf{V@R}_\alpha(\xi) = \mathsf{V@R}_{\frac{e-b}{e-r}}(\xi) = F_\xi^{-1}\left(\frac{e-b}{e-r}\right)$$

(see, e.g., Pflug and Römisch [2007, page 56]).

11.2 PROBLEMS

Exercise 11.1. *Show that (11.1) is convex.*

12 Duality For Convex Risk Measures

Risk measures, as introduced in Section 5, are convex. They hence have a representation $\mathcal{R}(Y) = \sup\{x^*(z) - \mathcal{R}^*(z^*) : z^* \in X^*\}$. To this end we specify the domain, its dual and the inner product $x^*(z)$.

Typical candidates for the domain of risk measures are L^p spaces, particularly L^∞. Recall that the duals are L^q. Note further that the canonical inner product for the $L^p - L^q$-duality is

$$L^p \times L^q \ni (Y, Z) \mapsto \mathbb{E}\, YZ.$$

Proposition 12.1. *Suppose that the risk measure $\mathcal{R}\colon L^p \to \mathbb{R} \cup \{\infty\}$ is positively homogeneous. Then $\mathcal{R}^*(Z) \in \{0, \infty\}$.*

Proof. Note that

$$\begin{aligned}\mathcal{R}^*(Z) &= \sup_{Y \in L^p} \mathbb{E}\, YZ - \mathcal{R}(Y)\\ &= \sup_{\lambda \in \mathbb{R}} \lambda \cdot \sup_{Y \in L^p} \left(\mathbb{E}\, YZ - \mathcal{R}(Y)\right) \in \{0, \infty\}.\end{aligned}$$

□

Proposition 12.2. *Suppose that the risk measure $\mathcal{R}\colon L^p \to \mathbb{R} \cup \{\infty\}$ is translation equivariant. Then $\mathcal{R}^*(Z) = \infty$ unless $\mathbb{E}\, Z = 1$.*

Proof.

$$\begin{aligned}\mathcal{R}^*(Z) &= \sup_{Y \in L^p} \mathbb{E}\, YZ - \mathcal{R}(Y)\\ &= \sup_{Y \in L^p} \sup_{c \in \mathbb{R}} \left(\mathbb{E}(Y + c)Z - \mathcal{R}(Y + c)\right) = \sup_{Y \in L^p} \mathbb{E}(Y)Z - \mathcal{R}(Y) + \sup_{c \in \mathbb{R}} c\left(\mathbb{E}\, Z - 1\right),\end{aligned}$$

from which the assertion follows. □

Proposition 12.3. *Suppose that the risk measure $\mathcal{R}\colon L^p \to \mathbb{R} \cup \{\infty\}$ is monotone. Then $\mathcal{R}^*(Z) = \infty$ unless $Z \geq 0$ almost surely.*

Proof. Suppose that $P(Z \leq 0) > 0$. Set $A := \{Z \leq 0\}$ and $Y_0 := \mathbb{1}_A$. Note, that $-Y_0 \leq 0$, hence $\mathcal{R}(-Y_0) \leq \mathcal{R}(0)$ and

$$\begin{aligned}\mathcal{R}^*(Z) &= \sup_{Y \in L^p} \mathbb{E}\, YZ - \mathcal{R}(Y)\\ &\geq \sup_{\lambda < 0} \left(\mathbb{E}\, \lambda Y_0 Z - \mathcal{R}(\lambda Y_0)\right) \geq \sup_{\lambda < 0} \mathbb{E}\, \lambda\, \mathbb{1}_A\, Z - \mathcal{R}(0) = \infty.\end{aligned}$$

Hence, $Z \geq 0$ a.s. □

Definition 12.4. The support function of a set $\mathcal{Z}$ is $s_{\mathcal{Z}}(Y) := \sup_{Z \in \mathcal{Z}} \mathbb{E}\, YZ$.

Theorem 12.5. *Define $\mathcal{Z} := \{Z : \mathcal{R}^*(Z) < \infty\}$.*

13 Stochastic Optimization: Terms, and Definitions, and the Deterministic Equivalent

Kein Geld ist vorteilhafter angewandt als das, um welches wir uns haben prellen lassen; denn wir haben dafür unmittelbar Klugheit eingehandelt.

Arthur Schopenhauer, 1788–1860

13.1 Expected Value of Perfect Information (EVPI) and Value of Stochastic Solution (VSS)

The *expected value of perfect information (EVPI)* is the price that one would be willing to pay in order to gain access to perfect information, that is to say the difference SP – wait-and-see,

$$\underbrace{\mathbb{E}\left[\min_x f(x,\xi)\right]}_{\text{wait-and-see}} \le \underbrace{\min_x \mathbb{E}\, f(x,\xi)}_{\text{SP}} \le \mathbb{E}\, f(x,\xi). \tag{13.1}$$

Both inequalities hold always.

For $f(x,\cdot)$ *concave*, Jensen's inequality continues the sequence of inequalities above with

$$\mathbb{E}\, f(x,\xi) \le f\left(x, \mathbb{E}\,\xi\right),$$

drawing some attention to the strategy $x_0 \in \arg\min f\left(x, \mathbb{E}\,\xi\right)$ (if this exists at all): Given this reference strategy x_0 the distance RHS – SP in (13.1) is called *Value of the Stochastic Solution* (VSS). However, in a general context $f(x,\cdot)$ are rather convex and no comparison of $f\left(x_0, \mathbb{E}\,\xi\right)$ with (SP) is possible in (13.1) for this case (counter-example: Farmer Ted).

13.2 The Farmer Ted

See Jeff's lecture, http://homepages.cae.wisc.edu/~linderot/classes/ie495/lecture2.pdf.

13.3 The Risk-Neutral Problem

Two-stage stochastic linear program with fixed recourse:

$$\begin{array}{lllr}
\text{minimize} & c^\top x + \mathbb{E}_\xi\left[q_\xi^\top y_\xi\right] = c(x) + \mathbb{E}\, Q(x,\cdot) & & \\
\text{(in } x \text{ and } y\text{)} & & & \\
\text{subject to} & Ax = b & & 1^{\text{st}} \text{ stage constraints} \\
& T_\xi x + W y_\xi = h_\xi \quad \text{for a.e. } \xi \in \Xi & & 2^{\text{nd}} \text{ stage constraints} \\
& x \in X,\ y_\xi \in Y \quad \text{for a.e. } \xi \in \Xi & &
\end{array} \tag{13.2}$$

Note, that minimization is done over a deterministic x and a *random variable* y.

13.4 GLOSSARY/ CONCEPT/ DEFINITIONS:

- $(x,\ y_\xi) \mapsto c^\top x + \mathbb{E}_\xi \left[q^\top y_\xi\right]$ is the objective function;
  - x is called *here-and-now decision* (solution), 1^{st} stage decision;
  - the (optimal) random variable y_ξ is called *wait-and-see decision* (solution), 2^{nd} stage decision or recourse action;
- c (deterministic) costs;
- q: vector of recourse costs, which is sometimes considered random as well;
- W is the *recourse matrix*. *Fixed recourse* is given, if – as in (13.2) – $W = W(\xi)$ (i.e., the matrix is deterministic/ *non*random);
- T_ξ are sometimes called *technology matrices;*
- Y: feasible set of recourse actions;
- the function
$$v_q(z) := \begin{cases} \min_{y\in Y}\left\{q^\top y\colon Wy = z\right\} & \text{if feasible} \\ +\infty & \text{else}\end{cases}$$
is called s*econd stage value function* or *recourse (penalty) function*;
- then define
$$Q(x,\xi) := v\left(h_\xi - T_\xi x\right) = \min_{y\in Y}\left\{q^\top y\colon Wy = h_\xi - T_\xi x\right\},$$
(notice: $x \mapsto Q(x,\xi)$ is lsc.)
- and
$$\mathcal{Q}(x) := \mathbb{E}_\xi\left[Q(x,\xi)\right] = \mathbb{E}_\xi\left[v\left(h_\xi - T_\xi x\right)\right]$$
is called *expected value function*, or *expected minimum recourse function*.
- A *recourse is relatively complete* if $Ax = b$, $x \ge 0$ implies $Q(x,\xi) < \infty$ for a.e. $\xi \in \Xi$.
- A *recourse is complete* if $\forall z\colon v(z) < \infty$ (i.e., there always exists a feasible recourse action, $\forall z\, \exists y\colon Wy = z$). As a consequence, $Q(x,\xi) < \infty$.

13.5 KKT FOR (13.2)

- v is a LP itself and consequently, from duality,
$$\begin{aligned} v(z) &= \min_{y\ge 0}\left\{q^\top y\colon Wy = z\right\} \\ &= \max_\lambda\left\{\lambda^\top z\colon \lambda^\top W \le q^\top\right\} \end{aligned} \tag{13.3}$$
where additionally $\lambda^{*\top} \in \partial v(z)$ (λ^* being the optimal ($\arg\max$) solution of the dual problem).
- From the chain rule, $\partial_x Q(x,\xi) = \partial_x v\left(h_\xi - T_\xi x\right) \ni -\lambda_\xi^{*\top} T_\xi$.
- Suppose further the probability space is discrete, that is $\mathbb{P} = \sum_{\xi\in\Xi} p_\xi \cdot \delta_\xi$ ($p_\xi := \mathbb{P}[\{\xi\}]$), then
$$\mathcal{Q}(x) = \mathbb{E}_\xi\left[Q(x,\xi)\right] = \sum_{\xi\in\Xi} p_\xi Q(x,\xi). \tag{13.4}$$
For $u_\xi^\top := -\lambda_\xi^{*\top} T_\xi \in \partial_x Q(x,\xi)$ thus $u^\top := \mathbb{E}_\xi\left[u_\xi^\top\right] = \sum_{\xi\in\Xi} p_\xi u_\xi^\top \in \partial\mathcal{Q}(x)$.

KKT, applied to the problem (13.2):

x^* is an optimal solution of (13.2) iff $\exists \lambda^*, \mu^* \geq 0$ st.

(i) $0 \in c^\top + \partial Q(x^*) + \lambda^{*\top} A - \mu^{*\top}$,

(ii) $\mu^{*\top} x^* = 0$.

13.6 DETERMINISTIC EQUIVALENT

Given the situation, that Ξ consists of finitely many ($S := |\Xi|$) atoms, Equation (13.2) can be reformulated in its *deterministic equivalent*, i.e.

$$
\begin{array}{lcccccccc}
\text{minimize} & c^\top x + & p_{\xi_1} q^\top y_{\xi_1} + & p_{\xi_2} q^\top y_{\xi_2} + & \dots & + p_{\xi_S} q^\top y_{\xi_S} & & \\
\text{(in x and y)} & & & & & & & \\
\text{subject to} & Ax & & & & & = & b \\
& T_{\xi_1} x & + W y_{\xi_1} & +0 & \dots & +0 & = & h_{\xi_1} \\
& T_{\xi_2} x & +0 & + W y_{\xi_2} & \ddots & \vdots & = & h_{\xi_2} \\
& \vdots & \vdots & \ddots & \ddots & +0 & \vdots & \vdots \\
& T_{\xi_S} x & +0 & \dots & +0 & + W y_S & = & h_{\xi_S} \\
& x \in X & y_{\xi_1} \in Y & y_{\xi_2} \in Y & & y_{\xi_S} \in Y & &
\end{array}
\tag{13.5}
$$

where $p_\xi := \mathbb{P}[\{\xi\}]$.

NB: (13.5) is a big LP (often too big, indeed), but *linear* and *sparse*. The size increases, as the number of atoms (scenarios) S increases. However, we expect that a lot of these constraints are redundant and we want to exploit this presumption.

13.7 L-SHAPED METHOD

i.e., Bender's decomposition, applied to (13.5).

Let $\lambda^*_\xi(\hat{x})^\top \in \arg\max_\lambda \left\{\lambda^\top \left(h_\xi - T_\xi \hat{x}\right) : \lambda^\top W \leq q\right\}$ be an optimal, dual solution to the recourse problem in scenario ξ, then $u(\hat{x})^\top := -\sum_\xi p_\xi \lambda^{*\top}_\xi(\hat{x}) T_\xi \in \partial Q(\hat{x})$. Thus,

$$Q(\hat{x}) + u(\hat{x})^\top (x - \hat{x}) \leq Q(x),$$

hence $x \mapsto Q(\hat{x}) + u(\hat{x})^\top (x - \hat{x})$ is a supporting hyperplane, supporting Q from below. So is the bundle,

$$Q_L(x) := \max_{l \in L} Q(x_l) + u_l^\top (x - x_l) \leq Q(x)$$

(where $u_l := u(x_l)$).

13.8 FARKAS' LEMMA

Lemma 13.1 (A Theorem on the Alternative)**.** *Exactly* one *of these following two statements holds true:*

- There exists y such that $Wy = z$ and $y \geq 0$;
- There exists σ such that $\sigma^\top W \leq 0$ and $\sigma^\top z > 0$.

13.9 L-SHAPED ALGORITHM.

The initial problem (13.2) can be restated equivalently, getting rid of the random variable in the objective function at the same time, as

$$\begin{aligned}
(13.2) \quad &\iff \begin{cases} \text{minimize (in x)} & c^\top x + \mathcal{Q}(x) \\ \text{subject to} & Ax = b, \\ & x \in X \end{cases} \\
&\iff \min_{x\in X} \left\{ c^\top x + \mathcal{Q}(x) : Ax = b \right\} \\
&\iff \begin{cases} \text{minimize (in x, } \theta) & c^\top x + \theta \\ \text{subject to} & Ax = b, \\ & \mathcal{Q}(x) \le \theta, \\ & x \in X \end{cases}
\end{aligned}$$

Algorithm 13.1 is based on the previous observation of supporting hyperplanes and the latter, equivalent formulation:

13.10 VARIANTS OF THE ALGORITHM.

- Multicut: In view of linearity of (13.4) and (13.7) we may equally well start with the set

  $$\mathcal{B} := \{(x, \theta_1, \dots \theta_S) : x \ge 0,\ Ax = b,\ \theta_i \ge \theta_0\}$$

  and solve the problem

  $$\min \left\{ c^\top x + \sum_\xi p_\xi \theta_\xi : (x, \theta_1, \dots, \theta_S) \in \mathcal{B} \right\}$$

  instead of (13.6) – the problem is said to be *separable* into scenario sub-problems.

  – The abort criterion reads: $\mathcal{Q}(\hat{x}) \le \sum_\xi p_\xi \hat{\theta}_\xi$?
  – The feasibility cut (singular!) remains unchanged, as they do not involve θs.
  – The new optimality cuts (plural!) read

  $$\mathcal{B} \leftarrow \mathcal{B} \cap \bigcap_{\xi :\ \mathcal{Q}(\hat{x}, \xi) > \hat{\theta}_\xi} \left\{ \left(x,\ \theta_1, \dots \theta_\xi \dots, \theta_S\right) : \theta_\xi \ge \mathcal{Q}(\hat{x}, \xi) + \hat{u}_\xi^\top (x - \hat{x}) \right\}.$$

- Chunked multicut: The same idea as multicut, but with a few ξ-clusters instead of the entire Ξ: $\Xi = \{\xi_1, \dots \xi_S\} = \dot{\bigcup}_{k=1}^C S_k$. Define $\mathcal{Q}_{[S_k]}(x) := \sum_{\xi \in S_k} p_\xi \mathcal{Q}(x, \xi)$ and proceed as for the multicut version.

- Numerical experiments show that the algorithm is sometimes flipping around without improving the solution significantly. In order to stop this misbehavior search for an improved *local* solution, by modifying the objective function as follows:

  – Regularization

  $$\min \left\{ c^\top x + \sum_k p_{S_k} \theta_k : (x, \theta_1, \dots, \theta_C) \in \mathcal{B},\ \left\|x - x^i\right\| \le \Delta_i \right\}$$

  – Regularized decomposition method

  $$\min \left\{ c^\top x + \sum_k p_{S_k} \theta_k + \frac{\left\|x - x^i\right\|^2}{2\rho} : (x, \theta_1, \dots, \theta_C) \in \mathcal{B} \right\}$$

Algorithm 13.1 L-Shaped Method

(i) Find θ_0 such that $\theta_0 \le Q(x)$ (for all x) and define

$$\mathcal{B} := \{(x,\theta) : x \ge 0,\ Ax = b,\ \theta \ge \theta_0\}.$$

($\mathcal{B}$ – to some extent – characterizes the epi-graph of the approximate Q_L and we have $\mathcal{B} \supseteq \mathrm{epi} Q$; if not available then choose $\theta_0 := -\infty$.)

(ii) Solve the problem

$$\min\left\{c^\top x + \theta \colon (x,\theta) \in \mathcal{B}\right\} \tag{13.6}$$

and call the solution found $(\hat{x},\ \hat{\theta})$.

(iii) Compute $Q(\hat{x})$.

(a) If $Q(\hat{x}) \le \hat{\theta} < \infty$, then $\hat{x}$ is the best (i.e., minimal) solution of our original problem.

(b) *Optimality cut*: if $\hat{\theta} < Q(\hat{x}) < \infty$, then put

$$\hat{u}^\top := -\sum_{\xi\in\Xi} p_\xi \lambda_\xi^*(\hat{x})^\top T_\xi \in \partial Q(\hat{x}) \tag{13.7}$$

and send the additional hyperplane, that is

$$\mathcal{B} \leftarrow \mathcal{B} \cap \left\{(x,\ \theta) : \theta \ge Q(\hat{x}) + \hat{u}^\top (x - \hat{x})\right\}.$$

Continue with step (ii).

(c) *Feasibility cut*: It holds that $Q(\hat{x}) = \infty$. In this case there exists $\hat{\xi}$, such that $v\left(h_{\hat{\xi}} - T_{\hat{\xi}}\hat{x}\right) = \infty$, i.e., $\left\{q^\top y \colon Wy = h_{\hat{\xi}} - T_{\hat{\xi}}\hat{x}\right\} = \{\}$, i.e. $\left\{y \colon Wy = h_{\hat{\xi}} - T_{\hat{\xi}}\hat{x}\right\} = \{\}$. Hence, $h_{\hat{\xi}} - T_{\hat{\xi}}\hat{x}$ is *not feasible* for v and thus $\hat{x}$ is not feasible for Q. Thus, by Farkas' lemma (Lemma 13.1), find an appropriate $\hat{\sigma}$, such that $\hat{\sigma}^\top W \le 0$ and $\hat{\sigma}^\top\left(h_{\hat{\xi}} - T_{\hat{\xi}}\hat{x}\right) > 0$. To exclude this particular $\hat{x}$ for the future send the additional condition

$$\mathcal{B} \leftarrow \mathcal{B} \cap \left\{(x,\ \theta) : \hat{\sigma}^\top\left(h_{\hat{\xi}} - T_{\hat{\xi}}x\right) \le 0\right\}$$

and continue with step (ii).

Additional controls on Δ and ρ are available here to (intuitively) enforce a direction of (significantly) good descent. The same techniques as for (unconstrained), nonlinear optimization apply here.

- ▹ Parallelizing: $u_\xi^\top := -\lambda_\xi^{*\top} T_\xi \in \partial Q(x, \xi)$ has to be evaluated, which is an LP for all $\xi \in \Xi$ (or chunks). This work can be parallelized, reducing (optimistically) the total time by the factor $\frac{1}{\text{\# Computers}}$.
- ▹ Computing $\lambda_\xi^{*\top} \in \partial v\left(h_\xi - T_\xi \hat{x}\right)$ is almost the same LP for all $\xi \in \Xi$, but without involving S as a dimension: the constrains of the dual function stay unchanged, only the objective function varies (if q is nonrandom). The idea of *bunching* is to avoid all those evaluations and instead build a basis of representative directions such that $\lambda^{*\top} = q_B^\top W_B^{-1}$ (cf. (13.3)). This is particularly advantageous if $\dim \Xi \gg \dim q$. The statement is based on the following
  - Theorem: Let x be optimal for (LP'),
    - then there is a basis such that $q_N^\top \geq q_B^\top W_B^{-1} W_N$ for the appropriate decomposition $W = (W_B, W_N)$ etc.;
    - moreover, $\lambda^{*\top} := q_B^\top W_B^{-1}$ is optimal for (DP').

14 Co- and Antimonotonicity

> Ich habe elende Millionäre und glückliche Tagelöhner gesehen.
>
> Johann Nestroy, 1801–1862

14.1 REARRANGEMENTS

Theorem 14.1 (Generalized Chebyshev's sum inequality)**.** *Let $p_i \geq 0$ with $\sum_{i=1}^n p_i = 1$. Then, for $x_1 \leq x_2 \cdots \leq x_n$ and $y_1 \leq y_2 \cdots \leq y_n$ it holds that*

$$\left(\sum_{i=1}^n p_i x_i\right) \cdot \left(\sum_{i=1}^n p_i y_i\right) \leq \sum_{i=1}^n p_i x_i y_i.$$

Proof. Note that

$$\sum_{j=1}^n \sum_{k=1}^n p_j p_k \underbrace{(x_j - x_k)(y_j - y_k)}_{\geq 0} \geq 0,$$

as the components are increasing. Hence, by expanding,

$$0 \leq \sum_{j=1}^n \sum_{k=1}^n p_j p_k x_j y_j - p_j p_k x_j y_k - p_j p_k x_k y_j + p_j p_k x_k y_k = 2 \sum_{j=1}^n p_j x_j y_j - 2 \sum_{j=1}^n p_j x_j \sum_{k=1}^n p_k y_k,$$

from which the result is immediate. □

Theorem 14.2 (Chebyshev's sum inequality, the continuous version)**.** *Let $f, g\colon [0,1] \to \mathbb{R}$ be nondecreasing. Then it holds that*

$$\int_0^1 f(x)\,\mathrm{d}x \cdot \int_0^1 g(x)\,\mathrm{d}x \leq \int_0^1 f(x)g(x)\,\mathrm{d}x.$$

Theorem 14.3 (The rearrangement inequality)**.** *Let $x_1 \leq x_2 \cdots \leq x_n$ and $y_1 \leq y_2 \cdots \leq y_n$. Then*

$$\sum_{i=1}^n x_{n+1-i} y_i \leq \sum_{i=1}^n x_{\sigma(i)} y_i \leq \sum_{i=1}^n x_i y_i \tag{14.1}$$

for every permutation $\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\}$.

Proof. Suppose the permutation σ maximizing (14.1) were not the identity. Then find the smallest j so that $\sigma(j) \neq j$. Note, that $\sigma(j) > j$ and there is $k > j$ so that $\sigma(j) = k$. Now

$$j < k \implies y_j \leq y_k \text{ and } j < \sigma(j) \implies x_j \leq x_{\sigma(j)}$$

and thus $0 \leq \left(x_{\sigma(j)} - x_j\right)\left(y_k - y_j\right)$, i.e.,

$$x_{\sigma(j)} y_j + x_j y_k \leq x_j y_j + x_{\sigma(j)} y_k. \tag{14.2}$$

Define the permutation exchanging the values $\sigma(j)$ and $\sigma(k)$, i.e., $\tau(i) := \begin{cases} i & \text{for } i \in \{1, \dots, j\} \\ \sigma(j) & \text{if } i = k \\ \sigma(i) & \text{else} \end{cases}$ and observe that the right hand side of (14.2) is better for $\tau(\cdot)$ and $\tau(j) = j$. It follows that $\sigma(\cdot)$ is the identity. □

14.2 COMONOTONICITY

Definition 14.4. The random variables X_i, $i = 1, \dots, n$ are *comonotonic* (aka. *nondecreasing*), if

$$\big(X_i(\omega) - X_i(\tilde{\omega})\big) \cdot \big(X_j(\omega) - X_j(\tilde{\omega})\big) \ge 0 \quad \text{for all } \omega, \tilde{\omega} \in N \text{ and } i, j \le n \tag{14.3}$$

and anti-monotone, if

$$\big(X_i(\omega) - X_i(\tilde{\omega})\big) \cdot \big(X_j(\omega) - X_j(\tilde{\omega})\big) \le 0 \quad \text{for all } \omega, \tilde{\omega} \in N \text{ and } i, j \le n,$$

where $P(N) = 1$.

Theorem 14.5 (Cf. Denneberg [1994]). *Let X and Y be $\mathbb{R}$-valued random variables. The following are equivalent:*

(i) X and Y are comonotonic;

(ii) there exists an $\mathbb{R}$-valued random variable Z and nondecreasing functions $v, w\colon \mathbb{R} \to \mathbb{R}$ so that

$$X = v(Z) \text{ and } Y = w(Z);$$

(iii) there are nondecreasing functions $v, w\colon \mathbb{R} \to \mathbb{R}$ so that

$$X = v(X+Y) \text{ and } Y = w(X+Y).$$

Remark. The subsequent proof verifies that v and w are monotone on the range of Z.

Proof. $(i) \implies (iii)$: Define $Z := X + Y$. For $z = Z(\omega)$ define $v(z) := X(\omega)$ and $w(z) := Y(\omega)$. To see that $v(\cdot)$ and $w(\cdot)$ are well-defined choose $\omega_1, \omega_2 \in Z^{-1}(\{z\})$, then $X(\omega_1) + Y(\omega_1) = Z(\omega_1) = z = Z(\omega_2) = X(\omega_2) + Y(\omega_2)$, and thus

$$X(\omega_1) - X(\omega_2) = -\big(Y(\omega_1) - Y(\omega_2)\big). \tag{14.4}$$

If $X(\omega_1) - X(\omega_2) \le 0$, then $Y(\omega_1) - Y(\omega_2) \ge 0$ by (14.4) and $Y(\omega_1) - Y(\omega_2) \le 0$ by comonotonicity, thus $Y(\omega_1) - Y(\omega_2) = 0$ and consequently $X(\omega_1) = X(\omega_2)$ by (14.4). Hence, $v(\cdot)$ and $w(\cdot)$ are well-defined.

To see that $v(\cdot)$ and $w(\cdot)$ are monotonic pick $\omega_1, \omega_2 \in \Omega$ with $z_1 := Z(\omega_1) \le Z(\omega_2) =: z_2$, then we find

$$X(\omega_1) - X(\omega_2) \le -\big(Y(\omega_1) - Y(\omega_2)\big). \tag{14.5}$$

If $X(\omega_1) > X(\omega_2)$, then $Y(\omega_1) < Y(\omega_2)$ by (14.5), which is in contrast to our assumption on comonotonicity and hence $X(\omega_1) \le X(\omega_2)$; similarly we find that $Y(\omega_1) \le Y(\omega_2)$. It follows that

$$\begin{aligned} v(z_1) &= X(\omega_1) \le X(\omega_2) = v(z_2) \text{ and} \\ w(z_1) &= Y(\omega_1) \le Y(\omega_2) = w(z_2), \end{aligned}$$

i.e., $v(\cdot)$ and $w(\cdot)$ are nondecreasing.

We shall verify next that $v(\cdot)$ and $w(\cdot)$ are Lipschitz with constant 1. Note that we have $v(z) + w(z) = X(\omega) + Y(\omega) = Z(\omega) = z$.

If $z_1 \le z_2$, then

$$z_2 - z_1 = v(z_2) + w(z_2) - v(z_1) - w(z_1) \ge v(z_2) - v(z_1) = |v(z_2) - v(z_1)|$$

by monotonicity of $w(\cdot)$; if $z_1 \geq z_2$, then

$$z_1 - z_2 = v(z_1) + w(z_1) - v(z_2) - w(z_2) \geq v(z_1) - v(z_2) = |v(z_2) - v(z_1)|,$$

i.e., $v(\cdot)$ and $w(z) = z - v(z)$ are both Lipschitz on $Z(\Omega)$.

Finally note that a Lipschitz function $f(\cdot)$ with Lipschitz constant L can be extended to the entire domain by setting $\tilde{f}(z) := \inf_{t \in Z(\Omega)} f(t) + L\,|t - z|$ while keeping the Lipschitz constant L (cf. Exercise 14.1).

(iii) $\implies$ *(ii)* is evident by choosing the random variable $Z := X + Y$.

(ii) $\implies$ *(i)*: Let $\omega_1, \omega_2 \in \Omega$. To prove (i) we may assume that $X(\omega_1) > X(\omega_2)$, i.e., by assumption, $v(Z(\omega_1)) > v(Z(\omega_2))$. As $v(\cdot)$ is monotone we deduce that $Z(\omega_1) \geq Z(\omega_2)$. As $w(\cdot)$ is monotone as well it follows that $Y(\omega_1) = w(Z(\omega_1)) > w(Z(\omega_2)) = Y(\omega_2)$ and hence (i), i.e., X and Y are comonotonic. □

Corollary 14.6. *The random variables X_i are comonotonic iff*

$$(X_1, \dots X_n) \sim \left(F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U)\right)$$

for one *uniform random variable U.*

Proof. It is evident that $\left(F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U)\right)$ are comonotonic, as $F_{X_i}^{-1}(\cdot)$ are nondecreasing and $F_{X_i}^{-1}(U) \sim X_i$.

As for the converse apply Theorem 14.5 (ii) and (4.3). Then $X = F_X^{-1}(U) = u \circ F_Z^{-1}(U)$. □

Proposition 14.7 (Upper Fréchet[1] bound)**.** *If X_i are pairwise comonotonic, then*

$$F_{X_1, \dots X_n}(x_1, \dots, x_n) = \min_{i=1,\dots n} F_{X_i}(x_i).$$

Proof. By Theorem 14.5 (ii) we have

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = P(v(Z) \leq x_1, w(Z) \leq x_2).$$

As $v(\cdot)$ and $w(\cdot)$ are monotone we have either $\{v(Z) \leq x_1\} \subset \{w(Z) \leq x_2\}$ or $\{v(Z) \leq x_1\} \supset \{w(Z) \leq x_2\}$.

If $\{v(Z) \leq x_1\} \subset \{w(Z) \leq x_2\}$, then

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = P(v(Z) \leq x_1) = F_{X_1}(x_1) \leq F_{X_2}(x_2),$$

and if $\{v(Z) \leq x_1\} \supset \{w(Z) \leq x_2\}$, then

$$F_{X_1,X_2}(x_1, x_2) = P(v(Z) \leq x_1, w(Z) \leq x_2) = P(w(Z) \leq x_2) = F_{X_2}(x_2) \leq F_{X_1}(x_1).$$

The assertion follows. □

Remark 14.8. It is *always true* that $F_{X_1, \dots X_n}(x_1, \dots, x_n) \leq \min_{i=1,\dots n} F_{X_i}(x_i)$. For comonotonic random variables, however, the upper Fréchet bound is attained.

Corollary 14.9. *Let Y and Z be comonotonic. Then $\mathbb{E}\, Y \cdot \mathbb{E}\, Z \leq \mathbb{E}\, YZ$.*

Proof. Integrate (14.3) with respect to $P(\mathrm{d}\omega) \otimes P(\mathrm{d}\omega')$ and proceed as in Chebyshevs's sum inequality, Theorem 14.1. □

Corollary 14.10. *The covariance* $\operatorname{cov}(\tilde{X}, \tilde{Y})$ *among all random variables with $\tilde{X} \sim X$ and $\tilde{Y} \sim Y$ is maximal, if $\tilde{X}$ and $\tilde{Y}$ are comonotonic.*

[1] Maurice René Fréchet, 1878–1973

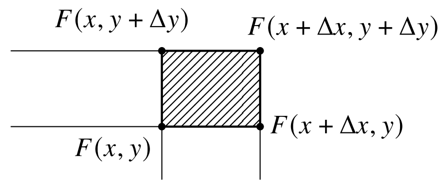


Figure 14.1: Probability of an area

14.3 INTEGRATION OF RANDOM VECTORS

For $\mathbb{R}$-valued random variables we have

$$\mathbb{E}\, g(X) = \int_\Omega g(\omega) P(\mathrm{d}\omega) = \int_{-\infty}^{\infty} g(x)\, \mathrm{d}F_X(x) = \int_{-\infty}^{\infty} g(x) f_X(x)\, \mathrm{d}x,$$

where the latter is only possible if the derivative (density) $\mathrm{d}F_X(x) = f_X(x)\,\mathrm{d}x$ exists.

How do these formulae generalize for higher dimensions?

$$\mathbb{E}\, g(X,Y) = \int_\Omega g\big(X(\omega), Y(\omega)\big) P(\mathrm{d}\omega) = \iint_{\mathbb{R}^2} g(x,y)\, \mathrm{d}^2 F_{X,Y}(x,y) = \iint_{\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)\, \mathrm{d}x\mathrm{d}y. \tag{14.6}$$

To this end observe that

$$\begin{aligned} &P\big(X \in [x, x+\Delta x), Y \in [y, y+\Delta x)\big) \\ &\qquad = F_{X,Y}(x+\Delta x, y+\Delta y) - F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x+\Delta x, y) + F_{X,Y}(x,y) \end{aligned} \tag{14.7}$$

and it is thus evident what $\mathrm{d}^2 F_{X,Y}(x,y)$ in (14.6) has to stand for (cf. Figure 14.1).

Generalizations to random vectors in $\mathbb{R}^n$ are obvious, the general form for (14.7), however, involves 2^n evaluations of $F_{X_1,\dots X_n}(x_1,\dots,x_n)$.

14.4 COPULA

Definition 14.11. The copula function of a random vector $(X_1,\dots,X_n)$ is the cdf $C\colon [0,1]^n \to [0,1]$ on $[0,1]^m$ expressing the joint distribution function by all marginal distributions, i.e.,

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = C\left(F_{X_1}(x_1),\dots F_{X_n}(x_n)\right).$$

Remark 14.12*.* Note, that the copula in dimension 1 is trivial, as $C(u) = u$.

Remark 14.13 (Independence copula)*.* The independence copula

$$C(u_1,\dots u_n) = u_1 \cdot \ldots u_n$$

governs independent random variables $X_1,\dots X_n$.

Lemma 14.14. *Copulas functions can be assumed to be uniformly continuous; more precisely, it holds that*

$$C(u_1,\dots u_n) - C(v_1,\dots v_n) \le |v_1 - u_1| + \dots + |v_n - u_n|.$$

Proof. Just observe that

$$\begin{aligned} P\left(X \le x_2, Y \le y_2\right) - P\left(X \le x_1, Y \le y_1\right) &\le \left|P\left(X \le x_2, Y \le y_2\right) - P\left(X \le x_1, Y \le y_2\right)\right| \\ &\quad + \left|P\left(X \le x_1, Y \le y_2\right) - P\left(X \le x_1, Y \le y_1\right)\right| \\ &\le \left|P\left(X \le x_2\right) - P\left(X \le x_1\right)\right| + \left|P\left(Y \le y_2\right) - P\left(Y \le y_1\right)\right| \end{aligned}$$

and thus

$$C\left(u_1,u_2\right) - C\left(v_1,v_2\right) \le |u_1 - v_1| + |u_2 - v_2|.$$

This generalizes to higher dimensions. ◻

Lemma 14.15. *It holds that*

$$\mathbb{E}\, g(X_1, \dots X_n) = \int_0^1 \cdots \int_0^1 g\left(F_{X_1}^{-1}(u_1), \dots F_{X_1}^{-1}(u_1)\right) \mathrm{d}^n C(u_1, \dots u_n).$$

Proof. By (14.6) and substituting the marginals $x_i \leftarrow F_{X_i}^{-1}(u_i)$ we have

$$\begin{aligned}\mathbb{E}\, g(X_1, \dots X_n) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots x_n)\, \mathrm{d}^n F_{X_1,\dots X_n}(x_1, \dots x_n)\\ &= \int_0^1 \cdots \int_0^1 g\left(F_{X_1}^{-1}(u_1), \dots F_{X_1}^{-1}(u_1)\right) \mathrm{d}^n F_{X_1,\dots,X_n}\left(F_{X_1}^{-1}(u_1), \dots F_{X_1}^{-1}(u_1)\right)\\ &= \int_0^1 \cdots \int_0^1 g\left(F_{X_1}^{-1}(u_1), \dots F_{X_1}^{-1}(u_1)\right) \mathrm{d}^n C\left(u_1, \dots u_1\right).\end{aligned}$$

□

Example 14.16. The copula for comonotonic random variables is $C(u_1, \dots, u_n) = \min_{i=1,\dots n} u_i$.

Lemma 14.17. *For comonotonic random variables X_i, $i = 1, \dots, n$, we have*

$$\mathbb{E}\, g(X_1, \dots, X_n) = \int_0^1 g\left(F_{X_1}^{-1}(u), \dots F_{X_n}^{-1}(u)\right) \mathrm{d}u.$$

14.5 PROBLEMS

Exercise 14.1 (McShane's Lemma on Lipschitz extensions)**.** *Let (Z, d) be a set equipped with a metric and let $f\colon U \to \mathbb{R}$ be Lipschitz with Lipschitz constant L, where $U \subset Z$. Then $\tilde{f}(z) := \inf_{u \in U} f(u) + L d(u, z)$ is well-defined for $z \in Z$, $f(u) = \tilde{f}(u)$ for $u \in U$ and $\tilde{f}\colon Z \to \mathbb{R}$ has Lipschitz constant L.*

Kirszbraun's theorem provides an extension for vector-valued functions, although the assertion for general Lipschitz functions is false.

15 Convexity

Gentlemen, we have run out of money. It is time to start thinking.

Ernest Rutherford, 1871–1937

Some parts follow a lecture by Mete Soner, but the content can be found in many elementary textbooks on convex analysis, for example in Boţ et al. [2009].

In what follows X is a real topological vector space.

15.1 PROPERTIES OF CONVEX FUNCTIONS

We consider functions to the extended reals, $f\colon X \to \mathbb{R} \cup \{\pm\infty\}$.

Definition 15.1. The domain of f is $\operatorname{dom} f := \{f < \infty\}$. f is proper, if its domain is not empty.

Definition 15.2. We shall say that $f\colon X \to \mathbb{R} \cup \{\pm\infty\}$ is *lower semicontinuous* (lsc.) if the sets $\{f > \lambda\}$ are open for all $\lambda \in \mathbb{R}$. The sets $\{f \le \lambda\}$ are called *lower levelsets*, *sublevel sets* or *trenches*.

Lemma 15.3. *The following are equivalent.*

(i) f *is lsc. at* $x_0 \in X$*;*

(ii) *for every* $\varepsilon > 0$ *there exists a neighborhood so that* $f(x) > f(x_0) - \varepsilon$ *for every* $x \in U$*;*

(iii) *then* $\liminf_{x \to x_0} f(x) \ge f(x_0)$.

Lemma 15.4. *The following are equivalent.*

(i) f *is lsc.;*

(ii) *the epigraph epi*$f := \{(x, \alpha) \in X \times \mathbb{R}\colon \alpha \ge f(x)\}$ *is closed in* $X \times \mathbb{R}$*;*

(iii) *the level sets* $\{f \le \lambda\}$ *are closed for all* $\lambda \in \mathbb{R}$.

Example 15.5. $\delta_A := \begin{cases} 0 & x \in A \\ +\infty & \text{else} \end{cases}$ is lsc., iff A is closed.

Definition 15.6. We shall call $f\colon X \to \mathbb{R}$

- convex, if $f\big((1-\lambda)x + \lambda y\big) \le (1-\lambda) f(x) + \lambda f(y)$ for $\lambda \in [0,1]$,
- concave, if $f\big((1-\lambda)x + \lambda y\big) \ge (1-\lambda) f(x) + \lambda f(y)$ for $\lambda \in [0,1]$, and
- affine, if $f\big((1-\lambda)x + \lambda y\big) = (1-\lambda) f(x) + \lambda f(y)$ for all $\lambda \in \mathbb{R}$.

Remark 15.7. Note, that f is affine iff $f(x) = d + x^*(x)$ for some linear x^*. Indeed, for f affine write $f(x) = a(0) + f(x) - a(0)$ and show that $f(x) - a(0)$ is linear.

Lemma 15.8. *If* f *is lsc. and convex, then*

$$f(x) = \sup_{a(\cdot) \le f(\cdot)} a(x) \qquad \text{for all } x \in X,$$

where a *is affine and* $a(\cdot) \le f(\cdot)$ *iff* $a(x) \le f(x)$ *for all* $x \in X$.

Proof. Note first that $\sup_{a(\cdot)\le f(\cdot)} a(x)$ is convex and lsc.

Conversely, consider

$$M := \{(x^*,\alpha) \in X^* \times \mathbb{R} \colon x^*(x) + \alpha \le f(x) \text{ for all } x \in X\}.$$

We show that M is not empty.

If $f \equiv +\infty$, the always $(x^*,\alpha) \in M$, hence $M \neq \emptyset$.

Otherwise, there is $y \in X$ so that $f(y) \in \mathbb{R}$. Then epi $f \neq \emptyset$ and $(y, f(y)-1) \notin$ epi f. As f is lsc., it follows from Lemma 15.4 that epi f is closed and convex. By the Hahn–Banach theorem there exists $(x^*,\alpha) \in X^* \times \mathbb{R}$ so that

$$x^*(y) + \alpha(f(y)-1) < x^*(x) + \alpha r \quad \text{for all } (x,r) \in \text{epi} f. \tag{15.1}$$

As $(y, f(y)) \in$ epi f it follows that $\alpha > 0$ and by rescaling (x^*, c), we may assume that $\alpha = 1$ and we get $x^*(y-x) + f(y) - 1 < r$ for all $(x,r) \in$ epi f. For $x \in$ dom f we have that $(x, f(x)) \in$ epi f, and thus $x^*(y-x) + f(y) - 1 < f(x)$, which actually holds for all $x \in X$. Consequently, the function $x \mapsto -x^*(x) + x^*(y) + f(y) - 1$ is a minorant and $M \neq \emptyset$.

We thus have

$$f(x) \ge \sup\{x^*(x) + \alpha \colon (x^*,\alpha) \in M\},$$

it remains to be shown that equality holds.

Assume there were $\tilde{x} \in X$ and $\tilde{r} \in \mathbb{R}$ such that

$$f(\tilde{x}) > \tilde{r} > \sup\{x^*(\tilde{x}) + \alpha \colon (x^*,\alpha) \in M\}. \tag{15.2}$$

Then $(\tilde{x},\tilde{r}) \notin$ epi f. Again, by Hahn–Banach theorem, there are $(\tilde{x}^*,\tilde{\alpha}) \in X^* \times \mathbb{R}$ and $\varepsilon > 0$ such that

$$\tilde{x}^*(x) + \tilde{\alpha} r > \tilde{x}^*(\tilde{x}) + \tilde{\alpha}\tilde{r} + \varepsilon \text{ for all } (x,r) \in \text{epi } f. \tag{15.3}$$

It follows that $\tilde{\alpha} \ge 0$, as $(\tilde{x}, r') \in$ epi f for every $r' \ge \tilde{r}$.

Assume that $f(\tilde{x}) \in \mathbb{R}$. Then $\tilde{\alpha}(r - \tilde{r}) > \varepsilon$ by (15.3), thus $\tilde{\alpha} > 0$. It follows from (15.3) that

$$f(x) > \frac{1}{\tilde{\alpha}} \tilde{x}^*(\tilde{x} - x) + \tilde{r} + \frac{\varepsilon}{\tilde{\alpha}}. \tag{15.4}$$

Hence, $x \mapsto \frac{1}{\tilde{\alpha}}\tilde{x}^*(\tilde{x}-x) + \tilde{r} + \frac{\varepsilon}{\tilde{\alpha}}$ is a minorant of $f(\cdot)$ which evaluates to $\tilde{r} + \frac{\varepsilon}{\tilde{\alpha}}$ at $\tilde{x}$, so that we get from (15.2) that $f(\tilde{x}) > \tilde{r} > \tilde{r} + \frac{\varepsilon}{\tilde{\alpha}}$, which is a contradiction.

Hence, $f(\tilde{x}) = \infty$, i.e., $\tilde{x} \notin$ dom X. As the domain of X is convex, we may separate $\tilde{x}$ from dom f, i.e., there is $\tilde{x}^*$ so that $\tilde{x}^*(x - \tilde{x}) > \varepsilon > 0$ (i.e., we may choose $\tilde{\alpha} = 0$ in (15.3)).

Consider the function $z^*(x) := -\tilde{x}^*(x - \tilde{x}) + \varepsilon$. By (15.3) we get that $z^*(x) \le 0$ for every $x \in$ dom f. As $M \neq \emptyset$, there are $y^* \in X^*$ and $\beta \in \mathbb{R}$ such that $y^*(x) + \beta \le f(x)$ for all $x \in X$. It follows from (15.2) that $\tilde{r} > y^*(\tilde{x}) + \beta$, so $\gamma := \frac{1}{\varepsilon}(\tilde{r} - y^*(\tilde{x}) - \beta) > 0$.

The function $a(x) := y^*(x) - \gamma\tilde{x}^*(x) + \gamma\tilde{x}^*(\tilde{x}) + \beta + \gamma\varepsilon$ is affine. For $x \in$ dom f we have that $a(x) = y^*(x) + \beta + \gamma\left(\underbrace{-\tilde{x}^*(x-\tilde{x}) + \varepsilon}_{z^*(x)}\right) \le y^*(x) + \beta \le f(x)$. Hence $a(\cdot)$ is an affine minorant of f and for $x = \tilde{x}$ one gets $a(\tilde{x}) := y^*(\tilde{x}) - \gamma\tilde{x}^*(\tilde{x}) + \gamma\tilde{x}^*(\tilde{x}) + \beta + \gamma\varepsilon = \tilde{r}$, which contradicts (15.4).

Hence, f is the pointwise supremum of affine functions. □

15.2 DUALITY

Definition 15.9 (Legendre–Fenchel[1] transformation)**.** The convex conjugate function is

$$\begin{aligned} f^* \colon X^* &\to \mathbb{R} \cup \{\infty\} \\ f^*(x^*) &:= \sup_{x \in X} x^*(x) - f(x) \end{aligned} \tag{15.5}$$

[1]Werner Fenchel, 1905–1988

and the bi-conjugate is

$$\begin{aligned} f^{**}\colon X &\to \mathbb{R}\cup\{\infty\} \\ f^{**}(x) &:= \sup_{x^*\in X^*} x^*(x) - f^*(x^*). \end{aligned}$$

Remark 15.10 (Fenchel's inequality, or Fenchel–Young[2] inequality)*.* By (15.5) it holds that

$$x^*(x) \le f(x) + f^*(x^*) \tag{15.6}$$

for all $x \in X$ and $x^* \in X^*$.

Lemma 15.11. *We have that $f \le g$ implies that $f^* \ge g^*$.*

Proof. Cf. Exercise 15.1. □

Lemma 15.12. *If $a(x) = d + y^*(x)$ is affine linear, then $a^*(x^*) = \begin{cases} -d & \text{if } x^* = y^* \\ +\infty & \text{else.}\end{cases}$ and $a^{**}(x) = a(x)$.*

Proof. Observe that

$$a^*(x^*) = \sup_{x\in X} x^*(x) - d - y^*(x) = \begin{cases} -d & \text{if } x^* = y^* \\ +\infty & \text{else.}\end{cases}$$

Further,

$$a^{**}(x) = \sup_{x^*\in X^*} x^*(x) - a^*(x^*) = \sup\{\underbrace{y^*(x) + d}_{x^*=y^*},\ \underbrace{x^*(x) - \infty}_{x^*\ne y^*}\} = y^*(x) + d = a(x)$$

for every $x \in X$, i.e., $a = a^{**}$. □

Example 15.13. On $X = \mathbb{R}$ let $f(x) := \frac{1}{p}|x|^p$, then $f^*(y) = \frac{1}{q}|y|^q$, where $\frac{1}{p} + \frac{1}{q} = 1$.

Indeed, the maximum is attained at $0 = \frac{\mathrm{d}}{\mathrm{d}x} xy - \frac{1}{p}|x|^p = y - x^{p-1}$, so $x^* = y^{\frac{1}{p-1}}$ and thus

$$f^*(y) = x^* y - \frac{1}{p}|x^*|^p = y^{\frac{1}{p-1}} y - \frac{1}{p} y^{\frac{p}{p-1}} = \frac{p-1}{p} y^{\frac{p}{p-1}} = \frac{1}{q} y^q.$$

Remark 15.14*.* f^* and f^{**} are lsc.

Theorem 15.15 (Fenchel–Moreau Theorem, Rockafellar)**.** *Let X be a Banach space. Let $f\colon X \to \overline{\mathbb{R}}$ be a proper, extended real valued lsc. and convex, function. Then $f = f^{**}$, where*

$$f^{**}(x) := \sup_{x^*\in X^*} x^*(x) - f^*(x^*).$$

Proof. By Fenchel's inequality (15.6) we have that that $f(x) \ge x^*(x) - f^*(x^*)$, and thus

$$f(x) \ge \sup_{x^*\in X^*} x^*(x) - f^*(x^*) = f^{**}(x),$$

i.e., $f \ge f^{**}$.

Let a be affine so that $a \le f$. Then $a^* \ge f^*$ and $a^{**} \le f^{**}$. Now by Lemma 15.12 we have that $a = a^{**}$, hence

$$f(x) = \sup_{a\le f} a(x) \le \sup_{a\le f^{**}} a(x) = f^{**}(x),$$

which is the converse inequality. □

Corollary 15.16 (The bipolar theorem)**.** *The polar cone is*

$$C^\circ := \{y \in X^*\colon y(c) \le 0 \text{ for all } c \in C\}.$$

Let C be a cone, then $C^{\circ\circ} := (C^\circ)^\circ = \overline{\operatorname{conv}\{\lambda c\colon \lambda \ge 0, c \in C\}}$.

[2] William Henry Young, 1863–1942

Proof. Consider the indicator function $f(c) := \delta_C(c) := \begin{cases} 0 & \text{if } c \in C, \\ +\infty & \text{else.} \end{cases}$ Then $\delta_C^*(y) = \sup_{c\in C} y(c)$ and $\delta_C^{**}(c) = \delta_{C^{\circ\circ}}(c)$ iff $C = C^{\circ\circ}$. □

Corollary 15.17 (Young's inequality)**.** *For $g(\cdot)$ strictly increasing it holds that*

$$xy \le \int_0^x g(u)\,\mathrm{d}u + \int_0^y g^{-1}(v)\,\mathrm{d}v,$$

where $x > 0$ and $y \in [0, g(x)]$; particularly

$$\int_0^y g^{-1}(v)\,\mathrm{d}v = \sup_x \left\{ xy - \int_0^x g(u)\,\mathrm{d}u \right\} \text{ and } \int_0^x g(u)\,\mathrm{d}u = \sup_y \left\{ xy - \int_0^y g^{-1}(v)\,\mathrm{d}v \right\}. \tag{15.7}$$

Proof. Set $f(x) := \int_0^x g(u)\,\mathrm{d}u$, then $0 = \frac{\mathrm{d}}{\mathrm{d}x} xy - \int_0^x g(u)\,\mathrm{d}u = y - g(x)$, i.e., $x^* = g^{-1}(y)$. Hence $f^*(y) = yg^{-1}(y) - \int_0^{g^{-1}(y)} g(u)\,\mathrm{d}u = \int_0^y g^{-1}(v)\,\mathrm{d}v$, and hence the assertion. □

15.3 PROBLEMS

Exercise 15.1. *Verify Lemma 15.11.*

Exercise 15.2. *For a family f_ι it holds that $(\inf_\iota f_\iota)^*(x^*) = \sup_\iota f_\iota^*(x^*)$, but $\left(\sup_\iota f_\iota\right)^*(x^*) \le \inf_\iota f_\iota^*(x^*)$.*

Exercise 15.3. *Show that $((1-\lambda) f_0 + \lambda f_1)^* \le (1-\lambda) f_0^* + \lambda f_1^*$ for $\lambda \in [0, 1]$.*

Exercise 15.4. *Set $g(x) := \alpha + \beta \cdot x + \gamma f(\lambda x + \delta)$, then $g^*(x^*) = -\alpha - \delta \frac{x^*-\beta}{\lambda} + \gamma f^*\left(\frac{x^*-\beta}{\lambda\gamma}\right)$, where $\lambda \ne 0$ and γ >0.*

Exercise 15.5 (Infimal convolution)**.** *Define the infimal convolution*

$$(f \Box g)(x) := \inf\{f(x-y) + g(y) : y \in \mathbb{R}^n\}$$

and more generally, $(f_1 \Box \ldots \Box f_m)(x) := \inf\left\{\sum_{i=1}^m f_i(x_i) : \sum_{i=1}^m = x\right\}$. Then, for f_i proper, convex and lsc., $(f_1 \Box \ldots \Box f_m)^* = f_1^* + \cdots + f_m^*$.

Exercise 15.6. *Show that the conjugate of $f(x) = e^x$ is $f^*(x^*) = \begin{cases} x^* \log x^* - x^* & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \\ +\infty & \text{if } x^* < 0 \end{cases}$.*

Sample Average Approximation (SAA) 16

This chapter is based on Shapiro et al. [2021]

16.1 SAA

Let $X \subset \mathbb{R}^n$ be closed and $X \neq \emptyset$. Consider the problem $\vartheta^* := \min_{x \in X} \underbrace{\mathbb{E}\, F(x,\xi)}_{=: f(x)}$ which we compare with $\hat{\vartheta}_N := \min_{x \in X} \underbrace{\frac{1}{N} \sum_{i=1}^{N} F(x, \xi_j)}_{=: \hat{f}_N(x)}$ for the empirical measure $P_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}$ and iid samples ξ_i.

16.1.1 Pointwise LLN

Suppose that $\mathbb{E}\, F(x,\xi) < \infty$.

Lemma 16.1. *The following hold true:*

(i) $\mathbb{E}\, \hat{f}_N(x) = f(x)$*, i.e.,* $\hat{f}_N(x)$ *is an unbiased estimator for* $f(x)$*;*

(ii) (LLN) For every $x \in X$ *it holds that* $\hat{f}_N(x) \to f(x)$*, as* $N \to \infty$ *with probability* 1*.*

Proposition 16.2. *The estimator* $\hat{\vartheta}_N$ *is not necessarily consistent, it holds in general that*

$$\limsup_{N \to \infty} \hat{\vartheta}_N \leq \vartheta^*.$$

Proof. We have that $\hat{\vartheta}_N \leq \hat{f}_N(x)$ for every $x \in X$, thus

$$\limsup_{N \to \infty} \hat{\vartheta}_N \leq \lim_{N \to \infty} \hat{f}_N(x) = f(x)$$

by the Law of Large Numbers (ii). Thus

$$\limsup_{N \to \infty} \hat{\vartheta}_N \leq \inf f(x) = \vartheta^*.$$

$\square$

Proposition 16.3. *The estimator* $\hat{\vartheta}_N$ *is downside biased, it holds that* $\mathbb{E}\, \hat{\vartheta}_{N+1} \leq \mathbb{E}\, \hat{\vartheta}_N \leq \vartheta^*$*.*

Proof. It holds that

$$\begin{aligned} \mathbb{E}\, \hat{\vartheta}_N &= \mathbb{E} \min_{x \in X} \frac{1}{N} \sum_{j=1}^{N} F(x, \xi_j) \leq \min_{x \in X} \mathbb{E}\, \frac{1}{N} \sum_{j=1}^{N} F(x, \xi_j) \\ &= \min_{x \in X} \frac{1}{N} \sum_{j=1}^{N} \mathbb{E}\, F(x, \xi_j) = \min_{x \in X} f(x) = \vartheta^*. \end{aligned}$$

Further, note that $\hat f_{N+1}(x) = \frac{1}{N+1}\sum_{i=1}^{N+1} F(x,\xi_i) = \frac{1}{N+1}\sum_{i=1}^{N+1}\frac{1}{N}\sum_{j\neq i} F(x,\xi_j)$, thus

$$\begin{aligned}\mathbb{E}\,\hat\vartheta_{N+1} &= \mathbb{E}\min_{x\in X}\hat f_{N+1}(x) = \mathbb{E}\min_{x\in X}\frac{1}{N+1}\sum_{i=1}^{N+1}\frac{1}{N}\sum_{j\neq i}F(x,\xi_j)\\ &\geq \mathbb{E}\,\frac{1}{N+1}\sum_{i=1}^{N+1}\underbrace{\min_{x\in X}\frac{1}{N}\sum_{j\neq i}F(x,\xi_j)}_{\hat\vartheta_N} = \mathbb{E}\,\frac{1}{N+1}\sum_{i=1}^{N+1}\hat\vartheta_N = \mathbb{E}\,\hat\vartheta_N,\end{aligned}$$

thus the assertion. □

16.1.2 Pointwise and Functional CLT

Suppose here that

(i) $\sigma(x)^2 := \operatorname{var} F(x,\Xi) < \infty$ and

(ii) $|F(x,\xi) - F(x',\xi)| \leq C(\xi)\,\|x-x'\|$ a.s. and $\mathbb{E}\,C(\xi)^2 < \infty$.

Lemma 16.4. *$f(\cdot)$ is Lipschitz with constant $\mathbb{E}\,C(\xi)$.*

Proof. It follows from (ii) by taking expectations that $f(x)-f(x') = \mathbb{E}\,F(x,\xi) - \mathbb{E}\,F(x',\xi) \leq \mathbb{E}\,C\,\|x-x'\|$, from which the assertion follows. □

We have that

$$\sqrt{N}\,(f_N(x) - f(x)) \xrightarrow{\mathcal{D}} Y(x) \sim \mathcal{N}\left(0,\sigma(x)^2\right).$$

More generally,

$$\sqrt{N}\,(f_N(x_1) - f(x_1),\dots f_N(x_n) - f(x_n)) \xrightarrow{\mathcal{D}} Y(x) \sim \mathcal{N}\,(0,\Sigma)\,,$$

where $\Sigma = \operatorname{cov}\,(F(x_i,\xi), F(x_i,\xi))_{i,j=1}^n$.

In a functional way,

$$\sqrt{N}\,(f_N(\cdot) - f(\cdot)) \xrightarrow{\mathcal{D}} Y\colon \Omega \to C(X),$$

where $Y\colon\Omega\to C(X)$ is called a random element in $C(X)$.

Theorem 16.5. *If (i) and (ii), then*

(i) $\hat\vartheta_N = \inf_x \hat f_N(x) + o\left(N^{-1/2}\right)$ and

(ii) $N^{1/2}\left(\hat\vartheta_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \inf_{s\in S} Y(s)$, where $S = \arg\min_{x\in X} f(x) \subset X$.

Proof. The proof uses the Δ-method described in Section 16.2 below for finite dimensions. □

Remark 16.6. We obtain from (ii) that $\hat\vartheta_N = \vartheta^* + N^{-1/2}\inf_{s\in S}Y(s) + o\left(N^{-1/2}\right)$. For $s=\{x^*\}$ we have that $\inf_{s\in S}Y(s) = Y(x^*) \sim N\left(0,\sigma(x^*)^2\right)$ and hence $\mathbb{E}\,\hat\vartheta_N = \vartheta^* + o\left(N^{-1/2}\right)$.

However, convergence is slower, in general, if S consists of more than 1 point.

16.2 THE Δ-METHOD

Proposition 16.7. *Let $Y_N \in \mathbb{R}^d$ be random vectors with, $Y_N \to \mu \in \mathbb{R}^d$ in probability and $\mathbb{R} \ni \tau_N \nearrow \infty$ deterministic numbers such that $\tau_N\,(Y_N - \mu) \xrightarrow{\mathcal{D}} Y$. Further, let $G\colon \mathbb{R}^d \to \mathbb{R}^n$ be differentiable at μ. Then $\tau_N\,(G\,(Y_N) - G(\mu)) \xrightarrow{\mathcal{D}} J\cdot Y$, where $J = \nabla G(\mu)$ is the $n\times d$ Jacobian matrix at μ.*

Proof. Notice, that $G(y)-G(\mu) = J(y-\mu)+r(y)$, where $r(y) = o(\|y-\mu\|)$, so that we have $\tau_N\left(G\left(Y_N\right)-G(\mu)\right) = J\underbrace{\tau_N\left(Y_N-\mu\right)}_{\xrightarrow{\mathcal{D}} Y}+\tau_N r(Y_N)$. We have that $\tau_N\left(Y_N-\mu\right) = \mathcal{O}(1)$ (as it converges in distribution), hence $\|Y_N-\mu\| = \mathcal{O}(1)$ and thus $r(Y_N) = o\left(\|Y_N-\mu\|\right) = o\left(\tau_N^{-1}\right)$. Thus the result. □

Claim 16.8. For $N^{1/2}\left(Y_N-\mu\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\Sigma\right)$ we have particularly that $N^{1/2}\left(G(Y_N)-G(\mu)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, J\Sigma J^\top\right)$.

17 Weak Topology of Measures

17.1 GENERAL CHARACTERISTICS

Definition 17.1. Let (X, d) be a metric space. The *weak topology* of probability measures is characterized by

$$\int_X h \,\mathrm{d}P_n \xrightarrow[n\to\infty]{} \int_X h \,\mathrm{d}P \qquad \text{for all bounded and continuous functions } h\colon X \to \mathbb{R}.$$

Theorem 17.2 (Riesz representation theorem)**.** *For any continuous linear functional $\psi\colon C_0(X) \to \mathbb{R}$ (the bounded functions vanishing at infinity on a locally compact Hausdorff space X) there is a regular, countatbly additive measure μ on the Borels so that*

$$\psi(h) = \int_X h \,\mathrm{d}\mu \quad \textit{for all } h \in C_0(X).$$

Definition 17.3. Let μ, ν be (probability) measures. The *Lévy–Prokhorov* metric is

$$\pi(\mu,\nu) := \inf\{\varepsilon > 0\colon \mu(A) \le \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \le \mu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathscr{B}(X)\}, \tag{17.1}$$

where $A^\varepsilon := \bigcup_{a\in A} B_\varepsilon(a)$ is the ε-fattening (or ε-enlargement) of $A \subset X$.

Remark 17.4. Notice, that $\pi(P,Q) \le 1$ for probability measures.

Definition 17.5. A collection $M \subset \mathcal{P}(X)$ of probability measures on (X,d) is *tight* iff for every $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset X$ so that

$$\mu(K_\varepsilon) > 1 - \varepsilon \quad \text{for all } P \in M.$$

Example 17.6. The collection $\{\delta_n\colon n = 1,2,\dots\}$ on $\mathbb{R}$ is not tight, while $\{\delta_{1/n}\colon n = 1,2,\dots\}$ is.

Example 17.7. A collection of Gaussian measures $\{\mathcal{N}(\mu_i,\Sigma_i)\colon i \in I\}$ is tight, if $\{\mu_i\colon i \in I\}$ and $\{\Sigma_i\colon i \in I\}$ are uniformly bounded.

Theorem 17.8 (Prokhorov's theorem)**.** *The following hold true:*

(i) *The metric π in (17.1) metrizes the weak topology of measures.*

(ii) *A set $\mathcal{K} \subset \mathcal{P}(X)$ is tight iff $\overline{\mathcal{K}}$ is sequentially compact.*

(iii) *If $\{P_n\colon n = 1,2,\dots\} \subset \mathcal{P}(\mathbb{R}^d)$ is tight, then there is a subsequence and a measure $P \in \mathcal{P}(\mathbb{R}^d)$ so that $P_n \to P$ weakly.*

Properties

(i) If (X,d) is separable, convergence of measures in the Lévy–Prokhorov metric is equivalent to weak convergence of measures. Thus, π is a metrization of the topology of weak convergence on $\mathcal{P}(X)$.

(ii) The metric space $(\mathcal{P}(X),\pi)$ is separable if and only if (X,d) is separable.

(iii) If $(\mathcal{P}(X),\pi)$ is complete then (X,d) is complete. If all the measures in $\mathcal{P}(X)$ have separable support, then the converse implication also holds: if (X,d) is complete then $(\mathcal{P}(X),\pi)$ is complete.

(iv) If (X,d) is separable and complete, a subset $\mathcal{K} \subseteq \mathcal{P}(X)$ is relatively compact if and only if its π-closure is π-compact.

17.2 THE WASSERSTEIN DISTANCE

Some points here follow [Pflug and Pichler, 2014].

Definition 17.9 (Optimal transportation cost)**.** Given two probability spaces $(\Xi, \mathcal{F}, P)$ and $\left(\tilde{\Xi}, \tilde{\mathcal{F}}, \tilde{P}\right)$, the Wasserstein distance of order $r \geq 1$ (optimal transportation costs) is

$$\mathsf{d}_r\left(P, \tilde{P}\right) = \inf_{\pi} \left(\iint_{\Xi \times \tilde{\Xi}} d\left(\xi, \tilde{\xi}\right)^r \pi\left(d\xi, d\tilde{\xi}\right) \right)^{1/r}, \tag{17.2}$$

where the infimum is taken over all (bivariate) probability measures π on $\Xi \times \tilde{\Xi}$ having the marginals P and $\tilde{P}$, that is

$$\pi\left(A \times \tilde{\Xi}\right) = P(A) \text{ and } \pi\left(\Xi \times B\right) = \tilde{P}(B) \tag{17.3}$$

for all measurable sets $A \in \mathcal{F}$ and $B \in \tilde{\mathcal{F}}$. The optimal measure π is called the *optimal transport plan*.

Remark 17.10. Occasionally, the Wasserstein distance is also considered for a (convex) function $c(x, y)$ instead of the distance $d(x, y)^r$.

Proposition 17.11 (Embedding)**.** *It holds that*

$$\mathsf{d}_r\left(P, \delta_{\xi_0}\right)^r = \int_{\Xi} \mathsf{d}\left(\xi, \xi_0\right)^r P\left(\mathrm{d}\xi\right),$$

and the mapping

$$\begin{aligned} i\colon\ (\Xi, \mathsf{d}) &\to \left(\mathcal{P}_r\left(\Xi; \mathsf{d}\right), \mathsf{d}_r\right), \\ \xi &\mapsto \delta_{\xi}(\cdot) \end{aligned}$$

assigning to each point $\xi \in \Xi$ its point measure δ_ξ located on ξ (Dirac measure[1]) is an isometric embedding for all $1 \leq r < \infty$ ($(\Xi, \mathsf{d}) \hookrightarrow \mathcal{P}_r\left(\Xi; \mathsf{d}\right)$).

Proof. There is just one single measure with marginals P and δ_{ξ_0}, which is the transport plan $\pi = P \otimes \delta_{\xi_0}$. Hence

$$\mathsf{d}_r\left(P, \delta_{\xi_0}\right)^r = \int_{\Xi} \int_{\Xi} \mathsf{d}\left(\xi, \tilde{\xi}\right)^r \delta_{\xi_0}\left(d\tilde{\xi}\right) P\left(d\xi\right) = \int_{\Xi} \mathsf{d}\left(\xi, \xi_0\right)^r P\left(d\xi\right),$$

the first assertion.

For the particular choice $P = \delta_{\xi_0}$ the latter formula simplifies to

$$\mathsf{d}_r\left(\delta_{\tilde{\xi}_0}, \delta_{\xi_0}\right)^r = \int_{\Xi} \mathsf{d}\left(\xi, \xi_0\right)^r \delta_{\tilde{\xi}_0}\left(d\xi\right) = \mathsf{d}\left(\tilde{\xi}_0, \xi_0\right)^r,$$

and hence $\xi \mapsto \delta_\xi$ is an isometry. □

Notice that if d is inherited by ξ, then $\mathsf{d}_r(P, \delta_{\xi_0})^r = \int_{\Xi} \|\xi - \xi_0\|^r P(d\xi)$.

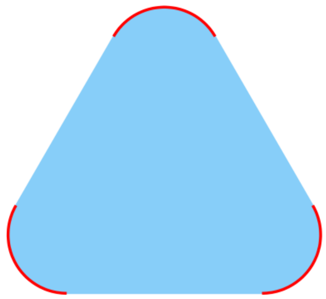
17.3 THE REAL LINE

Theorem 17.12 (Cf. Rachev and Rüschendorf [1998, Theorem 2.18])**.** *Let P and $\tilde{P}$ be probability measures on the real line with "cdf" $F(x) := P\left((-\infty, x]\right)$ and $G(x) := \tilde{P}\left((-\infty, x]\right)$. Let π be the measure on $\mathbb{R}^2$ with cdf. $H(x, y) := \min\{F(x), G(y)\}$. Then π is optimal for the Kantorovich transportation problem between P and $\tilde{P}$ for every cost function $c(x, y) = c(x - y)$ where $c(\cdot)$ is convex. Further,*

$$\mathsf{d}_c(P, \tilde{P}) = \int_0^1 c\left(F^{-1}(u) - G^{-1}(u)\right) \mathrm{d}u.$$

[1] $\delta_\xi(A) := \mathbb{1}_A(\xi) = \begin{cases} 1 & \text{if } \xi \in A \\ 0 & \text{if } \xi \notin A \end{cases}$ is the usual Dirac measure.

Corollary 17.13. *For the cost function* $c(\cdot) = |\cdot|$ *we have further that* $\mathrm{d}(P, \tilde{P}) = \int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right| \mathrm{d}u = \int_{-\infty}^{\infty} |F(x) - G(x)| \, \mathrm{d}x$.



Remark 17.14.

(i) If G does not give mass to points, then one may define $T := G^{-1} \circ F$ and it holds that

$$\int_{-\infty}^{x} \mathrm{d}P = F(x) = G(T(x)) = \int_{-\infty}^{T(x)} \mathrm{d}\tilde{P} \tag{17.4}$$

The transport map T is a *monotone rearrangement* of P to $\tilde{P}$.

(ii) Suppose P and $\tilde{P}$ have the densities $f = F'$and $g = G'$. Then differentiating (17.4) gives

$$f(x) = g(T(x)) \cdot T'(x).$$

18

Topologies For Set-Valued Convergence

18.1 TOPOLOGICAL FEATURES OF MINKOWSKI ADDITION

Theorem (Topological properties of Minkowski addition in locally compact vector spaces)**.** *Let A and B be sets.*

(i) $\mathring{A} + B$ *is open and* $\mathring{A} + B \subset (A + B)^\circ \subset A + B$*. If A or B is open, then* $A + B = (A + B)^\circ$*;*

(ii) $\bar{A} + \bar{B} \subset \overline{A + B}$*. If A or B is bounded, then* $\overline{A + B} = \bar{A} + \bar{B}$*;*

(iii) For A and B closed and A (or B) bounded, $\partial (A + B) \subset \partial A + \partial B$*.*

Proof. Let $x \in \mathring{A} + B$ have the composition $x = a + b$ with $a \in B_r (a) \subset \mathring{A}$ for some $r > 0$ and $b \in B$. Then $x \in B_r (x) = B_r (a + b) = B_r (a) + \{b\} \subset \mathring{A} + B$, so $\mathring{A} + B$ is open. As $\mathring{A} + B \subset A + B$ and $\mathring{A} + B$ open it is immediate that $\mathring{A} + B \subset (A + B)^\circ$, which is ((i)) (the rest being obvious).

Let $a \in \bar{A}$ and $b \in \bar{B}$. Choose $a_k \in A$ with $a_k \to a \in \bar{A}$ and $b_k \in B$ with $b_k \to b \in \bar{B}$. Obviously $a_k + b_k \in A + B$ and thus $a + b \in \overline{A + B}$, whence $\bar{A} + \bar{B} \subset \overline{A + B}$.

As for the converse let $x \in \overline{A + B}$, so there is a sequence $x_k = a_k + b_k$ with $a_k \in A$ and $b_k \in B$ and $x_k \to x$. Assume (wlog.) A bounded, thus there is a subsequence such that $a_k \to a \in \bar{A}$, and thus $b_k = x_k - a_k \to x - a$ converges as well with $b_k \to x - a =: b \in \bar{B}$. That is $x = a + b \in \bar{A} + \bar{B}$ and thus $\overline{A + B} \subset \bar{A} + \bar{B}$.

Observe first that $\partial (A + B) \subset \overline{A + B} \subset \bar{A} + \bar{B} = A + B$. Suppose that $x \in \partial (A + B)$ can be written as $x = a + b$ for some $a \in \mathring{A}$ and $b \in B$. Then $x = a + b \in \mathring{A} + B \subset (A + B)^\circ$, whence $x \notin \partial (A + B)$. This is a contradiction, so $a \notin \mathring{A}$, that is $a \in \partial A$. By similar reasoning (A and B reversed) we find that $b \in \partial B$ as well, which is the desired assertion. □

The assertion *bounded* in (ii) may not be dropped: To see this consider the *closed* sets $A := \mathbb{R} \times \{0\} \subset \mathbb{R}^2$ and $B := \left\{\left(x, \mathrm{e}^{-x^2}\right) : x \in \mathbb{R}\right\}$. $A + B$ is *open* though, and $\bar{A} + \bar{B} \subsetneq \overline{A + B}$.

18.1.1 Topological features of convex sets

Theorem (Topological properties of convex sets)**.**

- *If A is open, then* $\operatorname{conv} A$ *is open;*[1]
- *If A is bounded, then* $\operatorname{conv} A$ *is bounded;*
- *If A is closed and bounded, then* $\operatorname{conv} A$ *is closed and bounded;*
- $\operatorname{conv} (A + B) = \operatorname{conv} A + \operatorname{conv} B$*.*

Proof. Let $a \in \operatorname{conv} A$ have a representation $a = \sum_{i=1}^n \lambda_i a_i$ with $a_i \in A$, whence $a \in \sum_{i=1}^n \lambda_i A \subset \operatorname{conv} A$. As A is open it follows from Theorem (18.1) ((i)) that $\sum_{i=1}^n \lambda_i A$ is open, whence $\operatorname{conv} A$ is open.

Boundedness is obvious.

Let $a \in \overline{\operatorname{conv} A}$. Then there is a sequence $a_k = \sum_{i=1}^{d+1} \lambda_i^{(k)} a_i^{(k)} \in \operatorname{conv} A$ with $a_k \to a$ (here we use Carathéodory's theorem; the statement is wrong in non-finite dimensions). By picking subsequences we may assume that $a_1^{(k)}$ converges, $a_2^{(k)}$ converges, etc. and finally $a_k^{(d+1)}$, and moreover all $\lambda_i^{(k)}$. □

[1] The statement *if A is closed, then* $\operatorname{conv} A$ *is closed* is wrong (why?).

18.2 PRELIMINARIES AND DEFINITIONS

In a vector space X the Minkowski sum (also known as dilation) of two sets A and B is $A+B := \{a+b\colon a\in A,\ b\in B\}$, and the product with a scalar p is $p\cdot A := \{p\cdot a\colon a\in A\}$.

18.2.1 Convexity, and Conjugate Duality

The support function of a set $A\subset X$ is

$$s_A(x^*) := \sup_{a\in A} x^*(a), \tag{18.1}$$

where $x^*\in X^*$ is from the dual of a Banach space $(X,\|\cdot\|)$ with norm $\|\cdot\|$; its dual we will denote as $(X^*,\|\cdot\|)$, as no confusion with denoting the norm in the dual again by $\|\cdot\|$ will be possible anyway.

Remark 18.1 (A collection of properties). Important properties of the support function include

(i) $s_A\le s_B$ whenever $A\subset B$ (more specifically, $s_A(x^*)\le s_A(x^*)$ for all x^*),

(ii) $s_{\lambda\cdot A}(x^*) = s_A(\lambda\cdot x^*) = \lambda\cdot s_A(x^*)$ for $\lambda>0$ (positive homogeneity) and $s_A(0)=0$,

(iii) $s_{A+B} = s_A+s_B$,

(iv) $s_{\overline{\operatorname{conv} A}} = s_A$[2] and

(v) s_A is convex, that is $s_A\left((1-\lambda)x_0^*+\lambda x_1^*\right)\le (1-\lambda)s_A\left(x_0^*\right)+\lambda s_A\left(x_1^*\right)$ whenever $0\le\lambda\le 1$.

By employing the indicator function of the set A, $\mathbb{I}_A(a) := \begin{cases} 0 & \text{if } a\in A\\ \infty & \text{else}\end{cases}$, it is immediate that

$$s_A(x^*) = \sup_{x\in X} x^*(x) - \mathbb{I}_A(x),$$

where the supremum ranges over all $x\in X\supset A$ now. The support function itself thus is the usual *convex conjugate function* of $\mathbb{I}_A$, which we denote $s_A=\mathbb{I}_A^*$. The *bi-conjugate function* of $\mathbb{I}_A$ (the conjugate of s_A) is the function

$$s_A^*(a) := \sup_{x^*\in X^*} x^*(a) - s_A(x^*) = \begin{cases} 0 & \text{if } x^*(a)\le s_{\overline{\operatorname{conv} A}}(x^*) \text{ for all } x^*\in X^*\\ \infty & \text{else,}\end{cases}$$

and by the Rockafellar-Fenchel-Moreau-duality Theorem (cf. Rockafellar [1974]) one further infers that $s_A^* = \mathbb{I}_{\overline{\operatorname{conv} A}}$.

This also reveals the relation

$$\overline{\operatorname{conv} A} = \left\{s_A^*<\infty\right\} = \bigcap_{x^*\in X^*}\{a\colon x^*(a)\le s_A(x^*)\} = \bigcap_{x^*\in X^*}\{x^*\le s_A(x^*)\},$$

from which follows that the correspondence $A\mapsto s_A$, restricted to *convex*, compact sets $A\in C$, is one-to-one (injective).

18.2.2 Pompeiu–Hausdorff Distance

Having addition and multiplication available for sets an adequate and fitting notion of distance is useful. For this define the distance from a to a set B as $d(a,B) := \inf_{b\in B} d(a,b)$, where d is the distance function. The *deviation* of the set A from the set B is $\mathbb{D}(A,B) := \sup_{a\in A} d(a,B)$,[3] and the *Pompeiu-Hausdorff distance* is $\mathbb{H}(A,B) := \max\{\mathbb{D}(A,B),\,\mathbb{D}(B,A)\}$ (cf. Rockafellar and Wets [1997]).

[2] $\operatorname{conv} A := \{\sum_i \lambda_i a_i\colon \lambda_i\ge 0, \sum_i\lambda_i = 1 \text{ and } a_i\in A\}$ is the convex hull of A.
[3] in some references Hess [2002] also called excess of A over B.

Note that $\mathbb{D}(A,B)=0$ iff A is contained in the topological closure, $A\subset\bar{B}$, and $\mathbb{H}(A,B)=0$ iff $\bar{A}=\bar{B}$; moreover $\mathbb{H}(A,B)=\mathbb{H}\left(\overline{A},B\right)$ and obviously $\mathbb{H}(A,B)\le\sup_{a\in A,b\in B}d(a,b)$.

In a normed space with $d(a,b)=\|b-a\|$ it is enough to consider the boundaries, as we have in addition that $\mathbb{H}(A,B)=\mathbb{H}(\partial A,\partial B)$ if $\overline{A}$ and $\overline{B}$ are (sequentially) compact (i.e., A and B are relatively compact); moreover

$$\mathbb{H}(A,B)=\|b-a\| \tag{18.2}$$

for some $a\in\partial A$ and $b\in\partial B$ in this situation.

Lemma 18.2. *The deviation $\mathbb{D}$ and the Pompeiu-Hausdorff distance $\mathbb{H}$ satisfy the triangle inequality, $\mathbb{D}(A,C)\le\mathbb{D}(A,B)+\mathbb{D}(B,C)$ and $\mathbb{H}(A,C)\le\mathbb{H}(A,B)+\mathbb{H}(B,C)$.*

$(\mathcal{C},\mathbb{H})$, where $\mathcal{C}$ is the set of all nonempty, compact and convex subsets of X, is a Polish space (i.e. a complete, separable and metric space), provided that (X,d) is Polish.

Proof. See, e.g., Castaing and Valadier [1977]. □

The concept of the Hausdorff distance and the support functions introduced above link as follows to a nice ensemble: in a normed space ($d(a,b)=\|b-a\|$) the deviation $\mathbb{D}$, using Minkowski addition, rewrites as $\mathbb{D}(A,C)=\inf\{r>0\colon A\subset C+r\cdot B_X\}$ where $B_X=\{x\colon \|x\|\le 1\}$ is the unit ball and $C_r:=C+r\cdot B_X$ is the r-*fattening* of C. If A and C are convex, then $\mathbb{D}(A,C)=\inf\left\{r>0\colon s_A\le s_C+r\cdot s_{B_X}\right\}$, where "$\le$" is the usual "$\le$"-comparison of functions ($s_A\le s_C+r\cdot s_{B_X}$ iff $s_A(x^*)\le s_C(x^*)+r\cdot s_{B_X}(x^*)$ for all $x^*\in X^*$). As $s_{B_X}(x^*)=\sup_{b\in B_X}x^*(b)=\|x^*\|$, the norm of x^* in the dual $(X^*,\|.\|)$ by the Hahn-Banach Theorem, this simplifies further and it follows for general sets that

$$\mathbb{D}(\operatorname{conv}A,\ \operatorname{conv}C)=\inf\left\{r>0\colon s_A-s_C\le r\cdot s_{B_X}\right\}=\sup_{\|x^*\|\le 1}s_A(x^*)-s_C(x^*),$$

and the Pompei-Hausdorff distance thus is

$$\mathbb{H}(\operatorname{conv}A,\ \operatorname{conv}C)=\sup_{\|x^*\|\le 1}\left|s_A(x^*)-s_C(x^*)\right| \tag{18.3}$$

in terms of seminorms. These observations convincingly relate the Pompeiu-Hausdorff distance with Minkowski addition of convex sets.

It follows from the preceding discussion and remarks that for relatively compact sets A and C there are $a\in\partial A$, $c\in\partial C$ and $\|x^*\|\le 1$ such that $\mathbb{D}(A,C)=\|c-a\|=x^*(a-c)$. x^* is an outer normal for both sets, $\operatorname{conv}A$ and $\operatorname{conv}C$.

18.3 LOCAL DESCRIPTION

The sub-differential of a real-valued function $f\colon X^*\to\mathbb{R}$ at a point $x^*\in X^*$ is the set [4]

$$\partial f(x^*):=\{u\in X\colon f(z^*)-f(x^*)\ge z^*(u)-x^*(u)\ \text{ for all } z^*\in X^*\}\subset X.$$

Notably $\partial f(x^*)$ is a subset of X, so ∂f is a set-valued mapping which is expressed by writing

$$\begin{aligned}\partial f\colon X^*&\rightrightarrows X\\ x^*&\mapsto\partial f(x^*)\,.\end{aligned}$$

The symbol $\rightrightarrows X$ indicates that the outcomes are subsets – a collection of elements – of X.

With the sub-differential at hand we may add the following standard characterization of the support function s_A of a set A, which will turn out useful for our purpose:

Lemma 18.3. *The support function s_A has the sub-differential $\partial s_A(x^*)=\arg\max_{\overline{\operatorname{conv}A}}x^*$.[5] Moreover $\partial s_A(x^*)\subset\partial A$.*

[4] note, that $\partial f(x^*)\subset X$ is a subset in the pre-dual X rather than X^{**}.

[5] We shall abbreviate the argument of the maximum of a function f restricted to D by $\arg\max_D f:=\arg\max\{f(x): x\in D\}$.

Proof. With $u \in \arg\max_{\operatorname{conv} A} x^* \subset \operatorname{conv} A$, for any $z^* \in X^*$ we have that $s_A(z^*) \geq z^*(u) = s_A(x^*) + z^*(u) - x^*(u)$ and hence $u \in \partial s_A(x^*)$.

Conversely, with $a \in \partial s_A(x^*)$ we have that $s_A(z^*) - s_A(x^*) \geq z^*(a) - x^*(a)$ or

$$x^*(a) \geq s_A(x^*) + z^*(a) - s_A(z^*) \tag{18.4}$$

for all z^*. For the particular choice $z^* = 0$ we find that $x^*(a) \geq s_A(x^*)$ and it remains to show that $a \in \overline{\operatorname{conv} A}$. Suppose that $a \notin \overline{\operatorname{conv} A}$, then – by the Hahn-Banach Theorem – there is a z^* such that $z^*(a) > \sup\left\{z^*(a') : a' \in \overline{\operatorname{conv} A}\right\} = s_A(z^*)$. This same equation holds for multiples $\lambda \cdot z^*$ $(\lambda > 0)$, hence (18.4) cannot hold in general; thus, $a \in \overline{\operatorname{conv} A}$.

The second statement is obvious. □

Index