Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

RNDr. Jana Čerbáková

INCOMPLETE INFORMATION IN STOCHASTIC PROGRAMMING PROBLEMS

Department of Probability and Mathematical Statistics

Supervisor:	Prof. RNDr. Jitka Dupačová, DrSc.
Branch:	M5 – Econometrics and Operations Research

Prague, 2008

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DISERTACE

RNDr. Jana Čerbáková

NEÚPLNÁ INFORMACE V ÚLOHÁCH STOCHASTICKÉHO PROGRAMOVÁNÍ

Katedra pravděpodobnosti a matematické statistiky

Školitel:	Prof. RNDr. Jitka Dupačová, DrSc.
Studijní obor:	M5 – Ekonometrie a operační výzkum

Praha, 2008

Acknowledgements

I am very much indebted to Prof. Jitka Dupačová for her excellent guidance and huge support during my research at Charles University. Special thanks belong to David Polách for the implementation of developed algorithm.

This work was supported by the Grant Agency of the Czech Republic (grants 201/05/H007, 402/05/0115).

Prague, May 21, 2008 RNDr. Jana Čerbáková

Contents

Chapter 1

Introduction

Real life situations including economical practice are typically dealing with one common element — randomness. We base our decisions on outcomes of critical variables (demand, return, loss, etc.) without knowing their precise future development. The decision making process is confronted with uncertainty and our effort consists of finding the best alternative under some given restrictions (e.g. technical or legislative limits). In words of a mathematician, we solve a decision problem of stochastic optimization.

Stochastic programs are optimization problems where some variables in objective or in constraints are random. If the probability model is known (i.e. all the probability distributions of underlying random variables are completely known), the decision maker has only to face the unknown realization of random variables — he solves the *uncertainty problem*. Stochastic programming is also called decision making under risk. In such models one frequently uses expectations at the place of random variables when the problem is solved repeatedly. This framework is motivated by the law of large numbers, see Dantzig [16]. The second possible approach involves probabilistic constraints, i.e. requirements that some events occur only with small probabilities. This leads to chance-constrained optimization, see Charnes and Cooper [15]. Another applied way of solution concerns models where infeasibility is penalized, called problems with recourse. Suggestion of Prékopa [41] is to apply probabilistic constraints and at the same time, to extend objective function for expected penalty term:

> *...we are convinced that the best way of operating a stochastic system is to operate it with a prescribed (high) reliability and at the same time use penalties to punish discrepancies.*

In reality we frequently do not know the exact probability distributions of random variables. Therefore, we also have to deal with an *ambiguity* (see e.g.

Pflug and Wozabal [39]) caused by an incomplete knowledge of probability distributions of underlying random variables.

The knowledge of probability distribution is often limited only to a certain set of feasible distributions - the ambiguity set, which can be characterized e.g. by its support, moments, symmetry, unimodality, qualitative information or by a parametric family of distributions, etc. For example, Pflug and Wozabal [39] determine the ambiguity set to which the modeler is indifferent as the set of all probability distributions whose distance from the hypothetical distribution is smaller than a given value. In their paper, the distance of two probability distributions is measured by Wasserstein distance.

Such types of ambiguity sets occur in stability analysis. The incomplete or inaccurate knowledge of input information influences the quality of the obtained optimal decisions which may be then quite different from the truly optimal actions, specially when estimated distributions computed from historical data are used. Hence, it is important to study the behavior of optimal solutions and of the corresponding optimal decisions with respect to small changes in the probability distributions, to the perturbed input or to a new information, see Römisch [50], Heitsch et al. [27], where the extensions to the multi-stage programs are made. Stability then ensures that small modifications of the underlying distributions or of the problem formulation cause only small changes of solutions.

The basic choice of the set of feasible distributions corresponding to the distributions specified by their moments and unimodality can be found in Dupačová [17] and Dupačová [19]. Combination of moment conditions and properties like symmetry and/or unimodality is presented in Popescu [40] or Čerbáková [12], where applications to the worst-case VaR and CVaR are derived. A family of unimodal probability distributions is also explored in Shapiro [55].

Literature offers many approaches how to deal with ambiguity. The most widespread application is the use of *minimax*, resp. *maximin*, decision rule, pioneered in Žáčková [63] and further researched in Dupačová [17], Dupačová [18], Dupačová [19]. Given an ambiguity set, the decision maker searches for the best protection against the worst alternative of possible probability distributions of random variables, he/she looks at the *worst-case* situation and minimizes the maximum of expected costs, resp. maximizes the minimum of expected returns, over the set of feasible probability distributions. The worst-case approach contemplates all possible probability distributions and scenarios, including those that are extremely unlikely to happen. Such a decision rule is necessary in situations where even rare events may lead to disastrous consequences.

An alternative treatment of ambiguity like Bayes or maximal entropy approaches is discussed e.g. in Sengupta [53] and Jagannathan [30]. The Bayes approach is usually applied to ambiguity sets characterized by a parametric family of distributions. It is assumed that the distribution functions have a specified functional form. The decision maker is supposed to have some prior information about statistical properties of unknown random parameters. This information is updated as further observations are obtained. The decision maker chooses his/her decisions in order to minimize the expected value of the objective function.

Using the maximal entropy principle we suppose that all possible future realizations have the same probability. The decision maker is indifferent towards any event probability. This approach corresponds to the situation where no information is available and represents a counterpart to the qualitative information approach where some realizations are favored to others.

The main goal of this work is to introduce and compare two different ways how to deal with incomplete information on probability distribution which is often involved in real stochastic programs. The first one works with estimated distributions and then studies what may happen if the *believed* distribution or problem formulation is changed. We study there necessary and sufficient conditions for a solution to be stable. The second robust approach includes the incomplete knowledge of probability distribution to the problem formulation and leads to the minimax decision rule. In virtue of the chosen approaches we limit our attention to stochastic programs whose sets of feasible solutions do not depend on the probability distribution.

In this thesis we shall deal mainly with the set of discrete probability distributions defined by a *qualitative information* — the case when the decision maker has some qualitative ideas about the future development of random events. The knowledge of possible realizations of random variables is assumed, i.e. the set of uncertain future states of nature is fixed. Further, we assume that the available information about future realizations can be described by a weak partial order "not less probable than". It corresponds to an educated guess. Such an ambiguity set consistent with a qualitative information can be represented by a convex polyhedron, see Bühler [7], Bühler [8]. Its extremal points are generalized discrete uniform distributions, see Bühler [6].

The paper is organized as follows. In chapter 2, we start with basic continuity and convexity properties. Then our attention is turned out to a general formulation of stochastic programming problem and to important results from linear programming theory, which are then applied in the next chapters, namely, results from duality theory, including Farkas lemma, serving a derivation of feasibility and optimality cuts of algorithms presented in chapters 5 and 6. More about these basic findings can be found e.g. in Kall and Mayer [34] or Prékopa [42].

Chapter 3 is devoted to the Bayes decision problem. The goal is to find an action/decision minimizing the expected loss on the set of plausible actions, studied also in Berger [2]. The loss function expresses the consequences of choosing a particular decision for a concrete realization of random parameter. The main goal is to study the behavior of minimal losses and of the corresponding optimal actions with respect to small changes in the probability distribution. This helps us to evaluate the error caused by using an approximated or perturbed distribution. We state conditions guaranteeing that small changes of probability distribution cause only small changes of solutions. We are also able to measure the maximal distance between optimal values and optimal decisions with respect to the considered probability distributions. This concept of qualitative and quantitative stability comes from the work of Römisch [50], where the general concept of stability for problems with the set of feasible solutions depending on a probability distribution is derived. In the Bayes decision problem it is supposed instead that the set of all feasible decisions does not depend on the choice of probability distribution. This simplifying assumption facilitates to improve the bounds for the maximal distance between optimal decisions presented in Römisch [50]. Our results are introduced in theorems 3.1 and 3.2 and published in Čerbáková [9]. We also show how these results can be related to stability of Bayes decisions with respect to weak convergence of probability measures. We present selected necessary and sufficient conditions for a Bayes action to be stable. The essential contribution can be found in Salinetti [52], Kadane and Chuang [31].

Chapter 4 is inspired by the minimax approach and its connection with the moment problem. After a historical introduction to minimax ideas we present the general moment problem and its dual formulation utilizing the work of Popescu [40] and Shapiro [55]. In the moment problem it is assumed that the knowledge of probability distributions of random variables is limited to a set of possible probability distributions defined by prescribed values of their moments. Subsequently, we suppose that except the first two moments, we know further qualitative characteristic of the class of distributions — symmetry and unimodality. Under various assumptions on probability distributions we derive and compare upper bounds for two well-known risk measures Value-at-Risk and Conditional Value-at-Risk. These bounds are also illustrated numerically on the case of interbank exchange rate. The chapter is based on papers Čerbáková [10], Čerbáková [11] and Čerbáková [12]. A new result, the worst-case Conditional Value-at-Risk for symmetric distributions with given expectation and variance, is presented in section

4.1.2. We also corrected the proof of lemma 4.6, published in Popescu [40], about the upper bounds for probability of loss for symmetric and unimodal distributions.

In the remaining part of the thesis we restrict our attention to the minimax approach applied on a set of possible discrete probability distributions consistent with some type of qualitative information. In section 4.2 we study extreme points of this set since they help to simplify the calculation of minimax solutions. The considered problem of finding the worst-case probability distribution becomes a linear programming problem on a bounded polyhedron. Therefore, it is sufficient to search for the worst-case distribution among the extreme points of the set of possible probability measures. In the case of qualitative information we are able to specify explicit form of such extreme points. This results were published in Bühler [6], Bühler [7], Bühler [8] and detailed in Čerbáková [9].

In chapter 5 we modify the L-shaped algorithm for discrete two-stage minimax stochastic programs with a linear recourse presented in Riis [45], Riis and Andersen [47]. We simplify the algorithm by maximizing only over extreme points of the set of feasible probability distributions and use their special structure in the case of a qualitative information. Two versions of L-shaped algorithm are presented. The first one solves the two-stage minimax problem under the assumption of relatively complete recourse, i.e. the second-stage problem is always feasible. In each step the algorithm adds only new optimality cuts to the solved problem. The algorithm ends by finding an optimal decision and the worst-case probability distribution. The second version of the algorithm is derived with relaxed assumption of relatively complete recourse. We allow infeasibility of both the first-stage and the second-stage problems. If the first-stage problem is infeasible the algorithm ends with no feasible solution, otherwise in each step new feasibility and optimality cuts are added until an optimal solution is found.

In chapter 6, we consider a multi-stage stochastic program with a linear recourse. In the multi-stage stochastic program we assume that information is revealed over time and the decision maker has to take an action before knowing the actual realizations of random variables. The real probability distribution is approximated via scenarios. Specially, we deal with a problem with a known event tree structure having only a qualitative information about events' probabilities, see also Bühler [8] and Čerbáková [9]. We work with Markov type event probabilities, i.e. the probabilities of realizations at a given time depend only on the probability of the preceding event. This gives a possibility to characterize a non-recombining scenario tree by a special transition matrix of a Markov chain defined on scenario nodes.

For the sake of robustness, the use of minimax approach is crucial. The possibility to describe a scenario tree by a Markov chain and the special properties of its transition matrix are fundamental for a numerical solution of a multi-stage minimax stochastic problem. Due to the Markov property we are able to disassemble a scenario node probability to a multiplication of transition probabilities between nodes at a scenario path. This allows to specify the multi-stage minimax stochastic problem with regard to maximizations over transition probabilities and customize the nested decomposition algorithm for the minimax approach. We utilize the nested decomposition algorithm described in Kall and Mayer [34] and evolve a special form of the algorithm for minimax problems consistent with some qualitative information. We also employ results from Bühler [6], Bühler [7] and Bühler [8]. As in the two-stage model we present two versions of nested decomposition with and without the assumption of relatively complete recourse. The first one has been already implemented in programming language C# using .NET platform. The algorithm performances are illustrated on numerical examples. We also introduce a possible application to portfolio selection problem. The chapter is based on the work of Čerbáková [14].

The developed algorithms can be applied to multi-stage stochastic problems with a linear recourse and with a given scenario tree structure but with (incomplete) qualitative information on probability distributions. This is typical when we work with expert's opinions.

Finally, chapter 7 summarizes the presented results and adverts to some possible improvements and extensions. There are still several open problems like stability of minimax solutions, development of extreme points generator, comparison of numerical properties of the developed algorithms with already existing solution techniques, etc. Some of these problems have been already studied. As an example we mention the work of Riis [45], Riis and Schultz [46] deriving the results on stability of two-stage minimax problems.

Chapter 2

Preliminaries

In this chapter we mention the selected essential mathematical principles and definitions needed for the presented results. First of all we recall basic definitions of convex sets and continue with continuity properties of extended real valued functions and multifunctions. Section 2.1 concludes with introduction to weak and uniform convergences. The second section is devoted to basic formulations of stochastic programming problems. In the third part the duality statements in linear programs are presented. Finally, the list of selected symbols used in the following chapters is introduced.

All introduced findings in the thesis are defined on some subset of $\bar{\mathbb{R}}^n$ for $n \in \mathbb{N}$ and $\bar{\mathbb{R}} := [-\infty, +\infty]$. Therefore, also the definitions and results provided in this chapter are supposed to be defined on some subset of $\bar{\mathbb{R}}^n$ although a generalization exists.

2.1 Basic definitions

We start with several basic convexity definitions:

Definition 2.1 (Convex set)**.** A subset $\mathbb{X}$ of $\mathbb{R}^n$ is *convex* if for every choice of $x, \bar{x} \in \mathbb{X}$ one has $[x, \bar{x}] \in \mathbb{X}$, i.e.

$$\alpha x + (1-\alpha)\bar{x} \in \mathbb{X} \text{ for all } \alpha \in (0,1).$$

It can be proved, see e.g. Rockafellar and Wets [48], theorem 2.2, that a set $\mathbb{X}$ is convex if and only if $\mathbb{X}$ contains all convex combinations of its elements.

Definition 2.2 (Polyhedral set)**.** A set $\mathbb{X} \subset \mathbb{R}^n$ is said to be a *polyhedral set* if it can be specified by finitely many linear constraints.

Definition 2.3 (Convex hull)**.** A *convex hull* of $\mathbb{X} \subset \mathbb{R}^n$, denoted by conv $\mathbb{X}$, is the smallest convex set that includes $\mathbb{X}$.

Convex hull of $\mathbb{X}$ consists of all convex combinations of elements of $\mathbb{X}$, i.e.

$$\text{conv } \mathbb{X} := \{x : x = \sum_{k=0}^{K} \alpha_j x_k : x_k \in \mathbb{X}, \alpha_k \geq 0, \sum_{k=0}^{K} \alpha_k = 1, K \geq 0\},$$

and thus every point of conv $\mathbb{X}$ can be expressed as a convex combination of $K+1$ points of $\mathbb{X}$ (not necessarily different). See e.g. Rockafellar and Wets [48], theorem 2.27 and theorem 2.29.

Definition 2.4 (Convex polyhedron)**.** A set $\mathbb{X} \subset \mathbb{R}^n$ is said to be a *convex polyhedron* if there exists a finite set $\{x_0, \ldots, x_K\}$ such that $\mathbb{X} = \text{conv } \{x_0, \ldots, x_K\}$.

Definition 2.5 (Positive hull)**.** By *positive hull* of $\mathbb{X} \subset \mathbb{R}^n$, denoted by pos $\mathbb{X}$, we understand the smallest convex cone that includes $\mathbb{X}$.

And similarly to the case of convex hull we can prove that pos $\mathbb{X}$ consists of all nonnegative linear combinations of elements of $\mathbb{X}$, i.e.

$$\text{pos } \mathbb{X} := \{x : x = \sum_{k=0}^{K} \alpha_k x_k : x_k \in \mathbb{X}, \alpha_k \geq 0, K \geq 0\}.$$

Definition 2.6 (Convex polyhedral cone)**.** A set $\mathbb{X} \subset \mathbb{R}^n$ is said to be a *convex polyhedral cone* if there exists a finite set $\{x_0, \ldots, x_K\}$ such that $\mathbb{X} = \text{pos } \{x_0, \ldots, x_K\}$.

Definition 2.7 (Extreme point)**.** An element $x \in \mathbb{X}, \mathbb{X} \subset \mathbb{R}^n$, is said to be an extreme point of $\mathbb{X}$ if there do not exist any two different elements $\bar{x}, \tilde{x} \in \mathbb{X}$ such that

$$x = \alpha\bar{x} + (1-\alpha)\tilde{x} \text{ for } \alpha \in (0,1).$$

During reading the thesis we will meet the following properties of real-valued functions:

Definition 2.8 (Continuity and convexity)**.** The extended real valued function $f : \mathbb{X} \to \bar{\mathbb{R}}$ is

(i) *lower semicontinuous* at $\bar{x} \in \mathbb{X}$ if

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x});$$

(ii) *upper semicontinuous* at $\bar{x} \in \mathbb{X}$ if

$$\limsup_{x \to \bar{x}} f(x) \leq f(\bar{x});$$

(iii) *continuous* at $\bar{x} \in \mathbb{X}$ if it is both lower semicontinuous and upper semicontinuous at $\bar{x}$;

(iv) *uniformly continuous* if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, \bar{x} \in \mathbb{X}$, $|\bar{x} - x| < \delta$ implies $|f(\bar{x}) - f(x)| < \varepsilon$;

(v) *Lipschitz continuous* on $\mathbb{K} \subset \mathbb{X}$ if there exist $\kappa \in \mathbb{R}_+$ with

$$|f(\bar{x}) - f(x)| \leq \kappa |\bar{x} - x| \text{ for all } \bar{x}, x \in \mathbb{K};$$

(vi) *convex* on a nonempty convex set $\mathbb{K} \subset \mathbb{X}$ if the inequality

$$f(\alpha x + (1 - \alpha)\bar{x}) \leq \alpha f(x) + (1 - \alpha) f(\bar{x})$$

holds for any two different points $x, \bar{x} \in \mathbb{K}$ and any $\alpha \in (0, 1)$;

(vii) *concave* on a nonempty convex set $\mathbb{K} \subset \mathbb{X}$ if the inequality

$$f(\alpha x + (1 - \alpha)\bar{x}) \geq \alpha f(x) + (1 - \alpha) f(\bar{x})$$

holds for any two different points $x, \bar{x} \in \mathbb{K}$ and any $\alpha \in (0, 1)$.

Definition 2.9 (Equi-continuity and equi-boudedness)**.** We say that a class $\mathcal{C}$ of extended real valued functions $f : \mathbb{X} \to \bar{\mathbb{R}}$ is

(i) a class of functions *equi-continuous* at $\bar{x} \in \mathbb{X}$ if for each neighbourhood $\mathcal{V}$ of $f(\bar{x})$ exists a neighbourhood $\mathcal{U}$ of $\bar{x}$ such that for all $f \in \mathcal{C}$ is $f(\mathcal{U}) \subset \mathcal{V}$;

(ii) a class of *equi-bounded* functions if there exists a bounded subset $\mathbb{B} \subset \bar{\mathbb{R}}$ such that for all $f \in \mathcal{C}$ and for all $x \in \mathbb{X}$ is $f(x) \in \mathbb{B}$.

In chapter 3 we will examine the stability of Bayes decision problem. To this purpose we define multifunctions (or multivalued mappings, set-valued mappings) and study their continuity properties:

Definition 2.10 (Multifunction)**.** By a *multifunction* we understand a set-valued mapping $f : \mathbb{X} \to 2^{\mathbb{U}}$ with

$$2^{\mathbb{U}} := \text{ collection of all subsets of } \mathbb{U}.$$

The domain and range of f are taken to be the sets

$$\operatorname{dom} f := \{x \in \mathbb{X} : f(x) \neq \emptyset\} \text{ and } \operatorname{rge} f := \{u \in \mathbb{U} : \exists x \in \mathbb{X} \text{ with } u \in f(x)\}.$$

The multifunction f is fully described by its graph $\operatorname{gph} f := \{(x, u), u \in f(x)\} \subset \mathbb{X} \times \mathbb{U}$. If f is single-valued everywhere on $\mathbb{X}$ we say that f is a function. Before defining the semicontinuity properties of multifunctions it is necessary to say that it can be very confusing as there exist many definitions of these terms in literature. Our definitions come from Rockafellar and Wets [48].

Definition 2.11 (Semicontinuity)**.** We say that a multifunction $f : \mathbb{X} \to 2^{\mathbb{U}}$ is

(i) *upper semicontinuous* at $\bar{x}$ if

$$\limsup_{x \to \bar{x}} f(x) \subset f(\bar{x});$$

(ii) *lower semicontinuous* at $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \supset f(\bar{x});$$

It is called *continuous* if both conditions hold.

Upper semicontinuity everywhere corresponds to $f^{-1}(\mathbb{C})$ being closed whenever $\mathbb{C}$ is closed, whereas lower semicontinuity everywhere corresponds to $f^{-1}(\mathbb{O})$ being open whenever $\mathbb{O}$ is open.

The main part of the thesis deals with probability distributions belonging to some set $\mathcal{P}$ of possible probability distributions. It is assumed that these distributions are defined on a measurable space $(\Omega, \mathcal{B}(\Omega))$ with nonempty closed set $\Omega \subset \mathbb{R}^k$ and Borel σ-algebra $\mathcal{B}(\Omega)$ of Ω. On this set we have to define a topology in order to use general results. For our purposes the most convenient is the topology of weak convergence:

Definition 2.12 (Weak convergence)**.** Consider $p, \{p_\nu\}_{\nu=1}^\infty \in \mathcal{P}$. The sequence $\{p_\nu\}_{\nu=1}^\infty$ is said to *converge weakly* to p (written $p_\nu \xrightarrow[\nu \to \infty]{w} p$) if for any bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$ we have

$$\int_\Omega f(\bar{\omega}) \, \mathrm{d}p_\nu(\bar{\omega}) \xrightarrow[\nu \to \infty]{} \int_\Omega f(\bar{\omega}) \, \mathrm{d}p(\bar{\omega}).$$

In section 3.2 we also employ other type of convergence:

Definition 2.13 (Uniform convergence)**.** Suppose $f, f_\nu : \mathbb{X} \to \mathbb{R}$ are real-valued functions. We say that the sequence $\{f_\nu\}_{\nu=1}^{\infty}$ is *uniformly convergent* with limit f if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $x \in \mathbb{X}$ and all $\nu \geq N$, $|f_\nu(x) - f(x)| < \varepsilon$.

2.2 Introduction to stochastic programs

By a stochastic programming problem we usually understand an optimization problem of the form

$$\begin{aligned} &\inf_{x\in\mathbb{X}} E_p f(x,\omega) \\ &\text{s.t.} \quad E_p g_j(x,\omega) \geq 0, \quad j = 0,\dots,J, \end{aligned} \tag{2.1}$$

where ω is a random parameter, i.e. a measurable mapping $\omega(\xi)$ on probability space (Ξ, Σ, P) with values in Ω, where Ω is nonempty closed subset of $\mathbb{R}^k$. For the sake of simplicity we omit the argument ξ of ω. By $\bar{\omega}$ we will denote a realization $\omega(\xi) \in \Omega, \xi \in \Xi$.

Note that the mapping ω generates the probability distribution $p := P \circ \omega^{-1}$ on $(\Omega, \mathcal{B}(\Omega))$ (where $\mathcal{B}(\Omega)$ denotes Borel σ-algebra of Ω), which provides all relevant probabilistic information about the considered random parameter. The probability distribution p is known and does not depend on decision x.

By $\mathbb{X} \subset \mathbb{R}^n$ we denote a given set of decisions not depending on ω. The random objective function f and functions $g_j, j = 0,\dots,J$, of a decision $x \in \mathbb{X}$ are defined on $\mathbb{X} \times \Omega$ with values in extended reals $\bar{\mathbb{R}}$. It is assumed that $f, g_j, j = 0,\dots,J$, are measurable in ω and lower semicontinuous in x. By $E_p f(x,\omega)$ we understand the expectation under p, i.e.

$$E_p f(x,\omega) := \int_\Omega f(x,\bar{\omega})\,\mathrm{d}p(\bar{\omega}).$$

The expectation in the constraints and in the objective function can be replaced by other function of decision and random variable nonlinear in p.

A very important class of stochastic problems penalizes the infeasibility by defining a new function $\Phi(x,\omega)$, which quantifies the violation of the stochastic constraints $g_j(x,\omega) \geq 0, j = 0,\dots,J$. In this approach we take $\Phi(x,\omega) = 0$ if x and ω satisfy all the above mentioned constraints. Then we reformulate the problem (2.1) as being

$$\inf_{x\in\mathbb{X}} f(x) + E_p \Phi(x,\omega). \tag{2.2}$$

Models where penalties are used are also called stochastic problems with recourse, see e.g. Prékopa [42]. These stochastic programs are considered in chapters 5 and 6 of the thesis.

For stochastic programming the timing and sequence of decisions and revealing of specific realizations of random variable is crucial. Therefore, in stochastic programming we define stage as a point in time where some decision variable is set. A stage is followed by an event epoch where some random variables are fixed according to their distribution. Some further decision can be made at the next stage, etc. We talk about multi-stage programming, for more details we refer e.g. to Prékopa [42] and to chapters 5 and 6 of the thesis for the exact formulations and applications.

If the probability distribution p of the random variable ω is discrete, we deal with a finite set of possible realizations of ω, i.e. $\Omega = \{\bar{\omega}_1, \ldots, \bar{\omega}_S\}$, where S denotes the number of all possible realizations (scenarios). Then the distribution p is of the form $p = (p_1, \ldots, p_S)^T$ with $p_s \geq 0, s = 1, \ldots, S$, and $\sum_{s=1}^{S} p_s = 1$. Problem (2.1) becomes a deterministic optimization program

$$\begin{aligned} &\inf_{x \in \mathbb{X}} \sum_{s=1}^{S} p_s f(x, \bar{\omega}_s) \\ &\text{s.t.} \quad \sum_{s=1}^{S} p_s g_j(x, \bar{\omega}_s) \geq 0, \quad j = 1, \ldots, J. \end{aligned} \tag{2.3}$$

If the functions $f, g_j, j = 1, \ldots, J$, and all constraints defining $\mathbb{X}$ are linear, we solve a problem of linear programming.

2.3 Linear programming and duality

As a standard formulations of a linear programming problem we find an optimization problem like

$$\begin{aligned} &\min c^T x \\ &\text{s.t.} \quad Ax = b, \\ &\qquad\; x \geq 0, \end{aligned} \tag{2.4}$$

with the matrix $A \in \mathbb{R}^{m \times n}$, the objective's coefficient $c \in \mathbb{R}^n$, the right-hand side vector $b \in \mathbb{R}^m$ and the decision vector $x \in \mathbb{R}^n_+$.

We say that a linear problem is *feasible* if the set $\mathbb{X} := \{x \in \mathbb{R}^n_+ : Ax = b\}$ is nonempty.

Lemma 2.1. *The set* $\mathbb{X} := \{x \in \mathbb{R}^n_+ : Ax = b\}$ *is either empty or it can be expressed as a sum of convex polyhedron* $\mathbb{P}$ *and convex polyhedral cone* $\mathbb{C}$*, i.e.*

$$\mathbb{X} = \mathbb{P} + \mathbb{C} := \{z : z = x + y, x \in \mathbb{P}, y \in \mathbb{C}\},$$

where $\mathbb{C} := \{y \in \mathbb{R}^n_+ : Ay = 0\}$ *and* $\mathbb{P} := conv\,\{x_0, \ldots, x_K\}$ *with* $x_0, \ldots, x_K$ *being extreme points of* $\mathbb{X}$ *also called* feasible basic solutions *of problem (2.5).*

Lemma 2.2. *Assuming that* $\mathbb{X} \neq \emptyset$ *the problem (2.5) has an optimal solution if and only if* $c^T y \geq 0$ *for all* $y \in \mathbb{C}$*. In this case an optimal solution can be chosen among the extreme points* $x_0, \ldots, x_K$ *of* $\mathbb{X}$ *as the point minimizing the objective function, i.e. the optimal solution* $\hat{x} \in \underset{k\in\{1,\ldots,K\}}{\text{argmin}}\; c^T x_k$*.*

We define the dual problem to problem (2.5) as follows

$$\begin{aligned} &\max b^T y \\ &\text{s.t.} \quad A^T y \leq c, \end{aligned} \tag{2.5}$$

where $y \in \mathbb{R}^m$ stands for the dual decision variable. The set of feasible solutions of the dual problem is denoted by $\mathbb{Y} := \{y \in \mathbb{R}^m : A^T y \leq c\}$. Note, the dual of the dual problem is the primal problem again.

Between the primal and the dual problem there exist several relations. We present theorems applied in the next chapters of the thesis. For the proofs we refer e.g. to Kall and Mayer [34].

Lemma 2.3 (Weak duality)**.** *For any pair of feasible solutions* $x \in \mathbb{X}$ *and* $y \in \mathbb{Y}$ *it holds that*

$$b^T y \leq c^T x.$$

Lemma 2.4 (Strong duality)**.** *If the primal problem is solvable, i.e. the optimal solution of primal problem exists, then so is the dual problem, and the optimal values of the two problems coincide, i.e.*

$$\min_{x\in\mathbb{X}} c^T x = \max_{y\in\mathbb{Y}} b^T y.$$

Lemma 2.5 (Farkas lemma)**.** *The set of feasible solutions of primal problem* $\mathbb{X}$ *is nonempty if and only if*

$$A^T y \leq 0 \quad \textit{implies} \quad b^T u \leq 0.$$

2.4 Notation

Chapter 3 Stability of Bayes actions:

x	decision variable;
$\mathbb{X}$	set of all feasible decisions/actions, $\mathbb{X} \neq \emptyset$, $\mathbb{X} \subset \mathbb{R}^n$;
(Ξ, Σ, P)	probability space;
$(\Omega, \mathcal{B}(\Omega))$	measurable space with $\mathcal{B}(\Omega)$ Borel σ-algebra of Ω, $\Omega \neq \emptyset$, $\Omega \subseteq \mathbb{R}^k$;
ω	random loss vector, measurable mapping $\omega : \Xi \to \Omega$;
$\bar{\omega}$	realization $\omega(\xi) \in \Omega$, $\xi \in \Xi$;
$p, q, \{p_\nu\}_{\nu=1}^{\infty}$	probability distributions of ω defined on $(\Omega, \mathcal{B}(\Omega))$;
$\mathcal{P}$	set of all probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$;
$L(x, \omega)$	loss random lower semicontinuous function, $L : \mathbb{X} \times \Omega \to \bar{\mathbb{R}}$;
$f(x, \omega)$	auxiliary function defined by $f(x, \omega) := L(x, \omega) - L(\hat{x}, \omega)$;
$\hat{x}$	Bayes (optimal) decision/action;
$\hat{\mathbb{X}}(p)$	the set of all Bayes (optimal) actions with respect to probability distribution p;
$\vartheta_(p)$	optimal value of Bayes decision problem with respect to probability distribution p, i.e. $\vartheta_(p) := \inf \left\{ \int\limits_{\Omega} L(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}) : x \in \mathbb{X} \right\}$;
$\mathcal{U}$	nonempty open subset of $\mathbb{R}^n$;
$\mathbb{B}$	Euclidean unit ball;
$\hat{\mathbb{X}}_{\mathcal{U}}(q)$	the set of all Bayes (optimal) actions of a perturbed model (defined on $\mathbb{X} \cap \mathrm{cl}\,\mathcal{U}$) with respect to probability distribution q;
$\vartheta_{\mathcal{U}}(q)$	optimal value of a perturbed Bayes decision problem (defined on $\mathbb{X} \cap \mathrm{cl}\,\mathcal{U}$) with respect to probability distribution q;
$\mathcal{P}_{L_{\mathcal{U}}}$	subset of $\mathcal{P}$ ensuring that Bayes optimization problem is well-defined for all $p \in \mathcal{P}_{L_{\mathcal{U}}}$;
$d_{L_{\mathcal{U}}}(p, q)$	probability pseudometric measuring the uniform distance of $p, q \in \mathcal{P}_{L_{\mathcal{U}}}$;
$\zeta_h(p, q)$	Fortet-Mourier metric of h-order, $h \geq 1$;
$\psi_p(\tau)$	growth function defined by $\psi_p(\tau) := \min \left\{ \int\limits_{\Omega} L(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}) - \vartheta(p) : d(x, \hat{\mathbb{X}}(p)) \geq \tau, x \in (\mathbb{X} \cap \mathrm{cl}\,\mathcal{U}) \right\}$, $p \in \mathcal{P}_{L_{\mathcal{U}}}$.

Chapter 4 Minimax approach:

(Ξ, Σ, P)	probability space;
$(\Omega, \mathcal{B}(\Omega))$	measurable space with $\mathcal{B}(\Omega)$ Borel σ-algebra of Ω, $\Omega \neq \emptyset$, $\Omega \subseteq \mathbb{R}$;
ω	random variable, measurable mapping $\omega : \Xi \to \Omega$;
$\bar{\omega}$	realization $\omega(\xi) \in \Omega$, $\xi \in \Xi$;
p	probability distribution of ω defined on $(\Omega, \mathcal{B}(\Omega))$;
F_ω	left-continuous distribution function of random variable ω;
$\mathcal{P}$	set of all probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$;
$\mathcal{P}^m$	set of all probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$ fulfilling the moment conditions $E[\omega] = \mu_\omega$ and $E[(\omega - \mu_\omega)^2] = \sigma_\omega^2$;
$\mathcal{P}^s$	set of all symmetric probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$;
$\mathcal{P}^{m,s}$	set of all symmetric probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$ fulfilling the moment conditions $E[\omega] = \mu_\omega$ and $E[(\omega - \mu_\omega)^2] = \sigma_\omega^2$;
$\mathcal{P}^{s,u}$	set of all symmetric and unimodal probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$;
$\mathcal{P}^{m,s,u}$	set of all symmetric and unimodal probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$ fulfilling the moment conditions $E[\omega] = \mu_\omega$ and $E[(\omega - \mu_\omega)^2] = \sigma_\omega^2$;
α	probability level $\alpha \in (0, 1]$;
J	number of moment constraints;
f	value function, $f : \Omega \to \mathbb{R}$;
g	vector of moment functions, $g := (g_0, \ldots, g_J)$, $g_j : \Omega \to \mathbb{R}$, $j = 0, \ldots, J$;
q	vector of moment constraints, $q \in \mathbb{R}^{J+1}$;
y	vector of dual decisions, $y \in \mathbb{R}^{J+1}$;
$\mathrm{VaR}_\alpha(\omega)$	Value-at-Risk at probability level α of random variable ω;
$\mathrm{CVaR}_\alpha(\omega)$	Conditional Value-at-Risk at probability level α of random variable ω;
$\mathrm{VaR}_{\alpha,N}(\omega)$	Value-at-Risk at probability level α of normally distributed random variable ω;

$\mathrm{VaR}^{wc}_{\alpha,\mathcal{P}}(\omega)$	worst-case Value-at-Risk at probability level α of random variable ω with distribution in $\mathcal{P}$;
$\mathrm{CVaR}^{wc}_{\alpha,\mathcal{P}}(\omega)$	worst-case Conditional Value-at-Risk at probability level α of random variable ω with distribution in $\mathcal{P}$.

Chapter 5 Minimax rule in two-stage programs:

c	first-stage cost vector, $c \in \mathbb{R}^{n_1}$;
x	first-stage decision;
$\hat{x}$	optimal first-stage decision;
$\mathbb{X}$	set of feasible first-stage decisions, $\mathbb{X} := \{x \in \mathbb{R}^{n_1}_{+} : Ax = d\}$;
θ	auxiliary first-stage decision variable, $\theta \in \mathbb{R}$;
$\hat{\theta}$	optimal value of auxiliary first-stage decision variable θ;
(Ξ, Σ, P)	probability space;
$(\Omega, \mathcal{B}(\Omega))$	measurable space with $\mathcal{B}(\Omega)$ Borel σ-algebra of Ω, $\Omega \neq \emptyset$, $\Omega \subseteq \mathbb{R}^k$;
ω	random second-stage vector, measurable mapping $\omega : \Xi \to \Omega$;
$\bar{\omega}$	realization $\omega(\xi) = \{q(\xi), b(\xi), W(\xi), T(\xi)\} \in \Omega$, $\xi \in \Xi$;
q	random second-stage cost vector, $q(\xi) \in \mathbb{R}^{n_2}$;;
b	random second-stage right-hand side vector, $b(\xi) \in \mathbb{R}^m$;
W	random second-stage recourse matrix, $W(\xi)_{m \times n_2}$;
T	random second-stage technology matrix, $T(\xi)_{m \times n_1}$;
S	number of scenarios;
s	scenario index, $s = 1, \ldots, S$;
q_s	realization of second-stage cost vector under scenario s, $s \in \{1, \ldots, S\}$;
b_s	realization of second-stage right-hand side vector under scenario s, $s \in \{1, \ldots, S\}$;
W_s	realization of second-stage recourse matrix under scenario s, $s \in \{1, \ldots, S\}$;
T_s	realization of second-stage technology matrix under scenario s, $s \in \{1, \ldots, S\}$;
p	probability distribution of second-stage random vector ω defined on $(\Omega, \mathcal{B}(\Omega))$;
$\mathcal{P}$	set of feasible second-stage probability distributions of ω on $(\Omega, \mathcal{B}(\Omega))$;

K	number of extreme points of $\mathcal{P}$;
k	index of extreme points of $\mathcal{P}$;
p^k	k-th extreme point of $\mathcal{P}$, $k = 1, \ldots, K$;
$\mathcal{T}_k$	admissible support corresponding to p^k, k-th extreme point of $\mathcal{P}$, $k = 1, \ldots, K$;
y	second-stage decision vector, $y \in \mathbb{R}_+^{n_2}$;
π	second-stage dual decision variable, $\pi \in \mathbb{R}^m$;
$\hat{\pi}_s$	optimal second-stage dual decision for scenario s, $s \in \{1, \ldots, S\}$;
$\Phi(x, \omega)$	second-stage value function.

Chapter 6 Minimax rule in multi-stage programs:

T	number of stages;
t	stage index, $t = 1, \ldots, T$;
ω	stochastic data process, $\omega = (\omega_1, \ldots, \omega_{T-1})$;
$\omega^{t-1,\bullet}$	path of stochastic process ω preceding stage t, $t = 2, \ldots, T$, $\omega^{t-1,\bullet} := (\omega_1, \ldots, \omega_{t-1})$;
p_t	marginal probability distribution of ω_t, $t = 1, \ldots, T-1$;
x	decision process, $x = (x_1, \ldots, x_T)$;
K_t	index of the last scenario tree node at stage t, $t = 1, \ldots, T$;
$\mathcal{N}$	set of all scenario tree nodes, $\mathcal{N} = \{1, \ldots, K_T\}$;
$\mathcal{N}_t$	set of all scenario tree nodes at stage t, $\mathcal{N}_t = \{K_{t-1}, \ldots, K_t\}$, $t = 1, \ldots, T$;
$\mathcal{G}(n)$	set of all nodes corresponding to a subtree rooted at the node n, $n \in \mathcal{N}$;
n	index of node, $n \in \mathcal{N}$;
$a(n)$	index of an unique ancestor of node n, $n \in \mathcal{N} \setminus \{1\}$;
$D(n)$	set of node n descendants, $n \in \mathcal{N} \setminus \mathcal{N}_T$;
x_n	decision vector corresponding to node n, $0 \le x_n \le u_n$, $x_n \in \mathbb{R}^{k_n}$, $n \in \mathcal{N}$;
u_n	upper bound on decision x_n corresponding to node n, $u_n \in \mathbb{R}^{k_n}$, $n \in \mathcal{N}$;
c_n	cost vector corresponding to node n, $c_n \in \mathbb{R}^{k_n}$, $n \in \mathcal{N}$;

b_n	right-hand side vector corresponding to node n, $b_n \in \mathbb{R}^{h_n}$, $n \in \mathcal{N}$;
W_n	recourse matrix corresponding to node n, $W_n \in \mathbb{R}^{h_n} \times \mathbb{R}^{k_n}$, $n \in \mathcal{N}$;
T_n	technology matrix corresponding to node n, $T_n \in \mathbb{R}^{h_n} \times \mathbb{R}^{k_n}$, $n \in \mathcal{N}$;
p_n	marginal probability of node n, $n \in \mathcal{N}$;
$p_{n,m}$	transition probability of getting from node n to node m, $n \in \mathcal{N}_{t-1}$, $m \in \mathcal{N}_t$, $t = 2, \ldots, T$;
l_n	number of immediate descendants of node n, $n \in \mathcal{N} \setminus \mathcal{N}_T$, $l_n = \|D(n)\|$;
p^n	vector of transition probabilities $\{p_{n,m}\}_{m \in D(n)}$, $p^n \in \mathbb{R}^{l_n}_+$, $\sum\limits_{m \in D(n)} p_{n,m} = 1$;
$\boldsymbol{M_t}$	matrix of transition probabilities between nodes $n_{t-1} \in \mathcal{N}_{t-1}$ and $n_t \in \mathcal{N}_t$, $t = 1, \ldots, T-2$;
P_n	set of all feasible probability distributions p^n, $n = 1, \ldots K_{T-1}$;
e_{P_n}	number of extreme points of P_n ,$n = 1, \ldots, K_{T-1}$;
$\mathcal{T}_k^{P_n}$	admissible support corresponding to k-th extreme point of P_n, $k = 1, \ldots, e_{P_n}$, $n = 1, \ldots, K_{T-1}$;
F_1	optimal value of multi-stage minimax stochastic program;
$F_n(x_{a(n)})$	optimal value of descendant problem defined on a subtree rooted at node n, $n \in \mathcal{N} \setminus \{1\}$;
θ_n	auxiliary decision variable at node n, $n \in \mathcal{N} \setminus \mathcal{N}_T$;
v_n	dual variable at node n corresponding to the constraint $W_n x_n = B_n - T_n x_n$, $n \in \mathcal{N}$;
w_n	dual variable at node n corresponding to the constraint obtained by optimality cut, $n \in \mathcal{N} \setminus \mathcal{N}_T$;
λ_n	dual variable at node n corresponding to the upper bound on decision variable x_n, $n \in \mathcal{N}$;
z_n	dual variable at node n corresponding to the constraint obtained by feasibility cut, $n \in \mathcal{N} \setminus \mathcal{N}_T$;
s_n	number of optimality cuts added at node n, $n \in \mathcal{N} \setminus \mathcal{N}_T$;
r_n	number of feasibility cuts added at node n, $n \in \mathcal{N} \setminus \mathcal{N}_T$.

Chapter 3

Stability of Bayes actions

Incomplete or unprecise knowledge of input parameters of solved models influences the quality of the obtained optimal decisions which may be then quite different from the truly optimal actions. In Bayes models, the uncertainties are incorporated into the model and there is a chance to analyze stability of decisions with respect to the perturbed input, new information, etc. Simple economic applications of Bayesian methods have been frequently used in practice, see e.g. Wonnacott and Wonnacott [62]. The remaining part of this chapter is devoted to stability analysis for Bayes decision models and based on Čerbáková [13].

In Bayes decision model, see Berger [2], the only unknown quantity is the parameter $\omega(\xi) \in \Omega$, where the set of admissible values Ω is a non-empty closed subset of $\mathbb{R}^k$. For the simplicity we will omit the argument ξ of ω and write $\bar{\omega}$ for a realization $\omega(\xi)$. We assume that ω is random with probability distribution p belonging to a class of all probability distributions $\mathcal{P}$ defined on $(\Omega, \mathcal{B}(\Omega))$, where $\mathcal{B}(\Omega)$ denotes Borel σ-algebra of Ω. The decision maker chooses his action (decision) x from the set of all admissible actions $\mathbb{X}$, where $\mathbb{X}$ is supposed to be a non-empty closed subset of $\mathbb{R}^n$. He makes his decision on the basis of random lower semicontinuous loss function $L : \mathbb{X} \times \Omega \to \bar{\mathbb{R}}$ which represents the loss caused by action x when the true value of random parameter is $\bar{\omega}$.

Definition 3.1 (Bayes action)**.** An action $\hat{x} \in \mathbb{X}$ is called Bayes if and only if it minimizes the expected loss

$$\hat{x} \in \operatorname*{argmin}_{x \in \mathbb{X}} \int_{\Omega} L(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}), \tag{3.1}$$

where p is the assumed probability distribution at the time of decision making. The set of all Bayes actions, i.e. the set of optimal solutions of (3.1) with respect to p, is denoted by $\hat{\mathbb{X}}(p)$.

Distribution p can represent the prior probability distribution or, in statistical decision problems, the posterior probability distribution after observing the data. The posterior distribution combines the prior information with the sample information represented by the likelihood function according to the Bayes theorem, see Berger [2]. In view of stability discussed in the thesis, it is not important to distinguish between p representing a prior or a posterior distribution.

In real problems we usually do not know the exact probability distribution of random parameters. We have to estimate them. Therefore, it is very important to be able to calculate an error, which can be caused by using estimated distribution or be confident the error will be sufficiently small. And this is the problem of stability, i.e. small modifications of the underlying probability distribution or problem formulation are supposed to cause only small changes of solutions.

In this chapter we shall give not only the usual continuity results, see definition 3.2 below, but we shall also quantify the errors in minimal expected loss and in the Bayes actions due to perturbations. Such results are of importance in real-life problems, e.g. in robustness analysis of the obtained results.

Definition 3.2 (Stable action). We say that a Bayes action $\hat{x} \in \hat{\mathbb{X}}(p)$ is stable if for every sequence of probability distributions $\{p_\nu, \nu \in \mathbb{N}\}$ weakly converging to p, $p_\nu \xrightarrow[\nu\to\infty]{w} p$, where $p, p_\nu \in \mathcal{P}, \forall\nu$, and for every sequence of loss functions $\{L_\nu, \nu \in \mathbb{N}\}$ converging (in some topology) to L

$$\left[\int_\Omega L_\nu(\hat{x},\bar{\omega})\,\mathrm{d}p_\nu(\bar{\omega}) - \inf_{x\in\mathbb{X}} \int_\Omega L_\nu(x,\bar{\omega})\,\mathrm{d}p_\nu(\bar{\omega})\right] \xrightarrow[\nu\to\infty]{} 0 \tag{3.2}$$

holds true.

It was shown in Kadane and Chuang [31] that for $\{L_\nu, \nu \in \mathbb{N}\}$ converging to L uniformly in x and $\bar{\omega}$, the condition (3.2) is equivalent to

$$\left[\int_\Omega L(\hat{x},\bar{\omega})\,\mathrm{d}p_\nu(\bar{\omega}) - \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar{\omega})\,\mathrm{d}p_\nu(\bar{\omega})\right] \xrightarrow[\nu\to\infty]{} 0, \tag{3.3}$$

which can be rewritten as

$$\inf_{x\in\mathbb{X}} \int_\Omega \left[L(x,\bar{\omega}) - L(\hat{x},\bar{\omega})\right]\,\mathrm{d}p_\nu(\bar{\omega}) \xrightarrow[\nu\to\infty]{} \inf_{x\in\mathbb{X}} \int_\Omega \left[L(x,\bar{\omega}) - L(\hat{x},\bar{\omega})\right]\,\mathrm{d}p(\bar{\omega}) = 0. \tag{3.4}$$

The stability of $\hat{x}$ then becomes the question of the continuity, at p, of the infimum integral functional

$$\inf\left\{\int_{\Omega} f(x,\bar{\omega})\,\mathrm{d}p(\bar{\omega}) : x \in \mathbb{X}\right\}, \tag{3.5}$$

where $f(x,\bar{\omega}) := L(x,\bar{\omega}) - L(\hat{x},\bar{\omega})$. We will assume that the loss function L does not depend on ν, thus (3.3) can be used in definition of stability instead of (3.2). All presented results can be extended to uniformly convergent $\{L_\nu, \nu \in \mathbb{N}\}$.

Other formulations of stability of the Bayes decision problem can be found in Kadane and Srinivasan [32]. The authors introduce two definitions of the Strong Stability for ε-minimal solutions of the Bayes problem and derive sufficient conditions for their equivalence. They also prove stability results with respect to weak convergence of probability distributions based on the work of Berger and Salinetti [3], Billingsley and Topsøe [4], Kadane and Chuang [31] and Sallinetti [52]. The most important findings are mentioned in section 3.2.

Using general stability results of Römisch [50] in the context of Bayes decision analysis we shall be able to obtain error bounds for optimal values (minimal expected losses) and for the solution sets (Bayes actions) caused by perturbations of the underlying distribution. This concept of stability is formulated in section 3.1, the main results on improved distances of solutions sets are presented in theorems 3.1 and 3.2.

3.1 Stability theorems

According to definition 3.1, Bayes decision analysis deals with the problem

$$\inf\left\{\int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}p(\bar{\omega}) : x \in \mathbb{X}\right\}, \tag{3.6}$$

namely with the behavior of the set of optimal solutions $\hat{\mathbb{X}}(p)$ and optimal values $\vartheta(p)$ in dependence on small changes of probability distribution p. Together with the original problem (3.6) we consider a perturbed model with another distribution $q \in \mathcal{P}$ instead of p. We apply the following notation:

$$\vartheta_{\mathcal{U}}(q) := \inf\left\{\int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}q(\bar{\omega}) : x \in \mathbb{X} \cap \mathrm{cl}\,\mathcal{U}\right\}$$

for the optimal value of the perturbed model and

$$\hat{\mathbb{X}}_{\mathcal{U}}(q) := \left\{ x \in \mathbb{X} \cap \mathrm{cl}\,\mathcal{U} : \int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}q(\bar{\omega}) = \vartheta_{\mathcal{U}}(q) \right\}$$

for the set of optimal solutions of the perturbed model.

To measure the distance of probability distributions we define for any nonempty and open subset $\mathcal{U}$ of $\mathbb{R}^n$ the set

$$\mathcal{P}_{L_{\mathcal{U}}} := \left\{ q \in \mathcal{P} : \quad -\infty < \int_{\Omega} \inf_{x \in \mathbb{X} \cap r\mathbb{B}} L(x,\bar{\omega})\,\mathrm{d}q(\bar{\omega}) \quad \forall r > 0, \right. \\ \left. \sup_{x \in \mathbb{X} \cap \mathrm{cl}\,\mathcal{U}} \int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}q(\bar{\omega}) < \infty \right\} \tag{3.7}$$

to ensure all mentioned optimization problems are well defined. On $\mathcal{P}_{L_{\mathcal{U}}}$ we establish the following probability pseudometric

$$d_{L_{\mathcal{U}}}(p,q) := \sup_{x \in \mathbb{X} \cap \mathrm{cl}\,\mathcal{U}} \left| \int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}p(\bar{\omega}) - \int_{\Omega} L(x,\bar{\omega})\,\mathrm{d}q(\bar{\omega}) \right|. \tag{3.8}$$

A uniform distance of the form (3.8) is called a distance having ζ-structure.

Example 3.1. An important class of probability metrics with ζ-structure are the *Fortet-Mourier metrics* defined for $h \geq 1$ by

$$\zeta_h(p,q) := \sup_{L \in \mathcal{L}_h} \left| \int_{\Omega} L(\bar{\omega})\,\mathrm{d}p(\bar{\omega}) - \int_{\Omega} L(\bar{\omega})\,\mathrm{d}q(\bar{\omega}) \right|,$$

where

$$p, q \in \mathcal{P}_h := \left\{ q \in \mathcal{P} : \int_{\Omega} \|\bar{\omega}\|^h \,\mathrm{d}q(\bar{\omega}) < \infty \right\}$$

and $\mathcal{L}_h$ denotes the classes of locally Lipschitz continuous functions that increase with h, i.e.

$$\mathcal{L}_h := \{ \quad L : \Omega \to \mathbb{R} : |L(\bar{\omega}_1) - L(\bar{\omega}_2)| \leq \max\{1, \|\bar{\omega}_1\|, \|\bar{\omega}_2\|\}^{h-1} \|\bar{\omega}_1 - \bar{\omega}_2\|, \\ \forall \bar{\omega}_1, \bar{\omega}_2 \in \Omega \, \}.$$

In the one-dimensional case we can use the following explicit formula

$$\zeta_h(p,q) = \int_{-\infty}^{\infty} \max\{1, |t|^{h-1}\} |G(t) - H(t)| \,\mathrm{d}t,$$

where G, H are distribution functions associated with p, q, see Dupačová and Römisch [20], Rachev [44]. For example, for two 0-1 random variables

$$X_1 = \begin{cases} 0 & \text{with probability } p_1, \\ 1 & \text{with probability } 1 - p_1, \end{cases}$$

and

$$X_2 = \begin{cases} 0 & \text{with probability } q_1, \\ 1 & \text{with probability } 1 - q_1, \end{cases}$$

$p_1, q_1 \in [0, 1]$, with probability distributions p, q we obtain $\zeta_h(p, q) = |p_1 - q_1|$ for $h \geq 1$.

To state the main stability results for optimal decisions we need to introduce the *growth function*

$$\psi_p(\tau) := \min \left\{ \int_\Omega L(x, \bar{\omega})\, \mathrm{d}p(\bar{\omega}) - \vartheta(p) : d(x, \hat{\mathbb{X}}(p)) \geq \tau, x \in (\mathbb{X} \cap \operatorname{cl} \mathcal{U}) \right\}$$

and its inversion $\psi_p^{-1}(t) := \sup \{t \in \mathbb{R}_+ : \psi_p(\tau) \leq t\}$.

Theorem 3.1. *Let $L : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ be a random lower semicontinuous function, $\hat{\mathbb{X}}(p) \neq \emptyset$ and $\mathcal{U} \subset \mathbb{R}^n$ be an open bounded neighbourhood of $\hat{\mathbb{X}}(p)$, where $p \in \mathcal{P}_{L_\mathcal{U}}$.*

Then the multifunction $\hat{\mathbb{X}}_\mathcal{U} : (\mathcal{P}_F, d_F) \to \mathbb{R}^n$ is upper semicontinuous at p and for any $q \in \mathcal{P}_{F_\mathcal{U}}$, the following properties hold

$$|\vartheta(p) - \vartheta_\mathcal{U}(q)| \leq d_{L_\mathcal{U}}(p, q), \tag{3.9}$$

$$\emptyset \neq \hat{\mathbb{X}}_\mathcal{U}(q) \subset \hat{\mathbb{X}}(p) + \psi_p^{-1}(2d_{L_\mathcal{U}}(p, q))\mathbb{B} \tag{3.10}$$

with $\mathbb{B}$ denoting the Euclidean unit ball.

For more general problem a similar result is proved in Römisch [50], theorem 5 and theorem 9. We present here a version of proof for our special problem where the set $\mathbb{X}$ does not depend on probability distribution and we obtain a tighter bound (3.10) for optimal decisions.

Proof. For $x \in \mathbb{X} \cap \operatorname{cl} \mathcal{U}$ and $q \in \mathcal{P}_{L_\mathcal{U}}$ define the function f from $(\mathbb{X} \cap \operatorname{cl} \mathcal{U}) \times \mathcal{P}_{L_\mathcal{U}}$ to $\mathbb{R}$ by $f(x, q) := \int_\Omega L(x, \bar{\omega})\, \mathrm{d}Q(\bar{\omega})$. The function is lower semicontinuous and finite with respect to (3.7), see theorem 3 in Römisch [50]. Hence, $\hat{\mathbb{X}}_\mathcal{U}(q)$ is nonempty for each $q \in \mathcal{P}_{L_\mathcal{U}}$. For $\hat{x} \in \hat{\mathbb{X}}(p)$ and $\tilde{x} \in \hat{\mathbb{X}}_\mathcal{U}(q)$ the inequalities (3.9) follows:

$$\begin{aligned} |\vartheta(p) - \vartheta_\mathcal{U}(q)| &\leq \max\left\{ |\int\limits_\Omega L(\hat{x},\bar{\omega})(q-p)\,(\mathrm{d}\bar{\omega})|, |\int\limits_\Omega L(\tilde{x},\bar{\omega})(p-q)\,(\mathrm{d}\bar{\omega})| \right\} \\ &\leq d_{L_\mathcal{U}}(p,q). \end{aligned}$$

The mapping $\hat{\mathbb{X}}_\mathcal{U}$ is closed at $p \in \mathcal{P}_{L_\mathcal{U}}$ and, hence, upper semicontinuous at p.

By definition of ψ, $d_{L_\mathcal{U}}(p,q)$ and with $\tilde{x} \in \hat{\mathbb{X}}_\mathcal{U}(q) \subset (\mathbb{X} \cap \mathrm{cl}\,\mathcal{U}) =: \mathbb{X}_\mathcal{U}(p)$ we derive

$$\begin{aligned} \psi(d(\tilde{x}, \hat{\mathbb{X}}(p))) &\leq |\int\limits_\Omega L(\tilde{x},\bar{\omega})\,\mathrm{d}p(\bar{\omega}) - \vartheta(q)| \\ &\leq |\int\limits_\Omega L(\tilde{x},\bar{\omega})(p-q)\,(\mathrm{d}\bar{\omega}) + \vartheta_\mathcal{U}(q) - \vartheta(p)| \\ &\leq |\int\limits_\Omega L(\tilde{x},\bar{\omega})(p-q)\,(\mathrm{d}\bar{\omega})| + |\vartheta_\mathcal{U}(q) - \vartheta(p)| \\ &\leq 2d_{L_\mathcal{U}}(p,q). \end{aligned}$$

From here we obtain $d(\tilde{x}, \hat{\mathbb{X}}(p)) \leq \psi_p^{-1}(2d_{L_\mathcal{U}}(p,q))$, which implies (3.10). □

Theorem 3.1 stands as a basic tool for measuring errors caused by employing an inaccurate probability distribution. We illustrated under which conditions it can be declared that small changes of the underlying distribution do neither evoke a significant distance of Bayes actions (3.10) nor difference in suffered losses (3.9).

If, in particular, the problem (3.6) has *k-order growth* at the solution set $\hat{\mathbb{X}}(p)$ for some $k \geq 1$, i.e. $\psi_p(\tau) \geq \gamma\tau^k$ for each small $\tau \in \mathbb{R}^+$ and some $\gamma > 0$, then for $\tilde{x} \in \hat{\mathbb{X}}_\mathcal{U}(q)$ and $p, q \in \mathcal{P}_{L_\mathcal{U}}$,

$$\gamma d(\tilde{x}, \hat{\mathbb{X}}(p))^k \leq \psi(d(\tilde{x}, \hat{\mathbb{X}}(p)) \leq 2d_{L_\mathcal{U}}(p,q).$$

Hence,

$$\emptyset \neq \hat{\mathbb{X}}_\mathcal{U}(q) \subset \hat{\mathbb{X}}(p) + \left(\frac{2}{\gamma} d_{L_\mathcal{U}}(p,q)\right)^{\frac{1}{k}} \mathbb{B}.$$

Localized optimal values $\vartheta_\mathcal{U}(q)$ and solution sets $\hat{\mathbb{X}}_\mathcal{U}(q)$ can be replaced by their global versions $\vartheta(q)$ and $\hat{\mathbb{X}}(q)$, e.g. if the problem (3.6) is convex, $\hat{\mathbb{X}}_\mathcal{U}(q) \subset \mathcal{U}$ and $\exists\delta > 0$ such that $\forall q \in \mathcal{P}_{L_\mathcal{U}} : d_{L_\mathcal{U}}(p,q) < \delta$ (cf. Römisch [50]).

In the next theorem we combine convexity with properties of locally Lipschitz functions. A similar theorem can be also found in Römisch [50].

Theorem 3.2. *Let the assumptions of theorem 3.1 be satisfied. Furthermore, let* $\mathbb{X}$ *be convex and* $L(\cdot, \bar{\omega})$ *be convex on* $\mathbb{X}$ *for each* $\bar{\omega} \in \Omega$. *If there exist constants* $K > 0, h \geq 1$ *such that* $\frac{1}{K} L(x, \cdot) \in \mathcal{L}_h$ *for each* $x \in \mathbb{X} \cap cl\mathcal{U}$ *then* $\exists \delta > 0$ *such that*

$$|\vartheta(p) - \vartheta(q)| \leq K\zeta_h(p, q),$$
$$\emptyset \neq \hat{\mathbb{X}}(q) \subset \hat{\mathbb{X}}(P) + \psi_p^{-1}(2K\zeta_h(p, q))\mathbb{B},$$

whenever $p, q \in \mathcal{P}_h$ *and* $\zeta_h(p, q) < \delta$.

Proof. The statement follows by application of theorem 3.1 and the fact that $\frac{1}{K} L(x, \cdot) \in \mathcal{L}_h$ implies $d_{L_\mathcal{U}}(p, q) \leq K\zeta_h(p, q)$. □

For convex model (3.6) it can be proved, see Rockafellar and Wets [48], theorem 7.69 and Römisch [50], theorem 13, that the ε-minimal solution sets behave Lipschitz continuously in terms of the Pompeiu-Hausdorff distance

$$\mathbb{D}_\infty(C, D) := \inf\{\eta \geq 0 : C \subset D + \eta\mathbb{B}, D \subset C + \eta\mathbb{B}\}$$

defined for nonempty closed sets $C, D \subset \mathbb{R}^n$. By ε-minimal solution set we understand, for some $\varepsilon > 0$, the set

$$\hat{\mathbb{X}}_\varepsilon(p) := \{x \in \mathbb{X} : \int_\Omega L(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}) \leq \vartheta(p) + \varepsilon\}.$$

Theorem 3.3. *Let* L *be a random lower semicontinuous convex function,* $\mathbb{X}$ *closed convex,* $p \in \mathcal{P}_{L_\mathcal{U}}$ *and* $\hat{\mathbb{X}}(p)$ *be nonempty and closed. Then there exist constants* $\rho > 0$ *and* $\bar{\varepsilon} > 0$ *such that the estimate*

$$\mathbb{D}_\infty(\hat{\mathbb{X}}_\varepsilon(p), \hat{\mathbb{X}}_\varepsilon(q)) \leq \frac{2\rho}{\varepsilon} d_{L_\mathcal{U}}(p, q)$$

holds for $\mathcal{U} := (\rho + \bar{\varepsilon})\mathbb{B}$ *and any* $\varepsilon \in (0, \bar{\varepsilon})$, $q \in \mathcal{P}_{L_\mathcal{U}}$ *such that* $d_{L_\mathcal{U}}(p, q) < \varepsilon$.

3.2 Stability with respect to weak convergence of probability measures

Let us return to the definition 3.2 of stability of Bayes actions with respect to weak convergence of $\{p_\nu \in \mathcal{P}, \nu \in \mathbb{N}\}$ to $p \in \mathcal{P}$. We derived that the stability of Bayes actions is under uniform convergence of loss functions equivalent to the convergence

$$\inf_{x \in \mathbb{X}} \int_\Omega f(x, \bar{\omega}) \, \mathrm{d}p_\nu(\bar{\omega}) \xrightarrow[\nu \to \infty]{} \inf_{x \in \mathbb{X}} \int_\Omega f(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}). \tag{3.11}$$

In previous section we introduced how we can measure the distance of the two optimal values from (3.11), cf. (3.9). Now we show under which assumptions this distance converge to 0, i.e. when $\hat{x} \in \hat{\mathbb{X}}(p)$ is stable.

The most cited conditions of stability come from Salinetti [52], we present them in the next theorem.

Theorem 3.4. *Assume that*

(i) $L : \mathbb{X} \times \Omega \to \mathbb{R}$ *is lower semicontinuous on* $\mathbb{X} \times \Omega$*,*

(ii) $L(\hat{x}, \cdot)$ *is continuous on* Ω*,*

(iii) L *has locally equi-lower-bounded growth, i.e.* $\forall x \in \mathbb{X}$ *there exist a neighbourhood* $\mathcal{U}(x)$ *of* x *and* $b(x) \in \mathbb{R}$ *such that for all* $\tilde{x} \in \mathcal{U}(x)$*,*

$$L(\tilde{x}, \bar{\omega}) - L(\hat{x}, \bar{\omega}) \geq b(x), \quad \forall \bar{\omega} \in \Omega.$$

Then $\hat{x} \in \hat{\mathbb{X}}(p)$ *is stable if and only if for any sequence* $p_\nu \xrightarrow[\nu \to \infty]{w} p$ *and every* $\varepsilon > 0$ *the sequence*

$$\left\{ \inf_{x \in \mathbb{X}} \int_{\Omega} f(x, \bar{\omega}) \, dp_\nu(\bar{\omega}), \nu \in \mathbb{N} \right\}$$

has a bounded sequence of ε*-minimal solutions. It means that for any* $\varepsilon > 0$ *there exist a compact subset* $K_\varepsilon \subset \mathbb{X}$ *and a sequence* $\{x_\nu \in K_\varepsilon, \nu \in \mathbb{N}\}$ *such that for all* n

$$\int_{\Omega} f(x_\nu, \bar{\omega}) \, dp_\nu(\bar{\omega}) < \inf_{x \in \mathbb{X}} \int_{\Omega} f(x, \bar{\omega}) \, dp_\nu(\bar{\omega}) + \varepsilon.$$

Assumptions (i) and (ii) imply that f (defined in (3.5)) is lower semicontinuous on $\mathbb{X} \times \Omega$. Condition (iii) is trivially satisfied for L bounded on $\mathbb{X} \times \Omega$. The existence of bounded sequence of ε-minimal decisions is guaranteed e.g. when $\mathbb{X}$ is compact.

To derive other sufficient conditions for stability in sense of definition 3.2 we employ the following representation

$$\int_{\Omega} L(\hat{x}, \bar{\omega}) \, \mathrm{d}p_\nu(\bar{\omega}) - \inf_{x \in \mathbb{X}} \int_{\Omega} L(x, \bar{\omega}) \, \mathrm{d}p_\nu(\bar{\omega}).$$

Adding terms $\pm \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega)$ and $\pm \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega)$ we can write

$$\begin{aligned}
\int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) &- \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) \\
&= \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega) - \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega) \\
&+ \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) - \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega) \\
&+ \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega) - \inf_{x\in\mathbb{X}} \int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) \\
&= \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) - \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega) \\
&- \sup_{x\in\mathbb{X}} \left(- \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega)\right) + \sup_{x\in\mathbb{X}} \left(- \int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega)\right) \\
&\le \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) - \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega) \\
&+ \sup_{x\in\mathbb{X}} \left[\left(- \int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega)\right) - \left(- \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega)\right)\right] \\
&\le |\int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) - \int_\Omega L(\hat x,\bar\omega)\,\mathrm{d}p(\bar\omega)| \\
&+ \sup_{x\in\mathbb{X}} |\int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega) - \int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega)| \\
&\le 2d_L(p_\nu,p),
\end{aligned}$$

where

$$d_L(p_\nu,p) := \sup_{x\in\mathbb{X}} |\int_\Omega L(x,\bar\omega)\,\mathrm{d}p_\nu(\bar\omega) - \int_\Omega L(x,\bar\omega)\,\mathrm{d}p(\bar\omega)|.$$

The problem of stability of Bayes action $\hat x \in \mathbb{X}$ becomes now the task of p-uniformity of a class $\mathcal{L}_x := \{L_x(\cdot) := L(x,\cdot)\ ,\ x \in \mathbb{X}\}$, i.e. under which conditions

$$\lim_{\nu\to\infty} d_L(p_\nu,p) = 0 \tag{3.12}$$

holds true for every p_ν weakly convergent to p.

In Kadane and Srinivasan [32] , theorem 5.1, necessary (A1) and sufficient (A2) conditions ensuring (3.12) can be found for $\mathcal{L}_x$, a class of bounded, real valued, measurable functions defined on Ω:

(A1) $$\sup_{\substack{x\in\mathbb{X}\\ \bar\omega_1,\bar\omega_2\in\Omega}} |L(x,\bar\omega_1) - L(x,\bar\omega_2)| < \infty,$$

(A2) $$\lim_{\varepsilon\downarrow 0} \sup_{x\in\mathbb{X}} \left[\int_\Omega \sup_{\bar\omega_1,\bar\omega_2\in\mathbb{B}(\bar\omega,\varepsilon)} |L(x,\bar\omega_1) - L(x,\bar\omega_2)|\,\mathrm{d}p(\bar\omega)\right] = 0$$

with $\mathbb{B}(\bar\omega,\varepsilon)$ the ball of radius δ centered at $\bar\omega$.

Condition (A1) is satisfied if $\mathcal{L}_x$ is the class of equi-bounded functions. The assumption of equi-continuity of $\mathcal{L}_a$ then implies (A2). Equi-continuity is fulfilled for $L(x, \bar{\omega})$ continuous in $\bar{\omega}$ uniformly in x. Other sufficient condition for equi-continuity is local Lipschitz continuity, i.e. existence of function $g(x, \bar{\omega}), \alpha > 0, \varepsilon > 0$:

(a) $|L(x, \bar{\omega}_1) - L(x, \bar{\omega}_2)| \leq g(x, \bar{\omega}_1) \parallel \bar{\omega}_1 - \bar{\omega}_2 \parallel^{\alpha}$,

(b) $\sup_{x \in \mathbb{X}} \int_{\Omega} g(x, \bar{\omega}) \, \mathrm{d}p(\bar{\omega}) < \infty$.

hold true $\forall \bar{\omega}_2 \in \mathbb{B}(\bar{\omega}_1, \varepsilon)$.

For a detailed discussion of above mentioned requirements on stability see Billingsley and Topsøe [4], Kadane and Chuang [31], Kadane and Srinivasan [32], Lucchetti and Salinetti [37] and Salinetti [52].

Moreover, results presented in section 3.1 can be also applied to sample based Bayes actions (solutions of (3.6) with respect to empirical probability distributions), see Berger and Salinetti [3].

Chapter 4

Minimax approach

Consider a general framework for stochastic programs

$$\inf_{x \in \mathbb{X}} E_p f(x, \omega), \tag{4.1}$$

where $\mathbb{X} \subset \mathbb{R}^n$ denotes a given set of decisions, ω is a random parameter with values in Ω, where Ω is closed subset of $\mathbb{R}^k$, and with a known probability distribution p which does not depend on x. The random objective function f of a decision $x \in \mathbb{X}$ is defined on $\mathbb{X} \times \Omega$, E_p denotes the expectation under p. Assume now that all infima are attained, which is related to the relatively complete recourse, and that all expectations exist.

In many applications of stochastic programs there is some ambiguity about the probability distribution p. The available information on probability distribution can be described by assuming that p belongs to a specific class $\mathcal{P}$ of feasible probability distributions. The most common choices of $\mathcal{P}$ are:

- $\mathcal{P}$ consists of probability distributions which fulfill certain moment conditions, see Dupačová [17], Dupačová [18];

- $\mathcal{P}$ contains probability distributions with additional information, such as symmetry or unimodality, see Čerbáková [12], Dupačová [19], Popescu [40], Shapiro [55];

- $\mathcal{P}$ is a neighborhood of a hypothetical probability distribution p^0, see Pflug and Wozabal [39];

- $\mathcal{P}$ is the set of probability distributions consistent with some qualitative information, see Bühler [6], Bühler [7], Bühler [8], Čerbáková [9];

and their combinations as well. For other examples see Dupačová [21] and references therein.

In this thesis we contemplate on sets of distributions determined by moment conditions, on symmetric and unimodal distributions and on probability distributions consistent with some qualitative information. As a solution technique we choose the minimax approach introduced e.g. in Žáčková [63]. There the decision maker searches for the best protection against the worst-case probability distribution, i.e. he/she solves the problem

$$\min_{x\in\mathbb{X}} \sup_{p\in\mathcal{P}} E_p f(x,\omega). \tag{4.2}$$

In 1963 Iosifescu and Theodorescu [28] defined the solution of stochastic minimax program by the first player optimal mixed strategy in a two-person zero-sum game

$$(\mathbb{X}, \mathcal{P}, F(x,p)) \tag{4.3}$$

with $F(x,p) := E_p f(x,\omega)$. Žáčková [63] introduced the notation of minimax solution as an optimal pure strategy of the first player in the game (4.3). In the most cases the corresponding mixed strategy is unknown and we only have to settle for the knowledge of a set of possible probability distributions.

The minimax solutions exist under quite general assumptions, see e.g. Simons [54]:

Theorem 4.1 (Minimax theorem)**.** *Let $\mathbb{X}$ be a nonempty and convex, compact subset of some linear topological space $\mathcal{X}$ and $\mathcal{P}$ is a nonempty, convex subset of some linear topological space $\mathcal{Y}$. Assume that $F : \mathbb{X}\times\mathcal{P} \to \mathbb{R}$ is lower-semicontinuous and quasi-convex on $\mathbb{X}$ and upper-semicontinuous and quasi-concave on $\mathcal{P}$. Then*

$$\min_{x\in\mathbb{X}} \sup_{p\in\mathcal{P}} F(x,p) = \sup_{p\in\mathcal{P}} \min_{x\in\mathbb{X}} F(x,p).$$

We can apply the minimax decision rule and solve the problem (4.2) also in cases when the Minimax theorem does not hold true. This allows us to construct bounds for the optimal value of (4.1) valid for all $p \in \mathcal{P}$ and provides the worst-case analysis. In literature there are various applications of the minimax approach, we refer e.g. to Dupačová [21] and reference therein.

The following parts of the thesis are denoted just to the minimax approach. In section 4.1 we present possible applications on the moment problem to derive risk measures the worst-case Value-at-Risk and the worst-case Conditional Value-at-Risk under different assumptions on probability distribution of a loss random variable. In section 4.2 and chapters 5 and 6 we consider a set of discrete probability distributions consistent with some qualitative information and derive algorithms for the two-stage and the multistage stochastic minimax problems with linear recourse.

4.1 Moment problem

This part is devoted to possible applications of minimax approach to the moment problem, discussed e.g. in Dupačová [20]. We derive and compare upper bounds for two well-known risk measures Value-at-Risk and Conditional Value-at-Risk under different available information on a probability distribution of the underlying random variable. To overcome the incomplete information about the distribution we will apply the worst-case strategy and define the worst-case VaR and the worst-case CVaR.

We assume that the incomplete information on probability distribution can be described by the set $\mathcal{P}$ of all possible distributions. Specially, our choice of $\mathcal{P}$ will always correspond to specification of the distributions by its first and second order moments. Therefore, the application of general moment problem theory will be basic. The most important results of this chapter come from the dual formulation of moment problem. Essential outcomes in duality are derived in Popescu [40], Shapiro [55], Smith [59]. As the next step, besides the knowledge of expected value and variance, we will suppose other properties of considered distributions — symmetry and unimodality. Results are applied and illustrated on interbank exchange rate data.

Recall the formulation of a general moment problem. We deal with univariate random variable $\omega(\xi)$ with values in Ω, where Ω is nonempty closed subset of $\mathbb{R}$. For the simplicity we omit the argument ξ of ω. A realization $\omega(\xi)$ is denoted by $\bar{\omega}$. The set of all probability distributions of ω defined on $(\Omega, \mathcal{B}(\Omega))$ is denoted by $\mathcal{P}$ which is supposed to be a convex set. Existence of all mentioned expected values is assumed.

Suppose that every finite subset of Ω is $\mathcal{B}(\Omega)$-measurable. For a given vector of functions $g = (g_0, \dots, g_J) : \Omega \to \mathbb{R}$ and a sequence $q = (q_0, \dots, q_J)$ we define the moment problem as

$$\begin{aligned} &\sup_{p \in \mathcal{P}} E_p[f(\omega)] \\ &\text{s.t.} \quad E_p[g(\omega)] = q. \end{aligned} \tag{4.4}$$

Real valued functions $f, g_i, j = 0, \dots, J$, are assumed to be $\mathcal{B}(\Omega)$-measurable and p-integrable. We explicitly include the probability mass constraint by setting $q_0 \equiv 1, g_0 \equiv \mathbb{I}_\Omega$, where $\mathbb{I}$ denotes the indicator function.

Actually, it suffices to solve the problem (4.4) with respect to discrete probability distributions with a finite support on at most $J+1$ points. First results of this type date to the mid 1950.

Lemma 4.1. *Let p be a probability distribution on $(\Omega, \mathcal{B}(\Omega))$ such that real valued functions $g_j, j = 0, \dots, J$, are p-integrable and measurable on $(\Omega, \mathcal{B}(\Omega))$. Then there exists a probability distribution $\hat{p}$ on $(\Omega, \mathcal{B}(\Omega))$ with a finite support of at most $J+1$ points such that $E_p[g_j(\omega)] = E_{\hat{p}}[g_j(\omega)]$ for all $j = 1, \dots, J$.*

For the proof see e.g. Rogosinky [49] or Shapiro [55].

The moment problem restricted to probability distributions with a finite support of at most $S = J+1$ points can be written in the form of a linear programming problem

$$\begin{array}{cl} \sup\limits_{\substack{p \in \mathbb{R}^S_+ \\ \{\bar{\omega}_1, \dots, \bar{\omega}_S\} \subset \Omega}} & \sum\limits_{s=1}^{S} p_s f(\bar{\omega}_s) \\ \text{s.t.} & \sum\limits_{s=1}^{S} p_s g_j(\bar{\omega}_s) = q_j, \quad j = 1, \dots, J, \\ & \sum\limits_{s=1}^{S} p_s = 1. \end{array} \tag{4.5}$$

If we designate Υ_δ as the set of all Dirac distributions $\delta_{\bar{\omega}}, \bar{\omega} \in \Omega$[1], then we can also write

$$\mathcal{P} = \mathrm{cl}(\mathrm{conv}(\Upsilon_\delta)). \tag{4.6}$$

Due to lemma 4.1 and representation (4.6) we can formulate the corresponding dual problem in the following way

$$\begin{array}{l} \inf\limits_{y \in \mathbb{R}^{J+1}} y^T q \\ \text{s.t.} \quad y' g(\bar{\omega}) - f(\bar{\omega}) \geq 0 \quad \forall \bar{\omega} \in \Omega, \end{array} \tag{4.7}$$

as the expected values from (4.7) are related now to Dirac distributions $\delta_{\bar{\omega}}, \bar{\omega} \in \Omega$.

The weak duality holds, i.e. the optimal value of dual problem (4.7) is greater or equal to the optimal value of primal problem (4.5). To fulfil the strong duality we need some additional assumptions, e.g. Slater condition (see Smith [59]):

[1]Recall that Dirac distribution $\delta_{\bar{\omega}}$ is the measure of mass one at $\bar{\omega}$.

Lemma 4.2. *If q is an interior point of $\{q \in \mathbb{R}^{J+1} : E_p[g(\omega)] = q, p \in \mathcal{P}\}$, then the optimal value of problem (4.5) is equal to the optimal value of the corresponding dual problem (4.7).*

For more details about duality and its application on the moment problem we refer to Shapiro [55], Shapiro and Kleywegt [56], Smith [59].

4.1.1 Worst-case VaR and CVaR

Let ω represent random loss variable. We define by $\mathrm{VaR}_\alpha(\omega)$ the minimal level γ such that the probability the random loss ω achieves or exceeds γ is not greater than given $\alpha \in (0, 1]$. We are interested in small α close to 0. Then we solve the problem

$$\begin{aligned} &\mathrm{VaR}_\alpha(\omega) := \min_{\gamma \in \mathbb{R}} \gamma \\ &\text{s.t.} \quad P(\omega \geq \gamma) \leq \alpha. \end{aligned} \tag{4.8}$$

When the distribution of ω is perfectly known, using its left continuous distribution function F_ω we obtain the optimal value

$$\hat{\gamma} = F_\omega^{-1}(1 - \alpha),$$

where F_ω^{-1} is the inverse of F_ω, i.e. $F_\omega^{-1}(\alpha) := \inf\{x : F_\omega(x) \geq \alpha\}$. Specially if ω is normally distributed with expected value μ_ω and variance σ_ω^2 we get

$$\mathrm{VaR}_{\alpha,N}(\omega) = \mu_\omega + \Phi^{-1}(1 - \alpha) \cdot \sigma_\omega,$$

where Φ is the distribution function of univariate normal distribution $N(0, 1)$.

The second considered risk measure is Conditional Value at Risk. CVaR expresses *the expectation of values beyond the VaR*. For $\alpha \in (0, 1)$ it can be defined by

$$\mathrm{CVaR}_\alpha(\omega) := \inf_{a \in \mathbb{R}} \{a + \frac{1}{\alpha} E[(\omega - a)^+]\}, \tag{4.9}$$

where $(c)^+ := \max\{0, c\}$. For more details about VaR and CVaR see e.g. Acerbi and Tasche [1], Pflug [38] or Uryasev and Rockafellar [61].

In most cases we do not exactly know the distribution of random variable ω. We are only able to specify the set $\mathcal{P}$ of feasible probability distributions (e.g. the set of all distributions that fulfil some moment conditions, or the set of symmetric or unimodal distributions, see the introduction to chapter 4). Then we can apply the minimax strategy and define the worst-case VaR, resp. the worst-case CVaR, with respect to the set of considered probability distributions $\mathcal{P}$ by

$$\begin{aligned} \mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}}(\omega) &:= \min_{\gamma\in\mathbb{R}} \gamma \\ \text{s.t.} \quad & \sup_{p\in\mathcal{P}} P_p(\omega \geq \gamma) \leq \alpha. \end{aligned} \tag{4.10}$$

and

$$\mathrm{CVaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}}(\omega) := \sup_{p\in\mathcal{P}} \inf_{a\in\mathbb{R}} \{a + \frac{1}{\alpha} E_p[(\omega - a)^+]\}. \tag{4.11}$$

The worst-case CVaR remains a coherent risk measure, for more details and other possible approach see Zhu and Fukushima [64].

Worst-case Value-at-Risk

Denote the set of all probability distributions of ω fulfilling the first two moment conditions by $\mathcal{P}^m$, i.e. for given μ_ω and $\sigma^2_\omega, \sigma_\omega > 0$, we have

$$\mathcal{P}^m := \{p \in \mathcal{P} : E_p[\omega] = \mu_\omega, E_p[\omega - \mu_\omega]^2 = \sigma^2_\omega\}.$$

Lemma 4.3 (one-sided Chebyshev bound)**.** *Consider the moment problem (4.4) with the set of feasible probability distributions $\mathcal{P}^m$. Then the upper bound for the upper tail $P_p(\omega \geq \gamma)$, $p \in \mathcal{P}^m$, is given by*

$$\sup_{p\in\mathcal{P}^m} P_p(\omega \geq \gamma) = \begin{cases} \frac{\sigma^2_\omega}{\sigma^2_\omega + (\mu_\omega - \gamma)^2} & \text{for } \gamma > \mu_\omega, \\ 1 & \text{for } \gamma \leq \mu_\omega. \end{cases} \tag{4.12}$$

The lemma follows from Isii [29]. Applying this result to the definition of the worst-case VaR (4.10) we get for $\gamma > \mu_\omega$

$$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^m} = \mu_\omega + \sqrt{\frac{1-\alpha}{\alpha}}\sigma_\omega. \tag{4.13}$$

Worst-case Conditional Value-at-Risk

For the given expected value μ_ω and variance σ^2_ω we can rewrite the definition (4.11) of the worst-case CVaR in successive steps. Firstly, due to the finite expected value and convexity of the inner objective function in (4.11) we can interchange supremum and infimum. From Dupačová [20] we obtain

$$\max_{p\in\mathcal{P}^m} E_p[(\omega - a)^+] = \frac{1}{2}[\mu_\omega - a + \sqrt{\sigma^2_\omega + (\mu_\omega - a)^2}].$$

Then by solving the problem

$$\min_{a\in\mathbb{R}} \{a + \frac{1}{2\alpha}[\mu_\omega - a + \sqrt{\sigma^2_\omega + (\mu_\omega - a)^2}]\}$$

we obtain for $\alpha \in (0, \frac{1}{2}]$ and $a > \mu_\omega$ the optimal solution

$$\hat{a} = \mu_\omega + \frac{1-2\alpha}{2\sqrt{\alpha(1-\alpha)}}\sigma_\omega$$

and then the formula

$$\mathrm{CVaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^m}(\omega) = \mu_\omega + \sqrt{\frac{1-\alpha}{\alpha}}\sigma_\omega. \tag{4.14}$$

For a probability distribution identified by its first two moments and for $\gamma > \mu_\omega$ the expression (4.14) coincides with that for $\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^m}(\omega)$.

4.1.2 Worst-case VaR and CVaR for symmetric distributions

We focus on the possibility to improve the obtained upper bounds for the worst-case VaR and the worst-case CVaR with given the first two moments by adding supplemental information on symmetry of the underlying distribution p of ω.

Definition 4.1. Let $\Omega = I \subseteq \mathbb{R}$ be either a compact interval, or $I = \mathbb{R}$. We say that a distribution p of random variable ω defined on $(\Omega, \mathcal{B}(\Omega))$ is μ_ω-symmetric if $p[\mu_\omega - \bar{\omega}, \mu_\omega] = p[\mu_\omega, \mu_\omega + \bar{\omega}]\ \forall \bar{\omega} \in I_{\mu_\omega}$, where $I_{\mu_\omega} := \{\bar{\omega} \geq 0 : \mu_\omega - \bar{\omega} \in I \text{ and } \mu_\omega + \bar{\omega} \in I\}$.

The set $\mathcal{P}^s_{\mu_\omega}$ of all μ_ω-symmetric probability distributions is convex and closed under weak limits. We can also write $\mathcal{P}^s_{\mu_\omega} = \mathrm{cl}(\mathrm{conv}(\Upsilon^s_{\mu_\omega}))$, where

$$\Upsilon^s_{\mu_\omega} = \{p = \frac{1}{2}\delta_{\mu_\omega+\bar{\omega}} + \frac{1}{2}\delta_{\mu_\omega-\bar{\omega}}, \bar{\omega} \in I_{\mu_\omega}\}$$

is the set of μ_ω-symmetric Dirac distributions. For more details see Popescu [40].

We can rewrite the dual problem (4.7) for μ_ω-symmetric distributions in the following way

$$\begin{aligned} &\inf_{y\in\mathbb{R}^{J+1}} y'q \\ &\text{s.t.} \quad y'[g(\mu_\omega - \bar{\omega}) + g(\mu_\omega + \bar{\omega})] - [f(\mu_\omega - \bar{\omega}) + f(\mu_\omega + \bar{\omega})] \geq 0 \quad \forall \bar{\omega} \in I_{\mu_\omega}. \end{aligned} \tag{4.15}$$

The expectations are calculated with respect to distributions from the set $\Upsilon^s_{\mu_\omega}$. If the bound obtained by solving (4.15) is achievable, then there exists an optimal distribution which is a convex combination of $J+1$ μ_ω-symmetric Dirac distributions. It holds under Slater condition, see Popescu [40], Shapiro [55].

Worst-case Value-at-Risk

Lemma 4.4. *Consider the moment problem (4.4) with*

$$\mathcal{P} \equiv \mathcal{P}^{m,s}_{\mu_\omega} := \{p \in P^s_{\mu_\omega} : E_p[\omega] = \mu_\omega, E_p[(\omega-\mu_\omega)^2] = \sigma^2_\omega\}.$$

Then the upper bound for the upper tail $P(\omega \geq \gamma)$ identified by μ_ω and σ^2_ω is given by

$$\sup_{p \in \mathcal{P}^{m,s}_{\mu_\omega}} P_p(\omega \geq \gamma) = \begin{cases} \frac{1}{2}\min\{1, \frac{\sigma^2_\omega}{(\mu_\omega-\gamma)^2}\} & \textit{for } \gamma > \mu_\omega, \\ 1 & \textit{for } \gamma \leq \mu_\omega. \end{cases}$$

The proof can be found in Popescu [40]. Application of results from lemma 4.4 to the definition of the worst-case VaR (4.10) for $\gamma > \mu_\omega$ leads to the problem

$$\begin{array}{ll} \min\limits_{\gamma>\mu_\omega} \gamma & \\ \text{s.t.} & \frac{1}{2}\min\{1, \frac{\sigma^2_\omega}{(\mu_\omega-\gamma)^2}\} \leq \alpha. \end{array}$$

The optimal value is attained for $\alpha < \frac{1}{2}$ and is equal to

$$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s}_{\mu_\omega}} = \mu_\omega + \sqrt{\frac{1}{2\alpha}}\sigma_\omega. \tag{4.16}$$

For $\alpha \geq \frac{1}{2}$ there exists only the infimum μ_ω. If $\gamma \leq \mu_\omega$ then the infimum exists only for $\alpha = 1$ and its value is $-\infty$.

Worst-case Conditional Value-at-Risk

To compute the worst-case CVaR for symmetric probability distributions with the given first two moments we need to calculate an upper bound for the expectation of the term $(\omega - a)^+$.

Lemma 4.5. *Let ω be a random variable with symmetric probability distribution and given finite expected value μ_ω and variance σ^2_ω then*

$$\sup_{\mu \in \mathcal{P}^{m,s}_{\mu_\omega}} E_p[(\omega-a)^+] = \begin{cases} \frac{\sigma^2_\omega}{8(a-\mu_\omega)} & \textit{for } a > \mu_\omega + \frac{\sigma_\omega}{2}, \\ \mu_\omega - a + \frac{\sigma^2_\omega}{8(\mu_\omega - a)} & \textit{for } a < \mu_\omega - \frac{\sigma_\omega}{2}, \\ \frac{\sigma_\omega - a + \mu_\omega}{2} & \textit{for } \mu_\omega - \frac{\sigma_\omega}{2} \leq a \leq \mu_\omega + \frac{\sigma_\omega}{2}, \end{cases} \tag{4.17}$$

Proof. Consider the dual formulation (4.15) of moment problem and substitute $f(\omega) := (\omega - a)^+$, $g(\omega) := (\mathbb{I}_\Omega, \omega, \omega^2)^T$ and $q = (1, \mu_\omega, \sigma_\omega^2 + \mu_\omega^2)$. Then we solve the problem

$$\begin{aligned} &\min_{y\in\mathbb{R}^3} y_0 + \mu_\omega y_1 + (\sigma_\omega^2 + \mu_\omega^2)y_2 \\ &\text{s.t.} \quad 2y_0 + 2\mu_\omega y_1 + 2(\mu_\omega + \bar{\omega})^2 y_2 \geq (\mu_\omega - \bar{\omega} - a)^+ + (\mu_\omega + \bar{\omega} - a)^+ \; \forall \bar{\omega} \geq 0. \end{aligned} \tag{4.18}$$

In order to simplify let $\mu_\omega = 0$. The results for $\mu_\omega \neq 0$ can be derived by substituting $a - \mu_\omega$ for a.

We obtain the following problem

$$\begin{aligned} &\min_{y\in\mathbb{R}^2} \quad y_0 + y_2\sigma_\omega^2 \\ &\text{s.t.} \quad 2y_0 + 2y_2\bar{\omega}^2 \geq (-\bar{\omega} - a)^+ + (\bar{\omega} - a)^+ \qquad \forall \bar{\omega} \geq 0. \end{aligned} \tag{4.19}$$

Evidently $y_2 > 0$. We distinguish three cases:

- $a > 0$:
  Constraints of the problem (4.19) imply
  $$2y_0 + 2y_2\bar{\omega}^2 \geq \begin{cases} 0 & \text{for } 0 \leq \bar{\omega} \leq a, \\ \bar{\omega} - a & \text{for } \bar{\omega} > a. \end{cases} \tag{4.20}$$
  First suppose that the parabola tangents the ray $\bar{\omega} - a$, i.e.
  $$2y_0 + 2y_2 x\bar{\omega}^2 = \bar{\omega} - a.$$
  We get $y_0 = \frac{1}{16y_2} - \frac{a}{2} > -\frac{a}{2}$. In that case the parabola cannot have more than one common point with the $\bar{\omega}$-axis, therefore $2y_0 + 2y_2\bar{\omega}^2 \geq 0$. The discriminant cannot be greater than zero. We realize the condition $y_0 \geq 0$ and thereout $y_2 \leq \frac{1}{8a}$. Problem (4.19) reduces to
  $$\begin{aligned} &\min \frac{1}{16y_2} - \frac{a}{2} + y_2\sigma_\omega^2 \\ &\text{s.t.} \quad 0 < y_2 \leq \frac{1}{8a}. \end{aligned}$$
  In case when Lagrangian multiplier corresponding to the upper constrain is equal to zero we the obtain solution $\hat{y}_2 = \frac{1}{4\sigma_\omega}$, $a \leq \frac{\sigma_\omega}{2}$ and the optimal value $\frac{\sigma_\omega - a}{2}$. Otherwise for a nonzero multiplier is $\hat{y}_2 = \frac{1}{8a}$, $a > \frac{\sigma_\omega}{2}$ and the optimal value is equal to $\frac{\sigma_\omega^2}{8a}$. If we assume that the parabola $2y_0 + 2y_2\bar{\omega}^2$ tangents the $\bar{\omega}$-axis, we obtain the same results.

- $a < 0$:
  Constraints of the problem (4.19) are reduced to
  $$2y_0 + 2y_2\bar{\omega}^2 \geq \begin{cases} -2a & \text{for } 0 \leq \bar{\omega} \leq -a, \\ \bar{\omega} - a & \text{for } \bar{\omega} > -a. \end{cases} \tag{4.21}$$

At first let the parabola $2y_0 + 2y_2\bar{\omega}^2$ tangent the ray $\bar{\omega} - a$. We realize the condition $y_0 = \frac{1}{16y_2} - \frac{a}{2} > -\frac{a}{2}$. The parabola cannot have more than one common point with the ray $-2a$. By analogy with the case $a > 0$ we get $y_0 \geq -a$ and $y_2 \leq -\frac{1}{8a}$. We obtain the optimal values $\frac{\sigma_\omega - a}{2}$ for $-a \leq \frac{\sigma_\omega}{2}$ and $-a - \frac{\sigma_\omega^2}{8a}$ for $-a > \frac{\sigma_\omega}{2}$. The same result is obtained when assuming that the parabola tangents the ray $-2a$.

- $a = 0$:
  We solve problem (4.19) under the condition

$$2y_0 + 2y_2\bar{\omega}^2 \geq \bar{\omega} \quad \text{for } \bar{\omega} \geq 0.$$

The parabola cannot have more than one common point with the axis of the first quadrant, from here it follows $y_0 \geq \frac{1}{16y_2}$. The problem (4.19) reduces to

$$\min_{y_2 > 0} \frac{1}{16y_2} + y_2\sigma_\omega^2.$$

The minimum is achieved for $\hat{y}_2 = \frac{1}{4\sigma_\omega}$. The optimal value is equal to $\frac{\sigma_\omega}{2}$.

□

The proof can be also found in Čerbáková [10]. If the bounds of lemma 4.4 are achievable, then they are attained for a discrete distribution concentrated on a finite support.

The worst-case CVaR for symmetric distribution identified by its first two moments is then calculated as

$$\mathrm{CVaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s}_{\mu_\omega}}(\omega) := \inf_{a\in\mathbb{R}}\{a + \frac{1}{\alpha}\sup_{\mu\in\mathcal{P}^{m,s}_{\mu_\omega}} E_\mu[(\omega - a)^+]\}.$$

Applying lemma 4.5 we obtain the solution

$$\mathrm{CVaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s}_{\mu_\omega}}(\omega) = \begin{cases} \mu_\omega + \sqrt{\frac{1}{2\alpha}}\sigma_\omega & \text{for } \alpha < \frac{1}{2}, \\ \mu_\omega + \sqrt{\frac{1-\alpha}{2}}\frac{\sigma_\omega}{\alpha} & \text{for } \alpha > \frac{1}{2}, \\ \mu_\omega + \sigma_\omega & \text{for } \alpha = \frac{1}{2}. \end{cases} \tag{4.22}$$

Note, that for $\alpha < \frac{1}{2}$ the worst-case VaR and the worst-case CVaR are identical.

4.1.3 Worst-case VaR for symmetric and unimodal distributions

Assume that except the knowledge of the first two moments of ω we have the information p is symmetric and unimodal.

Definition 4.2. Let $\Omega = I \subseteq \mathbb{R}$ be either a compact interval, or $I = \mathbb{R}$. Distribution ω of random variable X defined on $(\Omega, \mathcal{B}(\Omega))$ is said to be μ_ω-unimodal on $I \ni \mu_\omega$ if the corresponding distribution function is convex on the left of μ_ω and concave on the right of μ_ω.

Let $\mathcal{P}^u_{\mu_\omega}$ denotes the set of continuous μ_ω-unimodal probability distributions. Then $\mathcal{P}^u_{\mu_\omega}$ is convex and $\mathrm{cl}(\mathcal{P}^u_{\mu_\omega})$ is the set of all μ_ω-unimodal distributions.

For univariate real random variables we can find continuous transformation from the set of all probability distributions $\mathcal{P}$ to the set of all unimodal distributions $\mathrm{cl}(\mathcal{P}^u_{\mu_\omega})$. This result has been presented e.g. in Dupačová [20]. In consequence, we can write $\mathrm{cl}(\mathcal{P}^u_{\mu_\omega}) = \mathrm{cl}(\mathrm{conv}(\mathcal{T}^u_m))$, where

$$\Upsilon^u_{\mu_\omega} = \{\delta_{[\bar{\omega},\mu_\omega]} : \bar{\omega} \in I, \bar{\omega} \neq \mu_\omega\}.$$

By $\delta_{[a,b]}$ we denote probability distribution with uniform density on $[a, b]$. For other properties of unimodal distributions, see Shapiro [57].

We obtain the convex set of all μ_ω-symmetric unimodal distributions as $\mathrm{cl}(\mathcal{P}^{s,u}_{\mu_\omega}) = \mathcal{P}^s_{\mu_\omega} \cap \mathrm{cl}(\mathcal{P}^u_{\mu_\omega}) = \mathrm{cl}(\mathrm{conv}(\Upsilon^{s,u}_{\mu_\omega}))$, where

$$\Upsilon^{s,u}_{\mu_\omega} = \{\delta_{[\mu_\omega-\bar{\omega},\mu_\omega+\bar{\omega}]} : \bar{\omega} \in I_{\mu_\omega}, \bar{\omega} \neq 0\}.$$

For details see Popescu [40].

The corresponding dual problem (4.7) is

$$\begin{aligned} &\inf_{y\in\mathbb{R}^{J+1}} y'q \\ &\text{s.t.} \quad y' \int_{\mu_\omega-\bar{\omega}}^{\mu_\omega+\bar{\omega}} g(z)\,\mathrm{d}z - \int_{\mu_\omega-\bar{\omega}}^{\mu_\omega+\bar{\omega}} f(z)\,\mathrm{d}z \geq 0 \quad \forall \bar{\omega} \in I_{\mu_\omega}, \bar{\omega} \neq 0, \\ &\qquad y'g(\mu_\omega) - f(\mu_\omega) \geq 0. \end{aligned} \tag{4.23}$$

The last condition arises from adding δ_{μ_ω} to the generating set $\Upsilon^{s,u}_{\mu_\omega}$ in order to obtain closure.

By analogy with symmetric case, the optimal distribution exists if the bound is achievable. Then it is a convex combination of $J+1$ μ_ω-symmetric unimodal distributions, possibly including a Dirac distribution at μ_ω.

Denote the set of all μ_ω-unimodal and symmetric probability distribution fulfilling the first two moment conditions by

$$\mathcal{P}^{m,s,u}_{\mu_\omega} := \{p \in P^s_{\mu_\omega} : E_p[\omega] = \mu_\omega, E_p[(\omega-\mu_\omega)^2] = \sigma^2_\omega\}.$$

Lemma 4.6. *Let ω be a random variable with symmetric unimodal distribution and given finite expected value μ_ω and variance σ_ω^2 then*

$$\sup_{\mu \in cl(\mathcal{P}_{\mu_\omega}^{m,s,u})} P_p[\omega \geq \gamma] = \begin{cases} \frac{1}{2}\min\{1, \frac{4}{9}\frac{\sigma_\omega^2}{(\gamma-\mu_\omega)^2}\} & \text{for } \gamma > \mu_\omega, \\ 1 & \text{for } \gamma \leq \mu_\omega. \end{cases} \tag{4.24}$$

Proof. We substitute $q = (1, \mu_\omega, \sigma_\omega^2)$, $f(\omega) = \mathbb{I}_{[\omega\geq\gamma]}$, $g_0(\omega) = \mathbb{I}_{[\omega\in\Omega]}$, $g_1(\omega) = \omega$ and $g_2(\omega) = (\omega - \mu_\omega)^2$ to (4.23) and obtain the problem

$$\begin{aligned} &\min_{y\in\mathbb{R}^3} y_0 + \mu_\omega y_1 + \sigma_\omega^2 y_2 \\ &\text{s.t.} \quad 2\bar{\omega}y_0 + 2\bar{\omega}\mu_\omega y_1 + \tfrac{2}{3}\bar{\omega}^3 y_2 \geq \int_{\mu_\omega-\bar{\omega}}^{\mu_\omega+\bar{\omega}} \mathbb{I}_{[z\geq\gamma]}\,\mathrm{d}z \quad \forall\bar{\omega} > 0, \bar{\omega} \neq 0 \\ &\qquad y_0 + \mu_\omega y_1 \geq \mathbb{I}_{[\mu_\omega\geq\gamma]}. \end{aligned} \tag{4.25}$$

Without loss of generality, we may assume that $\mu_\omega = 0$. The results for $\mu_\omega \neq 0$ can be derived by substituting $\gamma - \mu_\omega$ for γ. We distinguish two cases:

- $\gamma \leq 0$:
  We solve the following problem

$$\begin{aligned} &\min_{y\in\mathbb{R}^2} y_0 + \sigma_\omega^2 y_2 \\ &\text{s.t.} \quad y_0 + \tfrac{1}{3}\bar{\omega}^2 y_2 \geq 1 \qquad \text{for } 0 < \bar{\omega} \leq -\gamma, \\ &\qquad 2\bar{\omega}y_0 + \tfrac{2}{3}\bar{\omega}^3 y_2 \geq \bar{\omega} - \gamma \quad \text{for } \bar{\omega} > -\gamma, \\ &\qquad y_0 \geq 1. \end{aligned} \tag{4.26}$$

  By the first condition we must have $y_2 \geq 0$. The optimum is achieved for $y_0 = 1$, $y_2 = 0$. The bound is equal to 1.

- $\gamma > 0$:
  We reformulate the problem (4.25) as

$$\begin{aligned} &\min_{y\in\mathbb{R}^2} y_0 + \sigma_\omega^2 y_2 \\ &\text{s.t.} \quad \bar{\omega}y_0 + \tfrac{1}{3}\bar{\omega}^3 y_2 \geq 0 \qquad \text{for } 0 < \bar{\omega} < \gamma, \\ &\qquad \bar{\omega}(2y_0 - 1) + \tfrac{2}{3}\bar{\omega}^3 y_2 \geq -\gamma \quad \text{for } \bar{\omega} \geq \gamma, \\ &\qquad y_0 \geq 0. \end{aligned} \tag{4.27}$$

  The first constraint implies $y_2 \geq 0$. For $y_0 \geq \frac{1}{2}$, or $y_2 = 0$, the optimum is achieved for $y_0 = \frac{1}{2}$, $y_2 = 0$, and the bound is $\frac{1}{2}$.
  In the case when $0 \leq y_0 < \frac{1}{2}$ and $y_2 > 0$ the second condition holds for $\bar{\omega} \geq \gamma$ iff $g(\tilde{\omega}) \geq 0$, where $\tilde{\omega} = \sqrt{\frac{1-2y_0}{2y_2}}$ is the non-negative local

minimum of the function $g(\omega) = \gamma + \bar{\omega}(2y_0 - 1) + \frac{2}{3}\bar{\omega}^3 y_2$. The requirement $g(\tilde{\omega}) \geq 0$ is equivalent to the constraint $y_2 \geq \frac{2}{9}\frac{(1-2y_0)^3}{\gamma^2}$. The minimum is achieved at $y_0 = 0$ and $y_2 = \frac{2}{9\gamma^2}$, the optimal value is $\frac{2}{9}\frac{\sigma_\omega^2}{\gamma^2}$.

□

The proof was presented in Popescu [40] and corrected in Čerbáková [9]. By analogy with symmetric case, the discrete optimal distribution exists if the bound is achievable.

Now we can handle the univariate worst-case VaR for symmetric unimodal distribution of loss random variable ω with known expected value μ_ω and variance σ_ω^2. The optimal value is solution of the problem

$$\begin{aligned} &\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s,u}_{\mu_\omega}}(\omega) := \min_{\gamma > \mu_\omega} \\ &\text{s.t.} \quad \tfrac{1}{2}\min\{1, \tfrac{4}{9}\tfrac{\sigma_\omega^2}{(\mu_\omega - \gamma)^2}\} \leq \alpha. \end{aligned} \tag{4.28}$$

For $\alpha < \frac{1}{2}$ the minimum is

$$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s,u}_{\mu_\omega}}(\omega) = \mu_\omega + \sqrt{\frac{2}{9\alpha}}\sigma_\omega.$$

For $\alpha \geq \frac{1}{2}$ there exists only the infimum μ_ω. If $\gamma \leq \mu_\omega$ then the infimum exists only for $\alpha = 1$ and its value is $-\infty$.

4.1.4 Numerical illustrations

The assumption of normality is frequently used in practice. When it is not fulfilled, the maximal error in VaR, resp. CVaR, caused by holding to the fixed normal distribution can be evaluated as the difference between the worst-case VaR, resp. the worst-case CVaR, and VaR, resp. CVaR, value provided by a specific distribution.

This numerical study compares the worst-case VaR for arbitrary, normal, symmetric and symmetric unimodal distributions with the given first two moments. Results for a random variable ω with expected value $\mu_\omega = 0$ and variance $\sigma_\omega^2 = 1$ are summarized in Table 4.1 Figure 4.1.

In the next example we study interbank exchange rate increments between Norwegian Kroner NOK and Swedish Krona SEK during the time period 1.1.2004 – 31.12.2004, see Figure 4.2. Our data represent increments of daily closing values to insure data independence. We have chosen these two currencies because of their long-standing stability. Hence, there is a perspective to fulfil the assumptions of symmetry, respectively unimodality.

α	$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^m}$	$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s}_{\mu_\omega}}$	$\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s,u}_{\mu_\omega}}$	$\mathrm{VaR}_{\alpha,N}$
0.01	9.9499	7.0711	4.7140	2.3263
0.02	7.0000	5.0000	3.3333	2.0537
0.03	5.6862	4.0825	2.7217	1.8808
0.04	4.8990	3.5355	2.3570	1.7507
0.05	4.3589	3.1623	2.1082	1.6449
0.1	3.0000	2.2361	1.4907	1.2816

Table 4.1: Comparison of the worst-case VaR for random variable ω with $\mu_\omega = 0$ and $\sigma^2_\omega = 1$, where $\mathrm{VaR}_{\alpha,N}$ is VaR computed for normal distribution.

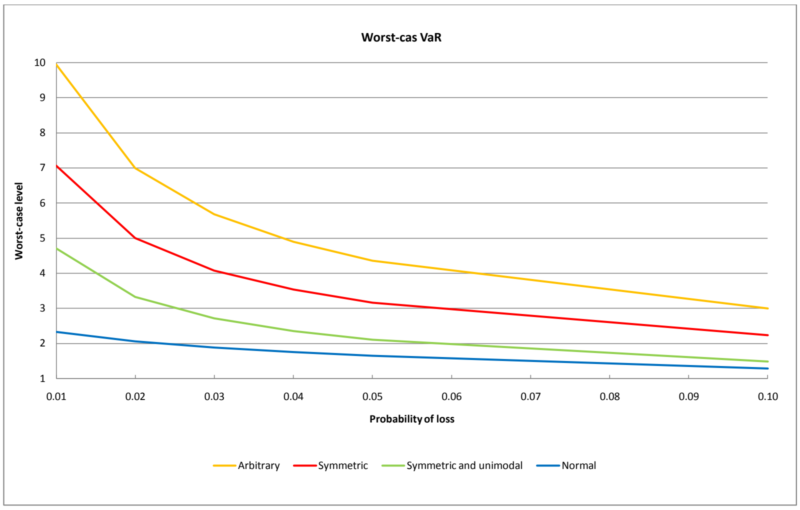

Figure 4.1: Comparison of the worst-case VaR for random variable ω with $\mu_\omega = 0$ and $\sigma^2_\omega = 1$ in dependence on α.

We apply the last 30 observations to estimate expected value and variance for a calculation of the next worst-case level. The results are presented in Figure 4.3. Real values of increments achieve or exceed the 5% worst-case value $\mathrm{VaR}^{\mathrm{wc}}_{\alpha,\mathcal{P}^{m,s,u}_{\mu_\omega}}$ only in 3%, whereas the level $\mathrm{VaR}_{\alpha,N}$ in 6%. A possible improvement of these bounds is in using robust estimations of expected value and variance.

We tested symmetry by a test derived in Gupta [25]. This test is one of the classical tests of symmetry based on the observed skewness. We test null hypothesis that the coefficient of skewness is equal to zero against the alternative that it is different from zero. The null hypothesis is rejected at the probability level $\alpha = 0.05$ only in few cases. For testing unimodality we use a Hartigan's DIP test of unimodality derived in Hartigan and Har-

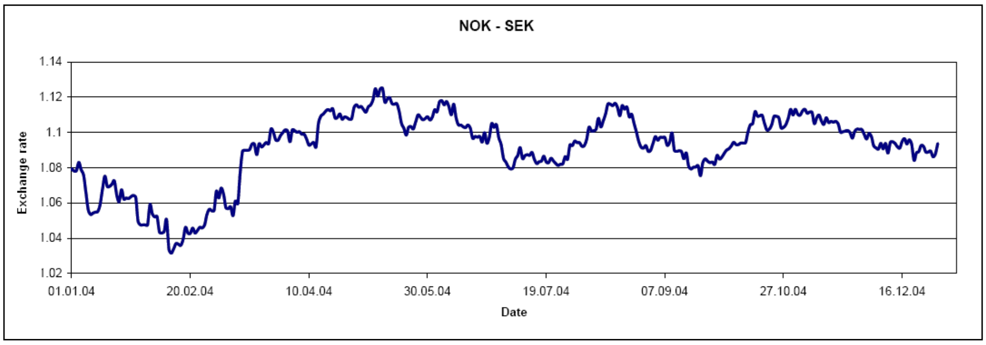

Figure 4.2: Interbank exchange rate between Norwegian Kroner NOK and Swedish Krona SEK.

tigan [26]. DIP is defined as the maximum distance between the empirical distribution and the best fitting unimodal distribution. We test unimodality against the alternative of multimodal probability distribution. We do not reject unimodality at the probability level $\alpha = 0.05$ in these intervals: 1.1.2004 – 27.5.2004, 24.6.2004 – 27.8.2004, 19.9.2004 – 17.10.2004, 7.11.2004 – 26.11.2004. We also tested normality by D'Agostino test. The normality was rejected at probability level $\alpha = 0.05$ approximately in one third of all cases. Therefore, it makes sense to study the release of assumptions about the underlying probability distribution.

The numerical experiments and tests of symmetry, unimodality and normality have been computed in R 1.7.0 using packages base and diptest, see [43].

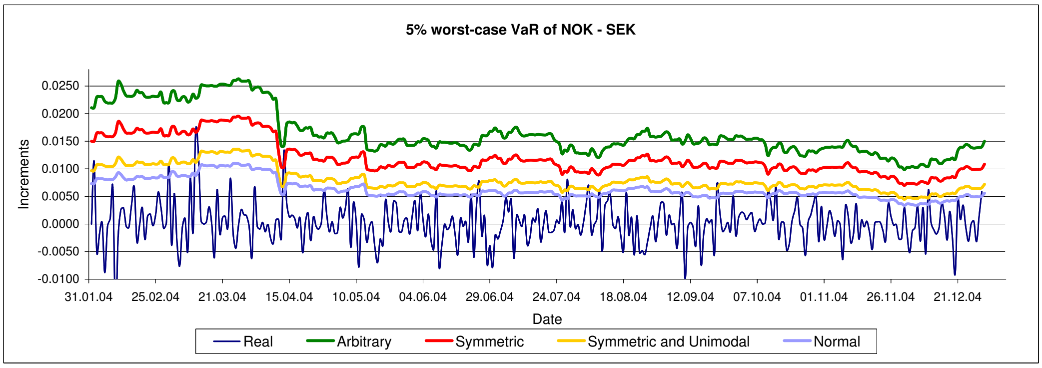

Figure 4.3: The worst-case VaR for increments of NOK - SEK interbank exchange rate.

4.2 Qualitative information

Assume that $\mathcal{P}$ is the set of probability distributions with a finite support $\Omega = \{\bar{\omega}_1, \ldots, \bar{\omega}_S\}$ which are consistent with some qualitative information. By qualitative information we understand an expert's opinion formulating probability relations of the type "realization $\bar{\omega}_i$ is not less probable than realization $\bar{\omega}_j$" $(1 \leq, i, j \leq S)$. This probability relation will be denoted by $\succeq$.

From the intuitive meaning of $\succeq$ it is clear that $\succeq$ should be reflexive and transitive, i.e. we assume:

1. There exist at least two comparable realizations $\bar{\omega}_i, \bar{\omega}_j \in \Omega$. We are able to decide, wheatear the realization $\bar{\omega}_i$ is not less probable than the realization $\bar{\omega}_j$ or vice-versa the realization $\bar{\omega}_j$ is not less probable than the realization $\bar{\omega}_i$.

2. Consider $\bar{\omega}_i \in \Omega$ which is not less probable than $\bar{\omega}_j \in \Omega$ and in addition $\bar{\omega}_j \in \Omega$ is not less probable than $\bar{\omega}_k \in \Omega$. Then $\bar{\omega}_i$ is not less probable than $\bar{\omega}_k$.

3. For all realizations $\bar{\omega} \in \Omega$ it holds that $\bar{\omega}$ is not less probable than $\bar{\omega}$.

Hence, the qualitative information $\succeq$ is a weak partial order on $\Omega \times \Omega$. Each partial order can be represented by a directed graph. We denote the directed graph corresponding with $\succeq$ by Γ.

The set of probability distributions consistent with a qualitative information is then defined by

$$\mathcal{P} = \Big\{ p = (p_1, \ldots, p_S)^T \in \mathbb{R}^S_+ : \ p_i \geq p_j \text{ if } \bar{\omega}_i \succeq \bar{\omega}_j, \\ \bar{\omega}_i, \bar{\omega}_j \in \Omega, i, j = 1, \ldots, S, \sum_{j=1}^{S} p_j = 1 \Big\}.$$

By p_i we denote the probability of realization $\bar{\omega}_i$, i.e. $p_i = p(\bar{\omega}_i)$ for $i = 1, \ldots, S$.

Such a set is a bounded polyhedron and a maximum of linear function of p is attained at least at one of its extreme points, see lemma 2.2. Therefore a characterization of its extreme points is crucial. We remind the definition of extreme points:

Definition 4.3. The probability distribution $p \in \mathcal{P}$ is an *extreme point* of $\mathcal{P}$ if there do not exist any probability distributions $\tilde{p}, \hat{p} \in \mathcal{P}, \tilde{p} \neq \hat{p}$ and $\lambda \in (0, 1)$ such that

$$p = \lambda \tilde{p} + (1 - \lambda)\hat{p}.$$

To simplify the computation of inner optimization problem of (4.2) we define an *admissible support* of $p \in \mathcal{P}$ by

$$\mathcal{T}(p) := \{i \in \{1, \dots, S\} : p_i > 0\}. \tag{4.29}$$

If $p = (p_1, \dots, p_S)^T \in \mathcal{P}$ is an extreme point then

$$p_s = \begin{cases} \frac{1}{|\mathcal{T}(p)|} & \text{for } s \in \mathcal{T}(p), \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

Moreover, the set $\mathcal{T}(p)$ is connected in the sense, that between two different realizations $\bar{\omega}_i \neq \bar{\omega}_j$ such that $i, j \in \mathcal{T}(p)$ (also nodes of graph Γ) there exists a connecting undirected path in Γ. For more details see also Bühler [6] and Bühler [8]. The following theorem summarizes properties of extreme points of the set of discrete probability distributions consistent with qualitative information $\succeq$.

Theorem 4.2. *Probability distribution $p = (p_1, \dots, p_S)^T \in \mathcal{P}$ defined on a finite set $\Omega = \{\bar{\omega}_1, \dots, \bar{\omega}_S\}$ is an extreme point of $\mathcal{P}$, the set of all probability distributions consistent with qualitative information $\succeq$, iff for all $\bar{\omega}_i, \bar{\omega}_j \in \Omega$ the following conditions hold*

(i) $p_i > 0$ *and* $p_j > 0 \Rightarrow p_i = p_j$,

(ii) $\bar{\omega}_i \succeq \bar{\omega}_j \Rightarrow p_i \geq p_j$,

(iii) $p_i = p_j > 0 \Rightarrow$ *there exists an undirected path in graph Γ connecting nodes $\bar{\omega}_i, \bar{\omega}_j$. Each node on this path has nonzero probability.*

Proof. Condition (ii) ensures that the distribution p is consistent with the qualitative information $\succeq$.

$\Leftarrow$: Firstly assume that the distribution p is an extreme point of the set $\mathcal{P}$. From the equality (4.30) it follows

$$p_s = \begin{cases} \frac{1}{|\mathcal{T}(p)|} & \text{for } s \in \mathcal{T}(p), \\ 0 & \text{otherwise.} \end{cases} \tag{4.31}$$

Hence, for all $\bar{\omega}_i, \bar{\omega}_j \in \Omega$ such that $p_i > 0, p_j > 0$, we obtain $\bar{\omega}_i, \bar{\omega}_j \in \mathcal{T}(p)$ and consequently

$$p_i = p_j = \frac{1}{|\mathcal{T}(p)|}.$$

The condition (i) is satisfied.

Let $p_i = p_j > 0$. The set $\mathcal{T}(p)$ is connected, i.e. between any two realizations $\bar{\omega}_i \neq \bar{\omega}_j$ such that $i, j \in \mathcal{T}(p)$, there exists an undirected path $\{i, i_1\}, \dots, \{i_m, j\}$ in graph Γ connecting realizations $\bar{\omega}_i, \bar{\omega}_{i_j}, \dots, \bar{\omega}_{i_m}, \bar{\omega}_j$ with $i_1, \dots, i_m \in \mathcal{T}(p)$. Therefore, $p_{i_1} > 0, \dots, p_{i_m} > 0$. The condition (iii) is fulfilled.

$\Rightarrow$: Suppose that $p \in \mathcal{P}$ agrees with assumptions (i) – (iii), but it is not an extreme point of $\mathcal{P}$, i.e. there exist $\tilde{p}, \hat{p} \in \mathcal{P}$ and $\lambda \in (0,1) : \tilde{p} \neq \hat{p}$ such that $p = \lambda \hat{p} + (1-\lambda)\tilde{p}$. It follows that

$$\mathcal{T} := \{i \in \mathcal{T}(p) : \tilde{p}_i > p_i\} = \{i \in \mathcal{T}(p) : \hat{p}_i < p_i\} \neq \emptyset.$$

Evidently, $\mathcal{T} \subset \mathcal{T}(p)$. Consider $j \in \mathcal{T}(p) \setminus \mathcal{T} \neq \emptyset$ and $i \in \mathcal{T}$. In graph Γ there exists an undirected path $\{i, i_1\}, \dots, \{i_n, j\}$ between nodes i, j such that all nodes on this path have nonzero probability. Assumptions (i) then implies $p_i = p_{i_1} = \dots = p_{i_n} = p_j > 0$.

From $i \in \mathcal{T}$ it follows

$$\tilde{p}_i > p_i = p_{i_1} = \dots = p_{i_n} = p_j,$$

$$\hat{p}_i < p_i = p_{i_1} = \dots = p_{i_n} = p_j.$$

Hence, $\mathcal{T} = \mathcal{T}(p)$ and it must hold

$$\sum_{i=1}^{S} \tilde{p}_i > 1 = \sum_{i=1}^{S} p_i > \sum_{i=1}^{S} \hat{p}_i.$$

This is a contradiction with $\tilde{p}, \hat{p} \in \mathcal{P}$ and hence p must be an extreme point of $\mathcal{P}$.

□

These results were discussed in Bühler [7] and detailed in Čerbáková [9]. Theorem 4.2 implies that all extreme points of $\mathcal{P}$ are generalized discrete distributions defined on a connected undirected subgraph of Γ.

Let $\{p^1, \dots, p^K\}$ be the set of all extreme points of $\mathcal{P}$ and define $\mathcal{T}_k := \mathcal{T}(p^k), k = 1, \dots, K$. Then we may significantly simplify the calculation of problem (4.2) as follows:

$$\min_{x \in \mathbb{X}} \max_{p \in \mathcal{P}} \sum_{i=1}^{S} p_i f(x, \bar{\omega}_i) = \min_{x \in \mathbb{X}} \max_{k \in \{1,\dots,K\}} \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} f(x, \bar{\omega}_s). \tag{4.32}$$

Example 4.1. Consider the set $\mathcal{P}$ defined on $\Omega = \{\bar\omega_1, \bar\omega_2, \bar\omega_3, \bar\omega_4\}$ of probability distributions consistent with the qualitative information

$$\bar\omega_1 \succeq \bar\omega_3, \bar\omega_4 \succeq \bar\omega_3.$$

Then $\mathcal{P} := \{p = (p_1, p_2, p_3, p_4)^T \in \mathbb{R}^4_+ : p_1 \geq p_3, p_4 \geq p_3, \sum_{i=1}^{4} p_i = 1\}$.

This information structure can be described by the directed graph Γ presented in the figure 1. Each oriented path in the graph corresponds to one probability relation.

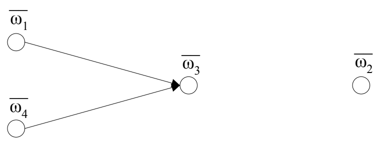

Figure 4.4: Graph Γ.

There exist 4 extreme points which fulfil the conditions of theorem 4.2. Namely, $p^1 = (1, 0, 0, 0)^T, p^2 = (0, 1, 0, 0)^T, p^3 = (0, 0, 0, 1)^T, p^4 = (\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3})^T$ with the corresponding admissible supports $\mathcal{T}_1 = \{1\}$, $\mathcal{T}_2 = \{2\}$, $\mathcal{T}_3 = \{3\}$, $\mathcal{T}_4 = \{1, 3, 4\}$.

No other admissible support generating an extreme point of $\mathcal{P}$ can be obtained. Confront e.g. $p^5 = (\frac{1}{2}, \frac{1}{2}, 0, 0)^T \in \mathcal{P}$ and $\mathcal{T}_5 = \{1, 2\}$. This vector cannot be an extreme point of $\mathcal{P}$ because there does not exist any undirected path in Γ connecting nodes $\bar\omega_1$ and $\bar\omega_2$.

In the following chapters 5 and 6 we introduce algorithms for solving the two-stage and the multi-stage stochastic minimax programs with linear recourse with the set of probability distributions consistent with the qualitative information $\succeq$ introduced in this section. However, the presented algorithms work also for more general sets of probability distributions whenever $\mathcal{P} = \text{conv}\{p^1, \ldots, p^K\}$, where $p^i, i = 1, \ldots, K$, are extreme points of $\mathcal{P}$. Such an example is

$$\mathcal{P} = \{p \in \mathbb{R}^S_+ : Dp \leq 0, \sum_{j=1}^{S} p_j = 1\},$$

where D is a fixed matrix composed of $0, 1$ and -1. However, in general the characterization of extreme points of $\mathcal{P}$ can be difficult, cf. Bühler [6] and references therein.

Chapter 5

Minimax rule in two-stage programs

This chapter deals with the minimax approach applied to the two-stage stochastic programs with linear recourse and with discrete probability distributions of random parameters consistent with the qualitative information $\succeq$ introduced in the section 4.2. We deal with two different situations. Firstly, it is assumed that the considered problem always has an optimal solution, i.e. the assumption of relatively complete recourse holds. The second situation does not require this assumption any more. For both cases a modified multi-cut L-shaped decomposition method for solving two-stage minimax problems is developed. Finally, other possible solution techniques described in literature are discussed.

Consider the minimax two-stage problem with linear recourse, i.e.

$$\min_{x \in \mathbb{X}} \sup_{p \in \mathcal{P}} \left\{ c^T x + E_p \Phi(x, \omega) \right\}, \tag{5.1}$$

where the random vector $\omega(\xi) = \{q(\xi), b(\xi), W(\xi), T(\xi)\}$ constitutes of the components of the second-stage cost $q(\xi) \in \mathbb{R}^{n_2}$, the second-stage right-hand side $b(\xi) \in \mathbb{R}^m$, the recourse matrix $W(\xi)_{m \times n_2}$ and the technology matrix $T(\xi)_{m \times n_1}$. The second-stage value function is given by

$$\Phi(x, \omega(\xi)) := \min \left\{ q(\xi)^T y : W(\xi) y = b(\xi) - T(\xi) x, y \in \mathbb{R}^{n_2}_+ \right\}.$$

The nonempty set of feasible first-stage decisions is denoted by $\mathbb{X} = \{x \in \mathbb{R}^{n_1}_+ : Ax = d\}$, where A and d are known matrix and vector with corresponding dimensions. It is also assumed that $c \in \mathbb{R}^{n_1}$ is a known vector. The set of all feasible probability distributions of ω defined on $(\Omega, \mathcal{B}(\Omega))$ is denoted by $\mathcal{P}$. For the sake of simplicity the argument ξ of ω will be omitted where it is not necessary and we will write $\bar{\omega}$ for a realization $\omega(\xi)$.

Let the following assumptions hold:

(A1) The set of feasible distributions $\mathcal{P}$ is defined as the set of all probability distributions p with finite support in fixed finite set $\Omega = \{\bar{\omega}_1, \ldots, \bar{\omega}_S\}$, which are consistent with the qualitative information $\succeq$. We establish the notation $\bar{\omega}_s = \{q_s, b_s, W_s, T_s\}$, $s = 1, \ldots, S$.

(A2) For all $x \in \mathbb{X}$ and $s \in \{1, \ldots, S\}$ there exists $y \in \mathbb{R}^{n_2}_+$ such that $W_s y = b_s - T_s x$.

(A3) For all $s \in \{1, \ldots, S\}$ there exists $\pi \in \mathbb{R}^m$ such that $\pi^T W_s \le (q_s)^T$.

Relatively complete recourse assumption (A2) ensures feasibility of the second-stage problem. Assumption (A3) of dual feasibility is applied to ensure boundedness of the second-stage problem.

Applying assumption (A1) we may reformulate the problem (5.1) in terms of the scenario probabilities $p_1, \ldots, p_S$, as follows

$$\min_{x \in \mathbb{X}} \left\{ c^T x + \max_{p \in \mathcal{P}} \sum_{s=1}^{S} p_s \Phi(x, \bar{\omega}_s) \right\}, \tag{5.2}$$

where

$$\mathcal{P} = \left\{ p \in \mathbb{R}^S_+ : p_i \ge p_j \text{ if } \bar{\omega}_i \succeq \bar{\omega}_j, \bar{\omega}_i, \bar{\omega}_j \in \Omega, i, j = 1, \ldots, S, \sum_{j=1}^{S} p_j = 1 \right\}. \tag{5.3}$$

The inner maximization problem is now a linear programming problem over a bounded polyhedron and hence the maximum value is always attained at one of extreme points of the set $\mathcal{P}$, see lemma 2.2. The objective function of the first-stage problem (5.2) is a real-valued, piecewise linear and convex function on $\mathbb{X}$, see Riis and Andersen [47].

Let us now focus on the set of probability distributions consistent with some qualitative information. Let K be the number of extreme points of $\mathcal{P}$ and $\{\mathcal{T}_1, \ldots, \mathcal{T}_K\}$ the set of corresponding admissible supports defined in (4.29).

Due to (4.32) it is sufficient to solve problem (5.2) only at extreme points of $\mathcal{P}$, i.e.

$$\min_{x \in \mathbb{X}} \left\{ c^T x + \max_{k \in \{1, \ldots, K\}} \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} \Phi(x, \bar{\omega}_s) \right\}. \tag{5.4}$$

This is equivalent to the form with auxiliary decision variable θ

$$\begin{aligned}
&\min c^T x + \theta \\
&\text{s.t.} \quad \theta \geq \sum_{s \in \mathcal{T}_k} \tfrac{1}{|\mathcal{T}_k|} \Phi(x, \bar{\omega}_s), \quad k \in \{1, \dots, K\}, \\
&\qquad x \in \mathbb{X}, \\
&\qquad \theta \in \mathbb{R}.
\end{aligned} \tag{5.5}$$

This approach leads directly to the following modification of the L-shaped algorithm.

5.1 L-shaped algorithm – with the assumption of relatively complete recourse

The L-shaped decomposition method works as follows. We solve problem (5.5) relaxing the first group of constraints. During the iteration process we add to this problem further linear constraints as optimality cuts.

Optimality cuts ensue from dual formulation of the second-stage problem

$$\Phi(x, \bar{\omega}_s) := \max \left\{ \pi^T (b_s - T_s x) : \pi^T W_s \leq (q_s)^T \right\}, \tag{5.6}$$

where π is vector of dual decision variables. By a linear programming duality, we have for all $x \in \mathbb{X}$ and $s = 1, \dots, S$, that

$$\Phi(x, \bar{\omega}_s) \geq (\hat{\pi}_s)^T (b_s - T_s x),$$

where $\hat{\pi}_s$ are optimal dual solutions for $s = 1, \dots, S$.

Defining $Q(x, k) := E_{p^k} \Phi(x, \omega)$ (p^k is the k-th extreme point of $\mathcal{P}, k = 1, \dots, K$) we obtain for all $x \in \mathbb{X}$

$$Q(x, k) \geq \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} (\hat{\pi}_s)^T (b_s - T_s x).$$

If the solutions $(\hat{x}, \hat{\theta})$ of problem (5.5) and $\hat{\pi}_s, s = 1, \dots, S$, of the corresponding dual problems (5.6) are such that $\hat{\theta} < Q(\hat{x}, k)$ for some $k = 1, \dots, K$, we obtain the optimality cut for the first-stage decision $\hat{x}$

$$\hat{\theta} \geq \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} (\hat{\pi}_s)^T (b_s - T_s \hat{x}). \tag{5.7}$$

Algorithm 5.1. Multicut L-shaped algorithm for the two-stage minimax problem with qualitative information — with the assumption of relatively complete recourse:

Step 1 (*Initialization*) Let $\nu = 0$ be the number of optimality cuts. Add the constraint $\theta = 0$ to the master problem

$$\begin{aligned} \min\, & c^T x + \theta \\ \text{s.t.} \quad & x \in \mathbb{X}, \\ & \theta \in \mathbb{R}. \end{aligned} \tag{5.8}$$

Step 2 (*Solve master and dual problems*)

1. Solve the master problem (5.11) and store its solution $(\hat{x}, \hat{\theta})$.
2. For all $s \in \mathcal{T}_k, k = 1, \ldots, K$, solve the dual problem (5.6) with $x = \hat{x}$ and store the solution $\hat{\pi}_s$.

Step 3 (*Add optimality cuts*) Set γ = True[1]. For all $k = 1, \ldots, K$, check whether $\hat{\theta} \geq Q(\hat{x}, k)$ holds for $(\hat{x}, \hat{\theta})$:

Yes The current solution is optimal. STOP.

No Add optimality cuts:

1. Set γ = False.
2. If $\nu = 0$ then drop the constraint $\theta = 0$.
3. For those $k \in \{1, \ldots, K\}$ (let their number be $\bar{K}$) for which $\hat{\theta} < Q(\hat{x}, k)$ add optimality cuts (5.7) to the master problem (5.11).
4. Set $\nu = \nu + \bar{K}$.
5. Return to Step 2.

The algorithm terminates in a finite number of iterations whenever a solution to the minimax problem exists (e.g. problem is feasible and bounded). This is a consequence of the fact that the set of dual feasible solutions for scenarios $s \in \{1, \ldots, S\}$ defined by

$$\{\pi \in \mathbb{R}^m : \pi^T W_s \leq (q_s)^T\}$$

is due to assumption (A3) nonempty convex polyhedral set with a finite number of extreme points. Linear objective function of dual problem is then maximized in one of these extreme points. Therefore, only a finite number of optimality cuts can be generated since only a finite number of different optimal dual solutions of problem (5.6) exists. Similar proof can be found

[1]The Boolean variable γ = True indicates that no optimality cut has been added in this pass.

in Riis and Andersen [47], Riis [45]. Assumptions (A1) - (A3) and $\mathbb{X} \neq \emptyset$ ensure the existence of the optimal solution.

For classical L-shaped decomposition in the two-stage stochastic programming problems with known probability distribution we refer e.g. to Kall and Wallace [35].

5.2 L-shaped algorithm – without the assumption of relatively complete recourse

Let us now consider the situation in the two-stage minimax problem without the assumption (A2) of relatively complete recourse, i.e. we allow empty set of feasible second-stage decisions for some $x \in \mathbb{X}$. We have to adapt the algorithm 5.1 by incorporating feasibility cuts to cut off such first-stage decisions which cause an infeasibility of the second-stage problem. We also accept empty set $\mathbb{X}$ of feasible primal solutions. In such a situation the algorithm will terminate with problem infeasibility.

Farkas lemma 2.5 gives us a necessary and sufficient condition for a set of feasible solution to be nonempty. Therefore, if for any first-stage decision $\hat{x} \in \mathbb{X}$ and scenario $s \in \{1, \ldots, S\}$ the set of the second-stage feasible solutions is empty, i.e.

$$\{y \in \mathbb{R}_+^{n_2} : W_s y = b_s - T_s \hat{x}\} = \emptyset$$

then there exist a vector π_s such that

$$\pi_s^T W_s \leq 0 \quad \text{and} \quad \pi_s^T (b_s - T_s \hat{x}) > 0. \tag{5.9}$$

Adding feasibility cut of the form

$$\pi_s^T (b_s - T_s \hat{x}) \leq 0 \tag{5.10}$$

to the first-stage problem we cut off the first-stage decision $\hat{x} \in \mathbb{X}$ which has led to the infeasible dual problem.

Algorithm 5.2. Multicut L-shaped algorithm for the two-stage minimax problem with qualitative information — without the assumption of relatively complete recourse:

Step 1 (*Initialization*) Let $\nu = 0$ be the number of optimality cuts and $n = 0$ be the number of feasibility cuts. Add the constraint $\theta = 0$ to the master problem

$$\begin{aligned} \min\, & c^T x + \theta \\ \text{s.t.}\quad & x \in \mathbb{X}, \\ & \theta \in \mathbb{R}. \end{aligned} \tag{5.11}$$

Step 2 (*Solve master and dual problems*)

1. Solve the master problem (5.11) and store its solution $(\hat{x}, \hat{\theta})$. If infeasible then STOP. The problem has no solution.
2. For all $s \in \mathcal{T}_k, k = 1, \ldots, K$, solve dual problem (5.6) with $x = \hat{x}$ and store its solution $\hat{\pi}_s$. If one of them is unbounded then continue with Step 4. Otherwise go to Step 3.

Step 3 (*Add optimality cuts*) Set $\gamma = \text{True}^2$. For all $k = 1, \ldots, K$, check whether $\hat{\theta} \geq Q(\hat{x}, k)$ holds for $(\hat{x}, \hat{\theta})$:

Yes The current solution is optimal. STOP.

No Add optimality cuts:

1. Set $\gamma = \text{False}$.
2. If $\nu = 0$ then drop the constraint $\theta = 0$.
3. For those $k \in \{1, \ldots, K\}$ (let their number be $\bar{K}$) for which $\hat{\theta} < Q(\hat{x}, k)$ add optimality cuts (5.7) to the master problem (5.11).
4. Set $\nu = \nu + \bar{K}$.
5. Return to Step 2.

Step 4 (*Add feasibility cuts*) For those $s \in \{1, \ldots, S\}$ (let them be $\bar{S}$) for which the dual second-stage problem is unbounded add feasibility cuts (5.10). Set $n = n + \bar{S}$. Return to Step 2.

Theorem 5.1. *If the problem is solvable, then Algorithm 5.2 terminates with an optimal solution in a finite number of iterations.*

Proof. The assumption (A3) ensures that for all $s \in \{1, \ldots, S\}$ the dual problem (5.6) is always feasible. This implies that the dual objective function $\Phi(\hat{x}, \bar{\omega}_s)$ is either finite or equal to $+\infty$ (i.e. the primal second-stage problem is infeasible).

If $\Phi(\hat{x}, \bar{\omega}_s) = +\infty$ then by Farkas lemma 2.5

$$\{\pi_s : \pi_s^T W_s \leq 0, (b_s - T_s\hat{x})^T \pi_s > 0\} \neq \emptyset.$$

By the definition 2.6 we have finitely many generating elements for the convex polyhedral cone $\{\pi_s : \pi_s^T W_s \leq 0\}$ such that, after having used all of them to construct feasibility cuts, for all feasible x, we should have

$$(b_s - T_s x)^T \pi_s \leq 0 \quad \forall \pi_s : \pi_s^T W_s \leq 0$$

[2] The Boolean variable $\gamma = \text{True}$ indicates that no optimality cut has been added in this pass.

and hence solvability of the second-stage problem. Hence, $\Phi(\hat{x}, \bar{\omega}_s) = +\infty$ may appear only finitely many times for a finite number of scenarios. And also only a finite number of optimality cuts can be generated, see discussion below the algorithm 5.1.

It remains to prove that the algorithm terminates with the optimal solution $\hat{x}$. In any iteration i for the optimal solution $(\hat{x}^i, \hat{\theta}^i)$ we have

$$c^T \hat{x}^i + \hat{\theta}^i \leq c^T \hat{x} + \max_{k \in \{1,\dots,K\}} \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} \Phi(\hat{x}, \bar{\omega}_s),$$

This is a consequence of the fact that the master problem is a relaxation of problem (5.5). Optimality cuts cut of those solutions for which

$$\hat{\theta}^i < \max_{k \in \{1,\dots,K\}} \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} \Phi(\hat{x}^i, \bar{\omega}_s).$$

Therefore, after a finite number of iterations (let them be I) we have

$$\hat{\theta}^I = \max_{k \in \{1,\dots,K\}} \sum_{s \in \mathcal{T}_k} \frac{1}{|\mathcal{T}_k|} \Phi(\hat{x}^I, \bar{\omega}_s)$$

and hence $\hat{x} = \hat{x}^I$. □

See also Kall and Wallace [35] for more details about L-shaped decomposition algorithm. Its modification in case of minimax approach is discussed in Riis and Andersen [47], Riis [45].

There are also other techniques for solving minimax stochastic programs (with incomplete information). The one of numerical method for the stochastic minimax problem was proposed by Ermoliev et al. in [23]. They combine a projected quasi sub-gradient approach and generalized linear programming techniques. Their procedure can be applied to general probability distributions and two-stage problems.

Later, Breton and Hachem [5] proposed two algorithms for the solution of multi-periods dynamic minimax problems based on projected sub-gradient and bundle methods. They suppose that the set of possible realizations of random data is discrete and known. The set of possible probability distributions is assumed to satisfy a number of general moment conditions. Under the same assumptions Takriti and Ahmed [60] developed a cut-and-branch procedure for a two-stage stochastic minimax mixed-integer problem with an application to electricity trading.

Stochastic minimax problems have also been considered in approximation techniques based on discretizations of general probability distributions. They can be used to provide bounds for the original problem. See e.g. Freuendorfer and Marohn [24] and references therein.

Applications of the worst-case approach to risk management are described in Rustem and Howe [51].

Chapter 6

Minimax rule in multi-stage programs

Consider a multi-stage linear stochastic programming model having decision variables and constraints divided into groups corresponding to stages $t = 1, \dots, T$. Information structure of such models is crucial.

We adopt the general T-stage stochastic program formulation from Dupačová [22]. It can be defined by a stochastic data process $\omega = (\omega_1, \dots, \omega_{T-1})$ and by a decision process $x = (x_1, \dots, x_T)$. The components $\omega_1, \dots, \omega_{T-1}$ of ω and the decisions $x_2, \dots, x_T$ are assumed to be random vectors, not necessarily of the same dimension, defined on some probability space (Ξ, Σ, P), while x_1 is non-random vector-valued variable. The decision process is *non-anticipative*.[1] The sequence of decisions and observations is

$$x_1, \omega_1, x_2(x_1, \omega_1), \omega_2, \dots, x_T(x_1, \omega_1, \dots, \omega_{T-1}).$$

Let us denote $\omega^{t-1,\bullet} := (\omega_1, \dots, \omega_{t-1})$ the path of the stochastic data process ω which precedes stage t. By $p_t, t = 1, \dots, T-1$, we understand the marginal probability distribution of ω_t and $p_t(\omega^{t-1,\bullet}), t = 2, \dots, T-1$, its conditional probability distribution.

For applications we approximate the true probability distribution p of ω by a discrete probability distribution carried by a finite number of atoms. Hence, the supports of $p_t, t = 1, \dots, T-1$, and $p_t(\omega^{t-1,\bullet}), t = 2, \dots, T-1$, are finite sets.

For disjoint sets of indices $\mathcal{N}_t = \{K_{t-1}+1, \dots, K_t\}, t = 2, \dots, T$, let us list $\bar{\omega}_{n_t}$ all possible realizations of $\omega^{t-1,\bullet}, n_t \in \mathcal{N}_t$. The total number of scenarios

[1] By non-anticipative we understand such conditions that our decisions cannot be influenced by the future realizations of random outcomes. Our decisions depend only on past decisions and past realizations of random variables involved in the model.

is equal to the number of elements of $\mathcal{N}_T$. Each scenario thus generates sequences of coefficients $\{c_{n_t}\}_{t=2}^T$, $\{b_{n_t}\}_{t=2}^T$, $\{T_{n_t}\}_{t=2}^T$, $\{W_{n_t}\}_{t=2}^T$, $\{u_{n_t}\}_{t=2}^T$. For the first stage, vectors and matrices c_1, b_1, u_1 and T_1, W_1 are known.

This information structure of multi-stage problem is represented by a non-recombining scenario tree. Its nodes are determined by all considered realizations $\bar{\omega}_{n_t}, n_t \in \mathcal{N}_t, t = 2, \dots, T$, and by the root indexed as $n_1 = 1 \in \mathcal{N}_1$. Each realization $\bar{\omega}_{n_{t+1}}$ of $\omega^{t,\bullet}, t = 1, \dots, T-1$, has an unique ancestor (a realization of $\omega^{t-1,\bullet}$) denoted by $a(n_{t+1})$ and a finite number of descendants (realizations of $\omega^{t+1,\bullet}$); their set will be denoted by $D(n_{t+1})$. The set of all nodes is represented by $\mathcal{N} = \{1, \dots, K_T\}$.

The path probabilities $p_{n_t} > 0, n_t \in \mathcal{N}_t, \sum\limits_{n_t \in N_T} p_{n_t} = 1, t = 2, \dots, T$, of n_t-th node realizations (i.e. realizations of $\omega^{t-1,\bullet}$) may be obtained by stepwise multiplication of the marginal probabilities $p_{n_2}, n_2 \in \mathcal{N}_2$, by the conditional arc (transition) probabilities, say, $p_{n_{\tau-1},n_\tau}, \tau = 3, \dots, t, n_\tau \in \mathcal{N}_\tau$, i.e.

$$p_{n_t} = p_{n_2} p_{n_2,n_3} \cdots p_{n_{t-1},n_t}.$$

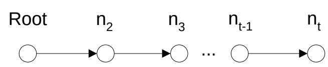

Figure 6.1: Scenario path.

Example 6.1. Consider the scenario tree in the figure 6.2. The number of stages is $T = 4$ with $K_1 = 1$, $K_2 = 4$, $K_3 = 11$ and $K_4 = 24$.

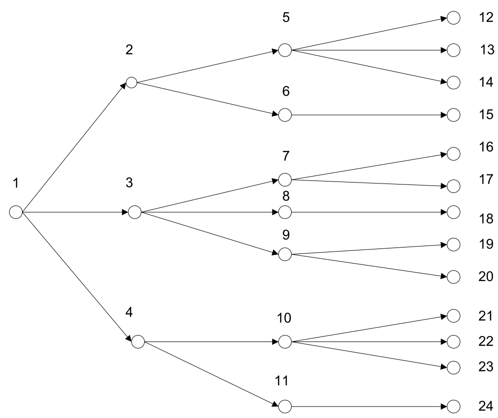

Figure 6.2: Scenario tree.

We introduce the following *arborescent form* of the T-stage stochastic program:

$$\min_{\substack{x_{n_t},\, n_t \in \mathcal{N}_t \\ t=1,\dots,T}} \left\{ c_1^T x_1 + \sum_{n_2=2}^{K_2} p_{n_2} c_{n_2}^T x_{n_2} + \cdots + \sum_{n_T=K_{T-1}+1}^{K_T} p_{n_T} c_{n_T}^T x_{n_T} \right\}$$

$$\begin{array}{llllllll}
\text{s.t.} & W_1 x_1 & & & & & = & b_1, \\
& T_n x_1 & + & W_{n_2} x_{n_2} & & & = & b_{n_2}, \quad n_2 \in \mathcal{N}_2, \\
& & & T_{n_3} x_{a(n_3)} & + & W_{n_3} x_{n_3} & = & b_{n_3}, \quad n_3 \in \mathcal{N}_3, \\
& & & & \ddots & & & \\
& & & & & T_{n_T} x_{a(n_T)} \; + \; W_{n_T} x_{n_T} & = & b_{n_T}, \quad n_T \in \mathcal{N}_T,
\end{array}$$

$$0 \le x_{n_t} \le u_{n_t}, n_t \in \mathcal{N}_t, t = 1, \dots, T, \tag{6.1}$$

in which the non-anticipativity constraints are included implicitly.

Let us assume:

(B1) There exists $0 \le x_1 \le u_1$ such that $W_1 x_1 = b_1$ and for all $x_{a(n_t)}$ and $n_t \in \mathcal{N}_t, t = 2, \dots, T$, there exists $0 \le x_{n_t} \le u_{n_t}$ such that $W_{n_t} x_{n_t} = b_{n_t} - T_{n_t} x_{a(n_t)}$. The ancestors $a(n_t)$ and the corresponding decision variables $x_{a(n_t)}$ come from constraints for the corresponding indices n_{t-1} and variables $x_{a(n_t)} \sim x_{n_{t-1}}$.

Due to this assumption of relatively complete recourse we guarantee the problem feasibility, i.e. the set of feasible solutions is nonempty. The problem boundedness is ensured by finite (real) upper bounds u_{n_t} (of x_{n_t}), $n_t \in \mathcal{N}_t, t = 2, \dots, T$. Therefore, the optimal solution always exists and hence the problem (6.1) is also dual feasible.

6.1 Probability distributions consistent with a qualitative information

The data process ω is modeled as a Markov chain with the corresponding state space $\mathcal{N}$, marginal probabilities $p_{n_2}, n_2 \in \mathcal{N}_2$ and transition probabilities p_{n_{t-1},n_t}, $n_{t-1} \in \mathcal{N}_{t-1}, n_t \in \mathcal{N}_t, t = 2, \dots, T$, of moving from state n_{t-1} to state n_t. The Markov property ensures that the conditional probability distribution of future states of the process, given the present state and all past states, depends only upon the present state and not on any past states, i.e. it is conditionally independent of the past states given the present state.

Any scenario tree and the corresponding transition probabilities can be unambiguously described by a transition matrix with p_{n_{t-1},n_t}, $n_{t-1}, n_t \in \mathcal{N}\backslash\{1\}$, representing the transition probabilities between nodes n_{t-1} and n_t and p_{n_2}, $n_2 \in \mathcal{N}_2$, standing for marginal distributions.

We may write the transition matrix as follows:

$$\boldsymbol{M} = \begin{pmatrix} (p^1)^T & 0 & 0 & \cdots & 0 \\ \mathbf{0} & \boldsymbol{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{M}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{M}_{T-2} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{I} \end{pmatrix},$$

where

$$\boldsymbol{M_t} = \begin{pmatrix} p_{K_t+1,K_{t+1}+1} & \cdots & p_{K_t+1,K_{t+2}} \\ \vdots & & \vdots \\ p_{K_{t+1},K_{t+1}+1} & \cdots & p_{K_{t+1},K_{t+2}} \end{pmatrix}.$$

and $(p^1)^T = (p_2, \ldots, p_{K_2})^T$.

We also assume that there exists only one nonzero element in each column of $\boldsymbol{M_t}, t = 1, \ldots, T-2$, to guarantee a non-recombining scenario tree.

We will simplify the notation of tree nodes in situations where it is not important to distinguish particular stages of the tree; we will then write just $n \in \mathcal{N}$ instead of $n_t \in \mathcal{N}$.

Example 6.2. The scenario tree in the figure 6.2 can be described by a transition matrix with the following nonzero blocks $(p^1)^T$, $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$:

	2	3	4
1	p_2	p_3	p_4

	5	6	7	8	9	10	11
2	$p_{2,5}$	$p_{2,6}$	0	0	0	0	0
3	0	0	$p_{3,7}$	$p_{3,8}$	$p_{3,9}$	0	0
4	0	0	0	0	0	$p_{4,10}$	$p_{4,11}$

	12	13	14	15	16	17	18	19	20	21	22	23	24
5	$p_{5,12}$	$p_{5,13}$	$p_{5,14}$	0	0	0	0	0	0	0	0	0	0
6	0	0	0	$p_{6,15}$	0	0	0	0	0	0	0	0	0
7	0	0	0	0	$p_{7,16}$	$p_{7,17}$	0	0	0	0	0	0	0
8	0	0	0	0	0	0	$p_{8,18}$	0	0	0	0	0	0
9	0	0	0	0	0	0	0	$p_{9,19}$	$p_{9,20}$	0	0	0	0
10	0	0	0	0	0	0	0	0	0	$p_{10,21}$	$p_{10,22}$	$p_{10,23}$	0
11	0	0	0	0	0	0	0	0	0	0	0	0	$p_{11,24}$

Due to the form of the transition matrix $\boldsymbol{M}$ the objective of problem (6.1) can be reformulated as follows

$$\min_{x_n, n\in\mathcal{N}} \left\{ c_1^T x_1 + (p^1)^T \begin{pmatrix} c_2^T x_2 \\ \vdots \\ c_{K_2}^T x_{K_2} \end{pmatrix} + (p^1)^T \boldsymbol{M_1} \begin{pmatrix} c_{K_2+1}^T x_{K_2+1} \\ \vdots \\ c_{K_3}^T x_{K_3} \end{pmatrix} + \ldots + \right.$$
$$\left. + \ldots + (p^1)^T \boldsymbol{M_1} \cdots \boldsymbol{M_{T-2}} \begin{pmatrix} c_{K_{T-1}+1}^T x_{K_{T-1}+1} \\ \vdots \\ c_{K_T}^T x_{K_T} \end{pmatrix} \right\}. \quad (6.2)$$

Let $l_n := |D(n)|$ denote the number of immediate descendants of node n, $n \in \mathcal{N}\backslash\mathcal{N}_T$. Assume now that our information on probability distribution of process ω is incomplete. The only available information is described by a tree structure (e.g. as in the figure 6.1) and by the following sets of feasible distributions:

P_1 represents the set of possible marginal distributions $p^1 = \{p_{n_2}\}_{n_2 \in \mathcal{N}_2} \in \mathbb{R}_+^{l_1}$ with finite support Ω_1 (Ω_1 is finite set of all possible realizations of ω_1, i.e. $\Omega_1 = \{\bar{\omega}_{n_2} : n_2 \in \mathcal{N}_2\} = \{\bar{\omega}_2, \ldots, \bar{\omega}_{K_2}\}$).

P_n, $n = 2, \ldots, K_{T-1}$, stand for the sets of feasible conditional probability distributions $p^n = \{p_{n,m}\}_{m\in D(n)} \in \mathbb{R}_+^{l_n}$, with finite supports $\Omega_n := \{\bar{\omega}_m : m \in D(n)\}$, $n = 2, \ldots, K_{T-1}$. Note that $\sum_{m\in D(n)} p_{n,m} = 1$ for all $n = 2, \ldots, K_{T-1}$.

We further suppose:

(B2) All sets $P_k, k = 1, \ldots .K_{T-1}$, are consistent with some qualitative information $\succeq$, i.e. the sets can be described in the following way

$$P_1 = \Big\{ p^1 = \{p_n\}_{n\in\mathcal{N}_2} \in \mathbb{R}_+^{l_1} : p_{n_1} \geq p_{n_2} \text{ if } \bar{\omega}_{n_1} \succeq \bar{\omega}_{n_2}, \bar{\omega}_{n_1}, \bar{\omega}_{n_2} \in \Omega_1,$$
$$n_1, n_2 = 1, \ldots, l_1, \sum_{n=1}^{l_1} p_n = 1 \Big\},$$

and for $k = 2, \ldots, K_{T-1}$,

$$P_n = \Big\{ p^n = \{p_{n,m}\}_{m\in D(n)} \in \mathbb{R}_+^{l_n} : p_{n,m_1} \geq p_{n,m_2} \text{ if } \bar{\omega}_{m_1} \succeq \bar{\omega}_{m_2},$$
$$\bar{\omega}_{m_1}, \bar{\omega}_{m_2} \in \Omega_n, m_1, m_2 = 1, \ldots, l_n, \sum_{m=1}^{l_n} p_{n,m} = 1 \Big\}.$$

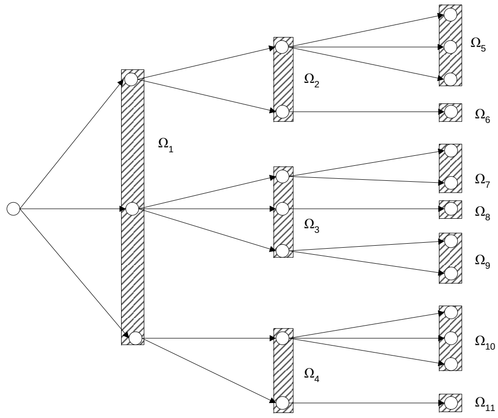

Figure 6.3: Admissible supports.

Example 6.3. Consider the scenario tree in the figure 6.2. Corresponding supports are illustrated in the figure 6.3. Then we have $\Omega_1 = \{\bar{\omega}_2, \bar{\omega}_3, \bar{\omega}_4\}$, $\Omega_2 = \{\bar{\omega}_5, \bar{\omega}_6\}$, $\Omega_3 = \{\bar{\omega}_7, \bar{\omega}_8, \bar{\omega}_9\}$, $\Omega_4 = \{\bar{\omega}_{10}, \bar{\omega}_{11}\}$, $\Omega_5 = \{\bar{\omega}_{12}, \bar{\omega}_{13}, \bar{\omega}_{14}\}$, $\Omega_6 = \{\bar{\omega}_{15}\}$, $\Omega_7 = \{\bar{\omega}_{16}, \bar{\omega}_{17}\}$, $\Omega_8 = \{\bar{\omega}_{18}\}$, $\Omega_9 = \{\bar{\omega}_{19}\bar{\omega}_{20}\}$, $\Omega_{10} = \{\bar{\omega}_{21}, \bar{\omega}_{22}, \bar{\omega}_{23}\}$, $\Omega_{11} = \{\bar{\omega}_{24}\}$.

Suppose that the available qualitative information is of the form $\bar{\omega}_1 \succeq \bar{\omega}_2$, $\bar{\omega}_1 \succeq \bar{\omega}_3$, $\bar{\omega}_5 \succeq \bar{\omega}_6$, $\bar{\omega}_7 \succeq \bar{\omega}_8$, $\bar{\omega}_8 \succeq \bar{\omega}_9$, $\bar{\omega}_{11} \succeq \bar{\omega}_{10}$, $\bar{\omega}_{12} \succeq \bar{\omega}_{14}$, $\bar{\omega}_{20} \succeq \bar{\omega}_{19}$, $\bar{\omega}_{23} \succeq \bar{\omega}_{21}$.

We obtain sets of probability distributions consistent with the above mentioned qualitative information:

$$
\begin{array}{rcl}
P_1 & = & \left\{p^1 \in \mathbb{R}^3_+ : p_1 \geq p_2, p_1 \geq p_3, p_1 + p_2 + p_3 = 1\right\}, \\
P_2 & = & \left\{p^2 \in \mathbb{R}^2_+ : p_{2,5} \geq p_{2,6}, p_{2,5} + p_{2,6} = 1\right\}, \\
P_3 & = & \left\{p^3 \in \mathbb{R}^3_+ : p_{3,7} \geq p_{3,8} \geq p_{3,9}, p_{3,7} + p_{3,8} + p_{3,9} = 1\right\}, \\
P_4 & = & \left\{p^4 \in \mathbb{R}^2_+ : p_{4,11} \geq p_{4,10}, p_{4,10} + p_{4,11} = 1\right\}, \\
P_5 & = & \left\{p^5 \in \mathbb{R}^3_+ : p_{5,12} \geq p_{5,14}, p_{5,12} + p_{5,13} + p_{5,14} = 1\right\}, \\
P_6 & = & P_8 = P_{11} = \{1\}, \\
P_7 & = & \left\{p^7 \in \mathbb{R}^2_+ : p_{7,16} + p_{7,17} = 1\right\}, \\
P_9 & = & \left\{p^9 \in \mathbb{R}^2_+ : p_{9,20} \geq p_{9,19}, p_{9,19} + p_{9,20} = 1\right\}, \\
P_{10} & = & \left\{p^{10} \in \mathbb{R}^3_+ : p_{10,23} \geq p_{10,21}, p_{10,21} + p_{10,22} + p_{10,23} = 1\right\}.
\end{array}
$$

The sets $P_n, n = 1, \ldots, K_{T-1}$, are bounded polyhedrons with generalized uniform distributions as extreme points, see section 4.2.

6.2 Minimax formulation

Let us now focus on a reformulation of problem (6.1) according to the incomplete information on underlying probabilities as defined in the previous section. We apply the minimax decision rule, i.e. we consider the worst-case probability distribution. The minimax objective function of T-stage stochastic program takes the form

$$\min_{x_n, n\in\mathcal{N}} \max_{\substack{p^1 \in P_1 \\ p^n \in P_n, \\ n=2,\dots,K_{T-1}}} \left\{ c_1^T x_1 + (p^1)^T \left[\begin{pmatrix} c_2^T x_2 \\ \vdots \\ c_{K_2}^T x_{K_2} \end{pmatrix} + \boldsymbol{M_1} \left[\begin{pmatrix} c_{K_2+1}^T x_{K_2+1} \\ \vdots \\ c_{K_3}^T x_{K_3} \end{pmatrix} + \right.\right.\right.$$
$$\left.\left.\left. + \dots + \boldsymbol{M_{T-2}} \begin{pmatrix} c_{K_{T-1}+1}^T x_{K_{T-1}+1} \\ \vdots \\ c_{K_T}^T x_{K_T} \end{pmatrix} \right]\right]\right\}, \tag{6.3}$$

where the non-zero elements of particular rows of matrices $\boldsymbol{M}_t, t = 1, \dots, T-2$, consist of vectors p^n.

The inner optimization problem of (6.3) has an optimal solution, since the objective function is continuous with respect to $p^1, p^n, n = 2\dots, K_{T-1}$, and the sets of probability distributions consistent with a qualitative information are compact and nonempty.

Since the transition matrix $\boldsymbol{M_t}$ enters only the last $T-t-1$ terms of the objective function (6.3), and since the products $(p^1)^T \boldsymbol{M_1} \cdots \boldsymbol{M_{t-1}}$ are nonnegative, the worst transition matrix $\boldsymbol{M_t^*}$ (i.e. the worst vectors $p^{n*}, n = K_t + 1, \dots, K_{t+1}$) does not depend on $(p^1)^T \boldsymbol{M_1} \cdots \boldsymbol{M_{t-1}}$ and we may rewrite the objective (6.3) in the following way

$$\min_{x_n, n\in\mathcal{N}} \left\{ c_1^T x_1 + \max_{p^1\in P_1} (p^1)^T \left[\begin{pmatrix} c_2^T x_2 \\ \vdots \\ c_{K_2}^T x_{K_2} \end{pmatrix} + \max_{\substack{p^n \in P_n, \\ n=2,\dots,K_2}} \boldsymbol{M_1} \left[\begin{pmatrix} c_{K_2+1}^T x_{K_2+1} \\ \vdots \\ c_{K_3}^T x_{K_3} \end{pmatrix} + \right.\right.\right.$$
$$\left.\left.\left. + \dots + \max_{\substack{p^n \in P_n, \\ n=K_{T-2}+1,\dots,K_{T-1}}} \boldsymbol{M_{T-2}} \begin{pmatrix} c_{K_{T-1}+1}^T x_{K_{T-1}+1} \\ \vdots \\ c_{K_T}^T x_{K_T} \end{pmatrix} \right]\right]\right\}. \tag{6.4}$$

If we return to the formulation based on transition probabilities we obtain the following minimax problem

$$\min_{\substack{x_{n_t}, n_t \in \mathcal{N}_t \\ t=1,\dots,T}} \left\{ c_1^T x_1 + \max_{\substack{p^1 \in P_1 \\ p^1 = \{p_{n_2}\}_{n_2 \in \mathcal{N}_2}}} \sum_{n_2=2}^{K_2} p_{n_2} \left[c_{n_2}^T x_{n_2} + \right.\right.$$

$$\max_{\substack{p^{n_2} \in P_{n_2} \\ p^{n_2} = \{p_{n_2,n_3}\}_{n_3 \in D(n_2)}}} \sum_{n_3=2}^{K_2} \left[p_{n_2,n_3} c_{n_3}^T x_{n_3} + \cdots + \right.$$

$$\left.\left.\left. + \max_{\substack{p^{n_{T-1}} \in P_{n_{T-1}} \\ p^{n_{T-1}} = \{p_{n_{T-1},n_T}\}_{n_T \in D(n_{T-1})}}} \sum_{n_T=K_{T-1}+1}^{K_T} p_{n_2} p_{n_2,n_3} \cdots p_{n_{T-1},n_T} c_{n_T}^T x_{n_T} \right]\right]\right\}$$

$$\begin{array}{lllllll} \text{s.t.} & W_1 x_1 & & & & = & b_1, \\ & T_{n_2} x_1 & + & W_{n_2} x_{n_2} & & = & b_{n_2}, \quad n_2 \in \mathcal{N}_2, \\ & & & T_{n_3} x_{a(n_3)} & + \quad W_{n_3} x_{n_3} & = & b_{n_3}, \quad n_3 \in \mathcal{N}_3, \\ & & & & \ddots & & \\ & & & & T_{n_T} x_{a(n_T)} + W_{n_T} x_{n_T} & = & b_{n_T}, \quad n_T \in \mathcal{N}_T, \end{array}$$

$$0 \le x_{n_t} \le u_{n_t}, n_t \in \mathcal{N}_t, t = 1, \dots, T. \tag{6.5}$$

Henceforward, we will follow the derivation of nested decomposition algorithm for the multi-stage stochastic problem with a linear recourse presented in Kall and Mayer [34] and modify it for the multi-stage minimax problems with a qualitative information.

The properties mentioned above allow as to formulate the multi-stage minimax stochastic program (6.3) in the nested form based on the following sequence of programs:

- for node $n = 1$

$$F_1 = \min_{x_1} \left[c_1^T x_1 + \max_{\substack{p^1 \in P_1 \\ p^1 = \{p_n\}_{n \in \mathcal{N}_2}}} \sum_{n=2}^{K_2} p_n F_n(x_1) \right]$$

$$\begin{array}{lrcl} \text{s.t.} & W_1 x_1 & = & b_1, \\ & 0 \le x_1 & \le & u_1; \end{array} \tag{6.6}$$

- generally for any node in stage $t = 2, \dots, T-1$, i.e. $n = K_{t-1} + 1, \dots, K_t$,

$$F_n(x_{a(n)}) = \min_{x_n} \left[c_n^T x_n + \max_{\substack{p^n \in P_n \\ p^n = \{p_{n,m}\}_{m \in D(n)}}} \sum_{m \in D(n)} p_{n,m} F_m(x_n) \right]$$

$$\begin{aligned} \text{s.t.} \quad W_n x_n &= b_n - T_n x_{a(n)}, \\ 0 \le x_n &\le u_n; \end{aligned} \tag{6.7}$$

- for nodes in stage T, i.e. $n = K_{T-1} + 1, \dots, K_T$,

$$F_n(x_{a(n)}) = \min_{x_n} c_n^T x_n$$

$$\begin{aligned} \text{s.t.} \quad W_n x_n &= b_n - T_n x_{a(n)}, \\ 0 \le x_n &\le u_n. \end{aligned} \tag{6.8}$$

Note that the functions $F_n(x_{a(n)})$ are piecewise linear and convex in x_n and the additive terms $\sum_{m \in D(n)} p_{n,m} F_m(x_n)$ are linear in $p_{n,m}$. Hence, we maximize the linear function over a bounded polyhedron P_n. The maximum will be attained at one of extreme points of the set P_n.

Maximization over set P_n can be replaced by a new condition applied to all $p^n \in P_n$. We can rewrite problem (6.7) to the form (denoted by $\text{Dest}(n, x_{a(n)})$):

$$F_n(x_{a(n)}) = \min_{x_n} c_n^T x_n + \theta_n$$

$$\begin{aligned} \text{s.t.} \quad \theta_n &\ge \sum_{m \in D(n)} p_{n,m} F_m(x_n), \quad \forall p^n = \{p_{n,m}\}_{m \in D(n)} \in P_n, \\ W_n x_n &= b_n - T_n x_{a(n)}, \\ 0 \le x_n &\le u_n, \quad n \in \mathcal{N}_t, \quad t \in \{2, \dots, T-1\}. \end{aligned} \tag{6.9}$$

The first group of constraints is satisfied whenever it holds true for all extreme probability distributions of P_n indexed by $k = 1, \dots, e_{P_n}$, that are fully characterized by admissible supports $\mathcal{T}_k^{P_n}, k = 1, \dots, e_{P_n}$. Hence, the first group of constraints can be replaced by

$$\theta_n \ge \sum_{m \in \mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|} F_m(x_n), \quad \forall k = 1, \dots, e_{P_n}, n \in \mathcal{N}_t, t \in \{2, \dots, T-1\}. \tag{6.10}$$

The main idea of the forthcoming algorithm is to solve (6.9) relaxing the first group of constraints. During the iteration process we add to this problem further linear constraints as optimality cuts.

We will solve the following sequence of successively built relaxed master problems denoted by Mast$(n, x_{a(n)})$:

$$\tilde{F}_n(x_{a(n)}) = \min_{x_n} c_n^T x_n + \theta_n$$

$$\begin{array}{rrcl} \text{s.t.} & W_n x_n & = & b_n - T_n x_{a(n)}, \\ & d_{nl}^T x_n + \theta_n & \geq & \delta_{nl}, \quad l = 1, \ldots, s_n, \\ & 0 \leq x_n & \leq & u_n, \quad n \in \mathcal{N}_t, \quad t \in \{2, \ldots, T-1\}, \end{array} \tag{6.11}$$

with the parameter vector $x_{a(n)}$ and the optimal-value function $\tilde{F}_n$. For the root node we define $\tilde{F}_1(x_{a(1)}) \equiv \tilde{F}_1$. Constraints in the second group of constraints are called optimality cuts. Their number is denoted by s_n.

Due to (B1), Mast$(n, x_{a(n)})$ has an optimal solution. It is also dual feasible for all $n \in \mathcal{N}$, for all $x_{a(n)}$.

6.3 Construction of optimality cuts

Let $(\hat{x}_m, \hat{\theta}_m)$ be an optimal solution of Mast$(m, \hat{x}_n)$, $m \in D(n)$. Take into consideration the dual formulation of problem Mast$(m, \hat{x}_n)$:

$$\begin{array}{rl} \max & (b_m - T_m \hat{x}_n)^T v_m + \sum_{l=1}^{s_m} \delta_{ml} w_{ml} + u_m^T \lambda_m \\ \text{s.t.} & W_m^T v_m + \sum_{l=1}^{s_m} d_{ml} w_{ml} + 1^T \lambda_m \leq c_m, \\ & \sum_{l=1}^{s_m} w_{ml} = 1, \\ & w_{ml} \geq 0, \quad l = 1, \ldots, s_m, \\ & \lambda_j \leq 0, \quad j = 1, \ldots, k_m, \end{array}$$

where k_m is the dimension of vector x_m.

If $(\hat{v}_m, \hat{w}_m, \hat{\lambda}_m)$ are the corresponding dual solutions then we have

$$\tilde{F}_m(\hat{x}_n) = c_m^T \hat{x}_m + \hat{\theta}_m = (b_m - T_m \hat{x}_n)^T \hat{v}_m + \sum_{l=1}^{s_m} \delta_{ml} \hat{w}_{ml} + u_m^T \hat{\lambda}_m, \tag{6.12}$$

for all $m \in D(n)$.

Weak duality ensures that for any feasible x_n

$$\tilde{F}_m(x_n) \geq (b_m - T_m x_n)^T \hat{v}_m + \sum_{l=1}^{s_m} \delta_{ml} \hat{w}_{ml} + u_m^T \hat{\lambda}_m. \tag{6.13}$$

Hence, any current solution $(\hat{x}_n, \hat{\theta}_n)$ of $\text{Mast}(n, x_{a(n)})$ has to fulfil the following conditions for all $k \in \{1, \ldots, e_{P_n}\}$

$$\hat{\theta}_n \geq \sum_{m \in \mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|} \left[(b_m - T_m \hat{x}_n)^T \hat{v}_m + \sum_{l=1}^{s_m} \delta_{ml} \hat{w}_{ml} + u_m^T \hat{\lambda}_m \right]. \tag{6.14}$$

For those k (let their number be $\bar{k}_{e_{P_n}}$) for which the condition (6.14) does not hold we add the following optimality cuts to the master problem $\text{Mast}(n, x_{a(n)})$:

$$d_{n\nu}^T x_n + \theta_n \geq \delta_{n\nu}, \quad \nu = s_n + 1, \ldots, s_n + \bar{k}_{e_{P_n}}, \tag{6.15}$$

where

$$\begin{aligned} d_{n\nu} &:= \sum_{m \in \mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|} T_m^T \hat{v}_m, \\ \delta_{n\nu} &:= \sum_{m \in \mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|} \left[b_m^T \hat{v}_m + \sum_{l=1}^{s_m} \delta_{ml} \hat{w}_{ml} + u_m^T \hat{\lambda}_m \right]. \end{aligned} \tag{6.16}$$

Definition 6.1. An optimality cut $d_{nl}^T x_n + \theta_n \geq \delta_{nl}$ in $\text{Mast}(n, x_{a(n)})$ will be called valid, if the inequality

$$\delta_{nl} - d_{nl}^T x_n \leq \max_{\substack{p^n \in P_n \\ p^n = \{p_{n,m}\}_{m \in D(n)}}} \sum_{m \in D(n)} p_{n,m} F_m(x_n)$$

holds for any feasible x_n.

Validity of optimality cuts ensures that the objective function of the relaxed master problem provides a lower bound to the objective function of the descendant recourse problem, i.e. $\tilde{F}_n(x_{a(n)}) \leq F_n(x_{a(n)})$ holds for any feasible x_n. Optimality cuts generated by the algorithm presented below are valid, see theorem 6.1 below.

6.4 Nested decomposition – with the assumption of relatively complete recourse

Algorithm 6.1. Multicut nested decomposition for the multi-stage minimax problem with a qualitative information — with the assumption of relatively complete recourse:

Step 1 (*Initialization*) Let $s_n = 0$, $\gamma_n =$ False[2] and add the constraint $\theta_n = 0$ to $\text{Mast}(n, x_{a(n)})$ for all $n \in \mathcal{N}$. Set $t = 1$.

Step 2 (*Forward pass*) For $n \in \mathcal{N}_t : \gamma_n =$ False do:

1. Solve $\text{Mast}(n, \hat{x}_{a(n)})$.
2. Store the solution $(\hat{x}_n, \hat{\theta}_n)$. If $t = T$ then store also the dual solution $(\hat{v}_n, \hat{w}_n, \hat{\lambda}_n)$.
3. Set $\gamma_n =$ True and $\gamma_\nu =$ False, $\forall \nu \in \mathcal{G}(n) \setminus \{n\}$[3].
4. Take the next node in $\mathcal{N}_t$.

If $t = T$ then continue with Step 3, else set $t = t + 1$ and repeat Step 2.

Step 3 (*Backward pass*) Set $\gamma =$ True[4]. For $n \in \mathcal{N}_{t-1}$ check whether (6.14) holds for $(\hat{x}_n, \hat{\theta}_n)$:

Yes Take the next node in $\mathcal{N}_{t-1}$.

No Add optimality cuts:

1. Set $\gamma =$ False.
2. If $s_n = 0$ then drop the constraint $\theta_n = 0$.
3. Add optimality cuts (6.15) to $\text{Mast}(n, \hat{x}_{a(n)})$ (let them be $\bar{e}_{p_n}$) with $\nu = s_n + 1, \ldots, s_n + \bar{e}_{P_n}$. Set $s_n = s_n + \bar{e}_{P_n}$.
4. Solve $\text{Mast}(n, \hat{x}_{a(n)})$ and temporarily store the dual solution $(\hat{v}_n, \hat{w}_n, \hat{\lambda}_n)$. Set $\gamma_\nu =$ False, $\nu \in \mathcal{G}(n) \setminus \{n\}$.

If $t = 1$ and $\gamma =$ True then no optimality cut has been added and the current solution is optimal. STOP.
If $t > 1$ then set $t = t - 1$ and repeat Step 3.
If $t = 1$ and $\gamma =$ False then return to Step 2.

Theorem 6.1. *The following statements hold:*

(i) The optimality cuts generated by the algorithm are valid. Moreover, $\tilde{F}_n(x_{a(n)}) \leq F_n(x_{a(n)})$ *for all* $n \in \mathcal{N}$ *and all* $x_{a(n)}$.

---

[2]The Boolean variable $\gamma_n =$ True indicates that the current master problem $\text{Mast}(n, \hat{x}_{a(n)})$ has a solution. The current solution $(\hat{x}_n, \hat{\theta}_n)$ can be used whenever node n is encountered during the subsequent iterations. The value $\gamma_n =$ False indicates that $\text{Mast}(n, \hat{x}_{a(n)})$ has to be solved whenever node n is encountered.

[3]$\mathcal{G}(n)$ represents the set of all nodes corresponding to a subtree rooted at the node n.

[4]The Boolean variable $\gamma =$ True indicates that no optimality cut has been added in this backward pass.

(ii) *The algorithm terminates in a finite number of iterations finding an optimal solution* $(\hat{x}_n, n \in \mathcal{N})$ *and extremal probability distributions* $\hat{p}^1 \in P_1, \hat{p}^n \in P_n, n \in \mathcal{N} \setminus \{1\}$.

Proof. (i) Let $(\hat{x}_n, \hat{\theta}_n)$ be the current solution of $\mathrm{Mast}(n, \hat{x}_{a(n)})$. Take into consideration inequalities (6.13) and (6.15). We may derive

$$p_{n,m}\left[(b_m - T_m\hat{x}_n)^T\hat{v}_m + \sum_{l=1}^{s_m}\delta_{ml}\hat{w}_{ml} + u_m^T\hat{\lambda}_m\right] \leq p_{n,m}\tilde{F}_m(\hat{x}_n)$$

and hence, for all feasible x_n,

$$\delta_{nl} - d_{nl}^T x_n \leq \sum_{m\in D(n)} p_{n,m}\tilde{F}_m(x_n) \leq \max_{p^n \in P_n} \sum_{m\in D(n)} p_{n,m}\tilde{F}_m(x_n). \quad (6.17)$$

For $n \in \mathcal{N}_T$ the problems $\mathrm{Mast}(n, x_{a(n)})$ and $\mathrm{Desc}(n, x_{a(n)})$ are identical and we have $\tilde{F}_n(x_{a(n)}) = F_n(x_{a(n)})$ for all feasible $x_{a(n)}$.

(1) Assume $n \in \mathcal{N}_{T-1}$. We may write

$$\begin{aligned} c_n^T x_n + \theta_n &\leq c_n^T x_n + \max_{l\in\{1,\dots,s_n\}}\left(\delta_{nl} - d_{nl}^T x_n\right) \\ &\leq c_n^T x_n + \max_{p^n \in P_n}\sum_{m\in D(n)} p_{n,m}\tilde{F}_m(x_n) \\ &= c_n^T x_n + \max_{p^n \in P_n}\sum_{m\in D(n)} p_{n,m}F_m(x_n). \end{aligned} \quad (6.18)$$

The last equality comes from the identity of $\mathrm{Mast}(m, x_{a(m)})$ and $\mathrm{Desc}(m, x_{a(m)})$ for $m \in \mathcal{N}_T$. Thus taking minima on both sides of (6.18) over the feasible domain we obtain $\tilde{F}_n(x_{a(n)}) \leq F_n(x_{a(n)})$ for $n \in \mathcal{N}_{T-1}$. Using formula (6.17) we have the validity of cuts

$$\delta_{nl} - d_{nl}^T x_n \leq \max_{p^n \in P_n}\sum_{m\in D(n)} p_{n,m}F_m(x_n), n \in \mathcal{N}_{T-1}.$$

(2) Let us now assume $n \in \mathcal{N}_{T-2}$. From previous results we may derive

$$\begin{aligned} c_n^T x_n + \theta_n &\leq c_n^T x_n + \max_{l\in\{1,\dots,s_n\}}\left(\delta_{nl} - d_{nl}^T x_n\right) \\ &\leq c_n^T x_n + \max_{p^n \in P_n}\sum_{m\in D(n)} p_{n,m}\tilde{F}_m(x_n) \\ &\leq c_n^T x_n + \max_{p^n \in P_n}\sum_{m\in D(n)} p_{n,m}F_m(x_n). \end{aligned} \quad (6.19)$$

The last inequality is the consequence of (1), i.e. $\tilde{F}_m(x_{a(m)}) \leq F_m(x_{a(m)})$ for $m \in \mathcal{N}_{T-1}$. Thus taking again minima on both sides of (6.19) we conclude $\tilde{F}_n(x_{a(n)}) \leq F_n(x_{a(n)})$ for $n \in \mathcal{N}_{T-2}$. Using formula (6.17) we have

$$\delta_{nl} - d_{nl}^T x_n \leq \max_{p^n \in P_n} \sum_{m \in D(n)} p_{n,m} F_m(x_n), n \in \mathcal{N}_{T-2}.$$

(3) By backward induction we derive these results for all $n \in \{1, \ldots, K_{T-1}\}$.

(ii) The algorithm terminates in a finite number of iterations due to the fact that for any node $n \in \{1, \ldots, K_{T-1}\}$ there exist finitely many different dual solutions of relaxed master problems Mast$(m, \hat{x}_n)$ associated with the child-node. Its dual feasible region does not depend on $\hat{x}_n$ and therefore only a finite number of different optimality cuts can be generated, see discussion in chapter 5 and the proof of theorem 5.1; we also refer to Kall and Mayer [34].

Assumptions (B1) and (B2) ensure the existence of feasible solutions of the multi-stage minimax problems consistent with the qualitative information $\succeq$. Hence, the optimal solutions always exist.

Let $(\hat{x}_n, \hat{\theta}_n)$ be the optimal solution of Mast$(n, \hat{x}_{a(n)})$ in which the algorithm terminates. Considering (6.14) we obtain

$$\hat{\theta}_n \geq \max_{k=\{1,\ldots,e_{P_n}\}} \sum_{m \in \mathcal{I}_k^{p_n}} \frac{1}{|\mathcal{I}_k^{P_n}|} \tilde{F}_m(\hat{x}_n) = \max_{p^n \in P_n} \sum_{m \in D(n)} p_{n,m} \tilde{F}_m(\hat{x}_n). \tag{6.20}$$

Applying (6.20) to an arbitrary node $n \in \mathcal{N}$ we get

$$\begin{aligned}
\tilde{F}_n(x_{a(n)}) &= c_n^T \hat{x}_n + \hat{\theta}_n \\
&\geq c_n^T \hat{x}_n + \max_{p^n \in P_n} \sum_{m \in D(n)} p_{n,m}(c_m^T \hat{x}_m + \hat{\theta}_m) \\
&\geq c_n^T \hat{x}_n + \max_{p^n \in P_n} \sum_{m \in D(n)} p_{n,m} \Bigg[c_m^T \hat{x}_m + \\
&\qquad + \max_{p^m \in P_m} \sum_{\mu \in D(m)} p_{m,\mu}(c_\mu^T \hat{x}_\mu + \hat{\theta}_\mu) \Bigg] \\
&= c_n^T \hat{x}_n + \max_{\substack{p^n \in P_n \\ p^m \in P_m, m \in D(n)}} \Bigg[\sum_{m \in D(n)} p_{n,m} c_m^T \hat{x}_m + \\
&\qquad + \sum_{m \in D(n)} \sum_{\mu \in D(m)} p_{n,m} p_{m,\mu}(c_\mu^T \hat{x}_\mu + \hat{\theta}_\mu) \Bigg] \\
&\vdots \\
&\geq c_n^T \hat{x}_n + \max_{\substack{p^n \in P_n \\ p^m \in P_m, m \in D(n) \\ \vdots \\ p^\sigma \in P_\sigma, \sigma \in D(a(\sigma))}} \sum_{\nu \in \mathcal{G}(n)} p_{n,m} p_{m,\mu} \cdots p_{\sigma,\nu} c_\nu^T \hat{x}_\nu \\
&\geq F_n(x_{a(n)}).
\end{aligned}$$

Together with results from (i) we conclude $\tilde{F}_n(x_{a(n)}) = F_n(x_{a(n)})$ for any node $n \in \mathcal{N}$ and finish the proof. □

For a more detailed discussion see Kall and Mayer [34], where the general nested decomposition algorithm for the multi-stage stochastic programs is presented. The proof technique used for the minimax approach is quite similar.

The algorithm is implemented in .NET platform using programming language C#. We also use the existing linear programming solver lpsolve55 downloadable from [36]. Since the solution procedure is a modification of a well-known algorithm, its efficiency is immediate.

The algorithm works also for more general sets of probability distributions, see the discussion at the end of section 4.2.

6.5 Construction of feasibility cuts

Let us consider the multi-stage minimax stochastic program in the nested form as presented in section 6.2 but now without the assumption (B1) of relatively complete recourse. We allow infeasibility of all stage primal problems, i.e. unboundedness of the corresponding dual objective functions.

We will have to modify Algorithm 6.1 by incorporating feasibility cuts into the solution procedures. If the problem $\text{Mast}(m, \hat{x}_n)$ is infeasible for $m \in D(n)$ then we add some further linear constraints

$$a_{nj}^T x_n \geq \alpha_{nj}, \quad j = 1, \ldots, r_n, \tag{6.21}$$

into the relaxed master problem $\text{Mast}(n, x_{a(n)})$ to ensure the finiteness of objective function $\tilde{F}_m, m \in D(n)$. By r_n we denote number of feasibility cuts added to the node n.

The sequence of successively built relaxed master problems $\text{Mast}(n, x_{a(n)})$ now takes the form:

$$\tilde{F}_n(x_{a(n)}) = \min_{x_n} c_n^T x_n + \theta_n$$

$$\begin{array}{rrcl} \text{s.t.} & W_n x_n & = & b_n - T_n x_{a(n)}, \\ & d_{nl}^T x_n + \theta_n & \geq & \delta_{nl}, \quad l = 1, \ldots, s_n, \\ & a_{nk}^T x_n & \geq & \alpha_{nj}, \quad k = 1, \ldots, r_n, \\ & 0 \leq x_n & \leq & u_n, \quad n \in \mathcal{N}_t, \quad t \in \{2, \ldots, T-1\}. \end{array} \tag{6.22}$$

Let $(\hat{x}_m, \hat{\theta}_m)$ be an optimal solution of $\text{Mast}(m, \hat{x}_n), m \in D(n)$. Take into consideration the dual formulation of problem $\text{Mast}(m, \hat{x}_n)$:

$$\begin{array}{ll} \max & (b_m - T_m \hat{x}_n)^T v_m + \sum_{l=1}^{s_m} \delta_{ml} w_{ml} + \sum_{k=1}^{r_m} \alpha_{mk} z_{mk} + u_m^T \lambda_m \\ \text{s.t.} & W_m^T v_m + \sum_{l=1}^{s_m} d_{ml} w_{ml} + \sum_{k=1}^{r_m} a_{mk} z_{mk} + 1^T \lambda_m \leq c_m, \\ & \sum_{l=1}^{s_m} w_{ml} = 1, \\ & w_{ml} \geq 0, \quad l = 1, \ldots, s_m, \\ & z_{mk} \geq 0, \quad k = 1, \ldots, r_m, \\ & \lambda_j \leq 0, \quad j = 1, \ldots, k_m. \end{array} \tag{6.23}$$

The infeasibility of $\text{Mast}(m, \hat{x}_n), m \in D(n)$ implies that the objective function of problem (6.23) is unbounded from above. We will follow the derivation of feasibility cuts presented in Kall and Mayer [34].

Farkas lemma 2.5 gives us the following conditions on empty set of feasible solutions of problem $\text{Mast}(m, \hat{x}_n)$:

(C1) There must exist a vector $(\tilde{v}_m, \tilde{w}_m, \tilde{z}_m, \tilde{\lambda}_m)$ such that

$$(b_m - T_m \hat{x}_n)^T \tilde{v}_m + \sum_{l=1}^{s_m} \delta_{ml} \tilde{w}_{ml} + \sum_{k=1}^{r_m} \alpha_{mk} \tilde{z}_{mk} + u_m^T \tilde{\lambda}_m \leq 0 \tag{6.24}$$

and

$$W_m^T\tilde{v}_m + \sum_{l=1}^{s_m} d_{ml}\tilde{w}_{ml} + \sum_{k=1}^{r_m} a_{mk}\tilde{z}_{mk} + 1^T\tilde{\lambda}_m > 0, \qquad (6.25)$$

$$1^T\tilde{w}_m = 0, \tilde{w}_m \in \mathbb{R}_+^{s_m}, \qquad (6.26)$$

$$\tilde{z}_m \in \mathbb{R}_+^{r_m}, \tilde{\lambda}_m \in \mathbb{R}_-^{k_m} \text{ and } \tilde{v}_m \in \mathbb{R}^{h_m}\text{[5]}.$$

Condition (6.26) implies $\tilde{w}_{ml} = 0$ for all $l = 1, \ldots, s_m$.

Hence, during the iteration process we add to the Mast$(n, x_{a(n)})$ a feasibility cut of the form (6.21), where

$$a_n := T_m^T\tilde{v}_m$$

and

$$\alpha_n := b_m^T\tilde{v}_m + \sum_{k=1}^{r_m} \alpha_{mk}\tilde{z}_{mk} + u_m^T\tilde{\lambda}_m,$$

to cut off $\hat{x}_n$, which has led to the infeasible problem Mast$(m, \hat{x}_n)$.

Before we formulate the algorithm for the multi-stage minimax program without the assumption of relatively complete recourse, we have to modify optimality cuts by involving dual variables corresponding to feasibility cuts.

Any current solution $(\hat{x}_n, \hat{\theta}_n)$ of Mast$(n, x_{a(n)})$ has to fulfil the following conditions for all $k \in \{1, \ldots, e_{P_n}\}$

$$\hat{\theta}_n \geq \sum_{m\in\mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|}\left[(b_m - T_m\hat{x}_n)^T\hat{v}_m + \sum_{l=1}^{s_m}\delta_{ml}\hat{w}_{ml} + \sum_{k=1}^{r_m}\alpha_{mk}\hat{z}_{mk} + u_m^T\hat{\lambda}_m\right]. \qquad (6.27)$$

For those k (let their number be $\bar{k}_{e_{P_n}}$) for which the condition (6.27) does not hold we add the following optimality cuts to the problem Mast$(n, x_{a(n)})$:

$$d_{n\nu}^T x_n + \theta_n \geq \delta_{n\nu}, \quad \nu = s_n + 1, \ldots, s_n + \bar{k}_{e_{P_n}}, \qquad (6.28)$$

where

$$\begin{aligned} d_{n\nu} &:= \sum_{m\in\mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|} T_m^T\hat{v}_m, \\ \delta_{n\nu} &:= \sum_{m\in\mathcal{T}_k^{P_n}} \frac{1}{|\mathcal{T}_k^{P_n}|}\left[b_m^T\hat{v}_m + \sum_{l=1}^{s_m}\delta_{ml}\hat{w}_{ml} + \sum_{k=1}^{r_m}\alpha_{mk}\hat{z}_{mk} + u_m^T\hat{\lambda}_m\right]. \end{aligned} \qquad (6.29)$$

According to the definition 6.1, such optimality cuts stay valid. The proof would be similar to the proof of theorem 6.1.

Definition 6.2. A feasibility cut $a_{nk}^T x_n \geq \alpha_{nk}$ in Mast$(n, x_{a(n)})$ will be called valid, if for any feasible solution $\bar{x}_\nu, \nu \in \mathcal{G}(n)$, of the descendant recourse problem rooted at node n, the inequality $a_{nk}\bar{x}_n \geq \alpha_{nk}$ holds.

[5] By h_m we denote the number of matrix W_m rows.

6.6 Nested decomposition – without the assumption of relatively complete recourse

Feasibility cuts generated by the algorithm presented below are valid. The algorithm will terminate in a finite number of iterations finding the optimal solutions if the solved problem is feasible. The argumentation is similar as in the two-stage problem, see the proof of theorem 5.1. The proof of validity of feasibility cuts generated by the algorithm is similar as in the multi-stage stochastic linear program, we refer e.g. to Kall and Mayer [34].

Algorithm 6.2. Multicut nested decomposition for the multi-stage minimax problem with a qualitative information — without the assumption of relatively complete recourse:

Step 1 (*Initialization*) Let $s_n = 0$, $r_n = 0$, $\gamma_n =$ False and add the constraint $\theta_n = 0$ to $\mathrm{Mast}(n, x_{a(n)})$ for all $n \in \mathcal{N}$. Set $t = 1$.

Step 2 (*Forward pass*) For $n \in \mathcal{N}_t : \gamma_n =$ False do:

1. Solve $\mathrm{Mast}(n, \hat{x}_{a(n)})$. If the problem is infeasible then store dual solution $(\tilde{v}_n, \tilde{z}_n, \tilde{\lambda}_n)$ which fulfils condition (C1) and continue with Step 4.
2. Store the solution $(\hat{x}_n, \hat{\theta}_n)$. If $t = T$ then store also the dual solution $(\hat{v}_n, \hat{w}_n, \hat{z}_n, \hat{\lambda}_n)$.
3. Set $\gamma_n =$ True and $\gamma_\nu =$ False, $\forall \nu \in \mathcal{G}(n) \backslash \{n\}$.
4. Take the next node in $\mathcal{N}_t$.

If $t = T$ then continue with Step 3, else set $t = t + 1$ and repeat Step 2.

Step 3 (*Backward pass*) Set $\gamma =$ True. For $n \in \mathcal{N}_{t-1}$ check whether (6.27) holds for $(\hat{x}_n, \hat{\theta}_n)$:

Yes Take the next node in $\mathcal{N}_{t-1}$.

No Add optimality cuts:

1. Set $\gamma =$ False.
2. If $s_n = 0$ then drop the constraint $\theta_n = 0$.
3. Add optimality cuts (6.28) to $\mathrm{Mast}(n, \hat{x}_{a(n)})$ (let them be $\bar{e}_{p_n}$) with $\nu = s_n + 1, \ldots, s_n + \bar{e}_{P_n}$. Set $s_n = s_n + \bar{e}_{P_n}$.
4. Solve $\mathrm{Mast}(n, \hat{x}_{a(n)})$ and temporarily store the dual solution $(\hat{v}_n, \hat{w}_n, \hat{z}_n, \hat{\lambda}_n)$. Set $\gamma_\nu =$ False, $\nu \in \mathcal{G}(n) \setminus \{n\}$.

If $t = 1$ and $\gamma =$ True then no optimality cut has been added and the current solution is optimal. STOP.
If $t > 1$ then set $t = t - 1$ and repeat Step 3.
If $t = 1$ and $\gamma =$ False then return to Step 2.

Step 4 (*Backtracking*) If $n = 1$ the STOP. Then the multi-stage minimax problem is infeasible. Otherwise

1. Set $m := n$ and $n := a(n)$.
2. Add feasibility cut (6.21) to Mast$(n, \hat{x}_{a(n)})$. Set $r_n = r_n + 1$. Set $\gamma_\nu :=$ False for all $\nu \in \mathcal{G}(n)$.
3. Solve Mast$(n, \hat{x}_{a(n)})$.

If the problem is infeasible then store the dual solution $(\tilde{v}_n, \tilde{z}_n, \tilde{\lambda}_n)$ which fulfils condition (C1) and continue with Step 4. Otherwise set $\gamma_n :=$ True, store the solution $(\hat{x}_n, \hat{\theta}_n)$ and return to Step 1.

6.7 Numerical study

In the next three examples we apply the algorithm 6.1 to simple illustrative problems. The fourth example shows a possible application to the multi-stage stochastic minimax problem of portfolio selection having only a qualitative information about the probability distributions. Both examples were computed by the above mentioned developed solver.

Example 6.4. Consider a scenario tree in the figure 6.4, with $\mathcal{N} = \{1, \ldots, 7\}$ and $T = 3$. The only available information on probability distribution is of the form $p_2 \geq p_3$, $p_{2,4} \geq p_{2,5}$ and $p_{3,7} \geq p_{3,6}$. We obtain the following extreme probability distributions

$$(1,0)^T, (\frac{1}{2}, \frac{1}{2})^T \in P_1, (1,0)^T, (\frac{1}{2}, \frac{1}{2})^T \in P_2, (0,1)^T, (\frac{1}{2}, \frac{1}{2})^T \in P_3, \tag{6.30}$$

with the corresponding admissible sets $\mathcal{T}_1^{P_1} = \{2\}, \mathcal{T}_2^{P_1} = \{2,3\}$, $\mathcal{T}_1^{P_2} = \{4\}, \mathcal{T}_2^{P_2} = \{4,5\}$, $\mathcal{T}_1^{P_3} = \{7\}, \mathcal{T}_2^{P_3} = \{6,7\}$.

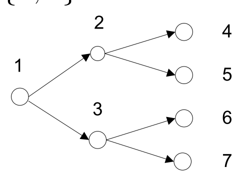

Figure 6.4: Scenario tree.

The problem to be solved is

$$\min_{\substack{x_i \in \mathbb{R}^2 \\ i=1,\ldots,7}} \Bigg[\begin{pmatrix} 1 \\ 0.2 \end{pmatrix}^T x_1 + p_2 \begin{pmatrix} 1.5 \\ 0.2 \end{pmatrix}^T x_2 + p_3 \begin{pmatrix} 0.9 \\ 0.2 \end{pmatrix}^T x_3 +$$

$$+ p_2 p_{2,4} \begin{pmatrix} 1.8 \\ 0.2 \end{pmatrix}^T x_4 + p_2 p_{2,5} \begin{pmatrix} 1.4 \\ 0.2 \end{pmatrix}^T x_5 + p_3 p_{3,6} \begin{pmatrix} 1.4 \\ 0.2 \end{pmatrix}^T x_6 +$$

$$+ p_3 p_{3,7} \begin{pmatrix} 0.9 \\ 0.2 \end{pmatrix}^T x_7 \Bigg]$$

$$\begin{aligned}
\text{s. t.} \qquad\qquad\qquad\qquad \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_1 &= 10, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_1 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_2 &= 15, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_1 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_3 &= 8, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_2 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_4 &= 17, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_2 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_5 &= 10, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_3 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_6 &= 10, \\
\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T x_3 + \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T x_7 &= 5, \\
\begin{pmatrix} 0 \\ 0 \end{pmatrix} \le x_i \le \begin{pmatrix} 100 \\ 3 \end{pmatrix}&, \; i = 1, \ldots, 7.
\end{aligned}$$

During the first forward pass we calculate solutions

	Objective	Primal	Dual
$\tilde{F}_1$	10	(10,0,0)	(1,0,0),
$\tilde{F}_2(\hat{x}_1)$	22.5	(15,0,0)	(1.5,0,0),
$\tilde{F}_3(\hat{x}_1)$	7.2	(8,0,0)	(0.9,0,0),
$\tilde{F}_4(\hat{x}_2)$	30.6	(17,0)	(1.8,0,0),
$\tilde{F}_5(\hat{x}_2)$	14	(10,0)	(1.4,0,0),
$\tilde{F}_6(\hat{x}_3)$	14	(10,0)	(1.4,0,0),
$\tilde{F}_7(\hat{x}_3)$	4.5	(5,0)	(0.9,0,0).

Primal solutions in sequence represent the optimal values of $x_i, i = 1, \dots, 3$, and $\theta_i, i = 1, \dots, 7$. Dual solutions correspond to the equality constraints and upper bounds on $x_i, i = 1, \dots, 7$.

In backward pass we obtain optimality cuts

$$\text{Mast}(1): \quad \begin{pmatrix} 0 \\ 1.5 \end{pmatrix}^T x_1 + \theta_1 \geq 52.8, \begin{pmatrix} 0 \\ 1.2 \end{pmatrix}^T x_1 + \theta_1 \geq 34.55,$$

$$\text{Mast}(2, \hat{x}_1): \quad \begin{pmatrix} 0 \\ 1.8 \end{pmatrix}^T x_2 + \theta_2 \geq 30.6, \begin{pmatrix} 0 \\ 1.6 \end{pmatrix}^T x_2 + \theta_2 \geq 22.3,$$

$$\text{Mast}(3, \hat{x}_1): \quad \begin{pmatrix} 0 \\ 0.9 \end{pmatrix}^T x_3 + \theta_3 \geq 4.5, \begin{pmatrix} 0 \\ 1.15 \end{pmatrix}^T x_3 + \theta_3 \geq 9.25,$$

and updated optimal solutions

	Objective	Primal	Dual
$\tilde{F}_1$	61.9	(13,3,48.3)	(1,0,-0.3,1,0),
$\tilde{F}_2(\hat{x}_1)$	48.3	(15,3,25.2)	(1.5,0,-0.1,1,0),
$\tilde{F}_3(\hat{x}_1)$	13.6	(8,3,5.8)	(0.9,0,-0.05,0,1),
$\tilde{F}_4(\hat{x}_2)$	25.2	(14,0)	(1.8,0,0),
$\tilde{F}_5(\hat{x}_2)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_6(\hat{x}_3)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_7(\hat{x}_3)$	1.8	(2,0)	(0.9,0,0).

Dual solutions are extended by the values corresponding to new optimality cuts.

In the next backward pass no optimality cut is added. The current solution is optimal.

From the dual solutions we obtain the coefficients of optimality cuts (6.29), which are active for the probability distributions

$$(1, 0)^T \in P_1, (1, 0)^T \in P_2, (\frac{1}{2}, \frac{1}{2})^T \in P_3, \tag{6.31}$$

among all extreme probability distributions (6.30). The active cuts are

$$\begin{pmatrix} 0 \\ 1.5 \end{pmatrix}^T x_1 + \theta_1 = 52.8, \begin{pmatrix} 0 \\ 1.8 \end{pmatrix}^T x_2 + \theta_2 = 30.6, \begin{pmatrix} 0 \\ 1.15 \end{pmatrix}^T x_3 + \theta_3 = 9.25,$$

corresponding to the worst-case probability distributions (6.31).

Paths with nonzero probabilities are illustrated in the figure 6.5 with a blue color. Only the scenario $1 - 2 - 4$ is significant for the optimal value

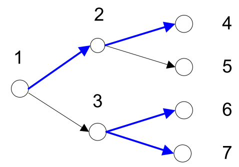

Figure 6.5: Final scenario tree.

of the objective function. The extremal transition probabilities $(\frac{1}{2}, \frac{1}{2})^T \in P_3$ between node 3 and its child nodes will not influence the optimal value. The contribution of a subtree rooted at node 3 will be erased by the fact that there is zero probability for getting to node 3 from node 1.

In fact, the algorithm returns the worst-case probability distributions for all paths of the considered scenario tree. Due to the Markov property the worst-case transition probabilities from node 3 to node 7 cannot depend on the worst-case marginal probability of node 3. Therefore, we also obtain the nonzero worst-case transition probabilities for nodes, which cannot influence the value of the objective function.

Example 6.5. Consider the scenario tree and the optimization problem as in example 6.4 with the different qualitative information $p_3 \geq p_2$, $p_{2,4} \geq p_{2,5}$ and $p_{3,7} \geq p_{3,6}$.

We also obtain a different optimal value of the objective function and different worst-case probability distributions. There exist two optimal decision strategies:

	Objective	Primal	Dual
$\tilde{F}_1$	44.55	(13,3,30.95)	(1,0,0,0,1),
$\tilde{F}_2(\hat{x}_1)$	48.3	(15,3,25.2)	(1.5,0,-0.1,1,0),
$\tilde{F}_3(\hat{x}_1)$	13.6	(8,3,5.8)	(0.9,0,-0.05,0,1),
$\tilde{F}_4(\hat{x}_2)$	25.2	(14,0)	(1.8,0,0),
$\tilde{F}_5(\hat{x}_2)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_6(\hat{x}_3)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_7(\hat{x}_3)$	1.8	(2,0)	(0.9,0,0),

and

	Objective	Primal	Dual
$\tilde{F}_1$	44.55	(10,0,34.55)	(1,0,0,0,1),
$\tilde{F}_2(\hat{x}_1)$	52.8	(18,3,25.2)	(1.5,0,-0.1,1,0),
$\tilde{F}_3(\hat{x}_1)$	16.3	(11,3,5.8)	(0.9,0,-0.05,0,1),
$\tilde{F}_4(\hat{x}_2)$	25.2	(14,0)	(1.8,0,0),
$\tilde{F}_5(\hat{x}_2)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_6(\hat{x}_3)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_7(\hat{x}_3)$	1.8	(2,0)	(0.9,0,0).

The worst-case probability distributions

$$(\frac{1}{2}, \frac{1}{2})^T \in P_1, (1,0)^T \in P_2, (\frac{1}{2}, \frac{1}{2})^T \in P_3.$$

are illustrated in the figure 6.6:

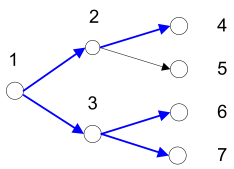

Figure 6.6: Final scenario tree.

Example 6.6. Consider again the example 6.4 with no additional qualitative information on probability distributions, i.e. using only constraints

$$P_1 = \{p \in \mathbb{R}^2_+ : p_1 + p_2 = 1\},$$

$$P_2 = \{p^2 \in \mathbb{R}^2_+ : p_{2,4} + p_{2,5} = 1\},$$

$$P_3 = \{p^3 \in \mathbb{R}^2_+ : p_{3,6} + p_{3,7} = 1\}.$$

We obtain the following optimal solutions:

	Objective	Primal	Dual
$\tilde{F}_1$	61.9	(13,3,48.3)	(1,0,-0.3,1,0),
$\tilde{F}_2(\hat{x}_1)$	48.3	(15,3,25.2)	(1.5,0,-0.1,1,0),
$\tilde{F}_3(\hat{x}_1)$	17.6	(8,3,9.8)	(0.9,0,-0.3,1,0),
$\tilde{F}_4(\hat{x}_2)$	25.2	(14,0)	(1.8,0,0),
$\tilde{F}_5(\hat{x}_2)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_6(\hat{x}_3)$	9.8	(7,0)	(1.4,0,0),
$\tilde{F}_7(\hat{x}_3)$	1.8	(2,0)	(0.9,0,0).

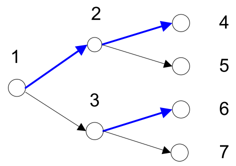

Figure 6.7: Final scenario tree.

The worst-case probability distributions

$$(1,0)^T \in P_1, (1,0)^T \in P_2, (1,0)^T \in P_3$$

are illustrated in the figure 6.7:

Compare these results with example 6.4. The additional information could not improve the optimal value in the example 6.4. Such an information can be also interpreted as *worthless*.

Example 6.7. We review a multi-stage stochastic minimax model for portfolio selection. The model we present determines an expected minimal transaction and borrowing costs strategy for bonds whose incomes are used to meet given cash requirements or liabilities. We apply the following notation:

- x_{j0} initial holding of bond j at the beginning of period 1 ($j = 1, \dots, J$);
- z_0 initial cash at the beginning of period 1;
- x_{jn_t} quantity of bond j held at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- b_{jn_t} quantity of bond j purchased at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- s_{jn_t} quantity of bond j sold at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- $z^b_{jn_t}$ unit purchasing price of bond j at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- $z^s_{jn_t}$ unit selling price of bond j at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- f_{jn_t} cash flow produced by bond j held at node n_t at the beginning of period t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- y_{n_t} additional cash requirements at node n_t ($n_t \in \mathcal{N}_t, t = 2, \dots, T$);
- u_{n_t} total cumulative borrowed cash up to node n_t ($n_t \in \mathcal{N}_t, t = 2, \dots, T$);
- z_{n_t} cash surplus at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);
- L_{n_t} cash requirements or liabilities at node n_t ($n_t \in \mathcal{N}_t, t = 1, \dots, T$);

c_{jn_t} transaction costs for buying/selling bond j at node n_t ($j = 1, \ldots, J, n_t \in \mathcal{N}_t, t = 1, \ldots, T$);

d_{n_t} present value factor at node n_t ($n_t \in \mathcal{N}_t, t = 2, \ldots, T$);

$i^+_{n_t}$ interest rate factor for savings at node n_t ($n_t \in \mathcal{N}_t, t = 2, \ldots, T$);

$i^-_{n_t}$ interest rate factor for borrowing at node n_t ($n_t \in \mathcal{N}_t, t = 2, \ldots, T$);

p_{n_t} probability of node n_t ($n_t \in \mathcal{N}_t, t = 1, \ldots, T$).

The problem is to minimize over $y_{n_t}, z_{n_t}, u_{n_t}, x_{jn_t}, b_{jn_t}, s_{jn_t}$ and maximize over p_n the expected present value of total transaction and borrowing costs, i.e.

$$\sum_{j=1}^{J} c_{j1}(b_{j1} + s_{j1}) + \sum_{n_2 \in \mathcal{N}_2} p_{n_2} d_{n_2} \left[y_{n_2} + \sum_{j=1}^{J} c_{jn_2}(b_{jn_2} + s_{jn_2}) \right] + \cdots + \\ + \sum_{n_T \in \mathcal{N}_T} p_{n_T} d_{n_T} \left[y_{n_T} + \sum_{j=1}^{J} c_{jn_T}(b_{jn_T} + s_{jn_T}) \right]$$

s.t.

$$x_{j0} + b_{j1} - s_{j1} = x_{j1}, \quad j = 1, \ldots, J, \tag{6.32}$$

$$\sum_{j=1}^{J} z^s_{j1} s_{j1} + \sum_{j=1}^{J} f_{j1} x_{j0} + z_0 = L_1 + z_1 + \sum_{j=1}^{J} z^b_{j1} b_{j1} + \sum_{j=1}^{J} c_{j1}(b_{j1} + s_{j1}), \tag{6.33}$$

$$x_{ja(n_t)} + b_{jn_t} - s_{jn_t} = x_{jn_t}, \quad j = 1, \ldots, J, \tag{6.34}$$

$$u_1 = 0, u_{n_t} = u_{a(n_t)} + y_{n_t}, \tag{6.35}$$

$$\begin{aligned} &\sum_{j=1}^{J} z^s_{jn_t} s_{jn_t} + \sum_{j=1}^{J} f_{jn_t} x_{ja(n_t)} + (1 + i^+_{n_t}) z_{a(n_t)} + y_{n_t} = \\ &= L_{n_t} + z_{n_t} + \sum_{j=1}^{J} z^b_{jn_t} b_{jn_t} + \sum_{j=1}^{J} c_{jn_t}(b_{jn_t} + s_{jn_t}) + i^-_{n_t} u_{a(n_t)}, \end{aligned} \tag{6.36}$$

for $n_t \in \mathcal{N}_t, t = 2, \ldots, T$.

We also assume all decision variables are nonnegative and upper bounded. Conditions (6.32) and (6.34) represent inventory constraints. Budget constraints are involved in (6.33) and (6.36). Condition (6.35) defines total cumulative borrowed cash.

We apply the model on the scenario tree with a given qualitative information about probability distributions presented in the example 6.3 with the following input data:

- initial values $T = 4,\ J = 3,\ z_0 = 100$, and $x_{j0} = 30, j = 1, 2, 3$;

- upper bounds on decision variables:

node	u_{n_t}	y_{n_t}	z_{n_t}	b_{1n_t}	b_{2n_t}	b_{3n_t}
1			10000	100	100	100
2 - 4	10000	10000	10000	100	100	100
5 - 11	10000	10000	10000	5	5	5
12 - 24	10000	10000	10000	10	10	10

node	s_{1n_t}	s_{2n_t}	s_{3n_t}	x_{1n_t}	x_{2n_t}	x_{3n_t}
1	100	100	100	100	100	100
2 - 4	100	100	100	100	100	100
5 - 11	5	5	5	100	100	100
12 - 24	10	10	10	100	100	100

- bid and ask bonds' prices:

node	$z^b_{1_{n_t}}$	$z^b_{2_{n_t}}$	$z^b_{3_{n_t}}$	$z^s_{1_{n_t}}$	$z^s_{2_{n_t}}$	$z^s_{3_{n_t}}$
1	102	92	83	98	88	79
2	103	89	81	99	85	77
3	101	94	88	97	90	84
4	102	88	72	98	84	68
5	92	92	73	88	88	69
6	93	81	76	89	77	72
7	104	88	77	100	84	73
8	107	86	75	103	82	71
9	111	93	83	107	89	79
10	101	98	84	97	94	80
11	100	87	71	96	83	67
12	93	83	75	89	79	71
13	96	79	79	92	75	75
14	92	91	80	88	89	76
15	110	94	83	106	90	79
16	112	95	73	108	91	69
17	106	89	74	102	85	70
18	99	88	79	95	84	75
19	100	83	69	96	79	65
20	105	79	79	101	75	75
21	104	86	74	100	82	70
22	98	90	66	94	86	62
23	95	84	79	91	80	75
24	97	88	80	93	84	76

- transaction costs, cash flows produced by bonds holding and cash requirements:

node	$i^-_{n_t}$	$i^+_{n_t}$	d_{n_t}	f_{1n_t}	f_{2n_t}	f_{3n_t}	L_{n_t}
1						110	
2	0.026	0.026	0.9876	0.05	0.05	0043	1010
3	0.024	0.024	0.9768	0.055	0.055	0.0443	1010
4	0.03	0.03	0.9788	0.06	0.06	0.063	1010
5	0.02	0.02	0.7657	0.054	0.055	0.048	1100
6	0.018	0.018	0.6785	0.045	0.055	0.044	1100
7	0.026	0.026	0.8765	0.066	0.066	0.066	1100

node	$i^-_{n_t}$	$i^+_{n_t}$	d_{n_t}	f_{1n_t}	f_{2n_t}	f_{3n_t}	L_{n_t}
8	0.016	0.016	0.7444	0.05	0.05	0.04	1100
9	0.017	0.017	0.7564	0.045	0.056	0.049	1100
10	0.017	0.017	0.7566	0.055	0.054	0.044	1100
11	0.016	0.016	0.7456	0.054	0.055	0.048	1100
12	0.02	0.02	0.5643	0.0554	0.0555	0.0444	1250
13	0.026	0.026	0.5678	0.0478	0.0578	0.0478	1250
14	0.022	0.022	0.4567	0.049	0.059	0.049	1250
15	0.032	0.032	0.5443	0.057	0.057	0.047	1250
16	0.025	0.025	0.4786	0.039	0.039	0.039	1250
17	0.04	0.04	0.6785	0.06	0.06	0.06	1250
18	0.028	0.028	0.4899	0.062	0.062	0.062	1250
19	0.031	0.031	0.5666	0.056	0.056	0.046	1250
20	0.033	0.033	0.5866	0.053	0.053	0.043	1250
21	0.018	0.018	0.4788	0.055	0.055	0.044	1250
22	0.038	0.038	0.6522	0.048	0.058	0.048	1250
23	0.022	0.022	0.5324	0.067	0.067	0.067	1250
24	0.029	0.029	0.5866	0.066	0.066	0.066	1250

The algorithm terminated in 7 iterations with the following results:

- nonzero optimal decisions:

node	z_{n_t}	s_{1n_t}	s_{2n_t}	s_{3n_t}	x_{1n_t}	x_{2n_t}	x_{3n_t}
1	1110.65	11.44	0	0	18.56	30	30
2	211.72	0.79	0	0	17.77	30	30
3	477.08	3.56	0	0	15	30	30
4	138.68	0	0	0	18.56	30	30
5	0	5	5	0	12.77	25	30
6	0	5	5	0.70	12.77	25	29.30
7	0	5	1.26	0	10	28.74	30
8	0	5	1.18	0	10	28.82	30
9	0	5	0.85	0	10	29.15	30
10	0	5	5	0	13.56	25.00	30
11	0	5	5	0.90	13.56	25	29.10
12	0	10	4.51	0	2.77	20.49	30

node	z_{n_t}	s_{1n_t}	s_{2n_t}	s_{3n_t}	x_{1n_t}	x_{2n_t}	x_{3n_t}
13	0	10	4.35	0	2.77	20.65	30
14	0	4.05	10	0	8.72	15	30
15	0	10	2.07	0	2.77	22.93	29.30
16	0	10	1.84	0	0	26.90	30
17	10	0	2.66	0	0	26.09	30
18	0	10	3.52	0	0	25.30	30
19	0	10	3.63	0	0	25.52	30
20	0	10	3.16	0	0	25.99	30
21	0	10	3.01	0	3.56	21.99	30
22	0	10	3.56	0	3.56	21.44	30
23	0	10	4.19	0	3.56	20.81	30
24	0	10	3.76	0	3.56	21.24	29.10

- optimal values of objectives functions and worst-case nodes probabilities:

node	obj.	$p_{a(n_t)n_t}$	node	obj.	$p_{a(n_t)n_t}$	node	obj.	$p_{a(n_t)n_t}$
1	28.26	1	9	12.15	0	17	8.59	1
2	16.63	0.33	10	16.41	0.5	18	6.62	1
3	17.55	0.33	11	16.19	0.5	19	7.72	0.5
4	16.30	0.33	12	8.19	1	20	7.72	0.5
5	15.85	1	13	8.15	0	21	6.23	0
6	13.83	0	14	6.42	0	22	8.85	1
7	14.07	1	15	6.57	1	23	7.56	0
8	11.22	0	16	5.67	0	24	8.07	1

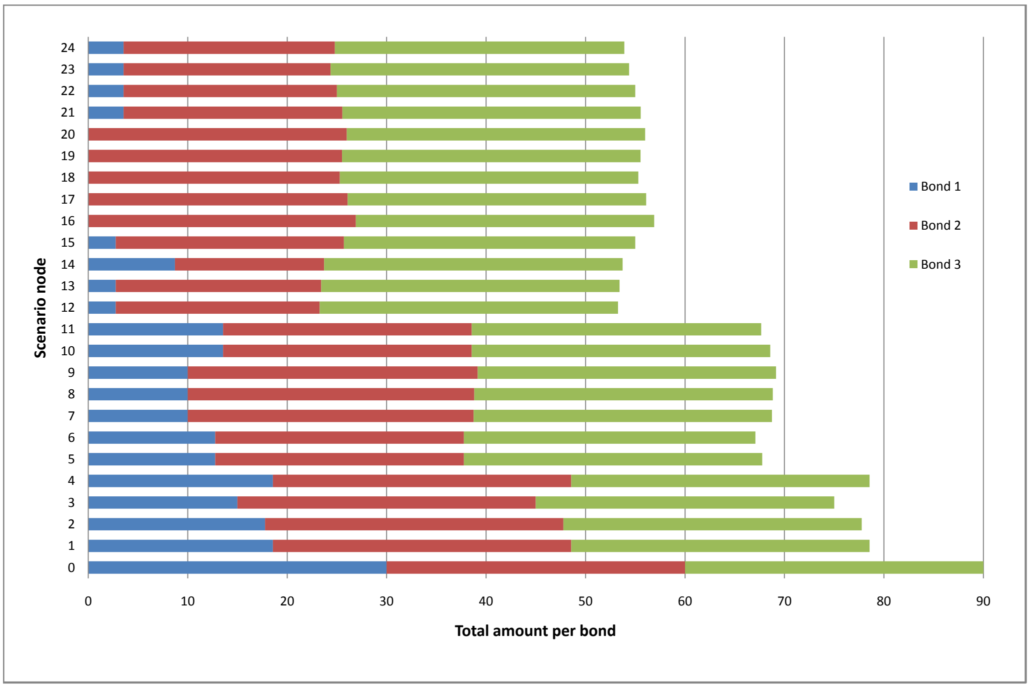

Figure 6.8: Optimal asset allocations at scenario nodes (Node 0 represents the initial allocation).

Chapter 7

Summary and open problems

In the thesis, the stochastic programming problems with incomplete information on probability distributions were discussed. The incomplete information inheres many real applications, where the perfect knowledge of probability distribution is concealed from us. Therefore, it is very important to understand and deal with such situations. We considered the following ways of solutions.

Firstly, we dealt with problems with estimated or believed probability distributions. For obtained solutions the analysis of stability becomes crucial task ensuring that small input modifications cause only small changes in optimal values and optimal decisions. General stability results were applied to the Bayes decision problem and improved error bounds for the optimal Bayes actions related to perturbations of the input were provided. In addition, classical stability properties with respect to weak convergence were analyzed.

The second possible approach incorporates the incomplete knowledge of probability distribution to the problem formulation. The decision maker includes all possible probability distributions of modeled random variables to a set of feasible distributions. His optimal decision then reflects all the considered distributions in an effort to protect himself against the worst-case alternative. This leads to the minimax approach. We mentioned the most common choices of distributions' sets and studied two of them in detail.

The worst-case approach applied on the set of probability distributions fulfilling certain moment conditions was used on two well-known risk measures — Value-at-Risk and Conditional Value-at-Risk. We derived upper bound for these measures under different assumptions on probability distributions. Except the knowledge of the first two moments we supposed subsequently other properties like symmetry and unimodality. Newly, we introduced the worst-case CVaR for symmetric distributions and correctly deduced the worst-case VaR for symmetric and unimodal distributions, both for the given mean and variance. The duality statements in the moment problem theory were crucial for presented results.

The set of probability distributions consistent with a special type of qualitative information was the second choice. Such a set is a bounded polyhedron and a linear function of probability distribution is maximized at least at one of its extreme points. For the considered type of qualitative information defined on a given finite set of possible realizations and corresponding to the educated guess of the form that "one realization is more probable than the other" we were able to precisely express all extreme probability distributions and significantly simplify computations of such minimax problems. We derived algorithms with and without the assumption of relatively complete recourse for the two-stage and the multi-stage minimax stochastic problems with linear recourse. The algorithms for the two-stage problems are based on the L-shaped algorithm. In the multi-stage case the nested decomposition was modified.

Considering the presented algorithms some improvements and extensions are possible:

- Nodes with zero probability in each extreme probability distribution can be excluded from the computation. This should make the algorithms faster.

- Other sequencing protocols in the multi-stage case can be considered. The variant which has been discussed in the thesis implements the FFFB (fast-forward-fast-backward) protocol.

- The algorithms have not been tested on very large multi-stage stochastic programs yet. The comparison of efficiency with other existing methods would be also an interesting task.

- Development of an extreme points generator and considering other types of distributions' sets can be the subject of further research.

Eventually, some open theoretical problems still remain:

- Definitely, other necessary and sufficient conditions for the Bayes decision problem to be stable with respect to a weak convergence of probability measures can be find in literature. It makes sense to study relations among them and their possible compatibility.

- Stability analysis of the two-stage minimax problem is rare; see e.g. in Riis [45], Riis and Andersen [47] for case of the two-stage minimax (integer) problems.

- The asymptotic properties of optimal values of the minimax stochastic programming problem is open for a further research, see Shapiro [58].

Bibliography

[1] Acerbi, C., Tasche, D. (2002): On the Coherence of Expected Shortfall. *J. of Bank. Fin.*, **26**, 1487 – 1503.

[2] Berger, J. O. (1980): Statistical Decision Theory: Foundations, Concepts and Methods. Springer Series in Statistics, Springer Verlag, New York.

[3] Berger, J. O., Salinetti, G. (1995): Approximations of Bayes Decision Problems: The Epigraphical Approach. *Annals of Operations Research* **56**, 1 – 13.

[4] Billingsley, P.,Topsøe, F. (1967): Uniformity in Weak Convergence. *Z. Wahrsheinlichkeitstheorie verw. Gebiete*, **7**, 1 – 16.

[5] Breton, M., Hachem, S. E. (1994): Algorithms for the Solutions of Stochastic Dynamic Minimax Problems. *Computational Optimization and Applications*, **4**, 317 – 345.

[6] Bühler, W. (1975): Characterization of the Extreme Points of a Class of Special Polyhedra. *Zeitschrift für Operations Research*, **19**, 131 – 137.

[7] Bühler, W., Gehring, H. (1978): Short-Term Financial Planning with Uncertain Receipts and Disbursements. *European Journal of Operations Research*, **2**, 158 – 167.

[8] Bühler, W. (1981): Capital Budgeting under Qualitative Data Information. Capital Budgeting under Conditions of Uncertainty, Crum, R. L., Denkinderen, F. G. J. (eds.), Martinus Nijhoff Publishing, Boston, 81 – 117.

[9] Čerbáková, J. (2004): Minimaxové kritérium ve finančním rozhodování. Master thesis, Charles University, Prague.

[10] Čerbáková, J. (2005): Moment Problem and Value-at-Risk. *WDS'05 Proceedings of Contributed Papers*, Part I, Šafránková, J. (ed.), Matfyzpress, 102-107.

[11] Čerbáková, J. (2005): Moment Problem and Worst-Case Value-at-Risk. *Proceedings of the 23rd International Conference MME2005*, Skanská, H. (ed.), Gaudeaumus, 33-38.

[12] Čerbáková, J. (2006): Worst-case VaR and CVaR. *Operation Research Proceedings 2005: Selected Papers of the Annual International Conference of the German Operations Research Society (GOR)*, Haasis, H. D, Kopfer, H., Schönberger, J. (eds.), Springer Verlag, New York, 817 – 822.

[13] Čerbáková, J. (2006): Stability of Bayesian Actions. *Proceedings of the 24th International Conference MME2006*, L. Lukáš (ed.), 99-104.

[14] Čerbáková, J. (2008): Input Information for Multi-Stage Stochastic Programs. Accepted in IMA Journal of Management Mathematics.

[15] Charnes, A., Cooper, W. W. (1959): Chance Constrained Programming. *Management Science*, **6**, 73 – 79.

[16] Dantzig, G. (1995): Linear Programming under Uncertainty. *Management Science*, **1**, 197-206.

[17] Dupačová, J. (1977): Minimaxová úloha stochastického lineárního programování a momentový problém. *Ekonomicko-matematický obzor*, **13** (3), 279 – 307.

[18] Dupačová, J. (1980): On Minimax Decision Rule in Stochastic Linear Programming. Studies in Mathematical Programming: A. Prékopa (eds.), Akadémiai Kiadó, 47 – 60.

[19] Dupačová, J. (1987): The Minimax Approach to Stochastic Programming and an Illustrative Application. *Stochastics*, **20**, 73 – 88.

[20] Dupačová, J., Römisch, W. (1998): Quantitative Stability for Scenario-based Stochastic Programs. Prague Stochastics '98, Hušková, M. et al. (eds.), JČMF, 119 – 124.

[21] Dupačová, J. (2001): Stochastic Programming: Minimax Approach. Encyclopedia of Optimization, Floudas, Ch. A., Pardalos, P. M. (eds.), Kluwer Academic Publishers, **V**, 327 – 330.

[22] Dupačová, J. (2006): Contamination for Multistage Stochastic Programs. Proc. of Prague Stochastics, Hušková, M., Janžura, M. (eds.), Matfyzpress, 91 – 101.

[23] Ermoliev, Y., Gaivoronsky, A., Nedeva, C. (1985): Stochastic Optimization Problems with Incomplete Information on Distribution Functions. *SIAM Journal on Control and Optimization*, **23**, 697 – 716.

[24] Freuendorfer, K., Marohn, Ch. (1996): Refinement Issues in Stochastic Multistage Linear Programming. Stochastic Programming Methods and Technical Applications, *Lecture Notes in Economics and Mathematical Systems*, Springer Verlag, New York, **458**, 305 – 328.

[25] Gupta, M. K. (1967): An Asymptotical Nonparametric Test of Symmetry. *Ann. Math. Stat*, **3**, 846 – 866.

[26] Hartigan, J. A., Hartigan, P. M. (1985): The DIP Test of Unimodality. *Annals of Statistics*, **13** (1), 70 – 74.

[27] Heitsch, H., Römisch, W., Strugarek, C. (2006): Stability of Multistage Stochastic Programs. *SIAM Journal on Control and Optimization*, **17**, 511 – 525.

[28] Iosifescu, M., Theodorescu, R. (1963): Sur la Programmation Linéaire. *C. R. Acad. Sci. Paris*, **256**, 4831 – 4833.

[29] Isii, K. (1963): On the Sharpness of Chebyshev-Type Inequalities. *Ann. Inst. Stat. Math.*, **14**, 185 – 197.

[30] Jagannathan, R. (1985): Use of Sample Information in Stochastic Recourse and Chance-Constrained Programming Models. *Management Science*, **31** (1), 96 – 108.

[31] Kadane, J., Chuang, D. T. (1978): Stable Decision Problems. *Annals of Statistics*, **10**, 1095 – 1110.

[32] Kadane, J., Srinivasan, C. (1998): Bayes Decision Problems and Stability. *Indian Journal of Statistics*, **60**, 383 – 404.

[33] Kadane, J., Salinetti, G., Srinivasan, C. (2000): Stability of Bayes Decisions and Applications. Robust Bayesian Analysis, Lecture Notes in Statistics, Springer Verlag, New York, **152**, 187 – 196.

[34] Kall, P., Mayer, J. (2005): Stochastic Linear Programming: Models, Theory, and Computations. Springer's International Series in Operations Research & Management Science, Kluwer Academic Publishers.

[35] Kall, P., Wallace, S. W. (1995): Stochastic Programming. John Wiley & Sons.

[36] Open source Linear Programming system lpsolve 5.5. URL http://lpsolve.sourceforge.net.

[37] Lucchetti, R., Salinetti, G., Wets, R, J-B (1994): Uniform Convergence of Proability Measures: Topological Criteria. *Journal of Multivariate Analysis* **51**, 252 – 264.

[38] Pflug, G., Ch. (2000): Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. Probabilistic Constrained Optimization: Methodology and Applications, Uryasev, S. (ed.), Kluwer Academic Publishers.

[39] Pflug, G. Ch., Wozabal, D. (2007): Ambiguity in Portfolio Selection. *Quantitative Finance*, **7** (4), 435 – 442.

[40] Popescu, I. (2005): A Semidefinite Programming Approach to Optimal Moment Bounds for Convex Classes of Distributions. *Mathematics of Operations Research*, **30** (3), 632 – 657.

[41] Prékopa, A. (1973): Contribution to the Theory of Stochastic Programming. *Math. Prog*, **4**, 202 – 221.

[42] Prékopa, A. (1995): Stochastic Programming. Kluwer Academic Publishers, Dordrecht and Akadémiai Kiadó, Budapest.

[43] R Development Core Team (2003): R: A Language and Environment for Statistical Computing. R Found. for Stat. Comp., Vienna, Austria, URL http://www.R-project.org.

[44] Rachev, S. T. (1991): Probability Metrics and the Stability of Stochastic Models, John Wiley & Sons.

[45] Riis, M. (2003): Classical and Non-Classical Stochastic Recourse Programs with Applications in Telecommunications. PhD thesis. Department of Operation Research, University of Aarhus.

[46] Riis, M., Schultz, R. (2003): Applying the Minimum Risk Criterion in Stochastic Recourse Programs. *Computational Optimization and Applications*, **24**, 267 – 287.

[47] Riis, M., Andersen, K. A. (2005): Applying the Minimax Criterion in Stochastic Recourse Programs. *European Journal of Operation Research*, **165**, 569 – 584.

[48] Rockafellar, R. T., Wets, R. J-B (1998): Variational Analysis. A Series of Comprehensive Studies in Mathematics **317**, Springer Verlag, New York.

[49] Rogosinsky, W. W. (1958): Moments of Non-Negative Mass. *Proc. Roy. Soc. London*, Ser. A, **245**, 1 – 27.

[50] Römisch, W. (2003): Stability of Stochastic Programming Problems. Handbooks in Operation Research and Management Science: Stochastic Programming, Ruszczyński, A., Shapiro, A. (eds.), Elsevier Science B. V., **10**, 486 – 554.

[51] Rustem, B., Howe, M. (2002): Algorithms for Worst-Case Design and Applications to Risk Management. Princeton University Press.

[52] Salinetti, G. (1994): Stability of Bayesian Decisions. *Journal of Statistical Planning and Inference* **40**, 313 – 329.

[53] Sengupta, J. K. (1991): Robust Decisions in Economic Models. *Computers and Operations Research*, **18**, 221 – 232.

[54] Simons, J. (2001): Minimax Theorems. Encyclopedia of Optimization, Ch. A. Floudas, P. M. Pardalos (eds.), **III**, 284 – 289.

[55] Shapiro, A. (2001): On Duality Theory of Conic Linear Problems. Semi-Infinite Programming: Recent Advances, Goberna, M. A., Lopez, M. A. (eds.), Kluwer Academic Publishers, 165 – 165.

[56] Shapiro, A., Kleywegt, A. (2002): Minimax Analysis of Stochastic Problems. *Optimization Methods and Software*, **17**, 523 – 542.

[57] Shapiro, A. (2006): Worst-Case Distribution Analysis of Stochastic Programs. *Mathematical Programming*, Ser. B, **107**, 91 – 96.

[58] Shapiro, A. (2008): Asymptotics of Minimax Stochastic Programs. *Statistics and Probability Letters*, **78**, 150 – 157.

[59] Smith, J. E. (1995): Generalized Chebychev Inequalities: Theory and Applications in Decision Analysis. *Operations Research*, **43** (5), 807 – 825.

[60] Takriti, S., Ahmed, S. (2002): Managing Short-Term Electricity Contracts under Uncertainty: A Minimax Approach. ISyE Technical Report.

[61] Uryasev, S., Rockafellar, R. T. (2002): Conditional Value-at-Risk for General Loss Distributions. *J. of Bank. Fin.*, **26**, 1443 – 1471.

[62] Wonnacott, T. H., Wonnacott, R. J. (1990): Decision Trees. Introductory Statistics for Business and Economics, 4th Edition, John Wiley & Sons, **21**, 639 – 663.

[63] Žáčková, J. (1966): On Minimax Solution of Stochastic Linear Programming Problems. *Čas. Pěst. Mat.*, **91**, 423 – 430.

[64] Zhu, S. S., Fukushima, M. (2005): Worst-Case Conditional Value-at-Risk with Application to Robust Portfolio Management. Technical Report 2005-2006 (July 05, 2005), Department of Applied Mathematics & Physics, Kyoto University.