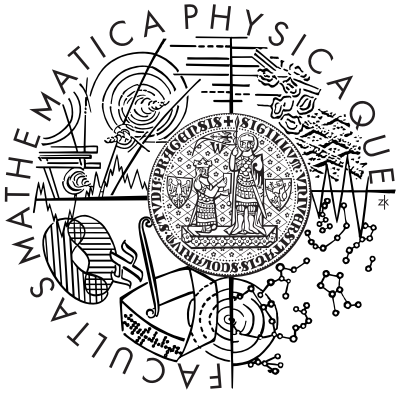
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DISERTAČNÍ PRÁCE



Mgr. Michal Houda

Stabilita a aproximace pro úlohy stochastického programování

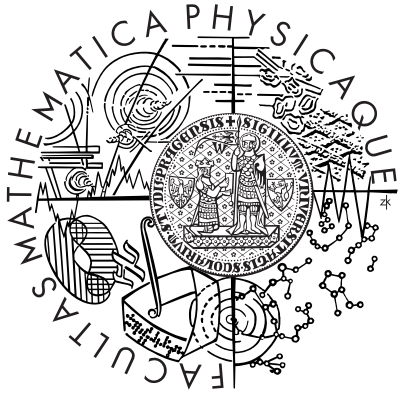
Katedra pravděpodobnosti a matematické statistiky

Školitel: RNDr. Vlasta Kaňková, CSc.
Studijní obor: M5 – Ekonometrie a operační výzkum

Praha, 2009

Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Mgr. Michal Houda

Stability and approximations for stochastic programs

Department of Probability and Mathematical Statistics

Supervisor: RNDr. Vlasta Kaňková, CSc.
Branch: M5 – Econometrics and Operations Research

Prague, 2009

Acknowledgments

I am very much indebted to Dr. Vlasta Kaňková from the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic for her invaluable support and guidance during all the time of my studies at Charles University in Prague. Special thanks belong to many other people at Charles University, Academy of Sciences and University of South Bohemia for their helpful comments and ideas, and to my parents especially for their moral support.

This work was supported by the Czech Science Foundation under the grants No. 402/03/H057, 402/04/1294, 402/05/0115, and 402/07/1113.

Prague, March 23, 2009 Mgr. Michal Houda

Contents

Chapter 1

Introduction

To select among several decisions the best one is an ancient and "obviously very simple" task that people have been solving for centuries. Verily, to find the shortest route to the office, the cheapest TV set in the town, the biggest mountain of the country, the best student in the class, you don't have to be mathematician and you can still answer the questions without trouble. Without the same trouble, we hear our representatives speaking about the best decision for the society the Party has taken or would take if it could; and obviously, the best solutions are "our" solutions... No, simple things are not simple.

The World War II is now widely considered as a starting point of a new mathematical discipline which adopted the name of *mathematical programming*. Mathematical programming is a discipline offering us a toolbox to solve optimization problems, i. e., the problems of selecting the "best" solution. The breakthrough was certainly the invention of the simplex method for solving linear optimization problems by George B. Dantzig [13] – he provided an instrument for automated searching of solutions of such problems and also its applications to solve real-world problems, first of all coming of course from military area.

The methods of linear and nonlinear programming developed later deal with *deterministic* problems. But the *uncertainty* is natural property of all real-world models. In his early paper, Dantzig [14] himself realized very large importance of the uncertainty in optimization. Sometimes, uncertain parameters can be replaced without worry, i.e., approximated with some "good" estimates, but otherwise such simplification is not possible and leads to false conclusions. Thus, in the fifties, works on the new area of science was originating: the *stochastic programming* was born.

Stochastic programming is large scientific area of the theory of optimization in which the problems with random parameters occur. We search for

solutions of such problems, investigate their properties and properties of the problems themselves. All mentioned is important especially in cases in which the randomness cannot be neglected. To begin our investigation at a mathematically correct base let us start with a formal description of the uncertainty in mathematical programming.

1.1 Uncertainty in mathematical programming

As said above, the uncertainty is natural property of mathematical optimization problems. As the starting point, consider the following general optimization problem

$$\text{minimize } c(x;\xi) \text{ subject to } x \in X,\ f(x;\xi) \leq 0 \tag{1.1}$$

where

- $\xi \in \mathbb{R}^s$ is a data element of the problem,
- $x \in X \subset \mathbb{R}^n$ is a decision vector,
- the dimensions n, s, m, and the mappings $c\colon \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ and $f\colon \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ are structural elements of the problem.

All decision-making problems follow such a frame but different methods have to be used due to the nature of structural and data elements of the problems. For our purpose, we extend the description of the framework by the following characterization.

1. The knowledge of the data is insufficient (uncertain). All that is known about the data vector ξ (at least at the very beginning) is that it belongs to a given *uncertainty set* $\Xi \subset \mathbb{R}^s$.
2. The objective function of (1.1) is required to be the best possible given the actual realization (instance) of $\xi \in \Xi$.
3. The constraints of problem (1.1) are required to be satisfied as much as possible given the actual realization of $\xi \in \Xi$.

If the realization of ξ is known and fixed in advance (before the decision has to be taken), and the elements of the problem have suitable properties, standard algorithms of deterministic optimization can be used to solve problem (1.1). For example, if X is a closed convex set, ξ is known and fixed, c and the components f_j of f are convex functions in x for all j, then (1.1)

is a classical case of a convex programming problem and it can be solved by traditional tools of mathematical (convex) programming.

This is rarely the case; in practice, *uncertainty* of the data is typical property and inevitably has to be considered during the course of building the modelling framework. Let us see several examples how the uncertainty occurs.

Example 1.1.

1. The data ξ is not known at the time when a decision has to be taken, and it is observed later in the future. For example, x is a vector of production variables and ξ is an unknown value of future demand or market price.

2. The data ξ cannot be measured or estimated exactly even if it is realized before or at the time when the concrete decision is taken. This is typical for many physical applications where for example material properties are measured with random errors.

3. The data is certain and the optimal solution of the problem can be computed exactly, but such solution cannot be implemented exactly due to physical characteristics of the solution. For example, if the decision is an amount of some commodity to be produced but the production cannot be operated precisely, the actual production is uncertain. Such uncertain implementation can be easily modeled via uncertainty in parameters of the model, not necessarily in the decision vector x itself.

4. The model itself is an approximation of a complicated real-world phenomenon and uncertainty comes directly from the reality and the modeling process.

Dealing with uncertainty is a kind of bread-and-butter problems that classical optimization try to solve. During the history of mathematical programming, several approaches were developed.

1.1.1 Sensitivity analysis – the traditional way

First, the uncertainty can be simply *ignored* at the stage of building the model or at the stage of searching for the optimal solution. The uncertain value of data parameter is replaced by some nominal value, usually by an average, expected value, expert estimate, etc. The accuracy of the optimal solution is (or should be) inspected ex-post by *sensitivity analysis*. This is a traditional way to control the stability of the model but it is limited only

to a single, already generated solution. Moreover, the sensitivity analysis deals only with small perturbations of the nominal data. There are simple examples where the approach described above fails entirely. We would cite famous Peter Kall's linear programming example.

Example 1.2 (Kall [35])**.** Consider a linear optimization problem

$$\text{minimize } x + y \text{ subject to } ax + y \geq 7,\ bx + y \geq 4$$

where a, b are random coefficients with uniform distribution on $[1; 4]$ and $[1/3; 1]$ respectively. After replacing a and b with their expected values and solving the deterministic optimization problem, we obtain the solution which is not feasible with 75% probability!

The classical sensitivity analysis is an adequate method only for simple models or for applications where the uncertainty of the data does not play a crucial role. Otherwise, we are necessarily compelled to use more sophisticated methods.

1.1.2 Parametric programming

The uncertainty can be introduced to the model via a parameter – member of a specified parametric space. The problem (precisely, a family of problems) is then solved over the parametric space. We introduce this technique here only for completeness and do not pursue this direction anymore in the thesis.

1.1.3 Stochastic programming

Stochastic programming handles the uncertainty of stochastic nature. More precisely, we consider ξ in (1.1) to be a random vector and assume that we are able to identify its underlying probability distribution. The idea of stochastic programming approach is to incorporate available information about data through its probability distribution and solve the new model by means of deterministic optimization (the new model was said to be a "deterministic equivalent" in early works on stochastic programming). There are various ways of doing that and there are many papers and books dealing with particular branches of stochastic programming. The stochastic programming community recognizes Dantzig's paper [14] as the initial work in the area; there are also a large number of books devoted to stochastic programming and its applications (Birge and Louveaux [8], Kall and Wallace [36], Prékopa [50], Ruszczyński and Shapiro [64], Wallace and Ziemba [72], and others).

Formally, we assume that ξ is a random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and its probability distribution $\mu \in \mathcal{P}(\Xi)$ is known and fixed.

Here, $\mathcal{P}(\Xi)$ denotes the space of all probability measures defined on the set Ξ. We assume that c and f are $\mathcal{A}$-measurable functions for all $x \in X$ and that μ does not depend on $x \in X$. We have to make a decision before the realization of ξ is taken. Information about probabilistic nature of ξ can be incorporated into the model through various techniques and so we arrive at particular variants of stochastic programming problems. Let us describe some of them by the following examples.

Example 1.3 (Chance-constrained problem)**.** We assume that the objective $c(x;\xi) = c(x)$ does not depend on ξ and that we require the constraints $f(x;\xi) \leq 0$ to be fulfilled with the prescribed level of probability $1-\varepsilon$, where $\varepsilon \in [0;1]$. In fact, non-linear and/or random objective of (1.1) can be moved to the constraint part of the problem considering the following problem

$$\text{minimize } t \text{ subject to } x \in X,\ f(x;\xi) \leq 0,\ c(x;\xi) \leq t,\ t \in \mathbb{R}.$$

(see Section 7.1 for details) The following stochastic programming problem is known as *chance* (or *probability*) *constrained problem* and is one of particular problems treated by the stochastic programming theory and practice:

$$\text{minimize } c'x \text{ subject to } x \in X,\ \mu\{\xi \in \Xi : f(x;\xi) > 0\} \leq \varepsilon.$$

We will treat particular cases of chance-constrained programming in Sections 2.1.2 and 4.3 and in Chapters 6 and 7.

Example 1.4 (Recourse problem)**.** We assume that the constraints set does not depend on ξ; it means, we can somehow incorporate all available constraints into the (deterministic) set X. Possible deviations from the model requirements are compensated by additional recourse action $g(x;\xi)$ which is often the optimal value of a so-called second stage optimization program; its solution depends on the action x from the first stage program and on the realization of the random vector ξ. All possible compensation actions are incorporated to the first stage program through some selected criterium. In many cases, it is done by taking expected value of the first-stage objective $c(x;\xi)$ and the second-stage value $g(x;\xi)$; the problem reads

$$\text{minimize } \mathbb{E}\big[c(x;\xi) + g(x;\xi)\big].$$

Expectation is taken with respect to the probability distribution of ξ and we assume that the functions c, g are measurable and that their expectations exist. This problem is called (stochastic programming) *problem with recourse* (or with *penalization*). Other criteria are sometimes used (var, $\mathbb{P}-$, quantile criteria). We treat recourse problems more deeply in Sections 2.1.1 and 4.2 and partially in Chapter 5.

Example 1.5 (Mixed case)**.** Mixed case where both the compensation g and the constraints f depend on the uncertainty parameter are common in practice. From the theoretical point of view, they are not so important because one random part of the problem can be easily converted to the other, as mentioned in Example 1.3 for the case of chance-constrained problem. We therefore do not pay any special attention to these problems.

1.1.4 Robust programming

Unlike stochastic programming, techniques of *robust programming* are looking for a solution to optimization program (1.1) that satisfies its constraints for all possible realizations of ξ, i. e., that is feasible for any member of the problems belonging to the family (1.1). The problem can be rewritten as

$$\text{minimize } c'x \text{ subject to } x \in X, f(x;\xi) \leq 0 \text{ for all } \xi \in \Xi. \tag{1.2}$$

The symbol Ξ is overloaded here: first, in (1.1), it represents the uncertainty set which the (unknown) parameter ξ is known to belong to, and second, in (1.2), it is the set of parameters for which the constraints must be fulfilled. In fact, this overloading is not too much important: the two sets usually coincide for the reason that information we have about the uncertain parameter is the same as the risk we want to hedge against. Even if it is not the case, the role of the original uncertainty set Ξ is not recognized so important in robust programming problems.

For c, f and X convex, (1.2) is a convex program but it is numerically hard to solve because of an infinite number of constraints. There are several relaxation techniques to deal with this issue, see e. g. Ben-Tal and Nemirovski [5], Bertsimas and Sim [7], and references therein. We describe a so-called "randomized" approach Calafiore and Campi [10] in Chapter 7.

The dissertation thesis is divided into three main parts. The first part is devoted to the stability in stochastic programming problems, involving the necessary theoretical background on probabilistic metrics and the survey of literature. This part ends with Chapter 5 concerning approximations in stochastic programming. Second part numbered as Chapter 6 deals with convexity in chance-constrained problems, especially for the case of weakly dependent constraint rows. Last part of the thesis, Chapter 7, compares stochastic and robust programming as the two disciplines solving one original optimization problem by different means.

Chapter 2

Stochastic programming problems

We introduced in Chapter 1 stochastic programming as a mathematical discipline that could help to solve optimization problems in which the uncertainty is of probabilistic nature. We continue in this chapter by specifying the general form of stochastic programming problems and two main classes of stochastic programming problems that meet this general form and to which we refer later in the thesis. Second part of the chapter is devoted to an introduction to the stability of stochastic programming problems; it will be continued by Chapter 4.

2.1 Classes of stochastic programming problems

In Examples 1.3 and 1.4 we started with two specific cases of stochastic programming problems – chance-constrained problem and recourse (penalization) problem. In the following, we generalize the concept by the following general formulation of stochastic programming problem:

$$\begin{aligned} &\inf_{x\in X} \int_{\Xi} F_0(x;\xi)\ \mu(\mathrm{d}\xi) \\ &\text{subject to} \\ &\int_{\Xi} F_j(x;\xi)\ \mu(\mathrm{d}\xi) \leq 0,\ j = 1,\dots,d \end{aligned} \tag{2.1}$$

In addition to general assumptions from (1.1), we assume that

- $\Xi \subset \mathbb{R}^s$ is a closed set, representing here the support of μ (cf. the note on overloaded symbol Ξ in Section 1.1.4),
- $\mu \in \mathcal{P}(\Xi)$ is a probability measure defined on Ξ,
- X is a closed set not depending on μ,
- $F_j : \mathbb{R}^n \times \Xi \to \bar{\mathbb{R}}$ are for all $j = 0, \ldots, d$ extended real random lower semicontinuous functions.

Recall that a function $F : \mathbb{R}^n \times \Xi \to \bar{\mathbb{R}}$ is *random lower semicontinuous* if its epigraphical mapping

$$\xi \mapsto \operatorname{epi} F(\cdot, \xi) := \big\{(x; r) \in \mathbb{R}^n \times \mathbb{R} : F(x; \xi) \leq r\big\}$$

is closed-valued and measurable. In the literature, random lower semicontinuous functions are also referred as *normal integrands*; a detailed analysis of random lower semicontinuous functions can be found in Rockafellar and Wets [58], Section 14.D.. From the above definition it follows that

1. $F(\cdot; \xi)$ is lower semicontinuous for every fixed $\xi \in \Xi$ (closed-valuedness of $\operatorname{epi} F$ means that for every ξ the epigraph $\operatorname{epi} F(\cdot, \xi)$ is closed subset of $\mathbb{R}^{n+1}$, i.e. $F(\cdot; \xi)$ is lower semicontinuous);
2. $F(x; \cdot)$ is measurable for every fixed $x \in X$.

Note that not every multifunction possessing the last two properties is necessarily random lower semicontinuous; the counterexample is due to the infinity nature of the range of F and the existence of a non-measurable set well known from the measure theory.

In what follows we use the following notation:

$$X(\mu) := \left\{ x \in X : \int_\Xi F_j(x; \xi)\ \mu(\mathrm{d}\xi) \leq 0,\ j = 1, \ldots, d \right\},$$

$$\varphi(\mu) := \inf_{x \in X(\mu)} \int_\Xi F_0(x; \xi)\ \mu(\mathrm{d}\xi),$$

$$\psi(\mu) := \left\{ x \in X(\mu) : \int_\Xi F_0(x; \xi)\ \mu(\mathrm{d}\xi) = \varphi(\mu) \right\}.$$

$X(\mu)$ is the constraint set of (2.1); it depends on the probability distribution μ. $\varphi(\mu)$ is the optimal value and $\psi(\mu)$ is the optimal solution set of the problem (2.1). X (without reference to to μ) is used to describe deterministic constraints, i.e. all constraints that does not depend on the probability measure μ. We will also refer to μ as "original distribution" for the sequel.

Different classes of stochastic programming problems, not only well-known and already presented recourse and chance-constrained programs fall into this general framework. Let nevertheless start again with recourse problems.

2.1.1 Recourse (penalization) problems

Let $d = 0$. Here, the constraint part of the problem does not exhibit any probabilistic nature – the only constraints we have in hand are already included in X. Problem (2.1) then reduces to

$$\inf_{x \in X} \int_{\Xi} F_0(x; \xi) \ \mu(\mathrm{d}\xi). \tag{2.2}$$

This problem covers among others a so called *stochastic programming problem with recourse*. In the usual application framework, the function F_0 is split up into two parts: first part is the first-stage objective function $c(x)$; it does not depend on ξ and gives us the (first stage) decision x. After some decision x is chosen, random variable ξ is realized and some compensation/penalization action is taken evaluated by the cost function $\mathcal{Q}(x; \xi)$. The expected value of these additional costs is taken into consideration when deciding about first-stage decision. The whole problem then reads:

$$\inf_{x \in X} \left(c(x) + \int_{\Xi} \mathcal{Q}(x; \xi) \ \mu(\mathrm{d}\xi) \right) \tag{2.3}$$

where $c \colon \mathbb{R}^n \to \mathbb{R}$, and $\mathcal{Q} \colon \mathbb{R}^n \times \mathbb{R}^s \to \bar{\mathbb{R}}$

For theoretical purposes, only properties of $\mathcal{Q}$ are of special consideration in stochastic programming (the function c is considered not to be of random nature). This cost function is usually given throughout a second-stage optimization problem depending on the realization of ξ as well as on the first-stage decision x. The recourse value function can be read, for example, as

$$\mathcal{Q}(x, \xi) := \inf\{q(y) : y \in Y(x, \xi)\}, \tag{2.4}$$

$Y : \mathbb{R}^n \times \mathbb{R}^s \rightrightarrows \mathbb{R}^m$, $q \colon \mathbb{R}^m \to \bar{\mathbb{R}}$. Of course, properties of the multifunction Y are crucial for properties of second-stage value function $\mathcal{Q}$ and thus for properties of the stochastic program (2.3) as a whole. Most of the literature is concerned with the problem of (fixed) linear recourse:

$$\begin{aligned} &\min \left(c'x + \mathbb{E}\mathcal{Q}(x; \xi) \right) \text{ subject to } x \in X \\ &\text{where} \\ &\mathcal{Q}(x; \xi) = \min\{q(\xi)'y : Wy = b(\xi) - T(\xi)x, y \geq 0\}, \end{aligned} \tag{2.5}$$

$c \in \mathbb{R}^n$ (first-stage costs), $y \in \mathbb{R}^m$ (second-stage solution), and $W \in \mathbb{R}^{\bar{m} \times m}$ (fixed recourse matrix). The vectors q (second-stage objective), b (right-hand side), and matrix T (technology matrix), all of them of appropriate sizes, depend on ξ affine linearly. To have the problem in (2.5) well defined we need the following assumptions to be fulfilled:

(A1) foreach $t \in \mathbb{R}^{\bar{m}}$ there exists $y \in \mathbb{R}^m_+$ such that $Wy = t$ (the recourse is complete);

(A2) there exists $u \in \mathbb{R}^{\bar{m}}$ such that $W'u \leq q$ (dual feasibility);

(A3) $\xi \in \mathcal{P}_2(\Xi)$, i.e. $\int_\Xi \|\xi\|^2 \mu(\mathrm{d}\xi) < +\infty$ (finite second-order moments).

A number of papers and books concern stochastic linear programming problems with recourse; among all we refer to the book Birge and Louveaux [8] where the basic theory and numerical methods for stochastic programming problems with expectation is treated.

2.1.2 Chance-constrained programming problems

Consider $d = 1$ and set $F_0(x;\xi) = c(x)$, where $c\colon \mathbb{R}^n \to \mathbb{R}$. In this case, there is no stochastic element in the objective function of the problem. Let further $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^s$ be a multifunction representing stochastic constraints of the problem, and define F_1 from (2.1) as

$$F_1(x;\xi) = p - \chi_{H(x)}(\xi),$$

with $p \in [0;1]$ being a prescribed (sufficiently high) probability level, and $\chi_A(\cdot)$ the characteristic function of a set A (equal to 1 if $\xi \in A$, 0 otherwise). The easy transformation of (2.1) (having in mind that integral of the characteristic function of a set A is equal to the probability of the set) leads to the formulation of so-called *(joint) chance-constrained* (or *probabilistic*) *programming problem:*

$$\min c(x) \text{ subject to } x \in X,\ \mu(H(x)) \geq p \tag{2.6}$$

Example 2.1. To get a particular formulation of Example 1.3 on page 10 in our form we have to set up

$$\begin{aligned} c(x) &= c'x, \\ H(x) &= \{\xi \in \Xi; f(x;\xi) \leq 0\}, \text{ and} \\ p &= 1 - \varepsilon. \end{aligned}$$

If this example, no matter if the function f is considered one- or multidimensional. Assume for now the constraint function f in Example 2.1 to be one-dimensional and only one component of ξ applies in the definition of f.

If we now have a number of $d = s$ of such functions (each of them with its own component of ξ), define for all $i = 1, \ldots, s$

$$\begin{aligned} &H_i(x) = \{\xi \in \Xi \colon f_i(x; \xi_i) \leq 0\}, \\ &F_i(x; \xi) = p_i - \chi_{H_i(x)}(\xi), \text{ with} \\ &p_i \in [0; 1]. \end{aligned}$$

Using these definitions and assuming stochastically independent components of ξ, (2.1) results in the formulation of stochastic programming problem with *individual probabilistic constraints*:

$$\min c(x) \text{ subject to } x \in X,\ \mu(f_i(x; \xi_i) \leq 0) \geq p_i \text{ for } i = 1, \ldots, s. \qquad (2.7)$$

From the probabilistic point of view, each constraint in the last model is examined separately, it means that the model ignores (possible) stochastic dependence between the constraints. This could be considered as an issue and an advantage at the same time for practical problems because numerical treatment of individual probabilistic constraints is much more simple than of joint ones. We will focus on another issue of individual/joint chance-constrained problems in Chapter 6.

As for recourse problems, many papers and books are devoted to chance-constrained programming problems. The starting reference could be the book of András Prékopa [50] with thorough analysis of many aspects of chance-constrained programming. We mention some additional literature in the upcoming chapters devoted to stability and structural dependence of chance-constrained problems.

2.1.3 Other formulations

More complicated models falling into the frame of (2.1) and involving probabilistic terms can be formulated; we now will mention only a few of them as illustration of possible extensions to the area.

1. **Block-joint probabilistic constraints.** Consider $d \geq 1$ and multi-dimensional functions f_i in (2.7); we obtain a problem that is composed of several blocks of joint probabilistic constraints; each block itself is considered independent on other blocks.

2. **Probability maximization.** Consider a problem

$$\max \mu(c(x; \xi) \leq b) \text{ subject to } x \in X.$$

The above problem is sometimes referred as problem of maximizing the probability: we require the objective c to attain some prescribed (sufficiently small) treshold b with as high probability as possible. To rewrite the problem in terms of (2.1) we set $d = 0$ and

$$H(x) = \{\xi \in \Xi \colon c(x;\xi) \leq b\},$$
$$F_0(x;\xi) = 1 - \chi_{H(x)}(\xi).$$

3. **Hybrid linear problem.** Probabilistic constraints and recourse can be combined together, as in the following model example with simple linear recourse with probabilistic constraints:

$$\min \left(c'x + \sum_{j=1}^{m} q_j \mathbb{E}[\xi_j - T_j x]_+ \right)$$

subject to

$$x \in X,\ \mu\{Tx \geq \xi\} \geq p,$$

where T_j are the rows of matrix T, q_i are penalization constants, $[\cdot]_+$ stands for positive part function. The second-stage optimization problem is a classical linear problem with probabilistic constraints, i.e. we require the constraints to be satisfied with sufficiently high probability, nevertheless possible constraint violation is in addition penalized by unit costs q_j.

2.2 Stability of stochastic programming problems

Almost all methods of stochastic programming rely on the fact that the distribution μ is known and fixed in advance. In practice, this is not the case, the true distribution is rarely completely known:

- often, the true distribution is not actually known but we have some historical data at our disposition, hence some statistical *estimate* of μ can (and should) be used;
- even if the distribution is completely known, its properties could forbid to solve the problem efficiently and some simple (e.g. discrete) *approximation* has to be used;
- the most unpleasant is the case when we dispose no information at all (nor the historical data) and one has to use some kind of *expert estimate*.

All these cases (and possibly other of yours) have common mathematical interpretation: in (2.1), we replace the "original" distribution μ by the "new" one (estimated, approximated) which we will denote ν for the remaining parts of the thesis.

The key question in the theory of stability of stochastic programming is: how the optimal value and optimal solution of (2.1) *change* if we replace the original measure μ by its estimate (approximation, ...) ν. We can split up such large question into several areas of interest:

1. **Qualitative stability:** we are looking for suitable qualitative properties of the optimal value and the optimal solution set, such as *(semi-) continuity*, *differentiability*, or *persistence* (existence of a solution on a neighbourhood).

2. **Quantitative stability:** the matter of thing is to quantify previous properties; for example to find a convenient upper bounds for

$$|\varphi(\mu) - \varphi(\nu)| \quad \text{and} \quad \mathbb{D}(\psi(\mu), \psi(\nu))$$

where $\mathbb{D}$ is suitably selected set-distance, e.g. Pompeiu-Hausdorff distance, ρ-distances, etc. (for thorough analysis see Rockafellar and Wets [58], Chapter 4). More formally we look for functions m_φ, m_ψ having convenient properties (e.g. Lipschitz, Hölder continuity, etc.) such that

$$|\varphi(\mu) - \varphi(\nu)| \leq m_\varphi(d(\mu, \nu))$$
$$\mathbb{D}(\psi(\mu), \psi(\nu)) \leq m_\psi(d(\mu, \nu))$$

where d is some function measuring "difference" between distributions μ and ν.

Natural question arises when reading previous sentences: how to measure (quantify) the "difference" between two distributions; this question was also investigated by several authors and is to be exploited in the next chapter.

Chapter 3

Probability metrics

3.1 Minimal and ideal probability metric

3.1.1 Minimal information metric

Stating quantitative stability of optimal values and/or optimal solution sets in the form

$$\begin{aligned} |\varphi(\mu) - \varphi(\nu)| &\leq m_\varphi(\mu, \nu) \\ \mathbb{D}(\psi(\mu), \psi(\nu)) &\leq m_\psi(\mu, \nu) \end{aligned} \tag{3.1}$$

the key is to find an explicit form of functionals m_φ, m_ψ for μ and ν being not too far each other. Hence, our aim is to introduce a notion of "distance between probability distributions". Denote $\mathcal{P}(\Xi)$ the space of all probability measures defined on $\Xi \subset \mathbb{R}^s$ and $\mathcal{P}_\mathcal{F} \subset \mathcal{P}(\Xi)$ its (somehow determined) subset; the choice of $\mathcal{P}_\mathcal{F}$ would depend on a set $\mathcal{F}$ of nonlinear real functions defined on Ξ on so that the subclass $\mathcal{P}_\mathcal{F}$ has suitable properties (existence of finite moments is such a common property).

Definition 3.1 (Rachev [53], Definition 2.3.1)**.** A mapping $d_\mathcal{F} : \Xi \times \Xi \to [0; +\infty]$ is said to be a *probability pseudometric* (also *semimetric*) on the space $\mathcal{P}_\mathcal{F}(\Xi)$ if it has the following properties:

(P1) (identity) $\mu = \nu \Rightarrow d_\mathcal{F}(\mu, \nu) = 0$

(P2) (symmetry) $d_\mathcal{F}(\mu, \nu) = d_\mathcal{F}(\nu, \mu)$

(P3) (triangle inequality) $d_\mathcal{F}(\mu, \nu) \leq d_\mathcal{F}(\mu, \pi) + d_\mathcal{F}(\pi, \nu)$ for all $\pi \in \mathcal{P}_\mathcal{F}(\Xi)$

It is said to be a *probability metric* if the identity property has a form of equivalence:

(P1*) (identity) $\mu = \nu \Leftrightarrow d_{\mathcal{F}}(\mu, \nu) = 0$.

Natural choice of a distance for the general stochastic programming problem (2.1) is a distance that uniformly compares expectations on a variety of nonlinear function $\mathcal{F}$. Such distance is defined for $\mu, \nu \in \mathcal{P}_{\mathcal{F}}$ by

$$d_{\mathcal{F}}(\mu, \nu) := \sup_{F \in \mathcal{F}} \Big| \int_{\Xi} F(\xi)\mu(\mathrm{d}\xi) - \int_{\Xi} F(\xi)\nu(\mathrm{d}\xi) \Big| \tag{3.2}$$

and known as the *distance having ζ-structure* or *Zolotarev pseudometric* on the space $\mathcal{P}_{\mathcal{F}}$ of probability measures. The concept of such a distance was introduced by Zolotarev [75], the question is examined also in Zolotarev [76] and Rachev [53].

For the distance (3.2), the properties (P1), (P2), and (P3) are always satisfied, hence $d_{\mathcal{F}}$ is always a pseudometric. If it has finite value and the class $\mathcal{F}$ is rich enough to ensure that $d_{\mathcal{F}}(\mu, \nu) = 0$ implies $\mu = \nu$, then it is a metric on $\mathcal{P}_{\mathcal{F}}$.

The choice of functions that will belong to $\mathcal{F}$, so forth defining the class $\mathcal{P}_{\mathcal{F}}$ and the (pseudo-) metric $d_{\mathcal{F}}$, would closely depend on structural properties of the problem. Smaller the number of members of $\mathcal{F}$, smaller the distance is and hence more stringent bounds could be found. The "smallest possible" $\mathcal{F}$ contains only the functions $F_j(x, \cdot)$ – see the upcoming Definition 3.2.

For reasons of localization (this will be treated later in more depth), we consider an open set $U \subset \mathbb{R}^n$ (without any additional specification here) and define the class of functions

$$\mathcal{F}_U := \big\{ F_j(x; \cdot) : x \in X \cap \operatorname{cl} U,\ j = 0, \dots, d \big\}$$

and the space of probability metrics corresponding to $\mathcal{F}_U$ as

$$\begin{aligned} \mathcal{P}_{\mathcal{F}_U}(\Xi) := \Bigg\{ & \nu \in \mathcal{P}(\Xi) : \\ & -\infty < \int_{\Xi} \inf_{x \in X \cap r\mathbb{B}} F_j(x; \xi) \mathrm{d}\nu(\xi) \ \text{ for each } r > 0, \\ & \sup_{x \in X \cap \operatorname{cl} U} \int_{\Xi} F_j(x; \xi) \mathrm{d}\nu(\xi) < +\infty \ \text{ for each } j = 0, \dots, d \Bigg\}. \end{aligned}$$

The conditions on ν in $\mathcal{P}_{\mathcal{F}_U}(\Xi)$ are such that expectations in (3.2) exist and are finite so the distance is well defined.

Definition 3.2. The distance

$$d_{\mathcal{F}_U}(\mu,\nu) := \sup_{\substack{j=0,\dots,d\\ x\in X\cap \mathrm{cl}\, U}} \Big| \int_\Xi F_j(x;\xi)(\mu)(\mathrm{d}\xi) - \int_\Xi F_j(x;\xi)(\nu)(\mathrm{d}\xi)\Big|$$

is called *minimal information (m. i.) metric*.

The distance $d_{\mathcal{F}_U}$ is a distance having ζ-structure. It is the "minimal distance" for the stability of the model (2.1) in sense that it takes information only from specific properties of the functions F_j. General stability theorems deal with this probability distance.

3.1.2 Ideal probability metric

Minimal information metric is difficult to handle. In order to find more friendly results we look for another distance having ζ-structure by enlarging the class $\mathcal{F}$ and thus bounding $d_{\mathcal{F}_U}$ from above. Then, all bounds of type (3.1) found originally with $d_{\mathcal{F}_U}$ remain still valid for any new distance from $\mathcal{P}_{\mathcal{F}}$. Sometimes, it is also necessary to reduce the class $\mathcal{P}_{\mathcal{F}}$ of acceptable probability measures to ensure existence and finiteness of expectations. The common strategy of constructing an enlarged class $\mathcal{F}_{id}$ is such that $\mathcal{F}_{id}$ contains all the functions $CF_j(x;\cdot)$ for all $x \in X\cap \mathrm{cl}\, U$ and some $C > 0$ and share some typical analytical properties of F_j. We call such probability metric as an *ideal probability metric* for the predefined class of stochastic programming problems.

3.2 Probability metrics for recourse functionals

3.2.1 Wasserstein metric

Functionals of stochastic programs with linear recourse as well as some other types of recourse functionals exhibit a local Lipschitz property which we are going to characterize now.

Definition 3.3. Consider a function $F : \Xi \to \mathbb{R}$. We define 1-*Lipschitz constant* of F

$$L_1(F) := \inf\big\{L : |F(\xi) - F(\tilde\xi)| \le L\|\xi - \tilde\xi\| \quad \forall \xi,\tilde\xi \in \Xi\big\},$$

where $\|\cdot\|$ is the usual Euclidian norm. We define the class of 1-*Lipschitz continuous functions* as

$$\mathcal{F}_1 := \{F{:}\Xi \to \mathbb{R} : L_1(F) \le 1\}.$$

Denote

$$\mathcal{P}_1(\Xi) = \mathcal{P}_{\mathcal{F}_1}(\Xi) := \left\{ \nu \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\| \nu(\mathrm{d}\xi) < +\infty \right\}$$

the class of probability measures having finite the first absolute moment.

Definition 3.4. Let $\mu, \nu \in \mathcal{P}_1(\Xi)$. We define so-called *Wasserstein metric* by

$$W_1(\mu, \nu) := d_{\mathcal{F}_1} = \sup_{F \in \mathcal{F}_1} \Big| \int_\Xi F(\xi)\mu(\mathrm{d}\xi) - \int_\Xi F(\xi)\nu(\mathrm{d}\xi) \Big|.$$

Wasserstein metric is a distance that has ζ-structure. It is used in many areas of probability and statistics where it is also known as *Kantorovich* or *Mallows metric*.

Application of the Kantorovich-Rubinstein theorem gives us a dual representation of W_1. The exact formula states

$$W_1(\mu, \nu) := \inf \left\{ \int_{\Xi \times \Xi} \|\xi - \tilde{\xi}\| \ \eta(\mathrm{d}\xi \times \mathrm{d}\tilde{\xi}) : \eta \in D(\mu, \nu) \right\},$$

where $D(\mu, \nu)$ is the set of all probability measures (from $\mathcal{P}(\Xi \times \Xi)$) for which μ and ν are marginal distributions.

Calculation of the value of Wasserstein metric is rather troublesome because it involves computation of a multidimensional integral. The situation is much more simpler in one-dimensional space: let $\mu, \nu \in \mathcal{P}(\Xi)$ be one-dimensional distributions with distribution functions F, G respectively. Then the Wasserstein metric simplifies to (see Vallander [70])

$$W_1(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| \mathrm{d}t$$

or, using duality,

$$W_1(\mu, \nu) = \int_{-\infty}^{+\infty} |F(t) - G(t)| \mathrm{d}t.$$

The calculation is demonstrated in Figure 3.1.

The one-dimensional Wasserstein metric is plausible metric concerning the easiness of calculation – it is a simple one-dimensional integral. But there are special cases where obtained results are useless as in the following example.

Example 3.1 (Houda [32])**.** Consider two one-dimensional distributions μ, ν with distribution functions F, G defined by

$$F(t) := \begin{cases} \dfrac{\varepsilon}{1-t} & \text{for } t < 0 \\ \text{arbitrary} & \text{for } t \geq 0 \end{cases}$$

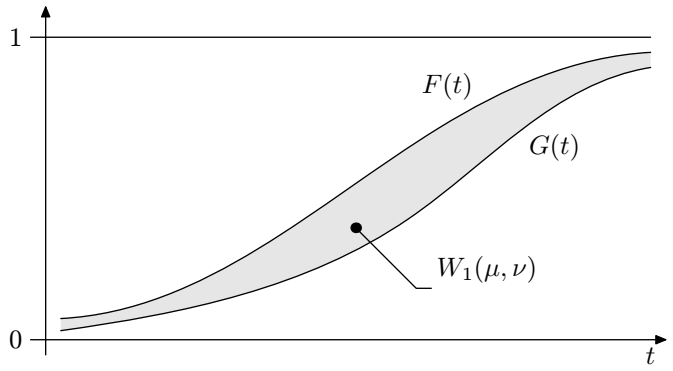


Figure 3.1: One-dimensional Wasserstein metric

with some $\varepsilon \in (0;1)$, and

$$G(t) := \begin{cases} 0 & \text{for } t < 0 \\ F(t) & \text{for } t \geq 0 \end{cases}$$

The distribution μ is an example of a probability distribution having *heavy tails*; although the two distribtions appear close each to other, the value of the Wasserstein metric between μ and ν is infinite. The problem is illustrated on Figure 3.2.

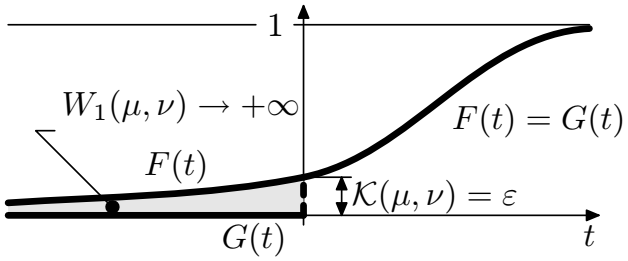


Figure 3.2: Wasserstein metric and heavy tails

In the literature there also appears a notion of higher-order Wasserstein metrics: consider $p \geq 1$, and denote

$$\mathcal{P}_p(\Xi) = \mathcal{P}_{\mathcal{F}_p}(\Xi) := \left\{ \nu \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^p \nu(\mathrm{d}\xi) < +\infty \right\}$$

a class of distributions having finite at least p-th moment. For $\mu, \nu \in \mathcal{P}_p(\Xi)$, we define (starting with the dual representation) the *p-Wasserstein metric*

by

$$W_p(\mu,\nu) := \left[\inf\left\{\int_{\Xi\times\Xi} \|\xi-\tilde{\xi}\|^p \,\eta(\mathrm{d}\xi\times\mathrm{d}\tilde{\xi}) : \eta\in D(\mu,\nu)\right\}\right]^{1/p},$$

where $D(\mu,\nu)$ is again the set of all probability measures (of $\mathcal{P}(\Xi\times\Xi)$), for which μ and ν are marginals. The p-order Wasserstein metric was historically used for stability of recourse models with random both recourse costs and technology matrix with right-hand side. Nowadays, the actual importance of these metrics diminishes in favour of the Fortet-Mourier metrics (see below). Let us end with saying that $(\mathcal{P}_p(\Xi), W_p)$ is metric space and W_p metrizes the topology of weak convergence on $\mathcal{P}_p(\Xi)$.

3.2.2 Fortet-Mourier metric

General recourse functionals can grow faster than linearly. To generalize the notion of Lipschitz continuity, we define for $p \geq 1$ a *p-Lipschitz constant* of the function $F\colon \Xi \to \mathbb{R}$ by

$$L_p(F) := \inf\left\{L : |F(\xi)-F(\tilde{\xi})| \leq L\|\xi-\tilde{\xi}\| \max\{1, \|\xi\|^{p-1}, \|\tilde{\xi}\|^{p-1}\}\ \forall \xi,\tilde{\xi}\in\Xi\right\}$$

and the class of *p-Lipschitz continuous functions* by

$$\mathcal{F}_p := \{F\colon \Xi\to\mathbb{R} : L_p(F)\leq 1\}.$$

Following the similar construction of the Wasserstein metric, we denote

$$\mathcal{P}_p(\Xi) = \mathcal{P}_{\mathcal{F}_p}(\Xi) := \left\{\nu\in\mathcal{P}(\Xi) : \int_\Xi \|\xi\|^p \nu(\mathrm{d}\xi) < +\infty\right\}$$

the class of probability measures having finite the p-th absolute moment, let $\mu,\nu\in\mathcal{P}(\Xi)$, and define the metric

$$\zeta_p(\mu,\nu) := d_{\mathcal{F}_p}(\mu,\nu) = \sup_{F\in\mathcal{F}_p}\left|\int_\Xi F(\xi)\mu(\mathrm{d}\xi) - \int_\Xi F(\xi)\nu(\mathrm{d}\xi)\right|.$$

The last one is called the *p-th order Fortet-Mourier metric* (see Fortet and Mourier [23], Rachev [53]). We immediately see that $W_1(\mu,\nu)$ coincides with $\zeta_1(\mu,\nu)$.

When constructing a scenario tree within a discretization process, one discusses the problem how to minimize the chosen metric. In the case of the p-Fortet–Mourier metric with $p>1$, the problem could be transformed to the minimization of the 1-Fortet–Mourier metric, i. e. the 1-Wasserstein metric. For details about these techniques see Pflug [47].

3.2.3 Bounded Lipschitz β-metric

In early papers concerning the stability with respect to the probability measures, a more complicated distance was considered, but which also holds the ζ-structure. It deals with bounded Lipschitz functions defining "bounded Lipschitz" constant by

$$\|f\|_{\mathrm{BL}} := \sup_{\xi\in\Xi}\|f(\xi)\| + \sup_{\substack{\xi,\tilde{\xi}\in\Xi\\ \xi\neq\tilde{\xi}}} \frac{|f(\xi)-f(\tilde{\xi})|}{\|\xi-\tilde{\xi}\|}$$

This leads to the class $\mathcal{F}_{\mathrm{BL}} := \{f\colon \Xi \to \mathbb{R} : \|f\|_{\mathrm{BL}} \leq 1\}$ and the metrics $\beta := d_{\mathcal{F}_{\mathrm{BL}}}$ called the *bounded Lipschitz β-metrics*. Bounds found using this metric has later been replaced by better results with the Wasserstein metric. This metric also metrizes the topology of weak convergence.

3.3 Probability metrics for chance constraints

3.3.1 $\mathcal{B}$-discrepancies

Recall here the definition of the chance-constrained model to see how an "ideal" probability metric is constructed for chance-constrained functionals; the model is

$$\min c(x) \text{ subject to } x \in X, \mu(H_j(x)) \geq p_j \text{ for all } j \tag{3.3}$$

(c.f. (2.6)). Functionals F_j are defined here by $F_j(x;\xi) = p_j - \chi_{H_j(x)}(\xi)$. Consider a class $\mathcal{B} = \mathcal{B}(\Xi)$ of Borel subsets of Ξ such that it contains all multifunctions H_j (i.e., constraint multifunctions). We define the class $\mathcal{F}$ by

$$\mathcal{F} = \mathcal{F}_{\mathcal{B}} := \{\chi_B : B \in \mathcal{B}\}$$

Due to the nature of indicator function χ the corresponding class of convenient probability measures is the whole set $\mathcal{P}(\Xi)$.

Definition 3.5. The distance

$$\alpha_B = d_{\mathcal{F}_{\mathcal{B}}} = \sup_{B\in\mathcal{B}}|\mu(B)-\nu(B)|$$

is called α-(pseudo-)metric or, alternatively, $\mathcal{B}$-discrepancy.

α-metric is not, in fact, the only (pseudo-) metric but the whole class of pseudometrics. A specific distance can be defined according to the nature and properties of $H_j(x)$.

Example 3.2 (polyhedral discrepancy)**.** Consider the following linear extenension of Example 2.1 where

$$H(x) := \{\xi \in \Xi : T(\xi)x \geq h(\xi)\}$$

with X and Ξ being polyhedra, h and T depending on ξ affine linearly. The sets $H(x)$ are polyhedra with a uniformly bounded number of faces, so considering the class

$$\mathcal{B}(\Xi) = \mathcal{B}_{\mathrm{ph},k}(\Xi) := \{B \subset \Xi : B \text{ is a polyhedra with at most } k \text{ faces}\}$$

we arrive at so-called *polyhedral discrepancy*

$$\alpha_{\mathrm{ph},k}(\mu,\nu) = d_{\mathcal{F}_{\mathcal{B}_{\mathrm{ph},k}(\Xi)}}(\mu,\nu) = \sup_{B\in\mathcal{B}\mathrm{ph},k} |\mu(B) - \nu(B)|.$$

In mixed-integer two-stage stochastic programming problems (we allow for integer decisions in both stages) with rational recourse matrix, recourse value function is discontinuous but can be partitioned into countable many regions of continuity. These regions splits into disjoints Borel subset whose closures are polyhedra and polyhedral discrepancies apply again. We do not further pursue this direction in the thesis; for details see Rachev and Römisch [54], Section 3.2.

Example 3.3 (half-space discrepancy)**.** Consider the chance-constrained problem with a single constraint:

$$\min c(x) \text{ subject to } x \in X,$$
$$\mu\left(\sum_{j=1}^{m} \xi_j x_j \geq \xi_{m+1}\right) \geq p$$

where ξ has a multivariate normal distribution with nonsingular covariance matrix. The set $H(x)$ defined as

$$H(x) := \{\xi \in \mathbb{R}^{m+1} : \sum_{j=1}^{m} \xi_j x_j \geq \xi_{m+1}\}$$

is closed half-space in $\mathbb{R}^{m+1}$; hence, the appropriate class (with $\Xi = \mathbb{R}^{m+1}$) would be

$$\mathcal{B}(\Xi) = \mathcal{B}_{\mathrm{h}}(\Xi) := \{B \subset \Xi : B \text{ is a closed half-space of } \Xi\}$$

and we arrive at so-called *half-space discrepancy*

$$\alpha_{\mathrm{h}}(\mu,\nu) = d_{\mathcal{F}_{\mathcal{B}_{\mathrm{h}}(\Xi)}}(\mu,\nu) = \sup_{B\in\mathcal{B}_{\mathrm{h}}(\Xi)} |\mu(B) - \nu(B)|.$$

Special attention is paid here to another case of $\mathcal{B}$-discrepancy distance: Kolmogorov metric. Due to wide usage of Kolmogorov metric we devote a special subsection to it even it does not differ from other discrepancies and follows the general scheme.

3.3.2 Kolmogorov metric

Consider a chance-constrained problem where the right-hand side only is random:

$$\min c'x \text{ subject to } x \in X, \mu\{\xi \in \Xi; g(x) \geq \xi\} \geq p$$

where $\Xi \subset \mathbb{R}^s$, $g\colon \mathbb{R}^n \to \mathbb{R}^s$, $p \in [0;1]$. The corresponding set-valued mapping determining the discrepancy distance is defined by

$$H(x) := \{\xi \in \Xi : \xi \leq g(x)\};$$

that are s-dimensional "semi-closed intervals"; we form a class $\mathcal{B}$ by exactly such intervals:

$$\mathcal{B}(\Xi) = \mathcal{B}_{\mathrm{K}}(\Xi) := \{(-\infty;\xi], \xi \in \Xi\}.$$

The corresponding distance is called *Kolmogorov* or *Kolmogorov-Smirnov metric*

$$\mathcal{K}(\mu,\nu) = \alpha_{\mathrm{K}}(\mu,\nu) = d_{\mathcal{F}_{\mathcal{B}_{\mathrm{K}}(\Xi)}}(\mu,\nu) = \sup_{B \in \mathcal{B}_{\mathrm{K}}(\Xi)} |\mu(B) - \nu(B)|.$$

The last expression is simply the difference of probability distributions functions F, G corresponding to μ, ν

$$\mathcal{K}(\mu,\nu) := \sup_{z \in \Xi} |F(z) - G(z)|.$$

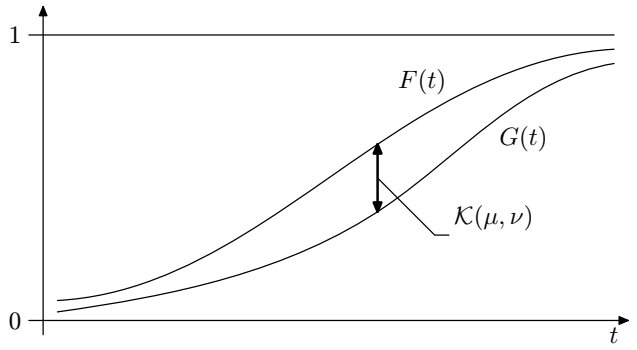


Figure 3.3: One-dimensional Kolmogorov metric

A considerable advantage of this metrics is the computational simplicity when it is applied to workaday optimization problems. Hence, it is often used in practice instead more sophisticated distances even it is not always an ideal metric. See Figure 3.3 for a better idea.

On the other hand, the Kolmogorov metric is not good choice when we deal, for example, with approximation of unknown mass points of a discrete distribution. This issue is illustrated in Figure 3.4; the Wasserstein metric works well in this case. For another case we refer to Dupačová and Römisch [20] where the authors measure changes in optimal value arising from including an additional scenario to the problem; the Kolmogorov metrics does not reflect (under the cited conditions) the position of the new scenario, compared to expected behaviour of the suitable metric.

In the case of Example 3.1 where the Wasserstein metric was found not to work, the Kolmogorov metric works well; it deals without problem even for distributions having heavy tails.

There are many results concerning stability of stochastic programming problems with respect to Kolmogorov metric. Section 4.3 references some pivotal literature devoted to this question.

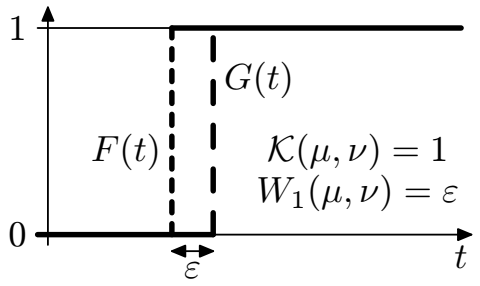


Figure 3.4: Kolmogorov metric and discrete distribtions

Chapter 4

Stability in stochastic programming problems

4.1 General stability theorem

We introduced in Section 2.1 a general form of the stochastic programming problem. Recall that definition: the mathematical formulation of the problem is to find

$$\begin{aligned} &\inf_{x\in X} \int_{\Xi} F_0(x;\xi)\ \mu(\mathrm{d}\xi) \\ \text{subject to} \quad & \\ &\int_{\Xi} F_j(x;\xi)\ \mu(\mathrm{d}\xi) \leq 0,\ j = 1,\ldots,d \end{aligned} \tag{4.1}$$

where $\xi \in \Xi$ is a random vector defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\Xi \subset \mathbb{R}^s$ is a closed support of $\mu \in \mathcal{P}(\Xi)$, a probability distribution of ξ, $X \subset \mathbb{R}^n$ is a closed constraint set not depending on μ, and $F_j : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ are random lower semicontinuous functions.

Before stating a general stability theorem we need some additional notions to be defined.

4.1.1 Complete local minimizing set

Stochastic programming problems often lack convexity properties. Even if the original problem is convex, perturbations and approximations can lead to optimization problems that are not convex. In order to find well-behaved properties of the optimal solutions, we need some concept of localization

that overcome this difficulty. To achieve this, consider a nonempty open set $U \subset \mathbb{R}^n$ (cf. the definition of the ideal probability metric in Section 3.1.1), $\nu \in \mathcal{P}_{\mathcal{F}_U}(\Xi)$, and denote

$$\begin{aligned} X_U(\nu) &:= \left\{ x \in X \cap \operatorname{cl} U : \int_\Xi F_j(x;\xi)\ \nu(\mathrm{d}\xi) \le 0,\ j = 1,\dots,d \right\} \\ \varphi_U(\nu) &:= \inf_{x \in X_U(\nu)} \int_\Xi F_0(x;\xi)\nu(\mathrm{d}\xi) \\ \psi_U(\nu) &:= \left\{ x \in X_U(\nu) : \int_\Xi F_0(x;\xi)\nu(\mathrm{d}\xi) = \varphi_U(\nu) \right\}. \end{aligned} \tag{4.2}$$

$X_U(\nu)$ is the set of feasible solutions (with respect to integral constraints) belonging to U, $\varphi_U(\nu)$ is the optimal value, and $\psi_U(\nu)$ the optimal solution set of the corresponding (localized) optimization problem.

Definition 4.1. A nonempty set $S \subset \mathbb{R}^n$ is called *complete local minimizing (CLM) set* of the optimization problem (4.2) relative to U if $S = \psi_U(\nu) \subset U$.

Local optimizing sets are examples of CLM sets (U will be chosen such a way that the local optimizing set is contained in U). Nonempty global optimizing sets are also CLM sets. The idea under the notion of CLM set is to restrict our consideration to a neighbourhood of a selected optimum instead to examine all possible feasible solutions of perturbed problem. Now only such a way localized solutions are of our interest. The idea of CLM sets was introduced in Robinson [56] and Klatte [44].

4.1.2 Semicontinuity of set-valued mappings

Dealing with stability of optimal solution sets, we need a notion of continuity of set-valued functions (multifunctions). We will use Berge's terminology; for the following definition see e.g. book Bank et al. [3] where also necessary and sufficient conditions for the property are stated.

Definition 4.2. Let $(\mathcal{P}, d)$ be a (semi-) metric space. We say that a set-valued mapping $S\colon (\mathcal{P}, d) \rightrightarrows \mathbb{R}^n$ is (*Berge*) *upper semicontinuous* at the point $\mu \in \mathcal{P}$ if for every open set $\mathcal{O} \subset \mathbb{R}^n$ with $S(\mu) \subset \mathcal{O}$ it holds that $S(\nu) \subset \mathcal{O}$ for each $\nu \in \mathcal{P}$ with sufficiently small value of $d(\mu, \nu)$.

Remark 4.1. Rockafellar and Wets [58] introduced a similar notion of *outer semicontinuous* set-valued mappings. This notion is not in fact identical: the multifunction S from Definition 4.2 is outer semicontinuous at a point μ if it is upper semicontinuous at this point, $S(\mu)$ is closed set, and S is locally

bounded at μ, i.e., for some (sufficiently small) neighbourhood V of μ the set $S(V)$ is bounded.

In our settings, it will be the set-valued mapping ψ_U which will be examined for upper semicontinuity with respect to a particular, selected probability (semi-) metric. It is closed-valued function, and local boundedness will be introduced through theorem assumptions. So, in all theorems of this thesis, we can replace the notion of upper semicontinuity by Rockafellar's outer semicontinuous alternative.

4.1.3 Growth conditions

Quantitative stability properties of the optimal solution set cannot be stated without knowing behaviour of the objective function near the optimal solution set: we need some growth condition of the objective function. Here we adopt the settings of Römisch [59] inspired by Rockafellar and Wets [58], Section 7.J:

Definition 4.3. For $\tau \geq 0$ we define the *growth function* by

$$\psi_\mu(\tau) := \min\left\{\int_\Xi F_0(x;\xi)\mu(\mathrm{d}\xi) - \varphi(\mu) : \mathrm{dist}(x;\psi(\mu)) \geq \tau,\ x \in X_U(\mu)\right\}$$

and the *associated growth function* by

$$\Psi_\mu(\eta) := \eta + \psi_\mu^{-1}(2\eta), \quad \eta \in \mathbb{R}_+$$

where

$$\psi_\mu^{-1}(2\eta) := \sup\{\tau \geq 0 : \psi_\mu(\tau) \leq 2\eta\}.$$

Both these functions describe the behaviour of the objective function: $\psi_\mu(\tau)$ is non-decreasing lower-semicontinuous function, vanishing at 0, and it can be considered as a "minimal growth" in sense that for non-optimal solutions it describes distance of the value of the objective function to the optimal value of the problem. $\Psi_\mu(\eta)$ is increasing lower-semicontinuous function also vanishing at 0 that will give us the perturbation estimate for optimal solutions (higher the growth, lower the associated function and better estimates we will obtain).

In the definition of associated growth function we are not restricted by the "minimal" growth function. In fact, we can use any convenient continuous increasing function (at least for small values of τ) in order to guarantee sufficient growth of the objective function. The results derived for Ψ can be simply redeclared for these special growth functions.

Example 4.1 (Linear growth)**.** If $\psi_\mu(\tau) \geq \gamma\tau$ for some $\gamma > 0$ and each *small* $\tau \geq 0$ we say the objective function has *linear growth* at its optimal solution set. In this case the associated function Ψ is bounded above by

$$\Psi_\mu(\eta) \leq \eta + \frac{2\eta}{\gamma} = C\eta \tag{4.3}$$

with $C = (\gamma + 2)/\gamma$ and sufficiently small $\eta \geq 0$.

Example 4.2 (k-th order growth)**.** Similarly, if $\psi_\mu(\tau) \geq \gamma\tau^k$ for $k \geq 1$ (the other remains unchanged) we say that the objective function satisfies *k-th order growth condition* at the solution set. Here, the upper bound to the associated function Ψ is modified to

$$\Psi_\mu(\eta) \leq \eta + \left(\frac{2\eta}{\gamma}\right)^{1/k} \leq C\eta^{1/k} \tag{4.4}$$

with some constant $C > 0$. The linear growth is included in this definition via $k = 1$, the case of $k = 2$ is called *quadratic growth condition.*

Strong convexity condition. In early papers devoted to stability of optimal solutions in stochastic programming, the strong convexity property of the objective function is widely used.

Definition 4.4. Given some subset $U \subset \mathbb{R}^n$, the function $c\colon \mathbb{R}^n \to \mathbb{R}$ is said to be *strongly convex* if there exists some $\kappa > 0$ such that for all $x_1, x_2 \in U$ and all $\lambda \in [0; 1]$,

$$c(\lambda x_1 + (1-\lambda)x_2) \leq \lambda c(x_1) + (1-\lambda)c(x_2) - \frac{1}{2}\kappa\lambda(1-\lambda)\|x_1 - x_2\|^2$$

This is actually very strong condition; for example if applied near the optimal solution set of the problem (2.2) it ensures the solution to be unique. Further, the strong convexity condition is sufficient for the quadratic growth condition. We refer to Schultz [65] to more details about conditions ensuring strong convexity property.

4.1.4 Metric regularity

In problems with probabilistic constraints the stability property for constraint sets $X(\mu)$ are needed. To state it we will use a notion of the special metric regularity condition as introduced in Rockafellar and Wets [58], Section 9.G.

Denote

$$X_y(\mu) := \left\{ x \in X : \int_\Xi F_j(x;\xi)\mu(\mathrm{d}\xi) \le y_j,\ j = 1,\dots,d \right\}$$
$$X_x^{-1}(\mu) := \left\{ y \in \mathbb{R}^d : x \in X_y(\mu) \right\}$$

Definition 4.5. We say that the mapping $x \mapsto X_x^{-1}(\mu)$ is *metrically regular* at some pair $(\bar{x}, 0) \in \mathbb{R}^n \times \mathbb{R}^d$, $\bar{x} \in X_0(\mu)$ if there are $a \ge 0$ and $\varepsilon > 0$ such that for all $x \in X$ and $y \in \mathbb{R}^d$ with $\|x - \bar{x}\| \le \varepsilon$ and $\max\limits_{j=1,\dots,d} |y_j| \le \varepsilon$ it holds that

$$\mathrm{dist}(x, X_y(\mu)) \le a \max_{j=1,\dots,d} \max \left\{ 0, \int_\Xi F_j(x;\xi)\mu(\mathrm{d}\xi) - y_j \right\}. \qquad (4.5)$$

Metric regularity condition relates the distance of some (unfeasible) solution x from the perturbed set X_y, and the "value" y of this perturbation. The relation is considered of Lipschitz nature again, giving us a base for useful estimation for the optimal value and optimal solution set changes. See Figure 4.1 for better idea about the notion. Note that for $y = 0$ we obtain $X_0(\mu) = X(\mu)$, the constraint set of the unperturbed problem.

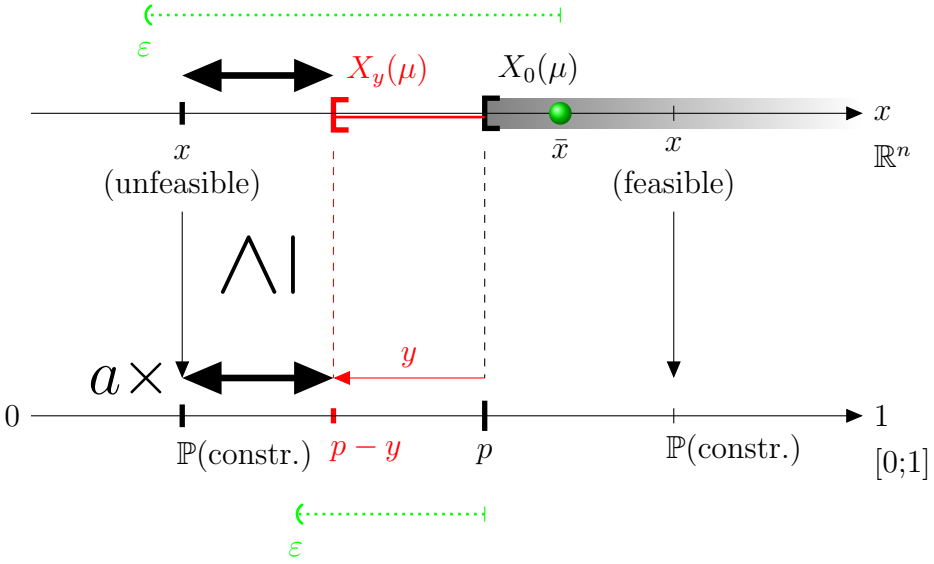


Figure 4.1: Concept of metric regularity condition.

Remark 4.2. In earlier works, the notion of "metric regularity" of $X_x(\mu)$ is known under the term of "pseudo-Lipschitz continuity" of the multifunction $y \mapsto X_y(\mu)$ (see Rockafellar [57]). Now we prefer the contemporary notion of metric regularity, as proposed by the same author later; see Rockafellar and Wets [58] and the bibliographical commentary for Chapter 9 of this book.

4.1.5 General stability theorem

We are now ready to formulate a general stability theorem for the problem (4.1); the theorem was introduced in Henrion and Römisch [28].

Theorem 4.1 (Henrion and Römisch [28])**.** *Consider the stochastic programming problem* (4.1) *with its general assumptions. Let* $P \in \mathcal{P}_{\mathcal{F}_U}$ *and assume that*

- $\psi(\mu) \neq \emptyset$, U *is open* bounded *neighbourhood of* $\psi(\mu)$*;*
- *if* $d \geq 1$, $x \mapsto \int_\Xi F_0(x;\xi)\mu(\mathrm{d}\xi)$ *is Lipschitz on* $X \cap \operatorname{cl} U$*;*
- $x \mapsto X_x^{-1}(\mu)$ *is metrically regular at each pair* $(\bar{x};0)$, $\bar{x} \in \psi(\mu)$.

Then

- $\psi_U \colon (\mathcal{P}_{\mathcal{F}_U}, d_{\mathcal{F}_U}) \rightrightarrows \mathbb{R}^m$ *is (Berge) upper semicontinuous at* μ
- *there exist* $\delta > 0$, $L > 0$, $\hat{L} \geq 1$ *such that for* $\nu \in \mathcal{P}_{\mathcal{F}_U}(\Xi)$, $d_{\mathcal{F}_U}(\mu,\nu) < \delta$*, we have that*
  - $|\varphi_U(\mu) - \varphi_U(\nu)| \leq L d_{\mathcal{F}_U}(\mu,\nu)$
  - $\psi_U(\nu) \neq \emptyset$ *and it is a CLM set wrt.* U*, i. e.,* $\psi_U(\nu) \subset U$
  - $\psi_U(\nu) \subset \psi(\mu) + \Psi_\mu\left(\hat{L} d_{\mathcal{F}_U}(\mu,\nu)\right)\mathbb{B}$

This general theorem states several stability results with respect to the minimal information metric. First, persistence and qualitative property of continuity (upper semicontinuity at the unperturbed problem) for localized optimal solution sets. Second, quantitative stability property of Lipschitz type for optimal values. And third, quantitative stability property for optimal solution sets relying on the growth properties of the objective function.

As mentioned earlier, the minimal information metric is not easy to handle. So different estimates for upper bounds with more suitable metrics for specific structures of stochastic programming problem are derived in the literature; some of them we will mention in the two following sections. Also the condition of special metric regularity is crucial in this setting and sometimes hard to check. We refer to Rockafellar and Wets [58] where some sufficient conditions for metric regularity are given.

4.2 Stability of stochastic programming problems with recourse

Early stability results for recourse problems was obtained by Römisch and Wakolbinger [63] with the β-metrics applied. The results was later extended in Römisch and Schultz [60]. Among others, the Hölder continuity of optimal values was proved. We illustrate these results by citing two theorems where the complete linear recourse is assumed. Recall thus the problem introduced by (2.5)

$$\begin{aligned} &\min\left(c'x + \mathbb{E}Q(x;\xi)\right) \text{ subject to } x \in X \\ \text{where} \quad &\\ &\mathcal{Q}(x;\xi) = \min\{q(\xi)'y : Wy = b(\xi) - T(\xi)x, y \geq 0\}, \end{aligned} \tag{4.6}$$

and $X \subset \mathbb{R}^n$ beeing nonempty closed polyhedron, $c \in \mathbb{R}^n$ and $W \in \mathbb{R}^{\bar{m}\times m}$ having constant elements, q, b, and T depending on ξ affine linearly.

For $p \in [1;+\infty)$ and $K > 0$, denote $\mathcal{P}_{p,K}(\Xi)$ a space of probability measures with uniformly bounded p-th moment:

$$\mathcal{P}_{p,K}(\Xi) := \left\{\nu \in \mathcal{P}(\Xi); \int_\Xi \|\xi\|^p \mu(\mathrm{d}\xi) \leq K\right\}.$$

Theorem 4.2 (Römisch and Schultz [60])**.** *Consider the problem* (4.6) *with assumption (A1), (A2) from Section 2.1.1 fulfilled. Let $p > 1$, $K > 0$ and fix $\mu \in \mathcal{P}_{2p,K}(\Xi)$. Assume further that $\psi(\mu)$ is nonempty and bounded. Then*

1. *ψ is upper semicontinuous at μ with respect to the metric space $(\mathcal{P}_{2p,K}(\Xi), \beta)$, and*
2. *there exist constants $\delta > 0$, $L > 0$ such that for any $\nu \in \mathcal{P}_{2p,K}(\Xi)$ fulfilling $\beta(\mu,\nu) < \delta$ we have $\psi(\nu) \neq \emptyset$ and*

$$|\varphi(\mu) - \varphi(\nu)| \leq L\beta(\mu,\nu)^{1-\frac{1}{p}}.$$

Later, Römisch and Schultz [61] have weakened significantly the assumptions and got better – Lipschitz estimates with the Wasserstein metric. The case of the complete linear recourse is studied in Römisch and Schultz [62], where the next theorem is proved. Note, if we do not take into account the assumption (A3), we would get a similar proposition with $\mathcal{P}_2(\Xi)$ and the metrics W_2.

Theorem 4.3. *Consider the problem* (4.6) *with (A1), (A2) and assume (A3): q or (a, T) are non-stochastic. Let $\mu \in \mathcal{P}_1(\Xi)$ and assume that $\psi(\mu)$ is nonempty and bounded set. Then ψ is upper semicontinuous at μ with respect to $(\mathcal{P}_1(\Xi), W_1)$, and there exist constants $\delta > 0$, $L > 0$ such that for any $\nu \in \mathcal{P}_1(\Xi)$ for which $W_1(\mu, \nu) < \delta$ we have $\psi(\nu) \neq \emptyset$ and*

$$|\varphi(\mu) - \varphi(\nu)| \leq L\, W_1(\mu, \nu).$$

The following theorem is a possible non-linear modification based on the previous theorem: we slightly modified its assumptions and the proof. Further, the continuity properties and compactness ensure existence and boundedness of optimal solution of the original (unperturbed) problem and, at the same time, we can leave out the localization property.

Theorem 4.4 (Houda [31],Houda [32])**.** *Consider the program* (2.2)*, where the following assumption are fulfilled:*

- *X is a compact set, and the function F_0 is uniformly continuous on $\mathbb{R}^n \times \mathbb{R}^s$;*
- *$F_0(x; \cdot)$ is Lipschitz continuous for all $x \in X$ with constant L not depending on x.*

Then ψ is (Berge) upper semicontinuous at μ with respect to $(\mathcal{P}_1(\Xi), W_1)$, and for any $\nu \in \mathcal{P}_1(\Xi)$ we have that

$$\psi(\nu) \neq \emptyset \text{ and}$$
$$|\varphi(\mu) - \varphi(\nu)| \leq L\, W_1(\mu, \nu).$$

Proof. 1. Denote $h_\mu(x) := \int_\Xi F_0(x, \xi)\mu(\mathrm{d}\xi)$; h_μ is continuous on a compact set X, thus $\psi(\nu)$ is nonempty set for all $(\nu \in \mathcal{P}_1)$.

2. Let $\nu \in \mathcal{P}_1$, $x_\mu \in \psi(\mu)$ and $x_\nu \in \psi(\nu)$. The two inequalities are valid:

$$\varphi(\mu) \leq h_\mu(x_\nu) \leq \varphi(\nu) + |h_\mu(x_\nu) - h_\nu(x_\nu)| \quad \text{(a)}$$
$$\varphi(\nu) \leq h_\nu(x_\mu) \leq \varphi(\mu) + |h_\mu(x_\mu) - h_\nu(x_\mu)| \quad \text{(b)}$$

The left inequality in (a) is due to the fact that $\varphi(\mu)$ is the optimal value of the problem, the right one to the equality $h_\nu(x_\nu) = \varphi(\nu)$ and a trivial fact that $a + b \leq |a + b|$. Arguments for (b) are similar.

3. From (a) and (b) we get

$$-|h_\mu(x_\mu) - h_\nu(x_\mu)| \leq \varphi(\mu) - \varphi(\nu) \leq |h_\mu(x_\nu) - h_\nu(x_\nu)|$$

and so

$$|\varphi(\mu) - \varphi(\nu)| \leq \sup_{x \in X_0} |h_\mu(x) - h_\nu(x)|.$$

4. The function $F_0(x,\xi)$ is Lipschitz in ξ thus for $x \in X$ one has

$$\begin{aligned}\left|\int_\Xi F_0(x,\xi)\mu(\mathrm{d}\xi) - \int_\Xi F_0(x,\xi)\nu(\mathrm{d}\xi)\right| \\ \leq \int_{\Xi\times\Xi} |F_0(x,\xi) - F_0(x,\xi')|\,\eta(\mathrm{d}\xi\times\mathrm{d}\xi') \\ \leq L\int_{\Xi\times\Xi} \|\xi-\xi'\|\eta(\mathrm{d}\xi\times\mathrm{d}\xi')\end{aligned}$$

for arbitrary $\eta \in D(\mu,\nu)$. So,

$$|h_\mu(x) - h_\nu(x)| \leq L \inf_{\eta\in D(\mu,\nu)} \int_{\Xi\times\Xi} \|\xi-\xi'\|\eta(\mathrm{d}\xi\times\mathrm{d}\xi') = LW_1(\mu,\nu)$$

and finally

$$|\varphi(\mu) - \varphi(\nu)| \leq \sup_{x\in X}|h_\mu(x,\xi) - h_\nu(x)| \leq L\,W_1(\mu,\nu).$$

5. The continuity of $\varphi(\cdot)$ at μ follows from the previous point. Furthemore, the continuity of $h_\mu(\cdot)$ follows from the theorem continuity assumptions. The set X is compact, so Proposition 4.2.1 from Bank et al. [3] leads to the upper semi-continuity of the mapping $\psi(\cdot)$ at μ. This completes the proof.

□

The Wasserstein metric is an ideal metric for recourse problems where F_0 is Lipschitz continuous function. Apparently, not all recourse functionals used are Lipschitz continuous. We still have a possibility to elevate order of Lipschitz continuity and get estimates with respect to the Fortet-Mourier metric (see e.g. Dupačová and Römisch [20], Dupačová et al. [19], Rachev and Römisch [54]). In Pflug [47], the method to find optimal approximations (scenarios) via minimization of the Wasserstein metric are proposed; in the same paper it is shown that the result obtained via mimimization of the Fortet-Mourier metric of higher order are equivalent to that one with Wasserstein metric. We will also refer the result of Theorem 4.4 with slightly different context in Chapter 5 concerning empirical distributions and processes.

4.2.1 Quantitative stability of optimal solutions

Quantitative stability of optimal solution sets can be achieved by direct application of Theorem 4.1. For example, Theorem 4.3 can be extended directly by

Theorem 4.5 (Römisch [59], Corolally 25)**.** *Consider the assumptions of Theorem 4.3 to be fulfilled. Then*

$$\emptyset \neq \psi_(\nu) \subset \psi(\mu) + \Psi_\mu(L\,W(\mu,\nu)) \tag{4.7}$$

where $L > 0$ and $\nu \in \mathcal{P}_1(\Xi)$ are the same as in Theorem 4.3 and Ψ_μ is associated growth function as defined in Definition 4.3.

Due to the definition of the growth function, the upper bound in (4.7) is the best possible. Stringent assumptions as the strong convexity condition were also used earlier in the literature; generally, such bounds are not the best but could have better computation aspects.

Theorem 4.6 (Houda [31], Kaňková and Houda [42])**.** *Let assumptions of Theorem 4.4 be fulfilled. Let X be convex and for all $\xi \in \Xi$ let $F_0(\cdot;\xi)$ be strongly convex on $\mathbb{R}^n$ with parameter $\kappa > 0$. Then*

$$\|\psi(\mu) - \psi(\nu)\|^2 \leq \frac{8}{\kappa} L\,W(\mu,\nu) \tag{4.8}$$

Proof. Denote $h_\mu(x) := \int_\Xi F_0(x,\xi)\mu(\mathrm{d}\xi)$; h_μ is strongly convex with parameter κ. Denote further $\hat{x}_\mu := \operatorname{argmin}_{x\in X} h_\mu(x)$; it is a single vector due to the strong convexity property again.

As $h_\mu(x)$ is strongly convex function on compact convex X, the following inequality is valid for all $x \in X$ (see Kaňková and Lachout [43], Lemma 2, proved in Kaňková [38], Lemma 6 for strongly concave functions):

$$\|x - \hat{x}_\mu\|^2 \leq \frac{4}{\kappa}(h_\mu(x) - h_\mu(\hat{x}_\mu))$$

In particular,

$$\begin{aligned}
\|\hat{x}_\nu - \hat{x}_\mu\|^2 \leq \\
&\leq \frac{4}{\kappa}\big(h_\mu(\hat{x}_\nu) - h_\mu(\hat{x}_\mu)\big) \leq \\
&\leq \frac{4}{\kappa}\big|h_\mu(\hat{x}_\mu) - h_\nu(\hat{x}_\nu)\big| + \frac{4}{\kappa}\big|h_\nu(\hat{x}_\nu) - h_\mu(\hat{x}_\nu)\big| \leq \\
&\leq \frac{8}{\kappa}\big|\varphi(\mu) - \varphi(\nu)\big| \leq \\
&\leq \frac{8}{\rho} L\,W_1(\mu,\nu)
\end{aligned}$$

which yields the assertion of the theorem. □

Many other results concerning stability of optimal solution sets was found in the literature, among all let us cite Dupačová [18], Römisch and Schultz [62], Shapiro [66], Artstein and Wets [2], Schultz [65], Kaňková [41], Dentcheva [17]; see also Römisch [59] for references.

4.3 Stability of chance-constrained problems

Stability results for chance constrained programs are direct corollaries of Theorem 4.1 where appropriate $\mathcal{B}$-discrepancies are applied. The crucial assumption is the metric regularity; specific conditions implying the metric regularity are examined in context of chance-constrained programming by several authors. For example, consider a (possibly) nonlinear chance-constrained program with random right-hand side only:

$$\min c(x) \text{ subject to } x \in X,\ \mu\{\xi \in \Xi : g(x) \geq \xi\} \geq p \tag{4.9}$$

where $g\colon \mathbb{R}^n \to \mathbb{R}^s$, $c\colon \mathbb{R}^n \to \mathbb{R}$, and $p \in [0;1]$. X and Ξ is assumed to be as in the general stochastic programming model. The constraints of the problem are formed by semi-closed intervals $(-\infty;\xi]$ (c.f. Chapter 3) and thus, one is led to use the Kolmogorov metric as an "ideal" probability metric. In the following theorem due to Henrion [26] (see also Henrion and Römisch [29]), the r-concave measure is assumed. We return in detail to this property in Chapter 6 where also r-concavity will be formally defined.

Theorem 4.7 (Henrion [26])**.** *In* (4.9)*, assume*

- *c is convex, X is closed and convex;*
- *g have concave components;*
- *$\mu \in \mathcal{P}(\Xi)$ is r-concave for some $r < 0$;*

In addition, assume that

- *$\psi(\mu)$ is nonempty and bounded;*
- *there exists some $\hat{x} \in X$ such that $F_\mu(g(\hat{x}) > p$ where F_μ is the distribution function of μ (Slater condition);*

Then $\psi(\cdot)$ is (Berge) upper semicontinuous at μ, and there exist constants $L, \delta > 0$ such that for any $\nu \in \mathcal{P}(\Xi)$ with $\mathcal{K}(\mu,\nu) < \delta$ we have that

$$\psi(\nu) \neq \emptyset \text{ and}$$
$$|\varphi(\mu) - \varphi(\nu)| \leq L\mathcal{K}(\mu,\nu).$$

The Slater condition and convexity assumptions ensure the metric regularity condition from the general theorem. Applying a strong convexity condition one can arrive at the stability of optimal solution set that will be of Hölder type (Henrion and Römisch [29]).

Stability of chance-constrained programs was also treated many times in the literature. Selected papers include Römisch and Wakolbinger [63], Römisch and Schultz [61], [54], Kaňková [39], Gröwe [25], Henrion and Römisch [28], [29], Rachev and Römisch [54], and others.

Chapter 5

Approximations in stochastic programming problems with recourse

5.1 Motivation

Recall the result of Theorem 4.3 concerning optimal values of the problems:

$$|\varphi(\mu) - \varphi(\nu)| \leq L\, W_1(\mu, \nu). \tag{5.1}$$

This bound appears under various variations in many other stochastic programming results concerning stability of optimal values and solutions. Our motivation here is the following: we see the stability bound "double-structured". First is the Lipschitz constant, coming from the model structure itself and analytical properties of involved functionals.

Example 5.1. Consider a farmer's problem originating from Birge and Louveaux [8]: a farmer specializes in raising several crops on a land of limited size a. He requires some amount b[1] of each of the crops for his needs (raised himself or bought elsewhere). We assume that the production costs c, selling prices q^+, and purchase prices q^- are known and fixed. Finally consider the yields ξ for each crop to be independent random variables with some predefined distribution. The farmer's goal is to devote an optimal amount x of land to each crop in order to minimize his costs c modified by possible additional purchases y^- of missing crops and reduced by selling the spare crops y^+.

[1] with appropriate index in mathematical formulation (5.2) that we will omit here for simpler understanding

We can formulate the above problem as a two-stage stochastic linear program

$$\min\left(c'x + \mathbb{E}\mathcal{Q}(x;\xi)\right) \text{ subject to } \sum x \geq a$$
$$\text{where} \qquad (5.2)$$
$$\mathcal{Q}(x;\xi) = \min\{q'y : y^- - y^+ \geq b - \xi x, y^+, y^- \geq 0\},$$

with the last inequation formulated for each crop.

The favourable separable form of the second stage leads to rewriting the compensation $\mathcal{Q}$ as a sum of individual compensations $\mathcal{Q}_i$ for each separated crop. Each such individual compensation is linear function of ξ hence Lipschitz continuous with Lipschitz constant that can be found easily. The example is finished in Houda [33] by numerical quantification of the Lipschitz constants in Birge and Louveaux's formulation of this farmer problem.

Second structural part of the bound (5.1) is the randomness expressed by the value of the Wasserstein metric between distributions μ and ν. In addition, this value is in this special case independent from the value of the Lipschitz constant examined in the previous example. This is not a rule: there are many other examples where such separation will not be possible. Nevertheless we will take advantage of this result and will concentrate on specific convergence properties of Wasserstein (and in limited view also Kolmogorov) metrics.

5.2 Empirical distributions and processes

If the probability measure μ, needed for succesful solution of the stochastic optimization problem, is not available, but we have empirical data at our disposition we can use them instead and replace the original distribution with the empirical version. The *empirical distribution function* is a widely used statistical gadget applied both theoretically and practically in many mathematical, engineering, and economic models.

5.2.1 Empirical distribution function

Let $\xi_1, \xi_2, \ldots, \xi_N, \ldots$ be independent random variables with the same probability distribution μ. For notation simplicity, we denote its distribution function by F instead of F_μ. The true distribution μ is, as stated in the introduction, considered as an unknown distribution which is estimated by the sample $\xi_1,\ldots,\xi_N$.

Definition 5.1. The random function

$$F_N(t) = \frac{1}{N} \sum_{i=1}^{N} \chi_{(-\infty;t]}(\xi_i), \quad t \in \mathbb{R} \tag{5.3}$$

is called **empirical distribution function** based on the sample $\xi_1, \ldots, \xi_N$; χ_A is again the characteristic (indicator) function of the set A.

For each realization of the sample, $F_N(t)$ is actually a distribution function; we denote associated probability measure as μ_N and call it *empirical measure*.

By Glivenko-Cantelli theorem and the law of large numbers, a sequence of empirical distribution functions F_N converges almost surely to the distribution function F under general conditions as N goes to infinity. Considering the definition of the Wasserstein and Kolmogorov metric, values of these metric for F and F_N will converge too (see e.g. Shorack and Wellner [68]).

For example, to be able to explore convergence properties of the Wasserstein metric, we require $\mu \in \mathcal{P}_1$ (i.,e., all ξ_i being integrable). In that case

$$\int |t| \mu_N(\mathrm{d}t) \to \int |t| \mu(\mathrm{d}t) \text{ a.s.}$$

and by dominated convergence

$$\int W(\mu_N, \mu) \to 0 \text{ a.s.}$$

The almost sure convergence of Kolmogorov metric is straightforward from its definition.

5.2.2 Empirical processes

One-dimensional Wasserstein metric involving integral computation of difference between distribution functions is well known in statistics. To find its convergence rate at infinity we are interested in behaviour of the associated empirical process defined as follows.

Definition 5.2. The **integrated empirical process** is defined by

$$\sqrt{N}\, W(\mu_N, \mu) = \int_{-\infty}^{+\infty} \sqrt{N}\, |F_N(t) - F(t)|\, \mathrm{d}t. \tag{5.4}$$

The (integrated) empirical process is considered here in a more general sense than usual: we assume here only that $\mu \in \mathcal{P}_1(\Xi)$. If it is the uniform distribution on [0;1], (5.4) describes the (integrated) empirical process in the usual statistical sense. In this case, such process is also known as Mallows statistic. Recall now the notion of the Brownian bridge and other basic notions from the stochastic analysis which we need in what follows.

Definition 5.3. A **stochastic process** $\mathbb{X}$ is the family $\{\mathbb{X}(t) := \mathbb{X}(t,\omega) : t \geq 0\}$ of real-valued random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 5.4.

1. The stochastic process $\mathbb{X}$ is **continuous** if for each $\omega \in \Omega$ the function $\mathbb{X}(\cdot,\omega)$ (also known as the **trajectory** of the process) is continuous (everywhere in t).

2. If for any $0 \leq t_1 < \cdots < t_K$ the joint distribution of $(\mathbb{X}(t_1), \ldots, \mathbb{X}(t_K))$ is normal, then the process is called **Gaussian**.

3. If for any $0 \leq t_0 < t_1 < \cdots < t_K$ the random variables $\mathbb{X}(t_k) - \mathbb{X}(t_{k-1})$, $k = 1, 2, \ldots$ are independent then the process is said to have **independent increments.**.

4. The process $\mathbb{X}$ has **stationary increments** if for each $t, s, h \geq 0$ the distributions of $(\mathbb{X}(t+s+h) - \mathbb{X}(t+h))$ and $(\mathbb{X}(t+s) - \mathbb{X}(t))$ are the same (i.e., the distribution of increments depends only on t and s). In that case it is reasonable to speak about covariance function $\operatorname{cov}(\mathbb{X}(s), \mathbb{X}(t))$ as a function of s and t.

Definition 5.5. A stochastic process $\mathbb{U}$ is called **Brownian bridge** if it is continuous Gaussian process having mean function $\mathbb{E}\mathbb{U}(t) = 0$ and covariance function $\operatorname{cov}(\mathbb{U}(s), \mathbb{U}(t)) = \min(s,t) - st$ where $s, t \in [0;1]$.

The Brownian bridge is conditioned Wiener process given marginal conditions $\mathbb{U}(0) = 0$, $\mathbb{U}(1) = 0$ (the right marginal condition is that differing from the usual Wiener process). Its increments are still normal but not yet independent. The Brownian bridge will be denoted by $\mathbb{U}$ in this thesis.

We find the Brownian bridge as the (weak) limit of the integral (5.4) in case of the uniform distribution on $[0;1]$ (see Section 3.8 of Shorack and Wellner [68]):

$$\int_{-\infty}^{+\infty} \sqrt{N} \left| \frac{1}{N} \sum_{i=1}^{N} \chi_{(-\infty;t]}(\xi_i) - t \right| \mathrm{d}t \to_d \int_{-\infty}^{+\infty} |\mathbb{U}(t)| \mathrm{d}t \tag{5.5}$$

In this case the probability distribution of the limit of (5.5) is also known explicitly (see Section 3.8 of Shorack and Wellner [68]):

$$F\left(\int_{-\infty}^{+\infty}|\mathbb{U}(t)|\mathrm{d}t \le x\right) = \sqrt{\frac{\pi}{2}}\sum_{j=1}^{+\infty}\delta_j^{-3/2}\psi\left(\frac{x}{\delta_j^{-3/2}}\right) \tag{5.6}$$

where

$$\psi(t) = \frac{1}{t^{1/3}}\left(3^{2/3}\mathrm{e}^{-\frac{2}{27}t^2}\mathrm{Ai}(3t)^{-4/3}\right)$$
$$\delta_j = -\frac{a_j'}{2^{1/3}}$$

Ai is the usual Airy function (one of the solutions of the differential equation $y''-xy=0$), and a_j' are j-th zeros of Ai$'$, see e.g. Abramowitz and Stegun [1], Section 10.4. Formula (5.6) was derived by Johnson and Killeen [34] based on previous result by Shepp [67] and Rice [55].

To find a weak limit for distributions that are not uniform is not so easy. The Inverse Theorem with a simple substitution does not suffice because one has to involve the derivative of $F(x)$ into consideration. The condition under which the convergence is valid was found by del Barrio et al. [16].

Proposition 5.1 (del Barrio et al. [16], Theorem 2.1)**.** *The limit theorem*

$$\sqrt{N}(F_N(t)-F(t)) \to_w \mathbb{U}(F(t)) \text{ in } L_1(\mathbb{R})$$

is valid if and only if

$$\int_{-\infty}^{+\infty}\sqrt{F(t)(1-F(t))}\mathrm{d}t < +\infty$$

From this it easily follows that, under the last condition, the following convergence result is valid:

Theorem 5.2. *If*

$$\int_{-\infty}^{+\infty}\sqrt{F(t)(1-F(t))}\mathrm{d}t < +\infty$$

then

$$\int_{-\infty}^{+\infty}\sqrt{N}\left|\frac{1}{N}\sum_{i=1}^{N}\chi_{(-\infty;t]}(\xi_i)-F(t)\right|\mathrm{d}t \to_d \int_{-\infty}^{+\infty}|\mathbb{U}(F(t))|\mathrm{d}t \tag{5.7}$$

Proof. We have used the fact that if some process Y_n converge weakly in $L_1(\mathbb{R})$ to Y, then, in particular, $||Y_n||_{L_1} \to_d ||Y||_{L_1}$ where $||g||_{L_1} = \int_{-\infty}^{\infty} g(t)\mathrm{d}t$ for each non-negative $g \in L_1(\mathbb{R})$). □

The convergence rate of the Kolmogorov metric is a long-standing result originated by Kolmogorov [45] for independent identically distributed samples: if the sample $\xi_1, \xi_2, \ldots, \xi_N$ correspond to a probability measure that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, then

$$\mathbb{P}\Big\{\sqrt{N} \sup_{-\infty<t<\infty} |F_N(t) - F(t)| < x\Big\} \to \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} \text{ for } x > 0 \quad (5.8)$$

(the limit is obviously 0 for $x < 0$). The rate of convergence is exponential and independent on the original distribution. Such kind of result is not known in the case of the Wasserstein metric.

5.3 Comparison: independent and dependent data samples

We will illustrate convergence results from the previous section here for some "representative" distributions. Recall that in the case of the uniform distribution, the distribution function of the limit (5.7) is known, see (5.6). On the other hand, left hand side of (5.7) is easily evaluable for the one-dimensional case and will stand for us as a base for the following numerical results. Furthermore, to extend the results, we focus not only on the independent case but also on the case of weakly dependent data.

5.3.1 Weakly dependent samples

In economical and engineering applications, weakly dependent samples are of very practical interest. It is known that many economical and technical data series exhibit dependence properties although they are considered as independent. The background theory partially permits doing that such a way and we would give some hints that it will be also the case for convergence of empirical distributions and processes.

From the practical point of view, let us leave out assumptions of independence in Section 5.2 and replace them with some weak type of dependence. For simplicity we are not going in depth and choose most simple variant of dependency, M-dependent sequences.

Definition 5.6. Let $\{\xi_t\}_{-\infty}^{+\infty}$ be a random sequence defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let, moreover, $\mathcal{B}(-\infty, a)$ be the σ-algebra generated by $\ldots, \xi_{a-1}, \xi_a$, and $\mathcal{B}(b, +\infty)$ be the σ-algebra generated by $\xi_b, \xi_{b+1}, \ldots$. The sequence is said to be M**-dependent** if $\mathcal{B}(-\infty, a)$ and $\mathcal{B}(b, +\infty)$ are independent for $b - a > M$.

More sophisticated dependency conditions could be found in the literature, see e.g. book of Yoshihara [74]. Here we make only a note that results found for M-dependent sequences also applies to other weak-dependent sequences. For example, it is known (see Yoshihara [74] again) that every stationary ϕ-mixing normal distributed sequence is also M-dependent for some $M \in \mathbb{N}$. A stationary Gaussian random sequence $\{\xi_t\}_{-\infty}^{+\infty}$ is ϕ-mixing if and only if the σ-algebras $\mathcal{B}(-\infty, k)$ and $\mathcal{B}(k + N, +\infty)$ are independent for any N sufficiently large.

Empirical estimates has already been investigated (in the literature) for some types of weakly (say mixing) dependent random sequences (for some details see e.g. Dai et al. [12], Kaňková [39], or Wang and Wang [73]). We can prove by the techniques employed in Kaňková [39] that the case of M-dependent samples can be transformed to the independent one and the asymptotic results remain valid (of course, the convergence will be slower).

Theorem 5.3. *For every natural N there exists $k \in \{0, 1, \ldots\}$ and $r \in \{1, \ldots, M\}$ such that $N = Mk + r$, and*

$$|F_N(t) - F(t)| \leq \sum_{j=1}^{M} \frac{N_j}{N} |F_{N_j}(t) - F(t)|, \quad t \in \mathbb{R} \tag{5.9}$$

where F_{N_j} are empirical distribution functions determined by sequences of N_j independent random variables, and

$$N_j = \begin{cases} k+1 & \text{for } j = 1, \ldots, r \\ k & \text{for } j = r+1, \ldots, M \end{cases}$$

Proof. Let M, N, k, r fulfil the relation $N = Mk + r$ (existence of such partition is straightforward). We split the sample sequence $\{\xi_i\}_{i=1}^{N}$ into M groups of lengths N_j by selecting indexes modulo M:

$$\begin{aligned} j = 1 &: \xi_1, \xi_{M+1}, \ldots \\ j = 2 &: \xi_2, \xi_{M+2}, \ldots \\ &\ldots \\ j = M &: \xi_M, \xi_{2M}, \ldots \end{aligned}$$

Each of these new sequences has independent members due to M-dependence property. Denote F_{N_j} empirical distribution function of each of them; F_N stands for the empirical distribution function of the whole sample despite it is formed from dependent random variables. An easy transformation leads to the following conclusion:

$$\begin{aligned}
&|F_N(t) - F(t)| \\
&\qquad = \left| \frac{1}{N} \sum_{i=1}^{N} \chi_{(-\infty;t]}(\xi_i) - F(t) \right| \\
&\qquad = \left| \sum_{j=1}^{M} \frac{1}{N} \sum_{i=0}^{N_j} \chi_{(-\infty;t]}(\xi_{Mi+j}) - \sum_{j=1}^{M} \frac{N_j}{N} F(t) \right| \\
&\qquad = \left| \sum_{j=1}^{M} \frac{N_j}{N} (F_{N_j}(t) - F(t)) \right|
\end{aligned}$$

as (naturally) $\sum N_j = N$. The assertion of the theorem follows directly. □

5.3.2 Simulation study overview

According to our analysis above, let us now focus on the numerical illustration to results from Section 5.2. To do that we choose four "representative" one-dimensional probability measures: uniform, exponential, normal and Cauchy, and as a weakly dependent sequence we choose a simple 2-dependent sequence – MA(1) process. The exact procedure is as follows.

1. Set N to an appropriate sample length; we select $N = 100$ as a representative for "short" sample, and $N = 1000$ as for "long" sample (having in mind that these sample lengths could be somehow vague in various situations).

2. Generate independent samples $\zeta_1, \zeta_2, \ldots, \zeta_N$ from the given distribution.

3. Make up a new series defined as

$$\xi_k := 0.5\zeta_k + 0.5\zeta_{k-1}.$$

   The theoretical distribution of ξ_k is given by convolution. In particular, it is

- triangular (Simpson's) distribution in the case of uniform samples on $[0; 1]$;
- gamma distribution with the shape and rate parameters both equal to 2 in the case of exponential distribution with parameter $\lambda = 1$;
- normal distribution with zero mean and variance 0.5 in the case of normal distribution $N(0; 1)$;
- Cauchy distribution with actual parameters.

4. The empirical distribution function is calculated for the independent as for the dependent sample series. The Wasserstein and Kolmogorov distances are then calculated with respect to the theoretical measure.

5. The procedure is repeated 200 times to obtain an estimate of the "density" of both distances and the density of their associated empirical processes.

First graph set (Figures 5.1 and 5.2) illustrates almost sure convergence of Kolmogorov and Wasserstein metric. The left column of the set displays densities estimated from independent data, the right column one from dependent data. The dotted line in each of the graphs represents the "short" sample size, the solid line the "long" one.

Remark 5.1. Instead of the true Cauchy distribution a "cut" version is used. Cutting is necessary because Cauchy distribution does not have the first moment and the Wasserstein metric is undefined for it. Furthermore, the cut version also represents a distribution with **heavy tails**.

Our numerical results confirm the expected one: both Wasserstein and Kolmogorov metrics vanish to zero as number of samples grows. It is true even for cut Cauchy distribution. In all the pictures one could even start to compare rates of convergence between dependent and independent samples or across different distributions, although the better base for such comparison will follow in results on empirical processes.

The second graph set (Figures 5.3 and 5.4) is devoted to the convergence of the empirical process $\sqrt{N}\mathcal{K}(\mu_N, \mu)$ (the case of the Kolmogorov metric), and the integrated empirical process (the case of the Wasserstein metric), respectively. The density lines are (Gaussian) kernel estimates of the limiting distribution in each case.

Kolmogorov's empirical process is known to converge to the limit (5.8) regardless the original distribution. This is numerically confirmed by Figure 5.3

(even for Cauchy distribution). For Wasserstein's integrated empirical process, the rate of convergence is stabilizing quickly in the case of the uniform, normal, and exponential distributions. The problem of Cauchy distribution is easily justifiable: although cutting allowed us to use the Wasserstein distance, numerical properties of the distance would to be hardly favourable. This is another illustration of the fact mentioned earlier in Example 3.1.

As for other "standard" distributions, the results with dependent data are seen to differ slightly but one can appreciated the difference as not very important. As members of used MA(1) process are actually weakly dependent, this behaviour is expected again.

An answer to the question about other interesting properties of weakly-dependent sample series in comparison with independent ones (especially for another types of dependency) is not given in this thesis; the question remains still unanswered and open for the future research.

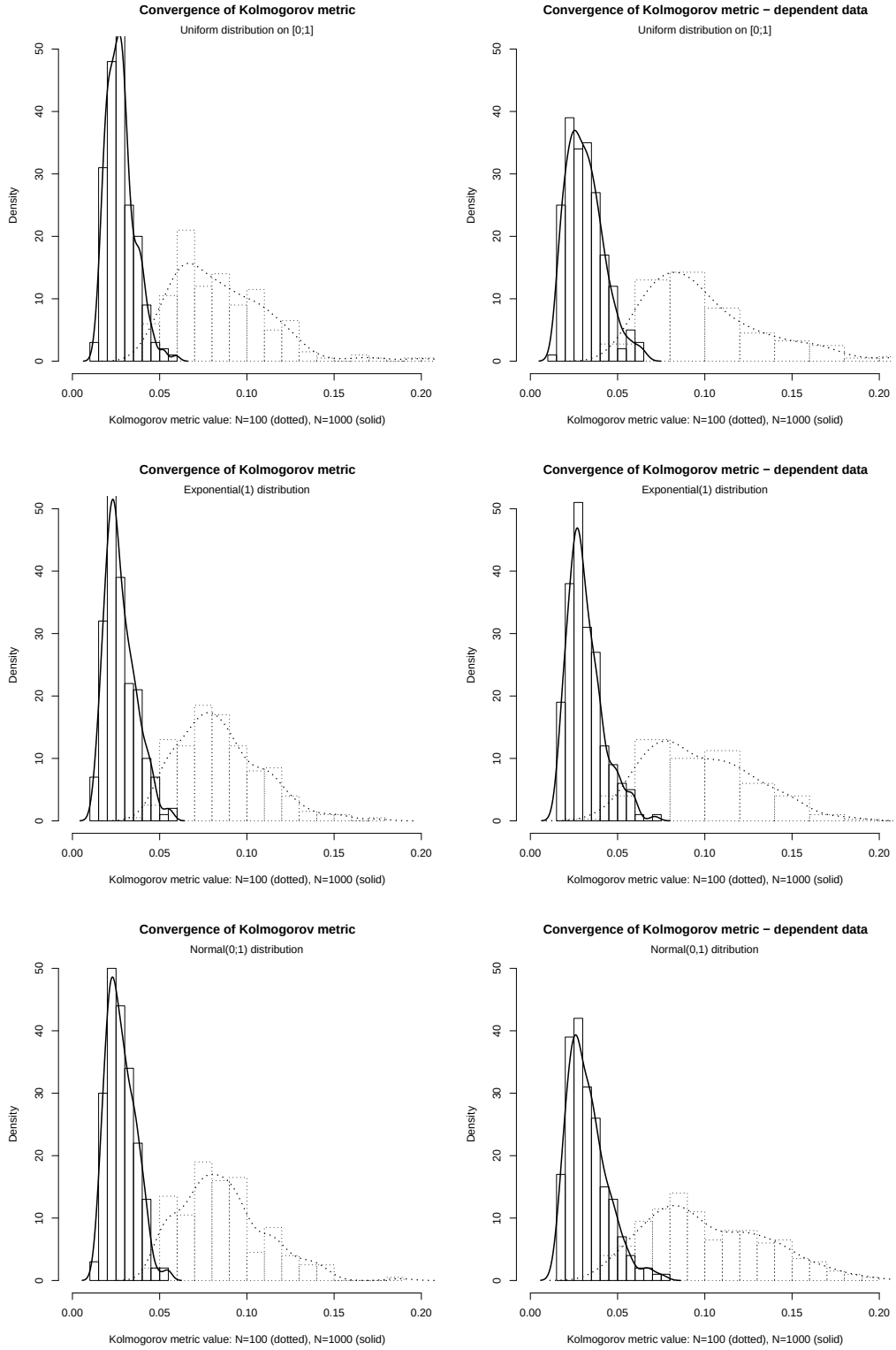


Figure 5.1: Convergence of the Kolmogorov metric

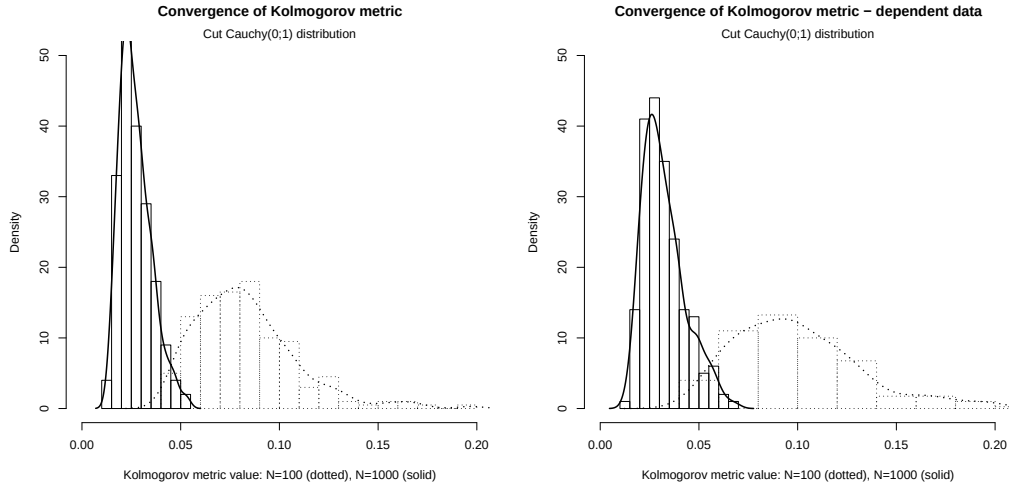


Figure 5.1: Convergence of the Kolmogorov metric (cont'd)

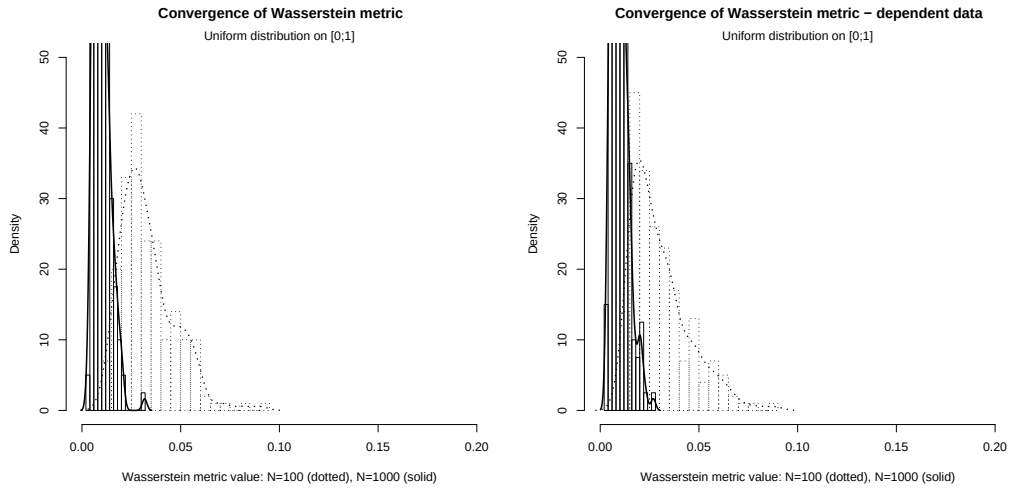


Figure 5.2: Convergence of the Wasserstein metric

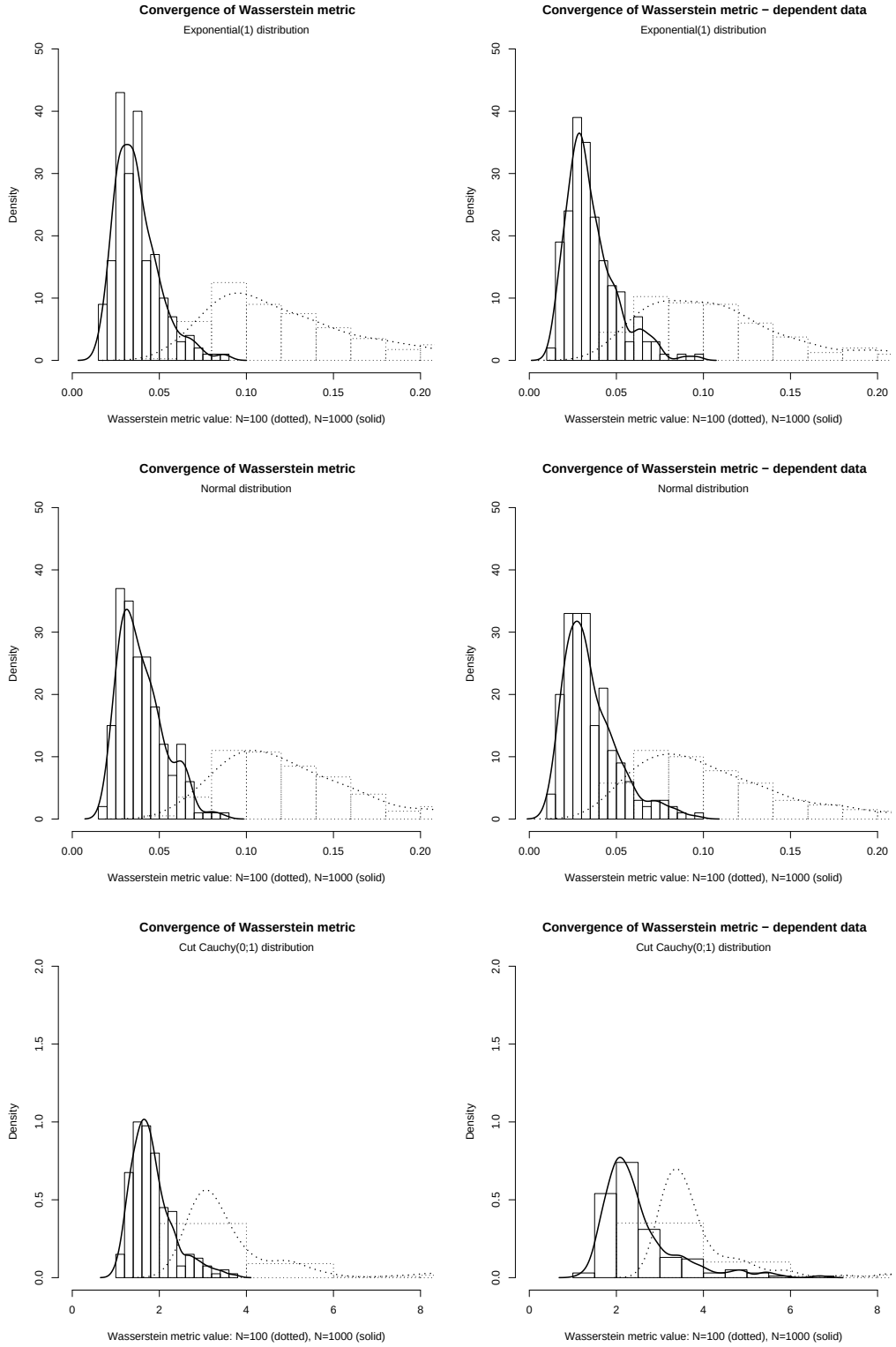


Figure 5.2: Convergence of the Wasserstein metric (cont'd)

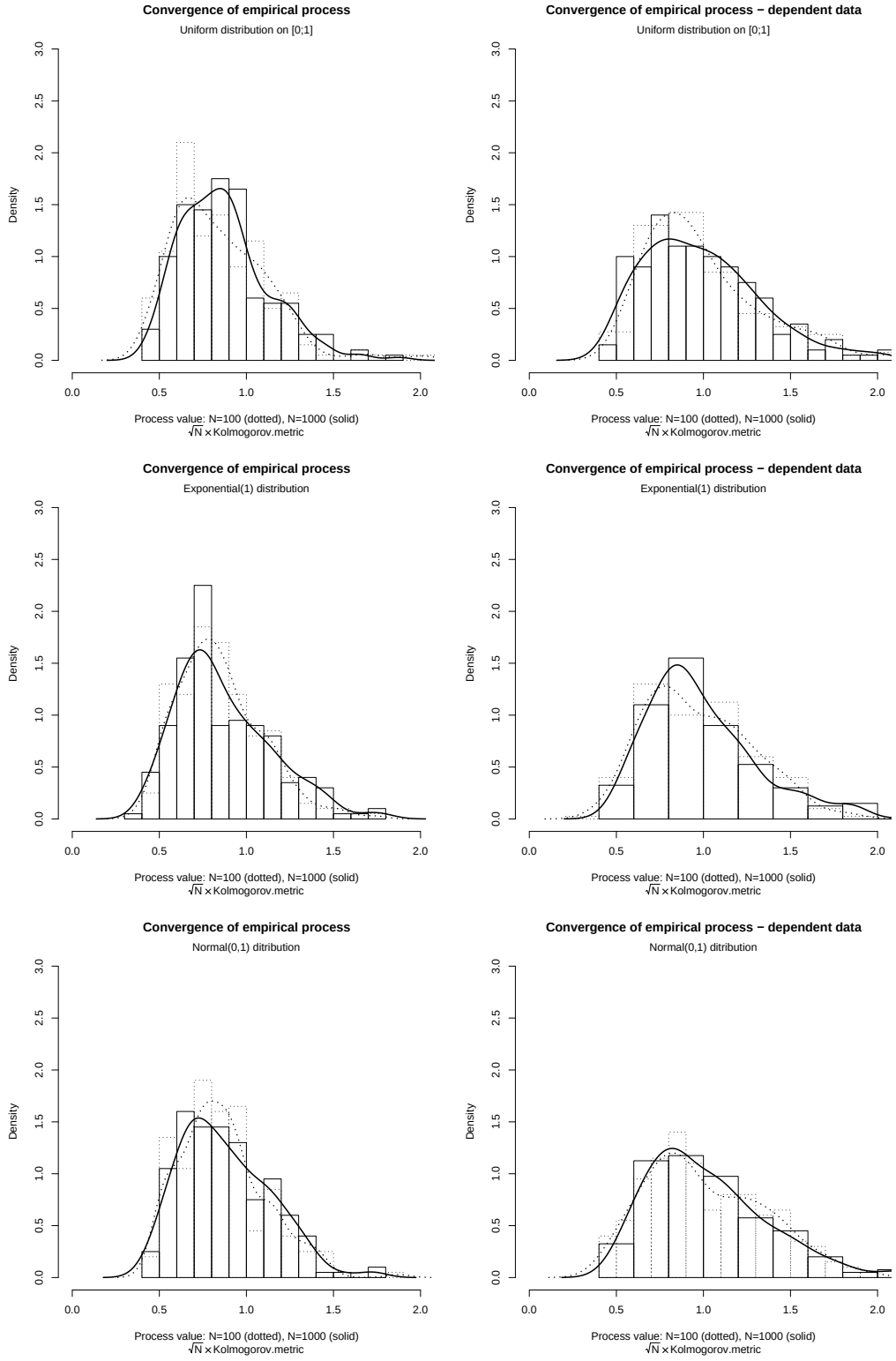


Figure 5.3: Convergence of the Kolmogorov empirical process

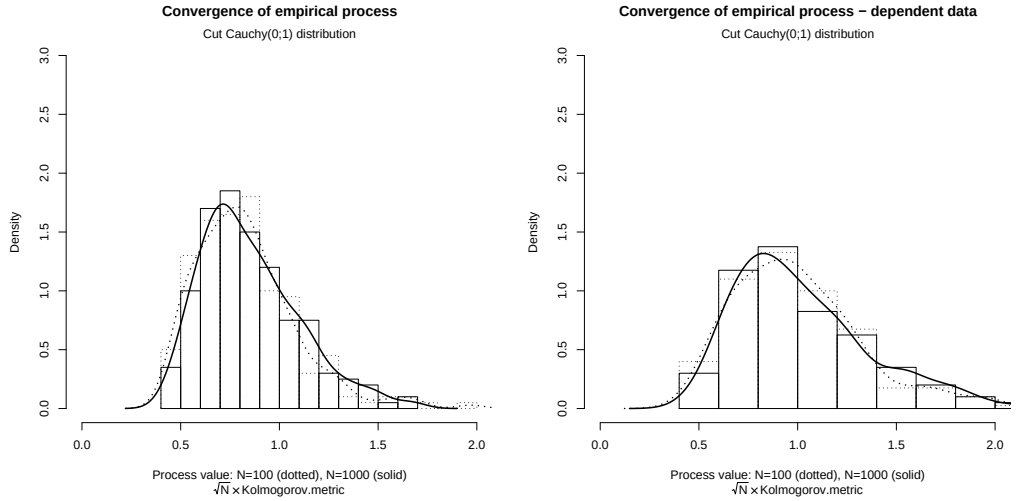


Figure 5.3: Convergence of the Kolmogorov empirical process (cont'd)

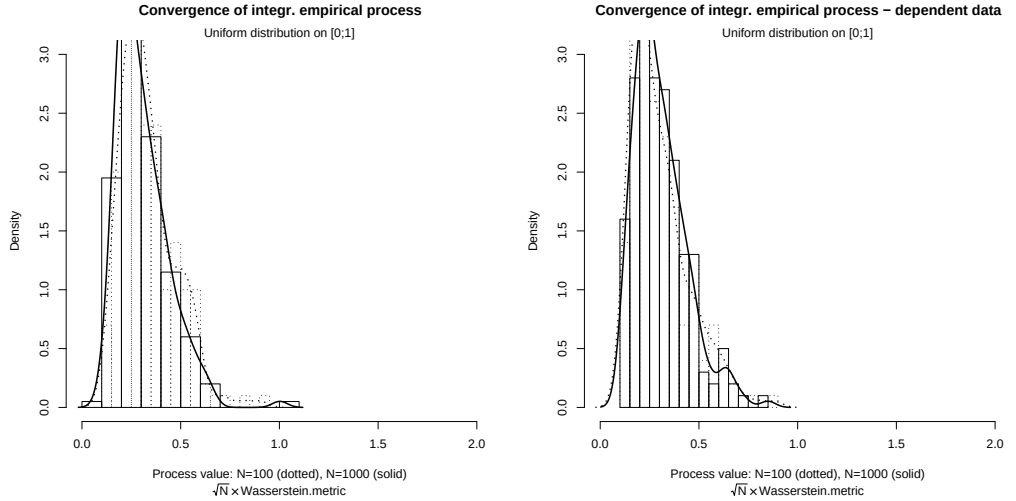


Figure 5.4: Convergence of the integrated empirical process

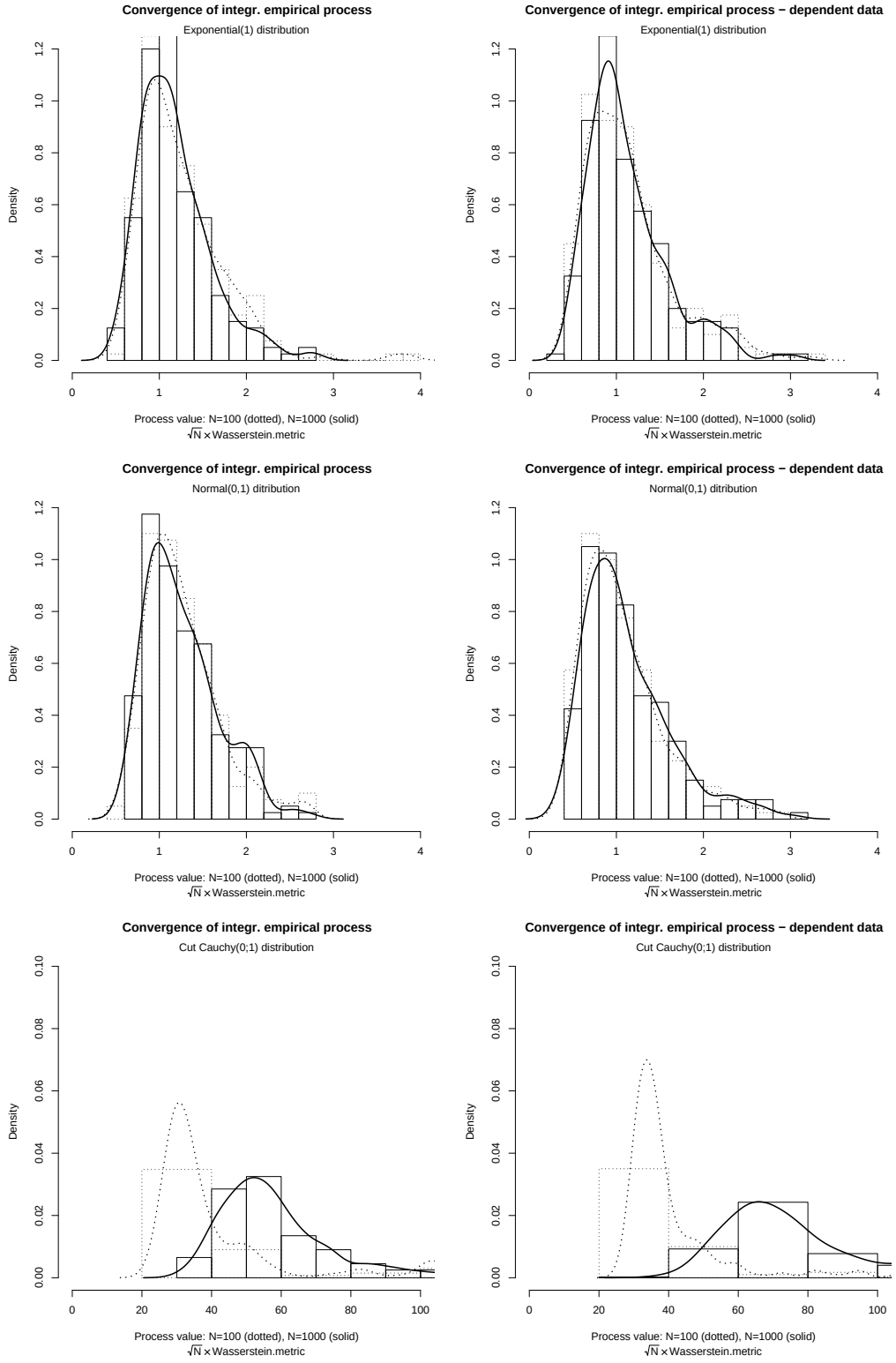


Figure 5.4: Convergence of the integrated empirical process (cont'd)

Chapter 6

Convexity and structural dependence in chance-constrained problems

In this chapter we focus our attention to a special case of chance-constrained problems. For computational and theoretical purposes, the convexity property of the constraint set is a highly important question: this favourable property has an indispensable impact to numerical properties of the problem as is the possibility to use well-known and highly developed convex programming methods. Also theoretical consequences of the convexity property of solution sets are of positive nature. In this chapter, another result on convexity of the set of feasible solutions is given, with an extension to stability of optimal solutions and a simple example.

6.1 Introduction

Let start our investigation with a slightly modified joint chance-constrained problem from Section 2.1.2:

$$\min c(x) \text{ subject to } \mathbb{P}\{h(x;\xi) \geq 0\} \geq p \tag{6.1}$$

where $x \in \mathbb{R}^n$ is again a decision vector, $\xi : \Omega \to \mathbb{R}^s$ is a random vector defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $c\colon \mathbb{R}^n \to \mathbb{R}$ is a vector real function, $h\colon \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^d$ is a vector-valued mapping, and $p \in [0;1]$ is a prescribed probability level. As in Chapter 2 we denote μ the distribution of the random vector ξ and $F = F_\mu$ its distribution function.

The problem (6.1) is equal to the problem (2.6) if $X = \mathbb{R}^n$ and for each $x \in \mathbb{R}^n$ we set $H(x) = \{\xi \in \mathbb{R}^s : h(x;\xi) \geq 0\}$, a set of realizations satisfying

the (let say original) constraints. To emphasize importance of the value p in this chapter, we denote the constraint set as

$$M(p) := \left\{x \in \mathbb{R}^n : \mathbb{P}\{\xi \in H(x)\} = \mu(H(x)) \geq p\right\}$$

and the whole problem simply as

$$\min F_0(x) \text{ subject to } x \in M(p).$$

Chance-constrained problems are common in engineering and economic practice and applications; well known are the applications in energy, water resources, production and inventory, telecommunications, and others. See Prékopa [51], Wallace and Ziemba [72], and references therein for details and further applications of chance-constrained programming.

A matter of importance both in theoretical and practical applications of the chance-constrained programming is to determine when the set $M(p)$ of feasible solutions is convex. It is trivially known that the sets $M(0)$, $M(1)$ (i. e., the set without any probabilistic constraint and the set of constraints satisfied almost surely) are convex if $h(\cdot, \xi)$ are concave functions for all $\xi \in \mathbb{R}^s$. Classical result of Prékopa Prékopa [48] (see also Prékopa [49], Tamm [69], Prékopa [50], Prékopa [51] and references therein) states the following:

Proposition 6.1 (Prékopa [51], Theorems 2.5 and 2.11)**.** *If μ is absolutely continuous (with respect to Lebesgue measure), log-concave measure (or r-concave with $r \geq -1/s$), and the one-dimensional components of h are quasi-concave functions of (x, ξ) then $M(p)$ is a convex set.*

A log-concave (or r-concave) measure is implied by a log-concave (or $\frac{r}{1-rs}$-concave) density. It is known (see the above references again) that many prominent multivariate distributions satisfy the condition, hence many chance-constrained problems can be solved by means of convex optimization.

The quasi-concavity property is not preserved under addition. For example, consider a problem with random right-hand side in the form

$$\min c(x) \text{ subject to } \mathbb{P}\{g(x) \geq \xi\} \geq p. \tag{6.2}$$

Problem (6.2) falls into the frame (6.1) if we set $h(x; \xi) = g(x) - \xi$. In order to have $h(x; \xi)$ quasi-concave, it is not sufficient to have $g(x)$ quasi-concave and $g(x)$ is usually required to be concave. Recently, Henrion and Strugarek [30] proposed an alternative approach to deal with this problem: their idea is to relax concavity condition of g and make more stringent concavity condition on the probability distribution μ. We recall this in Section 6.2. In Section

6.3 we further relax the condition of independence that authors of Henrion and Strugarek [30] require and show that, under modified assumptions, their results still remain valid. Section 6.4 introduces stability properties of the optimal values of the "dependent" and "independent" problems.

6.2 Concavity and density decrease

6.2.1 r-concave functions

We recall the definition of an r-concave function (see Prékopa [51], Definition 2.3):

Definition 6.1. A function $g : \mathbb{R}^n \to (0; +\infty)$ is called r-concave for some $r \in [-\infty; +\infty]$ if

$$g(\lambda x + (1-\lambda)y) \geq \left[\lambda g^r(x) + (1-\lambda)g^r(y)\right]^{1/r}$$

for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0; 1]$. Cases $r = -\infty, 0$, and $+\infty$ are treated by continuity.

We summarize some interesting cases in Table 6.2.1. For our purpose the most important are functions that are r-concave for $r \leq 1$. Note that if g is r^*-concave, it is also r-concave for all $r \leq r^*$.

$r = +\infty$	... $g(\lambda x + (1-\lambda)y) \geq \max\{g(x), g(y)\}$
$r \in (1; +\infty)$	... g^r is concave
$r = 1$	... g is concave
$r = 0$	... g is **log-concave** ($\log f$ is concave): $g(\lambda x + (1-\lambda)y) \geq g^\lambda(x) g^{1-\lambda}(y)$
$r < 0$	... g^r is convex
$r = -\infty$	... g is **quasi-concave:** $g(\lambda x + (1-\lambda)y) \geq \min\{g(x), g(y)\}$

Table 6.1: r-concave functions.

6.2.2 r-decreasing densities

Here we adopt the definition of Henrion and Strugarek [30] for a so-called r-decreasing function.

Definition 6.2 (Henrion and Strugarek [30], Definition 2.2)**.** A function $f\colon \mathbb{R} \to \mathbb{R}$ is called *r-decreasing* for some $r \in \mathbb{R}$ if

1. it is continuous on $(0; +\infty)$, and
2. there exists a threshold $t* > 0$ such that $t^r f(t)$ is strictly decreasing for all $t > t*$.

If $r = 0$ the function f is strictly decreasing in the classical sense. As for r-concave functions, if f is r^*-decreasing then it is r-decreasing for all $r \leq r^*$. In Henrion and Strugarek [30] it is shown that if marginal densities of the distribution μ are $(r+1)$-decreasing (for some $r > 0$), then the mapping $t \mapsto F(t^{-1/r})$ (F being the corresponding distribution function) is concave for $t \in (0; t^{*-r})$, which is further shown to be sufficient to ensure convexity property of the problem (6.2):

Theorem 6.2 (Henrion and Strugarek [30], Theorem 3.1)**.** *If*

1. *there exist $r_i > 0$ such that the components g_i of g are $(-r_i)$-concave,*
2. *the components ξ_i of ξ have $r_i + 1$-decreasing densities, and*
3. *the components ξ_i of ξ are independently distributed,*

then $M(p)$ is convex for all $p > p^* := \max_i F_i(t_i^*)$ where F_i denotes the distribution function of ξ_i and t_i^* refer to the definition of $r_i + 1$-decreasing probability density.

As shown in Henrion and Strugarek [30], required treshold constants p^* are not really high for prominent one-dimensional distributions and so Theorem 6.2 can be directly used in applications.

6.3 Weak dependence of the rows

In the sequel we ask for the relaxation of the independence condition (iii) in Theorem 6.2. To do this, we define an α' coefficient of dependence by the following definition.

Definition 6.3. For a random vector ξ we define a coefficient α' of (weak) dependence as

$$\alpha' := \sup_z \left| F(z) - \prod_i F_i(z_i) \right| \tag{6.3}$$

where F is the distribution function of the vector ξ, F_i are the corresponding one-dimensional marginal distribution functions and $z = (z_1, \ldots, z_s) \in \mathbb{R}^s$.

If $\alpha' = 0$, ξ_i's are independent random variables. Our α'-coefficient is inspired by and actually is a modified version of the known strong-mixing coefficient, see e. g. Yoshihara [74]. The classical strong-mixing (α-mixing) coefficient is generally defined for two σ-fields $\mathcal{B}_1$, $\mathcal{B}_2 \subset \mathcal{A}$ by

$$\alpha(\mathcal{B}_1, \mathcal{B}_2) := \sup_{A \in \mathcal{B}_1, B \in \mathcal{B}_2} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

If $\mathcal{B}_1, \mathcal{B}_2$ are σ-fields generated by two random vectors ξ_1, ξ_2, we say that the two vectors are α-dependent. For our purposes we extend the definition to more than two random variables (this is straightforward) and use their distribution functions instead of generated σ-fields. Of course, if some variables ξ_1, ξ_2 are α-dependent (in the strong-mixing sense), they are also α'-dependent in the sense of Definition 6.3.

Convention. To simplify the notation, we drop the prime symbol $'$ from Definition 6.3 and use the notion of α-dependence in the sense of Definition 6.3 for the remaining part of the chapter. As the original α strong-mixing coefficient is not used anymore in the thesis this will cause no additional problem.

In what follows we allow for a small structural dependence introduced by a small value of α. Recall that the set of feasible solutions of (6.2) can be written as

$$M(p) := \big\{x \in X : F(g(x)) \geq p\big\}$$

where F is the distribution function of the random right-hand side ξ. Further, denote

$$M'(p) = \big\{x \in X : \prod_{i=1}^{s} F_i(g_i(x)) \geq p\big\}$$

where F_i are one-dimensional marginal distribution functions of F.

If the components ξ_i of ξ are independently distributed, the two sets are equal:

$$M(p) = M'(p)$$

This is not true in case of weak dependence, but the following proposition is valid:

Proposition 6.3. *If the components ξ_i of ξ in* (6.2) *are α-dependent (in the sense of Definition 6.3) then*

$$M'(p + \alpha) \subset M(p) \subset M'(p - \alpha) \subset M(p - 2\alpha). \tag{6.4}$$

Proof. If the components of ξ are α-dependent, we have

$$|F(g(x)) - \prod_i F_i(g_i(x))| \leq \alpha, \text{ i.e.}$$

$$\prod_i F_i(g_i(x)) - \alpha \leq F(g(x)) \leq \prod_i F_i(g_i(x)) + \alpha.$$

The second inequation together with the definition of $M(p)$ implies

$$p \leq F(g(x)) \leq \prod_i F_i(g_i(x)) + \alpha$$

$$\prod_i F_i(g_i(x)) \geq p - \alpha,$$

that is

$$M(p) \subset M'(p - \alpha) \tag{6.5}$$

Similarly the first inequation of (6.3) yields

$$p - \alpha \leq \prod_i F_i(g_i(x)) \leq F(g(x)) + \alpha$$

$$F(g(x)) \geq p - 2\alpha$$

hence

$$M'(p - \alpha) \subset M(p - 2\alpha). \tag{6.6}$$

Combining equations (6.5) and (6.6) we obtain the whole chain of inequalities

$$M'(p + \alpha) \subset M(p) \subset M'(p - \alpha) \subset M(p - 2\alpha). \tag{6.7}$$

□

If p is sufficiently high, then possibly non-convex $M(p)$ is bounded from both side by convex sets:

Theorem 6.4. *If*

1. *there exist $r_i > 0$ such that the components g_i of g are $(-r_i)$-concave,*
2. *the components ξ_i of ξ have $r_i + 1$-decreasing densities,*
3. *the components ξ_i of ξ are α-dependently distributed, and*
4. *$p > \max_i F_i(t_i^*) + \alpha$ (with t_i^* and F_i as in Theorem 6.2),*

*then $M(p)$ is bounded (from both sides) by **convex** sets $M'(p+\alpha)$ and $M'(p-\alpha)$.*

Proof. The last will be proved modifying the proof of Theorem 6.2. Due to the assumptions, for $x, y \in M'(p-\alpha)$ we have

$$0 \leq F_i(t^*) < p - \alpha \leq F_i(g_i(x)) < 1$$

and the same inequalities are valid for x replaced by y. The second inequality follows simply from the assumption (iv), the last one is proved by Lemma 3.1 of Henrion and Strugarek [30]. By the continuity of marginal distributions functions and quantile properties we obtain

$$0 < t_i^* < F_i^{-1}(F_i(g_i(x))) \leq g_i(x)$$

and the same with y. The remaining part of the proof is the same as in Henrion and Strugarek [30], in particular we have for $\lambda \in [0;1]$

$$\begin{aligned}
&\prod_{i=1}^{m} F_i(g_i(\lambda x + (1-\lambda)y)) \\
&\quad \geq \prod_{i=1}^{m} [F_i(F_i^{-1}(F_i(g_i(x))))]^{\lambda} [F_i(F_i^{-1}(F_i(g_i(y))))]^{1-\lambda} \\
&\quad = [\prod_{i=1}^{m} F_i(g_i(x))]^{\lambda} [F_i(g_i(y))]^{1-\lambda} \\
&\quad \geq (p-\alpha)^{\lambda}(p-\alpha)^{1-\lambda} \\
&\quad = p - \alpha
\end{aligned}$$

hence $\lambda x + (1-\lambda)y \in M'(p-\alpha)$. The other bound follows directly from Proposition 6.3. □

Theorem 6.4 has great impact also on numerical solutions of "dependent problems." If we (somehow) know that the components of ξ are α-dependent so that α is small enough (at least such that condition 4 of 6.4 is fulfilled), then, solve two convex, "independent" problems

$$\min c(x) \text{ subject to } x \in M'(p-\alpha), \text{ and} \tag{6.8}$$

$$\min c(x) \text{ subject to } x \in M'(p+\alpha). \tag{6.9}$$

The optimal solutions of these problems (or the optimal solution of the "independent" problem with $M'(p)$) are good approximations of the original (dependent) problem in the meaning that both optimal values of the problems are lower and upper bounds of the optimal value of the original problem. An insight into stability properties of the problem is provided by the following section.

6.4 Stability properties

We have already stated that optimal value of the problem (6.2) is bounded by optimal values of the problems (6.8) and (6.9). Under additional assumptions on objective and constraint functions, we can prove Theorem 6.5.

Before do that we recall the notion of (special) metric regularity introduced in Section 4.1.4. Here it will be applied to the constraint set $M'(p)$: actually $X_0 = M'(p)$, $X_y = M'(p-y)$, and the inequality (4.5) is now rewritten as

$$\operatorname{dist}(x, M'(p-y)) \leq a \max\{0, p + y - \prod_{i=1}^{s} F_i(g_i(x))\}$$

valid for some $\varepsilon > 0$ and some $a > 0$, for all $x \in \mathbb{B}_\varepsilon(\bar{x}), y \in \mathbb{B}_\varepsilon(0)$ where $\mathbb{B}_\varepsilon(x)$ is ε-neighbourhood of x. The condition ensures the quantitative stability property for the optimal value of "dependent" and "independent" problem.

Theorem 6.5. *Consider the problem* (6.2) *and let*

1. *assumption 1–4 of Theorem 6.4 be fulfilled;*
2. *c be Lipschitz continuous function on $\mathbb{R}^n$;*
3. *the mapping $x \mapsto \{y \in \mathbb{R} | x \in M'(p + \alpha - y)\}$ be metrically regular at $(\bar{x}'(p+\alpha), 0)$;*
4. *α-dependence coefficient satisfies $\alpha < \frac{1}{2}\varepsilon$, where ε is provided by metric regularity condition (iii).*

Then there exists a constant $L > 0$ such that

$$|\varphi'(p+\alpha) - \varphi(p)| \leq L \max\{0, p + \alpha - \prod F_i(g_i(\bar{x}'(p-\alpha)))\} \tag{6.10}$$

is fulfilled. Here, $\varphi(p)$, $\varphi'(p+\alpha)$ are the optimal values of (6.2) *and* (6.9) *respectively, and $\bar{x}'(p-\alpha)$, $\bar{x}'(p+\alpha)$ are the optimal solutions of* (6.8) *and* (6.9) *respectively.*

Proof. Under the assumptions of Theorem 6.5

$$\begin{aligned} &|\varphi'(p+\alpha) - \varphi(p)| \\ &\quad \leq |\varphi'(p+\alpha) - \varphi'(p-\alpha)| \\ &\quad \leq L\|\bar{x}'(p+\alpha) - \bar{x}'(p-\alpha)\| \end{aligned}$$

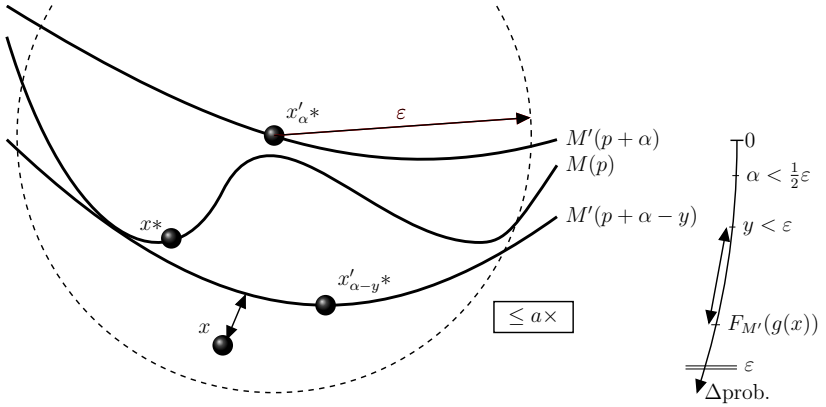


Figure 6.1: Metric regularity and bounded feasibility sets.

If $\alpha < \frac{1}{2}\varepsilon$, the metric regularity condition imply

$$\begin{aligned}&\|\bar{x}'(p+\alpha) - \bar{x}'(p-\alpha)\| \\ &\quad \leq a \max\{0, p+\alpha - \prod F_i(g_i(\bar{x}'(p-\alpha))\}\end{aligned}$$

Combining the two inequalities yields the assertion. □

The concept of Theorem 6.5 is illustrated in Figure 6.1. The stability of optimal solutions is more complicated thing that require some kind of growth condition (see Chapter 4.1.3); this question is not yet finished and will be subject of our further research.

6.5 Example: comparison

We are now ready to illustrate the result of this chapter on a simple optimization problem taken originally from Henrion and Strugarek [30]:

$$\text{minimize } x + y \text{ subject to} \tag{6.11}$$

$$\begin{aligned} g_1(x,y) &= \frac{1}{x^2+y^2+0.1} &&\geq \xi_1 \\ g_2(x,y) &= \frac{1}{(x+y)^2+0.1} &&\geq \xi_2 \end{aligned}$$

Assume that the random vector ξ is normally distributed with zero mean and the variance matrix Σ. Further, we consider two cases:

1. independent case with

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{6.12}$$

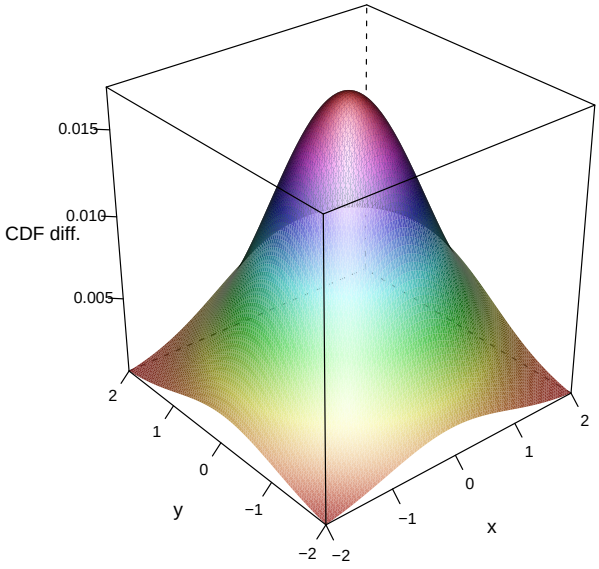


Figure 6.2: Estimation of the α coefficient

2. weak dependent case with

$$\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix} \tag{6.13}$$

In this example the g_i's are (-1)-concave, ξ_i's have 2-decreasing densities with the treshold $t^* = \sqrt{2}$, the critical probability level is $p^* = \Psi(\sqrt{2}) = 0.921$ (these values are taken from the original paper), and the weak-dependence coefficient for the dependent case is $\alpha = 0.017$ (this last is estimated value, see Figure 6.2).

The overall shape of the collection of the sets $M'(p)$ is given in Figure 6.3. Each individual set $M'(p)$ is given as horizontal cut at the specified level p (found on z-axis). The contour lines of these sets are depicted on Figure 6.4. As usual, the symbol φ denotes the optimal values of the corresponding problem; unfortunately, the right-hand side of (6.10) is of the existence type only, thus the actual value of the upper bound remains unknown.

For the chosen normal distribution, convexity of the feasible set is assured theoretically at the probability level of 0.921 for the independent case. As Figure 6.3 shows, the actual probability level in the example is much more smaller, around the value of 0.7.

For the weak dependent case, these tresholds (theoretical and actual) are shifted towards the center of feasibility sets (center of image), and the

optimal values and optimal solutions (depicted as points on the last figure) remain stable as the value of α-coefficient is small.

There are still two interesting but open questions: first, how is the actual treshold value for both independent and dependent case; second, if there is convexity property even on weakly dependent structures above the treshold value. This last is seen from Figure 6.4 but there is still no theoretical background to prove this hypothesis.

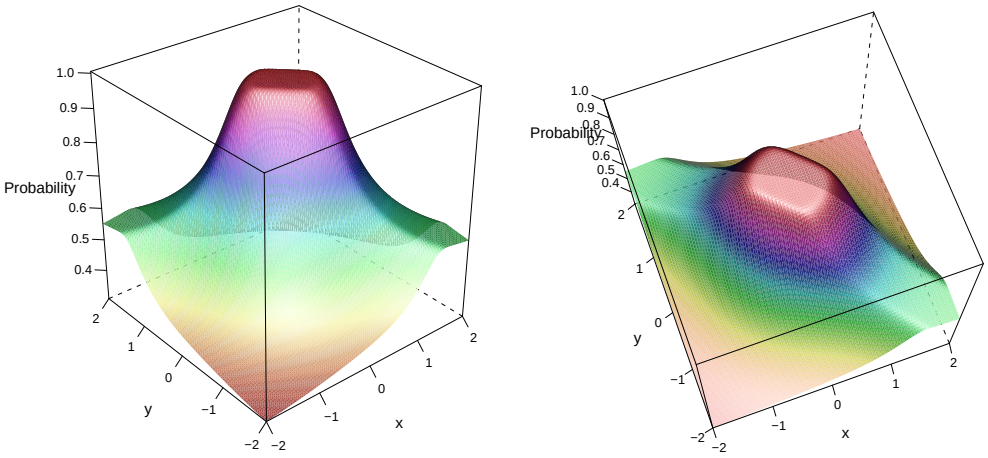


Figure 6.3: Collection of sets $M'(p)$

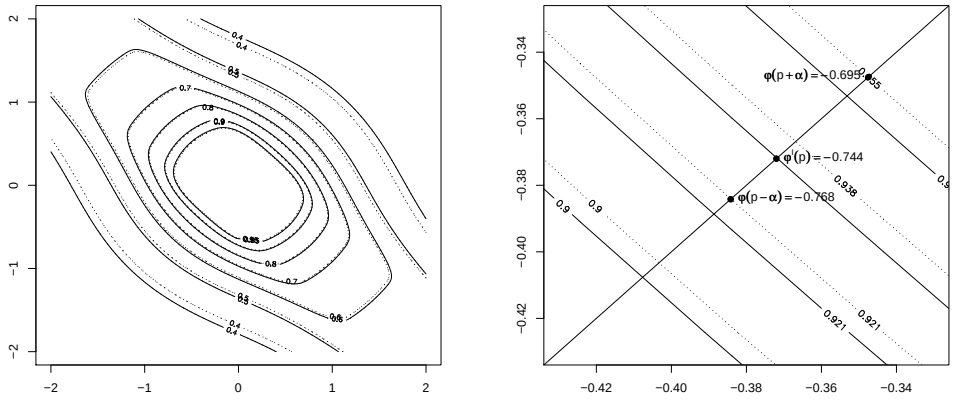


Figure 6.4: Contour lines for $M'(p)$ (solid) and $M(p)$ (dotted) sets

Chapter 7

Chance-constrained problems and robust programming approach

Techniques of *robust programming* have became a popular alternative to stochastic programming during the last two decades. The robust programming handles the uncertainty in an optimization model through the so-called "worst case" analysis: the constraints of the general optimization model (1.1) are required to be satisfied for all possible realizations of uncertainty parameter ξ (these realizations are called *instances* of ξ by the robust programming community), and we optimize the worst-case objective function among all robust solutions. Even if such paradigm is classical in the statistical decision theory, real development of this discipline in applied optimization dates only to the last decade starting with the works of Mulvey et al. [46] and Ben-Tal and Nemirovski [4]. On the other hand, robust optimization problems are not new (they form part of semi-infinite programming problems); also the influence of the robust control theory is evident and not negligible.

In this chapter we introduce some basic notions and techniques used to solve uncertain optimization programs and try to compare them with the stochastic programming – from a philosophical rather than quantitative point of view. Both techniques are eligible, each in its own context viewed from mathematical and applied (economic, engineering, managerial) point of view.

7.1 Uncertain convex program

Recall a simplified version of the general optimization problem introduced by (1.1). We call it an *uncertain convex program* (UCP). In fact, this is a *family*

of convex optimization programs parameterized by $\xi \in \Xi$. We consider the following "standardized" form of the uncertain convex program:

$$\text{minimize } c'x \text{ subject to } x \in X, f(x;\xi) \leq 0, \tag{7.1}$$

where

- $x \in X \subset \mathbb{R}^n$ is the decision vector,
- $c \in \mathbb{R}^n$ is a (known) objective coefficient, and
- $\xi \in \Xi \subset \mathbb{R}^s$ is a parameter (data) vector.

We assume further that

- X is convex and closed set,
- $f\colon X \times \Xi \to \mathbb{R}$ is convex in x for all $\xi \in \Xi$,
- Ξ is a prescribed set of instances (cf. the note concerning the overloaded symbol Ξ in Section 1.1.4).

As we stated already in Introduction (cf. Example 1.3), without lost of generality, the objective function is linear and the constraint function f is scalar-valued. More precisely, the linearity of the objective can be imposed by moving the non-linear (and possibly random) objective function to the constraint part of the problem:

$$\text{minimize } t \text{ subject to } x \in X, c(x;\xi) \leq t, f(x;\xi) \leq 0,$$

and multiple valued convex constraint functions $f_i(x;\xi)$ (incorporating the moved objective if needed) can be converted into a single scalar-valued constraint by the transformation

$$f(x;\xi) := \max_i f_i(x;\xi)$$

If the realization of ξ is known and fixed, the problem (7.1) is deterministic convex program and we can use techniques of convex programming to solve the problem. This corresponds to the approach of ignoring uncertainty as described in Chapter 1. In many cases, such solution is very sensitive to perturbations of ξ and other methods must be used.

7.1.1 Chance-constrained program

In this chapter, we compare the chance-constrained program (as a member of the family of stochastic programming problems) with a robust one. We recall here a definition of the *chance-constrained problem* (PCP), with slightly modified notation: assuming that ξ is a random vector defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with known probability distribution $\mu \in \mathcal{P}(\Xi)$, we require the constraints of (7.1) to be fulfilled with a prescribed level of probability $1 - \varepsilon$. The problem reads

$$\text{minimize } c'x \text{ subject to } x \in X_\varepsilon := \big\{x \in X : \mu\{\xi \in \Xi : f(x;\xi) > 0\} \leq \varepsilon\big\}. \tag{7.2}$$

There are several issues addressed to chance-constrained problems:

1. the problem (7.2) or its approximation need not to be convex even if f is convex in x for all ξ;
2. to evaluate the probability in the definition of X_ε one often has to calculate values of multidimensional integrals.

The first issue has been partially overcame using r-concave measures; Chapter 6 is a contribution of our thesis to the area. Many numerical techniques are used to overcome the second issue; we present one of possible approaches below.

The term of "chance-constrained" optimization is coined with an early work of Charnes and Cooper [11]; most of the relevant literature is resumed in Prékopa [50]. Main results of the topic concern conditions under which (7.2) is a convex program and how to convert the probability constraints into an explicit deterministic form.

7.1.2 Robust convex program

The robust programming approach is an alternative way to deal with uncertainty parameters in (7.1). It is also known as 'min-max' or 'worst-case' approach due to the nature of the problem. In robust optimization we look for a solution which is feasible for *all* possible instances of ξ; this approach leads to the *robust convex problem* (RCP):

$$\min_{x \in X} c'x \text{ subject to } f(x;\xi) \leq 0 \text{ for all } \xi \in \Xi. \tag{7.3}$$

Throughout, we assume that there exists a feasible solution to (7.3). The robust convex programming problem is convex but it has an infinite number of constraints and so it is numerically hard to solve. The robust optimization

methods propose some relaxation techniques to deal with such a problem. In what follows, we consider a solution method based on 'randomization' of the parameter ξ and the sampling techniques. Another disadvantage of robust programming in its original form is the fact that it allocate the same weight to all values of the parameter regardless possible different importance of individual instances of ξ.

The mathematical and computational framework of robust optimization problems was introduced by Mulvey et al. [46] and Ben-Tal and Nemirovski [4], and developed by other authors in various directions, see e. g. Ben-Tal and Nemirovski [5], [6], Bertsimas and Sim [7], El Ghaoui et al. [21], and references therein.

7.2 Approximations to stochastic and robust optimization programs

To solve the chance-constrained program (7.2), complete knowledge of probability distribution μ is needed. On the other way, robust programming problem (7.3) is not solvable at all as mentioned above (having infinite number of constraints in general). Thus various approximation and estimation techniques are adopted to overcome these difficulties, both in stochastic and robust programming. In this section we present two of them, one for stochastic and one for robust programming approach. Both have in common their sampling nature that involves the probability distribution of parameter ξ.

7.2.1 Chance-constrained "sampled" problem

Consider a set of independent samples $\xi_1, \ldots, \xi_N$ distributed according to μ, the original distribution of the parameter ξ. Recall that

$$F_N(t) = \frac{1}{N} \sum_{i=1}^{N} \chi_{(-\infty;t]}(\xi_i), \quad t \in \mathbb{R} \tag{7.4}$$

is the empirical distribution function defined as a (discrete) random variable for the given sample (see Section 5.2 to find more on notation and other properties of empirical distribution functions). The problem (7.2) is approximated, for the given sample, by replacing the original probability distribution μ by the empirical probability measure μ_N associated with F_N, and then the problem reads

$$\min_{x\in X} c'x \text{ subject to } x \in X[\varepsilon, N] := \left\{x \in X; \frac{1}{N}\,\mathrm{card}\{i; f(x;\xi_i) > 0\} \leq \varepsilon\right\} \tag{7.5}$$

where card denotes the cardinality of the argument. In this chapter, we call the problem (7.5) as the *chance-constrained sampled problem* (PCP_N). The essential idea of (7.5) is that the relative frequency of constraint violations corresponds to the desired upper level of infeasibility in (7.2). (7.5) is the program with a single constraint and in some simple cases it is computationally tractable.

There exists many results in the theory of stability of stochastic optimization problems dealing with a question how far is the optimal solution of (7.5) from the original optimal solution of (7.2); among all we refer to works Henrion and Römisch [28], Kaňková [37], Kaňková [40], Rachev and Römisch [54], Römisch [59], references therein, and many of other authors. Here, general stability theorem (Theorem 4.1) can be considered as a base for further special results. For example, Corollary 2 in Henrion and Römisch [28] states conditions under which the Hausdorff distance between optimal solution set of (7.2) and of (7.5) is Hölder continuous with respect to the Kolmogorov metric. As the Kolmogorov metric $K(\mu, \mu_N)$ converges almost surely to zero under general conditions, the distance between optimal solutions of (PCP) and (PCP_N) is expected to converge to zero too. This will be made in evidence on the numerical example in Section 7.3.

7.2.2 Robust sampled program

Recently, Calafiore and Campi [10] and de Farias and Van Roy [15] independently proposed the following approximation to the robust convex program (7.3). Consider again the set of independent samples $\xi_1, \ldots, \xi_N$ distributed according to given probability distribution μ. The robust convex problem (7.3) is then approximated by the following *robust sampled convex problem* (SCP_N):

$$\min_{x\in X} c'x \text{ subject to } X[N] := \left\{x \in X; f(x;\xi_i) \leq 0 \text{ for } i = 1, \ldots, N\right\}. \tag{7.6}$$

This is a relaxation of the original robust convex problem: we do not require the original constraints to be satisfied for all realizations of $\xi \in \Xi$ but only for a certain finite but sufficiently large number of samples which are moreover the most probable to happen. Calafiore and Campi [9] found a rule to set up N in order to have the optimal solution of (7.6) feasible in (7.2):

Proposition 7.1 (Calafiore and Campi [9], Theorem 2)**.** *For fixed $\varepsilon, \beta > 0$, the optimal solution of* (7.6) *is feasible in* (7.2) *with a probability at least* $1 - \beta$ *if*

$$N \geq \frac{2n}{\varepsilon} \ln \frac{1}{\varepsilon} + \frac{2}{\varepsilon} \ln \frac{1}{\beta} + 2n.$$

This approach (also called "randomized" by the authors) has several favorable impacts:

- the problem is convex, it has a finite number of constraints and it is effectively computable;
- it incorporates weights to the individual parameter instances of ξ – their absence in the original robust convex program is also the subject of criticism of the common robust framework;
- in addition, realizations of ξ used in (7.6) are such that are most probable to happen.

The randomized approach to robust convex program was proposed in Calafiore and Campi [10] and de Farias and Van Roy [15]. In Erdoğan and Iyengar [22] the idea was extended to the case of the so-called ambiguous chance-constrained programming where the distribution μ is known only approximately. The study of these authors is based on the Vapnik-Červonenkis (VC) theory, see e. g. Vapnik [71].

The solution of (7.6) approximates the solution of (RCP): higher the number of samples, closer the solutions are. In order to find the optimal solution of (SCP_N) that is sufficiently close to the optimal solution of (RCP) one need a rather high number of samples to be generated. The authors of above-cited papers have sought for a rule on the sample size N that assures the optimal solution of (SCP_N) to be ε-feasible, i. e. feasible in the chance-constrained problem (PCP). But one cannot expect that this solution is near to the optimal solution of (PCP). In fact, in the following section we give a comparative numerical study on a simple optimization problem where this conclusion is approved. We further develop this idea in detail from the economical and practical point of view.

7.3 Numerical study

7.3.1 Problem setting

On the following simple numerical example we illustrate the fact that the two above mentioned approaches to solve uncertain convex program are based on

a different "philosophy" how to understand the uncertainty. Thus, consider the following uncertain convex program

$$\text{minimize } x \text{ subject to } x \geq \xi, x \in \mathbb{R} \tag{7.7}$$

where $\xi \in \mathbb{R}$ is distributed according to the standard normal distribution $\mathrm{N}(0;1)$, and to the exponential distribution $\mathrm{Exp}(1)$ with rate one, respectively. We denote F the distribution functions for both cases (as we do not intend to differentiate between them).

According to Sections 7.1 and 7.2, we define the following deterministic programs

$$\min_{x\in\mathbb{R}} x \text{ subject to } x \geq F^{-1}(1-\varepsilon), \tag{7.8}$$

$$\min_{x\in\mathbb{R}} x \text{ subject to } \frac{1}{N}\operatorname{card}\{i; x < \xi_i\} \leq \varepsilon, \tag{7.9}$$

$$\min_{x\in\mathbb{R}} x \text{ subject to } x \geq \max_{i=1,\dots,N} \xi_i. \tag{7.10}$$

(7.8) is chance-constrained problem (PCP), (7.9) is sampled version of it (PCP_N), and (7.10) is robust sampled problem (SCP_N). Normal and exponential distribution functions are defined on an unbounded set, hence the missing robust program (RCP) is not well defined – there is no real solution feasible to all the instances of ξ. However, practical interest of this last fact is small; we can use some suitable transformation of the distribution in order to obtain a bounded support and define the new problem. We do not pursue this direction in what follows.

For each of the three above problems the optimal solution coincides with the lower boundary of their feasibility sets. That is:

- the optimal solution of the chance-constrained problem (PCP) related to (7.7) is $1-\varepsilon$ quantile of F;
- the optimal solution of the chance-constrained sampled problem (PCP_N) is computed as the lowest from N samples which are greater or equal to the 0.95 sample quantile;
- the optimal solution to (SCP_N) is the highest of N samples considered.

7.3.2 Comparison between the stochastic and robust approach

To compare the approaches we set up $\varepsilon = \beta = 0.05$ (probability level in (7.2) and confidence parameter in Proposition 7.1). Figure 7.1 illustrate the

convergence properties of robust and chance-constrained sampled problems. The optimal solution of (7.8), marked by a small tickmark on x-axis, is approximately 1.64 for the normal and 3.00 for the exponential distribution. The number of samples assuring that the optimal solution of (SCP_N) is 0.05-feasible with probability 0.95 (see Proposition (7.1)) is about 240. Thus, we consider three values of the sample size $N = 30, 240, 3000$, representing low, "accurate", and large sample size.

The sampling procedure is repeated 200 times for each sample size in order to estimate densities (histograms) for the optimal solutions which are themselves random variables. (These estimates are standard Gaussian kernel estimates as defined by default in `density` function of `R` programming language.) The optimal solution of (PCP_N) is represented by the dotted line, the optimal solution of (SCP_N) is represented by the solid one. The left-hand column of the array represents normal distribution, the right-hand column represents exponential distribution.

First, dotted histograms represent the fact that the optimal solution of the chance-constrained sampled problem (7.9) converges, as N goes to infinity, to the solution of (7.8) (marked by the tickmark). The sampling method related to (7.5) is useful especially if the number N of samples is high, as a possible error in estimating optimal solution of the chance-constrained problem (7.2) decreases.

On the other hand, the robustly sampled optimal solutions of (7.10) (solid histograms) are getting away from the point of the optimal solution of (PCP) as far as the number of samples increases, and are going to the upper boundary of the support of F (i. e. to the infinity in both of our cases), but with rapidly decreasing rate.

For the robust sampled problem, the tickmark represents the lower boundary of ε-feasibility set X_ε, i. e. the limiting point for which a solution is feasible in (PCP) with a high probability. You could observe the fact mentioned by Proposition 7.1 – if N is greater than 241, the optimal value of the (SCP_N) program is feasible in (PCP) with probability of 0.95.

The presented approaches exhibit very different numerical results. Before making a conclusion on this difference in Section 7.5 we explain the background of the approaches also by their possible real-world applications in the next section.

7.4 Applications of the chance-constrained and robust programming

7.4.1 Applications of the chance-constrained problems

There is a huge number of applications in stochastic programming due to the long history of the subject. A collection of the most important ones is given in the Wallace and Ziemba's book Wallace and Ziemba [72]. We give here only a short overview of selected particular tasks that was solved in real-world applications. Other items include applications in agriculture, power generation and electricity distribution, military, production control, telecommunications, transportation and many others.

- Chemical engineering (Henrion and Möller [27]). A continuous distillation process is frequently very dependent on a controlled rate of its inflow; if the last is of stochastic nature, it cannot be processed immediately but has to be stored in a feed tank. The objective is to find the optimal feed control with the prescribed lower and upper level of the inflow preventing the feed tank to be empty or full, together with the fact that costs compensating possible level violations are difficult to model.
- Finance – portfolio selection. The objective is to select the optimal portfolio of bonds in order to maximize final amount of money and to cover necessary payments in all years. The last is modeled via liquidity constraints we want to satisfy with some high level of probability.
- Water management (Prékopa and Szántai [52]): one of the very beginning application of chance-constrained problems. A number of reservoirs must be designed in order to control flooding due to random stream inflows.

These models have in common that we estimate the probability distribution (needed to solve the optimization problem) by means of observations from the past. The resulting solution in (PCP_N) is an approximation to the (unknown) solution of (PCP) and the approximation is better as the number of samples (observations) is higher. Furthermore, our solution of the chance-constrained "sampled" problem is only approximatively ε-feasible for a given level ε. On the other hand, this level is usually not crucial for real applications if our preferences are pointed towards costs saving solutions – this is also the case of all mentioned applications.

7.4.2 Applications of robust sampled problem

The number of real-world applications of the robust programming is a little more sparse. The most important are the robust truss topology design and the robust portfolio selection problems; there are also other applications, especially in finance (e. g. option modeling), management (supply chain management), or engineering (power supply).

- Robust Truss Topology Design (Ben-Tal and Nemirovski [4]). The problem is to select the optimal configuration of a structural system (mechanical, aerospace, ...) that is subjected to one ore more given loads (nominal loads) and an unspecified set of small uncertain loads. The goal is to find such configuration that the construction is rigid to all of the prescribed loads.
- Robust portfolio selection (Goldfarb and Iyengar [24]). Here, the uncertain parameters are the modeling errors in the estimates of market parameters and they are assumed to lie in a known and bounded uncertainty set. The robust portfolio is the solution to an optimization problem where the worst-case behaviour of parameters is assumed.

The optimal solutions to robust problems hedge against the worst-case realization of uncertain parameters regardless their "importance". The randomized (sampled) approach incorporate the information about importance to the model via probability distribution of samples so that the optimal solution of the sampled problem does not have to satisfy the constraints for all possible realizations of the parameter. At the same time, the probability of such violation is small and for a given ε one could easily compute the number of samples to generate in order to obtain an ε-feasible solution. Indeed, if the number of samples is significantly greater than required, the optimal solution of the sampled problem also hedges against the parameters with the smaller probability of occurrence. This could be the task if the risk of constraint violation has to be minimized as much as possible and costs of doing that are of smaller importance. This is usually the case of the truss constructions mentioned above.

7.5 Conclusion – stochastic vs. robust programming

Choosing the approximation method to an uncertain convex program is ambiguous. There is no reason to measure difference between the two optimal

solutions as they originate in different context. But the selected method has to fill up the needs of practical dimension of the problem:

- how much the probability of violation of the constraints is crucial,
- how many samples one has at disposition or can generate.

Getting an answer to the first question stronger, one's preferences have to be directed towards the robust sampled problems assuring the high probability of fulfilling the constraints. Chance to fulfill the constraints by the optimal solution of chance-constrained sampled problem is only *approximately* the desired value $1 - \varepsilon$, especially if the number of samples is low. But this solution could be useful in cases where the $1 - \varepsilon$ level is not crucial and our preferences are pointed more likely towards costs savings solutions.

We have illustrated how these general theses apply in a simple optimization problem and two rather 'representative' distributions. The generalization is possible – for other distributions it is straightforward, for the multidimensional case the problem is likely to be a problem of getting data and obtaining a clear representation.

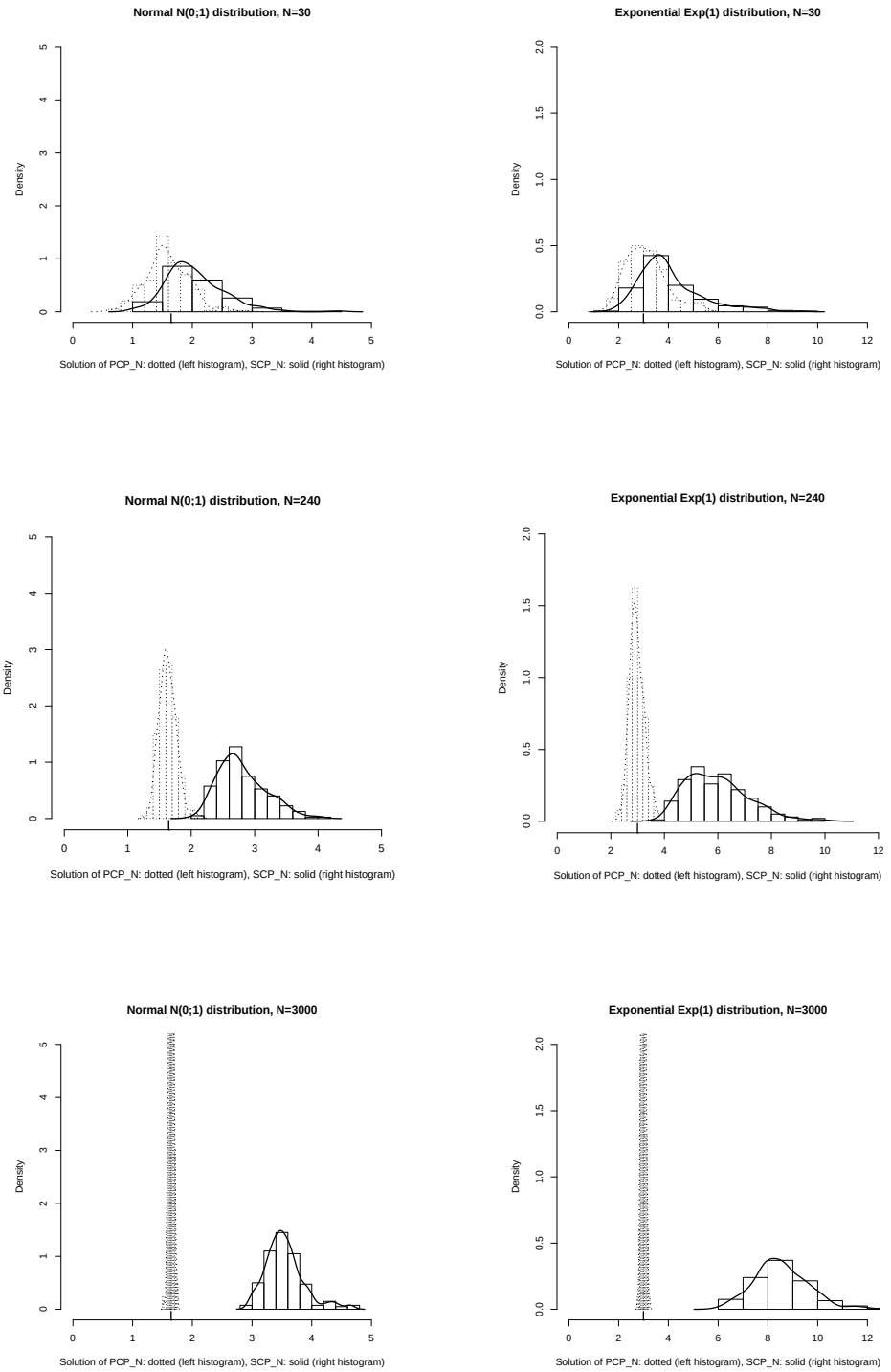


Figure 7.1: Convergence of optimal values for (SCP_N) and (PCP_N)

Chapter 8

Summary and open questions

This doctoral thesis focuses on the stability and approximations in stochastic programming. The first introducing part (Chapters 1 to 3) is devoted to present the concept of the uncertainty in mathematical programming, with presenting some of the two actually common approaches to deal with the uncertainty: the stochastic programming and the robust programming. The stochastic programming problems are presented in Chapter 2 and the concept of distance on the probability space with some prominent metrics (Wasserstein and Kolmogorov) appointed in Chapter 3. Chapter 4 presents selected known results from the theory of stability of stochastic, lead by Theorem 4.1 of Henrion and Römisch [28]. Some small modification to known results are given by Theorems 6.5 and 4.6.

Chapter 5 deals with approximations in stochastic programming problems. In detail, convergence rates of Wasserstein and Kolmogorov metrics and of associated empirical processes are studied. The theoretical results known for independent data series are extended (in numerical part of the chapter) by considering not only independent samples but also weakly-dependent ones (only the simplest M-dependence property is considered). The theoretical questions about observed results was answered only partially by Theorem 5.3.

Chapter 6 is devoted to the question of convexity in chance-constrained programming. The result of Henrion and Strugarek [30] is exploited and the result reproved with the original assumption of a structural weak-dependence property replacing the original independence of the constraint rows. The result is then connected with the stability theorem, but the obtained result still remain the result of existence characteristics only. Another extension that seems to be examined is apparent convexity property even for weakly-dependent problems, as this thesis only approximates this one by an independent problem.

Chapter 7 closing the thesis compares the stochastic programming (represented by the chance-constrained programming problem) with one of contemporary robust programming approaches (so-called randomized robust programming problem). On a simple numerical example we show a conceptual difference between the two approaches that has to be taken into account while applying one or other approach to the real-world application.

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.

[2] Artstein, Z. and Wets, R. J.-B. (1994). Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM Journal on Optimization*, 4(3):537–550.

[3] Bank, B., Guddat, J., Klatte, D., Kummer, B., and Tammer, K. (1982). *Non-Linear Parametric Optimization*. Akademie-Verlag, Berlin.

[4] Ben-Tal, A. and Nemirovski, A. (1997). Robust truss topology desing via semidefinite programming. *SIAM Journal on Optimization*, 7(4):991–1016.

[5] Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

[6] Ben-Tal, A. and Nemirovski, A. (2002). On tractable approximations of uncertain linear matrix inequalities affecetd by interval uncertainty. *SIAM Journal On Optimization*, 12(3):811–833.

[7] Bertsimas, D. and Sim, M. (2006). Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36.

[8] Birge, J. R. and Louveaux, F. (1997). *Introduction to stochastic programming*. Springer-Verlag, New York.

[9] Calafiore, G. and Campi, M. C. (2004). Decision making in an uncertain environment: the scenario based optimization approach. In Andrýsek, J., Kárný, M., and k, J. K., editors, *Multiple Participant Decision Making*, pages 99–111. Advanced Knowledge International.

[10] Calafiore, G. and Campi, M. C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming, Ser. A*, 102(1):25–46.

[11] Charnes, A. and Cooper, W. W. (1959). Chance constrained programming. *Management Science*, 6:73–79.

[12] Dai, L., Chen, C. H., and Birge, R. J. (2000). Convergence properties of two-stage stochastic programming. *Journal of Optimization Theory and Applications*, 106(3):489–509.

[13] Dantzig, G. B. (1949). Programming in a linear structure. *Econometrica*, 17:73–74.

[14] Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1:197–206.

[15] de Farias, D. P. and Van Roy, B. (2004). On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478.

[16] del Barrio, E., Giné, E., and Matrán, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *The Annals of Probability*, 27(2):1009–1071.

[17] Dentcheva, D. (1999). Regular castaing representations of multifunctions with applications to stochastic programming. *SIAM Journal on Optimization*, 10(3):732–749.

[18] Dupačová, J. (1990). Stability and sensitivity analysis for stochastic programming. *Annals of Operations Research*, 27(1):115–142.

[19] Dupačová, J., Gröve-Kuska, N., and Römisch, W. (2003). Scenario reduction in stochastic programming. An approach using probability metrics. *Mathematical Programming, Ser. A*, 95(3):493–511.

[20] Dupačová, J. and Römisch, W. (1988). Quantitative stability for scenario-based stochastic program. In sková, M. H., Lachout, P., and sek, J. A. V., editors, *Proceedings of the Prague Stochastics '98*, pages 119–124, Prague. Union of Czech Mathematicians and Physicists.

[21] El Ghaoui, L., Oustry, F., and Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal of Optimization*, 9(1):33–52.

[22] Erdoğan, E. and Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61.

[23] Fortet, R. and Mourier, E. (1953). Convergence de la rpartition empirique vers la rpartition thorique. *Annales Scientifiques de l'cole Normale Suprieure, Ser. 3*, 70(3):267–285.

[24] Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

[25] Gröwe, N. (1997). Estimated stochastic programs with chance constraints. *European Journal of Operational Research*, 101(2):285–305.

[26] Henrion, R. (2000). Qualitative stability of convex programs with probabilistic constraints. In Nguyen, V. H., Strodiot, J.-J., and Tossings, P., editors, *Optimization, Lecture Notes in Economics and Mathematical Systems*, volume 481, pages 164–180. Springer, Berlin.

[27] Henrion, R. and Möller, A. (2003). Optimization of a continuous distillation process under random inflow rate. *Computers & Mathematics with Applications*, 45(1-3):247–262.

[28] Henrion, R. and Römisch, W. (1999). Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84(1):55–88.

[29] Henrion, R. and Römisch, W. (2004). Hölder and Lipschitz stability of solutions sets in programs with probabilistic constraints. *Mathematical Programming, Ser. A*, 100(3):589–611.

[30] Henrion, R. and Strugarek, C. (2008). Convexity of chance constraints with independent random variables. *Computational Optimization and Applications*, 41(2):263–276.

[31] Houda, M. (2001). Stability and estimates in stochastic programming problems (special cases). Diploma thesis, Charles University in Prague, Faculty of Mathematics and Physics. In Czech.

[32] Houda, M. (2002). Probability metrics and the stability of stochastic programs with recourse. *Bulletin of the Czech Econometric Society*, 9(17):65–77.

[33] Houda, M. (2003). On quantitative stability in stochastic programming problems with recourse. In Houška, M., editor, *Proceedings of the 21st International Conference Mathematical Methods in Economics 2003*, pages 117–121, Prague. Czech University of Agriculture.

[34] Johnson, B. M. and Killeen, T. (1983). An explicit formula for the C.D.F. of the l_1 norm of the Brownian bridge. *The Annals of Probability*, 11(3):807–808.

[35] Kall, P. (1976). *Stochastic Linear Programming*. Springer-Verlag, Berlin.

[36] Kall, P. and Wallace, S. W. (1994). *Stochastic Programming*. Wiley, Chichester.

[37] Kaňková, V. (1990). On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika*, 26:449–461.

[38] Kaňková, V. (1993). Stability in stochastic programming - the case of unknown location parameter. *Kybernetika*, 29(1):80–101.

[39] Kaňková, V. (1994a). A note on estimates in stochastic programming. *Journal of Computational and Applied Mathematics*, 56(1-2):97–112.

[40] Kaňková, V. (1994b). A note on relationship between distribution function estimation and estimation in stochastic programming. In Lachout, P. and Víšek, J. Á., editors, *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 122–125, Prague. Academy of Sciences of the Czech Republic.

[41] Kaňková, V. (1998). A note on multifunctions in stochastic programming. In *Stochastic programming methods and technical applications, Lecture Notes in Economics and Mathematical Systems*, volume 458, pages 154–168, Berlin. Springer.

[42] Kaňková, V. and Houda, M. (2003). A note on quantitative stability and empirical estimates in stochastic programming. In Leopold-Wildburger, U., Rendl, F., and Wascher, G., editors, *Operations Research Proceedings 2002*, pages 413–418, Berlin. Springer.

[43] Kaňková, V. and Lachout, P. (1992). Convergence rate of empirical estimates in stochastic programming. *Informatica*, 3(4):497–523.

[44] Klatte, D. (1987). A note on quantitative stability results in nonlinear optimization. In Lommatzsch, K., editor, *Proceedings 19. Jahrestagung Mathematische Optimierung, Seminarbericht Nr. 90*, pages 77–86, Berlin. Humboldt University Berlin, Department of Mathematics.

[45] Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari*, 4:83–91.

[46] Mulvey, J. M., Vanderbei, R. J., and V., Z. S. (1995). Robust optimization of large-scale systems. *Operations Research*, 43(2):264–281.

[47] Pflug, G. C. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming, Ser. B*, 89:251–271.

[48] Prékopa, A. (1972). A class of stochastic programming decision problems. *Mathematische Operations forschung und Statistik*, pages 349–354.

[49] Prékopa, A. (1973). Stochastic programming models for inventory control and water storage. *Coll. Math. Soc. J. Bolyai*, 7:229–245.

[50] Prékopa, A. (1995). *Stochastic Programming*. Akadémiai Kiadó, Budapest.

[51] Prékopa, A. (2003). Probabilistic programming. In Ruszczynski, A. and Shapiro, A., editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, pages 267–352. Elsevier, Amsterdam.

[52] Prékopa, A. and Szántai, T. (1978). Flood control reservoir system design using stochastic programming. *Mathematical Programming Study*, 9:138–151.

[53] Rachev, S. (1991). *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester.

[54] Rachev, S. T. and Römisch, W. (2002). Quantitative stabilisy in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818.

[55] Rice, S. O. (1982). The integral of the absolute value of the pinned Wiener process – calculation of its probability density by numerical integration. *The Annals of Probability*, 10(1):240–243.

[56] Robinson, S. M. (1987). Local epi-continuity and local optimization. *Mathematical Programming*, 37(2):208–222.

[57] Rockafellar, R. T. (1985). Lipschitzian properties of multifunctions. *Nonlinear Analysis: Theory, Methods and Applications*, 9(8):867–885.

[58] Rockafellar, R. T. and Wets, R. J. B. (1997). *Variational Analysis*. Springer-Verlag, Berlin.

[59] Römisch, W. (2003). Stability of stochastic programming problems. In Ruszczynski, A. and Shapiro, A., editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, pages 483–554. Elsevier, Amsterdam.

[60] Römisch, W. and Schultz, R. (1991a). Distribution sensitivity in stochastic programming. *Mathematical Programming*, 50:197–226.

[61] Römisch, W. and Schultz, R. (1991b). Stability analysis for stochastic programs. *Annals of Operation Research*, 30:590–609.

[62] Römisch, W. and Schultz, R. (1993). Stability of solutions for stochastic programs with complete recourse. *Mathematics of Operations Research*, 18(3):590–609.

[63] Römisch, W. and Wakolbinger, A. (1987). Obtaining convergence rates for approximations in stochastic programming. In Guddat, J., Jongen, H. T., Kummer, B., and Nožička, F., editors, *Parametric Optimization and Related Topics*, pages 327–343. Akademie-Verlag, Berlin.

[64] Ruszczyński, A. and Shapiro, A., editors (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam.

[65] Schultz, R. (1994). Strong convexity in stochastic programs with complete recourse. *Journal of Computational and Applied Mathematics*, 56(1-2):3–22.

[66] Shapiro, A. (1994). Quantitative stability in stochastic programming. *Mathematical Programming*, 67(1):99–108.

[67] Shepp, L. A. (1982). On the integral of the absolute value of the pinned Wiener process. *The Annals of Probability*, 10(1):234–239.

[68] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

[69] Tamm, E. (1977). On g-concave functions and probability measures. *Eesti NSV Teaduste Akademia Toimetised Füüs. Mat.*, 26:376–379.

[70] Vallander, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and its Applications*, 18:784–786.

[71] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag, New York.

[72] Wallace, S. W. and Ziemba, W. T., editors (2005). *Applications of Stochastic Programming*, volume 5 of *MPS-SIAM Series on Optimization*. SIAM, Philadelphia.

[73] Wang, L. and Wang, J. (1999). Limit distribution of statistical estimates for stochastic programs with dependent samples. *ZAMM Journal of Applied Mathematics and Mechanics*, 79(4):257–266.

[74] Yoshihara, K. (1992). *Summation theory for weakly dependent sequences. Weakly Dependent Stochastic Sequences and Their Applications*, volume 1. Sanseido, Tokyo.

[75] Zolotarev, V. M. (1976). Metric distances in spaces of random variables and their distributions. *Mathematics of the USSR-Sbornik*, 30(3):373–401.

[76] Zolotarev, V. M. (1984). Probability metrics. *Theory of Probability and its Applications*, 28(2):278–302.