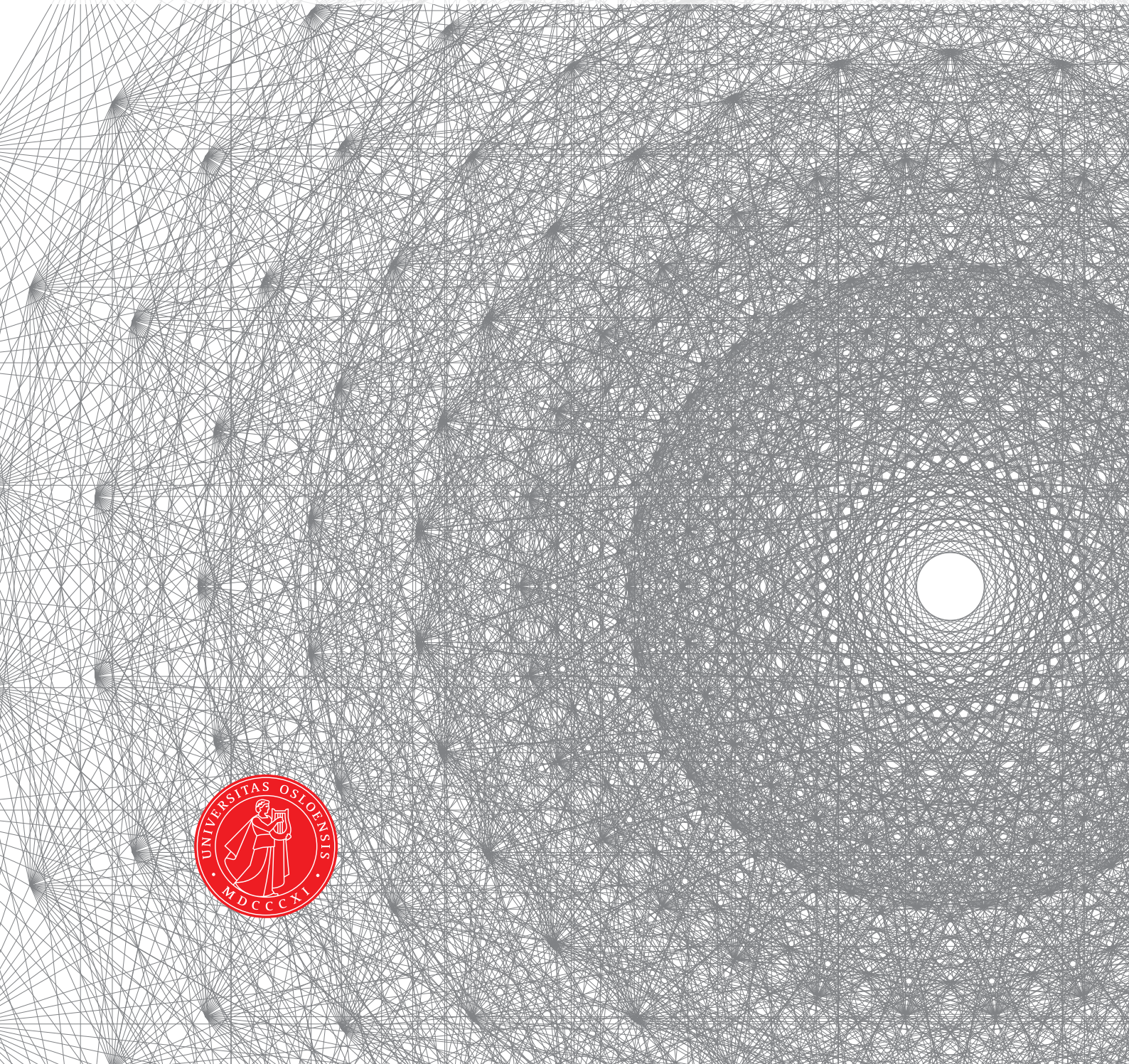
UiO : Department of Mathematics
University of Oslo

Optimization under uncertainty with conditional Value-at-Risk

Marius Havgar

Master's Thesis, Spring 2021



This master's thesis is submitted under the master's programme *Computational Science*, with programme option *Applied Mathematics and Risk Analysis*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 30 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

The objective of this thesis has been the study of risk analysis and optimization under uncertainty. Theoretical aspects of several risk measures have been contemplated, and we have discussed their strengths and weaknesses. Of special interest were coherent risk measures, and in particular conditional Value-at-Risk. We have shown that this is an important risk measure which is suitable for a wide range of applications. We have studied how Value-at-Risk and conditional Value-at-Risk can be used to estimate the future risk in a financial market, and how Value-at-Risk may give unsatisfactory results in this application. When applied to the optimization of structural design, CVaR leads to beneficial compared to similar optimization under Value-at-Risk. Moreover, we have considered the optimization of reinsurance contracts, where we have shown that multivariate reinsruance contracts can be optimized using CVaR.

Acknowledgements

First and foremost, I owe my deepest gratitude to my supervisor Kristina Rognlien Dahl. Thank you for introducing me to the wonderful field of risk analysis. But more than anything, thank you for your guidance and never ending positivity and support - it has bee invaluable during these challenging times. I would also like to thank all my professors and group teachers, who have all contributed to a wonderful study experience. Thank you also to Tuva. Thank you for always being there, for your supportiveness and patience. (Especially during mathematical discussions..)

And lastly, thank you to my friends and study colleagues in B1001 and B1002, for a never ending supply of interesting, albeit occasionally heated, discussions - and coffee breaks.

Contents

List of Figures

List of Tables

Contents

List of Figures

List of Tables

Introduction

Outline of the thesis

The thesis is organized as follows:

Chapter 1 Chapter 1 contains preliminary mathematics needed for the discussion of risk analysis. We introduce the concept of coherent risk measures and coherency in optimization under uncertainty.

Chapter 2 Chapter 2introduces the problem of reinsurance. We study optimization of reassurance contracts for univariate and multivaraite risks for Value-at-Risk and conditional Value-at-Risk. We present results and examples for optimal reinsurance for worst-case dependency between risks.

Chapter 3 Chapter 3 explores how these risk measures can be used to estimate the future risk in a financial market. We see through examples how conditional Value-at-Risk produces more conservative estimates than Value-at-Risk, and studied how we may use this in combination with a stochastic model to produce risk estimates.

Chapter 4 Chapter 4 presents the optimization of structural design using risk analysis. We study how failure probability and buffered failure probability may be applied, and study the gain in computability with the use of conditional Value-at-Risk.

My contributions

Throughout this thesis I have highlighted independent contributions with the symbol ✻. Such contributions include elaborating on missing details in proofs or examples from other sources, illustrations of concepts through constructed examples, or the extension of known results to new settings. Moreover, all the numerical examples and programming have been created by me. Most notably are the implementations and applications in Chapter 3. The python implementation, estimation method and analysis were all independent contributions. I cannot say with certainty that similar analysis has not been done previously - it probably has - but as I am not aware of any concrete examples.

CHAPTER 1

Risk Measures and Optimization

1.1 Notation and preliminaries

The purpose of this chapter is to introduce an appropriate framework to analyse the problems we will face the following sections. We therefore begin by recalling some basic mathematical theory that will be used throughout. We will also make particular choices regarding the spaces we work with. If it is not otherwise noted, we will assume that the choices made in this chapter apply everywhere in the text.

Measure theory, L^p-spaces and random variables

The notions of measure spaces and measure theory are of key importance for modern probability theory, and hence for the analysis of random variables.

Definition 1.1.1 (σ-algebra)**.** Let N be a countable index-set, and let a set Ω be given. A family $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if the following holds.

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Longrightarrow A^C \in \mathcal{F}$
- $\forall n \in N, n \geq\ 1 A_n \in \mathcal{F} \Longrightarrow \cup A_n \in \mathcal{F}$,

where A^C denotes the complement of A, relative to Ω.

As a direct consequence of closeness under countable unions and complements, it follows that a σ-algebra $\mathcal{F}$ also contains the empty set $\emptyset$, and that it is closed under countable intersections. σ-algebras will be used to represent the outcomes of "experiments", where random variables map outcomes of such experiments onto real numbers.

Definition 1.1.2 (Probability measure)**.** Let $\mathcal{F}$ be a sigma algebra on Ω. A function $P\colon \mathcal{F} \longrightarrow [0,1]$ is called a probability measure if the following holds.

- $P(\emptyset) = 0,\ P(\Omega) = 1$
- If $\forall n \geq 1\ A_n \in \mathcal{F}$ with $A_l \cap A_k$ when $k \neq l$, then $P(\cup_n A_n) = \sum_n P(A_n)$.

Definition 1.1.3 (Probability space)**.** If Ω is a set, $\mathcal{F}$ a σ-algebra on Ω and $P\colon \mathcal{F} \to [0,1]$ is a probability measure, the triple $(\Omega, \mathcal{F}, P)$ is called a probability

space. If we replaced the probability measure P with a general measure μ, the triple $(\Omega, \mathcal{F}, \mu)$ would simply be called a measurable space.

In this context, a random variable X is then a measurable function $(\Omega, \mathcal{F}, P) \mapsto (\mathbb{R}, \mathbb{B}(\mathbb{R}), \lambda)$, where $\mathbb{B}(\mathbb{R})$ is the Borel-σ-algebra on $\mathbb{R}$. Moreover, to say that this function is measurable is to require that for each Borel-set $B \in \mathbb{B}(\mathbb{R})$ it holds that $\{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}$, which will be denoted $\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{F}$.

Definition 1.1.4 (L^p-space)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $1 \leq p < \infty$. The space $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ consists of all measurable functions $f \colon \Omega \longrightarrow \mathbb{R}$ such that

$$\int_\Omega |f|^p \, d\mu < \infty. \tag{1.1}$$

If $f, g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ where $\mu(\{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}) = 0$, we say that $f = g$ almost everywhere (a.e.). By identifying a.e-equal functions in equivalence classes, we define $L^p(\Omega, \mathcal{F}, \mu)$ to be the space of such equivalence classes of functions in $\mathcal{L}^p$.

Definition 1.1.5 (p-norm)**.** Let $1 \leq p < \infty$ be given. For a measurable function f, the p-norm of f is defined as

$$\|f\|_p \coloneqq \left(\int_\Omega \|f\|^p \, d\mu \right)^{\frac{1}{p}} \tag{1.2}$$

Finding a space that accommodates all the needs for a particular problem can often prove difficult, and making such a choice will often involve some compromise. We have made the choice to work with the space $L^2(\Omega, \mathcal{F}, P)$, and it will be implicitly assumed that random variables belong there unless mentioned otherwise. We believe that this is a reasonable choice, as it fits well with the modelling of the problems at hand, while maintaining the benefits provided by L^2, as we will denote this space when there is no danger of confusion. We next state two important inequalities from analysis.

Proposition 1.1.6 (Minkowski's inequality [MW12])**.** *Let $1 \leq p \leq \infty$. Then,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \tag{1.3}$$

for all $f, g \in L^p$.

Proposition 1.1.7 (Hölder's inequality [MW12])**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $p, q \in [1, \infty)$ where $\frac{1}{p} + \frac{1}{q} = 1$. Then, for all measurable functions f, g on Ω,*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{1.4}$$

As a consequence of Hölder's inequality, we have that

$$L^1(\Omega, \mathcal{F}, \mu) \subset L^2(\Omega, \mathcal{F}, \mu). \tag{1.5}$$

This follows by applying the inequality with $p = q = 2$, and $g = \mathbf{1}_\Omega$. Here, $\mathbf{1}$ is the indicator function, which for a set $A \subseteq \Omega$ is defined as

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases} \tag{1.6}$$

Proposition 1.1.8. *A random variable $X \in L^2(\Omega, \mathcal{F}, P)$ has finite expectation and variance, denoted $E[X] \coloneqq \int_\Omega X \, dP$, $V[X] = \int_\Omega (X - E[X])^2 \, dP$, respectively, whenever these are finite numbers.*

Proof. The finiteness of $E[X]$ follows directly, since $X \in L^2 \implies X \in L^1$. Hence,

$$E[X] = \int_\Omega X \, dP \leq \int_\Omega |X| \, dP \leq \infty. \tag{1.7}$$

Moreover,

$$\begin{aligned} V[X] &= E\left[(X - E[X])^2\right] && (1.8)\\ &= \int_\Omega (X - E\,[X])^2 \, dP && (1.9)\\ &= \int_\Omega X^2 \, dP - 2E[X] \int_\Omega X \, dP + E[X]^2 \int_\Omega \mathbf{1}_\Omega \, dP && (1.10)\\ &\overset{(i)}{=} E[X^2] - E[X]^2 && (1.11)\\ &\overset{(ii)}{\leq} \infty, && (1.12) \end{aligned}$$

using in (i) that $P(\Omega) = 1$ and in (ii) that $X \in L^2$ ■

Convexity

The following results hold for a general vector space X.

Definition 1.1.9 (Convex set)**.** Let $S \subseteq X$ be a set. S is *convex* if for all $x, y \in S$,

$$\{\lambda x + (1 - \lambda)y \mid 0 \leq \lambda \leq 1\} \subseteq S. \tag{1.13}$$

If the above holds with strict inequalities for all $0 < \lambda < 1$, S is said to be *strictly convex*.

Visually, convex sets are those that contain the line segments connecting any two points in the set. Examples of convex sets

Definition 1.1.10 (Convex function)**.** Let $S \subset X$ be a convex set. A function $f : S \longrightarrow \mathbb{R}^+$ is *convex* if for all $x, y \in S$,

$$f(\lambda x + (1 - \lambda)y \leq \lambda f(x) + (1 - \lambda) f(y). \tag{1.14}$$

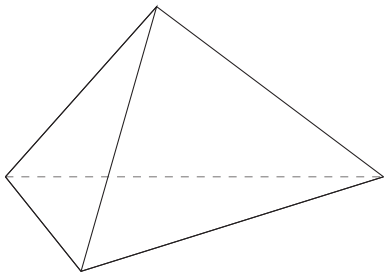
Strictly convex functions are defined analogously to strictly convex sets.

If a convex function $f : S \longrightarrow (\infty, +\infty]$ is not constantly equal to $+\infty$, it is a *proper convex function*.

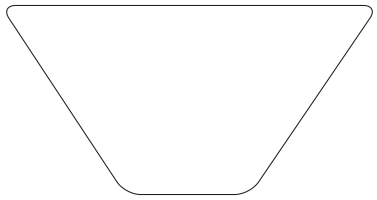
Convex functions on $\mathbb{R}$ appear as functions where the graph between two points is bounded above by the line connecting the function value at those points.

Definition 1.1.11 (Epigraph)**.** Let $f : S \subseteq X \to \mathbb{R}$ be a function. The epigraph of f, denoted epi$f$ is the set of all points that lie above the graph of f, and is defined as

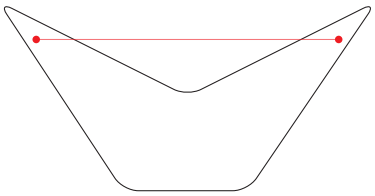
$$\text{epi} f \coloneqq \{(x, a) \mid x \in S, a \geq f(x)\} \tag{1.15}$$



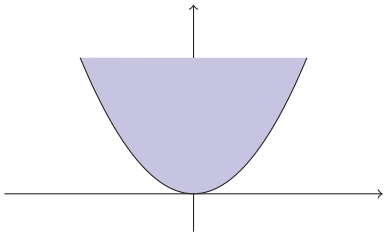
(a) A convex set in $\mathbb{R}^3$



(b) A convex set in $\mathbb{R}^2$



(c) A non-convex set in $\mathbb{R}^2$



(d) A convex function.

Figure 1.1: Example of convexity and non-convexity

Definition 1.1.12 (Lower semi-continuity)**.** Let $f : X \longrightarrow \mathbb{R}^+$ be a function and $x \in X$ a point. Then f is said to be *lower semi-continuous* at the point x if for each $y \in X$ there exists an open neighbourhood U containing y such that $f(x) > y$ for all $x \in U$. If f is lower semi-continuous at all points in its domain, it is said to be lower semicontinuous.

Figure 1.1 provides some examples of convexity and non-convexity. The shaded area in figure 1.1d indicates the epigraph of the function.

Definition 1.1.13. (Hyperplane, halfspace) A set $P \subseteq X$ is called a hyperplane in X if

$$P = \{x \in X \mid \langle x, a\rangle = c\}, \tag{1.16}$$

where $a \in X$ is a non-zero vector and $c \in \mathbb{R}$ is some scalar. Furhtermore, P divides X into two sets.

$$P^+ := \{x \in X \mid \langle x, a\rangle \geq c\}, \tag{1.17}$$
$$P^- := \{x \in X \mid \langle x, a\rangle \leq c\}. \tag{1.18}$$

These sets are called the upper and lower halfspace, respectively.

Definition 1.1.14 (Supporting hyperplane)**.** If $S \subseteq X$ is a non-empty set, a hyperplane P is called a *supporting hyperplane* to S if the following holds.

a) S is entirely contained in either P^+ or P^-

b) At least one point in S a boundary point of P.

Theorem 1.1.15 (Supporting hyperplane theorem)**.** *Let $S \subseteq X$ be a non-empty, convex set and $x \in X$ a point on its boundary. Then, there exists a supporting hyperplane P to S containing x.*

Proof. We refer to [BBV04, p 51] for the details. ■

Definition 1.1.16 (Level sets)**.** Let $f : \mathbb{V} \to \mathbb{R}$ be a functional defined on some set $\mathbb{V}$ and let $\alpha \in \mathbb{R}$. The α-level set of f is defined as

$$\{\mathbf{v} \in \mathbb{V} \mid g(\mathbf{x}, \mathbf{v}) \leq \alpha\} \tag{1.19}$$

Lemma 1.1.17. *A convex function has convex level sets*

✻✎ Let f be a convex function, and let $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ be in the α-level set for f for a fixed $\alpha \in \mathbb{R}$. Let $\lambda \in (0, 1)$. Then,

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{1.20}$$
$$\lambda\alpha + (1-\lambda)\alpha \tag{1.21}$$
$$= \alpha, \tag{1.22}$$

which shows that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is in the α-level set of f, which finishes the proof. ■

In comparing two random variables, we may be interested in comparing their expectations under convex transformations:

Definition 1.1.18 (Convex order)**.** Let X, Y be real valued random variables. X is said to be smaller than Y in convex order, provided

$$E[f(X)] \leq E[f(Y)], \tag{1.23}$$

fora all convex functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that the expectations exist. We denote this $X \leq_{\text{CX}} Y$.

While it is true that functions from $\mathbb{R}^n$ to $\mathbb{R}$ a positive semi-definite Hessian matrix are convex, convex functions are not limited to the class of such function. For example, the function $x \mapsto |x|$ is convex on $\mathbb{R}$, but not differentiable at the point $x = 0$.

Definition 1.1.19 (Subgradient, subdifferential)**.** Assume $f : X \subseteq \mathbb{R} \longrightarrow \mathbb{R}^+$ is a convex, lower semi-continuous function which is not constantly equal to $+\infty$. A vector $g \in S$ is said to be a *subgradient* of f at the point x_0 if for all $x \in S$

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle \tag{1.24}$$

Collecting all subgradients of f at a point x_0 gives the *subdifferential* of f at x_0, which we denote $\partial f(x_0)$. Formally,

$$\partial f(x_0) = \{g \in S \mid \forall x \in S \; f(x) \geq f(x_0) + \langle g, x - x_0 \rangle\}. \tag{1.25}$$

Example 1.1.20 (✻)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function with $\partial f(0) = [-1, 1]$. Then, $f(x)$ is bounded below by $|x|$. To see this, note that since $1 \in \partial f(0)$, the definition of a subgradient requires that

$$\forall x \in \mathbb{R}\ f(x) \geq x. \tag{1.26}$$

Similarly,

$$\forall x \in \mathbb{R}\ f(x) \geq -x. \tag{1.27}$$

Clearly, the smallest function which satisfies this is

$$f(x) = \begin{cases} x, & x \geq 0, \\ -x, & x < 0 \end{cases} \tag{1.28}$$

$$= |x|. \tag{1.29}$$

The reverse is also true. If $f(x) = |x|$, $\partial f(0) = [-1, 1]$. For all other $x \in \mathbb{R}$, $\partial f(x)$ is either $+1$ or -1, depending on the sign of x.

The subgradient g of f at a point x_0 where f is finite has a geometric interpretation. Namely that $h(z) = f(x_0) + \langle g, x - z\rangle$ is a non-vertical supporting hyperplane to the epigraph of f at the point $(x_0, f(x_0))$ [Roc70]. Figure 1.2 illustrates selected supporting hyperplanes for $f(x) = |x|$ to epi$f$ at the point $0, f(0)$ corresponding to elements in its subdifferential $\partial f(0)$.

Subdifferential may allow us to identify a global minimum of a non-differentiable function.

Proposition 1.1.21. *If f is a convex function with $\partial f(x^*) \neq \emptyset$ with $0 \in \partial f(x^*)$, then f attains a global minimum at x^*.*

✻✎ Assume f is a convex function with $\partial f(x^*) \neq \emptyset$ with $0 \in \partial f(x^*)$. Then,

$$0 \in \{g \in S \mid \forall x \in S\ f(x) \geq f(x^*) + \langle g, x - x_0\rangle\}. \tag{1.30}$$

This means that

$$\forall x \in S, \quad f(x) \geq f(x^*) + \langle 0, x - x_0\rangle\}, \tag{1.31}$$

which gives $\forall x \in S, \quad f(x) \geq f(x^*)$, which is what we wanted to show. ■

If we consider the case where $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a convex, differentiable function, f has exactly one supporting hyperplane at each point $x \in \mathbb{R}$, namely the tangent at the point $x, f(x)$. In this case

$$0 \in \partial f(x) \Leftrightarrow f'(x) = 0, \tag{1.32}$$

which we know to be sufficient for x to be a global minimum in this case.

Definition 1.1.22. Let f be a function with subdifferential ∂f. The graph of ∂f, denoted gph∂f is defined as follows.

$$\text{gph}\partial f = \{(x, p) \mid p \in \partial f(x)\} \tag{1.33}$$

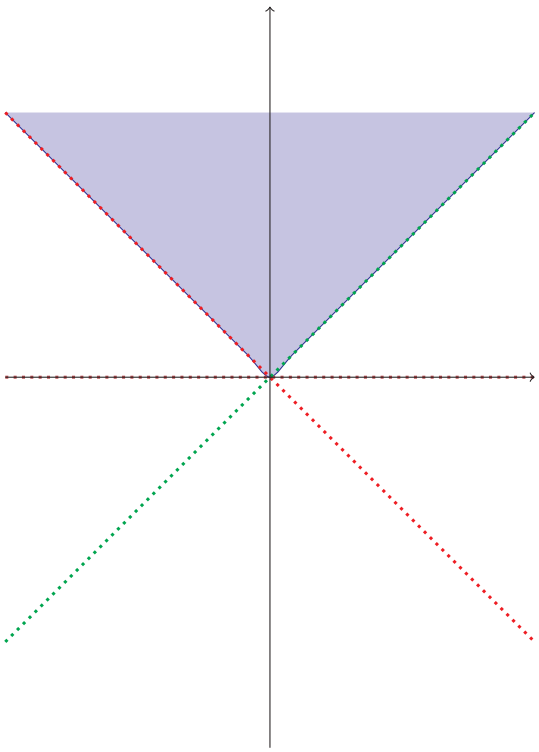


Figure 1.2: $f(x) = |x|$ and its epigraph with some supporting hyperplanes at $(0, f(0))$ corresponding to elements in $\partial f(0)$ indicated.

From [RR14, p. 12] we have that the subdifferential for a real valued, proper convex function f can be expressed in terms of it's right- and left derivatives. Indeed,

$$\partial f(x) = \begin{cases} \{p \in \mathbb{R} \mid f'^{-}(x) \leq p \leq f'^{+}(x)\}, & \text{for } x \in \text{ dom } f, \\ \emptyset & \text{otherwise} \end{cases} \tag{1.34}$$

We conclude this section by reviewing some results related to monotone relations and random variables. These results will later be used to derive results about particular risk measures, and they will help uncover a deeper connection between convex analysis and risk analysis. They are based largely on the results developed in [RR13].

Definition 1.1.23 (Monotone relations)**.** A set Γ of pairs $(x, p) \in \mathbb{R} \times \mathbb{R}$ is a *monotone relation* if

$$\text{for all } (x_1, p_1) \text{ and } (x_2, p_2) \in \Gamma, (x_1 - x_2)(p_1 - p_2) \geq 0 \tag{1.35}$$

whenever

$$(x_1, p_1) \leq (x_2, p_2) \text{ or } (x_1, p_1) \geq (x_2, p_2). \tag{1.36}$$

The inverse of a monotone relation Γ, denoted Γ^{-1} is defined as

$$\Gamma^{-1} := \{(p, x) \mid (x, p) \in \Gamma\}. \tag{1.37}$$

Visually, a monotone relation Γ can be viewed as a graph that tends monotonically from south-west to north-east without discontinuities, while the inverse relation Γ^{-1} appears as the reflection of Γ about the line $y = x$.

Definition 1.1.24 (Convex conjugate)**.** Let $f : X \to \mathbb{R}$ be a function, and X^* the dual space of X, with a (real-valued) bilinear mapping $\langle \cdot, \cdot \rangle : X \times X^* \longrightarrow \mathbb{R}$. The *convex conjugate* of f, denoted $f^* : X^* \longrightarrow \mathbb{R}$ is defined as

$$f^*(p) = \sup_{x \in X} \{\langle x, p \rangle\} \tag{1.38}$$

Particularly if $X = \mathbb{R}^n$, the convex conjugate of $f : X \longrightarrow \mathbb{R}$ at a point $p \in \mathbb{R}$ is

$$f^*(p) = \sup_{x \in \mathbb{R}^\kappa} \{\langle x, p \rangle - f(x) \mid p \in \mathbb{R}^n\}, \tag{1.39}$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product for vectors

The convex conjugate also relates to subdifferentials. The following result is from [RR14].

Proposition 1.1.25. *Let f be a convex function on $\mathbb{R}^+$ which is not constantly equal to $+\infty$, and let f^* be its convex conjugate. Then, the following holds.*

$$\partial f^* = (\partial f)^{-1}, \tag{1.40}$$

in the sense that

$$x \in \partial f^*(p) \Longleftrightarrow p \in \partial f(x). \tag{1.41}$$

While the facts regarding convex conjugates and monotone relations seem somewhat distant from the main topic of this thesis which is risk analysis, they will prove very efficient in helping us prove some important facts about risk measures.

1.2 Optimization

Problems within risk analysis often feature systems that involve uncertain quantities and potentially undesired consequences. The goal is typically to handle the uncertain system in a way that minimizes the risk of such consequences, or to maximize the systems utility while keeping the risk of consequences below at an acceptable level. In general terms, this will commonly result in a problem of optimizing some aspects of the system with respect to some measurement of the undesired resulting consequences. This section will briefly consider optimization in the deterministic case, and establish relevant results from convex analysis. Section 1.2 will couple these principles with random variables, where the notion of risk and risk measures get introduced. As will be revealed, convexity plays an integral part both to optimization and risk analysis.

Traditionally, constrained optimization aims to solve problems on the form

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c_i(x) \leq 0 \quad i = 1, \ldots, m, \end{aligned} \tag{1.42}$$

where $x \in S \subseteq \mathbb{R}^n$, and $f\colon S \to \mathbb{R}$ is a function that assigns to each choice of x a real-valued cost, and $c_i : S \longrightarrow \mathbb{R}$ are functions that represent constraints. We will follow the common practice of referring to the functions f and c_i as *the objective function* or sometimes simply *the objective*, and *the constraints* respectively. The goal is to identify x such that f is minimized while the constraints in equation (1.42) are not violated. An example of such a problem could be the design of a mechanical structure, where x represents choices of materials for the construction to which f assigns a cost. Then, c_i could represent specifications, e.g. of strength, that that the resulting construction needs to fulfill.If a point $x^* \in S$ satisfies the constraints in (1.42) and $f(x) \geq f(x^*)$ for all $x \in S$, we will say the problem has an *optimal value* of $f(x^*)$, and that x^* is an *optimal solution point*.

There are circumstances where we may want to transform a particular optimization problem into another form, and solve the transformed problem instead. This may be motivated by obtaining an alternative formulation of the problem that in some regard is easier to solve. Particularly, such reformulations can be of importance for algorithmic applications, where certain problem forms are better suited for algorithmic solving. We will informally say that two optimization problems equivalent if the solution of one provides a complete description of the solution of the other. The following proposition illustrates one such equivalent formulation.

Proposition 1.2.1. *Assume $x^* \in S \subseteq \mathbb{R}^n$ is an optimal solution to the optimization problem in* (1.42)*. Consider then the following optimization problem.*

$$\begin{aligned} &\underset{x \in X}{\textit{minimize}} \quad t \textit{ over all } (x,t) \in S \times \mathbb{R} \\ &\textit{subject to} \begin{cases} f(x) - x_{n-1} \leq 0, \\ c_i(x) \leq 0 \quad \textit{for } i = 1, \ldots, m. \end{cases} \end{aligned} \tag{1.43}$$

Then, an optimal solution (t^*, x^*) to (1.43) *provides a complete description of the optimal solution to* (1.42)*. Particularly, $t^* = f(x^*)$*

✻*Proof.* Assume that x^* is an optimal solution to (1.42), which gives an optimal value of $f(x^*)$. Consider the constraints

$$\begin{aligned} f(x) - t &\leq 0 \\ c_i(x) &\leq 0. \end{aligned} \tag{1.44}$$

For each x, t is bounded below by $f(x)$. Hence, t is minimal whenever $f(x)$ is minimal. Moreover, x^* satisfies (1.44). Hence, (1.43) has an optimal value of $t = f(x^*)$, attained at (t, x^*). For the opposite implication, assume instead that (t^*, x^*) is optimal for (1.44), and let

$$A = \{x \in X \mid c_i(x) \leq 0 \quad i = 1, 2, \ldots, n\}, \tag{1.45}$$

such that the constraints $c_i(x) \leq 0$ are satisfied on A. It is clear that for a given $x \in A$, $f(x)$ is a lower bound on t, which then attains its minimum at $t = f(x)$.

Since this is true for any $x \in A$, minimizing t coincides with minimizing $f(x)$ for $x \in A$. By assumption, x^* minimizes $f(x)$ for $x \in A$, so choosing

$$t^* = f(x^*), x \in A \tag{1.46}$$

is indeed optimal. This shows that when x^* is an optimal solution point to (1.42) with optimal value $f(x^*)$, then $(x^*, f(x^*))$ is an optimal solution point to (1.44), with optimal value $f(x^*)$.

■

This specific reformulation is sometimes called the *epigraphical form* of the optimization problem. The name stems from the fact that we have introduced an extra variable in the epigraph of f. Epigraphs play an important role in convex analysis, and we introduce them formally in the next section. As we will see in later chapters, this particular form of the problem plays an important role in the optimization with respect to certain risk measures.

If one has identified a candidate x^* as the optimal value for f, the question of whether better solutions exist and the uniqueness of x^* arises. As it turns out, both these properties can be obtained if f is *convex*.

Proposition 1.2.2. *Any local minimum x^* of a convex function f is a global minimum of f. If f is strictly convex, x^* is a unique.*

Proof. Let $f : A \to \mathbb{R}$ be a convex function, and assume for a contradiction that x^* minimizes f locally, but not globally. This means that there exists some $y^* \in A$ such that $f(y^*) > f(x^*)$. By convexity we have that for all $0 < \lambda < 1$,

$$\begin{aligned} f(\lambda x^* + (1-\lambda)y) &\leq \lambda f(x^*) + (1-\lambda)f(y) && (1.47)\\ &< f(x^*) + (1-\lambda)f(x^*) && (1.48)\\ &= f(x^*), && (1.49) \end{aligned}$$

which contradicts the assumption that x^* is a local minimum, because we can find points $\lambda x^* + (1-\lambda)y$ arbitrarily close to x^* where f takes smaller values. This cannot be the case, since x^* is a local minimum iff there exists some open ball with radius $r > 0$ where f does not take smaller values. For uniqueness of x^*, assume additionally that f is strictly convex, and that there exists some point $y^* \neq x^*$ such that $f(x^*) = f(y^*)$. Fix $\lambda \in (0, 1)$. By strict convexity we have that

$$\begin{aligned} f(\lambda x^*) + (1-\lambda)y^*) &< \lambda f(x^*) + (1-\lambda)f(y^*) && (1.50)\\ &= \lambda f(x^*) + (1-\lambda)f(x^*) && (1.51)\\ &= f(x^*), && (1.52) \end{aligned}$$

which is contradictory, and hence x^* is unique. ■

The previous result is of significant importance, both theoretically and for practical applications, and is among the motivating factors for the use of convex risk measures, which get introduced in Section 1.3. From a practical perspective, it can in certain cases reduce the computational efforts of solving an optimization problem by limiting it only to determine if given candidate solutions are locally optimal. In other instances where the optimization problem is such that gradient methods are applicable, it may suffice for the optimization scheme to identify a stationary point, which is then globally optimal.

Remark 1.2.3. We have so far only considered the optimization problem of minimizing a function. This is however not a restriction, as any optimization problem can be posed as a minimization problem of the negative function. Indeed,

$$\begin{aligned}
&\text{maximize} && g(x) && && (1.53)\\
&\text{subject to} && c_i(x) \leq 0 \quad i = 1, \ldots, m && && (1.54)\\
& && \Updownarrow && && (1.55)\\
&\underset{x \in X}{\text{minimize}} && -g(x) && && (1.56)\\
&\text{subject to} && c_i(x) \leq 0 \quad i = 1, \ldots, m. && && (1.57)
\end{aligned}$$

What's more, the problem constraints may safely be given in the above form, since any constraint $\widetilde{c_i}(x) \geq 0$ can be stated as $-\widetilde{c_i}(x) \leq 0$, resulting in the original problem statement.

Optimization under uncertainty

So far we have discussed constrained optimization of deterministic functions, but this does not let us capture problems of decision making and optimization when the objective function or the constraints are not certain, but depend on some state of the world. To get an intuitive understanding of what me mean by this, consider the following problem when deciding on the design of a mechanical structure, where $x \in X$ represents choices of materials for the construction to which f assigns a cost. Then, c_i could represent specifications, e.g. of material strength, which the resulting construction needs to fulfill. In a deterministic world, this is sensible, but it needs to be extended to capture the added difficulty of uncertainty. Instead, consider the functions $c_i(x, \omega)$, where ω represents possible future states of the construction's environment. With this formulation, $c_i(x, \omega)$ could represent the strength of a part of the construction given a choice of materials and a given state of its environment. As this view is adapted, $\underline{c}_i(x) \coloneqq c_i(x, \omega) : X \times \Omega \to \mathbb{R}$ can be regarded as a random variable.

To signify that we do not need to distinguish strictly between the objective function and the constraints in an optimization problem, we will often refer to the objective function as $\underline{c}_0(x) \coloneqq f(x, \omega)$. This is justified, as we sometimes may want to view the optimization from another angle, where a constraint can take the role of the objective and vice versa. For instance, in one formulation one might aim to minimize the cost under given a constraint on the weight, wile an alternative problem is to minimize the weight given a constraint on the cost.

In the following a brief account of some approaches to optimization under uncertainty discussed by Rockafellar in [Roc12] will be given. In all approaches, the aim is to characterize the random variables in the objective and the constraints to real numbers. By composing the random variables with functionals, we are left with deterministic optimiztion problems. The question then becomes: what are reasonable ways to attach real numbers to these random variables?

Approach 1: Guessing the Future

A simplistic approach is to chose $\omega' \in \Omega$ as a best estimate of the unknown information, and then to

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c_0\left(x, \omega'\right) \\ \text{subject to} \quad & c_i(x, \omega') \leq 0 \quad \text{for } i = 1, \ldots, m. \end{aligned} \tag{1.58}$$

There are clear drawbacks to this approach. Importantly, the solution can be very unstable even with respect to tiny changes in ω'. Additionally, ω' might not at all be a good estimate of the unknown information, and the probability of $c_i(x, \omega)$ being reasonably close to $c_i(x, \omega')$ can be very small.

Approach 2: Worst Case Analysis

Instead of fixing a best estimate in Ω, the worst case approach considers the constraints in Equation (1.42) under the worst possible circumstances in the following manner:

$$\text{minimize} \quad \sup_{\omega \in \Omega} c_0\left(x, \omega\right) \text{ over all } x \in X \tag{1.59}$$

$$\text{subject to} \quad \sup_{\omega \in \Omega} c_i(x, \omega) \leq 0 \quad \text{for } i = 1, \ldots, m. \tag{1.60}$$

It is true that this approach avoids many of the pitfalls of the previous approach, but there are immediate drawbacks to this approach too. Since we have chosen to consider L^2-random variables this approach will not be feasible if they are unbounded ones. This is, at least in theory, the case for many random variables in applications. While the risk of vastly underestimating the costs, or having the constraints violated are eliminated, this comes at a cost. The set of suitable ω might be very small, or even empty. Moreover, such a risk averse approach will be too conservative for many applications. Indeed, we are now only concerned with the absolutely worst possible states, without paying any attention to the performance under ordinary circumstances. In the case of the mechanical structure, it might not be possible to hedge against every possible storm or earthquake, or the cost for such a structure could become impossibly high.

Approach 3: Relying on Expectations

Passing to the expected value is another way of attaching real values to the random variables $c_i(x, \omega)$. The problem then becomes then to

$$\text{minimize} \quad E[\underline{c}_0(x)] \quad \text{over all } x \in X \tag{1.61}$$

$$\text{subject to} \quad E[\underline{c}_i(x)] \leq 0 \quad \text{for } i = 1, \ldots, m. \tag{1.62}$$

While this solves the problem of unbounded variables, a clear drawback to this approach is that it only requires the constraints to be satisfied on average. This can be hard to justify for many applications. In the case of our mechanical structure, it is not enough that it is acceptable only on the average. Additionally, by relying on the expected value, one does not place emphasis on the potentially large costs at the right tail of the distribution, and large costs can be mitigated

by anti-costs as long as the expected value is satisfactory. In short, if it is very important to maintain control over costs and constraints, we need a more risk averse approach.

Approach 4: Standard Deviation Units as Safety Margins

A way to improve on the expectation-approach is to introduce safety margins based on the standard deviation. We choose a set of positive numbers $\lambda_i > 0$, and define the constraints in the optimization problem as

$$\text{minimize} \quad E[\underline{c}_0(x)] + \lambda_0 \sigma(\underline{c}_0(x)), \quad \text{over all } x \in X \tag{1.63}$$

$$\text{subject to } E\left[\underline{c}_i(x)\right] + \lambda_i \sigma\left(\underline{c}_i(x)\right), \text{ for } i = 1, 2, \ldots, n, \tag{1.64}$$

where $\sigma(\underline{c}_i(x)) \coloneqq E\left[\left(\underline{c}_i(x) - E\left[\underline{c}_i(x)\right]\right)^2\right]^{\frac{1}{2}}$ is the standard deviation of $\underline{c}_i(x)$. This approach partially solves the problems of relying solely on the expectation. We are no longer content with $E[\underline{c}_i(x)] \leq 0$, but have also gained control over how likely it is that the constraints will be breached. The approach provides a more risk averse approach compared to the former. However, the point that anti-costs can offset costs still remains. Another drawback related to convexity will be revisited in the next section.

1.3 Risk measures

Much of the theory and literature regarding risk analysis has its origin in finance and financial mathematics. In finance, risk and its analysis have been of concern possibly as long as financial markets have existed, raising questions such as: "How much of a premium do you require to hold a commodity with a value that can possibly decrease over time", or "how much capital should be allocated to cover the position of a futures trade". While such questions are old, much of the literature in risk analysis is relatively recent. Many of the traditional approaches revolve around measurements of the deviations of the underlying asset. A unified framework for the analysis of financial risk was introduced in 1999 by Artzner et. al. [Art+99]. There, the authors gave a axiomatic definition of the *coherent risk measure*. This was done by imposing axioms that were thought to be natural conditions a sensible measure of risk should satisfy, according to financial understanding. In [FS02] it is suggested to relax the conditions of coherent risk measure slightly, which introduces the *convex risk measure*. In the following we will review some facts about coherent and convex risk measures. We will review the traditional approaches from the previous section in light of these risk measures.

To avoid confusion, we follow [RU+00] in pointing out a distinction between two concepts that are often associated with risk, namely uncertainty and cost. We will measure the uncertainty of a random variable as in terms of its deviation when assessing inconsistency, and its risk when assigning to it a single value as a representative of its cost. The term "cost" is general. Monetary losses, injuries, or contamination may all be examples of losses. Costs will take positive values, so greater positive costs are worse than smaller. When costs are below 0 they will be said to be *acceptable*. In this context, placing a large bet on a game with an almost certain, large negative outcome has very little uncertainty, since the

loss is almost certain, but it carries a great risk since the cost of the is likely large. Note that some texts about risk measures use an opposite orientation, where larger values of X are associated with positive gains and a lower levels of risk. We maintain the view that positive outcomes of X represent losses. There are no fundamental differences between these views, except from a change of signs.
The following definition of a coherent risk measure is from [RU+00].

Definition 1.3.1. $\mathcal{R} : L^2(\Omega, \mathcal{F}, P) \to (-\infty, \infty]$ is a *coherent measure of risk in the extended sense* if

(R1) (Constantness): $\mathcal{R}(C) = C$ for all constants C,

(R2) (Convexity): $\mathcal{R}((1-\lambda)X + \lambda X') \leq \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(X')$,

(R3) (Monotonicity): $\mathcal{R}(X) \leq \mathcal{R}(X')$ when $X \leq X'$,

(R4) (Closedness): $\mathcal{R}(X) \leq 0$ when $\|X_k - X\|_2 \to 0$ witxch $R(X_k) \leq 0$.

$\mathcal{R}$ is called a *coherent measure of risk in the basic sense* if additionally

(R5) (Homogeneity): $\mathcal{R}(\lambda X) = \lambda R(X)$ for $0 < \lambda \in \mathbb{R}$.

The coherent risk measure in the extended sense is equivalent to that of [Art+99], while the axioms are not the same. The latter chose axioms motivated by economic arguments, and required the risk measure to be translation invariant, i.e, that adding a risk-free gain to a risky asset resulted in an equal decrease in the risk. Moreover, Artzner et. al imposed *subadditivity*, formally

$$\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y). \tag{1.65}$$

Remark 1.3.2. Subadditivity together with $(R5)$(Homogeniety) gives $(R2)$(Convexity). Indeed, Let X, Y be random variables which satify subadditivity and $(R5)$. Then, for any $\lambda \in [0, 1]$,

$$\begin{aligned} \mathcal{R}(\lambda X + (1-\lambda)Y) &\leq \mathcal{R}(\lambda X) + \mathcal{R}((1-\lambda)Y) && (1.66)\\ &= \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(Y). && (1.67) \end{aligned}$$

From these axioms, several reasonable results can be gathered:

1. By combining $(R1)$ and $(R5)$, we get subadditivity:

$$\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X') \tag{1.68}$$

2. When X is bounded above,

$$\mathcal{R}(X) \leq \sup_{\omega \in \Omega} X. \tag{1.69}$$

3. The monotonicity property in (R3) is equivalent to the following.

$$X \leq 0 \implies \mathcal{R}(X) \leq 0. \tag{1.70}$$

Proof.

1. Let X and $X' \in L^2$, and define $Y = 2X$ and $Y' = 2X'$. Then,

$$\mathcal{R}(X + X') = \mathcal{R}\left(\frac{1}{2}Y + \frac{1}{2}Y\right) \tag{1.71}$$
$$\overset{(R2)}{\leq} \frac{1}{2}\mathcal{R}(Y) + \frac{1}{2}\mathcal{R}(Y') \tag{1.72}$$
$$\overset{(R5)}{=} \mathcal{R}\left(\frac{1}{2}Y\right) + \mathcal{R}\left(\frac{1}{2}Y'\right) \tag{1.73}$$
$$= \mathcal{R}(X) + \mathcal{R}(X'), \tag{1.74}$$

   which is what we wanted to show.

2. Let $X' = \sup_{\omega\in\Omega} X$. Appealing to (R3) gives

$$\mathcal{R}(X) \overset{(R3)}{\leq} \mathcal{R}(X') \tag{1.75}$$
$$= \mathcal{R}(\sup_{\omega\in\Omega} X) \tag{1.76}$$
$$\overset{(R1)}{=} \sup_{\omega\in\Omega} X, \tag{1.77}$$

   so the statement holds.

3. Assume that $(R3)$ holds, i.e,

$$X \leq Y \implies \mathcal{R}(X) \leq \mathcal{R}(Y). \tag{1.78}$$

   This means that if $Y \equiv 0$, we get that

$$\mathcal{R}(X) \leq \mathcal{R}(0) = 0. \tag{1.79}$$

   For the opposite implication, assume instead that

$$X \leq 0 \implies \mathcal{R}(X) \leq 0, \tag{1.80}$$

   and let $X' \leq Y'$. Define $Z = 0.5(X' - Y') \leq 0$. By the convexity in (R1) it follows that

$$0 \leq \mathcal{R}(Z) \leq 0.5\mathcal{R}(X') - 0.5\mathcal{R}(Y'), \tag{1.81}$$

   or equivalently $\mathcal{R}(X') \leq \mathcal{R}(Y')$, which finishes the proof.

In [FS02] the authors argue that scaling property that follows $(R5)$ is not necessarily reasonable in practice, pointing out that the risk associated with a financial position may increase non-linearly as the position becomes very large. They suggest instead to relax the conditions somewhat, introducing the *convex risk measure*.

Definition 1.3.3 (Convex Risk Measure)**.** A functional $\mathcal{R} : L^2(\Omega, \mathcal{F}, P)$ is a convex risk measure if

(R1') **(Convexity):** $\mathcal{R}((1-\lambda)X + \lambda X') \leq \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(X')$,

(R2') **(Monotonicity):** $\mathcal{R}(X) \leq \mathcal{R}(X')$ when $X \leq X'$,

(R3') **(Translation invariance):** If $m \in \mathbb{R}$, then $\mathcal{R}(X+m) = \mathcal{R}(X)+m$.

■

Proposition 1.3.4. *Under these axioms, any coherent risk measure in the extended sense is automatically a convex risk measure.*

✻✎ Let $\mathcal{R}$ be a coherent risk measure in the extended sense, and let X be a random variable and $m \in \mathbb{R}$ a constant. We must show that axioms $(R1')-(R3')$ hold.The axioms $(R2')$ and $(R3')$ follow by assumption, so we are left with deriving $(R3')$. We have that

$$\begin{aligned}\mathcal{R}(X) &= \mathcal{R}(\frac{1}{2}(2X+2m) - \frac{1}{2}2m) && (1.82)\\ &\overset{(R2)}{\leq} \mathcal{R}(X+m) - \mathcal{R}(m)\\ &\overset{(R1)}{=} \mathcal{R}(X+m) - m\\ &\leq \mathcal{R}(X) + \mathcal{R}(m) - m\\ &= \mathcal{R}(X). && (1.83)\end{aligned}$$

By rearranging the terms, we deduce from the inequalities

$$\mathcal{R}(X) \leq \mathcal{R}(X+m) - m \leq \mathcal{R}(X) \tag{1.84}$$

that

$$\mathcal{R}(X) = \mathcal{R}(X+m) - m \Longrightarrow \mathcal{R}(X+m) = \mathcal{R}(X) + m. \tag{1.85}$$

The result holds in the extended sense, since (1.82) also requires $(R5)$ to bring the scaling factor of $\frac{1}{2}$ outside the risk measure. This concluded the proof, and shows that the axioms for convex risk measures are in fact a relaxation of the axioms for coherent risk measure (in the extended sense). ■

Coherency in Optimization

With the axioms for risk measures established, a scheme for optimization under uncertainty analogous to the approaches from Section 1.3 is proposed. This was the topic of [Roc12]. In the following, we review some key points from this work, and elaborate on some of the details. In the previous discussion, several methods to associating a single value to the random variables $\underline{c}_i(x)$ were proposed. With the concept of risk measures in place, the idea is the following approach.

$$\begin{aligned}\text{minimize} \quad & \text{for } x \in X \mathcal{R}_0\left(\underline{c}_0(x)\right),\\ \text{subject to} \quad & \mathcal{R}_i\left(\underline{c}_i\left(x\right)\right) \leq 0 \quad i = 1, \ldots, m.\end{aligned} \tag{1.86}$$

Here, $\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_m$ are risk measures as proposed in Definition 1.3.1. To establish a theorem regarding important properties of coherent risk measures, the following lemma will be useful.

Lemma 1.3.5. *The composition of a convex function with a nondecreasing convex function is itself a convex function. In mathematical notation, suppose $f(x) \geq f(y)$ for all $x \geq y$, and suppose f, g are convex functions. Then, $f \circ g$ is a convex function.*

Proof. By convexity of g, $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$. Hence, since f is nondecreasing, and by applying the property of convexity of f in the last inequality, it follows that

$$f(g(\lambda x + (1-\lambda)y)) \leq f(\lambda g(x) + (1-\lambda)g(y)) \tag{1.87}$$
$$\leq \lambda f(g(x)) + (1-\lambda)f(g(y)) \tag{1.88}$$

■

With this in place, we can introduce the following theorem. This is Theorem 1.3.4 in [RU+00].

Theorem 1.3.6. *Suppose in problem (1.86) that for each $i = 0, 1, \ldots, m$, that each functional $\mathcal{R}_i$ is a coherent measure of risk in the extended sense. Then, the following properties hold:*

(a) ***Preservation of convexity:*** *If $c_i(x, \omega)$ is convex with respect to $x \in S$ for each $\omega \in \Omega$, then the function $\mathcal{R}_i(c_i(x))$ is convex.*

(b) ***Preservation of certainty:*** *If $\underline{c}_i$ is a constant random variable for each x, i.e., $c_i(x, \omega) = c_i(x)$ (i.e, no influence from $\omega \in \Omega$) , then $\mathcal{R}_i(\underline{c}_i(x)) = c_i(x)$.*

(c) ***Insensitivity to scaling:*** *If the risk measures $\mathcal{R}_i$ also satisfy (R5), then problem (1.86) remains the same when the units in which the values $c_i(x, \omega)$ are denominated are re-scaled.*

✻*Proof of Theorem 1.3.6.*

(a) Assume that $\underline{c}_i(x)$ is a convex function for each $\omega \in \Omega$. By (R1) and (R2) $\mathcal{R}_i$ is a non-decreasing, convex function. By Lemma 1.3.5, $\mathcal{R}_i(\underline{c}_i(x))$ is a convex function.

(b) For each x, $\underline{c}_i(x) \in \mathbb{R}$. By (R1) the conclusion follows.

(c) A rescaling of $\underline{c}_i$ is the transformation $\underline{c}_i(x) \mapsto \lambda \underline{c}_i(x)$ for some scalar $\lambda > 0$. Then, (R5) gives $R_i(\lambda \underline{c}_i(x)) = \lambda R_i(\underline{c}_i(x))$, and thus problem (1.86) remains the same.

■

Property (c) of 1.3.6 illustrates some of the difficulties in developing axioms for risk measures that are both sensible from an economical point of view and natural from a practical perspective. It is intuitive that the measured risk should not be perceived differently after a scaling by a change of units. At the same time, Föllmer et al. argue that the risk needs not increase linearly with the increase in a portfolio.

Traditional approaches to optimization under uncertainty revisited

We will now review the previous approaches from section 1.2 in light of coherent risk measures in the basic sense. As shown, coherence in this regard also implies convexity in the sense of definition 1.3.3.

Approach 1: Guessing the Future We assess the risk in $\underline{c}_i(x)$ as $\mathcal{R}(\underline{c}_i(x))$ by choosing some $\omega' \in \Omega$ with $P(\omega') > 0$:

$$\mathcal{R}(\underline{c}_i(x)) = c_i(x, \omega') \tag{1.89}$$

Defining $\mathcal{R}$ in this manner gives a coherent risk measure in the basic sense, but there are obvious drawbacks in terms of uncertainty. The risk is regarded as acceptable if it is acceptable in the state ω', while no other states are considered.

Proposition 1.3.7. *Fix $\omega' \in \Omega$. The risk measure $\mathcal{R}(\underline{c}_i(x)) = \underline{c}_i(x, \omega')$ is a coherent risk measure in the basic sense, according to Definition 1.3.1.*

✻*Proof.* We check that $\mathcal{R}$ as defined in Proposition 1.3.7 satisfies the axioms (R1) - (R5).

(R1) If $c_i(x, \omega')$ is a constant C, then $\mathcal{R}(c_i(x, \omega')) = C$.

(R2) Let $\lambda \in (0, 1)$. Then,

$$\begin{aligned} \mathcal{R}(\lambda \underline{c}_i(x) + (1-\lambda)\underline{c}_j(x)) &= \lambda c_i(x, \omega') + (1-\lambda) c_j(x, \omega') && (1.90)\\ &= \lambda \mathcal{R}(\underline{c}_i(x)) + (1-\lambda)\underline{c}_j(x). && (1.91) \end{aligned}$$

(R3) Assume $\underline{c}_i(x) \leq \underline{c}_j(x)$. Then,

$$\begin{aligned} \mathcal{R}(\underline{c}_i(x)) &= \underline{c}_i(x, \omega') && (1.92)\\ &\leq \underline{c}_j(x, \omega') && (1.93)\\ &= \mathcal{R}(\underline{c}_j(x)). && (1.94) \end{aligned}$$

(R4) Let $(\underline{c}_i(x))_i$ be a sequence of random variables converging in L^2-norm to $\underline{c}(x)$, with $\mathcal{R}(\underline{c}_i(x)) \leq 0$. Then $c_i(x, \omega') \leq 0$ for every i. Let's for convenience denote $c_i(x, \omega') = c_i$ and $c(x, \omega') = c$ for their scalar values, and assume for a contradiction that $c > 0$, which implies

$$\mathcal{R}(\underline{c}(x)) > 0. \tag{1.95}$$

Then, since

$$\|c_i - c\|_2 \to 0, \tag{1.96}$$

it follows that

$$\|c_i - c\|_1 = |c_i - c| \to 0, \tag{1.97}$$

since c_i and c are constants. Therefore, there exists for every $\epsilon > 0$ some $N \in \mathbb{N}$ such that $|c_n - c| < \epsilon$ for all $n \geq N$. Fix $\epsilon = \frac{c}{2}$, and observe that

$$|c_n - c| \leq |c| > \epsilon. \tag{1.98}$$

(R5) Let $\lambda \in \mathbb{R}$. Then,

$$\mathcal{R}(\lambda \underline{c}_i(x)) = \lambda \underline{c}_i(x)(\omega') = \lambda \mathcal{R}(\underline{c}_i(x)). \tag{1.99}$$

This shows that $\mathcal{R}$ is a coherent risk measure. ■

Approach 2: worst-case analysis The approach is to assess the risk in $\underline{c}_i(x)$ as $\mathcal{R}(\underline{c}_i(x))$ as $\sup \underline{c}_i(x)$, provided $\underline{c}_i(x)$ is bounded. Defining $\mathcal{R}$ in this way gives a coherent risk measure in the basic sense, but the same flaws as in the similar traditional approach remains.

Proposition 1.3.8. *Let $\underline{c}_i(x)$ be a bounded random variable, and let $\mathcal{R}(\underline{c}_i(x)) = \sup \underline{c}_i(x)$. This gives a coherent risk measure in the basic sense, according to definition 1.3.1.*

✻✎ Let $\underline{c}_i(x)$ and R be defined as above. The result follows by the same steps as in the proof of proposition 1.3.7 by replacing the constant $c_i(x, \omega)$ there with $\sup \underline{c}_i(x)$. ■

Approach 3: Relying on Expectation The approach is to rely on expectation by letting $\mathcal{R}(\underline{c}_i(x)) = E[\underline{c}_i(x)]$.

Proposition 1.3.9. *The risk measure $\mathcal{R}(\underline{c}_i(x)) \coloneqq E[\underline{c}_i(x)]$ is a coherent risk measure in the basic sense.*

✻*Proof.* We check that the risk measure as defined in Proposition 1.3.9 satisfies $(R1) - (R5)$.

(R1) If $\underline{c}_i(x) = C$ for some constant C, $\mathcal{R}(\underline{c}_i(x)) = E[C] = C$.

(R2) By linearity of expectations it follows that

$$\begin{aligned} \mathcal{R}(\lambda \underline{c}_i(x) + (1-\lambda)\underline{c}_j(x)) &= E[\lambda \underline{c}_i(x) + (1-\lambda)\underline{c}_j(x)] && (1.100)\\ &\lambda E[\underline{c}_i(x)] + (1-\lambda)E[\underline{c}_j(x)] && (1.101)\\ &\lambda \mathcal{R}(\underline{c}_i(x)) + (1-\lambda)\mathcal{R}(\underline{c}_j(x)). && (1.102) \end{aligned}$$

(R3) Let $\underline{c}_i(x) \leq \underline{c}_j(x)$. Then $\mathcal{R}(\underline{c}_i(x)) = E[\underline{c}_i(x)] \leq E[\underline{c}_j(x)] = \mathcal{R}(\underline{c}_j(x))$.

(R4) Assume $\|X_k - X\|_2 \to 0$ when $k \to \infty$ and $\mathcal{R}(x_k) \leq 0$ for each k. Then, the following holds.

$$\begin{aligned} 0 &\leq \|X_k - X\|_1 && (1.103)\\ &= \|(X_k - X) \cdot 1\|_1 && (1.104)\\ &= E[|X_k - X|] && (1.105)\\ &\overset{(i)}{\leq} \|X_k - X\|_2 \|1\|_2 && (1.106)\\ &= \|X_k - X\|_2, && (1.107) \end{aligned}$$

by using Hölder's inequality with p = q = 2 in (i). Since $\|X_k - X\|_2 \to 0$ as $k \to \infty$, it follows that $E[|X_k - X|] \to 0$ as $k \to \infty$. Finally, observe that

$$0 \leq |E[X_k] - E[X]| \tag{1.108}$$

$$\overset{(ii)}{\leq} |E[X_k - X]| \tag{1.109}$$

$$\overset{(iii)}{\leq} E[|X_k - X|] \to 0, \tag{1.110}$$

using linearity of the expectation in (i), Jensen's inequality in (ii) and the limit from above in the final step. This shows that $|E[X_k] - E[X]| \to 0$ as $k \to \infty$, and so, by assumption,

$$\mathcal{R}(X) = \lim_{k\to\infty} E[X_k] \leq 0. \tag{1.111}$$

(R5) By linearity of expectations, it follows that

$$\mathcal{R}(\lambda \underline{c}_i(x)) = E[\lambda \underline{c}_i(x)] \tag{1.112}$$

$$= \lambda E[\underline{c}_i(x)] \tag{1.113}$$

$$= \lambda \mathcal{R}(\underline{c}_i(x)). \tag{1.114}$$

This is what we wanted to show. ■

Approach 4: Standard Deviation as Safety Margins The approach is to use $\mathcal{R}(\underline{c}_i(x)) = E[\underline{c}_i(x)] + \lambda_i \sigma(\underline{c}_i(x))$ to assess the risk in $\underline{c}_i(x)$. While this approach seems intuitive and properly risk averse, it unfortunately does not constitute a coherent measure of risk, either in the basic nor extended sense, since monotonicity, $(R3)$ in Definition 1.3.1, does not hold in general. It suffices to find a counter-example.

Example 1.3.10 (✻ Illustrates lack of monotonicity)**.** Let $Y \equiv 0$, and let X be a non-constant random variable taking values in $(-1, 0)$. This means that

$$-1 < E[X] = \mu_X < 0 \text{ and } \sigma(X) = \sigma_X > 0. \tag{1.115}$$

On the other hand, $E[Y] = \sigma(Y) = 0$. Then, it is not hard to see that one can choose a suitable λ' such that

$$E[X] + \lambda' \sigma(X) = \mu_X + \lambda' \sigma_X > 0. \tag{1.116}$$

Defining $\mathcal{R}(\cdot) = E[\cdot] + \lambda' \sigma(\cdot)$ gives $\mathcal{R}(X) > 0$. On the other hand, $\mathcal{R}(Y) = 0$, which violates $(R3)$ In fact, the risk measure indicates that within this safety margin, the random variable X carries a positive risk. This is a surprising result in the assessment of a random variable that is guaranteed never to produce a costly outcome. This behaviour underlines the importance of distinguishing between the variability and the risk associated with a random variable. A random variable with great variability does in general not need to have great risk.

Remark 1.3.11 (✻). Let $\mathcal{R}$, defined as in the worst-case approach and ρ, defined as in the expected value-approach, be two risk measures. Then, $\mathcal{R}(X) \geq \rho(X)$ for any bounded random variable X.

Proof. Indeed, observe that for any $\omega \in \Omega$ we have $X(\omega) \leq \sup_{\omega \in \Omega} X(\omega)$, and hence

$$\rho(X) = \int_\Omega X \, dP \tag{1.117}$$

$$\leq \int_\Omega \sup_{\omega \in \Omega} X(\omega) \, dP \tag{1.118}$$

$$= \sup_{\omega \in \Omega} X(\omega) \tag{1.119}$$

$$= \mathcal{R}(X). \tag{1.120}$$

This is what we wanted to show. ■

As all but approach 4 produced a coherent risk measure in the basic sense, they also satisfy the axioms for a convex risk measure. However, their drawbacks were many and sometimes severe. This suggests that it is not in itself enough for a risk measure to be coherent or convex to be a sensible risk measure. What is sensible will vary from situation to situation, usually depending strongly of the tolerance for costs, which can require us to impose more restrictions than the axioms. One natural restriction is to require, in addition to the axioms of coherency or convexity, that $E[X] \leq \mathcal{R}(X)$ which gives an *averse risk measure*. If the inequality is strict, we say $\mathcal{R}$ is strictly averse.

Risk measure as probability of compliance An alternative to the approaches reviewed above is to measure risk in terms of the probability of exceeding a certain cost level. This approach will correspond with a risk measure that is widely used within the financial industry, known as *Value-at-Risk*. Recall that the *cumulative distribution function* for X, denoted F_X is given by

$$F_X(c) = P(X^{-1}((-\infty, c]). \tag{1.121}$$

This is well defined, since $(-\infty, c]$ is a Borel-set and X is measurable. Related to the cumulative distribution function is the *survival function*. It is denoted S_X, and defined by

$$S_X(c) = 1 - F_X(c). \tag{1.122}$$

We describe the *quantile* of X in terms of F_X.

Definition 1.3.12 (Quantile of a random variable)**.** Let $\alpha \in (0, 1)$. The α-quantile of X, denoted $q_\alpha(X)$ is defined as

$$q_\alpha(X) = \inf\{c \in \mathbb{R} \mid \alpha \leq F_X(c)\}. \tag{1.123}$$

The quantile of the random variable identifies the smallest value the random variable can take with probability less than or equal to α. If X represents costs, the α quantile identifies the level of cost such that any cost larger than or equal to this level occurs with a probability of α or less. For technical reasons, the α-quantile is expressed as an infimum. If Ω contains some ω with $P(\omega) > 0$, as is the case with e.g. discrete probability distributions and importantly empirical distributions, the quantile remains well defined. For such distributions the

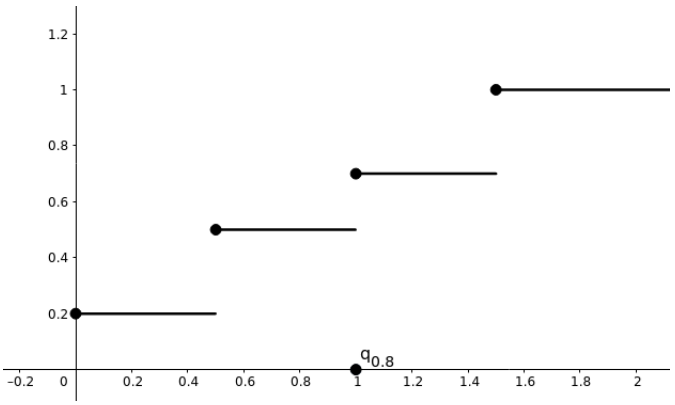


Figure 1.3: The 0.8-quantile of a random variable with piecewise cumulative distribution function. In this case it takes the value 1.

cumulative distribution function will have flat parts and points of discontinuity. In these cases, the α-quantile corresponds to selecting the left endpoint of an interval, on which the cumulative distribution function has equal values. Figure Figure 1.3 illustrates the 0.8-quantile of a random variable with a piecewise cumulative distribution function.

The idea is to use the α-quantile of X as a measure of risk. Therefore, let

$$\mathcal{R}_i(\underline{c}_i(x)) = q_{\alpha_i}(\underline{c}_i(x)). \tag{1.124}$$

By defining $\mathcal{R}_i$ this way, one is essentially assessing the risk in $\underline{c}_i(x)$ as the smallest cost that occurs with a probability of $1-\alpha$ or less.

Example 1.3.13 (✻Illustrates q_α as a measure of risk)**.** Let X be a random variable that represents the total net loss in a financial portfolio after some time. If $q_{0.95}(X) = 1\text{M\$}$, there is a probability of 5% or less that the loss will exceed 1M$

Using q_α as a measure of risk might seems compelling. In fact, q_α coincides the widely used Value-at-Risk (VaR). Nevertheless, there are some clear drawbacks to this approach. Indeed, by using $\mathcal{R}(\underline{c}_i(x)) = q_{\alpha_i}(\underline{c}_i(x))$, one neglects the very worst outcomes of $\underline{c}_i(x)$. For L^2-random variables this upper part of the tail could be very long, and the costs could become infinite even when $q_\alpha(X)$ is small for a reasonable α. Additionally, $\mathcal{R}$ is not a coherent risk measure, either in the basic nor the extended sense. Specifically, $\mathcal{R}$ is not convex ($R2$).

Proposition 1.3.14. *The risk measure $\mathcal{R}(X) = q_\alpha(X)$ is not a coherent risk measure. Is is also not a convex risk measure.*

✻*Proof.* It suffices to find a counterexample. Example 1.3.15 is one such counterexample. ■

Example 1.3.15 (✻)**.** Illustrates non-convexity of $q_\alpha(X)$] Assume there are two independent, financial portfolios. Both portfolios have a 95% probability of giving a 1M$ profit, and a 5% probability of a loss of 18M$. Let X and Y be

random variables that represent the loss in each portfolio respectively. I.e. X and Y can take the values -1M\$ or 18M\$. Suppose $\mathcal{R} = q_{0.93}$ is chosen to be the measure the risk. A combination portfolio $Z = 0.5X + 0.5Y$ is created. Then,

$$q_{\alpha_{0.93}}(X) = q_{\alpha_{0.93}}(Y) = -1\text{M\$},$$

while the combined portfolio $Z = 0.5X + 0.5Y$ has $q_{0.93}(Z) = 8.5$M\$ To summarize,

$$0.5\mathcal{R}(X) + 0.5\mathcal{R}(Y) \leq \mathcal{R}(0.5X + 0.5Y).$$

That is, the risk in the combined portfolio exceeds the sum of risk in each of the portfolios. This shows that a diversification of the portfolio does no reduce, but on the contrary, increases the total risk. This is not desirable behaviour of a risk measure, and is in violation with $(R2)$. Therefore, the α-quantile does in general not give a coherent measure of risk.

Under certain restrictions, $q_\alpha(X)$ remains a coherent risk measure, as the next proposition shows.

Proposition 1.3.16. *Let $X, Y \in L^2$ be independent, normally distributed random variables, $X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right), Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right)$. Let as before $\mathcal{R}$ be given by the α-quantile at level $\alpha \in (0,1)$. Then $\mathcal{R}$ is a coherent measure of risk.*

✻*Proof.* $(R1)$: Assume $X = c$ for a constant $c \in \mathbb{R}$. For each $\alpha \in (0,1)$, $q_\alpha(X) = c$. Hence $\mathcal{R}(X) = c$.

$(R2)$: Let Φ denote the cumulative distribution function of a normally distributed random variable with expectation 0 and variance 1. By a change of variables, one can easily show that $F_X(x) = \Phi\left(\frac{x-\mu_x}{\sigma_X}\right)$. Additionally, Φ is a monotonically increasing function, and is therefore invertible. Let Φ^{-1} denote its inverse. Finding $q_\alpha(X)$ is therefore reduced to finding c such that $\alpha = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right)$, or alternatively

$$q_\alpha(X) = \Phi^{-1}(\alpha)\,\sigma_X - \mu_X. \tag{1.125}$$

Note that when α is fixed, $\Phi^{-1}(\alpha)$ is a constant. Let

$$Z = \lambda X + (1-\lambda)Y,\ \lambda \in (0,1). \tag{1.126}$$

It is well-known that a linear combination of independent normally random variables again is a normally distributed random variable. Indeed,

$$Z \sim \mathcal{N}\left(\lambda\mu_X + (1-\lambda)\mu_Y, \lambda^2\sigma_X^2 + (1-\lambda)^2\sigma_Y^2\right). \tag{1.127}$$

Using (1.125), an expression for $\mathcal{R}(Z)$ is derived:

$$\begin{aligned}\mathcal{R}(Z) &= q_\alpha(Z)\\ &= \Phi^{-1}(\alpha)\sqrt{\lambda^2\sigma_X^2 + (1-\lambda)^2\sigma_Y^2}\\ &\quad - (\lambda\mu_X + (1-\lambda)\mu_Y).\end{aligned} \tag{1.128}$$

On the other hand,

$$\begin{aligned}\lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(Y) &= \lambda q_\alpha(X) + (1-\lambda) q_\alpha(Y) \\ &= \Phi^{-1}(\alpha)\left(\lambda\sigma_X + (1-\lambda)\sigma_Y\right) \\ &\quad - \left(\lambda\mu_X + (1-\lambda)\mu_Y\right).\end{aligned} \tag{1.129}$$

Comparing the expressions in (1.128) and (1.129) after subtracting equal terms and dividing by $\Phi^{-1}(\alpha)$ gives the inequality

$$\sqrt{\lambda^2\sigma_X^2 + (1-\lambda)^2\sigma_Y^2} \leq \left(\lambda\sigma_X + (1-\lambda)\sigma_Y\right), \tag{1.130}$$

which holds. Hence,

$$\mathcal{R}(\lambda X + (1-\lambda)Y) \leq \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(Y). \tag{1.131}$$

$(R3)$: Assume $X \leq Y$. Observe first that for every $c \in \mathbb{R}$,

$$\{\omega \in \Omega \mid Y(\omega) \in (-\infty, c]\} \subseteq \{\omega \in \Omega \mid X(\omega) \in (-\infty, c]\}, \tag{1.132}$$

which implies that

$$P\left(Y^{-1}(-\infty, c]\}\right) \leq P\left(X^{-1}(-\infty, c]\}\right), \tag{1.133}$$

by monotonicity of P. Therefore, if

$$q_\alpha(X) = \arg\inf_{c\in\mathbb{R}}\{\alpha \leq P\left(X^{-1}(-\infty, c]\right)\}, \tag{1.134}$$

it follows that

$$P\left(X^{-1}(-\infty, q_\alpha(X)]\right) \geq P\left(Y^{-1}(-\infty, q_\alpha(X)]\right) \geq \alpha. \tag{1.135}$$

This means that

$$q_\alpha(X) \geq \arg\inf_{c\in\mathbb{R}}\{\alpha \leq P\left(Y^{-1}(-\infty, c]\right)\}, \tag{1.136}$$

which shows that $\mathcal{R}(X) \leq \mathcal{R}(Y)$.

$(R4)$: Assume that X_k is a sequence of normally distributed random with $\mathcal{R}(X_k) = q_\alpha(X_k) \leq 0$ for each k and $\|X_k - X\|_2 \to 0$ as $k \to \infty$.
We will show that $R(X) \leq 0$ holds in three steps:

1. We show that $E[X_k] \longrightarrow E[X]$,
2. We show that $V[X_k] \longrightarrow V[X_k]$,
3. We finally use the above results in combination with equation (1.125) to show that $\mathcal{R}(X) \leq 0$.

The result in the first step is established by using Hölder's inequality and Jensen's inequality as in the proof of $(R4)$ in Prop. 1.3.9.
We show that step two holds as follows. By assumption, $E[(X_k - X)^2]^{1/2} \to 0$, which means that $E[(X_k - X)^2] \to 0$. Moreover, by Minkowski's inequality inequality it follows that

$$|\|X_k\|_2 - \|X\|_2| \leq \|X_k - X\|_2 \to 0, \tag{1.137}$$

which implies that

$$\|X_k\|_2 - \|X\|_2 \to 0, \tag{1.138}$$

so $E[X_k^2] \to E[X^2]$. We use the fact that $V[X] = E[X^2] - E[X]^2$ and apply the above results. This gives

$$\begin{aligned} \lim_{k\to\infty} V[X_k] &= \lim_{k\to\infty} \left(E[X_k^2] - E[X_k]^2\right) && (1.139)\\ &= E[X^2] - E[X]^2 && (1.140)\\ &= V[X]. && (1.141) \end{aligned}$$

For the third step, recall from (1.125) that

$$q_\alpha(X) = \Phi^{-1}(\alpha)\,\sigma_X - \mu_X. \tag{1.142}$$

Then,

$$\lim_{k\to\infty} |\mathcal{R}(X) - \mathcal{R}(X_k)| = \lim_{k\to\infty} |\Phi^{-1}(\alpha)\,(\sigma_X - \sigma_{X_k}) - (\mu_X - \mu_{X_k})| = 0, \tag{1.143}$$

which shows that $\mathcal{R}(X) \leq 0$.

$(R5)$: Let $Y = \lambda X$ and $\mathcal{R}(X) = q_\alpha(X)$.

$$\begin{aligned} \mathcal{R}(Y) &= \arg\inf_{c\in\mathbb{R}} \left\{\alpha \leq P\left(Y^{-1}(-\infty, c]\right)\right\} && (1.144)\\ &= \arg\inf_{c\in\mathbb{R}} \left\{\alpha \leq P\left(X^{-1}\left(-\infty, \frac{c}{\lambda}\right]\right)\right\} && (1.145)\\ &= \arg\inf_{\lambda s} \left\{\alpha \leq P\left(X^{-1}(-\infty, s]\right)\right\}, && (1.146) \end{aligned}$$

where the change of variable $s = \frac{c}{r}$ was used in the final equality. The conclusion $\mathcal{R}(Y) = \mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ follows. ■

Notably, the assumption of normally, independently distributed random variables is only needed in $(R2)$. This is indeed the only axiom that $q_\alpha(X)$ does not satisfy in the general case. As the result shows, the risk measure $q_\alpha(X)$ is indeed coherent and thus convex under certain restrictions. The restriction of the random variables to the class of independent and normally distributed ones, is not necessarily justifiable. When $\underline{c}_i(x)$ is parametrized by x, $\underline{c}_i(x)$ must belong to this class for each choice of x. In fact, the restrictions on $\underline{c}_i(x)$ can be relaxed further. If $\underline{c}_i(x)$ has a *elliptical* probability distribution, $q_\alpha(X)$ remains coherent and hence convex, see [EMS02], Theorem 1.

Approach 6: Safeguarding with Value-at-Risk

Let $\mathcal{R}_i(\underline{c}_i(x)) = q_{\alpha_i}(\underline{c}_i(x)) = q_{\alpha_i}(\underline{c}_i(x))$ for $\alpha_i \in (0,1),\ i = 0, 1, \ldots, m$. The approach is to

$$\begin{array}{ll} \text{minimize} & \mathrm{VaR}_{\alpha_0}(\underline{c}_0(x)) \qquad\quad \text{over all } x \in X, \\ \text{subject to} & \mathrm{VaR}_{\alpha_i}(\underline{c}_i(x)) \leq 0 \quad \text{for } i = 1, \ldots, m. \end{array} \tag{1.147}$$

In this approach, one insists that for each of the constraints, $\underline{c}_i(x)$ is acceptable with at least probability α_i. Then, within the set of feasible x, the smallest cost c_0 is identified such that $\underline{c}_0(x) \leq c_0$ with at least probability α_0. We will later show that many of the shortcomings of VaR as a risk measure can be resolved by extending the quantile-approach to considering the expected loss in the tail of the distribution.

Conditional Value-at-Risk

Conditional Value-at-Risk (CVaR), also known as *average value-at-risk* or *expected shortfall* is a measure of risk which can be expressed as the expectation of the random variable in the conditional distribution of its upper α-tail. If the probability distribution of a random variable $X \in L^2$ is continuous, then the above description of CVaR at level α coincides with with the expectation of X on the interval $[q_\alpha, \infty)$. If on the other hand X does not have a continuous distribution, we must consider the "the upper α-tail" of X more carefully, since this may not be the interval $[q_\alpha, \infty)$. Recall that if for an $\omega' \in \Omega$, $X(\omega') = q_\alpha$ with $P(\omega') > 0$, F_X will have a jump at ω'. In that case, $P\left(X^{-1}\left([q_\alpha, \infty)\right)\right)$ might not be $1 - \alpha$.

Example 1.3.17 (✻ Illustrates discontinuities in q_α)**.** Consider the finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_5\}$ with $P(\omega_i) = 0.2$ for $i = 1, \ldots, 5$, and the random variable X defined by the mapping

$$\omega_i \mapsto i, \text{ for } i = 1, \ldots, 5. \tag{1.148}$$

Then, $q_{0.85}(X) = \inf\{c \in \mathbb{R} \mid 0.85 \leq F_X(c)\} = 5$, but

$$P\left(X^{-1}[5, \infty)\right) = 0.2 \neq (1 - 0.85) = 0.15. \tag{1.149}$$

As the example shows, when calculating CVaR_α, we cannot simply use $(1 - \alpha)$ as a rescaling factor. The following definition from [Ace02] remains consistent.

The following defintion is from [RU+00, p. 12]

Definition 1.3.18 (CVaR)**.** The CVaR at level α, denoted $\bar{q}_\alpha(X)$

$$\bar{q}_\alpha(x) \coloneqq (X) = \frac{1}{1 - \alpha} \int_\alpha^1 q_\beta(X)\, d\beta. \tag{1.150}$$

The following proposition lists some key properties of CVaR. The property in (a) is found in [RU02b], while The result from (b) is given in [RU00], [RU02a].

Proposition 1.3.19.

(a) $\bar{q}(X) \geq q_\alpha(X)$.

(b) *CVaR can be expressed as the following minimization problem.*

$$CVaR_\alpha(X) = \inf_{C\in\mathbb{R}} \left\{ C + \frac{1}{1-\alpha} E\left[\max\left\{X - C, 0\right\}\right] \right\}. \tag{1.151}$$

(c) *The expression in* (1.151) *also calculates VaR.*

$$VaR_\alpha(X) = \arg\inf_{C\in\mathbb{R}} \left\{ C + \frac{1}{1-\alpha} E\left[\max\left\{X - C, 0\right\}\right] \right\}. \tag{1.152}$$

The formulation of CVaR in b is especially potent. It reduces the problem of calculating CVaR_α which simultaneously calculates the VaR_α.

✻*Proof.*

(a) Since $q_\alpha(X)$ is non-decreasing in α, it follows that

$$\begin{aligned}
\bar{q}_\alpha(X) &= \frac{1}{1-\alpha}\int_\alpha^1 q_\beta(X)d\beta && (1.153)\\
&\geq \frac{1}{1-\alpha}\int_\alpha^1 q_\alpha(X)d\beta && (1.154)\\
&= \frac{1}{1-\alpha}(1-\alpha)q_\alpha && (1.155)\\
&= \text{VaR}_\alpha(X). && (1.156)
\end{aligned}$$

for (c) we refer to [RU00] [RU02a] for the details. In what follows, we will use convex duality theory to show (b). ■

We will establish some helpful results, in particular Theorem 1 and 2, from [RR14] to deduce the minimization formula in (b) from 1.3.19. To this end, we introduce the following function.

Definition 1.3.20 (Superexpectation)**.** Let X be a random variable and $c \in \mathbb{R}$ a number. The *superexpectation* of X, denoted $E_X(c)$, is defined as

$$E_X(c) = \int_\Omega [X - c]^+ \, dP, \tag{1.157}$$

where $[X - c]^+ \coloneqq \max(X - c, 0)$.

We also need the following observation.

Remark 1.3.21. Let X be a random variable. In the case where X has a continuous distribution, the expectation of X can be calculated by means of its quantile function. Observe that for a probability α,

$$\alpha = P(\{(-\infty, x]\} = F_X(x), \tag{1.158}$$

so a change of variable gives $d\alpha = f(x)dx$. In this case, $q_\alpha(X) = F_X^{-1}(\alpha)$, giving

$$\begin{aligned}
q_\alpha(F(x)) &= q_\alpha(X) && (1.159)\\
x &= q_\alpha. && (1.160)
\end{aligned}$$

Using the last expression, we get that

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx. \tag{1.161}$$

$$\tag{1.162}$$

Using then the change of variable $f(x)dx = d\alpha$ and $x = q_\alpha$ shows that

$$E[X] = \int_0^1 q_\alpha(X)\ d\alpha. \tag{1.163}$$

In the case where X does not have a continuous distribution, a similar result can be achieved by the means of approximating the integral by a limiting sum.

We also need the notion of the right-sided quantile and the lef-continuous distribution function.

Definition 1.3.22. Let the right-sided quantile be defined as

$$q_\alpha^+(X) \coloneqq \lim_{\alpha' \searrow \alpha} q_{\alpha'}(X), \tag{1.164}$$

and

$$F_X^-(x) \coloneqq \lim_{x' \nearrow x} F_X(x'). \tag{1.165}$$

Recalling the definition of monotone relations, we define the following relations, introduced in [RR13].

Definition 1.3.23. Let F_X be the cumulative distribution function of X and E_X its superexpectation. Then,

$$\Gamma = \{(x,p) \in \mathbb{R} \times \mathbb{R} \mid F_X^-(x) \leq p \leq F_X(x)\}, \tag{1.166}$$

$$\Delta_X = \{(p,x) \in \mathbb{R} \times \mathbb{R} \mid q_p(X) \leq x \leq q_p(X^+)\} \tag{1.167}$$

In fact, these are maximal monotone relations, and it holds that $\Gamma_X = \Delta_X^{-1}$ and conversely, see the article for the full details. They are maximal in the sense that they cannot be enlarged without breaking the condition for a monotone relation.

We now introduce two main theorems from [RR14]. They will allow us to deduce the convexity and closeness of CVaR. The result is a part of Theorem 1 in [RR14]

Theorem 1.3.24. *The superexpectation function E_X for a random variable X with $E[|X|] < \infty$ if a nondecreasing finite convex function on $\mathbb{R}$ which has the following properties:*

1.3.24.1. $\Gamma_X = gph\ \partial E_X$.

The following is Theorem 2 in [RR14].

Theorem 1.3.25 (Dualization of superexpectations)**.** *The closed, proper convex function E_X^* on $(-\infty, \infty)$ that is conugate to the superexpectation function E_X*

on $(-\infty, \infty)$ *is given by*

$$E_X^*(p) = \begin{cases} -(1-p)\bar{q}_p(X) & \text{for } p \in (0,1), \\ -E[X] & \text{for } p = 0, \\ 0, & p = 1, \\ \infty, & \text{for } p \notin [0,1]. \end{cases} \tag{1.168}$$

It is continuous relative to [0, 1], entailing

$$\lim_{p \nearrow 1}(1-p)\bar{q}_p(X), \quad \lim_{p \searrow 0} \bar{q}_p(X) = E[X], \tag{1.169}$$

and it corresponds subdifferentially to the monotone relation $\Delta_X = \Gamma_X^{-1}$ *and the quantile function* $q_p(X)$ *through*

$$\Gamma_x = gph, \quad Q_p(X) = E_X^{*'-}(p). \tag{1.170}$$

On the other hand, any function g *on* $(-\infty, \infty)$ *that is finite convex and continuous on* $[0,1]$ *with* $g(1) = 0$ *but* $g(0) = \infty$ *for* $p \notin [0,1]$ *is* E_X^* *for some random variable* X*.*

We review the key ideas of the proof, and refer to [RR14, p.19] for the full details.

Proof. Let g be a function that equals the right hand side of (1.168). The goal is to show that g is proper convex and closed with the property $g^* = E_X^*$. Then, $g = (g^*)^* = E_X$ will follow, as the double conjugate of a closed, convex function is the function itself. The continuity on $[0,1]$ is then seen since $g(p) = \frac{-(1-p)}{1-p}\int_p^1 q_\beta(X)\, d\beta$, getting $\lim_{p\nearrow 1} g(p) = 0$ and $\lim_{p\searrow 0} 0 = -E[X]$ and using the fact that closed convex, proper functions are continuous on intervals where they are finite, [Roc70, p. 82]. Moreover, $g(p)$ is convex as $q_p(X)$ is non-decreasing on $(0,1)$, and the extension to $+\infty$ when $p \notin [0,1]$ means that it remains a closed convex, proper function.

Since g is a left-continuous function, $g'^-(p) = q_p(X)$, while $g'^+(p) = q_p^+(X)$. These right- and left-derivatives will in turn be used to relate gph∂g to ΔX in light of 1.34. In the meantime, the one-sided derivatives $g'^+(p)$ and $g'^-(p)$ are the functions $q_p^+(X)$ and $q_p(X)$. This means that gph ∂g is the relation Δ_X. Then, g^* has the relation $\delta_X^{-1} = \Gamma_X$ by proposition 1.1.22. From Theorem 1.3.24 we know that Γ_X is in fact the graph of ∂E_X, which shows that ∂E_X has the same monotone relation as g^* as its graph. This establishes that g is equal to E_X^* up to a constant. By evaluating $E_X^*(1)$ which coincides with $g(1) = 0$, we get that E_X^* indeed equals the right hand side function. ∎

In conclusion, the theorem has shown that for $p \in (0,1)$,

$$\bar{q}_\alpha(X) = -E_X^*(\alpha)\frac{1}{1-\alpha}, \tag{1.171}$$

where $E_X^*(p)$ is a closed, proper convex function. Moreover, for $\alpha \in (0,1)$ we get that by inserting the expression for convex conjugates from 1.1.24 that

$$\bar{q}_\alpha(X) = -\frac{1}{1-p} E_X^*(\alpha) \tag{1.172}$$

$$= -\frac{1}{1-\alpha} \sup_x \{xp - E_X(x)\} \tag{1.173}$$

$$= \inf_x \{x + \frac{1}{1-\alpha} E[X-x]^+\}. \tag{1.174}$$

From this we conclude that the minimization formula from (b) holds.

The following is Theorem 2 in [Roc12].

Theorem 1.3.26. *For any probability level $\alpha \in (0,1)$, the functional $\mathcal{R}(X) = \bar{q}_\alpha(X)$ is a coherent measure of risk in the basic sense.*

Proof. For axioms $(R2), (R4)$ we refer to [RU+00], Theorem 2. For the remaining axioms, we proceed directly. (R1): Let $X \equiv C \in \mathbb{R}$. Then,

$$\mathcal{R}(X) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(X)\, d\beta \tag{1.175}$$

$$= \frac{1}{1-\alpha} \int_\alpha^1 C\, d\beta \tag{1.176}$$

$$= C. \tag{1.177}$$

(R3): Let X, Y be random variables such that $X < Y$. Then,

$$\mathcal{R}(X) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(X)\, d\beta \tag{1.178}$$

$$\leq \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(Y)\, d\beta \tag{1.179}$$

$$= \mathcal{R}(Y). \tag{1.180}$$

(R5): Let X, Y be random variables, and let $\lambda \in \mathbb{R}$ be a constant. Then,

$$\mathcal{R}(\lambda X) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(\lambda X)\, d\beta \tag{1.181}$$

$$= \frac{\lambda}{1-\alpha} \int_\alpha^1 q_\beta(X)\, d\beta \tag{1.182}$$

$$= \lambda \mathcal{R}(X). \tag{1.183}$$

This shows that CVaR is a coherent risk measure in the basic sense. In turn, this shows that it is also a convex risk measure. ■

Penalizing with the L^p-norm

Recall the risk measure $X \mapsto E[X]$ which had the drawback of lacking in aversity for many applications. We will now present a method to extend this measure in a natural way, overcoming in part the problem with a lack of risk aversity. Recall that, for $p \in [1, \infty)$ the L^p-norm of $X \in L^p$ is $\|X\|_p = E\left[|X|^p\right]^{1/p}$, and

if $p = \infty$ we define $\|X\|_\infty = \sup_{\omega\in\Omega}\{|X(\omega)|\}$. Keeping in mind that we are restricted to random variables in L^2, the L^p-norm is not a norm on L^2 in general, since fineness is not guaranteed. It proves however to be a useful tool as a penalty term, helping to increase the risk aversity of certain risk measures. Let $\lambda \in [0, 1]$, and define

$$\mathcal{R}(X) \coloneqq E[X] + \lambda \| \max_{\omega\in\Omega}\{X - E[X], 0\}\|_p. \tag{1.184}$$

It is immediately obvious that $\mathcal{R}$ is more risk averse than the expected value, since $\| \cdot \|_p \geq 0$. Moreover, by adjusting λ and p, one can regulate the emphasis on the costs of X. On the other hand, the expectation-term can relax the view of the risk associated with positive costs of X if the expected outcome is desirable, while if $p = \infty, \mathcal{R}$ is a type of worst case-analysis.

Example 1.3.27 (✻)**.** Let $\lambda = 1$ and $p = \infty$. Let $X \in L^2$ be a bounded random variable with negative expectation which also takes positive values almost surely. In other words, X has a good outcome on average, but with a positive probability of costs. Noting the $\sup_{\omega\in\Omega} X(\omega) \geq E[X]$ gives that

$$\begin{aligned}
\mathcal{R}(X) &= E[X] + \| \max_{\omega\in\Omega}\{X - E[X], 0\}\|_p && (1.185)\\
&= E[X] + \sup_{\omega\in\Omega}\{X(\omega) - E[X]\} && (1.186)\\
&= \sup_{\omega\in\Omega}\{X(\omega)\}. && (1.187)
\end{aligned}$$

Example 1.3.28 (✻)**.** Let $\lambda = 1$ and $p = 1$. Assume $X \in L^2$ is a continuous random variable with cumulative distribution function F and probability density function f, and denote by $\mathcal{M}$ the set $\{\omega \in \Omega \mid X(\omega) \geq E[X]\}$. Using $\mathcal{R}(X) = E[X] + \| \max_{\omega\in\Omega}\{X - E[X], 0\}\|_1$ yields

$$\begin{aligned}
\mathcal{R}(X) &= E[X] + E\left[\max_{\omega\in\Omega}\{X - E[X], 0\}\right] && (1.188)\\
&= E[X] + \int_{\mathcal{M}} X(\omega) - E[X]\ dP && (1.189)\\
&= E[X] - E[X]P(\mathcal{M}) + \int_{E[X]}^{\infty} xf(x)\ dx && (1.190)\\
&= E[X] - E[X](1 - F(E[X]) + \int_{E[X]}^{\infty} xf(x)\ dx && (1.191)\\
&= E[X]F\left(E[X]\right) + \int_{E[X]}^{\infty} xf(x)\ dx. && (1.192)\\
& && (1.193)
\end{aligned}$$

If additionally X has a symmetrical distribution, the expression simplifies further to

$$\mathcal{R}(X) = \frac{1}{2}E[X] + \int_{E[X]}^{\infty} xf(x)\ dx, \tag{1.194}$$

since $F\left(E[X]\right) = \frac{1}{2}$, which means that the risk is evaluated as half of the expected value plus the expectation of X restricted to the tail part of the

distribution, where X takes larger values than its expectation. Furthermore, if X is normally distributed with mean μ and variance σ^2, (1.194) simplifies further, to

$$\mathcal{R}(X) = \mu + \frac{\sigma}{\sqrt{2\pi}}.$$

Since $\sigma > 0$, this approach is more risk averse than relying only on μ.

Corollary 1.3.29 (✻Aversity of $\mathcal{R}$)**.** *Let $\mathcal{R}$ be defined as in* (1.184)*. Then, $\mathcal{R}(X) \geq E[X]$, regardless of the distribution of X.*

Proof. For any choice of $\lambda \in [0, 1]$ and $p \in [1, \infty]$,

$$\lambda \| \max_{\omega \in \Omega} \{X(\omega) - E[X]\}, 0\|_p \geq 0,$$

which shows that $\mathcal{R}(X) \geq E[X]$ ■

In this section we have studied the property of various risk measures, and discussed a coherent approach to optimization under risk. We have reviewed drawbacks and advantages with different risk measures, and we have seen that conditional Value-at-Risk offers multiple attractive qualities. Particularly, it is convex, which tells us that diversification leads to a reduction in risk, and it is sensitive to extreme outcomes in the tail end of the risk distribution. In the next chapter, we will review how Value-at-Risk and conditional Value-at-Risk can be applied to optimization of reinsurance contracts, in the interest of risk reduction.

CHAPTER 2

Optimal reassurance contracts under conditional Value-at-Risk

2.1 Introduction

In this chapter we consider the optimization of reinsurance contracts, with respect to the various risk measures composed with the the reinsurance contract. In general, insurance is purchased in order to reduce the risk of large losses, where the insurer will cover certain losses. The party that hedges against such losses will normally pay a risk premium for this protection, typically a fraction of the expected value of the insurer's cost. The same practice is common among insurance companies. When an insurance company buys insurance for a part of an insurance contract from another company, this is known as reinsurance.

In the following chapter, we refer to the party which hedges against losses by purchasing reinsurance as the *cedent*. The cedent purchases reinsurance from the *reinsurer*. Let $X_1, X_2, \ldots, X_m$ be m positive, random variables which represent the losses (risks in [HC20]) which the cedent wishes to insure. These risks are assumed to have continuous and monotonically increasing cumulative distributions functions. In the interest of risk reduction, the cedent is willing to pay a premium to have a second insurance company cover parts of these risks. Such reassurance contracts can be structured in several ways. When the reinsurer covers all losses within an interval, it is known as an *insurance layer contract*, and the covering of all losses exceeding a given threshold is known as a *stop-loss contract*. In the context of an insurance layer, let a_i, b_i be real numbers, such that any loss in the interval $[a_i, b_i]$ for risk number i is covered by the reinsurer. Then, let I_i denote the loss borne by the cedent for risk number i and R_i be the risk covered by the reinsurer for the same contract. This means that

$$I_i(X_i) \coloneqq \begin{cases} X_i, & \text{for } X_i < a_i \\ a_i, & \text{for } a_i \leq X_i < b_i \\ X_i - (b_i - a_i), & \text{for } X_i > b_i, \end{cases} \tag{2.1}$$

$$R_i(X_i) \coloneqq \begin{cases} 0, & \text{for } X_i < a_i \\ X_i - a_i, & \text{for } a_i \leq X_i < b_i \\ b_i - a_i, & \text{for } X_i > b_i. \end{cases} \tag{2.2}$$

The premium payed by the cedent to the reinsurer is $(1+\theta)E[R_i(X_i)]$, where $\theta > 0$ is a risk premium. We will denote this quantity π_i

2.2 Optimal reinsurance in the univariate case

In [Che+14], Cheung et al. considered the optimizaion of reinsurance contracts under general law-invariant convex risk measures, and in particular for conditional Value-at-Risk, which is a special case of a law-invariant convex risk measure [Sha13, p. 143], where a general class of reinsurance contract structures was considered. There, the optimization was solved for a single risk X. The authors found that, under mild assumptions, a stop-loss contract was found to be optimal under law-invariant convex risk measures ([Che+14] Proposition 3.1), while an insurance layer contract was optimal under Value-at-Risk ([Che+14] Theorem 4.3). Moreover, explicit expressions for the optimal contract parameters were found. We repeat these results in the following theorems, adapted to our notation. In particular, notice that the value-at-risk in [Che+14, p.73] is defined in terms of $1-\alpha$, where it is used that the Value-at-Risk at level $1-\alpha$ is given by

$$V@R_\alpha(Y) = \inf_m\{m \mid S_Y(m) \leq \alpha\}. \tag{2.3}$$

In contrast, this text uses that

$$\begin{aligned} VaR_\alpha(X) &= q_\alpha(X) && (2.4)\\ &= \inf\{c \in \mathbb{R} \mid \alpha \leq F_X(c)\} && (2.5)\\ &= \inf\{c \in \mathbb{R} \mid 1-\alpha \geq S_X(c)\}. && (2.6) \end{aligned}$$

Thus, $V@R_\alpha(X) = VaR_{1-\alpha}(X)$.

Theorem 2.2.1. *Let I be a stop-loss reinsurance contract with parameter a and risk premium θ for the risk X (corresponding to $b = +\infty$ in Equations (2.1) and (2.2).) Assume that $\alpha \in (0,1)$ and $1-\alpha < (1+\theta)^{-1}$, where S_X is the survival function of X. Then,*

$$a^* = S_X^{-1}((1+\theta)) \tag{2.7}$$

is an optimal contract parameter for the optimization problem

$$\min_a CVaR_\alpha\left[I(X) + (1+\theta)E[R(X)]\right]. \tag{2.8}$$

We elaborate on the details of the proof from [Che+14].

✻*Proof.* Let $\alpha \in (0,1)$ be fixed. We consider two cases. Case 1: $a \leq q_\alpha(X)$, case 2: $a > q_\alpha(X)$. In case 1, note that

$$\begin{aligned} & CVaR_\alpha(I(X)) && (2.9)\\ &= \frac{1}{1-\alpha}\int_\alpha^1 q_\beta I((X))\ d\beta && (2.10)\\ &= \frac{1}{1-\alpha}\int_\alpha^1 a\ d\beta && (2.11) \end{aligned}$$

$$= a \tag{2.12}$$

while

$$(1+\theta)E[R(X)] = (1+\theta)\int_{-\infty}^{\infty} R(x)f(x)\ dx \tag{2.13}$$

$$= (1+\theta)\int_{a}^{\sup X} (x-a)f(x)\ dx. \tag{2.14}$$

Combining the expressions from Equations (2.12) and (2.14) we find that

$$\frac{d}{da}\left[CVaR_\alpha(I(X)) + (1+\theta)\int_a^{\sup X}(x-a)f(x)dx\right] \tag{2.15}$$

$$= 1(1+\theta)(-af(a) + af(a) - \int_a^{\sup X} f(x)dx) \tag{2.16}$$

$$= 1-(1+\theta)P(X \geq a). \tag{2.17}$$

In order to minimize the problem in Equation (2.8), we set the final expression in Equation (2.17) to 0:

$$\begin{aligned} 1-(1+\theta)P(X \geq a) &= 0 \\ P(X > a) &= (1+\theta)^{-1} \\ a &= S_X^{-1}((1+\theta)^{-1}). \end{aligned}$$

We now consider case 2: Note first that a change of variable yields

$$\int_\alpha^1 q_\beta\,(I)\ \ d\beta = \int_{q_\alpha}^{\sup X} I(x)dF_X(x) \tag{2.18}$$

$$= \int_{q_\alpha}^{\sup X} I(x)f(x)\ dx, \tag{2.19}$$

where F_X and f are the cumulative distribution function and the probability density function of X. In light of this,

$$CVaR_\alpha(I(X)) = \frac{1}{1-\alpha}\int_{q_\alpha}^{\sup X} I(x)dF_x(x) \tag{2.20}$$

$$= \frac{1}{1-\alpha}\int_{q_\alpha}^{\sup X} I(x)f(x)\ dx \tag{2.21}$$

$$= \frac{1}{1-\alpha}\left(\int_{q_\alpha}^{a} I(x)f(x)dx + \int_a^{\sup X} af(x)dx\right) \tag{2.22}$$

since $a > q_\alpha$ by assumption. Differentiation with respect to a gives:

$$\frac{d}{da}\left[\frac{1}{1-\alpha}\left(\int_{q_\alpha}^{a} I(x)f(x)dx + \int_{a}^{\sup X} af(x)dx\right)\right] \tag{2.23}$$

$$= \frac{1}{1-\alpha}\left(I(a)f(a) - af(a) + \int_{a}^{\sup X} f(x)dx\right) \tag{2.24}$$

$$= \frac{P(X>a)}{1-\alpha}, \tag{2.25}$$

using the fundamental theorem of calculus. Meanwhile, $\frac{d}{da}(1+\theta)E[R(X)] = (1+\theta)P(X > a)$ remains unchanged from Equation (2.17). From this we conclude that in case 2,

$$\frac{d}{da}\left[CVaR_\alpha(I(X)) + (1+\theta)E[R(X)]\right] \tag{2.26}$$

$$= \frac{P(X>a)}{1-\alpha} - (1+\theta)P(X>a) \tag{2.27}$$

$$= S_X(a)\left(-(1+\theta) + \frac{1}{1-\alpha}\right), \tag{2.28}$$

which is positive when $(1-\alpha) > (1+\theta)^{-1}$. This shows that if $(1-\alpha) > (1+\theta)^{-1}$,

$$a = S_X^{-1}((1+\theta)^{-1}), \tag{2.29}$$

which finishes the proof. ■

The next theorem corresponds to Theorem 4.3 in [Che+14], adapted to our notation.

Theorem 2.2.2. *Assume that $\alpha \in (0,1)$. The following statements concering the optimal solution for the VaR-minimization of*

$$\min_{a,b} VaR_\alpha\left(I + (1+\theta)\pi\right). \tag{2.30}$$

are true:

- *If $1-\alpha < (1+\theta)^{-1}$, then an insurance layer contract with $a = S_X^{-1}((1+\theta)^{-1}), b = S_X^{-1}(1-\alpha)$ are optimal contract parameters.*

- *If $1-\alpha \geq (1+\theta)^{-1}$, then having no reinsurance is optimal.*

Proof. See [Che+14, p. 85]. ■

We remark that the assumption that X is bounded is reasonable from a insurance perspective. Indeed, it is not easy to imagine situations in which in infinite financial loss occurs. Moreover, in the context of reinsurance, where the cedent's risk X represents the liability for a policy, such a contract will in most cases have a maximum coverage, in which case X is bounded by the structure of the policy. In the following example, we illustrate the previous result of optimal parameter a, using Monte Carlo sampling from a theoretical distribution. We could modify the sampling to accommodate the wish of boundedness e.g by

asserting that the no value is allowed to exceed a certain number of standard deviations, or that no value exceeds the largest value expected in a sample of size N. However, the conclusion for optimal parameter a could well be generalized to the unbounded case, by replacing $\sup X$ with ∞ in the previous calculations. The following example illustrates an optimal stop-loss reinsurance contract with respect to CVaR.

Example 2.2.3 (✻)**.** Let X be a log-normally distributed random variable with

$$E[X] = 50, SD[X] = 5. \tag{2.31}$$

Let $\alpha = 0.95$ be fixed, and the risk premium be $\theta = \frac{1}{9}$. Using Monte Carlo sampling, a sample of 10^5 observations of X was generated. Figure 2.1 illustrates how the risk loading on the cedent, $CVaR_{0.95}(I(X) + (1+\theta)E[R(X)])$, varies with a. From the generated sample

$$S_X^{-1}((1+\theta)^{-1}) \approx 43.80 \tag{2.32}$$

was estimated, using the empirical survival function of X. The corresponding risk value was 50.90.If $a = \infty$ were used, in which case $I(X) \equiv X$, $R(X) \equiv 0$, no part of the risk was reinsured. In this case, the conditional Value-at-Risk at .95-level is

$$\begin{aligned} CVaR_{0.95}(I(X) + (1+\theta)E[R(X)]) &= CVaR_{0.95}(X) && (2.33)\\ &= 61.09 && (2.34) \end{aligned}$$

For this risk, using a stop-loss reinsurance as descirbed above reduces the risk by 16.68% compared to having no reinsurance.

Lemma 2.2.4. *If the reinsurance contract is an insurance layer contract, i.e, I, R defined as in equations (2.1) and (2.2), the optimal value of the parameter a remains unchanged.*

✻*Proof.* We repeat the arguments from the proof of Theorem 2.2.1, where we consider two cases in turn. Case $1 : a \leq q_\alpha(X)$, case $2 : a > q_\alpha(X)$.
Case 1: Assume $a \leq q_\alpha(X)$. Then,

$$\begin{aligned} &\frac{d}{da} CVaR_\alpha(I(X)) && (2.35)\\ &= \frac{d}{da}\left[\frac{1}{1-\alpha}\int_\alpha^1 q_\beta(I(X))\, d\beta\right] && (2.36)\\ &= \frac{1}{1-\alpha}\frac{d}{da}\left[\int_{q_\alpha}^b \overbrace{I(x)}^{(i)} f(x)\, dx + \int_b^{\sup X} \overbrace{I(x)}^{(ii)} f(x)\, dx\right] && (2.37)\\ &= \frac{1}{1-\alpha}\left[\int_{q_\alpha}^b f(x)\, dx + \int_b^{\sup X} f(x)\, dx\right] && (2.38)\\ &= 1, && (2.39) \end{aligned}$$

where we used in (i) that $I(x) = a$ when $x \in [a, b)$ and in (ii) that $I(x) = x - (b - a)$ when $x > b$.

Case 2: Assume instead that $a \leq q_\alpha(X)$.

$$\frac{d}{da} CVaR_\alpha(I(X)) \tag{2.40}$$
$$= \frac{1}{1-\alpha}\frac{d}{da}\left[\int_{q_\alpha}^{a} I(x)f(x)\ dx\ + \int_a^b af(x)\ dx + \int_b^{\sup X} I(x)f(x)\ dx\right] \tag{2.41}$$
$$= \frac{1}{1-\alpha}\left[I(a)f(a) + \int_a^b f(x)\ dx - af(a) + \int_b^{\sup X} f(x)\ dx\right] \tag{2.42}$$
$$= \frac{1}{1-\alpha}\int_a^{\sup X} f(x)\ dx \tag{2.43}$$
$$= \frac{P(X>a)}{1-\alpha}. \tag{2.44}$$

Meanwhile, the derivative of the risk premium, $\frac{d}{da}E[R(X)]$ remains unchanged:

$$\frac{d}{da}\left[E[R(X)]\right] = \frac{d}{da}\left[\int_a^b (x-a)f(x)\ dx + \int_b^{\sup X}(b-a)f(x)\ dx\right] \tag{2.45}$$
$$= -af(a) - \int_a^b f(x)\ dx - \int_b^{\sup X} f(x)\ dx \tag{2.46}$$
$$= -P(X>a). \tag{2.47}$$

Thus

$$\frac{d}{da} CVaR_\alpha(I(X)+\pi) = \frac{P(X>a)}{1-\alpha} - (1+\theta)P(X>a) \tag{2.48}$$
$$= P(X>a)\left(\frac{1}{1-\alpha} - (1+\theta)\right), \tag{2.49}$$

which is greater than 0 when $1-\alpha < \frac{1}{1+\theta}$. This is consistent with the results from Equation (2.17) and Equation (2.28). From this we conclude that the optimal contract parameter a for a stop-loss contract also is optimal for the insurance layer contract. ■

So far we have only contemplated reinsurance of a single risk, for which optimal contract parameters are easily determined. In the recent work [HC20] by Huseby et al., the problem of optimal reinsurance in the multivariate case was studied. Here, m independent continuously distributed risks were considered, where the optimization problem was:

$$\min_{\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m} q_\alpha\left(\sum_{i=1}^m I_i(X_i) + (1+\theta)\sum_{i=1}^m E[R_i(X_i)]\right), \tag{2.50}$$

which was shown to have optimality criteria in line with those of the univariate problem from (2.30). While the latter problem had unique optimal parameters a^*, b^* when $1-\alpha < (1+\theta)^{-1}$, which could explicitly be identified from the survival functions of X_i. While this is not the case for multivariate risks, the authors derived necessary criteria for an optimal solution to problem (2.50),

together with a numerical scheme for identifying optimal solutions. Particularly, it was shown that

$$a_i = S_{X_i}((1+\theta)^{-1}), i = 1, \dots, m \tag{2.51}$$

and

$$P(\sum_{i=1}^{m} I_i(X_i) > \sum_{i=1}^{m} a_i) = \alpha \tag{2.52}$$

were necessary conditions for optmiality of $a_1, b_1, a_2 \dots, a_m, b_m$ provided $(\theta+1)^{-m} \geq \alpha$. It is clear that the constraint

$$P(\sum_{i=1}^{m} I_i(X_i) > \sum_{i=1}^{m} a_i) = \alpha \tag{2.53}$$

may have multiple or infinitely many solutions. To ease the search of optimal parameters $\{b_i\}_i$, one may impose the restriction that

$$P(X_i > b_i) \tag{2.54}$$

must have the same value for all i. Denoting this common value $B \coloneqq S_{X_i}(b_i)$, one finds parameters $\{b_i\}_i$ such that Equation (2.53) holds by choosing $b_i = S_{X_i}(B)$. Such solutions are called *balanced solutions* in [HC20]. However, while balanced solutions are convenient for identifying b_i, they are not guaranteed to be globally optimal. Alternatively, optimal contract parameters are determined by iterating over paramters b_i such that Equation (2.53) is satisfied, while minimizing the total cost from problem (2.50). Moreover, the authors found that there was a financial upside to optimizing the contract parameters over multiple risks with respect to the Value-at-Risk of the sum of residual risks, compared to optimizing each contract individually. Moreover, it was shown that by bundling risks together by forming the combined risk $X = X_1 + X_2 + \cdots + X_m$ further financial upside could be gained. This begs the question: Are there similar optimality criteria for reinsurance in the multivariate case, with respect to conditional Value-at-Risk? The next section is devoted to the study of this question.

2.3 Optimal reinsurance in the multivariate case, under conditional Value-at-Risk

Inspired by [HC20] we formulate the following minimization problem:

$$\min_{\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m} CVaR_\alpha \left[\sum_{i=1}^{m} I_i(X_i) + \sum_{i=1}^{m} \pi_i \right],$$

where maintain the assumption that X_i are positive and bounded, continuous stochastic variables with monotonically increasing cumulative distribution functions. The change from the univariate to a multivaraite setting causes the optimization problems to be considerable more complex. Particularly, we work with expressions on the form

$$\int_{\alpha}^{1} q_{\beta}(\sum_{i=1}^{m} I_i(X_i)\ d\beta. \tag{2.55}$$

The difficulty lies in finding an expression for the quantile of the sum of I_is. Even in the case where the underlying variables X_i are assumed to be independent, an analytical expression for $q_{\beta}\left(\sum_{i=1}^{m} I_i\right)$ may not be available. Moreover, we are perhaps not justified in assuming that the magnitudes of the losses are independent of each other. In a real life-scenario where X_i are losses claimed by different insurance holders for similar insurance products, the claims may vary together depending on external conditions such as the weather. Moreover, the dependency between risks may be non-linear, with a stronger dependence in the tail part of the distribution. For instance, water damages caused by bursting water pipes would probably be independent of each other, while large claims after a severe flood will probably not be independent. However, the exact dependency structure between the risks may be unknown, be dependent on the context of the insurance contracts, or too difficult to use in practice.

In light of this, we will first consider a pessimistic view, in which the risks are assumed to be *comonotonic*. Before defining comonotonicity of random variables, we need the notion of equality in distribution of random variables.

Definition 2.3.1 (Equality in distribution)**.** Two real valued random variables X_1, X_2 with cumulative distribution functions F_1, F_2 are equal in distribution (equal in law) if and only if the following holds:

$$\forall x \in \mathbb{R} \quad F_1(x) = F_2(x), \tag{2.56}$$

in which case we write

$$X_1 \overset{d}{=} X_2. \tag{2.57}$$

According to Theorem 4 in [Dha+02], the following is an equivalent definition of comonotonicity.

Definition 2.3.2 (Comonotonicity)**.** Let

$$\mathbf{X} = (X_1, X_2, \ldots, X_m) \tag{2.58}$$

be a random vector. Then, $\mathbf{X}$ is said to be *comonotonic* iff there exists a random variable Z and non-decreasging functions $f_1, f_2, \ldots, f_m$ such that

$$\mathbf{X} \overset{d}{=} (f_1(Z), f_2(Z), \ldots, f_m(Z)). \tag{2.59}$$

It is clear from the definition that comonotonicy represents a particularly strong type of positive dependence between random variables; any increase to the underlying random variable Z is associated with an increase in each X_i. The next theorem, which is Theorem 7 in [Kaa+02], shows that, from a risk analysis perspective, comonotonicity of random variables constitutes a "worst case" dependency in the case of VaR.

Theorem 2.3.3 (Quantile-additivity of comonotonic random variables)**.** *If the random vector $X_1, X_2, \ldots, X_m$ is comonotonic, then the α-quantiles of the sum of its components are equal to the sum of the α-quantiles of its components:*

$$q_(\alpha)\left(X_1 + \cdots + X_m\right) = q_\alpha(X_1) + \cdots + q_\alpha(X_m). \tag{2.60}$$

From this it follows immediately that CVaR is comonotonically additive.

Corollary 2.3.4 (CVaR is additive for comonotonic risks)**.** *Let $(X_1, X_2, \ldots, X_m)$ be comonotionic. Then, for any $\alpha \in (0, 1)$,*

$$CVaR_\alpha(X_1 + \cdots + X_m) = CVaR_\alpha(X_1) + \cdots + CVaR_\alpha(X_m) \tag{2.61}$$

This result is given in Theorem 4.1.1 in [Dha+06]. We expand on the details in the following proof.

✻*Proof.* The result follows directly from Theorem (2.3.3) and linearity of the integral: Assume $(X_1, \ldots, X_m)$ are comonotonic. From the definition of CVaR we have that

$$\begin{aligned}
CVaR_\alpha(X_1 + \cdots + X_m) &= \frac{1}{1-\alpha}\int_\alpha^1 q_\beta\left(X_1 + \cdots + X_m\right)\, d\beta && (2.62)\\
&= \int_\alpha^1 q_\beta(X_1) + \cdots + q_\beta(X_m)\, d\beta && (2.63)\\
&= \frac{1}{1-\alpha}\left(\int_\alpha^1 q_\beta(X_1)\, d\beta + \cdots + \int_\alpha^1 q_\beta(X_m)\, d\beta\right) && (2.64)\\
&= CVaR_\alpha(X_1) + \cdots + CVaR_\alpha(X_m), && (2.65)
\end{aligned}$$

which is what we wanted to show. ■

Recall from Equation (1.68) that a coherent risk measure in the basic sense satisfies subadditivity, i.e, $\mathcal{R}(X_1 + X_2) \leq \mathcal{R}(X_1) + \mathcal{R}(X_2)$. The previous proposition tells us that equality is attained in the case where X_1 and X_2 are comonotonic. This has the interpretation that the risk can not be decreased by means of diversification. We aim to show next that comonotonic risks have a worst case dependency with respect to conditional value at risk. We begin by stating a result from Proposition 1 in [Puc13], which shows that convex ordering can be characterized by CVaR.

Proposition 2.3.5.

Part a)

Let X, Y be real valued random variables. Then, the following are equivalent:

(i) $X \leq_{CX} Y$

(ii) $CVaR_\alpha(X) \leq CVaR_\alpha(Y)$ *for all* $\alpha \in (0,1)$.

Part b)
Let $(X_1, X_2, \ldots, X_m)$ *be real valued random variables which have marginal distributions* $(F_1, F_2, \ldots, F_m)$*, and let* $(X_1^*, X_2^*, \ldots, X_m^*)$ *be comonotonic random variables with the same distribution functions* $(F_1, F_2, \ldots, F_m)$*. Then,*

$$\sum_{i=1}^{m} X_i \leq_{CX} \sum_{i=1}^{m} X_i^* \tag{2.66}$$

We now ready to show that the comonotonc dependency is a worst case dependency with respect to CVaR. The following proposition is adapted to our setting from [Puc13].

Proposition 2.3.6. *Let* $\mathcal{X}$ *denote the collection of all random variables* $(X_1, X_2, \ldots, X_m)$ *which have marginal distributions* $(F_1, F_2, \ldots, F_m)$ *and final first moments. Denote the worst case CVaR of the sum* $X_1 + \cdots + X_m$ *as*

$$\overline{CVaR}_\alpha(F_1, F_2, \ldots, F_m) \coloneqq \sup_{(X_1,\ldots,X_m)\in\mathcal{X}} CVaR_\alpha(X_1 + \cdots + X_m).$$

Then,

$$\overline{CVaR}_\alpha(F_1, F_2, \ldots, F_m) = CVaR_\alpha(X_1^*) + \cdots + CVaR_\alpha(X_m^*) \tag{2.67}$$

is attained when $(X_1^*, X_2^*, \ldots, X_m^*) \in \mathcal{X}$ *is comonotonic*

We expand on the details of the proof:

✻*Proof.* Let $\mathcal{X}$ denote the collection of all random variables $(X_1, X_2, \ldots, X_m)$ which have marginal distributions $(F_1, F_2, \ldots, F_m)$ and final first moments, and assume $(X_1, X_2, \ldots, X_m) \in \mathcal{X}$. From Proposition 2.3.5 part b) it follows that

$$\sum_{i=1}^{m} X_i \leq_{\mathrm{CX}} \sum_{i=1}^{m} X_i^* \tag{2.68}$$

for some $(X_1^*, X_2^*, \ldots, X_m^*) \in \mathcal{X}$ which is comonotonic. Moreover, from Proposition 2.3.5 part a) we have that for any $\alpha \in (0,1)$,

$$(2.68) \Longrightarrow CVaR_\alpha(X_1 + \cdots + X_m) \leq CVaR_\alpha(X_1^* + \cdots + X_m^*), \tag{2.69}$$

so the right hand side in the inequality in Equation (2.68) is an upper bound on $CVaR_\alpha(X_1 + \cdots + X_m)$. Furthermore, from Corllary 2.3.4 we get that

$$CVaR_\alpha(X_1^* + \cdots + X_m^*) = CVaR_\alpha(X_1^*) + \cdots + CVaR_\alpha(X_m^*). \tag{2.70}$$

This concludes the proof. ■

The term $\sum_{i=1}^m \pi_i$ is not stochastic, and can hence be taken out of the risk measure, using the translation invariance property of CVaR (and of coherent risk measures in general). Moreover, it is expected that the total risk

$$\min_{\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m} CVaR_\alpha\left[\sum_{i=1}^{m} I_i(X_i)\right] + \pi_i \tag{2.71}$$

is less than or equal to the sum of individual risks

$$\sum_{i=1}^{m} \min_{a_i,b_i} CVaR_\alpha \left[I_i(X_i)\right] + \pi_i \tag{2.72}$$

as a consequence of convexity of CVaR. We will aim to evaluate how pronounced this difference is through concrete examples. Moreover, we look to determine if the optimal contract parameters in the univariate case will also be optimal in the multivariate case.

With this in mind, we turn out attention to the problem of finding optimal reinsurance parameters for multivariate risks under stop-loss insurance. In the following theorem, we show that the optimal contract parameters for stop loss reinsurance contracts in this setting remain unchanged.

Theorem 2.3.7 (✻)**.** *Assume that $(X_1, X_2, \ldots, X_m)$ are m continuous, bounded random variables with comonotonic dependency. For the optimization problem*

$$\min_{\{a_i\}_{i=1}^m} CVaR_\alpha \left[\sum_{i=1}^{m} I_i(X_i) + \sum_{i=1}^{m} \pi_i\right],$$

where I_i are stop loss reinsurance contracts with parameters a_i and risk premium θ. If $1-\alpha < (1+\theta)^{-1}$, the following parameters are optimal:

$$a_i = S_{X_i}^{-1}((1+\theta)^{-1}). \tag{2.73}$$

✻*Proof.* Let $(X_1, X_2, \ldots, X_m)$ be as stated above, and let $a_1, a_2, \ldots, a_m$ be arbitrary. Then it holds that for any $\alpha \in (0,1)$,

$$CVaR_\alpha \left[\sum_{i=1}^{m} I_i(X_i) + \sum_{i=1}^{m} \pi_i\right] = CVaR_\alpha \left[\sum_{i=1}^{m} I_i\right] + \sum_{i=1}^{m} \pi_i, \tag{2.74}$$

using the translation invariance property of CVaR. Moreover, using the fact that CVaR is comonotonically additive fom Corollary 2.3.4 yields

$$CVaR_\alpha \left[\sum_{i=1}^{m} I_i\right] + \sum_{i=1}^{m} \pi_i = \sum_{i=1}^{m} \left(CVaR_\alpha(I_i) + \pi_i\right). \tag{2.75}$$

For each i we then have that

$$\min_{a_i} CVaR_\alpha \left[I_i(X_i) + \pi_i\right]$$

has an optimal deductible $a_i^* = S_{X_i}^{-1}((1+\theta)^{-1})$ according to Theorem 2.2.1. From this we gather that

$$\min_{\{a_i\}_{i=1}^m} CVaR_\alpha \left[\sum_{i=1}^{m} I_i(X_i) + \sum_{i=1}^{m} \pi_i\right]$$

is attained when $a_i = S_{X_i}^{-1}((1+\theta)^{-1})$. This concludes the proof. ■

The assumed comonotonicity of risks in Theorem 2.3.7 may be subject to criticism. In Chapter 1 it was discussed that hedging against worst case scenarios may not be justifiable or feasible. We remarked there that optimizing under the constraint that the worst case risk should be acceptable - for instance, to design a structure that fails under no possible circumstance, may be impossible or too expensive. Does this criticism carry over to the optimization of reinsurance contracts in the multivariate case? Probably not. To better understand the implications of using the assumption of comonotonicity, we need to review how the assumption of comonotonicity has on the optimal value; could we achieve significantly better reinsurance contracts under milder assumptions on the dependency structures on the risks, provided that their true dependency was not comonotonic? Although we have not been successful in deriving optimal contract parameters for a wider class of dependency structures of the risks, the following proposition indicates that there could be similarities regarding the optimal contract parameters.

Proposition 2.3.8 (✻)**.** *Assume that $(X_1, X_2, \ldots, X_m)$ are m positive and continuous bounded random variables with marginal distributions $F_1, F_2, \ldots, F_m$, but with unknown dependency structure. Let $I_1, I_2, \ldots, I_m$ be stop-loss reinsurance contracts with parameters $a_1, a_2, \ldots, a_m$. Let $\alpha \in (0,1)$ be fixed, and assume further that $a_1, a_2, \ldots, a_m$ are chosen such that*

$$q_\alpha \left(I_1(X_1) + \cdots + I_m(X_m) \right) = a_1 + a_2 + \ldots, + a_m. \tag{2.76}$$

Then, for the optimization problem

$$\min_{a_i} CVaR_\alpha \left[\sum_{i=1}^m I_i(X_i) + \sum_{i=1}^m \pi_i \right], \tag{2.77}$$

$a_i = S_{X_i}((1+\theta)^{-1})$ are optimal contract parameters.

✻*Proof.* Notice first that it is possible to find a_i such that $q_{I_1+\cdots+I_m}(\alpha) = a_1 + \cdots + a_m$. For instance, if $a_i = 0$, then $I_1 + \cdots + I_m \equiv 0$, in which case $q_\alpha(I_1 + \cdots + I_m) = 0 = a_1 + \cdots + a_m$. Moreover, under this assumption,

$$CVaR_\alpha \left[\sum_{i=1}^m I_i(X_i) + \sum_{i=1}^m \pi_i \right] = CVaR_\alpha \left[\sum_{i=1}^m I_i(X_i) \right] + \sum_{i=1}^m \pi_i \tag{2.78}$$

$$= \frac{1}{1-\alpha} \left[\int_\alpha^1 q_{I_1+\cdots+I_m}(\beta)\ d\beta \right] + \sum_{i=1}^m \pi_i \tag{2.79}$$

$$= \frac{1}{1-\alpha} \left[\int_\alpha^1 a_1 + \cdots + a_m\ d\beta \right] + \sum_{i=1}^m \pi_i \tag{2.80}$$

$$= a_1 + \cdots + a_m + \sum_{i=1}^m \pi_i. \tag{2.81}$$

To optimize this, we find the partial derivative with respect to each contract parameter a_i. For the premium term π_j, we have that

$$\pi_j = (1+\theta) E[R_j(X_j)], \tag{2.82}$$

so

$$\frac{\partial \pi_i}{\partial a_j} = (1+\theta)\frac{\partial}{\partial a_j}\int_{a_i}^{\sup X_i}(x-a_i)f_{X_i}(x)\,dx \tag{2.83}$$

$$= \begin{cases} -(1+\theta)P(X_i > a_i) & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.84}$$

Altogether, this gives

$$\frac{\partial}{\partial a_j}\sum_{i=1}^{m}(a_i+\pi_i) = 1-(1+\theta)P(X_j > a_j). \tag{2.85}$$

Setting this equal to 0 yields the optimal solutions

$$a_i = S_{X_i}^{-1}(1+\theta)^{-1}, \tag{2.86}$$

which is what we wanted to show. ■

While we have derived some theoretical results regarding the optimal parameter choices for stop-loss reinsurance contracts under rather strong assumptions, namely either comonotonicity of the risks, or that $q_{I_1+\cdots+I_m}(\alpha) = a_1+\cdots+a_m$, there are still unanswered questions. Is the stop-loss contract prefreable over an insruance layer contract in the general multivaraite case? And if the risks are not comonotonic, are there other choices of a_i that yield better solutions than those derived above? The remainder of this chapter is dedicated to numerical examples, where the aim is to explore these questions.

2.4 Numerical examples

In this section, we will empirically study optimal reinsurance parameters for risks with different distributions and dependency structures. The theoretical results from the previous section have relied on continuously distributed bounded random variables. When using pseudo-randomly generated numbers on a computer, with final sample size, we will in practice not attain continuous distributions. The cumulative distribution functions and survival functions will therefore be approximated from the empirical distributions of the random variables when it is necessary. In the interest of computability and visualization, we will limit the number of risks to two. For the remainder of the section, let $\alpha = 0.95$ and $\theta = \frac{1}{9}$ be fixed, and denote the value of

$$CVaR_\alpha\left[\sum_{i=1}^{m} I_i(X_i) + \sum_{i=1}^{m}\pi_i\right] \tag{2.87}$$

as V_α. We will use numerical methods to estimate optimal solutions to study this optimization problem. The next example reveals that this problem leads to non-convex optimization, for which we are not guaranteed to find a global optimum. In order to increase the likelihood of finding satisfying solutions, we will use two different optimzation schemes. We generate a large sample from

the joint distribution of the risks. Then, we use a grid-based method, where we iteratively search over a mesh of points over in the parameter values. For each iteration, the grid is refined in a neighbourhood of the currently best solution, see implementation in Example 2, Example 3 and Example 4 in the appendix. Finally, we apply the *L-BFGS*-algorithm to attempt to identify optimal contract parameters, which uses an approximation of the Hessian matrix to guide the search. Since the problem is non-convex, its convergence to a global minimum is not guaranteed. However, the L-BFGS-algorithm is a popular choice also for non-convex problems, see for instance [LF01], [SYX16]. The implementation is found in the code listings mentioned above..

The examples in this section follow the same structure, where we complete the following steps:

- Step 1: Calculate the optimal worst case stop-loss reinsurance, according to Theorem 2.3.7.
- Step 2: Numerically estimate of the optimal stop-loss reinsurance contract parameters.
- Step 3: Numerically estimate an optimal insurance layer contract, given that deductibles as found in Step 1 are used.
- Step 4: Numerically estimate the optimal insurance layer contract, when all parameters are free.
- Step 5: Compare the above results.

Example 2.4.1 (✻Independent risks, with respect to CVaR.)**.** Let X_1 be as in Example 2.2.3, having a log-normal distribution with

$$E[X_1] = 50,\ SD[X_1] = 5, \tag{2.88}$$

and let X_2 be independent of X_1, having log-normal distribution with

$$E[X_2] = 100,\ SD[X_2] = 10. \tag{2.89}$$

Step 1: In Example 2.2.3, it was estimated that

$$a = 43.81, b = +\infty \tag{2.90}$$

were optimal, yielding an objective value of 50.90. If X_2 is treated separately from X_1, we would get optimal contract parameters

$$a_2 = S_{X_2}((1+\theta)^{-1}) = 87.52, b_2 = +\infty, \tag{2.91}$$

by Theorem 2.2.1, which yields the optimal value of 101.80. Using these parameters, an optimal value of $V_\alpha = 152.72$ was estimated. All in all, we have:

$$a_1 = 43.81,\ a_2 = 87.52,\ b_1 = b_2 = +\infty, \tag{2.92}$$

$$V_\alpha = 152.72 \tag{2.93}$$

Step 2: We now estimate the optimal parameters of a stop-loss contract, by letting

$$b_1 = b_2 = +\infty, \tag{2.94}$$

while searching for the best choices of (a_1, a_2) over a fine mesh of points. The optimal stop-loss contract parameters found were:

$$a_1^* = 43.80, a_2^* = 87.52, \tag{2.95}$$

with an optimal value of $V_\alpha = 152.72$. These results coincide with those found in step 1, where the result from Theorem 2.3.7 was applied. Figure 2.3a shows the objective value of V_α for each pair of (a_1, a_2) as a surface plot. As can be seen from the graph, the optimization problem appears non-convex in the parameter a_1.
Step 3: Next, we investigate if an insurance layer contract with the same deductibles as found in step 2 can give a smaller objective value. Fixing

$$a_1 = a_1^*, a_2 = a_2^* \tag{2.96}$$

and optimizing b_1, b_2 using a similar numerical scheme as for a_1, a_2, we find that optimal values for b_1, b_2 are approximately the largest values in the samples of X_1 and X_2 respectively.

$$b_1 \approx \sup X_1, b_2 \approx \sup X_2. \tag{2.97}$$

This pattern persists for any choice of (a_1, a_2) tested.
Step 4: Finally, we apply with the L-BFGS-algorithm to estimate the optimal reinsurance layer parameters. Using multiple different initial values parameter values, the algorithm converged to similar solutions, with

$$a_1 = 43.80, a_2 = 87.52, b_1 \approx \sup(X_1), b_2 \approx \sup(X_2), \tag{2.98}$$

and objective value of 152.72. Such parameters b_1, b_2 seem to indicate that a stop-loss contract is optimal or close to optimal in this case.
Step 5: We conclude this example by noting we were not able to improve upon the optimal value value $V_\alpha = 152.72$, which was found applying the result from Theorem 2.3.7, where comonotonic dependence was assumed, by also taking into account insurance layer contracts with a free choice of parameters. In this instance, falsely assuming comonotonicity of the risks did not lead to a detectably worse optimal value.

In the next two examples, we study the optimal reinsurance contract of correlated risks. We will first create a sample with the same marginal distributions as in Example 2.4.1. We begin by reviewing how we obtain such a sample in the bivariate case. Assume the target marginal distributions are F_1, F_2 with inverse cumulative distributions F_1^{-1}, F_2^{-1}.

1. Let Z_1, Z_2 have a bivariate normal distribution with covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where ρ is the desired correlation.
2. Let $U_1 = \Phi(Z_1), U_2 = \Phi(Z_2)$, where Φ is the cumulative distribution function of a standard normal random variable. Since Z_1, Z_2 were both standard normally distributed, U_1, U_2 are uniformly distributed on $[0, 1]$.

3. Let $X_1 = F_1^{-1}(Z_1), X_2 = F_2^{-1}(Z_2)$. Then, X_1, X_2 are correlated random variables with the desired marginal distributions.

This strategy utilizes the fact that generating a sample from the multivariate normal distribution is readily available in most statistics packages. Although analytical expressions for F_1^{-1}, F_2^{-1} may not be available, we will approximate these numerically by using a statistics package to generate a large number from their marginal distributions, and approximate F_1^{-1}, F_2^{-1} numerically.

Example 2.4.2 (✻Correlated risks, with respect to CVaR.)**.** We generate a large sample form X_1, X_2 where X_1, X_2 have the same marginal distributions as in Example 2.4.1, but where the variables have a positive correlation of 0.7, See the code listed in Example 2 in the appendix for the implementation details. Figure 2.2 shows a scatter-plot of this sample, and the marginal distributions of X_1, X_2. The empirical correlation was found to be 0.698.
Step 1: Using the result from Theorem 2.2.1, we get that optimal worst-case stop-loss parameters are

$$a_1^* = 43.78, a_2^* = 87.56, b_1 = b_2 = +\infty, \tag{2.99}$$

which gives an objective value of 152.734.
Step 2: We now estimate the optimal parameters of a stop-loss contract, by letting

$$b_1 = b_2 = +\infty, \tag{2.100}$$

optimizing over (a_1, a_2). It was estimated that

$$a_1 = 43.78, a_2 = 87.57 \tag{2.101}$$

were optimal, with an optimal value of 152.734. Figure 2.3a shows the objective value of V_α for each pair of (a_1, a_2) as a surface plot.
Step 3: Fixing

$$a_1 = a_1^*, a_2 = a_2^* \tag{2.102}$$

and optimizing b_1, b_2 using a similar numerical scheme as for a_1, a_2, the optimal values of b_1, b_2 were found to be

$$b_1 = 82.61 = \sup\tilde{(}X_1), b_2 = 168.11 = \sup\tilde{(}X_2), \tag{2.103}$$

where $\sup\tilde{(}X_1), \sup\tilde{(}X_2)$ denote the maximum elements of the generated samples of X_1, X_2. The associated objective value was 152.734.
Step 4: Applying the L-BFGS-algorithm to with different initial guesses to estimate the optimal reinsurance layer parameters yielded the following parameters:

$$a_1 = 43.78,\ a_2 = 87.56,\ b_1 = 116.32,\ b_2 = 172.83, \tag{2.104}$$

with an associate objective value of 152.734.
Step 5: There were essentially no difference in the attained optimal value between the previous steps. In all cases,

$$a_1^* = 43.78, a_2^* = 87.56 \tag{2.105}$$

were found to be optimal. The associated objective value was 152.734. We were not able to identify an insurance layer contract which gave a better objective value than a stop-loss contracts.

We have thus far studied numerical examples with uncorrelated risks and with positively correlated risks. We now consider sampling from a negatively correlated .

Example 2.4.3 (✻Negatively dependent risks, with respect to CVaR.)**.** We now sample X_1, X_2 negatively correlated joint distribution, where the X_1 has a gamma distribution with shape parameter 2 and scale parameter 20. This means that

$$E[X_1] = 40, SD[X_1] = 20\sqrt{2}. \tag{2.106}$$

while X_2 has a normal distribution with

$$E[X_2] = 50, SD[X_2] = 10. \tag{2.107}$$

To obtain a negatively dependent sample with the specified distributions, complete the same steps described in the prelude to example 2.4.2, using a correlation matrix with $\rho = -0.7$ and the inverse cumulative distributions of X_1 and X_2. From this procedure, a sample with negative dependency was generated, which is illustrated in Figure 2.4.
Step 1: Using the result from Theorem 2.2.1, we get that optimal worst-case stop-loss parameters are

$$a_1^* = 10.638, a_2^* = 37.188, b_1 = b_2 = +\infty, \tag{2.108}$$

which gives an objective value of 95.644.
Step 2: We now estimate the optimal parameters of a stop-loss contract, by letting

$$b_1 = b_2 = +\infty, \tag{2.109}$$

optimizing over (a_1, a_2). Both the optimal parameters and the objective value coincide with those in step 1. 2.3c shows the objective value of V_α for each pair of (a_1, a_2) as a surface plot.
Step 3: Fixing

$$a_1 = a_1^*, a_2 = a_2^* \tag{2.110}$$

and optimizing b_1, b_2 using a similar numerical scheme as for a_1, a_2, the optimal values of b_1, b_2 were found to be

$$b_1 = 114.59, b_2 = 66.83, \tag{2.111}$$

The associated objective value was 95.479, which is in fact slightly smaller than the similar stop-loss contract from step 2. Indeed, the optimal value exhibits a slight increase whenever either b_1 or b_2 are increased past these values, given that $a_1 = a_1^*, a_2 = a_2^*$. This is illustrated in Figure 2.5, where the behaviour in b_1 is shown. When b_1 is increased from 78 to 80 with $b_2 = 66.83$ and (a_1, a_2) as before, the objective value rapidly increases from 95.58 to 95.64, before the

graph subsequently flattens out.
Step 4: Applying the L-BFGS-algorithm to with different initial guesses to estimate the optimal reinsurance layer parameters yielded the following parameters:

$$a_1 = 11.13,\ a_2 = 37.22,\ b_1 = 184.70,\ b_2 = 64.88, \tag{2.112}$$

with an associate objective value of 95.512.
Step 5: In this example, the best parameters found yielded an insurance layer contract with

$$a_1 = 11.13,\ a_2 = 37.22,\ b_1 = 184.70,\ b_2 = 64.88, \tag{2.113}$$

with an associate objective value of 95.512. In comparison, the objective value achieved by using the result from Theorem 2.2.1, whereby a stop-loss contract is used, was 95.644, which constitutes a 0.138% reduction in objective value.

In the final example of this section, we study the case of negatively correlated normal random variables.

Example 2.4.4 (✻Negatively dependent risks, with respect to CVaR.)**.** We now sample X_1, X_2 from the bivariate normal distribution, where the correlation is -0.3,

$$E[X_1] = 100,\ SD[X_1] = 25,\ E[X_2] = 80,\ SD[X_2] = 30. \tag{2.114}$$

Step 1: Using Theorem 2.2.1, we get

$$a_1^* = 67.90, a_2^* = 41.60, b_1 = b_2 = +\infty, \tag{2.115}$$

which gives an objective value of 190.90.
Step 2: We now estimate the optimal parameters of a stop-loss contract, by letting

$$b_1 = b_2 = +\infty, \tag{2.116}$$

optimizing over (a_1, a_2). Both the optimal parameters and the objective value coincide with those in step 1.
Step 3: Fixing

$$a_1 = a_1^*, a_2 = a_2^* \tag{2.117}$$

and optimizing b_1, b_2 using a similar numerical scheme as for a_1, a_2, the optimal values of b_1, b_2 were found to be

$$b_1 = 143.70, b_2 = 141.56, \tag{2.118}$$

with an objective value of 190.83. In comparison, the maximal values in the samples of X_1, X_2 were $217.91, 218.10$. Again it appears that an insurance layer contract gives a slightly better result compared to a stop-loss contract.
Step 4: An application of the L-BFGS-algorithm to with different initial guesses yields the following estimates of optimal parameters:

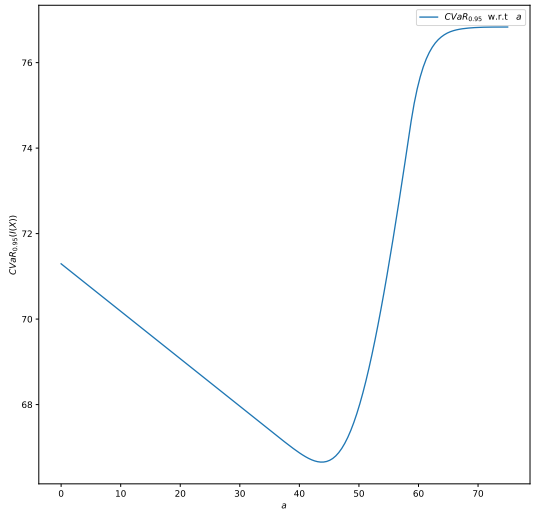
$$a_1 = 67.45,\ a_2 = 40.95,\ b_1 = 141.80,\ b_2 = 147.04, \tag{2.119}$$

with an associate objective value of 190.83.
Step 5: In this example, the best parameters found were

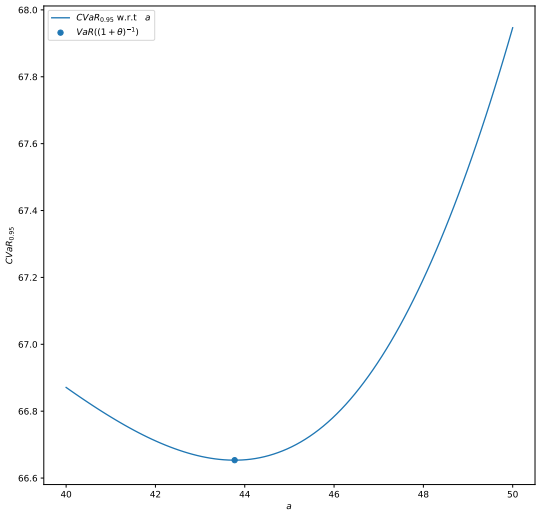
$$a_1 = 67.45,\ a_2 = 40.95,\ b_1 = 141.80,\ b_2 = 147.04, \tag{2.120}$$

with an associate objective value of 190.83. Compared to the results achieved by optimizing according to Theorem 2.2.1, we found a slight reduction in the objective value, of approximately 0.037%. The optimal deductibles were close in value to those found in step 1.

In this chapter, we have introduced the problem of optimal reinsurance under different risk measures. We have reviewed known results relating to optimal reinsurance in the univariate case under law-invariant risk measures and Value-at-Risk, together with recent developments in the multivaraite case under Value-at-Risk. We have then extended these result to the case of multivariate risks under CVaR, where we have optimized the contract parameters of a stop-loss reinsurance contract given a worst-case risk dependency structure. We have then applied this method of determining reinsurance contract parameters to risks which were not comonotonic, and compared the objective value to those found using numerical optimization methods. From the examples tested, we have seen that the objective value found using numeric optimization in all cases was very close to that found using the worst-case method. Similarly, the optimal deductibles were comparable between the methods. In the cases where the risks were negatively dependent, we were able to marginally improve on the solutions found using the worst-case approach, but not in the uncorrelated case. From this we may conclude that, in the case of bivariate risks, there are several situations where using a worst-case approach is viable. Moreover, in a real world application, the true dependency between the risks is probably not known. The historic claims data to infer such dependencies from may have a limited number of examples, which could give a false sense of dependency between the risks. In using a worst-case approach to the reinsurance optimization, one avoids the problem of choosing reinsurance parameters based on a wrongful assumption on the dependency of risks, which could lead to misunderstanding the true risk exposure.

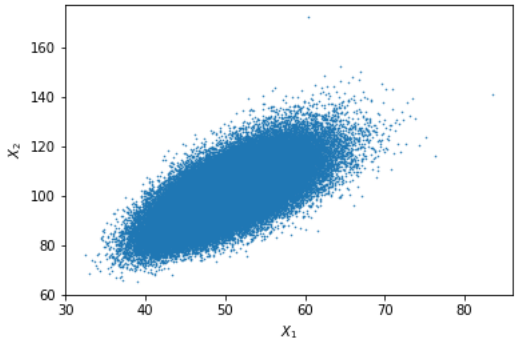


(a) $CVaR_{0.95}(I(X) + (1+\theta)E[R(X)])$ as a function of a.

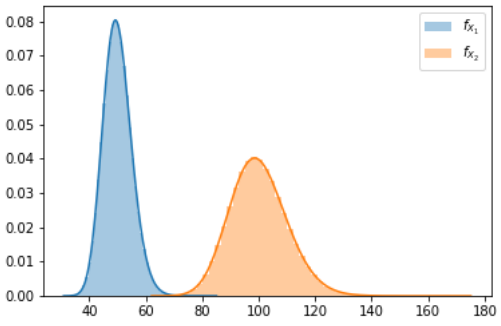


(b) Indicated optimal value for a.

Figure 2.1: Risk that cedent carries as a function of a.

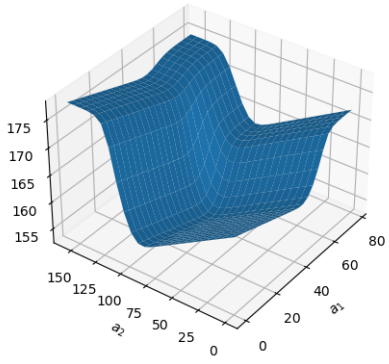


(a) Scatter-plot of X_1 vs X_2 when X_1, X_2 are log-normally distributed. A positive correlation of $\rho = 0.698$ was measured.

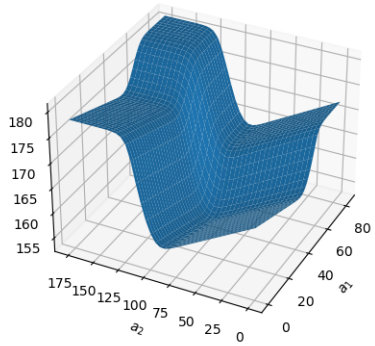


(b) Marginal distributions of X_1 and X_2.

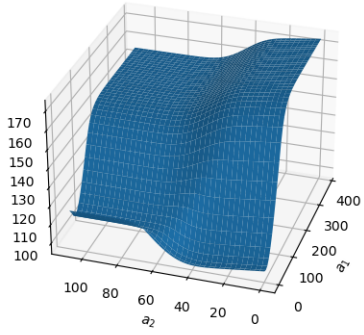
Figure 2.2: The distribution of X_1, X_2.



(a) Log-normally distributed, uncorrelated risks

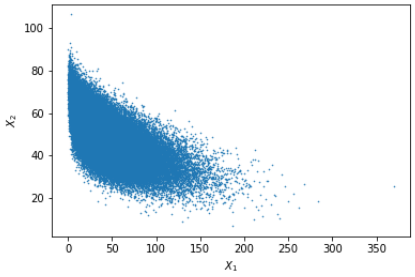


(b) Log-normally, positively correlated risks. $\rho = -0.66$

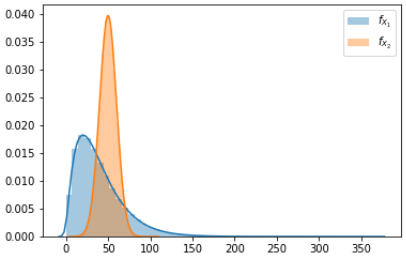


(c) Risks with beta and normal distributions, and negative dependency. $\rho = 0.698$

Figure 2.3: $CVaR_{0.95}(\sum_{i=1}^{m}[I_i(X) + (1+\theta)E[R_i(X_i)]])$ as a function of a_1, a_2 for various risk dependency structures

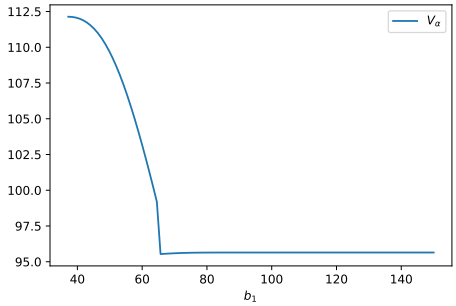


(a) Scatter-plot of X_1 vs X_2 when X_1, X_2 are negatively correlated, where X_1 has a beta distribution and X_2 has a normal distribution. A $\rho = -0.66$ was measured.

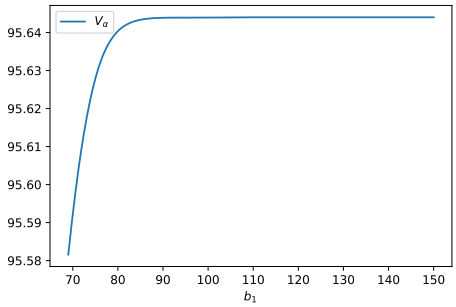


(b) Marginal distributions of X_1 and X_2.

Figure 2.4: The distribution of X_1, X_2 from Example 2.4.3



(a) A larger interval of b_1-values.



(b) A closer look at the b_1-values of interest.

Figure 2.5: V_α as a function of b_1, which illustrates behaviour described in Example 2.4.3, step 3.

CHAPTER 3

Estimating market risk

In this chapter we study the use of conditional Value-at-Risk to estimate risk in financial markets. In the previous chapters, we have pointed out shortcomings of Value-at-Risk, which will serve as a benchmark in this section. Important criticism has been Value-at-Risk's insensitivity to severe losses at the tail-end of the distribution. Nevertheless, Value-at-Risk continues to be a commonly used risk measure in finance. For instance, it is used to calculate regulatory capital requirements in *the Third Basel Accord*, which is a framework for banking regulation. The validity of Value-at-Risk for estimation of financial risk has been subject to extensive study. In [YY+02] the authors did a comparative analysis of Value-at-Risk and conditional Value-at-Risk for financial risk management. We quote the following summarizing remark from their work:

1. Information given by VaR may mislead rational investors who maximize their expected utility. In particular, rational investors employing only VaR as a risk measure are likely to construct a perverse position that would result in a larger loss in the states beyond the VaR level.
2. Investors can alleviate this problem by adopting expected shortfall as their conceptual viewpoint, since, by definition, it also takes into account the loss beyond the VaR level.
3. The effectiveness of expected shortfall, however, depends on the stability of estimation and the choice of efficient backtesting methods.

In this chapter, we will empirically compare the two risk measures through concrete numerical examples where we real financial market data. We will only consider financial assets from the stock market, where the portfolio of interest either consist of a single stock, or a combination of stocks in a stock index. There are in practice complicating factors in working with real market data, which is often referred to as market friction. For instance, selling a financial asset to reduce risk relies on the liquidity of the market. There is no guarantee that a motivated seller is able to sell their asset at a desired time, to the current market price. Although such factors will have an effect on the true risk of a financial position, they remain outside the scope of this chapter.

3.1 A model of financial assets

To model the price of financial asset mathematically, we will need the notion of a *stochastic process*

Definition 3.1.1 ([Øks03] Stochastic process)**.** A stochastic process is a parametrized collection of random variables

$$\{X_t\}_{t \in T} \tag{3.1}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$ and assuming values in $\mathbb{R}$.

As we desire to model the value of a portfolio of stocks at different times, we are interested in a suitable stochastic process

$$S_t, \tag{3.2}$$

such that S_t is a random variable representing the value of the portfolio at time t. Before presenting a candidate for such a process, we need a special stochastic process, namely the *Brownian motion* (GBM). The Brownian motion is continuous, and resembles a random walk at a large scale.

Definition 3.1.2 ([Øks03] 1-dimensional Brownian motions)**.** Let B_t be a stochastic process. Then, B_t is a Brownian motions if the following holds:

1. $B_0 = 0$.
2. For fixed $\omega' \in \Omega$, $t \mapsto B_t(\omega')$ is a continuous function.
3. $B_t - B_s$ is independent of $B_v - B_u$ for any $0 \leq s \leq t < v$.
4. $(B_t - B_s) \sim (0, (t - s))$ for any $0 \leq s \leq t$.

The Brownian motion has many interesting properties. As a consequence of points 1. and 3. in Definition 3.1.2, it follows that for $t \geq 0$,

$$B_t \sim N(0, t), \tag{3.3}$$

since

$$B_t = (B_t - B_0) \sim N(0, t - 0), \tag{3.4}$$

where $B_0 = 0$. Moreover, it can be shown that $\omega \mapsto B_t(\omega)$ it is a nowhere differentiable function in t. For our purposes, the random walk-like behaviour is of interest, and this feature will be used to create a stochastic process which mimics a similar random walk-like behaviour of the stock market. One can imagine that buyers and sellers constantly and randomly sell and buy stocks in a company, which nudges the price of the stock up and down in an erratic, random manner. Although this is a simple abstraction, the Brownian motion is an important building block for mathematical models of the prices of financial assets. However, the Brownian motion itself cannot be a good model, since it will eventually take negative values given enough time. Another stochastic process which exhibits a similar randomness, but which is always positive is the *geometric Brownian motion*.

Definition 3.1.3 ([Øks03])**.** Let μ and σ be positive real numbers, and let B_t be a Brownian motion. Then, the stochastic process, defined for $t \geq 0$,

$$S_t = S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t\right) \tag{3.5}$$

is a geometric Brownian motion, where S_0 is the initial value at time 0.

The geometric Brownian motion is the exponential of a Brownian motion, with drift. It is one of the most common models to describe the development of the price of a financial asset over time [BBK08, p. 19]. For example, the famous Black-Scholes model for option pricing assumes that the price of the underlying asset can be modeled by a geometric Brownian motion [BS73]. The geometric Brownian motion will serve as our model for the price development of stock portfolios. Note that since B_t has a normal distribution, so does $\ln(S_t)$, since

$$\ln(S_t) = \ln\left(S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t\right)\right) \tag{3.6}$$

$$= \ln(S_0) + (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t, \tag{3.7}$$

where σB_t has a normal distribution. This tells us that S_t has a log-normal distribution. It can be shown ([Øks03]) that S_t has a log-normal distribution with

$$E[S_t] = S_0 e^{\mu t}, \tag{3.8}$$

$$V[S_t] = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right). \tag{3.9}$$

3.2 Problem formulation

[✻] We will consider the financial risk of holding a stock portfolio for 1 trading day. Since the absolute value of the loss depends on the initial value of the portfolio, we will evaluate the loss as the relative difference between the value of the portfolio at time $t-1$ and time t. We let V_t denote the value of a the portfolio starting at time $t = 0$. Let for all t $= 1, 2, \ldots, T$

$$X_t = -\frac{V_t - V_{t-1}}{V_{t-1}}, \tag{3.10}$$

such that negative returns correspond to positive values. We can think of X_t as the stochastic process which represents the negative relative returns of the stock. We are interested in measuring the Value-at-Risk an conditional Value-at-Risk for X_t. However, if this analysis is only done retrospectively. If we were only to discuss discuss the Value-at-Risk of the portfolios past performance, it would be more of an exercise in statistics than in risk analysis. After all, we are most concerned with the future risk of holding the stock portfolio. In the following examples, we to use two approaches to estimating the financial risk. We will use the historic price data for a number of prior trading days to evaluate the empirical Value-at-Risk and conditional Value-at-Risk. Then, we will use the same historic data to estimate the parameters of a geometric Brownian motion, and use its probability distribution to estimate the Value-at-Risk and conditional Value-at-Risk. We will compare these estimates with the true observations of the returns.

3.3 Examples

[*]

Case 1: Standard & Poor's 500 index

Daily closing prices for the Standard & Poor's index were collected between the time periods January 1st 2015 and January 1st 2021. Let V_t denote the value of one share of the Standard & Poor's index at time t, where $t = 0$ is January 1st 2015. The negative relative returns

$$X_t = -\frac{V_t - V_{t-1}}{V_{t-1}} \tag{3.11}$$

were calculated for $t = 1, 2, \ldots$ up to the last day in the data set. The reason for the negative orientation is to fit with the positive orientation of our risk measures. Recall that have associated positive outcomes with positive risks in previous chapters. The distribution of these returns are plotted in Figure 3.1a. Starting at $t = 1$, we used the empirical Value-at-Risk and conditional Value-at-Risk of the sample

$$\{X_1, X_2, \ldots, X_{100}\} \tag{3.12}$$

to estimate $VaR_{0.95}(X_{101})$ and $VaR_{0.95}(X_{101})$. Similarly, we proceeded to repeat this process for

$$\{X_2, X_3, \ldots, X_{102}\} \tag{3.13}$$

to estimate the risk of $X_{103}, \ldots$, repeating the process to exhaust all the remaining observations in the dataset. In doing so, we have created empirical risk estimates for each of the returns $X_{101}, X_{102}, \ldots$, each time basing the estimate on the previous 100 trading days. Figure 3.1b displays the estimated risks for each of these returns, together with the true observed values of X_t for the same days. We then considered the following events: The 10% worst losses and the 10 worst losses. The estimates for $CVaR_{0.95}$ and $VaR_{0.95}$ for these days were compared to the observed losses. Figure 3.2 gives a visual representation of this. On the x-axis, the worst relative returns are plotted, against their corresponding risk estimates on the y-axis. Points that fall below the line through the origin represent states where the losses exceed the estimates, and vice versa for points above the line. From these figures, it is clear that the estimates using $VaR_{0.95}$ have a tendency of under-estimating the worst losses, and as the losses increase in magnitude, the tendency to fall further below the line increases. This is in contrast to the corresponding $CVaR_{0.95}$-estimates, where the points remain relatively centered about the line, even when the losses increase in magnitude.

Alternatively to empirically calculating the Value-at-Risk and conditional Value-at-Risk from the past returns, we use the assumption that the dynamics of the stock price is dictated by a geometric Brownian motion. The idea is again to use the historic data, namely to base our estimate of X_{100} on

$$\{V_0, V_1, \ldots, V_{99}\}, \tag{3.14}$$

$X_{101} on \{V_1, \ldots, V_{101}\}$ and so on. We do so by using $\{V_0, \ldots, V_{99}\}$ to estimate the parameters of a geometric Brownian motion, which we hope provides a good model for V_t for times $t = 0, 1, \ldots, 99, 100$. Having found suitable estimates for σ, μ, and knowing the current value V_{99}, we may then create a geometric Brownian motion starting at time $t = 99$, and infer the probability distribution of V_{100}. We know that this has a log-normal distribution, with expectation and listed in Equations (3.8), (3.9), calculated by setting $t = 1$, and inserting for S_0 the known value of V_{99} in these equations. Using this hypothesis of the distribution of V_{100}, we can find the $CVaR_\alpha$ and VaR_α of X_{100} from the theoretical distribution of V_{100}. From this distribution we can calculate the conditional Value-at-Risk. The implementation is found in the Appendix, A.7.

To estimate the parameters of the geometric Brownian motion, we use the estimation method 2.3 from section 2 of [CJM17]:

$$\bar{V} = \frac{1}{n} \sum_{t=1}^{n} \ln \left(\frac{V_t}{V_{t-1}} \right), \tag{3.15}$$

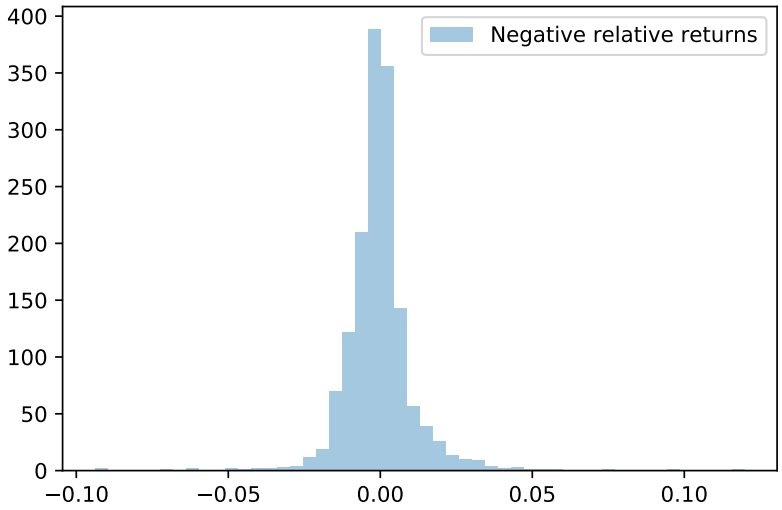
$$\hat{\mu} = \bar{V} + \frac{\hat{\sigma}^2}{2}, \tag{3.16}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} \left[\log \left(\frac{V_t}{V_{t-1}} \right) - \bar{V} \right]^2}. \tag{3.17}$$

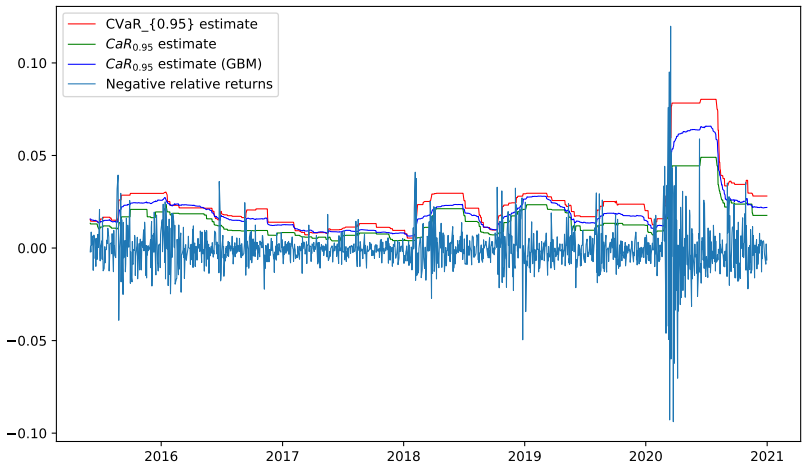
The blue line in Figure 3.1b is the $CVaR_{0.95}$-estimate based on this approach. As the figure shows, the risk estimate produces in this manner seems to almost always lie between the lower bound of the empirical $VaR_{0.95}$-estimate and the upper bound of the empirical $CVaR_{0.95}$-estimate. Figure 3.2c shows a scatter plot of the 10% worst losses, with negative relative returns on the x-axis and the $CVaR_{0.95}$ estimated through the GBM-method on the y-axis. We observe that this method resulted in fewer vast underestimates of the risk, compared to the empirical $VaR_{0.95}$-estimate, while it over-estimates the risks slightly less than the $CVaR_{0.95}$-estimates. Table 3.1 displays the risk estimates and observed losses for the 10 worst observed losses in the dataset. We see that the empirical estimation of $CVaR_{0.95}$ has produced the most conservative risk estimates, followed by the GBM-based $CVaR_{0.95}$ estimates before the empirical $VaR_{0.95}$-estimates.

Case 2: Microsoft

We repeat the same analysis as in Case 1 with daily stock prices for Microsoft (ticker code MSFT). Closing prices between the dates January 1st 2015 and January 1st 2021 were collected, and we proceeded to use a 100-day rolling window to create empirical estimates for the $CVaR_{0.95}$ and $VaR_{0.95}$ for the next day. We also estimated the next days $CVaR_{0.95}$ by the means of fitting a geometric Brownian motion to the previous 100 days of data, as was done in Case 1. In this instance, the estimates produces via the GBM-method were consistently more conservative than those of empirical Value-at-Risk and empirical conditional Value-at-Risk. From the scatter plots in Figure 3.2, we can see that even in the 10% worst cases, the GBM-based estimate still over-estimates the risk in most cases. In Figure 3.4a we can see a similar, but less pronounced tendency of over-estimating the risks in the case of empirical

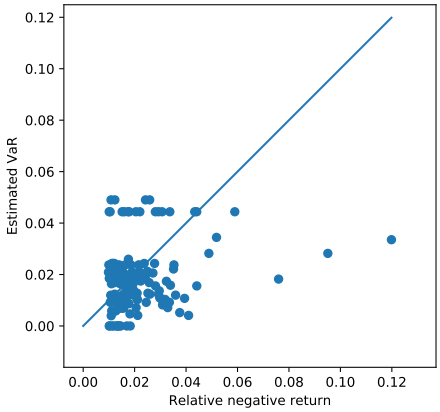


(a) Negative relative returns, when holding S&P 1 day.

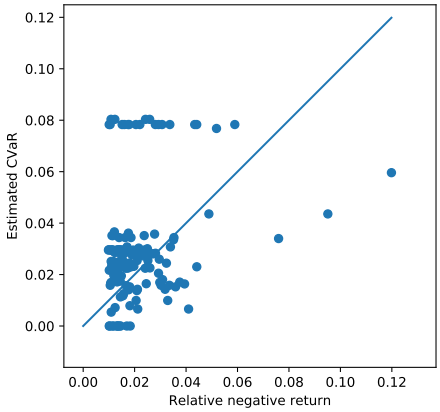


(b) Estimated $CVaR_{0.95}$ and $VaR_{0.95}$

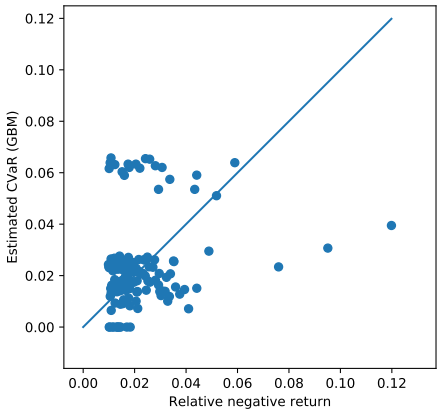
Figure 3.1: Estimated risk and measured negative relative returns from S&P using three methods, based on 100 days of prior market data: 1) Empirical estimate of VaR. 2) Empirical estimate of CVaR. 3) Estimating CVaR from a GBM fittet do the data. Measurement period was 2015.01.01-2021.01.01.



(a) Negative relative returns vs $CVaR_{0.95}$-estimates



(b) Negative relative returns vs $VaR_{0.95}$-estimates



(c) Negative relative returns vs $CVaR_{0.95}$, estimated from an inferred geometric Brownian motion.

Figure 3.2: The worst 10% negative relative returns of S&P, measured against the corresponding risk estimates for that day. Negative returns are plotted along the x-axis, with risk estimates on the y-axis.

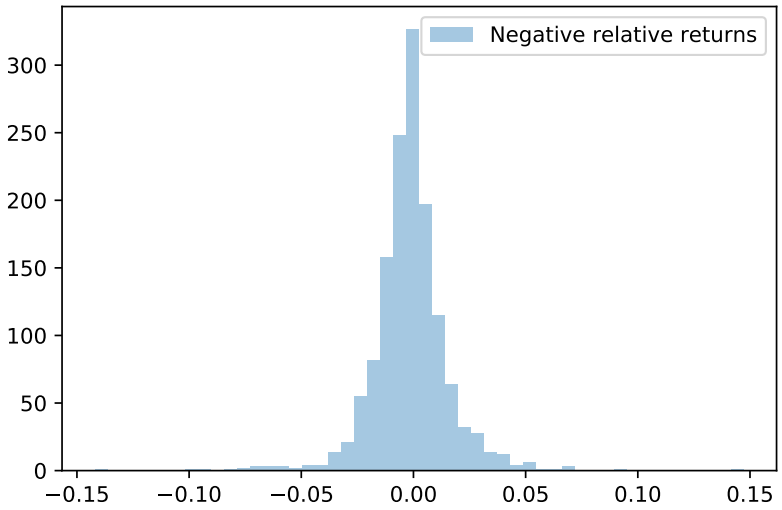
Loss	Empirical $CVaR_{0.95}$	Empirical $VaR_{0.95}$	GMB-estimated $CVaR_{0.95}$
0.119	0.059	0.033	0.039
0.095	0.043	0.028	0.030
0.075	0.033	0.018	0.023
0.058	0.078	0.044	0.063
0.051	0.076	0.034	0.051
0.048	0.043	0.028	0.029
0.044	0.023	0.015	0.015
0.044	0.078	0.044	0.059
0.043	0.078	0.044	0.053
0.040	0.006	0.004	0.007

Table 3.1: Risk estimates and observed losses for the worst 10 trading days of Standard & Poor's returns.

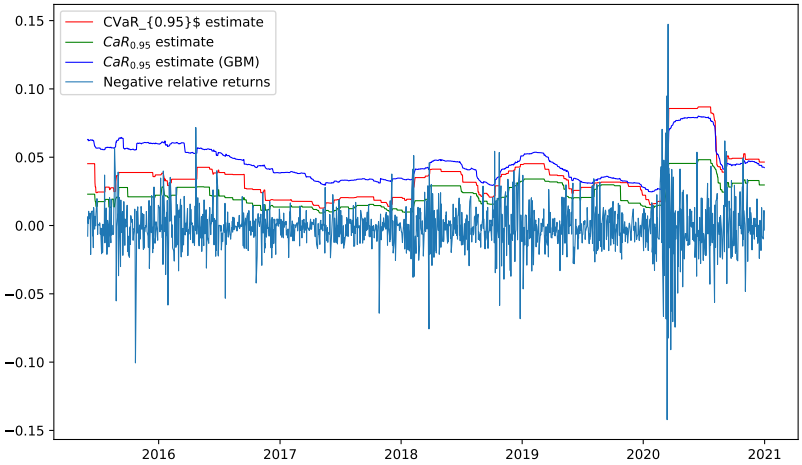
$CVaR$. In Figure 3.4b, the opposite is true. The risk is typically underestimated, especially for larger losses. The same is reflected in Table 3.2, which displays the losses and risk estimates for the worst 10 days of returns. At times, the VaR-estimate reports a risk estimate less than half of the GBM-based estimate, and less than one third of the observed loss. From this we can gather that VaR in certain markets provide poor risk estimates for exreme market movements.

Loss	Empirical $CVaR_{0.95}$	Empirical $VaR_{0.95}$	GMB-estimated $CVaR_{0.95}$
0.147	0.065	0.043	0.052
0.094	0.052	0.028	0.045
0.071	0.033	0.027	0.053
0.070	0.026	0.016	0.030
0.067	0.043	0.022	0.039
0.061	0.039	0.024	0.041
0.056	0.024	0.014	0.056
0.054	0.021	0.015	0.030
0.054	0.046	0.026	0.043
0.053	0.085	0.045	0.078

Table 3.2: Risk estimates and observed losses for the worst 10 days of Microsoft stock returns.

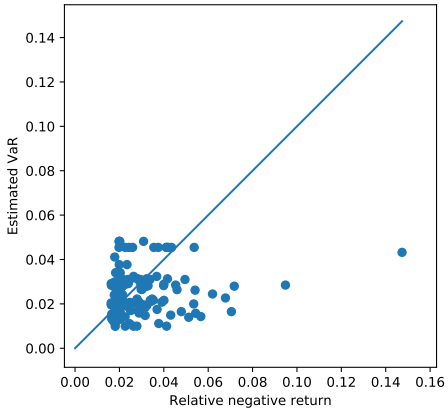


(a) Negative relative returns, when holding Microsoft stock 1 day.

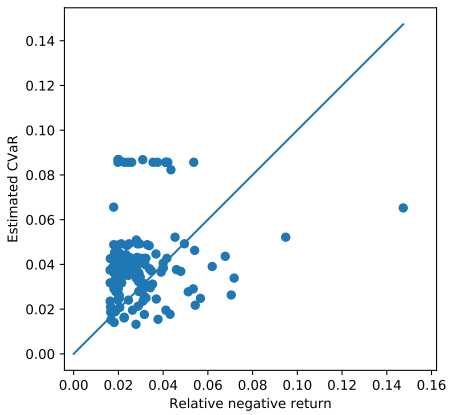


(b) Estimated $CVaR_{0.95}$ and $VaR_{0.95}$

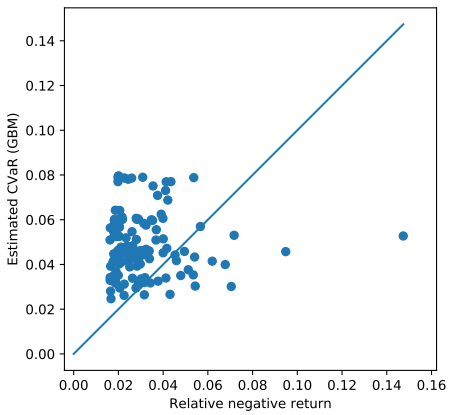
Figure 3.3: Estimated risk and measured negative relative returns from Microsoft using three methods, based on 100 days of prior market data: 1) Empirical estimate of VaR. 2) Empirical estimate of CVaR. 3) Estimating CVaR from a GBM fittet do the data. Measurement period was 2015.01.01-2021.01.01.



(a) Negative relative returns vs $CVaR_{0.95}$-estimates



(b) Negative relative returns vs $VaR_{0.95}$-estimates



(c) Negative relative returns vs $CVaR_{0.95}$, estimated from a inferred geometric Brownian motion.

Figure 3.4: The worst 10% negative relative returns of the stock Microsoft, measured against the corresponding risk estimates for that day. Negative returns are plotted along the x-axis, with risk estimates on the y-axis.

CHAPTER 4

Application of risk analysis in structural design

In the previous chapter, the main are of applications of risk analysis has been in the financial domain, and problems within finance lend themselves readily to risk analysis. After all, we have an immediate notion of how to understand undesired outcomes - they are typically simply a financial loss. In the following chapter, we aim to show that the framework developed in Chapter 1 easily carries over to applications within structural design and reliability. While the application of risk analysis in structural design and reliability is not a novel concept, we will see that conventional methods can be improved upon by applying results from Chapter 1, mainly by comparing traditional approaches to a modification that utilizes conditional Value-at-Risk. The chapter is to a large extent based on [RR10]. The use of an analogy to conditional Value-at-Risk will provide multiple benefits which will be discussed in this chapter. Most importantly, it will increase the control over and emphasis on risks associated with tail events, and it will in given circumstances be computationally favorable.

4.1 Introduction

Structural design and optimization is commonly concerned with analysing the reliability of a structure exposed to an environment. The word structure is not intended to be understood narrowly, and may include static structures such as buildings and bridges, machines, or windmills to mention a few examples. The way the structure functions, which can be thought of as the performance of the structure, will typically depend on one or several aspects of its environment. For the discussions ahead we will focus our attention to the case of mechanical structures in physical environments, but note that the scope of the techniques presented need not be limited to this setting. More specifically, given a mechanical structure which is exposed to an environment $\mathbf{V}$ we are interested in understanding when the structure will function, to what degree it functions, what states of $\mathbf{V}$ might lead to undesirable levels of function, and with what probability such states will occur. The vector $\mathbf{V}$ is here a collection of facts about the environment, perhaps including temperature, wind speed, humidity etc. Likewise, $\mathbf{x}$ is a vector holding data about the design of the structure. The following example illustrates this.

Example 4.1.1 (✻Motivating example for the usefulness of failure probability)**.** Assume the structure of interest is a floating wind turbine. The wind turbine is exposed to a multitude of forces from the environment which are acting on it, and of special concern are the wind-speed, the wave height and the temperature. Under ideal conditions the windmill produces electricity at full capacity. However, when the wind speed increases past a certain level, the wind turbine is forced to alter its angle to the wind and lean at an angle to avoid failing. Furthermore, if the temperature falls below a certain threshold, there will be an accumulation of ice on the blades of the turbine, which in turn halts the production rate. At a certain level, the forces apply stress which causes the turbine to be in a state we do not regard as acceptable. If both the wind and the wave height reach extreme levels, the combinations of forces acting on the turbine may become harmful to the construction, and may under dire circumstances cause a critical failure.

We could conceivably be in a situation to choose between several alternative designs of the wind turbine, and also be in possession of statistical data about the environment. Given knowledge about how the environment influences the state of the structure, we wish to make an informed decision about which structure to prefer. An attempt to solve this problem could utilize the concept of failure probabilities to analyse the performance associated with each of the proposed designs. The failure probability, which is yet to be precisely defined, would tell us which of the design options had the higher probability of failing, given that we had knowledge about the distribution of the environmental variables.

Definition 4.1.2 (Environmental Variables)**.** The vector

$$(V_1(\omega), V_2(\omega), \dots, V_k(\omega))^T \coloneqq \mathbf{V} \tag{4.1}$$

is a k-dimensional random variable V, which represents features of the environment or other factors that influence the structure, but that are not part of the design.

Definition 4.1.3 (Design Variables)**.** We will by $\mathbf{x} = (x_1, x_2, \dots, x_m)^T \in \mathbb{R}^m$ denote the *design variables* of a structure, where ^{T} signifies that the vector is transposed. The various components might encode a certain choice of material, strength of component or other aspects of the design of the structure, and are not regarded as random. As the name suggests, the design variables are variables that are within a designer's control to choose.

Definition 4.1.4 (Limit State Function)**.** The function $g\colon \mathbb{R}^m \times \mathbb{R}^k \mapsto \mathbb{R}$ denotes the *limit state function* of the structure, assigning a real value to each choice of design and state of the environment which represents the level of performance of the structure.

We will make a distinction in notation between $(V_1, V_2, \dots V_k)^T$ which are the environmental variables that are taken to be random, and $(v_1, v_2, \dots, v_k)^T \in \mathbb{R}^m$, which is a realization of $\mathbf{V}$. In this regard, $g(\mathbf{x}, \mathbf{V})$ is a random variable, whereas $g(\mathbf{x}, \mathbf{v})$ is a scalar value. If $g(\mathbf{X}, \mathbf{v}) < 0$, we say that the structure performs at an *acceptable* level, while $g(\mathbf{X}, \mathbf{v}) > 0$ indicated an *unacceptable* level of performance and $g(\mathbf{X}, \mathbf{v}) = 0$ is a *limit sate*.

4.2 Failure probability

Failure probability is a crucial concept in reliability theory, where the analyses aims to quantify the probability that a structure will fail. This is commonly used when the designer considers the return period of events such as earth quakes and floods. Specifying that the structure should withstand forces that occur less frequent than for instance once per century, is essentially equivalent to imposing a failure probability less than the probability that such forces occur withing a century. In this section we will formally introduce the failure probability, and its close relative, the *buffered failure probability*. We show that the latter is a more conservative choice and in some cases an improvement to the failure probability. We finally discuss aspects of computational costs.

Definition 4.2.1 (Failure probability)**.** The *failure probability*, of a structure, which is the probability that the structure is in failed state according to its limit state function, is defined as

$$p(\mathbf{x}, \mathbf{V}) \coloneqq P\left[g(\mathbf{x}, \mathbf{V}) > 0.\right]. \tag{4.2}$$

Note that although this definition does not mention time, time will be of importance in many practical applications. The probability that the structure, given a design choice, is exposed to an environment in which it fails, will in most circumstances depend on the duration of exposure. Other times, the degree of failure may depend on the amount of time the structure is exposed a sufficient level of stress. This means that one will often deal with problems related to the maximum value of $g(\mathbf{x}, \mathbf{V})$ over a period of time. In this discussion, we limit the scope to considering the canonical problem of $P[g(\mathbf{x}, \mathbf{V}) > 0]$. This is not a restriction, since the distribution of $\mathbf{V}$ can be modified to accommodate such time dependencies.

An immediate disadvantage of the failure probability is the lack of sensitivity to severe outcomes, which is analogous to the discussion on the use of the quantile as a risk measure in Section 1.3. This can be illustrated by an example.

Example 4.2.2 (✻)**.** Consider two alternative structures that are intended for the same environment. The environmental variables of concern are V_1, V_2 which are independent and standard normally distributed, and $\mathbf{x}$ represents material choices, and the choice $\mathbf{x} = (1, 1)^T$ is valid for both structures. (In the interest of concreteness, we can imagine that the first structure is a suspension bridge, while the second is a beam bridge). Their respective limit state functions are

$$g_1(\mathbf{x}, \mathbf{V}) = -|x_1 + x_2| + (v_1 + v_2)$$

$$g_2(\mathbf{x}, \mathbf{V}) = \begin{cases} -|x_1 + x_2| + (v_1 + v_2), & \text{if } -|x_1 + x_2| + (v_1 + v_2), < 0, \\ 1000 & \text{otherwise.} \end{cases}$$

This clearly means that both structures share the same failure probability of $p(\mathbf{x}) \approx 0.079$. On the other hand, if both structures are in a failed state, the second structure will almost always be in a more severe state than the first. This information is not revealed by the failure probability, and if the goal was to choose a structure with a failure probability of less than 0.1, both structures would appear equally well suited.

Estimation of failure probability

Assume that $\mathbf{V} = (V_1, \ldots, V_k)^T$ has a joint probability distribution $f_{\mathbf{v}}$ and $I(\cdot)$ is the indicator function for the event $g(\mathbf{x}, \mathbf{v}) > 0$. The failure probability $p(\mathbf{x})$ is given in integral form as

$$\int_{\Omega} I\big(g(\mathbf{x}, \mathbf{V}) > 0\big) dP(\omega), \tag{4.3}$$

which we evaluate as

$$\int_{\mathbb{R}^k} I\big(g(\mathbf{x}, \mathbf{v}) > 0\big) f_{\mathbf{v}}(\mathbf{v})\ d\mathbf{v}.$$

For complex structures, the number of random variables, k, of interest might be very large. As k grows, evaluating the failure probability analytically will in many cases be impractical or infeasible. Indeed, for many standard numerical integration rules, such as the Gauss quadrature extended to multiple dimensions, the error in the estimation decreases in the order of $n^{-C/k}$, where n is the number of partitions used in the integration and C is a constant depending on the method, c.f [Fau18][p.151]. This effect is often referred to as the dimensionality effect, or the curse of dimensionality. A numerical method for estimating high-dimensional integrals which exhibits significantly better scaling properties is the Monte Carlo method. In [RU+00] it is suggested that the evaluation of failure probability in high-dimensional cases should be done either by Monte Carlo integration, or via geometric considerations, although it is argued that neither of these approches are ideal. We will discuss these methods and their drawbacks in turn.

Monte Carlo estimation of the failure probability

To evaluate the failure probability $p(\mathbf{x})$ with Monte-Carlo simulation, one draws N samples $\{\mathbf{v}_i\}_{i=1}^N$ from the joint distribution $f_{\mathbf{V}}$. Then,

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} I\big(g(\mathbf{x}, \mathbf{v_i}) > 0\big)$$

is an unbiased estimator of $p(\mathbf{x})$. The main objection to this strategy is the computational cost involved when N becomes large. By the law of large numbers, the estimation error decreases proportional to $N^{\frac{-1}{2}}$, regardless of the dimensionality. In the meantime, the computational cost of sampling increases linearly as N grows. AA benefit of the Monte Carlo method is that it is well suited for parallel computing.

An alternative to numerical integration to evaluate the failure probability is a geometric consideration. In this approach, one considers a set consisting of states of $\mathbf{V}$ which are associated with a failed state, and uses the minimal distance from this set to the origin to infer the failure probability.

Definition 4.2.3 (Failure set)**.** Let a limit state function g and design variables $\mathbf{x}$ be given. Then, the set

$$\{\mathbf{v} \mid g(\mathbf{x}, \mathbf{v}) > 0\}$$

is the *failure set* of g given $\mathbf{x}$

The shortest distance from the origin to the boundary of this set is known as the *reliability index* of the design $\mathbf{x}$, and is denoted $\beta(\mathbf{x})$. Under certain conditions there is a one-to-one correspondence between the reliability index and the failure probability. In other cases, the reliability index can be used to give an estimate of failure probability. However, there is in general no guarantee that this estimate is a good approximation, and it might in fact underestimate the true failure probability. We illustrate the failure set and the reliability index with an example.

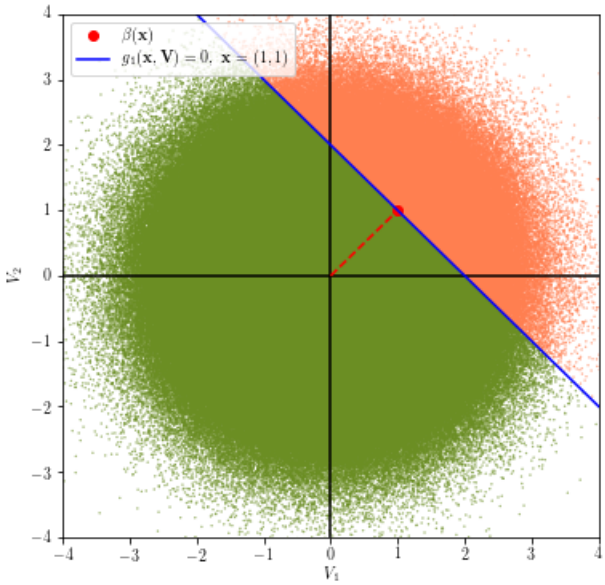


Figure 4.1: Monte-Carlo sampling from f. Green points indicate a safe state, red points a failed state. The blue line indicates the limit states where $g_1(\mathbf{x}, \mathbf{v}) = 0$. The red line indicates the shortest distance to the boundary of the set $\{\mathbf{v} \mid g_1(\mathbf{x}, \mathbf{v}) \leq 0\}$. The failure probability is approximately 7.86%.

Example 4.2.4. Example 4.2.2 featured a limit state function

$$g_1(\mathbf{x}, \mathbf{V}) = -|x_1 + x_2| + (V_1 + V_2),$$

with $\mathbf{x} = (1, 1)^T$. Figure 4.1 illustrates 2.5M samples of $(v_1, v_2)^T$ from a bivariate normal distribution. Red color indicates if the structure is failed, while green indicate safe. As indicated in the figure, the line $V_2 = 1 - V_1$ represents the boundary of the failure set, namely

$$\{\mathbf{v} \mid g_1(\mathbf{x}, \mathbf{v}) = 0, \mathbf{x} = (1, 1)^T\} \tag{4.4}$$

Points that fall above the line correspond to failed states, points below the line correspond to safe states.

To determine for which cases the failure probability corresponds to the failure probability we will need the following definition.

Definition 4.2.5 (Affine function)**.** Assume $\mathbf{v} \in \mathbb{R}^k$ is a realization of $\mathbf{V}$. A function g is affine in $\mathbf{v}$ if

$$g(\mathbf{x}, \mathbf{v}) = \langle \mathbf{a}(\mathbf{x}), \mathbf{v} \rangle + \mathbf{b}(\mathbf{x}),$$

where $\mathbf{a}$ is an k-valued function, $\mathbf{b}$ a real-valued function and $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^k$.

The following result is found in [RU+00, p. 7] (although not listed as a preposition).

Proposition 4.2.6. *Assume that $\{V_i\}_{i=1}^k$ are independent, standard normally distributed, and that the limit state function is affine in $\mathbf{v}$. Then, the failure probability $p(\mathbf{x})$ is given by*

$$p(\mathbf{x}) = \Phi(-\beta(\mathbf{x})),$$

whenever $p(\mathbf{x}) \leq 0.5$ where $\beta(-\mathbf{x})$ is the reliability index of the design $\mathbf{x}$.

Proof. Assume $g(\mathbf{x}, \mathbf{v})$ is affine in $\mathbf{v}$. Then, the boundary of the failure region, defined as

$$\{\mathbf{v} \mid g(\mathbf{x}, \mathbf{v}) = 0\},$$

is given by

$$\{\mathbf{v} \mid \langle \mathbf{a}(\mathbf{x}), \mathbf{v} \rangle + \mathbf{b}(\mathbf{x}) = 0\}.$$

It defines a hyperplane in $\mathbb{R}^k$, which we denote P. The failure probability therefore corresponds to the probability that $\mathbf{v}$ lies in the upper half-space defined by P, and is given by

$$p(\mathbf{x}) = P\big(\langle \mathbf{a}(\mathbf{x}), \mathbf{v} \rangle \geq \mathbf{b}(\mathbf{x})\big).$$

By assumption $\{V_i\}_{i=1}^k$ are independent, standard normally distributed, and the sum of independent, normally distributed random variables is again a normally distributed random variable. Hence

$$\langle \mathbf{a}(\mathbf{x}), \mathbf{v} \rangle \sim \mathcal{N}\left(0, \|\mathbf{a}(\mathbf{x})\|_2^2\right).$$

Therefore

$$\begin{aligned} P\left(g(\mathbf{x}, \mathbf{v}) \geq 0\right) &= P\left(\|\mathbf{a}(\mathbf{x})\|_2 Z \geq 0\right) \\ &= P\left(Z \geq \frac{\mathbf{b}(\mathbf{x})}{\|\mathbf{a}(\mathbf{x})\|_2}\right) \\ &= \Phi\left(\frac{-\mathbf{b}(\mathbf{x})}{\|\mathbf{a}(\mathbf{x})\|_2}\right), \end{aligned} \tag{4.5}$$

whenever $p(\mathbf{x}) \leq 0.5$ and Z has is standard normally distributed. Let the origin in $\mathbb{R}^k$ be denoted $\mathbf{0}$. The closest point $\mathbf{p}$ on P to $\mathbf{0}$ is the projection of $\mathbf{0}$ onto P.

We are interested in the distance from $\mathbf{0}$ to $\mathbf{p}$, which we call d. It is well-known from analytic geometry that d is given by

$$\frac{|\mathbf{b}(\mathbf{x})|}{\|\mathbf{a}(\mathbf{x})\|_2}.$$

Then, since d is equal to $\beta(\mathbf{x})$, it follows that

$$\Phi(-d) = \phi\left(-\beta(\mathbf{b}(\mathbf{x})\right) = P\left(g(\mathbf{x}, \mathbf{v}) \geq 0\right),$$

whenever $P(\mathbf{x}) \leq 0.5$, which is what we wanted to show. ■

Corollary 4.2.7 (✻)**.** *If the limit state function g is convex in $\mathbf{v}$, the approximation*

$$\hat{p}(\mathbf{x}) = \Phi(-\beta)\mathbf{x})) \tag{4.6}$$

is an upper bound on $p(\mathbf{x})$ whenever $p(\mathbf{x}) \leq 0.5$

Proof. Let

$$S = \partial\{\mathbf{v} \mid g(\mathbf{v}, \mathbf{x}) = 0\} \tag{4.7}$$

denote the boundary of the failure set. Since g is convex in $\mathbf{v}$, its level sets are by Lemma 1.1.17 again convex sets. Let $\mathbf{v}^*$ a be point in S which minimizes the distance from S to the origin in $\mathbb{R}^k$, and let the distance be denoted $\beta(\mathbf{x})$. By the supporting hyperplane theorem, there exists a supporting hyperplane P containing $\mathbf{v}^*$. Let P^+ denote the upper halfspace, which contains all $\mathbf{v}$ corresponding to a failed state. That is,

$$\mathbf{v} \in \{\mathbf{v} \mid g(\mathbf{v}, \mathbf{x}) > 0\} \implies \mathbf{v} \in P^+.$$

This implies that

$$\begin{aligned} P\left(\mathbf{v} \in \{\mathbf{v} \mid g(\mathbf{v}, \mathbf{x}) > 0\}\right) &= P\left(g(\mathbf{v}, \mathbf{x}) > 0\right) \\ &\leq P(\mathbf{v} \in P^+) \\ &\underset{(i)}{=} \Phi(-\beta(\mathbf{x})), \end{aligned}$$

using the result from Proposition 4.2.6 in (i). This concludes the proof. ■

It is worth noting that when g is not convex in $\mathbf{v}$ it can lead to computational and analytical problems. As is pointed out in [RR10] a lack of convexity might complicate the task of determining the closes point on the surface of the failure region to the origin in $\mathbb{R}^k$. Indeed, numerical methods might identify points that are locally closest to the origin, but fail to identify the globally closest ones. This can lead to errors in the estimation of $\beta(\mathbf{x})$ which influences the estimate of the failure probability. Another concern with non-convexity is the risk that $\Phi\big(-\beta(\mathbf{x})\big)$ incorrectly estimates the failure probability. We illustrate this issue in Example 4.2.8. Another concern is the assumption on the distribution of $\mathbf{V}$. While the assumption of normality may hold for many practical applications, it may not always be appropriate. Without any assumption on the distribution of $\mathbf{V}$, we cannot derive any bounds on the failure probability by means of the

reliability index. However, to overcome this issue, one can attempt apply a bijective transformation Ψ to $\mathbf{V}$ such that $\Psi(\mathbf{V})$ has the desired distribution [Ros52]. For this approach to be successful, it will however require that the transformed limit state function $(\Psi \circ g)(\mathbf{x}, \mathbf{v})$ is again affine or convex. This may not always be the case.

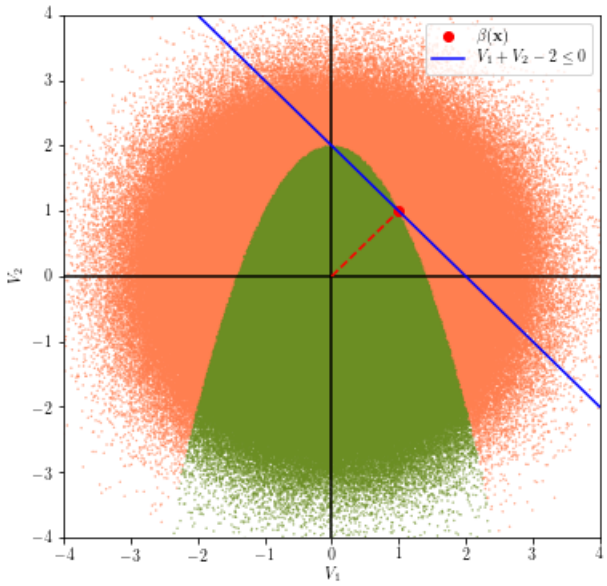


Figure 4.2: Monte-Carlo sampling from f. Green points indicate a safe state, red points a failed state. The limit state function is $V_1 \cdot x_1 + V_2^2 \cdot x_2 - 2$, where $\mathbf{x} = (1, 1)$. The blue line is equal to that in Figure 4.2. The red dotted line indicates the shortest distance to the boundary of the set $\{\mathbf{v} \mid V_1 + V_2^2 - 2 \leq 0\}$. The failure probability is approximately 21%. The probability of failure is larger than the probability of selecting a point above the blue line.

Example 4.2.8 (✻). Let $\mathbf{x} = (1, 1)$ be given, and the limit state function g_3 be defined as

$$g_3(\mathbf{x}, \mathbf{v}) = -2 + x_1 v_1 + x_2 v_2^2. \tag{4.8}$$

Using Monte Carlo sampling with $2.5M$ samples, we find an estimated failure probability of

$$p(\mathbf{v}, \mathbf{x}) \approx 21\%.$$

Meanwhile, $\beta(\mathbf{x}) = \sqrt{2}$, which is the same as the case with the limit state function in Figure 4.1 with $\mathbf{x} = (1, 1)$. The failure probability in 4.1 was calculated to approximately 7.86%, and was given as the probability that a sample lies above the indicated line. In comparison, we see that there is a large number of failed states which lie below the indicated blue line in Figure 4.2, and it is clear that $\Phi(-\beta(1, 1))$ vastly underestimates the failure probability.

4.3 Buffered failure probability

It can perhaps be argued that examples illustrating some of the shortcoming of the failure probability in the previous example were artificial. Nevertheless, they illustrate that the failure probability can sometimes exhibit undesirable properties. In light of this, [RR10] proposes the *buffered failure probability* as an alternative to the failure probability, which rectifies many of the shortcomings of the failure probability. We will follow the assumption in the article that the cumulative distribution function of the limit state function is continuous and strictly increasing for any choice of $\mathbf{x}$. This ensures that any the quaintly for any suitable α is given by the inverse of the cumulative distribution function of the limit state function at level α.
Recall Definiton 1.3.12 of the quantile of a random variable. By the assumption that the cumulative distribution $F(g(\mathbf{x}, \mathbf{V}))$ is monotonically increasing and continuous, it follows that for $\alpha \in (0, 1)$,

$$q_\alpha \left(g(\mathbf{x}, \mathbf{V})\right) = F^{-1}_{g(\mathbf{x},\mathbf{V})}(\alpha). \tag{4.9}$$

If we choose α_0 to be the (unique) value such that $q_{\alpha_0} = 0$, it is obvious that

$$0 = q_{\alpha_0} = F^{-1}_{g(\mathbf{x},\mathbf{V})}(\alpha_0), \tag{4.10}$$

which means

$$\begin{aligned} p(\mathbf{x}, \mathbf{V}) &= 1 - P(g \leq 0) && (4.11)\\ &= 1 - F_{g(\mathbf{x},\mathbf{V})}(F{-}1_{g(\mathbf{x},\mathbf{V})}(\alpha_0)) && (4.12)\\ &= 1 - \alpha_0. && (4.13) \end{aligned}$$

This is to say that the failure probability is given as 1 minus the alpha-level which makes the quantile attain the value zero. To avoid any confusion, we introduce the notion of the superquantile. However, this is nothing new, but simply the definition of CVaR from Defition 1.3.18 applied to the limit state function $g(\mathbf{x}, \mathbf{V})$. To lighten the notation, we use the following definition.

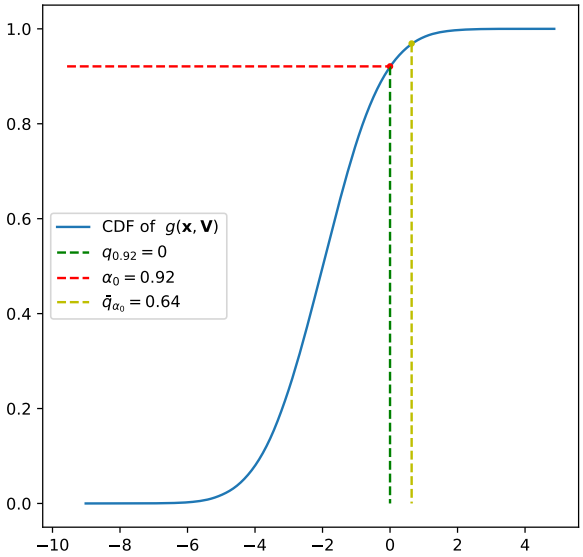
Definition 4.3.1 (Superquantile)**.** The superquantile at level α of the random variable $g(\mathbf{x}, \mathbf{V})$, denoted $\bar{q}_\alpha(\mathbf{x})$, is defined as

$$\bar{q}_\alpha(\mathbf{x}) \coloneqq E\left[g(\mathbf{x}, \mathbf{V}) \mid g(\mathbf{x}, \mathbf{V}) > q_\alpha(\mathbf{x})\right].$$

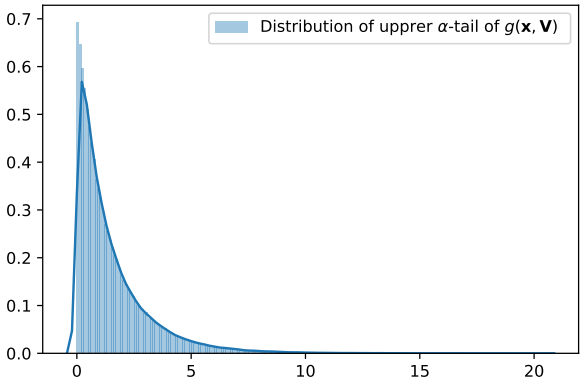
Example 4.3.2 (✻Illustrates superquantiles)**.** Let the limit state function be defined as

$$g(\mathbf{x}, \mathbf{V}) = V_1 x_1 + V_2 x_2 - 2,$$

where $\mathbf{x} = (1, 1)$. In Figure 4.3a we have indicated the quantile and the superquantile at level α_0 on the cumulative distribution of $g(\mathbf{x}, \mathbf{V})$. The cumulative distribution, the quantile and the superquantile were all estimated by Monte-Carlo sampling with $2.5M$ samples, although it is possible to calculate analytically. We observe that the superquantile is indeed larger than the quantile, with. $\bar{q}_{0.92}(\mathbf{x}) = 0.64$. We can interpret the superquantile at level α_0 as the average performance of the structure, given that the structure is in a failed state. Figure 4.3b illustrates the estimated probability density function of $g(\mathbf{x}, \mathbf{V})$, conditional on the event that $g(\mathbf{x}, \mathbf{V}) > 0$, again by by Monte-Carlo estimation, calculated on the same samples as Figure 4.3a.



(a) The level α_0 is chose such that $q_{\alpha_0}(\mathbf{x}) = 0$. This corresponds to $\alpha_0 = 0.92$, and hence $p(\mathbf{x}) = 1 - 0.92 = 0.08$. The superquantile at level $\alpha_0 = 0.92$ is $\bar{q}_{0.92}(\mathbf{x}) = 0.64$.



(b) A Monte-Carlo estimate of the probability density function of $g(\mathbf{x}, \mathbf{V})$, conditional on $g(\mathbf{x}, \mathbf{V}) \geq 0$

From Proposition 1.3.19, we know that CVaR is a more conservative risk measure than VaR, which is to say that $q_\alpha(\mathbf{x}) \leq \bar{q}_\alpha(\mathbf{x})$ in general. Where the failure probability is associated with the quantile $q_\alpha(\mathbf{x}) = 0$, the buffered failure probability is associated with the superquantile where $\bar{q}_\alpha(\mathbf{x}) = 0$.

Definition 4.3.3 (Buffered Failure Probability)**.** Let α_0' be selected such that

$\bar{q}_\alpha(\mathbf{x}) = 0$. Then, the buffered failure probability, denoted $\bar{p}(\mathbf{x})$ is given by

$$\bar{p}(\mathbf{x}) \coloneqq 1 - \alpha_0',$$

or alternatively,

$$P\left(g(\mathbf{x}, \mathbf{V}) \geq q_{\alpha_0'}(\mathbf{x})\right).$$

The next result shows that the buffered failure probability is more conservative than the failure probability.

Proposition 4.3.4. *For any* $\mathbf{x}$ *and* $\alpha \in (0, 1)$ *we have that* $p(\mathbf{x}) \leq \bar{p}(\mathbf{x})$

Proof. Let α_0 denote the alpha-level such that $\bar{q}_{\alpha_0}(\mathbf{x}) = 0$.

Proposition 1.3.19 shows that $\bar{q}_\alpha(\mathbf{x}) \leq q_\alpha(X)$ for any α, so in particular $\bar{q}_{\alpha_0}(\mathbf{x}) \leq \bar{q}_{\alpha_0}$, which gives

$$\begin{aligned}\bar{p}(\mathbf{x}) &= P\left(g(\mathbf{x}, \mathbf{V}) \geq \bar{q}_{\alpha_0}\right) \\ &\geq P\left(g(\mathbf{x}, \mathbf{V}) \geq q_{\alpha_0}(\mathbf{x})\right) \\ &= p(\mathbf{x}).\end{aligned}$$

This shows what $\bar{p}(\mathbf{x})$ is in fact more conservative than $p(\mathbf{x})$. ■

As the name buffered failure probability indicates, $\bar{p}(\mathbf{x})$ introduces in a sense a safety buffer. While the failure probability identifies the probability that the structure fails, i.e, $g(\mathbf{x}, \mathbf{V}) > 0$, the buffered failure probability is chosen such that while $g(\mathbf{x}, \mathbf{V})$ exceeds the value corresponding to the buffered failure probability, the structure remain safe on average. To estimate the buffered failure probability numerically, we need to solve

$$\bar{q}_\alpha(g(\mathbf{x}, \mathbf{V})) \stackrel{!}{=} 0$$

with respect to α. Since $g(\mathbf{x}, \mathbf{V})$ has a continuously and monotonly increasing cumulative distribution function by assumption, the expression can be evaluated as

$$\bar{q}_\alpha(g(\mathbf{x}, \mathbf{V})) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(g(\mathbf{x}, \mathbf{V}))d\beta \stackrel{!}{=} 0. \tag{4.14}$$

All in all, finding the buffered failure probability of $g(\mathbf{x}, \mathbf{V})$ simplifies to solving the equation

$$\int_\alpha^1 q_\beta(g(\mathbf{x}, \mathbf{V}))d\beta \stackrel{!}{=} 0 \tag{4.15}$$

with respect to α. This can be easily achieved, for instance by the means of Monte Carlo sampling. This is illustrated in Algorithm 0.

Algorithm 0: ✻Buffered failure probability

Generate N samples $\{x_i\}_{i=1}^N$ from the distribution of X;
Sort samples in increasing order;
S $\leftarrow 0$;
$i \leftarrow 0$;
while $S \geq 0$ *and* $i \leq N$ **do**
  $i \leftarrow i + 1$;
  $S \leftarrow S + x_{N-i}$;
end
if $S > 0$ **then**
  $\bar{p}(X) \leftarrow 1$;
else
  $\bar{p}(X) \leftarrow \frac{i}{N}$;
end

As an immediate consequence of $\bar{q}_\alpha(X) > q_\alpha(X)$ we have that the buffered failure probability is more conservative than the failure probability:

$$\bar{p}(\mathbf{x}) \geq p(\mathbf{x}). \tag{4.16}$$

In this regard, the buffered failure probability indeed provides a safety buffer in the sense that some states with $g(\mathbf{x}, \mathbf{V}) < 0$ will be above the value $q_{\alpha_0}(g(\mathbf{x}, \mathbf{V}))$, when $\bar{q}_{\alpha_0}(g(\mathbf{x}, \mathbf{V})) = 0$. We illuminate this in the next example.

Example 4.3.5 (✻). Let the environmental variables $\mathbf{V} = V_1, V_2$ be normally distributed with $\mu = -2, \sigma^2 = 1$ and Weibull-distributed with shape parameter $k = 1.5$ respectively. The design variables are $\mathbf{x}(1, 2)^T$. The limit state function is

$$g(\mathbf{x}, \mathbf{V}) = -3.5 + x_1 V_1 + x_2 V_2 \tag{4.17}$$

We use Monte Carlo sampling to estimate the failure probability and the buffered failure probability. The buffered failure probability is computed according to Algorithm 0. Using 2.5M samples, the estimated failure probability was 2.03% and the buffered failure probability was 5.26%. The latter corresponds to the structure having an average performance of 0 when $g(\mathbf{x}, \mathbf{V}) \geq -0.9348$. Figure 4.5 illustrates the distribution of $g(\mathbf{x}, \mathbf{V})$ for different realizations of $\mathbf{V}$, and shows how the buffered failure probability introduces a "buffer".

As we have seen, the buffered failure probability provides a more conservative assessment of which states are safe. Since it takes into account the performance of the structure also in failed states, it is sensitive to severe outcomes in the positive tail of the distribution of $g(\mathbf{x}, \mathbf{V})$. This may allow a designer to distinguish between designs where severity of the outcomes are important. In the next section, we analyse how buffered failure probability may be used in design optimization.

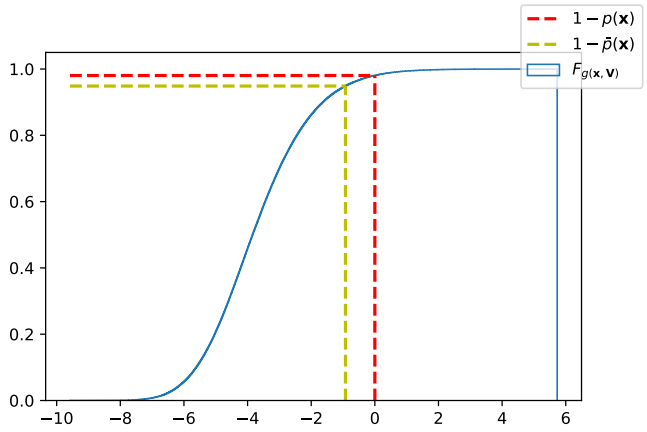


Figure 4.4: Monte-Carlo sampling from the distribution of $g(\mathbf{x}, \mathbf{V})$ with design- and environmental variables as described in Example 4.3.5. The yellow line indicates the buffered failure probability and its associated quantile, while the red line indicates the failure probability and its associated quantile.

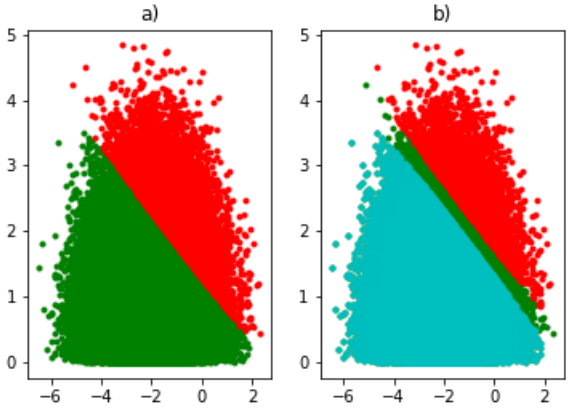


Figure 4.5: a) Red dots represent failed states. Green dots are "buffer" states. Here, $0 < g(\mathbf{x}, \mathbf{V}) \leq -0.9348$, corresponding to the quantile for the buffered failure probability. The figure in b) similarly show the failed and safe states, according to $g(\mathbf{x}, \mathbf{V})$

4.4 Design optimization with failure probability and buffered failure probability

The following is also based on [RR10] to a large extent. Failure probability and buffered failure probability can be useful tools in optimization of structural design. In this context, a designer wants to identify a design $\mathbf{x}$ or a class of designs by optimizing some function of the design, for instance the cost, which complies with a suitable upper bound on the failure probability or buffered

failure probability.

In the following, we assume that $\mathbb{X} \subseteq \mathbb{R}^k$ is a set of feasible design choices, and the function $f : \mathbb{X} \to \mathbb{R}$ which we with to minimize, has continuous partial derivatives. We will think of f as the cost of the design $\mathbf{x}$, but the function could be adapted to suit the needs of a practical application. For instance, penalties on physical attributes of the design, such as weight, could be included in the cost function, or alternatively as additional constraints.

If α is the desired upper bound on the failure probability, the problem to optimize the cost f of a feasible design x, which we denote $\mathbf{P}$, can be formulated as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & p(\mathbf{x}) \leq \alpha_0, \\ & \mathbf{x} \in \mathbb{X}. \end{aligned}$$

While this problem is easy to state and understand, it may have several problematic aspects which makes it difficult to work with in practical application. In particular, it can be the case that while $g(\mathbf{x}, \mathbf{V})$ is convex in $\mathbf{x}$ for any realization $\mathbf{v}$ of $\mathbf{v}$, $p(\mathbf{x})$ remains non-convex. This is illustrated in the following example.

Example 4.4.1 (✻)**.** Let the limit state function $g(\mathbf{x}, \mathbf{V})$ be defined as

$$g(x, v) = -x + v. \tag{4.18}$$

Then, g is a affine function, and in particular, it is convex in x for any v. In the meantime, assume that the environmental variable V has a standard normal distribution. Thus, the failure probability can be calculated as

$$p(x) = P(-x + V \geq 0) \tag{4.19}$$

$$= 1 - P(V \leq x). \tag{4.20}$$

Inserting for the probability density of a standard normal random variable in the last line yields

$$p(x) = 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \tag{4.21}$$

Clearly, p is twice differentiable with respect to x, and it is hence a sufficient condition for convexity that $\frac{\partial^2}{\partial x^2} p \geq 0$ for all x. However, this is not the case:

$$\frac{\partial^2}{\partial x^2} p = \frac{\partial^2}{\partial x^2} 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{4.22}$$

$$= \frac{\partial}{\partial x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{4.23}$$

$$= -x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{4.24}$$

which takes negative values for negative values of x. This shows that p is indeed not convex in x, even though g is.

The previous example underlines an important point. If the optimization constraint is not convex, it will result in a much harder optimization problem. We now compare the previous optimization problem to a similar one, where we use buffered failure probability constraints.

Optimization with buffered failure probability constraints

In the following we review a proposed method from [RR10] for using buffered failure probability in design optmization, and investigate how we may use the minimization formula for conditional Value-at-Risk to formulate a convex minimization problem. We denote the optimization problem with buffered failure probability constraints **BP**:

$$\min_{x} \quad f(\mathbf{x}) \tag{4.25}$$

$$\text{s.t} \quad \bar{p}(\mathbf{x}) \leq 1 - \alpha_0, \tag{4.26}$$

$$\mathbf{x} \in \mathbb{X}. \tag{4.27}$$

Observing that

$$\bar{p}(\mathbf{x}) \leq 1 - \alpha_0 \Leftrightarrow \bar{q}_{\alpha_0}(\mathbf{x}) \leq 0, \tag{4.28}$$

we can reformulate **BP** using Proposition 1.3.19 (b). This gives the following reformulation. **BP'**:

$$\min_{\mathbf{x},c} \quad f(\mathbf{x}) \tag{4.29}$$

$$\text{s.t } c + \frac{1}{1-\alpha} E[\max[0, g(\mathbf{x}, \mathbf{v}) - c]] \leq 0, \tag{4.30}$$

$$\mathbf{x} \in \mathbb{X}, c \in \mathbb{R}. \tag{4.31}$$

Although this problem cannot always be solved analytically, we can take a numeric approach. If we can sample from the joint distribution of $\mathbf{v}$, we can approximate the superexpectation $E[\max[0, g(\mathbf{x}, \mathbf{v}) - c]]$ numerically, which gives the following problem:
BP'$_\mathbf{N}$:
Draw a large sample $\{\mathbf{v}_i\}_{i=1}^N$,

$$\min_{\mathbf{x},c} \quad f(\mathbf{x}) \tag{4.32}$$

$$\text{s.t } c + \frac{1}{N(1-\alpha)} \sum_{j=1}^{N} \max[0, g(\mathbf{x}, \mathbf{v}_j) - c] \leq 0, \tag{4.33}$$

$$\mathbf{x} \in \mathbb{X}, c \in \mathbb{R}. \tag{4.34}$$

Furthermore, by introducing variables

$$c_1, c_2, \ldots, c_N,$$

setting

$$c_j = \max[0, g(\mathbf{x}, \mathbf{v}_j) - c],$$

and denoting

$$\mathbf{c} \coloneqq (c, c_1, c_2, \ldots, c_N),$$

BP'$_\mathbf{N}$ may be reformulated as

$$\min_{\mathbf{x},\mathbf{c}} \quad f(\mathbf{x}) \tag{4.35}$$

$$\text{s.t } c + \frac{1}{N(1-\alpha)} \sum_{j=1}^{N} c_j \leq 0, \tag{4.36}$$

$$\max[0, g(\mathbf{x}, \mathbf{v}_j) - c] = c_j, \tag{4.37}$$

$$\mathbf{x} \in \mathbb{X},\ \mathbf{c} \in \mathbb{R}^{N+1}. \tag{4.38}$$

We may further relax the condition

$$c_j = \max[0, g(\mathbf{x}, \mathbf{v}_j) - c]$$

to

$$c_j \geq \max[0, g(\mathbf{x}, \mathbf{v}_j) - c],$$

since having $c_j > \max[0, g(\mathbf{x}, \mathbf{v}_j) - c]$ for any $j = 1, \ldots, N$ will not improve the objective value. In doing so, the constraint

$$c_j \geq \max[0, g(\mathbf{x}, \mathbf{v}_j) - c],$$

may be decomposed into the two constraints

$$g(\mathbf{x}, \mathbf{v}_j) - c \leq c_j, \quad c_j \geq 0.$$

This finally yields the final equivalent optimization problem:
BP$_\mathbf{N}$:

$$\min_{\mathbf{x},\mathbf{c}} \quad f(\mathbf{x}) \tag{4.39}$$

$$\text{s.t } c + \frac{1}{N(1-\alpha)} \sum_{j=1}^{N} c_j \leq 0, \tag{4.40}$$

$$g(\mathbf{x}, \mathbf{v}_j) - c \leq c_j, \tag{4.41}$$

$$c_j \geq 0, \tag{4.42}$$

$$\mathbf{x} \in \mathbb{X},\ \mathbf{c} \in \mathbb{R}^{N+1}. \tag{4.43}$$

One immediate benefit of solving the optimization problem **BP$_\mathbf{N}$** in comparison to **P** is that **BP$_\mathbf{N}$** is more computationally tractable. As we have seen, **P** does not guarantee a convex optimization problem even when g was a convex function. Meanwhile, **BP$_\mathbf{N}$** is a convex optimization problem provided $g(\mathbf{x}, \mathbf{v}_j)$ are convex functions and $\mathbb{X}$ is a convex set, in which case **BP$_\mathbf{N}$** is solvable by standard optimization algorithms [RR13].

CHAPTER 5

Concluding remarks

We have in this thesis studied risk measures and optimization under uncertainty. In the first chapter, we considered theoretical aspects of various risk measures, discussed their advantages and disadvantages, and reviewed coherency in the sense of risk analysis. Of particular focus was the study of conditional Value-at-Risk, which has many attractive qualities. After laying the theoretical groundwork, we studied how concepts from risk analysis applies to different problem domains.

In Chapter 2 we studied how an insurer may reduce their risk-loading by the means of reinsurance contracts. We reviewed optimality conditions for the case of reinsuring a single risk, together with newly developed optimality conditions for multiple risks when the risk measure Value-at-Risk is used. We then aimed to develop similar results in the case of conditional Value-at-risk. The centerpiece of the discussion in this chapter was Theorem 2.3.7, which identifies optimal stop-loss contract parameters when risks have a worst-case dependency. We then examined how such solutions compared to those found using simulation and numerical optimization. In the examples studied, we were not able to improve significantly on solutions yielded by Theorem 2.3.7, even when the risks had weaker dependency.

Unfortunately, we were not successful in deriving similar results for multivariate risks with respect to conditional Value-at-Risk, tailored to other dependency structures among the risk. This question remains open for future work. Moreover, the problem of optimal reinsurance in the multivariate case under other coherent risk measures also needs further study. It is known that conditional Value-at-risk is a fundamental building block for the general classes of law-invariant risk measures and convex risk measures. Could results similar to those derived in this chapter be extended to such classes of risk measures?

In Chapter 3 the focus was on the application of Value-at-Risk and conditional Value-at-Risk to estimate stock market risk in the near future. We discovered that Value-at-Risk was far less risk averse than conditional Value-at-Risk through concrete examples. We then used a different technique to estimate the average Value-at-Risk, by estimating the parameters of a geometric Brownian motion from historic data, and using this process to estimate the risk, with adequate results.

To showcase the versatility of risk analysis, we investigated in Chapter 4 the use of risk analysis in the optimization of structural designs. We discussed first some aspect of the commonly used tool failure probability, which corresponded

to the risk measure Value-at-Risk. We then studied how the approach from [RR13] remedied some of these drawbacks. Here, the central element was the use of superquantiles, and their counterpart buffered failure probability. These share strong ties to conditional Value-at-Risk, and we discussed how using these lead to preferable optimization problems. However, the analysis in this chapter create an abstraction over the real world; we describe the state of a structure by a mapping g, from the design choice together with the state of its environment, to a number describing its performance. In a practical application, such a function is likely not known perfectly. We may only be able to observe measurements of g for given pairs of $\mathbf{x}$ and $\mathbf{v}$. In practice, we may need to create approximations to this function, inferred from laboratory tests. One interesting problem for future work could be to combine machine learning approaches for the approximation of g with concepts from risk analysis to be used in the optimization of structural design.

Appendices

APPENDIX A

Appendix

Listing A.1: Chapter 2, necessary functions.

```
import numpy as np
from tqdm import tqdm
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from statsmodels.distributions.empirical_distribution import ECDF
import seaborn as sns
#Common problem parameters
N = 100000
alpha = 0.95
theta = 1/9


def empirical_quantile(X, alpha):
    """ X : np.array
        alpha: probability level in (0, 1)

        Returns: the empirical quantile at level alpha.
    """
    Y = np.sort(X)
    index = np.ceil(Y.shape[0] * alpha).astype(int)
    return Y[index]

def empirical_cvar(X, alpha):
    """ X : np.array
        alpha: probability level in (0, 1)

        Returns: the empirical conditional value at risk, at level alpha.
    """
    quantile = empirical_quantile(X, alpha)
    return np.mean(X[np.where(X>=quantile)])

def S(X, x):
    """
    X : np.array
    x : value of X

    Returns: Probability that X > x.
    """
    Y = np.sort(X)
    return (X.shape[0] - np.where(Y>x)[0][0])/X.shape[0]


def _I(a, b, X):
```

```
    """
    Helper function which returns the retained risk I for an \

    insurance layer contract with parameters a, b, for the risk X.
    If b = np.inf, it is a stop-loss reinsurance contract.
    X: np.array.
    a: Contract praramter a
    b: Contract paramter b

    Returns: I(X) given a, b.
    """
    I = X.copy()
    I[(a <= I) & (I <= b)] = a
    I[I>b] = I[I>b]-(b-a)
    return I

def reinsurance_risk_value(risk_measure, risks, a_s, b_s, alpha, theta):
    """
    If b_i = np.inf, then reinsurance contract no. i is a stop-loss contract.
    Otherwise it is an insurance layer contract.

    Parameters:
    a_s: list of contract parameters a_1, ..., a_m
    b_s: list of contract parametrs b_1, ..., b_m.
    theta: Risk premium
    alpha: Probability level for the risk optimization.

    riks: list of realizations of X_i, where each X_i is of type np.array.

    Returns:
    The risk value rho_alpha(sum(I_i(X_i) + (1+theta)E[R_i(X_i)])) given contract parameters.
    """
    I = []
    R = []
    for a_i,b_i, X_i in zip(a_s, b_s, risks):
        I.append(_I(a_i, b_i, X_i))
        R.append(X_i - I[-1])
    return risk_measure(np.sum(I, axis = 0), alpha) + np.mean(np.sum(R, axis = 0))*(1+theta)


def optimize_gridsearch_a_param(X_1, X_2, a_min = (0, 0), \
a_max = (100, 100), max_iter = 10, tol = 10e-10, verbose = True):
    def f_bivariate(params):
        """A wrapper function to be with optimization packages.
        Requires X_1, X_2 to exist as global variables.
        """
        a_1 = params[0]
        a_2 = params[1]
        b_1 = params[2]
        b_2 = params[3]
        return reinsurance_risk_value(empirical_cvar, [X_1, X_2], \
        [a_1, a_2], [b_1, b_2], alpha = alpha, theta = theta)

    iterations = 1
    diff = np.inf
    max_iter = 30
    #Initialize
    _X = np.linspace(0, max(X_1), 10)
    _Y =  np.linspace(0, max(X_2), 10)
    dx = _X[1]-_X[0]
    dy = _Y[1]-_Y[0]
    X, Y = np.meshgrid(_X, _Y)
```

```
    Z = np.array([f_bivariate( (a_1 , a_2, np.inf, np.inf))  for a_1, a_2 in
zip(X.ravel(), Y.ravel())]).reshape(X.shape)
    current_best = np.min(Z)
    while diff > tol and iterations < max_iter:
        index = np.where(Z == np.min(Z))
        _a1 = X[index[0][0], index[1][0]]
        _a2 = Y[index[0][0], index[1][0]]

        _X = np.linspace(_a1-dx, _a1+dx, 10)
        _Y =  np.linspace(_a2-dy, _a2+dy, 10)
        dx = _X[1]-_X[0]
        dy = _Y[1]-_Y[0]
        X, Y = np.meshgrid(_X, _Y)
        Z = np.array([f_bivariate( (a_1 , a_2, np.inf, np.inf))  for a_1, a_2 in
zip(X.ravel(), Y.ravel())]).reshape(X.shape)
        diff = abs(np.min(Z) - current_best)
        current_best = np.min(Z)
        if verbose:
            print(f"Iteration:  {iterations}, current value:  \
            {current_best}, (a_1, a_2) : ({_a1}, {_a2}), difference from last iteration: {diff}")
        iterations += 1
    if verbose:
        print(f"Best parameters found: a_1 = {_a1}, a_2 = {_a2} \
        with  optimal value {f_bivariate((_a1, _a2, np.inf,  \
        np.inf))} after {iterations} iterations.")
    return _a1, _a2, current_best


def optimize_gridsearch_b_param(X_1, X_2, a = (0, 0), b_min = \
(0, 0), b_max = (100, 100), max_iter = 10, tol = 10e-10, \
verbose = True):
    def f_bivariate(params):
        """A wrapper function to be with optimization packages.
        Requires X_1, X_2 to exist as global variables.
        """
        a_1 = params[0]
        a_2 = params[1]
        b_1 = params[2]
        b_2 = params[3]
        return reinsurance_risk_value(empirical_cvar, [X_1, X_2], \
        [a_1, a_2], [b_1, b_2], alpha = alpha, theta = theta)

    iterations = 1
    diff = np.inf
    max_iter = 30
    #Initialize
    _X = np.linspace(b_min[0], b_max[0], 10)
    _Y =  np.linspace(b_min[1], b_max[1], 10)
    dx = _X[1]-_X[0]
    dy = _Y[1]-_Y[0]
    X, Y = np.meshgrid(_X, _Y)
    Z = np.array([f_bivariate( (a[0] , a[1], b_1, b_2))  for b_1, b_2 in
zip(X.ravel(), Y.ravel())]).reshape(X.shape)
    current_best = np.min(Z)
    while diff > tol and iterations < max_iter:
        index = np.where(Z == np.min(Z))
        _b1 = X[index[0][0], index[1][0]]
        _b2 = Y[index[0][0], index[1][0]]

        _X = np.linspace(_b1-dx, _b1+dx, 10)
        _Y =  np.linspace(_b2-dy, _b2+dy, 10)
        dx = _X[1]-_X[0]
```

```
        dy = _Y[1]-_Y[0]
        X, Y = np.meshgrid(_X, _Y)
        Z = np.array([f_bivariate( (a[0] , a[1], b_1, b_2))  for b_1, b_2 in
zip(X.ravel(), Y.ravel())]).reshape(X.shape)
        diff = abs(np.min(Z) - current_best)
        current_best = np.min(Z)
        if verbose:
            print(f"Iteration:  {iterations}, current value: \
            {current_best}, (b_0, b_1) : ({_b1}, {_b2}), \
            difference from last iteration: {diff}")
        iterations += 1
    if verbose:
        print(f"Best parameters found: b_1 = {_b1}, b_2 = {_b2} \
        with optimal value {f_bivariate((a[0], a[1], _b1, _b2))} \
        after {iterations} iterations.")
    return _b1, _b2, current_best

def loss_surface_b_param(X_1, X_2, a_1, a_2, n_points = (30, 30)):
    def f_bivariate(params):
        """A wrapper function to be with optimization packages.
        Requires X_1, X_2 to exist as global variables.
        """
        a_1 = params[0]
        a_2 = params[1]
        b_1 = params[2]
        b_2 = params[3]
        return reinsurance_risk_value(empirical_cvar, [X_1, X_2],\
        [a_1, a_2], [b_1, b_2], alpha = alpha, theta = theta)

    _X = np.linspace(a_1, max(X_1), n_points[0])
    _Y =  np.linspace(a_2, max(X_2), n_points[1])
    dx = _X[1]-_X[0]
    dy = _Y[1]-_Y[0]
    X, Y = np.meshgrid(_X, _Y)
    Z = np.array([f_bivariate( (a_1 , a_2, b_1, b_2))  \
    for b_1, b_2 in  zip(X.ravel(), Y.ravel())])
    Z = Z.reshape(X.shape)
    fig = plt.figure()
    ax = fig.add_subplot(111, projection='3d')
    Z = Z.reshape(X.shape)
    ax.plot_surface(X, Y, Z)

#Optimize over a_1, a_2, b_1, b_2, numerical example 1.
import numpy as np
from scipy.optimize import minimize
def optimize_with_LBFGSB(X_1, X_2, initial_guesses, \
max_evaluations = 10000, verbose = True):

    def f_bivariate(params):
        """A wrapper function to be with optimization packages.
        Requires X_1, X_2 to exist as global variables.
        """
        a_1 = params[0]
        a_2 = params[1]
        b_1 = params[2]
        b_2 = params[3]
        return reinsurance_risk_value(empirical_cvar, [X_1, X_2],\
        [a_1, a_2], [b_1, b_2], alpha = alpha, theta = theta)

    optimal_values = []
    params = []
    for initial_guess in initial_guesses:
```

```
        result = minimize(f_bivariate, initial_guess, method='L-BFGS-B',
                          options={'maxfun':max_evaluations,
                                   'disp': True})
        #Only positive values are admissable.
        if result.fun >= 0:
            optimal_values.append(result.fun)
            params.append(result.x)
            if verbose:
                print(f"""Initial guess: {initial_guess}, \n
                        parameters found: {result.x}, \\
                        Objective value: {result.fun}""")
    if len(params)>0:
        best_value = min(optimal_values)
        index = optimal_values.index(best_value)
        return params[index], optimal_values[index]
    else:
        print("No feasible solutions found. Try again with other initial values.")
        return none
```

Listing A.2: Python example

```
# Example 2

def generate_correlated_lognormal_variables(corr):
    N = int(10E6)
    means = (0, 0)
    stds = (1, 1)
    # Setting up the covariance matrix
    covariance_matrix  = [[stds[0]**2          , stds[0]*stds[1]*corr],
                          [stds[0]*stds[1]*corr,           stds[1]**2]]
    X_1, X_2 = np.random.multivariate_normal(means, covariance_matrix, 100000).T
    # Finding the empirical cumulative distribution functions.
    ecdf_X = ECDF(X_1)
    ecdf_Y = ECDF(X_2)
    # Generating uniform random variables on [0, 1] with correlation = 0.5
    X_inv = ecdf_X(X_1)
    Y_inv = ecdf_Y(X_2)

    alpha = 0.95
    theta = 1/9
    m_1 = np.log(500)-np.log(101)/2
    s_1 = np.sqrt(np.log(101/100))
    m_2  = np.log(1000) - np.log(101)/2
    s_2  = np.sqrt(np.log(101/100))
    lognormal_1 = np.random.lognormal(m_1, s_1, N)
    lognormal_2 = np.random.lognormal(m_2, s_2, N)
    inv_lognormal = lambda X, x : np.quantile(X, x)
    X_1 = inv_lognormal(lognormal_1, X_inv)
    X_2 = inv_lognormal(lognormal_2, Y_inv)
    return X_1, X_2
```

Listing A.3: Example 2, Chapter 2

```
#Example 2
L_1, L_2 = generate_correlated_lognormal_variables(0.7)
plt.scatter(L_1, L_2, s = .3)
plt.xlabel("$X_1$")
plt.ylabel("$X_2$")
plt.savefig("scatterplot_lognormal_positive_correlation.png")
plt.figure()
sns.distplot(L_1, label = "$f_{X_1}$")
sns.distplot(L_2, label = "$f_{X_2}$")
```

```
plt.legend()
plt.savefig("distribution_lognormal_positive_correlation.png")
print("Correlation: ", np.corrcoef(L_1, L_2)[0, 1])

#step 1
value = reinsurance_risk_value(empirical_cvar, [L_1, L_2], \
[ a_1, a_2], [np.inf, np.inf], alpha = alpha, theta = theta)
print(value)

#Step 2
optimize_gridsearch_a_param(L_1, L_2, a_min = (0, 0), a_max = \
(max(L), max(M)))

#Step 3
optimize_gridsearch_b_param(L_1, L_2, a = (a_1, a_2), b_min = \
(a_1, a_2), b_max = (max(L_1), max(L_2)))

#Step 4
initial_guess_1 = (0, 0, 100, 100)
initial_guess_2 = (0, 0, 50, 100)
initial_guess_3 = (43, 87, 200, 100)
initial_guesses = [initial_guess_1,
                   initial_guess_2,
                   initial_guess_3]
a_1l = empirical_quantile(L_1, 1-(1/(1+theta)))
a_2l = empirical_quantile(L_2, 1-(1/(1+theta)))
loss_surface_b_param(L_1, L_2, a_1l, a_2l, (20, 20))
params, value = optimize_with_LBFGSB(L_1, L_2, initial_guesses)
```

Listing A.4: Chapter 2, Example 3

```
#Example 3
W_1, W_2 = generate_correlated_weibull_variables(-.7)
plt.scatter(W_1, W_2, s = .3)
plt.xlabel("$X_1$")
plt.ylabel("$X_2$")
plt.savefig("figures/scatterplot_beta_negative_correlation.png")
plt.figure()
sns.distplot(W_1, label = "$f_{X_1}$")
sns.distplot(W_2, label = "$f_{X_2}$")
plt.legend()
plt.savefig("figures/distribution_beta_negative_correlation.png")
print("Correlation: ", np.corrcoef(W_1, W_2)[0, 1])
#step 1
a_1 = empirical_quantile(W_1, 1-(1/(1+theta)))
a_2 = empirical_quantile(W_2, 1-(1/(1+theta)))
value = reinsurance_risk_value(empirical_cvar, [W_1, W_2], \
[ a_1, a_2], [np.inf, np.inf], alpha = alpha, theta = theta)
print("a_1, a_2, value:", a_1, a_2, value)
#Step 2
optimize_gridsearch_a_param(W_1, W_2, a_min = (0, 0), a_max = \
(max(W_1), max(W_2)), b =  ( np.inf, np.inf))

b_values_1 = np.linspace(a_2, 150, 400)
v_values_1 = [reinsurance_risk_value(empirical_cvar, [W_1, W_2],\
(a_1, a_2),(np.inf, b_2), alpha = alpha, theta = theta) for b_2 in b_values_1]
b_values_2 = np.linspace(69, 150, 400)
v_values_2 = [reinsurance_risk_value(empirical_cvar, [W_1, W_2], \
(a_1, a_2),(np.inf, b_2), alpha = alpha, theta = theta) for b_2 in b_values_2]
plt.figure()
plt.plot(b_values, v_values, label = "$V_{|\alpha}$")
```

```
plt.legend()
plt.xlabel("$b_1$")
plt.savefig("b_1_loss_example_3.pdf")
plt.figure()
plt.plot(b_values_2, v_values_2, label = "$V_{\\alpha}$")
plt.xlabel("$b_1$")
plt.legend()
plt.savefig("b_2_loss_example_3.pdf")
#Step 3
optimize_gridsearch_b_param(W_1, W_2, a =  (a_1, a_2), b_min = (a_1, a_2), b_max = (max(W_1), max(W_2)))
#Step 4
initial_guess_1 = (0, 0, 100, 100)
initial_guess_2 = (0, 0, 50, 100)
initial_guess_3 = (a_1, a_2, max(W_1)/2, max(W_2)/2)
initial_guess_4 = (np.mean(W_1), np.mean(W_2), max(W_1), max(W_2))

initial_guesses = [initial_guess_1,
                   initial_guess_2,
                   initial_guess_3]
a_1l = empirical_quantile(W_1, 1-(1/(1+theta)))
a_2l = empirical_quantile(W_2, 1-(1/(1+theta)))
loss_surface_b_param(W_1, W_2, a_1, a_2, (20, 20))
params, value = optimize_with_LBFGSB(W_1, W_2, initial_guesses)
```

Listing A.5: Chapter 2, Example 4

```
# Example 4
corr = -.3
N = int(2E5)
means = (100, 80)
stds = (25, 30)
# Setting up the covariance matrix
covariance_matrix  = [[stds[0]**2          , stds[0]*stds[1]*corr],
 [stds[0]*stds[1]*corr,           stds[1]**2]]
N_1, N_2 = np.random.multivariate_normal(means, covariance_matrix, N).T

print("Correlation: ", np.corrcoef(N_1, N_2)[0, 1])

#step 1
a_1 = empirical_quantile(N_1, 1-(1/(1+theta)))
a_2 = empirical_quantile(N_2, 1-(1/(1+theta)))
value = reinsurance_risk_value(empirical_cvar, [N_1, N_2], \
[ a_1, a_2], [np.inf, np.inf], alpha = alpha, theta = theta)
print("a_1, a_2, value:", a_1, a_2, value)

#Step 2
optimize_gridsearch_a_param(N_1, N_2, a_min = (0, 0), \
a_max = (max(N_1), max(N_2)), b =  ( np.inf, np.inf))
#Step 3
optimize_gridsearch_b_param(N_1, N_2, a =  (a_1, a_2), b_min = \
(a_1, a_2), b_max = (max(N_1), max(N_2)))
#Step 4
initial_guess_1 = (0, 0, 100, 100)
initial_guess_2 = (50, 50, 150, 150)
initial_guess_3 = (a_1, a_2, max(N_1)/2, max(N_2)/2)
initial_guess_4 = (np.mean(N_1), np.mean(N_2), max(N_1), \
max(N_2))

initial_guesses = [initial_guess_1,
                   initial_guess_2,
                   initial_guess_3]
```

```
params, value = optimize_with_LBFGSB(N_1, N_2, initial_guesses)
```

Listing A.6: Chapter 2, loss surface

```
import numpy as np
from tqdm import tqdm
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from statsmodels.distributions.empirical_distribution import ECDF

def empirical_quantile(X, alpha):
    """ X : np.array
        alpha: probability level in (0, 1)

        Returns: the empirical quantile at level alpha.
    """
    Y = np.sort(X)
    index = np.ceil(Y.shape[0] * alpha).astype(int)
    return Y[index]

def empirical_cvar(X, alpha):
    """ X : np.array
        alpha: probability level in (0, 1)

        Returns: the empirical conditional value at risk, at level alpha.
    """
    quantile = empirical_quantile(X, alpha)
    return np.mean(X[np.where(X>=quantile)])

def S(X, x):
    """
    X : np.array
    x : value of X

    Returns: Probability that X > x.
    """
    Y = np.sort(X)
    return (X.shape[0] - np.where(Y>x)[0][0])/X.shape[0]


def _I(a, b, X):
    """
    Helper function which returns the retained risk I for an \
    insurance layer contract with parameters a, b, for the risk X.

    If b = np.inf, it is a stop-loss reinsurance contract.
    X: np.array.
    a: Contract praramter a
    b: Contract paramter b

    Returns: I(X) given a, b.
    """
    I = X.copy()
    I[(a <= I) & (I <= b)] = a
    I[I>b] = I[I>b]-(b-a)
    return I

def reinsurance_risk_value(risk_measure, risks, a_s, b_s, alpha, theta):
    """
    If b_i = np.inf, then reinsurance contract no. i is a stop-loss contract. Otherwise it is an insura

    Parameters:
```

```
    a_s: list of contract parameters a_1, ..., a_m
    b_s: list of contract parametrs b_1, ..., b_m.
    theta: Risk premium
    alpha: Probability level for the risk optimization.

    riks: list of realizations of X_i, where each X_i is of type np.array.

    Returns:
    The risk value rho_alpha(sum(I_i(X_i) + (1+theta)E[R_i(X_i)])) given contract parameters.
    """
    I = []
    R = []
    for a_i,b_i, X_i in zip(a_s, b_s, risks):
        I.append(_I(a_i, b_i, X_i))
        R.append(X_i - I[-1])
    return risk_measure(np.sum(I, axis = 0), alpha) +\
    np.mean(np.sum(R, axis = 0))*(1+theta)



def f_bivariate(params):
    """A wrapper function to be with optimization packages. """
    a_1 = params[0]
    a_2 = params[1]
    b_1 = params[2]
    b_2 = params[3]
    return reinsurance_risk_value(empirical_cvar, \
    [X_1, X_2], [a_1, a_2], [b_1, b_2], alpha = alpha, theta = theta)


N = 100000
alpha = 0.95
theta = 1/9


def generate_correlated_lognormal_variables(corr):
    N = int(10E6)
    means = (0, 0)
    stds = (1, 1)
    # Setting up the covariance matrix
    covariance_matrix  = [[stds[0]**2          , stds[0]*stds[1]*corr],
                          [stds[0]*stds[1]*corr,           stds[1]**2]]
    X_1, X_2 = np.random.multivariate_normal(means, covariance_matrix, 100000).T
    # Finding the empirical cumulative distribution functions.
    ecdf_X = ECDF(X_1)
    ecdf_Y = ECDF(X_2)
    # Generating uniform random variables on [0, 1] with\
    correlation = 0.5
    X_inv = ecdf_X(X_1)
    Y_inv = ecdf_Y(X_2)

    alpha = 0.95
    theta = 1/9
    m_1 = np.log(500)-np.log(101)/2
    s_1 = np.sqrt(np.log(101/100))
    m_2  = np.log(1000) - np.log(101)/2
    s_2  = np.sqrt(np.log(101/100))
    lognormal_1 = np.random.lognormal(m_1, s_1, N)
    lognormal_2 = np.random.lognormal(m_2, s_2, N)
    inv_lognormal = lambda X, x : np.quantile(X, x)
    X_1 = inv_lognormal(lognormal_1, X_inv)
    X_2 = inv_lognormal(lognormal_2, Y_inv)
```

```
    return X_1, X_2



def generate_correlated_beta_variables(corr):
    N = int(10E6)
    means = (0, 0)
    stds = (1, 1)
    # Setting up the covariance matrix
    covariance_matrix  = [[stds[0]**2          , stds[0]*stds[1]*corr],
                          [stds[0]*stds[1]*corr,           stds[1]**2]]
    X_1, X_2 = np.random.multivariate_normal(means, covariance_matrix, 100000).T
    # Finding the empirical cumulative distribution functions.
    ecdf_X = ECDF(X_1)
    ecdf_Y = ECDF(X_2)
    # Generating uniform random variables on [0, 1] with correlation = 0.5
    X_inv = ecdf_X(X_1)
    Y_inv = ecdf_Y(X_2)

    s_1 = 75
    s_2 = 5.402
    W_1 = np.random.gamma(2, 20, N)
    W_2 = np.random.normal(50, 10, N)
    inv_weibull = lambda X, x : np.quantile(X, x)
    X_1 = inv_weibull(W_1, X_inv)
    X_2 = inv_weibull(W_2, Y_inv)
    return X_1, X_2



X_1, X_2 = generate_correlated_beta_variables(-0.7)
X, Y = np.meshgrid(np.linspace(0, max(X_1), 40), np.linspace(0, max(X_2), 40))
Z = np.array([f_bivariate( (a_1 , a_2, np.inf, np.inf))  \
for a_1, a_2 in tqdm( zip(X.ravel(), Y.ravel()))])
Z = Z.reshape(X.shape)

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
Z = Z.reshape(X.shape)
ax.plot_surface(X, Y, Z)
plt.xlabel("$a_1$")
plt.ylabel("$a_2$")
plt.show()

fig, ax = plt.subplots()
plt.imshow(Z, extent=[ 0,max(X_2), max(X_1),0], interpolation="none")
ax.xaxis.tick_top()
plt.xlabel('$a_1$')
plt.ylabel('$a_2$')
plt.show()
from mpl_toolkits import mplot3d
fig = plt.figure()
ax = plt.axes(projection='3d')
ax.contour3D(X, Y, Z, 50)
plt.show()
```

Listing A.7: Chapter 4

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import yfinance
```

```
import seaborn as sns
from scipy.stats import norm
np.random.seed(2021)


def empirical_cvar(X, alpha):
    """ X : np.array
        alpha: probability level in (0, 1)

        Returns: the empirical conditional value at risk, at level alpha.
    """
    quantile = empirical_quantile(X, alpha)
    return np.mean(X[np.where(X>=quantile)])

def var_cvar_analysis(asset_ticker, name, lookback = 100, holding_period = 1, alpha = 0.95):
    """
    This function retrieves financial market data from yahoo finance.
    It calculates the relative losses over given holding periods,
    and uses them to estimate the CVaR and VaR for the next period.

    It also estimates the parameters of a gemetric brownian motion, which
    is used to estimate the CVaR of the next period.

    This is repeated for the entire dataset.
    """

    horizon = '7Y'
    starting_date = "2015-1-1"
    end_date = "2021-1-1"
    #Retrieving market data
    asset = yfinance.Ticker(asset_ticker)
    asset_history = asset.history(period = horizon)["Close"].loc[starting_date:end_date]
    asset_history = pd.DataFrame(asset_history)
    #Calculating the relative difference
    asset_history.columns = ["Price"]
    asset_history["Difference"] = asset_history["Price"].diff(holding_period)
    asset_history["Shifted"] = asset_history["Price"].shift(holding_period)
    asset_history["Returns"] = -1* asset_history["Difference"]/asset_history["Price"].shift(holding_period)
    asset_history = asset_history.dropna() #Shifting the data has caused a NAN-instance. This is dropped.
    asset_history["CVaR estimate"] = 0
    asset_history["VaR estimate"] = 0
    asset_history["CVaR estimate GBM"] = 0
    for day in range(asset_history.shape[0]-lookback-1):
        #Estimating the paramters of a GBM
        logdiff = np.log(asset_history[day:day + lookback].\ Price.shift(-1)/asset_history[day:day + lookback].Pric
        XH = np.mean(logdiff)
        quad = (logdiff - XH)**2
        sigma = np.sqrt(sum(quad)/(quad.shape[0]))
        mu = XH + sigma**2/2
        #Using GBM parameters to find expectation and
        #Variance for a log-normal RV
        E = asset_history.iloc[day].Price*np.exp(mu)
        var = asset_history.iloc[day].Price**2*np.exp(2*mu)*(np.exp(sigma**2)-1)
        #Estimating the decribing features of the log-normal RV
        sigma = np.sqrt(np.log(1 + (var/E**2)) )
        mu = np.log( E**2 / np.sqrt(E**2 + var))
        #Calculating the theoretical CVaR. Notice 1-alpha. This is because we are looking for the
        # 1-alpha worst prices, causing the smallest returns.
        cvar_est = -(1 - np.exp(mu+sigma**2/2) * ( norm.cdf( norm.ppf(1-alpha)-sigma)/(1-alpha) ))
        asset_history["CVaR estimate"].iloc[day+lookback+1] = \
        empirical_cvar(asset_history.Returns.iloc[day:day+lookback].values,alpha) \
```

```
        asset_history["VaR estimate"].iloc[day+lookback+1] = \
        np.quantile(asset_history.Returns.iloc[day:day+lookback].values,alpha)
        asset_history["CVaR estimate GBM"].iloc[day+lookback+1] = \
        -(cvar_est-asset_history.Price[day])/asset_history.Price[day]
    asset_history = asset_history.iloc[lookback+1:]
    plt.figure(figsize = (8, 6))
    plt.figure(figsize = (10, 6))
    plt.plot(asset_history["CVaR estimate"], linewidth = 0.8, color \
             = "Red", label = f"CVaR_{{{alpha}}}$ estimate")
    plt.plot(asset_history["VaR estimate"], linewidth = 0.8, color = \
             "Green", label = f"$CaR_{{{alpha}}}$ estimate")
    plt.plot(asset_history["CVaR estimate GBM"], linewidth = 0.8, color\
             = "Blue", label = f"$CaR_{{{alpha}}}$ estimate (GBM)")
    plt.plot(asset_history["Returns"],  linewidth = 0.8, \
             label = "Negative relative returns")
    plt.legend()
    plt.savefig(f"figures/{name}_risk_return.pdf")
    plt.figure()
    sns.distplot(asset_history["Returns"].values, kde = False)
    plt.legend(["Negative relative returns"])
    plt.savefig(f"figures/{name}_return_distribution.pdf")
    n_worst_days = int(asset_history.shape[0]*0.1)
    worst_indices = asset_history.Returns.argsort()[-n_worst_days:][::-1].values
    worst_returns = asset_history.iloc[worst_indices].Returns.values
    predicted_cvar = asset_history.iloc[worst_indices]["CVaR estimate"].values
    predicted_var = asset_history.iloc[worst_indices]["VaR estimate"].values
    predicted_cvar_GBM = asset_history.iloc[worst_indices]["CVaR estimate GBM"].values
    plt.figure(figsize = (5, 5))
    plt.scatter(worst_returns, predicted_cvar)
    plt.axis("equal")
    plt.plot( ( 0, max(worst_returns)), (0, max(worst_returns)))
    plt.xlabel("Relative negative return")
    plt.ylabel("Estimated CVaR")
    plt.savefig(f"figures/{name}_returns_vs_cvar.pdf")
    plt.figure(figsize = (5, 5))
    plt.scatter(worst_returns, predicted_var)
    plt.axis("equal")
    plt.plot( ( 0, max(worst_returns)), (0, max(worst_returns)))
    plt.xlabel("Relative negative return")
    plt.ylabel("Estimated VaR")
    plt.savefig(f"figures/{name}_returns_vs_var.pdf")
    plt.figure(figsize = (5, 5))
    plt.scatter(worst_returns, predicted_cvar_GBM)
    plt.axis("equal")
    plt.plot( ( 0, max(worst_returns)), (0, max(worst_returns)))
    plt.xlabel("Relative negative return")
    plt.ylabel("Estimated CVaR (GBM)")
    plt.savefig(f"figures/{name}_returns_vs_CVaR_GBM.pdf")
    return asset_history, asset_history.iloc[worst_indices[:10]]

asset_history_sp, worst_days_sp  = var_cvar_analysis("^GSPC", "SP", lookback = 100)
print(worst_days_sp[worst_days.columns[-4:]] + "SP")


asset_history_sp, worst_days_sp  = var_cvar_analysis("MSFT", "MS", lookback = 100)
print(worst_days_sp[worst_days.columns[-4:]] + "SP")
```

Bibliography

[Ace02] Acerbi, C. 'Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion'. In: *Journal of Banking & Finance* vol. 26 (July 2002), pp. 1505–1518.

[Art+99] Artzner, P. et al. 'Coherent measures of risk'. In: *Mathematical finance* vol. 9, no. 3 (1999), pp. 203–228.

[BBK08] Benth, F. E., Benth, J. S. and Koekebakker, S. *Stochastic modelling of electricity and related markets*. Vol. 11. World Scientific, 2008.

[BBV04] Boyd, S., Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

[BS73] Black, F. and Scholes, M. 'The Pricing of Options and Corporate Liabilities'. In: *Journal of Political Economy* vol. 81, no. 3 (1973), pp. 637–654.

[Che+14] Cheung, K. et al. 'Optimal reinsurance under general law-invariant risk measures'. In: *Scandinavian Actuarial Journal* vol. 2014, no. 1 (2014), pp. 72–91.

[CJM17] Croghan, J., Jackman, J. K. and Min, K. J. 'Estimation of geometric Brownian motion parameters for oil price analysis'. In: (2017).

[Dha+02] Dhaene, J. et al. 'The concept of comonotonicity in actuarial science and finance: theory'. In: *Insurance: Mathematics and Economics* vol. 31, no. 1 (2002), pp. 3–33.

[Dha+06] Dhaene, J. et al. 'Risk measures and comonotonicity: a review'. In: *Stochastic models* vol. 22, no. 4 (2006), pp. 573–606.

[EMS02] Embrechts, P., McNeil, A. and Straumann, D. 'Correlation and dependence in risk management: properties and pitfalls'. In: *Risk management: value at risk and beyond* vol. 1 (2002), pp. 176–223.

[Fau18] Faul, A. C. *A Concise introduction to numerical analysis*. CRC Press, 2018.

[FS02] Föllmer, H. and Schied, A. 'Convex measures of risk and trading constraints'. In: *Finance and stochastics* vol. 6, no. 4 (2002), pp. 429–447.

[HC20] Huseby, A. B. and Christensen, D. 'Optimal Reinsurance Contracts in the Multivariate Case'. In: Jan. 2020, pp. 465–472.

[Kaa+02] Kaas, R. et al. 'A simple geometric proof that comonotonic risks have the convex-largest sum'. In: *ASTIN Bulletin: The Journal of the IAA* vol. 32, no. 1 (2002), pp. 71–80.

[LF01] Li, D.-H. and Fukushima, M. 'On the global convergence of the BFGS method for nonconvex unconstrained optimization problems'. In: *SIAM Journal on Optimization* vol. 11, no. 4 (2001), pp. 1054–1064.

[MW12] McDonald, J. and Weiss, N. *A Course in Real Analysis*. Elsevier Science, 2012.

[Øks03] Øksendal, B. 'Stochastic differential equations'. In: *Stochastic differential equations*. Springer, 2003, pp. 65–84.

[Puc13] Puccetti, G. 'Sharp bounds on the expected shortfall for a sum of dependent random variables'. In: *Statistics & Probability Letters* vol. 83, no. 4 (2013), pp. 1227–1232.

[Roc12] Rockafellar, R. 'Coherent Approaches to Risk in Optimization Under Uncertainty'. In: *Tutorials in Operations Research* vol. 13 (Apr. 2012).

[Roc70] Rockafellar, R. T. *Convex analysis*. 28. Princeton university press, 1970.

[Ros52] Rosenblatt, M. 'Remarks on a multivariate transformation'. In: *The annals of mathematical statistics* vol. 23, no. 3 (1952), pp. 470–472.

[RR10] Rockafellar, R. T. and Royset, J. O. 'On buffered failure probability in design and optimization of structures'. In: *Reliability engineering & system safety* vol. 95, no. 5 (2010), pp. 499–510.

[RR13] Rockafellar, R. T. and Royset, J. O. 'Superquantiles and their applications to risk, random variables, and regression'. In: *Theory Driven by Influential Applications*. Informs, 2013, pp. 151–167.

[RR14] Rockafellar, R. T. and Royset, J. O. 'Random variables, monotone relations, and convex analysis'. In: *Mathematical Programming* vol. 148, no. 1-2 (2014), pp. 297–331.

[RU+00] Rockafellar, R. T., Uryasev, S. et al. 'Optimization of conditional value-at-risk'. In: *Journal of risk* vol. 2 (2000), pp. 21–42.

[RU00] Rockafellar, R. and Uryasev, S. 'Optimization of Conditional Value-At-Risk'. In: *Journal of risk* vol. 2 (Jan. 2000), pp. 21–42.

[RU02a] Rockafellar, R. and Uryasev, S. 'Conditional Value-At-Risk for General Loss Distributions'. In: *Journal of Banking & Finance* vol. 26 (July 2002), pp. 1443–1471.

[RU02b] Rockafellar, R. T. and Uryasev, S. 'Conditional value-at-risk for general loss distributions'. In: *Journal of banking & finance* vol. 26, no. 7 (2002), pp. 1443–1471.

[Sha13] Shapiro, A. 'On Kusuoka representation of law invariant risk measures'. In: *Mathematics of Operations Research* vol. 38, no. 1 (2013), pp. 142–152.

[SYX16] Shi, Z., Yang, G. and Xiao, Y. ‘A limited memory BFGS algorithm for non-convex minimization with applications in matrix largest eigenvalue problem’. In: *Mathematical Methods of Operations Research* vol. 83, no. 2 (2016), pp. 243–264.

[YY+02] Yamai, Y., Yoshiba, T. et al. ‘On the validity of value-at-risk: comparative analyses with expected shortfall’. In: *Monetary and economic studies* vol. 20, no. 1 (2002), pp. 57–85.