Ambiguous Chance Constrained Programs: Algorithms and Applications

Emre Erdoğan

Adviser: Associate Professor Garud Iyengar
Partial Sponsor: TÜBİTAK NATO-A1

Submitted in partial fulfilment of the
Requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2006

ABSTRACT

Ambiguous Chance Constrained Programs: Algorithms and Applications

Emre Erdoğan

Chance constrained problems are optimization problems where one or more constraints ensure that the probability of one or more events occurring is less than a prescribed threshold. Although it is typically assumed that the distribution defining the chance constraints are known perfectly; in practice this assumption is unwarranted. We study chance constrained problems where the underlying distributions are not completely specified and are assumed to belong to an uncertainty set $\mathcal{Q}$. We call such problems "ambiguous chance constrained problems." We focus primarily on the special case where the uncertainty set $\mathcal{Q}$ of the distributions is of the form $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$, where ρ_p denotes the Prohorov metric.

We study single and two stage ambiguous chance constrained programs. The single stage ambiguous chance constrained problem is approximated by a robust sampled problem where each constraint is a robust constraint centered at a sample drawn according to the central measure $\mathbb{Q}_0$. We show that the robust sampled problem is a good approximation for the ambiguous chance constrained problem with a high probability. This result is established using the Strassen-Dudley Representation Theorem. We also show that the robust sampled problem can be solved efficiently both in theory and in practice.

Nemirovski and Shapiro [61] formulated two-stage convex chance constrained programs and proposed an ellipsoid-like iterative algorithm for the special case where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine. We show that this algorithm extends to bi-convex $\mathbf{f}(\mathbf{x}, \mathbf{h})$ in a fairly straightforward fashion. The complexity of the solution algorithm as well as the quality of its output are functions of the radius r of the largest Euclidean ball that can be inscribed in the polytope defined by a random set of linear inequalities generated by the algorithm [61]. In this dissertation we provide some guidance for selecting r. We develop an approximation algorithm to two-stage ambiguous chance constrained programs when the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine and the extreme points of a certain "dual" polytope are known explicitly.

Contents

List of Figures

List of Tables

Notation

Sets will be denoted by calligraphic letters, e.g. $\mathcal{A}$, and $\mathcal{A}^c$ will denote the complement of the set $\mathcal{A}$. All (deterministic) vectors will be denoted by the boldface lowercase letters, e.g. $\mathbf{x}$. Random vectors and samples of random vectors will be denoted by the boldface uppercase letters, e.g. $\mathbf{H}$, and measures will be denoted by the mathematical boldface letters, e.g. $\mathbb{P}$. We will denote that a random vector $\mathbf{H}$ has distribution $\mathbb{Q}$ by $\mathbf{H} \sim \mathbb{Q}$ and a σ-algebra on a space $\mathcal{H}$ by $\mathcal{F}(\mathcal{H})$. The natural logarithm and logarithm with base 2 are denoted by $\ln(\cdot)$ and $\log(\cdot)$, respectively. The distribution of a Normal random variable with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ will be denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The norm $\|\cdot\|$ will denote the usual Euclidean norm $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{\sum_i x_i^2}$. $\mathcal{B}_r(\mathbf{y})$ will denote a Euclidean ball of radius r centered at $\mathbf{y}$, i.e. $\mathbf{B}_r(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| \leq r\}$, and the set $\mathcal{B}_r$ will denote a Euclidean ball of radius r, i.e. $\mathcal{B}_r = \mathcal{B}_r(\mathbf{y})$ for some fixed $\mathbf{y}$. Given an input $\mathbf{x} \in \mathbf{R}^n$, a separation oracle $S_{\mathcal{A}}$ for a convex set $\mathcal{A} \subset \mathbf{R}^n$ returns an affine function $L : \mathbf{R}^n \mapsto \mathbf{R}$ satisfying

$$S_{\mathcal{A}}(\mathbf{x}) = \begin{cases} L \text{ s.t. } L(\mathbf{z}) \leq 0,\ \forall \mathbf{z} \in \mathcal{A},\ L(\mathbf{x}) > 0 & \mathbf{x} \notin \mathcal{A}, \\ L \equiv \mathbf{0} & \text{otherwise} \end{cases}$$

The gradient of a function $L : \mathbf{R}^n \mapsto \mathbf{R}$ will be denoted by ∇L.

Acknowledgments

I would like to express my deepest gratitude to my advisor Garud Iyengar. His boundless support, friendship, and research ideas and interests made this work possible. Working with him was a great pleasure and he had a very important role in my career and my happy life at Columbia University.

I am deeply grateful to the IEOR faculty, especially Donald Goldfarb and the department chair Guillermo Gallego, the Ph.D. students, and the administrative stuff for their friendship, support, and help. Being a member of the IEOR community was a great experience. I am also indebted to my Ph.D. dissertation committee members Garud Iyengar, Donald Goldfarb, Ward Whitt, and Assaf Zeevi of Columbia University, and Reha Tütüncü of Goldman Sachs Asset Management.

Without my family's support, trust, and encouragement I would not be able to come this far. They deserve my gratitude also for instilling the work ethic and discipline into me. I would like to thank Leslie Smith who has been a source of my well being during my doctoral studies.

This doctoral dissertation is partially supported by TUBITAK NATO-A1 fellowship.

Chapter 1

Introduction

This doctoral dissertation is motivated by the fact that many optimization models that guarantee a significant theoretical improvement in business and service operations are still not widely used in practice. As an example, consider Markowitz's Nobel Prize winning work on the mean-variance framework for portfolio optimization [55, 56]. This framework is a quadratic optimization program and many influential financial models such as the CAPM are direct consequences of this model. However, "in practice it is an error prone procedure that often results in error-maximized and investment-irrelevant portfolios" [57]. The main reason for this behavior is that the solution of optimization models is very sensitive to uncertainty in the input data. Since models are approximations of "real" life problems and, moreover, the values defining the model parameters are set using statistical estimation procedures, eliminating uncertainty is almost impossible. Hence, the sensitivity of optimization algorithms to data perturbation has to be adequately addressed if optimization models are to be effective in solving real life problems.

1.1 Optimization under uncertainty

Optimization under uncertainty dates back to 1950s [19] and several strategies have been proposed to address parameter uncertainty in optimization problems. To summarize these approaches consider the following generic optimization problem:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & f(\mathbf{x},\mathbf{h}) \in \mathcal{C}, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{1.1}$$

where $\mathbf{x} \in \mathbf{R}^n$ is the decision vector, $\mathbf{h} \in \mathbf{R}^m$ is the problem parameter, $\mathcal{X} \subseteq \mathbf{R}^n$ is a convex set, $\mathcal{C} \subseteq \mathbf{R}^l$ is a set, and $f(\cdot, \mathbf{h}) : \mathbf{R}^n \mapsto \mathbf{R}^l$ is convex in $\mathbf{x}$ for each fixed $\mathbf{h}$. We will assume that the exact value of the parameter $\mathbf{h}$ is not known, i.e., the source of uncertainty in the generic optimization model (1.1) is the parameter $\mathbf{h}$.

The deterministic optimization approach to solving optimization problems computes a solution to (1.1) assuming that the problem parameters $\mathbf{h}$ are known and fixed. In practice, however, the parameters $\mathbf{h}$ are the result of some measurement or estimation process, and are, therefore, never known for certain. This uncertainty is of serious concern in applications because solutions of optimization problems are often very sensitive to fluctuations in the problem parameters. This phenomenon is well documented in several different application areas [5, 42, 18].

Sensitivity analysis [10] is a post-optimization tool that can be used to assess the effect of parameter uncertainty. This approach is only attractive when the underlying uncertainty is "small" in an appropriate sense. However, in many real life applications the sensitivity analysis is not effective and more sophisticated techniques are needed. The two major techniques that are currently popular are: robust optimization and stochastic programming. Both of these models have different strengths and weakness. Therefore, it is very hard to conclude that one model is better than the other one. The choice of the technique depends on the particular problem and the analyst.

Robust optimization: Robust optimization was introduced in Ben-Tal and Nemirovski [6, 7, 8]. There is also a parallel literature on robust formulations of optimization problems originating from the robust control literature [28, 30].

In this framework, the uncertain parameters $\mathbf{h}$ are assumed to belong to a bounded uncertainty set $\mathcal{H}$, possibly given by the confidence regions around the point estimates of the parameters. The robust optimization problem corresponding to the nominal problem (1.1) is defined to ensure that any optimal solution to the robust problem is feasible for all the parameters $\mathbf{h} \in \mathcal{H}$. Thus, the robust optimization problem is given by:

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & f(\mathbf{x}, \mathbf{h}) \in \mathcal{C}, \quad \forall\ \mathbf{h} \in \mathcal{H}, \\ & \mathbf{x} \in \mathcal{X}. \end{array} \tag{1.2}$$

Since, in many applications the uncertainty set $\mathcal{H}$ is given by the confidence region around the point estimates of the parameters, problem (1.2) allows one to provide probabilistic

guarantees on the performance of the optimal solutions.

The robust problem (1.2) is solved by reformulating the semi-infinite constraints, $f(\mathbf{x}, \mathbf{h}) \leq 0$, for all $\mathbf{h} \in \mathcal{H}$, as a finite collection of constraints. Such a reformulation is only possible when the uncertainty set $\mathcal{H}$ and the function $f(\mathbf{x}, \mathbf{h})$ satisfy some regularity conditions. See [6, 8, 9] for robust formulations that can be solved efficiently. Even when the reformulation is possible, the resulting problem is typically harder than the nominal problem (1.1) ([9] proposes a new framework wherein the robust problem remains in the same complexity class as the nominal problem). In general, however, the robust problem is intractable.

Another criticism of the robust framework is that it gives the same "weight" to all perturbations $\mathbf{h} \in \mathcal{H}$. Also, in certain applications one might have the flexibility of violating the constraints corresponding to a small fraction of the set $\mathcal{H}$ and this flexibility can drastically reduce the solution cost (see Chapter 2.) An alternative optimization framework that mitigates this criticism to some extent is called *stochastic programming*, in particular, *chance constrained programming*.

Stochastic programming: In the stochastic programming [68] approach the parameter $\mathbf{H}$, source of the uncertainty, is assumed to be random with a known distribution $\mathbb{Q}$, i.e., $\mathbf{H} \sim \mathbb{Q}$. There are a number of different formulations of stochastic programs. For simplicity we will consider the following version:

$$\begin{aligned} \min \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{E}_{\mathbb{Q}}[g(\mathbf{x}, \mathbf{H})] \leq \alpha, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{1.3}$$

where, $\mathcal{X} \subseteq \mathbf{R}^n$, $\alpha \in \mathbf{R}$ is a constant, and $\mathbf{E}_{\mathbb{Q}}$ denotes the expectation operator under the probability distribution $\mathbb{Q}$. Note that by introducing a new variable if necessary, one can assume that the objective function is linear and independent of the uncertain parameter.

Depending on the decision variables, stochastic programming models can be either single-stage or multi-stage decision problems. The variables in single-stage models correspond to those decisions that must be made here-and-now and cannot depend on the future observations or partial realizations of the random parameters. Multi-stage models can include wait-and-see decisions that are made after some/all of the random parameters are realized.

Stochastic programming models have a very rich literature and have many applications

in real life, including electric power generation [58], finance [21, 24], telecommunication networks [71, 78], oil industry [22], capacity planning [31]. References to many more applications can be found in [68, 62].

Chance constrained programming: Chance constrained program is a special case of the stochastic program (1.3) where the function $g(\cdot,\cdot)$ has the form

$$g(\mathbf{x},\mathbf{h}) = \mathbf{1}_{\{\mathbf{h}\in\mathcal{H}: f(\mathbf{x},\mathbf{h})\notin\mathcal{C}\}}$$

and $\alpha = \epsilon$, where $\mathbf{1}_{\mathcal{A}}$ is the indicator function of a set $\mathcal{A}$ and $0 \leq \epsilon \leq 1$ is a given constant. In particular, the chance constrained program is:

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbb{Q}(\mathbf{H}\in\mathcal{H} : f(\mathbf{x},\mathbf{H})\notin\mathcal{C}) \leq \epsilon, \\ & \mathbf{x}\in\mathcal{X}. \end{array} \tag{1.4}$$

The parameter ϵ controls the probability that the optimal solution of (1.4) violates the constraints – as $\epsilon \downarrow 0$ the chance-constrained problem starts to resemble the robust problem (1.2). Therefore, in our formulation the chance constrained formulation is less conservative when compared to robust optimization.

Although chance-constrained problems have a long history dating back to at least the work of Charnes and Cooper [16], they have not found wide applicability. This is primarily because computing the optimal solution for chance-constrained problems is extremely difficult. To begin with just evaluating $\mathbb{Q}(\mathbf{H}\in\mathcal{H} : f(\mathbf{x},\mathbf{H})\notin\mathcal{C})$ involves a multidimensional integral that becomes difficult as the number of parameters grows. Moreover, even if the function $f(\mathbf{x},\mathbf{h})$ is convex (or even linear) in $\mathbf{x}$ the feasible set $\mathcal{X}_\epsilon(\mathbb{Q})$ of (1.4) is in general not convex. A detailed discussion of the chance-constrained programs and, more generally, stochastic programs can be found in [68]. In Chapter 3 we will give a brief summary of chance constrained programming and some recent solution techniques.

Ambiguous chance-constrained programs: In the chance constrained program (1.4) we implicitly assumed that the distribution $\mathbb{Q}$ of the random parameters $\mathbf{H}$ is fixed. A major criticism raised against chance constrained problems and, more generally, stochastic programs is that, in practice, the underlying probability measure $\mathbb{Q}$ is never known exactly. Just as the point estimates for the parameters, the distribution $\mathbb{Q}$ is also estimated from data or measurements, and is, therefore, known only to within some error, i.e. the measure

$\mathbb{Q} \in \mathcal{Q}$ where $\mathcal{Q}$ is a set of measures. One approach to deal with the uncertainty in the probability measure $\mathbb{Q}$ is to consider the following optimization model.

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \bar{\mathcal{X}}_\epsilon, \end{array} \tag{1.5}$$

where

$$\bar{\mathcal{X}}_\epsilon = \Big\{\mathbf{x} \in \mathcal{X} : \mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon,\ \forall \mathbb{Q} \in \mathcal{Q}\Big\}. \tag{1.6}$$

We will call (1.5) an *ambiguous chance-constrained problem*. A problem of the form (1.5) has two sources of uncertainty: the distribution $\mathbb{Q}$ of the parameter $\mathbf{H}$ is uncertain, and, given a measure $\mathbb{Q}$, the particular realization of the parameter $\mathbf{H}$ is also uncertain. In the decision theory literature the uncertainty in the distribution is referred to as *ambiguity*, and hence the name for the problem.

Modeling ambiguity and its consequence has been received attention in several different fields. The minimax formulation has a long history in stochastic programing [80, 11, 26, 27, 47, 74, 72, 73]. Ruszczynski and Shapiro [69] show the equivalence between minimax stochastic programming and minimizing a convex risk measure [2, 37] of the second-stage cost. [70] extends the minimax approach to a multiperiod setting. The study of ambiguity in Economics began with the work of Gilboa and Schmeidler [40]. This work was extended to a multiperiod setting by Hansen and Sargent [44] and Epstein and his co-authors [17, 32, 33]. Ambiguity in the context of Markov decision processes was independently investigated by Iyengar [46] and El Ghaoui and Nilim [29].

The issue of ambiguity of measures was also raised in [15] where the authors considered a finite uncertainty set $\mathcal{Q}$. They proposed a solution strategy where one samples from *all* of these measures and showed that the samples from different measures "help" each other. In contrast, we consider the case where $\mathcal{Q}$ is uncountably infinite and we draw samples from *only* the central measure $\mathbb{Q}_0$.

1.2 Contributions

We develop approximation algorithms to single and two stage ambiguous chance constrained versions of program (1.5). Since techniques used in these two decision models are significantly different, we will consider them separately. The results of this dissertation have appeared in [34, 36].

1.2.1 Single stage ambiguous chance constrained programs

The single stage ambiguous chance constrained program is a special case of the problem (1.5) where the set $\mathcal{C}$ is the non-positive real line, i.e.,

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbb{Q}(\mathbf{H} : f(\mathbf{x},\mathbf{H}) > 0) \leq \epsilon, \ \forall \mathbb{Q} \in \mathcal{Q}, \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.7}$$

Our contributions to the literature of single stage ambiguous chance constrained programs are presented in Chapter 5 and summarized below.

We consider uncertainty sets $\mathcal{Q}$ of measures that are of the form $\mathcal{Q} = \{\mathbb{Q} : \rho(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$ where $\rho(\cdot,\cdot)$ denotes a suitable metric between probability measures, i.e. the uncertainty sets are "balls" centered around a central measure $\mathbb{Q}_0$. We approximate the ambiguous chance-constrained problem (1.7) by a *robust sampled problem* defined as follows

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & f(\mathbf{x},\mathbf{z}) \leq 0, \quad \forall \ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, \quad i = 1,\ldots,N, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{1.8}$$

where $\mathbf{H}_i^0$, $i = 1,\ldots,N$, denote IID samples drawn according to the central measure $\mathbb{Q}_0$ and the norm $\|\cdot\|$ on the $\mathcal{H}$ space is related to the probability metric $\rho(\cdot,\cdot)$ (details are given in Chapter 4). Results in [9] imply that for a large class of constraint functions $f(\mathbf{x},\mathbf{h})$ and suitably defined norms $\|\cdot\|$ the robust sampled problem (1.8) is in the same complexity class as the non-robust version of the problem (1.8), i.e., $\beta = 0$.

We combine results from Computational Learning Theory with results for *coupling* of random variables [77] to compute upper bounds on the number of samples N required to ensure that the feasible set of the robust sampled problem (1.8) is contained in $\mathcal{X}_\epsilon(\mathbb{Q})$, for any $\mathbb{Q} \in \mathcal{Q}$, with a high probability. This bound depends on the Vapnik-Chervonenkis (VC) dimension of the function $f(\mathbf{x},\mathbf{h})$.

We use coupling to extend the results of Calafiore and Campi [14, 15] to the ambiguous chance constrained problems, i.e., we compute upper bounds on the number of samples required to ensure that the optimal solution of the robust sampled problem (1.8) is contained in $\bar{\mathcal{X}}_\epsilon$ (1.6) with a high probability. The bound in this case depends on the number of "support" constraints, and is independent of the VC dimension of $f(\mathbf{x},\mathbf{h})$.

1.2.2 Two stage ambiguous and convex chance constrained programs

The model (1.7), while quite general in its ability to model constraints, is limited to the so-called single stage problems where decisions must be made before the uncertain parameter $\mathbf{H}$ is revealed. A natural extension is to consider two-stage problems where one has to commit to the first stage decision $\mathbf{x}$ *before* the realization of the uncertain parameter $\mathbf{H}$, and the second stage decision variable $\mathbf{v}$ can be chosen *after* observing the parameter $\mathbf{H}$. A prototypical example of a two-stage problem is the network design where the first stage variables are the capacities on the arcs and the second stage variables are the routing decisions. The simplest two-stage chance constrained problem is also a special case of the problem (1.4) and given as

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \left\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}\big(\mathbf{H} : \exists \mathbf{v} \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{y}, \mathbf{H})\big) > 1 - \epsilon\right\}, \end{array} \tag{1.9}$$

where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h}) : \mathbf{R}^n \times \mathbf{R}^m \mapsto \mathbf{R}^l$ is bi-affine, $\mathbf{v} \in \mathbf{R}^k$, and $\mathbf{W} \in \mathbf{R}^{l \times k}$. Since the matrix $\mathbf{W}$ does not depend on the realization of $\mathbf{H}$, problems of the form (1.9) are said to have a fixed recourse. Thus, (1.9) is a two-stage linear chance constrained problem with a fixed recourse. This model was introduced by Nemirovski and Shapiro [61].

Our contributions to the literature of two stage chance constrained programs are as follows. We extend the iterative algorithm proposed in [61] to solve *non-ambiguous* two-stage chance constrained problems where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ (see (1.9) for details) is bi-convex, i.e. convex in one variable when the other variable is held constant. We still assume a constant recourse matrix $\mathbf{W}$. This fairly straightforward extension is discussed in § 3.2 and the results in this section can increase the practical value of (1.9) significantly. For example, second-order cone constraint functions (see § 3.2) are bi-convex and have many applications in engineering and finance [42, 41, 53].

Since the iterative algorithm proposed in [61] closely resembles the ellipsoid algorithm, the number of iterations required to compute a feasible solution as well as the quality of the solution are functions of the radius r of a ball with the largest volume contained within the feasible set of the problem. We show that the value of the parameter r is determined by the degree of *robust* feasibility of the chance constrained problem. This is similar to the results relating the condition number of optimization problems to the complexity of solving them to optimality [65, 66, 38]. This result is proved in § 3.2 and discussed in § 3.2.1.

We formulate the ambiguous two-stage chance constrained problem and modify the

algorithm proposed in [61] to compute a good solution for this problem. Our extension is limited to the special case where the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine (i.e. the case considered in [61]) and the extreme points of a certain "dual" polytope are known explicitly. We extend all the known results for chance constrained problems to the ambiguous setting in Chapter 6.

Chapter 2

Chance Constrained Programming

General chance constrained problems (i.e., $f(\mathbf{x}, \mathbf{H})$ is possibly non-convex) have a very rich literature. In Chapter 1, we briefly introduced the single-stage and two-stage convex chance constrained programs. In this chapter, we will give examples and present solution methods. We will also discuss one of the main weakness of this approach and motivate the ambiguous chance constrained programs. See [63, 68] for a more information on chance constrained programs.

As we have mentioned, depending on the type of the decisions, the chance constrained programs are called either single stage or multi-stage chance constrained programs. Variables in the single-stage models correspond to those decisions that must be made here-and-now and cannot depend on observations or partial realizations of the random parameters. In the single stage models the set $\mathcal{C}$ in (1.4) is the non-positive real line. In particular, the single stage chance constrained program is:

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \mathcal{X}_\epsilon(\mathbb{Q}) = \Big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : f(\mathbf{y}, \mathbf{H}) > 0) \leq \epsilon\Big\}, \end{array} \tag{2.1}$$

where $f(\cdot, \mathbf{H}) : \mathbf{R}^n \mapsto \mathbf{R}$ is convex in $\mathbf{x}$ for each fixed $\mathbf{H}$. Since a collection of convex constraints $f_i(\mathbf{x}, \mathbf{H}) \leq 0$, $i = 1, \ldots, p$, can be formulated as a single convex constraint $f(\mathbf{x}, \mathbf{H}) = \max_{1 \leq i \leq p}\{f_i(\mathbf{x}, \mathbf{H})\} \leq 0$, restricting the range of constraint function $f(\cdot, \mathbf{H})$ to $\mathbf{R}$ does not represent any loss of generality.

We mentioned in Chapter 1 that in certain applications one might have the flexibility of violating the constraints corresponding to a small fraction of the set $\mathcal{H}$ and this flexibility of choosing a positive ϵ, even as small as $\epsilon = 10^{-12}$, can improve the objective function value

drastically. To see this let us consider the following simple chance constrained problem:

$$\begin{array}{ll} \max & x \\ \text{s.t.} & \mathbb{Q}(\mathbf{H} : x\mathbf{H} - 50 > 0) \leq \epsilon, \\ & x \geq 0, \end{array} \tag{2.2}$$

where $\mathbf{H}$ is a standard Normal random variable, $\mathbf{H} \sim \mathcal{N}(0,1)$. It is easy to see that feasible set of the problem (2.2) is $\{x \in \mathbf{R} : \Phi^{-1}(1-\epsilon)x \leq 50\}$ (see below for details), where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. When $\epsilon = 0$, the problem (2.2) has the optimal solution $x^* = 0$, which is the only feasible solution. However, for $\epsilon = 10^{-12}$, the optimal solution is $x^* = 7.1078$. As one can see from this example that we were able to improve the value optimal solution significantly without loosing much feasibility.

The model (2.1), while quite general in its ability to model constraints, is limited to the so-called single stage problems where decisions must be made before the uncertain parameter $\mathbf{H}$ is revealed. A natural extension is to consider two-stage problems where one has to commit to the first stage decision $\mathbf{x}$ *before* the realization of the uncertain parameter $\mathbf{H}$, and the second stage decision variable $\mathbf{v}$ can be chosen *after* observing the parameter $\mathbf{H}$. A prototypical example of a two-stage problem is the network design where the first stage variables are the capacities on the arcs and the second stage variables are the routing decisions. The simplest two-stage chance constrained problem is given by

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \left\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}\big(\mathbf{H} : \exists \mathbf{v} \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{y}, \mathbf{H})\big) > 1 - \epsilon\right\}, \end{array} \tag{2.3}$$

where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h}) : \mathbf{R}^n \times \mathbf{R}^m \mapsto \mathbf{R}^l$ is bi-affine, $\mathbf{v} \in \mathbf{R}^k$, and $\mathbf{W} \in \mathbf{R}^{l \times k}$. Since the matrix $\mathbf{W}$ does not depend on the realization of $\mathbf{H}$, problems of the form (2.3) are said to have a fixed recourse. Thus, (2.3) is a two-stage linear chance constrained problem with a fixed recourse. This model was introduced by Nemirovski and Shapiro [61].

The two-stage model (2.3) is a special case of multi-stage chance constrained programs where at each stage a portion of the data is revealed and decisions at that stage depend only on the data realized at the previous stages. It is known that two-stage and multi-stage chance constrained programs are NP-hard [61, 60]. Therefore, the best one can hope is to approximate these problems within a reasonable error. In [61, 36], approximations to the two-stage models are given (see also § 3.2). However, we are not aware of any good approximation scheme to general multistage chance constrained programs.

The chance constraints are also used in stochastic integer programs which have many applications including telecommunications, water resource management, bond portfolios. For solution methods and more applications see [68]. We will not consider stochastic integer programs in this dissertation.

2.1 Applications of Chance Constrained Models

Chance constrained programs can be used to model many applications in engineering, management science, and finance. In this section, we will show how chance constrained approach can be used in portfolio management with a Value at Risk (VaR) constraint, index tracking, supply chain, and network design. For more applications see [63, 68]. Also, in Chapter 7 we give numerical examples.

2.1.1 Portfolio selection with VaR constraints

Let us consider a discrete time market with n assets and assume that assets returns $\mathbf{H}$ are random with a distribution $\mathbb{Q}$, i.e., $\mathbf{H} \sim \mathbb{Q}$ and $\mathbf{E}[\mathbf{H}] = \boldsymbol{\mu}$. For simplicity, we will assume that the mean returns $\boldsymbol{\mu}$ are known.

Let the ith entry of the decision variable $\mathbf{x}$ denote the fraction of wealth invested in the ith asset. Suppose that we want to choose a portfolio with the maximum expected return among the ones that satisfy a Value at Risk (VaR). In our applications, we will say that a portfolio satisfies the VaR constraint if the portfolio has a return greater than a given constant η with a probability at least $1 - \epsilon$ for a pre-specified ϵ.

Since the random return and the expected return of a portfolio $\mathbf{x}$ are given by $\mathbf{H}^T\mathbf{x}$ and $\mathrm{E}[\mathbf{H}]^T\mathbf{x} = \boldsymbol{\mu}^T\mathbf{x}$, respectively, the portfolio selection problem with the VaR constraint is:

$$\begin{aligned} \max \quad & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbb{Q}(\mathbf{H} : \mathbf{H}^T\mathbf{x} \geq \eta) \geq 1 - \epsilon, \\ & \textstyle\sum_{i=1}^n \mathbf{x}_i = 1. \end{aligned}$$

The constraint $\sum_{i=1}^n \mathbf{x}_i = 1$ is a normalization constraint which states that the sum of the fractions $\mathbf{x}_i$ of the total wealth should be equal to 1. By setting $f(\mathbf{x}, \mathbf{H}) = -\mathbf{x}^T\mathbf{H} + \eta$ and $\mathcal{X} = \{\mathbf{x} \in \mathbf{R}^n : \sum_{i=1}^n \mathbf{x}_i = 1\}$, one can see that the portfolio section problem with a VaR constraint can reformulated a chance constrained problem (2.1).

2.1.2 Index Tracking

Index tracking is a form of passive portfolio management where the goal is to manage a portfolio with a return that is highly correlated with the return of a given index, e.g. the S&P 500, the NASDAQ, etc. Performance of the index tracking portfolio is measured by its return relative to that of the index (benchmark). Since investing in all of the assets that constitute the index is impractical, index tracking strategies attempt to reproduce return of the index by carefully selecting a subset of assets.

To model the index tracking problem as a chance constrained program, let us consider a discrete time market with n assets and a benchmark b. Let $\mathbf{H}$ be a random vector denoting returns of the assets, $\mathbf{B}$ be a random variable denoting the benchmark's return, and $\mathbb{Q}$ be their joint probability distribution.

In index tracking, we ideally would like to choose a portfolio $\mathbf{x}$ with the same return as that of benchmark. However, since the returns are random it is impossible to replicate the benchmark's return perfectly. Therefore, our goal is to choose a portfolio $\mathbf{x}$ with a tracking error less than a pre-specified positive constant η with a probability at least $1-\epsilon$. The tracking error of a portfolio $\mathbf{x}$ is defined as the deviation of its return $\mathbf{H}^T\mathbf{x}$ from the benchmark's return $\mathbf{B}$, i.e., $|\mathbf{H}^T\mathbf{x} - \mathbf{B}|$, where $|\cdot|$ denotes the absolute value. A chance constrained model that can be used to construct the index tracking portfolio defined above is:

$$\begin{aligned} \min \quad & \eta \\ \text{s.t.} \quad & \mathbb{Q}((\mathbf{B},\mathbf{H}) : |\mathbf{H}^T\mathbf{x} - \mathbf{B}| \leq \eta) > 1-\epsilon, \\ & \textstyle\sum_{i=1}^{n} \mathbf{x}_i = 1. \end{aligned}$$

It is easy to see that, this is also a single stage chance constrained program.

2.1.3 Supply chain

Consider a single period inventory management problem with n items. Suppose that demands for these items are given by the random vector $\mathbf{H} \sim \mathbb{Q}$ and we have to order these items to before the realization of the random demand $\mathbf{H}$. Our goal is to find optimal order quantities of these items that minimize our total loss, which is defined as the total cost (possibly including order cost, inventory cost, backorder cost, etc.) minus total revenue obtained from the sales, subject to some capacity or other constraints.

To model this problem as a chance constrained program, let i-th component of the decision vector $\mathbf{x}$ denote order quantity of the i-th item and the function $C(\mathbf{x},\mathbf{h})$ denote

the total loss when we order $\mathbf{x}$ and the realized demand is $\mathbf{h}$. Suppose the order quantity $\mathbf{x} \in \mathcal{X}$ for some convex set $\mathcal{X}$. Our goal is to choose an order quantity $\mathbf{x}$ with the total cost at most η with a probability at least $1-\epsilon$. Below is a more general chance constrained program that can be used to solve this problem. Note that this problem is also minimizing the total cost threshold η.

$$\begin{aligned} \min \quad & \eta \\ \text{s.t.} \quad & \mathbb{Q}(\mathbf{H} : C(\mathbf{x}, \mathbf{H}) \leq \eta) > 1 - \epsilon \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

2.1.4 Network Design

Suppose we want to design a directed network to satisfy demands of a number of nodes by shipping goods from a source node. Assume that the source node has infinite capacity and let the vector $\mathbf{D} \geq \mathbf{0}$ denote the random demands of the nodes. Let the ith element of the strictly positive vector $\mathbf{c}$ denote the cost of adding one unit capacity to the ith arc. For simplicity we ignore the set up costs and integrality constraints. Our goal is to find the minimum cost arc capacities while satisfying the demands of the sink nodes with probability at least $1-\epsilon$.

The network design problem for a fixed (non-random) demand vector $\mathbf{d}$ is given by

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{u} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{v} \geq \mathbf{d}, \\ & \mathbf{u} - \mathbf{v} \geq \mathbf{0}, \\ & \mathbf{u}, \mathbf{v} \geq \mathbf{0}, \end{aligned} \tag{2.4}$$

where $-\mathbf{A}$ denotes the node-arc incidence matrix of the network, the variable $\mathbf{u}$ denotes the capacities on the arcs, and the variable $\mathbf{v}$ denotes the flow on the network.

Once the network is constructed, i.e. the capacities $\mathbf{u}$ are installed on the arcs, we assume that it will be used over a reasonably long period over which the demand $\mathbf{d}$ can change. We model variation in demand by assuming that it is random. In particular, we assume that the random demand has a distribution $\mathbb{Q}$, i.e., $\mathbf{D} \sim \mathbb{Q}$. Note that the flow conservation constraints $\mathbf{A}\mathbf{v} \geq \mathbf{d}$ in (2.4) are formulated as inequalities instead of equalities as is usually the case – this is necessary to accommodate random demands $\mathbf{D}$.

Although the capacity cannot, typically, be altered over the life of the network, the flow $\mathbf{v}$ is chosen *after* observing the realization of the demand. Thus, the network design problem

is a two-stage optimization problem: the capacities $\mathbf{u}$ are the first stage variables, and the flows $\mathbf{v} \in \mathbf{R}^n$ are the second stage variables. The objective of the two-stage optimization problem is to find a minimal cost capacity allocation $\mathbf{u}$ that guarantees that at least $(1-\epsilon)$-fraction of the random demand $\mathbf{D}$ can be feasibly routed in the resulting network. One can transform this network design problem into a chance constrained program of the form (2.3) by setting

$$\mathbf{x} = \mathbf{u}, \quad \mathbf{H} = \mathbf{D}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{A} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}, \quad f(\mathbf{x}, \mathbf{H}) = \begin{bmatrix} \mathbf{H} \\ -\mathbf{x} \\ \mathbf{0} \end{bmatrix}, \quad \mathcal{X} = \mathbf{R}^n_+.$$

2.2 Solution Methods

Solving general chance constrained problems is computationally very difficult. The main difficulty comes from the fact that the feasible set of the program (1.4) is, in general, not convex and/or solution methods involve numerically computing many multidimensional integrals. Consequently, the solution techniques do not scale well as the number of parameters grows.

In this section, we will give a brief overview of the available solution techniques for the single stage (2.1) and the two stage (2.3) chance constrained programs. The solution methods that form the basis of the results in this dissertation will be introduced in this section and discussed in detail in Chapter 3. Since the techniques used in the single and the two stage problems are different, we will deal with them separately. We begin with the single stage programs.

Prekopa [63, 68] shows that when the random variable $\mathbf{H}$ has a log-concave probability distribution function and the constraint function $f(\mathbf{x}, \mathbf{h})$ satisfies some regularity conditions than the single stage chance constrained problem (2.1) is a convex problem. However, this does not mean that the problem (2.1) can be solved efficiently.

Efficient solution methods are known only for some restricted classes of single stage chance constrained problems. For example, when $f(\mathbf{x}, \mathbf{h})$ is separable or bi-affine and the underlying distribution $\mathbb{Q}$ is elliptical *or* symmetric and log-concave the chance constrained problem (2.1) is a convex program [51, 45, 48]. These classes of probability distributions include multivariate Normal, uniform distributions over symmetric, compact, convex sets, and multivariate student and exponential distributions as special cases. When the under-

lying distribution is elliptically symmetric and one is able to compute the quantile function analytically, the chance constrained problem (2.1) can be reformulated as a deterministic problem which can be solved efficiently.

These approaches use properties of the distributions and structure of the function $f(\mathbf{x}, \mathbf{h})$ to reformulate the chance constrained problems as a deterministic program. We will illustrate this idea when the underlying distribution is Normal with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, i.e., $\mathbf{H} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $f(\mathbf{x}, \mathbf{h}) = a + \mathbf{x}^T\mathbf{h}$, where $a \in \mathbf{R}$ is a constant [63, 68]. The chance constraint $\mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) \leq 0) \geq 1 - \epsilon$ can be reformulated as a deterministic constraint as follows:

$$\begin{aligned} 1 - \epsilon \leq \mathbb{Q}(\mathbf{H} : a + \mathbf{x}^T\mathbf{H} \leq 0) &= \mathbb{Q}(\mathbf{Z} : \mathbf{x}^T\boldsymbol{\mu} + \sqrt{\mathbf{x}^T\boldsymbol{\Sigma}\ \mathbf{x}}\mathbf{Z} \leq -a) \\ &= \mathbb{Q}(\mathbf{Z} : \mathbf{Z} \leq \frac{-a - \mathbf{x}^T\boldsymbol{\mu}}{\sqrt{\mathbf{x}^T\boldsymbol{\Sigma}\ \mathbf{x}}}) \\ &= \Phi(\frac{-a - \mathbf{x}^T\boldsymbol{\mu}}{\sqrt{\mathbf{x}^T\boldsymbol{\Sigma}\ \mathbf{x}}}), \end{aligned}$$

where $\mathbf{Z} \sim \mathcal{N}(0, 1)$ and $\Phi(\cdot)$ is the cumulative distribution of the standard normal random variable. Therefore, we have

$$1 - \epsilon \leq \mathbb{Q}(\mathbf{H} : a + \mathbf{x}^T\mathbf{H} \leq 0) \iff \Phi^{-1}(1 - \epsilon)\sqrt{\mathbf{x}^T\boldsymbol{\Sigma}\ \mathbf{x}} \leq -a - \mathbf{x}^T\boldsymbol{\mu} \tag{2.5}$$

It is easy to see that the constraint (2.5) is a convex constraint, in fact a second-order cone constraint, when $\epsilon \leq 0.5$. Hence, the chance constrained program (2.1) can be reformulated as a convex deterministic program when $\mathbf{H} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $f(\mathbf{x}, \mathbf{h}) = a + \mathbf{x}^T\mathbf{h}$. The formulations in [51, 45, 48] are extensions of this idea.

Recently there has been a renewed interest in the special case of convex chance constrained problems. One approximates the convex chance constrained problem (2.1) by the sampled optimization problem

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{Y}[\mathbf{H}_{1,N}] = \{\mathbf{y} \in \mathcal{X} \mid f(\mathbf{y}, \mathbf{H}_i) \leq 0, i = 1, ..., N\}, \end{aligned} \tag{2.6}$$

where $\mathbf{H}_i$, $i = 1, \ldots, N$, are N IID samples from $\mathbb{Q}$. de Farias and Van Roy [20] consider the special case where $f(\mathbf{x}, \mathbf{h}) = \mathbf{h}^T\mathbf{x} + a$ and use results from Computational Learning Theory [1, 49, 79] to show that a sample size of

$$N \geq \frac{4n}{\epsilon} \log\left(\frac{12}{\epsilon}\right) + \frac{4}{\epsilon} \log\left(\frac{2}{\delta}\right),$$

where $\log(\cdot)$ denotes the logarithm with base 2, ensures that the set of decision vectors feasible for the sampled problem (2.6) is contained in $\mathcal{X}_\epsilon(\mathbb{Q})$, the feasible set of (2.1), with a probability at least $1-\delta$. Thus, in this sampling based method there are two possible sources of errors: with a probability δ, the feasible set of (2.6) (and consequently, the optimal solution of (2.6)) may not be contained in $\mathcal{X}_\epsilon(\mathbb{Q})$; and, in the event that this is not the case, the feasible points of (2.6) can still violate the constraint $f(\mathbf{x}, \mathbf{H}) \leq 0$ with a probability at most ϵ.

Erdoğan and Iyengar [34] show a similar bound for general convex constraints with the constant n replaced by the Vapnik-Chervonenkis (VC) dimension d_f [1, 49, 79] of the constraint function. Calafiore and Campi [14, 15] show that for general convex functions $f(\mathbf{x}, \mathbf{h})$ when

$$N \geq \frac{2n}{\epsilon} \ln\left(\frac{2}{\epsilon}\right) + \frac{2}{\epsilon} \ln\left(\frac{1}{\delta}\right) + 2n,$$

where $\ln(\cdot)$ denotes the natural logarithm, the optimal solution of the sampled problem (2.6) is feasible for (2.1) with a probability at least $1-\delta$. On the one hand, this bound is weak in the sense that it is only valid for the optimal solution, and *not* the entire feasible set. On the other hand, since the VC dimension d_f of a class of functions can be orders of magnitude larger than the problem dimension n, the number of samples needed to ensure that the optimal solution is feasible for (2.1) with a high probability can be orders of magnitude smaller. The result in [14, 15] was proved using a fundamental fact that the optimal solution of a convex program is "supported" by at most n constraints. Since the results by Calafiore and Campi [14, 15] and de Farias and Van Roy [20] form the basis of our results related to the single stage ambiguous chance constrained programs, we will review their work in detail in Chapter 3.

Note that these results only provide *upper* bounds for the number of samples, i.e. only a sufficient condition. Thus, a natural question of the quality or "tightness" of the approximation arises. Recently, Nemirovski and Shapiro [61, 59] established logarithmically separated upper and lower bounds on the number of samples required to approximate a convex chance constrained problem when the measure $\mathbb{Q}$ has a certain concentration-of-measure property. See § 3.2 for details.

Clearly, the two-stage chance constrained program (2.3) is much harder than the single stage version (2.1). It is known that unless the rows of the constraint $\mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{H})$ is independent or $\mathbf{f}(\mathbf{x}, \mathbf{H})$ is separable, the problem (2.3) is NP-hard [61]. Dentcheva et al [23] approximate (2.3) via efficient points when the random vector $\mathbf{H}$ has a discrete distribution.

The computational effort required to solve (2.3) by this approach increases exponentially with the dimension of the parameter vector $\mathbf{H}$ and, therefore, becomes prohibitive very quickly. The chance constrained set in (2.3) can be approximated by hypergraph-based techniques [13]. However, computing accurate bounds using the hypergraph-based approach requires constructing multi-trees to estimate the probability of intersections of sets. Although the hypergraph-based approach is promising, we are not aware of any solution method that utilizes this approach to solve (2.3) efficiently. Furthermore, none of these approaches provide any worst-case performance guarantees.

One could attempt to approximate (2.3) by the sampled two-stage LP

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \mathcal{Y}[\mathbf{H}_{1,N}] = \{\mathbf{y} \in \mathcal{X} \mid \exists \mathbf{v}_i \text{ s.t. } \mathbf{W}\mathbf{v}_i \geq \mathbf{f}(\mathbf{y}, \mathbf{H}_i), i = 1 \ldots, N\}. \end{array} \tag{2.7}$$

However, note that since each scenario $\mathbf{H}_i$ has its own set of second-stage variables $\mathbf{v}_i$, $i = 1, \ldots, N$, the problem dimension grows with N, and the results of Calafiore and Campi [14, 15] no longer yield a good bound on the number of samples required to produce a good approximation. Also, since it is difficult to quantify the VC-dimension of the two-stage constraint, the results in [34] cannot be applied.

To the best of our knowledge, the most elegant approximation approach to (2.3) is given by Nemirovski and Shapiro [61] for the case where the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine. The authors provide nice convergence and performance guarantees by constructing an iterative solution algorithm for (2.3) that closely resembles the ellipsoid algorithm [8, 50, 75]. Moreover, the iterative algorithm is very easy to implement. Erdoğan and Iyengar [36] extend this algorithm to the case where the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-convex. See also Chapter 3.

2.3 Criticism of Chance Constrained Programs

Although the bounds on the sample size N in [20, 14, 15, 61] are distribution-free, in the sense that they do not depend on the underlying measure $\mathbb{Q}$, one must be able to sample from $\mathbb{Q}$ in order to construct the sampled problem (2.6). Also, there is an implicit assumption that the distribution $\mathbb{Q}$ of the random parameters $\mathbf{H}$ is fixed. A major criticism raised against chance constrained problems and, more generally, stochastic programs is that, in practice, the measure is never known exactly. Just as the point estimates for the parameters, the distribution $\mathbb{Q}$ is also estimated from data or measurements, and is, therefore, known only

to within some error, i.e. the measure $\mathbb{Q} \in \mathcal{Q}$ where $\mathcal{Q}$ is a set of measures. Furthermore, the underlying distribution might chance in the future after the decision is made.

Since our primary interest in the chance constrained problems was to use them to handle uncertainty in real life optimization problems, we also have to deal with the uncertainty in the probability distribution $\mathbb{Q}$. In particular we have to produce a solution that is feasible to the chance constraints for all the probability measures $\mathbb{Q} \in \mathcal{Q}$. In Chapter 1 we called this problem an *ambiguous chance-constrained problem*.

As one can see both the single and two-stage ambiguous chance-constrained problems are much harder than their non-ambiguous versions (2.1) and (2.3). This is mainly due to the fact that the ambiguous chance-constrained problems might have uncountably many chance constraints each of which is non-convex. The Chapters 5 and 6 of this dissertation deal with the single and two stage ambiguous chance constrained problems, respectively, and develop tractable approximations to these problems.

Chapter 3

Chance Constraints and Learning Theory

In the first part of this chapter, we will review the results by de Farias and Van Roy [20] and Calafiore and Campi [14, 15], which form the basis of the main results related to the single stage ambiguous chance constrained programs in Chapter 5. In the second part of this chapter, we will review the results by Erdoğan and Iyengar [36], which will set the stage for the results in Chapter 6. Throughout this chapter we assume that the distribution $\mathbb{Q}$ of the parameters $\mathbf{H}$ is known and fixed.

3.1 Single stage chance constrained programs

The goal of this section is to relate the sampled problem (2.6) to the single stage chance constrained problem (2.1). Let $\mathbf{H}_{1,N} = \{\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_N\}$ denote N IID samples of the random vector $\mathbf{H} \sim \mathbb{Q}$. Then the feasible set of the sampled problem (2.6) is given by

$$\mathcal{Y}[\mathbf{H}_{1,N}] = \{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \mathbf{H}_i) \leq 0, i = 1, ..., N\}. \tag{3.1}$$

In the sequel, we will often abbreviate $\mathcal{Y}[\mathbf{H}_{1,N}]$ as $\mathcal{Y}[N]$ with the understanding that the set $\mathcal{Y}[N]$ is defined using a *particular* sequence of IID samples $\mathbf{H}_{1,N}$ of length N. Next, we appropriately interpret concepts from Computational Learning Theory [1, 79, 49] to establish bounds on the number of samples N needed to ensure that $\mathcal{Y}[N]$ is contained in the feasible set $\mathcal{X}_\epsilon = \{\mathbf{x} \in \mathcal{X} : \mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0) \leq \epsilon\}$ of the chance constrained problem (2.1) with a prescribed confidence $1 - \delta$. Next, we briefly summarize the result in [14, 15]

that provides a bound for the number of samples required to ensure that the optimal solution of the sampled problem (2.6) is contained in $\mathcal{X}_\epsilon$ with a prescribed confidence $1-\delta$. As mentioned before, the results in this chapter are not new – they simply provide the context for the new results in Chapter 5.

3.1.1 Decision vectors, concepts, and Vapnik-Chervonenkis (VC) dimension

With each decision vector $\mathbf{x} \in \mathcal{X}$ we associate the *concept* or *classification* of $\mathcal{H}$ defined as follows

$$\mathcal{B}_x = \{\mathbf{h} \in \mathcal{H} : f(\mathbf{x}, \mathbf{h}) \leq 0\}.$$

Let $\mathcal{B}_f = \{\mathcal{B}_x : \mathbf{x} \in \mathcal{X}\}$ denote the class of all concepts induced on $\mathcal{H}$ as the decision vector $\mathbf{x}$ runs over the set $\mathcal{X}$. Thus, the set $\mathcal{X}_\epsilon = \{\mathbf{x} \in \mathcal{X} : \mathbb{Q}(\mathbf{H} \notin \mathcal{B}_x) \leq \epsilon\}$. To apply the results from Learning Theory to the problem at hand, we pretend that our goal is to learn the set $\mathcal{X}_0$ of concepts $\mathcal{B}_x$ that cover $\mathcal{H}$ with probability 1, i.e. $\mathcal{X}_0 = \{\mathbf{x} \in \mathcal{X} : \mathbb{Q}(\mathbf{H} \in \mathcal{B}_x) = 1\}$. Since Learning algorithms only have access to a finite number of samples of the random variable $\mathbf{H}$, it is impossible to learn the concepts in $\mathcal{X}_0$; instead, any such algorithm will have to be satisfied with learning a concept with a small error ϵ, i.e. a concept $\mathcal{B}_x$ with $\mathbb{Q}(\mathbf{H} \notin \mathcal{B}_x) \leq \epsilon$ or equivalently $\mathbf{x} \in \mathcal{X}_\epsilon$. Learning $\mathcal{X}_\epsilon$ is clearly equivalent to generating a large enough collection of samples $\mathbf{H}_{1,N}$ such that for all $\mathbf{x} \notin \mathcal{X}_\epsilon$ there exists at least one sample $\mathbf{H}_i \notin \mathcal{B}_x$.

For the particular case considered in this paper, learning $\mathcal{X}_\epsilon$ is equivalent to producing a good approximation for the function $f(\mathbf{x}, \cdot)$ using a finite number of samples. Thus, one should expect that the complexity of learning $\mathcal{X}_\epsilon$ when the function $f(\mathbf{x}, \mathbf{h}) = a + \mathbf{h}^T\mathbf{x}$ is linear, or equivalently the associated concept $\mathcal{B}_x$ is a half space, should be smaller than the complexity of learning $\mathcal{X}_\epsilon$ when the function $f(\cdot, \cdot)$ is nonlinear. Learning Theory quantifies the complexity of a concept class $\mathcal{B}_f$ by its *Vapnik-Chervonenkis* (VC) dimension [79].

Let $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\} \subset \mathcal{H}$ denote a finite subset of $\mathcal{H}$ with $|S| = N$. Define

$$\Pi_f(\mathcal{S}) = \left\{\left(\mathbf{1}_{\mathcal{B}_x}(\mathbf{s}_1), \ldots, \mathbf{1}_{\mathcal{B}_x}(\mathbf{s}_N)\right) : \mathbf{x} \in \mathcal{X}\right\}, \tag{3.2}$$

where $\mathbf{1}_{\mathcal{B}}(\cdot)$ denotes the characteristic function of the set $\mathcal{B}$. The set Π_f is the set of *dichotomies* or *behaviors* induced by the concept class $\mathcal{B}_f$, or equivalently the function $f(\cdot, \cdot)$. From (3.2), it is clear that the number of elements $|\Pi_f(\mathcal{S})| \leq 2^N$. We say that a set $\mathcal{S}$ is *shattered* by the concept class $\mathcal{B}_f$ if $\Pi_f(\mathcal{S}) = \{0,1\}^N$, or equivalently $|\Pi_f(\mathcal{S})| = 2^N$.

Note that if a set $\mathcal{S}$ is shattered by the concept class $\mathcal{B}_f$ it does not yield any information about the concept class. Thus, the size of the largest shattered set is a measure of the complexity of the concept class $\mathcal{B}_f$.

Definition 1 (VC dimension of $f(\cdot,\cdot)$) *The VC dimension d_f of the function $f(\cdot,\cdot)$ is the cardinality of the largest set $\mathcal{S} \subset \mathcal{H}$ that is shattered by the concept class $\mathcal{B}_f$, i.e.*

$$\begin{aligned} d_f &= \sup\left\{|\mathcal{S}| : \Pi_f(\mathcal{S}) = \{0,1\}^N\right\}, \\ &= \sup\left\{|\mathcal{S}| : |\Pi_f(\mathcal{S})| = 2^N\right\}. \end{aligned} \tag{3.3}$$

In the sequel we will find it convenient to work with the growth function $\pi_f(N)$ defined as follows.

$$\pi_f(N) = \max\left\{|\Pi_f(\mathcal{S})| : |\mathcal{S}| = N\right\}. \tag{3.4}$$

The growth function π_f is another measure of the complexity of the concept class: the faster this function grows, the more behaviors on sets of size m that can be realized by $\mathcal{B}_f$; consequently, the less is the information that this finite set conveys about the class $\mathcal{B}_f$. A surprising and fundamental result in Computational Learning Theory asserts that if the VC dimension $d_f < \infty$, the growth function $\pi_f(N)$ is bounded by a polynomial in N of degree d_f.

Proposition 1 (Sauer's Lemma [12, 1, 49]) *Suppose the VC dimension d_f of the function $f(\cdot,\cdot)$ is finite. Then*

$$\pi_f(N) \leq 1 + \binom{N}{1} + \binom{N}{2} + ... + \binom{N}{d_f} \leq \left(\frac{eN}{d_f}\right)^{d_f}, \tag{3.5}$$

where e denotes the base of natural logarithm.

In this paper we assume that the VC dimension $d_f < \infty$. This is not a very restrictive assumption since many functions $f(\cdot,\cdot)$ used in practice have finite VC dimension.

Proposition 2 *Let d_f denote the VC dimension of the function $f(\cdot,\cdot)$.*

(a) $\mathcal{X} = \mathbf{R}^n$, $\mathcal{H} = \{\mathbf{h} = (h_0, \bar{\mathbf{h}}) : h_0 \in \mathbf{R}, \bar{\mathbf{h}} \in \mathbf{R}^n\} = \mathbf{R}^{n+1}$ *and* $f(\mathbf{x}, \mathbf{h}) = h_0 + \bar{\mathbf{h}}^T\mathbf{x}$. *Then* $d_f = n + 1$.

(b) $\mathcal{X} = \mathbf{R}^n$, $\mathcal{H} = \{\mathbf{h} = (\mathbf{A}, \mathbf{b}, \mathbf{u}, v) : \mathbf{A} \in \mathbf{R}^{p\times n}, \mathbf{b} \in \mathbf{R}^p, \mathbf{u} \in \mathbf{R}^n, v \in \mathbf{R}\}$, *and* $f(\mathbf{x}, \mathbf{h}) = \sqrt{(\mathbf{A}\mathbf{x} + \mathbf{b})^T(\mathbf{A}\mathbf{x} + \mathbf{b})} - \mathbf{u}^T\mathbf{x} - v$. *Then* $d_f \leq O(n^2)$.

(c) Suppose the VC dimension of the function $f_i(\cdot,\cdot)$ is d_i, $i = 1,\ldots,l$. Then the VC dimension d_f of the function $f(\mathbf{x},\mathbf{h}) = \max_{1\le i\le l}\{f_i(\mathbf{x},\mathbf{h})\}$ is bounded above by $d_f \le \mathcal{O}(10^l \max_{1\le i\le l}\{d_i\})$.

Proof: Part (a) is proved on p.77 in [1] (see also [20]), part (b) is established in [12] and part (c) can be established using techniques in [54]. ■

Part (c) states that the best known bound on the VC dimension of a pointwise maximum of l functions grows *exponentially* in l. Thus, the VC dimension of the concept class induced by the constraint function $f(\cdot,\cdot)$ can be quite large. We will remark on this at the end of the next subsection.

3.1.2 Learning the chance constrained set $\mathcal{X}_\epsilon$

For $\mathbf{x}\in\mathcal{X}$ let $\mathrm{err}(\mathbf{x}) = \mathbb{Q}(\mathbf{H}\notin\mathcal{B}_x)$. Thus, $\mathcal{X}_\epsilon = \{\mathbf{x}\in\mathcal{X} : \mathrm{err}(\mathbf{x})\le\epsilon\}$. The feasible set $\mathcal{Y}[N]$ of the sampled problem (2.6) is the set of all decision vectors $\mathbf{x}$, or equivalently concepts $\mathcal{B}_x$, that are consistent with the given sample $\mathbf{H}_{1,N}$. Intuitively speaking, if the sample size is large enough one would expect that $\mathcal{Y}[N]$ is a reasonable estimate of the set $\mathcal{X}_\epsilon$. The next two results make this rigorous.

Lemma 1 *Fix $\epsilon>0$. Suppose $\bar{\mathbf{x}}\in\mathcal{X}$ with $err(\bar{\mathbf{x}})>\epsilon$. Then, for all $N\ge 1$,*

$$\mathbb{Q}^N\left(\mathbf{H}_{1,N} : \bar{\mathbf{x}}\in\mathcal{Y}[N]\right) \le e^{-\epsilon N}, \tag{3.6}$$

where $\mathbb{Q}^N$ denotes the product measure $\mathbb{Q}\times\mathbb{Q}\times\ldots\times\mathbb{Q}$ with N terms.

Proof: Recall that $\mathbf{H}_{1,N}$ are N IID samples of the random vector $\mathbf{H}\sim\mathbb{Q}$. Therefore,

$$\mathbb{Q}^N\left(\mathbf{H}_{1,N} : \bar{\mathbf{x}}\in\mathcal{Y}[N]\right) = \left(\mathbb{Q}(\mathbf{H} : f(\bar{\mathbf{x}},\mathbf{H})\le 0)\right)^N \le (1-\epsilon)^N \le e^{-\epsilon N},$$

where the last inequality follows from the fact that $1-\epsilon\le e^{-\epsilon}$. ■

Lemma 1 establishes that the probability that a given concept $\mathcal{B}_x$ with $\mathrm{err}(\mathbf{x})>\epsilon$ is contained in $\mathcal{Y}[N]$ decays exponentially with the number of samples N. Suppose the set $\mathcal{X}$ is finite. Then the union bound implies that $\mathbb{Q}^N\left(\mathbf{H}_{1,N} : \mathcal{Y}[N]\not\subseteq\mathcal{X}_\epsilon\right) \le |X|\, e^{-\epsilon N}\le\delta$, for all $N\ge\frac{1}{\epsilon}\ln\left(\frac{|X|}{\delta}\right)$, i.e $\mathcal{O}\left(\frac{1}{\epsilon}\ln\left(\frac{|X|}{\delta}\right)\right)$ samples are needed to learn $\mathcal{X}_\epsilon$ with a probability of error bounded by δ. Since the complexity of learning a concept is determined by the VC dimension of the concept class, we expect that a similar bound should hold with $|X|$ replaced by $\pi_f(N)$.

Lemma 2 (Proposition 8.2.3 in [1]) *Let π_f denote the growth function associated with the concept class $\mathcal{B}_f$ induced by $f(\cdot,\cdot)$. Then, for all $N \geq 8/\epsilon$,*

$$\mathbb{Q}^N\left(\mathbf{H}_{1,N} : \mathcal{Y}[N] \not\subseteq \mathcal{X}_\epsilon\right) \leq 2\pi_f(2N)2^{-\epsilon N/2}. \tag{3.7}$$

This result and the upper bound (3.5) imply the following corollary.

Corollary 1 *Fix $\epsilon, \delta > 0$. Suppose the VC dimension d_f of $f(\cdot,\cdot)$ is finite. Then*

$$\mathbb{Q}^N\left(\mathbf{H}_{1,N} : \mathcal{Y}[N] \not\subseteq \mathcal{X}_\epsilon\right) \leq \delta,$$

for all

$$N \geq \max\left\{\frac{8}{\epsilon}, \left(\frac{4d_f}{\epsilon}\lg\left(\frac{12}{\epsilon}\right) + \frac{4}{\epsilon}\lg\left(\frac{2}{\delta}\right)\right)\right\}.$$

Corollary 1 only provides a *sufficient* condition for the sample approximation $\mathcal{Y}[N]$ to be contained in the set $\mathcal{X}_\epsilon$ with a high probability. However, it is quite possible that the set $\mathcal{Y}[N]$ is very "small", and therefore, a very conservative estimate of $\mathcal{X}_\epsilon$. Consequently, the quality of the solution produced by the sample approximation can possibly be quite poor. Nemirovski and Shapiro [61] showed that when $\mathbb{Q}_0$ is symmetric with a concentration of measure property and $f(\mathbf{x},\mathbf{h})$ is bi-affine, the set $\mathcal{Y}[N]$ *contains*, with a high probability, a slightly more conservative version of the chance constrained set $\mathcal{X}_\epsilon$. (Erdoğan and Iyengar [36] extended the result to Lipschitz continuous $f(\mathbf{x},\mathbf{h})$). Together with Corollary 1, this result implies that for a properly chosen N, the set $\mathcal{Y}[N]$ is a reasonable approximation of $\mathcal{X}_\epsilon$. We conclude this section with the following lower bound.

Lemma 3 (Theorem 3.5 in [49]) *Suppose the VC dimension d_f of $f(\cdot,\cdot)$ is finite. Then the worst case number of samples N required to learn $\mathcal{X}_\epsilon$ with a probability at least $1-\delta$ is $N = \Omega(d_f/\epsilon)$.*

Corollary 1 and Lemma 3 establish that the number of samples N required to learn $\mathcal{X}_\epsilon$ is $N = \Theta(d_f/\epsilon)$ to within logarithmic terms. Since the best known bound for the VC dimension of the constraint function $f(\cdot,\cdot)$ is $10^l \max_{1\leq i\leq l}\{d_i\}$ where d_i is VC dimension of the functions f_i, $i = 1,\ldots,l$ (see Proposition 2) (c)), it is possible that the number of samples required to learn $\mathcal{X}_\epsilon$ could be prohibitive even for well-behaved constraint functions.

3.1.3 Quality of the optimal solution of the sampled problem

In this section the goal is more modest – we want to compute the number of samples required to ensure that only the *optimal* solution of the sampled problem (2.6), as opposed to the

entire set $\mathcal{Y}[N]$, is feasible for the chance constrained problem (2.1) with a high probability. Calafiore and Campi [14, 15] recently showed that $N = \mathcal{O}(\frac{n}{\epsilon}\log(\frac{1}{\epsilon}))$ suffices to achieve this goal. In this section we briefly review the results in [14, 15].

Let (P) denote the following convex program

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, N, \\ & \mathbf{x} \in \mathcal{X}, \end{array}$$

where $f_i(\mathbf{x})$ is a convex function of $\mathbf{x}$ for all $i = 1, \ldots, N$. Let $\widehat{\mathbf{x}}$ denote the *unique* optimal solution of (P). Let (P_k) denote the convex program obtained by dropping the k-th constraint, $k = 1, \ldots, N$, and let $\widehat{\mathbf{x}}_k$ denote the *unique* optimal solution of the problem (P_k). See [14, 15] for the case where the optimal solutions are not unique.

Definition 2 (Support constraint) *The k-th constraint $f_k(\mathbf{x}) \leq 0$ is called a support constraint for the problem (P) if $\mathbf{c}^T\widehat{\mathbf{x}}_k < \mathbf{c}^T\widehat{\mathbf{x}}$.*

Theorem 3 (Theorem 2 in [15]) *The convex program has at most n support constraints.*

This result immediately follows from Helly's Theorem (Corollary 21.3.2 in [67]).

Lemma 4 *Fix $\epsilon > 0$. Let $\widehat{\mathbf{x}}$ denote the optimal solution of the sampled problem (2.6). Then $\mathbb{Q}^N(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \mathcal{X}_\epsilon) \leq \binom{N}{n} e^{-\epsilon(N-n)}$.*

Proof: The sampled problem (2.6) is a convex program with N constraints. Let $\mathcal{I} \subseteq \{1, \ldots, N\}$ with $|\mathcal{I}| = n$. Let $\mathcal{H}^N_{\mathcal{I}} = \{(\mathbf{h}_1, \ldots, \mathbf{h}_N) : \text{ all the support constraints} \subseteq \mathcal{I}\}$. Then Theorem 3 implies that the cartesian product $\mathcal{H}^N (\equiv \bigotimes_{i=1}^N \mathcal{H}) = \cup_{\{\mathcal{I}\subseteq\{1,\ldots,N\}:|\mathcal{I}|=n\}} \mathcal{H}^N_{\mathcal{I}}$. Thus,

$$\begin{aligned} &\mathbb{Q}^N(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \mathcal{X}_\epsilon) \\ &\quad \leq \sum_{\{\mathcal{I}\subseteq\{1,\ldots,N\}:|\mathcal{I}|=n\}} \mathbb{Q}^N(\mathbf{H}_{1,N} \in \mathcal{H}^N_{\mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \mathcal{X}_\epsilon), \\ &\quad = \sum_{\{\mathcal{I}\subseteq\{1,\ldots,N\}:|\mathcal{I}|=n\}} \Big(\mathbb{Q}^n(\mathbf{H}_{i\in\mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \mathcal{X}_\epsilon) \prod_{i\notin\mathcal{I}} \mathbb{Q}(\mathbf{H}_i : f(\widehat{\mathbf{x}}_{\mathcal{I}}, \mathbf{H}_i) \leq 0 | \mathcal{A}_{\mathcal{I}})\Big), \end{aligned}$$

where $\widehat{\mathbf{x}}_{\mathcal{I}}$ denotes the optimal solution of the sampled problem (2.6) with only the samples $i \in \mathcal{I}$ present, $\mathcal{A}_{\mathcal{I}}$ is the event $\mathcal{A}_{\mathcal{I}} = \{\mathbf{H}_{i\in\mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \mathcal{X}_\epsilon\}$ and each probability in the sum

can be written as a product because $\mathbf{H}_{1,N}$ are IID samples. Since $\widehat{\mathbf{x}}_{\mathcal{I}} \notin \mathcal{X}_\epsilon$, it follows that $\mathbb{Q}(\mathbf{H}_i : f(\widehat{\mathbf{x}}_{\mathcal{I}}, \mathbf{H}_i) \leq 0 | \mathcal{A}_{\mathcal{I}}) \leq (1-\epsilon)$, for all $i \notin \mathcal{I}$. Thus,

$$\begin{aligned} \mathbb{Q}^N(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \mathcal{X}_\epsilon) &\leq (1-\epsilon)^{(N-n)} \sum_{\{\mathcal{I} \subseteq \{1,\ldots,N\} : |\mathcal{I}|=n\}} \mathbb{Q}^n(\mathbf{H}_{i \in I} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \mathcal{X}_\epsilon) \\ &\leq \binom{N}{n}(1-\epsilon)^{(N-n)} \leq \binom{N}{n} e^{-\epsilon(N-n)}. \end{aligned}$$

■

Lemma 4 immediately implies the following.

Corollary 2 *Fix $\epsilon, \delta > 0$. Let $\widehat{\mathbf{x}}$ denote the optimal solution of the sampled problem (2.6). Then*

$$\mathbb{Q}^N \left(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \mathcal{X}_\epsilon\right) \leq \delta,$$

for all

$$N \geq \frac{2n}{\epsilon} \ln\left(\frac{2}{\epsilon}\right) + \frac{2}{\epsilon} \ln\left(\frac{1}{\delta}\right) + 2n.$$

3.2 Two-stage convex chance constrained programs

In this section we develop an approximation algorithm for the two-stage convex chance constrained problem

$$\begin{array}{ll} \min & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \left\{ \mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon \right\}, \end{array} \tag{3.8}$$

where $\mathcal{C}$ denotes the polyhedron

$$\mathcal{C} = \left\{ \mathbf{z} \in \mathbf{R}^l \mid \exists \mathbf{v} \in \mathbf{R}^k \text{ s.t. } \mathbf{z} \leq \mathbf{W}\mathbf{v} \right\}, \tag{3.9}$$

$\mathbf{x} \in \mathbf{R}^n$ is the decision vector, $\mathbf{H} \in \mathcal{H}$ is a random parameter vector distributed according to $\mathbb{Q}$, $\mathbf{W} \in \mathbf{R}^{l \times k}$ is a recourse matrix, and $\mathbf{f}(\mathbf{x}, \mathbf{h}) : \mathcal{X} \times \mathcal{H} \mapsto \mathbf{R}^l$ is an impact function. Nemirovski and Shapiro [61] introduced the particular form for the set $\mathcal{C}$ and the associated chance constrained problem (3.8). To reiterate, the variable $\mathbf{x}$ denotes the first stage decisions and the variable $\mathbf{v} \in \mathbf{R}^k$ denotes the second stage decisions. We assume that the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ and the sets $\mathcal{X}$ and $\mathcal{H}$ satisfy the following conditions.

Assumption 1

(i) The sets $\mathcal{X}$ *and* $\mathcal{H}$ *are convex compact sets. Let* $\|\mathbf{x}\| \leq R_{\mathcal{X}} < \infty$ *(resp.* $\|\mathbf{h}\| \leq R_{\mathcal{H}} < \infty$*) for all* $\mathbf{x} \in \mathcal{X}$ *(resp.* $\mathbf{h} \in \mathcal{H}$*).*

(ii) The impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ *is bi-convex, i.e. for all fixed* $\mathbf{h} \in \mathcal{H}$ *(resp.* $\mathbf{x} \in \mathcal{X}$*), the function* $\mathbf{f}(\cdot, \mathbf{h}) : \mathcal{X} \mapsto \mathbf{R}^l$ *(resp.* $\mathbf{f}(\mathbf{x}, \cdot) : \mathcal{H} \mapsto \mathbf{R}^l$*) is a convex function.*

(iii) $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{f}^0(\mathbf{x}) + \mathbf{f}^1(\mathbf{x}, \mathbf{h})$*, where* $\mathbf{f}^1(\mathbf{x}, \alpha\mathbf{h}) = \alpha\mathbf{f}^1(\mathbf{x}, \mathbf{h})$*, for all* $\alpha \geq 0$*,* $(\mathbf{x}, \mathbf{h}) \in \mathcal{X} \times \mathcal{H}$*.*

(iv) There exists a non-decreasing (finite-valued) function $K_f^{\mathcal{H}} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ *such that* $\|\mathbf{f}(\mathbf{x}_1, \mathbf{h}) - \mathbf{f}(\mathbf{x}_0, \mathbf{h})\| \leq K_f^{\mathcal{H}}(\|\mathbf{h}\|)\|\mathbf{x}_1 - \mathbf{x}_0\|$*, for all* $\mathbf{h} \in \mathcal{H}$ *and* $\mathbf{x}_0$*,* $\mathbf{x}_1 \in \mathcal{X}$*. Let* $\bar{K}_f^{\mathcal{H}} \triangleq K_f^{\mathcal{H}}(R_{\mathcal{H}})$*.*

(v) There exists a non-decreasing (finite-valued) function $K_f^{\mathcal{X}} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ *such that* $\|\mathbf{f}(\mathbf{x}, \mathbf{h}_1) - \mathbf{f}(\mathbf{x}, \mathbf{h}_0)\| \leq K_f^{\mathcal{X}}(\|\mathbf{x}\|)\|\mathbf{h}_1 - \mathbf{h}_0\|$*, for all* $\mathbf{h}_1$*,* $\mathbf{h}_0 \in \mathcal{H}$ *and* $\mathbf{x} \in \mathcal{X}$*. Let* $\bar{K}_f^{\mathcal{X}} \triangleq K_f^{\mathcal{X}}(R_{\mathcal{X}})$*.*

Condition (i) is not essential for the results to hold. In most practical applications, the uncertain system parameter vector $\mathbf{H}$ is bounded and, therefore, (i) holds. That said, approximating the distribution $\mathbb{Q}$ of the uncertain parameter by a multivariate distribution with unbounded support, such as Normal, lognormal, Cauchy etc., often yields analytical tractability. When this is the case, one can set the parameter $R_{\mathcal{H}}$ to any large but finite value that ensures $\mathbb{Q}(\mathbf{H} : \|\mathbf{H}\| > R_{\mathcal{H}}) < \gamma << \epsilon$ and replace ϵ by $\epsilon - \gamma$ in the original problem. Note that this transformation leaves the probability of infeasibility unchanged and after this transformation all the results in this paper will hold. In the rest of the paper, to keep the presentation simple and intuitive we will assume that the underlying distribution has a compact support.

Assuming $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is convex in $\mathbf{x}$ for all fixed $\mathbf{h} \in \mathcal{H}$ is necessary to ensure that the sampled version of (3.8) is tractable. The assumption that $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is convex in $\mathbf{h}$ for all fixed $\mathbf{x} \in \mathcal{X}$ and has a component that is homogeneous allows one to leverage the concentration-of-measure property defined below in Definition 3. The assumption that $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is Lipschitz continuous individually in each variable is sufficient, though not necessary, to establish the approximation results. Assumption 1 may appear overly restrictive; however, there are many function classes that satisfy these. Two canonical examples are as follows.

(a) Affine constraints: $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{A}_0\mathbf{x} + \sum_{i=1}^m h_i\mathbf{A}_i\mathbf{x}$ with $\mathcal{X} \subset \mathbf{R}^n$, $\mathcal{H} \subset \mathbf{R}^m$, and $\mathbf{A}_i \in \mathbf{R}^{l \times n}$ for $i = 0, \ldots, m$. The growth functions $K_f^{\mathcal{H}}(\|\mathbf{h}\|) = O(1)(|||\mathbf{A}_0||| + \|\mathbf{h}\| \sum_{i=1}^m |||\mathbf{A}_i|||)$

and $K_f^{\mathcal{X}}(\|\mathbf{x}\|) = O(1)\|\mathbf{x}\| \sum_{i=1}^m |||\mathbf{A}_i|||$, with the constants depending on the particular choice of the vector norm $\|\cdot\|$ and the matrix norm $|||\cdot|||$, satisfy Assumption 1.

(b) Second-order cone constraints: Each component of $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is a conic quadratic representable function [8], e.g. $f_j(\mathbf{x}, \mathbf{h}) = \sqrt{(\Gamma \mathbf{A}_j \mathbf{x} + \boldsymbol{\rho})^T (\Gamma \mathbf{A}_j \mathbf{x} + \boldsymbol{\rho})} - \mathbf{t}_j^T \mathbf{x} - v_j$ with $\mathcal{X} \subset \mathbf{R}^n$, $\mathbf{A}_j \in \mathbf{R}^{k \times n}$, $\mathbf{t}_j \in \mathbf{R}^n$, $v_j \in \mathbf{R}$ for $j = 1, \ldots, l$, and $\mathcal{H} = \left\{\mathbf{h} = (\Gamma, \boldsymbol{\rho}) : \Gamma \in \mathbf{R}^{p \times k}, \boldsymbol{\rho} \in \mathbf{R}^p, \right\}$. In this case, we can set $K_f^{\mathcal{H}}(\|\mathbf{h}\|) = O(1)(|||\mathbf{T}|||+\|\mathbf{h}\|||\mathbf{A}|||)$ and $K_f^{\mathcal{X}}(\|\mathbf{x}\|) = O(1)\|\mathbf{x}\|||\mathbf{A}|||$, where $\mathbf{T} = [\mathbf{t}_1^T; \mathbf{t}_2^T; \ldots; \mathbf{t}_l^T]$, $\mathbf{A} = [\mathbf{A}_1; \mathbf{A}_2; \ldots; \mathbf{A}_l]$.

We assume that $\mathbb{Q}$ satisfies the $(\bar{\theta}, \psi)$-*concentration of measure* property defined as follows [61].

Definition 3 ([61]) *Let $\bar{\theta} \in (\frac{1}{2}, 1]$ and $\psi(\alpha, \theta) : [1, \infty) \times (\bar{\theta}, 1] \mapsto \mathbf{R}_+$ be a convex, non-decreasing and non-constant function of α. A distribution $\mathbb{Q}$ on $\mathbf{R}^m$ is said to have $(\bar{\theta}, \psi)$-concentration of measure property if for all $\alpha \geq 1$ and closed convex sets $\mathcal{A}$ with $\mathbb{Q}(\mathcal{A}) \geq \theta > \bar{\theta}$,*

$$\mathbb{Q}(\{\mathbf{H} \notin \alpha \mathcal{A}\}) \leq \exp\{-\psi(\alpha, \theta)\}.$$

This assumption essentially states that a small "blow-up" of the set $\mathcal{A}$ with a measure of at least $\bar{\theta}$ increases its measure exponentially. The prototypical example of a measure satisfying such a property is the multivariate Normal distribution, $\mathcal{N}(\mathbf{0}, \mathbf{I})$ – it satisfies the concentration property with $\psi(\alpha, \theta) = \frac{1}{2}\alpha^2 \Phi^{-1}(\theta)^2$, where $\Phi(\cdot)$ denotes the CDF of a $\mathcal{N}(0, 1)$ random variable. Some other generic distributions that have this property are general normal distributions, uniform distribution on the vertices of the unit cube, uniform distribution on the unit cube [61]. The assumption that the impact function has a homogeneous component (see Assumption 1 part (iii)) is made to leverage the concentration property of the measure $\mathbb{Q}$.

Erdoğan and Iyengar [34] (see also [20]) compute bounds for the number of samples required to reliably produce a solution for the chance constrained problem when the VC-dimension of the constraint function is known. Since the VC dimension of the constraint defining (3.8) is difficult to quantify, VC-dimension based bounds are not useful in solving (3.8). Calafiore and Campi [14, 15] bounded the number of samples required in terms of the number of decision variables when all the constraints are convex in $\mathbf{x}$ for any fixed $\mathbf{h}$. Since we need the second stage variables $\mathbf{v}$ to check feasibility for each sample, the number of decision variables grows linearly with the number of samples; this renders the bounds in [14, 15] worthless. We propose to approximately solve (3.8) using Algorithm

SolveChance shown in Figure 3.1. SolveChance is a simple extension of an algorithm proposed by Nemirovski and Shapiro [61] to solve the special case with bi-affine impact functions $\mathbf{f}(\mathbf{x}, \mathbf{h})$. The extension to the bi-convex case is fairly straightforward; our main contribution is to show that feasibility of an appropriately defined conservative version of (3.8) implies that SolveChance returns a "good" solution with a high probability. Next, we carefully describe the algorithm and then prove a series of intermediate results that are needed to establish the main result.

SolveChance uses two oracles, $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$. The oracle $S_{\mathcal{X}}$ is the separation oracle for the convex compact set $\mathcal{X}$ and the oracle $S_{\mathcal{R}}(\mathbf{x}; \mathbf{h})$, for a fixed $\mathbf{h} \in \mathbf{R}^m$, returns a linear inequality $L : \mathbf{R}^n \mapsto \mathbf{R}$ that separates $\mathbf{x} \in \mathbf{R}^n$ from the convex set

$$\mathcal{R} = \left\{\mathbf{x} \mid \mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}\right\} = \{\mathbf{x} \mid \exists \mathbf{v} \in \mathbf{R}^k \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h})\}. \tag{3.10}$$

Thus, $\mathbf{x} \in \mathcal{R}$ if, and only if, the value of the optimization problem

$$\begin{aligned} \min \quad & \|\mathbf{u} - \mathbf{x}\| \\ \text{s.t.} \quad & \mathbf{W}\mathbf{v} - \mathbf{f}(\mathbf{u}, \mathbf{h}) \geq \mathbf{0} \end{aligned} \tag{3.11}$$

is equal to 0. When the optimal value of (3.11) is strictly positive, any sub-gradient $\mathbf{d}$ at the optimal solution $\mathbf{u}^*$ satisfies $\mathbf{d}^T(\mathbf{u} - \mathbf{x}) > 0$ for all $\mathbf{u} \in \mathcal{R}$ and, therefore, serves as a separating hyperplane. Note that (3.11) is a convex optimization problem and can be solved very efficiently when $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is an affine function or a conic quadratic representable function [8]. The above discussion establishes that the separation oracle $S_{\mathcal{R}}(\mathbf{x}; \mathbf{h})$ is well-defined.

We will call an output $\widehat{\mathbf{x}}$ of SolveChance well-defined if $\widehat{\mathbf{x}} \neq \emptyset$. Nemirovski and Shapiro [61] established the following result (see also Theorem 12).

Theorem 4 *Suppose the output $\widehat{\mathbf{x}}$ of* SolveChance *is well-defined, (i.e. $\widehat{\mathbf{x}} \neq \emptyset$). Then $\widehat{\mathbf{x}}$ is infeasible for (3.8), i.e. $\widehat{\mathbf{x}} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})$, with a probability at most $M\delta$.*

Note that Theorem 4 does not consider the case $\widehat{\mathbf{x}} = \emptyset$. Next, we compute an upper bound on the probability that the output $\widehat{\mathbf{x}} = \emptyset$. Let $\mathcal{I} = \{L_1, \ldots, L_{MN}\}$ denote an ordered list of the MN linear inequalities generated by the calls to the oracle $S_{\mathcal{R}}$ over the course of one run of SolveChance. Let $\mathcal{X}_{\mathcal{I}}$ denote the convex compact set

$$\mathcal{X}_{\mathcal{I}} = \{\mathbf{x} \in \mathcal{X} : L_j(\mathbf{x}) \leq 0, j = 1, \ldots, MN\}. \tag{3.12}$$

Note that the set $\mathcal{I}$ and, therefore, $\mathcal{X}_{\mathcal{I}}$ depend on the IID samples $\mathbf{H}_{1,MN}$, where each $\mathbf{H}_i \sim \mathbb{Q}$. The following lemma was stated in [61] without a proof.

Input: $\epsilon > 0$, $\delta \in (0,1)$, separation oracles $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$
Output: $\widehat{\mathbf{x}}$
set $\mathcal{P} \leftarrow \emptyset$, $N \leftarrow \lceil \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right) \rceil$, $M \leftarrow \lceil 2n^2 \ln(\frac{nR_{\mathcal{X}}^2 \|\mathbf{c}\|_2}{r\omega} + 2) \rceil$
set $\mathbf{x}_0 \leftarrow \mathbf{0}$, $\mathbf{E}_0 \leftarrow R_{\mathcal{X}}\mathbf{I}$
for t = 1, ..., M **do**
    Construct a direction vector $\mathbf{d}_t$
        $s_t \leftarrow S_{\mathcal{X}}(\mathbf{x}_{t-1})$
        if $\mathbf{x}_{t-1} \notin \mathcal{X}$, **set** $\mathbf{d}_t \leftarrow \nabla s_t$
        else
            generate $\mathbf{H}_{1,N}$ IID $\mathbb{Q}$
            $L_{N(t-1)+i} \leftarrow S_{\mathcal{R}}(\mathbf{x}_{t-1}; \mathbf{H}_i)$, $i = 1, \ldots, N$.
        if $\exists\ j \in \{1, \ldots, Nt\}$ such that $L_j(\mathbf{x}_{t-1}) > 0$, **set** $\mathbf{d}_t \leftarrow \nabla L_j$
        else set $\mathbf{d}_t = \mathbf{c}$ and $\mathcal{P} \leftarrow \mathcal{P} \cup \mathbf{x}_{t-1}$
    Given $(\mathbf{x}_{t-1}, \mathbf{E}_{t-1})$ and $\mathbf{d}_t$, set $(\mathbf{x}_t, \mathbf{E}_t)$ by the Ellipsoid method update
if $\mathcal{P} = \emptyset$ **return** $\widehat{\mathbf{x}} \leftarrow \emptyset$; **else return** $\widehat{\mathbf{x}} \leftarrow \text{argmin}_{\mathbf{x} \in \mathcal{P}}\{\mathbf{c}^T\mathbf{x}\}$

Figure 3.1: Algorithm SolveChance

Lemma 5 *Suppose the set $\mathcal{X}_{\mathcal{I}}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r. Then the output $\widehat{\mathbf{x}}$ of* SolveChance *is well defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$.*

Proof: We will prove the result by contradiction. Suppose SolveChance returns $\widehat{\mathbf{x}} = \emptyset$. Let $\{(\mathbf{x}_{t-1}, \mathbf{d}_t) : t = 1, \ldots, M\}$ denote the iterates and the separating hyperplanes generated during one run of SolveChance. Since $\widehat{\mathbf{x}} = \emptyset$, for each $\mathbf{x}_t$, $t = 0, \ldots, M-1$, either $\mathbf{x}_t \notin \mathcal{X}$ or there exists some $j \in \{1, \ldots, N(t+1)\}$ such that $L_j(\mathbf{x}_t) > 0$. Thus, $\mathbf{x}_t \notin \mathcal{X}_{\mathcal{I}}$ for all $t = 0, \ldots, M-1$.

By the definition of $\mathcal{X}_{\mathcal{I}}$, it follows that $\{(\mathbf{x}_{t-1}, \mathbf{d}_t) : t = 1, \ldots, M\}$ is a set of iterates and separating hyperplanes that could have been generated while using the Ellipsoid algorithm to solve the convex optimization problem $\min\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in \mathcal{X}_{\mathcal{I}}\}$. Since $\mathbf{x}_t \notin \mathcal{X}_{\mathcal{I}}$ for all $t = 0, \ldots, M-1$, it follows that the Ellipsoid algorithm returns an empty solution. This is a contradiction because of the choice of the iteration count M [8]. ■

For $\alpha > 1$ and $r > 0$, define

$$\mathcal{X}_\epsilon(\alpha, r) = \left\{ \mathbf{x} \in \mathcal{X}^{-r} \,\middle|\, \mathbb{Q}\Big(\mathbf{H} : \mathbf{f}(\mathbf{x}, \alpha\mathbf{H}) \in \mathcal{C}^{-r\bar{K}_f^{\mathcal{H}}}\Big) > 1 - \epsilon \right\}, \tag{3.13}$$

where

$$\mathcal{A}^{-\gamma} \stackrel{\Delta}{=} \{\mathbf{y} \in \mathcal{A} \mid \mathbf{y} + \mathbf{u} \in \mathcal{A}, \text{ for all } \|\mathbf{u}\| \leq \gamma\}, \tag{3.14}$$

denote the *interior* γ-ball of the set $\mathcal{A}$. Recall that $\bar{K}_f^{\mathcal{H}} = K_f^{\mathcal{H}}(R_{\mathcal{H}})$ is the maximum value of the growth function $K_f^{\mathcal{H}}(\cdot)$. Definition (3.14) implies that

$$\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}^{-\mu} \quad \Leftrightarrow \quad \mathbf{f}(\mathbf{x}, \mathbf{h}) + \mathbf{u} \in \mathcal{C}, \ \forall \mathbf{u} : \|\mathbf{u}\| \leq \mu. \tag{3.15}$$

Since $\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}$ if, and only if, there exists a $\mathbf{v} \in \mathbf{R}^k$ such that $\mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h})$; we have that $\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}^{-\mu}$ if, and only if, for all $\mathbf{u}$ with $\|\mathbf{u}\| \leq \mu$, there exists a $\mathbf{v} \in \mathbf{R}^k$ (possibly a function of $\mathbf{u}$) such that $\mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h}) + \mathbf{u}$. The set $\mathcal{X}_\epsilon(\alpha, r)$ can be interpreted as the set of decision vectors that are *robustly* feasible for the chance constrained set $\widehat{\mathcal{X}}_\epsilon(\mathbb{Q})$ [6].

Theorem 5 *Fix $\mathbf{y} \in \mathcal{X}_\epsilon(\alpha, r)$. Then the Euclidean ball $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$ with a probability at least $1 - MNe^{-\psi(\alpha, 1-\epsilon)}$.*

Proof: Let $\mu_r = r\bar{K}_f^{\mathcal{H}}$. Then we have that

$$\begin{aligned} 1 - \epsilon &< \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), \\ &= \mathbb{Q}(\alpha^{-1}\mathbf{H} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{H}) \in \mathcal{C}^{-\mu_r}). \end{aligned} \tag{3.16}$$

Let $\mathcal{H}_\mathbf{y} = \{\mathbf{h} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{h}) \in \mathcal{C}^{-\mu_r}\}$. Then (3.16) and the concentration property of $\mathbb{Q}$ imply that $\mathbb{Q}(\mathcal{H}_\mathbf{y}) \geq 1 - e^{-\psi(\alpha, 1-\epsilon)}$ provided $1 - \epsilon > \bar{\theta}$.

Since $\mathbf{y} \in \mathcal{X}^{-r}$, it follows that $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}$, and Lipschitz continuity of $\mathbf{f}$ implies that for all $\mathbf{x} \in \mathcal{B}_r(\mathbf{y})$ and $\mathbf{h} \in \mathcal{H}_\mathbf{y}$,

$$\|\mathbf{f}(\mathbf{x}, \mathbf{h}) - \mathbf{f}(\mathbf{y}, \mathbf{h})\| \leq K_f^{\mathcal{H}}(\|\mathbf{h}\|)\|\mathbf{x} - \mathbf{y}\| \leq rK_f^{\mathcal{H}}(R_{\mathcal{H}}) = \mu_r,$$

i.e. $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{f}(\mathbf{y}, \mathbf{h}) + \mathbf{u}$ for some $\mathbf{u}$ with $\|\mathbf{u}\| \leq \mu_r$. Since $\mathbf{y} \in \mathcal{X}_\epsilon(\alpha, r)$, it follows that

$$\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}, \ \forall \mathbf{x} \in \mathcal{B}_r(\mathbf{y}), \ \mathbf{h} \in \mathcal{H}_\mathbf{y}. \tag{3.17}$$

Suppose $\mathbf{H}_i \in \mathcal{H}_\mathbf{y}$, for all $i = 1, \ldots, MN$. Then (3.17) implies that $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$. Thus,

$$\{\mathbf{h}_{1,MN} : \mathbf{h}_i \in \mathcal{H}_\mathbf{y}, i = 1, \ldots, MN\} \subseteq \{\mathbf{h}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}\}. \tag{3.18}$$

Consequently,

$$\begin{aligned} \mathbb{Q}^{MN}(\mathbf{H}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \not\subseteq \mathcal{X}_\mathcal{I}) &\leq \mathbb{Q}^{MN}((\mathbf{H}_{1,MN} : \mathbf{H}_i \in \mathcal{H}_\mathbf{y}, i = 1, \ldots, MN)^c), \\ &\leq MN\mathbb{Q}(\mathcal{H}_\mathbf{y}^c) < MNe^{-\psi(\alpha, 1-\epsilon)}. \end{aligned}$$

■

Theorem 5 implies the following result.

Corollary 3 *Suppose* $\mathcal{X}_\epsilon(\alpha, r) \neq \emptyset$. *Then the output* $\widehat{\mathbf{x}}$ *of* SolveChance *is well-defined (i.e.* $\widehat{\mathbf{x}} \neq \emptyset$*) with a probability at least* $1 - MNe^{-\psi(\alpha,1-\epsilon)}$.

Proof: Lemma 5 implies that the solution $\widehat{\mathbf{x}}$ is well defined if the set $\mathcal{X}_\mathcal{I}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r; and, by Theorem 5, the probability of this event is at least $1 - MNe^{-\psi(\alpha,1-\epsilon)}$. ■

Corollary 3 establishes that the output $\widehat{\mathbf{x}}$ of SolveChance is well defined with a high probability provided the chance constrained problem (3.8) is robustly feasible and the measure $\mathbb{Q}$ has a concentration of measure property. Next, we establish a bound on the value $\mathbf{c}^T\widehat{\mathbf{x}}$. We will call a well-defined output $\widehat{\mathbf{x}}$ (i.e. $\widehat{\mathbf{x}} \neq \emptyset$) of SolveChance an (α, r, ω)-approximation of (3.8) if

$$\mathbf{c}^T\widehat{\mathbf{x}} \leq \inf_{\mathbf{x}\in\mathcal{X}_\epsilon(\alpha,r)} \mathbf{c}^T\mathbf{x} + \omega. \tag{3.19}$$

This definition was introduced in [61].

Theorem 6 *Suppose* $\mathcal{X}_\epsilon(\alpha, r) \neq \emptyset$. *Then* SolveChance *returns an* (α, r, ω)*-approximate solution of (3.8) with a probability at least* $(1 - MNe^{-\psi(\alpha,1-\epsilon)})(1 - M\delta)$.

Proof: Fix $\kappa > 0$ and choose $\mathbf{y}^\kappa \in \mathcal{X}_\epsilon(\alpha, r)$ such that

$$\mathbf{c}^T\mathbf{y}^\kappa \leq \inf_{\mathbf{x}\in\mathcal{X}_\epsilon(\alpha,r)} \mathbf{c}^T\mathbf{x} + \kappa. \tag{3.20}$$

By Lemma 5, $\widehat{\mathbf{x}}$ is well defined on the event $\{\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_\mathcal{I}\}$ and by Theorem 5 the probability of this event is at least $1-MNe^{-\psi(\alpha,1-\epsilon)}$. Combining this with Theorem 4, we can show that SolveChance produces a well-defined output $\widehat{\mathbf{x}}$ that is feasible for (3.8) with a probability at least $(1 - MNe^{-\psi(\alpha,1-\epsilon)})(1 - M\delta)$.

Next, we show that $\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_\mathcal{I}$ implies

$$\mathbf{c}^T\widehat{\mathbf{x}} \leq \min_{\mathbf{x}\in\mathcal{X}_\mathcal{I}} \mathbf{c}^T\mathbf{x} + \omega. \tag{3.21}$$

This result is established by exploiting the close resemblance of SolveChance to the Ellipsoid algorithm. We closely follow the analysis of the Ellipsoid algorithm detailed in the proof of Theorem 5.2.1 in [8].

Let $\mathbf{x}^*_\mathcal{I} = \text{argmin}_{\mathbf{x}\in\mathcal{X}_\mathcal{I}}\{\mathbf{c}^T\mathbf{x}\}$. Let $(\mathbf{x}_t, \mathbf{E}_t)$, $t = 0, 1, \ldots, M-1$, denote the iterates generated by SolveChance . Let $\mathcal{E}_t$ denote the ellipsoid $\mathcal{E}_t = \{\mathbf{z} \mid (\mathbf{z}-\mathbf{x}_t)^T\mathbf{E}_t^{-1}(\mathbf{z}-\mathbf{x}_t) \leq 1\}$. The choice of the iteration count M ensures that there exists $\nu \leq 1$ and $\mathbf{z} \in \mathcal{X}_\mathcal{I}$ such that

(i) $\mathbf{vol}(\mathcal{E}_t)/r^n \leq \nu \leq 1$,

(ii) $\mathbf{y} = \mathbf{x}^*_{\mathcal{I}} + \nu(\mathbf{z} - \mathbf{x}^*_{\mathcal{I}}) \notin \mathcal{E}_M$.

Since $\mathbf{y} \in \mathcal{X}_{\mathcal{I}} \subset \mathcal{X}$, it follows that $\mathbf{y} \in \mathcal{E}_0$. Therefore, there exists $\tau < M$ such that $\mathbf{y} \in \mathcal{E}_\tau$ but $\mathbf{y} \notin \mathcal{E}_{\tau+1}$. Then it follows that $\mathbf{d}_\tau^T \mathbf{s} > \mathbf{d}_\tau^T \mathbf{x}_\tau$ for all $\mathbf{s} \in \mathcal{E}^c_{\tau+1} \cap \mathcal{E}_\tau$, and, in particular,

$$\mathbf{d}_\tau^T \mathbf{y} > \mathbf{d}_\tau^T \mathbf{x}_\tau. \tag{3.22}$$

We claim that the iterate $\mathbf{x}_\tau \in \mathcal{P}$, i.e. it is one of the candidate points for computing the output $\widehat{\mathbf{x}}$. Suppose this is not the case. Then we must have that $\mathbf{x}_\tau \notin \mathcal{X}_\tau$, where

$$\mathcal{X}_\tau = \{\mathbf{x} \in \mathcal{X} \mid L_j(\mathbf{x}) \leq 0, j = 1, \ldots, N(\tau + 1)\} \supseteq \mathcal{X}_{\mathcal{I}},$$

and the separating hyperplane $\mathbf{d}_\tau$ must satisfy $\mathbf{d}_\tau^T \mathbf{s} \leq \mathbf{d}_\tau^T \mathbf{x}_\tau$ for all $\mathbf{s} \in \mathcal{X}_\tau$. Thus, the bound (3.22) together with the fact that $\mathbf{y} \in \mathcal{X}_{\mathcal{I}} \subseteq \mathcal{X}_\tau$ leads to a contradiction. Now, the analysis in the proof of Theorem 5.2.1 in [8] implies that $\mathbf{c}^T \mathbf{x}_\tau \leq \min_{\mathbf{x} \in \mathcal{X}_{\mathcal{I}}} \mathbf{c}^T \mathbf{x} + \omega$ and the bound (3.21) follows.

Finally, $\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_{\mathcal{I}}$ implies $\min_{\mathbf{x} \in \mathcal{X}_{\mathcal{I}}} \mathbf{c}^T \mathbf{x} \leq \mathbf{c}^T \mathbf{y}^\kappa$, which together with (3.20) imply that

$$\mathbf{c}^T \widehat{\mathbf{x}} \leq \min_{\mathbf{x} \in \mathcal{X}_{\mathcal{I}}} \mathbf{c}^T \mathbf{x} + \omega \leq \mathbf{c}^T \mathbf{y}^\kappa + \omega \leq \inf_{\mathbf{x} \in \mathcal{X}_\epsilon(\alpha, r)} \mathbf{c}^T \mathbf{x} + \kappa + \omega$$

Since $\kappa > 0$ was arbitrary, the result follows. ■

3.2.1 Discussion of the approximation result

Algorithm SolveChance has three tunable parameters, namely δ, ω and r. Nemirovski and Shapiro [61] study the effects of these parameters on the running time and approximation quality of SolveChance. While the parameters δ and ω have a well-defined meaning, the parameter r is rather ad-hoc and it is not clear how to set its value. The parameter r is clearly very important for the performance of SolveChance: the iteration count M, the probability that the output $\widehat{\mathbf{x}}$ of SolveChance is well-defined and feasible for (3.8) (see Theorem 4 and Corollary 3), and the approximation guarantee on the output $\widehat{\mathbf{x}}$ (see Theorem 6) are all inversely proportional to the parameter r. Since the set $\mathcal{X}_{\mathcal{I}}$ is random, selecting r is difficult. Yet, SolveChance requires r as an input.

The main contribution of this section is to provide guidance in selecting r. The iteration count M as well as the probability that a well-defined $\widehat{\mathbf{x}}$ is infeasible are inversely proportional to r. This suggests that r should be set as high as possible. For any fixed $\alpha > 0$, the maximum allowed value $\bar{r}$ of r is limited by the requirement that the set $\mathcal{X}_\epsilon(\alpha, r)$ (see (3.13))

is non-empty, i.e. the set of decisions that are *robustly* feasible for (3.8) is non-empty [6]. Thus, the computational complexity of SolveChance is intimately related to the robust feasibility of (3.8) – the more robust the chance constrained problem, the easier it is to compute a feasible solution. This is similar to the relationship between the complexity of computing an optimal solution of a conic linear program and its condition number [65, 66, 38]. Although setting a high value for the parameter r induces SolveChance to efficiently produce a feasible solution, it results in a weak approximation guarantee (see (3.19)).

Let $p = (1 - M\delta)(1 - MNe^{-\psi(\alpha,1-\epsilon)})$ denote the probability that the output $\widehat{\mathbf{x}}$ is well-defined and feasible. If $p > 0$, then $T = \frac{-\ln(\gamma)}{p}$ independent replications of SolveChance ensure that with a probability $1 - \gamma$ at least one of the outputs is feasible. The requirement $p > 0$ implies an upper bound on M, and consequently, a lower bound $\underline{r}$ on r. Thus, it follows that there is a lower bound on the achievable approximation guarantee. Within the range $[\underline{r}, \overline{r}]$ one can trade-off the optimality with efficiency.

Chapter 4

Probability Metrics and Coupling

In Chapter 1 we introduced the following robust chance constrained set (see (1.6))

$$\bar{\mathcal{X}}_\epsilon = \Big\{ \mathbf{x} \in \mathcal{X} : \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon \Big\},$$

where $\mathcal{Q}$ is the uncertainty set for the distribution $\mathbb{Q}$. For an appropriately chosen metric ρ on the space $\mathcal{M}(\mathcal{H})$ of probability measures on $\mathcal{H}$, the set $\mathcal{Q}$ is defined as follows:

$$\mathcal{Q} = \big\{ \mathbb{Q} : \rho(\mathbb{Q}, \mathbb{Q}_0) \leq \beta \big\}. \tag{4.1}$$

Recall that $\mathcal{X} \subseteq \mathbf{R}^n$, $\mathcal{H} \subseteq \mathbf{R}^m$, and we denote the norm in $\mathcal{H}$ space by $\|\cdot\|$. The assumption here is that the uncertain parameter $\mathbf{H}$ is distributed according to some *fixed* distribution $\mathbb{Q} \in \mathcal{Q}$; however, the decision maker can only estimate the distribution to within the error β. The goal is to compute a solution $\widehat{\mathbf{x}}$ that performs "well" for all distributions in the set $\mathcal{Q}$. We will characterize the details of the approximation later.

In this chapter we first review properties of some basic probability metrics and discuss how the uncertainty set $\mathcal{Q}$ can be chosen. Next, we introduce the concept of *coupling* of random variables, which plays an important role in constructing approximations of the single stage version of the robust chance constrained set $\bar{\mathcal{X}}_\epsilon$ via samples.

In this dissertation we will be primarily using the probability metric called Prohorov metric which is discussed in detail next.

4.1 The Prohorov Metric

The Prohorov metric, ρ_p, is defined as follows.

$$\rho_p(\mathbb{Q}_1, \mathbb{Q}_2) = \inf \left\{ \epsilon : \mathbb{Q}_1(\mathcal{B}) \leq \mathbb{Q}_2(\mathcal{B}^\epsilon) + \epsilon, \forall \mathcal{B} \in \mathcal{F}(\mathcal{H}) \right\}, \tag{4.2}$$

where

$$\mathcal{B}^\epsilon = \left\{ \mathbf{x} \in \mathcal{X} : \inf_{\mathbf{z} \in \mathcal{B}} \|\mathbf{x} - \mathbf{z}\| \leq \epsilon \right\}.$$

Although the definition appears asymmetric, ρ_p is a metric. It plays an important role in probability because it metrizes weak convergence. Moreover, $\rho_p(\mathbb{Q}_1, \mathbb{Q}_2)$ is the minimum distance "in probability" between random variables distributed according to $\mathbb{Q}_i$, $i = 1, 2$.

The Prohorov metric is admittedly a difficult metric to work with and it often complicates problem analysis. However, there are many compelling reasons for using this metric. In any practical application, the true distribution $\mathbb{Q}$ is unknown and is often estimated from samples. Let $\widehat{\mathbb{Q}}_n$ denote the empirical distribution constructed from n IID samples of $\mathbf{H}$. Since $\mathbb{Q}_n \Rightarrow \mathbb{Q}$ and the Prohorov metric metrizes weak convergence of probability measures, it follows that sets of the form (4.1) defined using the Prohorov metric ρ_p is are the natural candidates for the "confidence region" around the point estimate $\mathbb{Q}_n$. The size of the uncertainty set, i.e. the parameter β, can also be estimated from the samples. See § 4.2 for details.

Uncertainty sets defined in terms of the Prohorov metric allow one to model situations where the one makes arbitrarily large errors in estimating the true distribution $\mathbb{Q}$, albeit with a small probability. Suppose the parameter $\mathbf{H}$ denotes the return on a collection of assets. With probability $1 - \beta$ the return distribution $\mathbb{Q} = \mathbb{Q}_1$ and with probability β the return distribution is $\mathbb{Q} = \mathbb{Q}_2$, where $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are any two arbitrary measures. Then, for all sufficiently small β, we have $\rho_p(\mathbb{Q}_1, \mathbb{Q}) \leq \beta$, i.e. if the return distribution $\mathbb{Q}$ is arbitrarily different from the nominal distribution $\mathbb{Q}_1$ but with a small probability, the measures are still close in Prohorov metric. This feature of the Prohorov metric implies that the corresponding uncertainty sets are likely to be more conservative than those corresponding to other metrics. For more information on Prohorov metric, see [64, 43].

4.2 Choosing the uncertainty set

In the context of ambiguous chance constrained problems, the choice of the Prohorov metric can be justified by the fact that it is the natural metric for defining weak convergence of

measures. It is, however, not clear how one should compute the constant β. We propose the following strategy. Suppose we assume that $\mathbf{H}$ is described by a parametric family of distributions $\mathcal{F}(\theta)$, $\theta \in \Theta$. Suppose we estimate θ using an estimator $T_N(\mathbf{H}_1, \ldots, \mathbf{H}_N)$ and let $\mathbb{Q}_0 = \mathcal{F}(\theta_0)$, where $\theta_0 = T_N(\mathbf{H}_1, \ldots, \mathbf{H}_N)$. In the robust statistics literature, there is a "breakdown point" $\epsilon(T)$ associated with every estimator beyond which the estimator is completely unreliable [43]. Heuristically, the estimator is said to perform well for all measures $\mathbb{Q}$ such that $\rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \frac{1}{4}\epsilon(T)$. Thus, we could set $\beta = \frac{1}{4}\epsilon(T)$.

4.3 Some other metrics of interest and their relation with the Prohorov metric

Some other metrics of interest are as follows.

(a) Wasserstein or Kantorovich metric ρ_w:

$$\rho_w(\mathbb{Q}_1, \mathbb{Q}_2) = \sup\left\{ \left| \int_{\mathcal{H}} g(\mathbf{h}) \big(\mathbb{Q}_1(d\mathbf{h}) - \mathbb{Q}_2(d\mathbf{h}) \big) \right| : g \in C_{1,1}(\mathcal{H}) \right\},$$

where $C_{1,1}(\mathcal{H})$ denotes the set of Lipschitz continuous functions with Lipschitz constant at most 1.

(b) Total variation metric ρ_{tv}:

$$\rho_{tv}(\mathbb{Q}_1, \mathbb{Q}_2) = \sup\left\{ |\mathbb{Q}_1(\mathcal{B}) - \mathbb{Q}_2(\mathcal{B})| : \mathcal{B} \in \mathcal{F}(\mathcal{H}) \right\}. \tag{4.3}$$

(c) Hellinger metric ρ_h: Let f_i, $i = 1, 2$ denote the densities of measures $\mathbb{Q}_i$, $i = 1, 2$, with respect to a common dominating measure (e.g. $\mathbb{Q} = (\mathbb{Q}_1 + \mathbb{Q}_2)/2$). Then

$$\rho_h(\mathbb{Q}_1, \mathbb{Q}_2) = \Big(\int_{\mathcal{H}} \big(\sqrt{f_1} - \sqrt{f_2} \big)^2 \mathbb{Q}(d\mathbf{h}) \Big)^{\frac{1}{2}}.$$

(d) Relative entropy distance ρ_e: Let f_i, $i = 1, 2$ denote the densities of measures $\mathbb{Q}_i$, $i = 1, 2$, with respect to a common dominating measure (e.g. $\mathbb{Q} = (\mathbb{Q}_1 + \mathbb{Q}_2)/2$). Then

$$\rho_e(\mathbb{Q}_1, \mathbb{Q}_2) = \int_{\mathcal{H}} f_1(\mathbf{h}) \lg \Big(\frac{f_1(\mathbf{h})}{f_2(\mathbf{h})} \Big) d\mathbf{h}$$

The relative entropy distance ρ_e is *not* a metric because it is not symmetric and does not satisfy the triangle inequality.

The following lemma relates the Prohorov metric ρ_p to the other distance functions.

Lemma 6 ([39]) *The distances ρ_w, ρ_h, ρ_{tv} and ρ_e are related to the Prohorov metric as follows.*

(a) Prohorov and Wasserstein metrics: $\rho_p^2 \leq \rho_w \leq (\mathbf{diam}(\mathcal{H}) + 1)\rho_p$, where $\mathbf{diam}(\mathcal{H}) = \sup\{\|\mathbf{h}_1 - \mathbf{h}_2\| : \mathbf{h}_i \in \mathcal{H}, i = 1, 2\}$.

(b) Prohorov and Total variation metrics: $\rho_p \leq \rho_{tv}$.

(c) Prohorov and Hellinger metrics: $\rho_p \leq \rho_h$.

(d) Prohorov metric and the relative entropy distance: $\rho_p \leq \sqrt{\rho_e/2}$.

These bounds imply that for any uncertainty set of the form $\mathcal{Q} = \{\mathbb{Q} : \rho(\mathbb{Q}, \mathbb{Q}_0) \leq \gamma\}$, where the metric ρ is given by ρ_w, ρ_{tv}, ρ_h or ρ_e, one can choose $\beta(\gamma) > 0$ such that $\mathcal{Q} \subseteq \tilde{\mathcal{Q}} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta(\gamma)\}$, i.e. $\tilde{\mathcal{Q}}$ is a *conservative* approximation of $\mathcal{Q}$.

4.4 Coupling of random variables

In this section, we introduce the concept of *coupling* of random variables and relate it to the probability metrics.

Definition 4 (Coupling of random variables) *A random variable $\tilde{\mathbf{X}}$ is said to be a* copy *or a* representation *of the random variable $\mathbf{X}$ if and only if they have the same distribution, i.e. $\tilde{\mathbf{X}} \stackrel{D}{=} \mathbf{X}$. A collection of random variables $\{\tilde{\mathbf{X}}^\alpha : \alpha \in \mathcal{A}\}$ defined on a common probability space $(\Omega, \mathcal{F}(\Omega), \mathbb{P})$ is said to be a* coupling *of the collection $\{\mathbf{X}^\alpha : \alpha \in \mathcal{A}\}$ if and only if $\tilde{\mathbf{X}}^\alpha \stackrel{D}{=} \mathbf{X}^\alpha$, for all $\alpha \in \mathcal{A}$.*

Note that only the individual $\tilde{\mathbf{X}}^\alpha$ are copies of the individual $\mathbf{X}^\alpha$, the whole collection $\{\tilde{\mathbf{X}}^\alpha : \alpha \in \mathcal{A}\}$ is *not* a copy of $\{\mathbf{X}^\alpha : \alpha \in \mathcal{A}\}$, i.e. the joint distribution of $\{\tilde{\mathbf{X}}^\alpha : \alpha \in \mathcal{A}\}$ need not be the same as that of $\{\mathbf{X}^\alpha : \alpha \in \mathcal{A}\}$.

Theorem 7 (Strassen-Dudley) *Let $\mathbf{X}^1 \sim \mathbb{Q}_1$ and $\mathbf{X}^2 \sim \mathbb{Q}_2$ be two random variables taking values in $\mathcal{H}$. Suppose $\rho_p(\mathbb{Q}_1, \mathbb{Q}_2) \leq \beta$. Then there exists a coupling $(\tilde{\mathbf{X}}^1, \tilde{\mathbf{X}}^2)$ of $(\mathbf{X}^1, \mathbf{X}^2)$ such that*

$$\mathbb{P}\left(\|\tilde{\mathbf{X}}^1 - \tilde{\mathbf{X}}^2\| > \beta\right) \leq \beta. \tag{4.4}$$

Proof: This result was established by Strassen [76] for complete separable metric spaces and extended to arbitrary separable metric spaces by Dudley [25]. See also Rachev [64]. ■
Theorem 7 states that if two probability measures $\mathbb{Q}_i$, $i = 1, 2$, are "close" in the Prohorov metric then there exists a coupling $(\tilde{\mathbf{X}}^1, \tilde{\mathbf{X}}^2)$ such that the samples are "close" with a high probability. This result can be improved if the random variables $\mathbf{X}^i$, $i = 1, 2$, are bounded w.p.1.

Theorem 8 *Let* $\mathbf{X}^1 \sim \mathbb{Q}_1$ *and* $\mathbf{X}^2 \sim \mathbb{Q}_2$ *are two random variables taking values in* $\mathcal{H}$. *Suppose* $\rho_p(\mathbb{Q}_1, \mathbb{Q}_2) \leq \beta$ *and* $\|\mathbf{X}^i\| \leq R$ *a.s.,* $i = 1, 2$. *Then there exists a coupling* $(\tilde{\mathbf{X}}^1, \tilde{\mathbf{X}}^2)$ *of* $(\mathbf{X}^1, \mathbf{X}^2)$ *such that*

$$\mathbb{E}\left(\|\tilde{\mathbf{X}}^1 - \tilde{\mathbf{X}}^2\|\right) \leq (1 + 2R)\beta, \tag{4.5}$$

where the expectation is with respect to the common probability measure $\mathbb{P}$.

Proof: The Wasserstein metric $\rho_w(\mathbb{Q}_1, \mathbb{Q}_2)$ between probability measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ can be equivalently characterized as follows.

$$\begin{aligned} &\rho_w(\mathbb{Q}_1, \mathbb{Q}_2) \\ &\quad = \inf\left\{\mathbb{E}\left[\|\tilde{\mathbf{X}}^1 - \tilde{\mathbf{X}}^2\|\right] : \mathbf{X}^i \sim \mathbb{Q}_i, i = 1, 2,\ (\tilde{\mathbf{X}}^1, \tilde{\mathbf{X}}^2) \text{ is a coupling of } (\mathbf{X}^1, \mathbf{X}^2)\right\}. \end{aligned}$$

Since $\|\mathbf{X}^i\| \leq R$ a.s., one can without loss of generality assume that $\mathbf{diam}(\mathcal{H}) \leq 2R$. Thus, the bound $\rho_w \leq (\mathbf{diam}(\mathcal{H})+1)\rho_p$ together with the characterization above, yields the result. ■

Chapter 5

Single-Stage Ambiguous Chance Constraints

As we have mentioned in § 2.3 that a major criticism raised against chance constrained problems is the implicit assumption that the distribution $\mathbb{Q}$ of the random parameters $\mathbf{H}$ is fixed. However, in practice, the measure $\mathbb{Q}$ is never known exactly. Just as the point estimates for the parameters, the distribution $\mathbb{Q}$ is also estimated from data or measurements, and is, therefore, known only to within some error, i.e. the measure $\mathbb{Q} \in \mathcal{Q}$ where $\mathcal{Q}$ is a set of measures. See Chapter 4 for detailed information about the uncertainty set $\mathcal{Q}$.

Since our primary interest in the chance constrained problems was to use them to handle the uncertainty in real life optimization problems, we also have to deal with the uncertainty in the probability distribution $\mathbb{Q}$. A natural problem to consider when the measure $\mathbb{Q}$ is uncertain is given by

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \bar{\mathcal{X}}_\epsilon, \end{aligned} \tag{5.1}$$

where

$$\bar{\mathcal{X}}_\epsilon = \Big\{\mathbf{x} \in \mathcal{X} : \mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0) \leq \epsilon,\ \forall \mathbb{Q} \in \mathcal{Q}\Big\}. \tag{5.2}$$

In Chapter 1 we called (5.1) an *ambiguous chance-constrained problem*. To reiterate a problem of the form (5.1) has two sources of uncertainty: the distribution $\mathbb{Q}$ of the parameter $\mathbf{H}$ is uncertain, and, given a measure $\mathbb{Q}$, the particular realization of the parameter $\mathbf{H}$ is also uncertain. In the decision theory literature the uncertainty in the distribution is referred to as *ambiguity*, and hence the name for the problem. Modeling ambiguity and its consequence

has been received attention in several different fields. See Chapter 1 for a brief history of the minimax formulation in the literature.

5.1 Sample Approximations to the Ambiguous Chance Constrained Programs

In this chapter we show how to construct a sampling-based approximation for the robust chance constrained set

$$\bar{\mathcal{X}}_\epsilon = \Big\{\mathbf{x} \in \mathcal{X} : \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{Q}(\mathbf{H} : f(\mathbf{x},\mathbf{H}) > 0) \leq \epsilon\Big\},$$

where $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q},\mathbb{Q}_0) \leq \beta\}$, and ρ_p denotes the Prohorov metric. Note that the bounds in Lemma 6 in Chapter 4 imply that one can conservatively approximate an uncertainty set defined in terms of any of the metrics discussed in Chapter 4 by a set defined in terms of the Prohorov metric. The main results of this chapter are the robust analogs of Lemma 1, Lemma 2 and Lemma 4 in Chapter 3.

In this chapter we define the error of a solution $\mathbf{x}$, $\mathrm{err}(\mathbf{x})$, as follows.

$$\mathrm{err}(\mathbf{x}) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{Q}(\mathbf{H} : f(\mathbf{x},\mathbf{H}) > 0). \tag{5.3}$$

Thus, $\bar{\mathcal{X}}_\epsilon = \{\mathbf{x} \in \mathcal{X} : \mathrm{err}(\mathbf{x}) \leq \epsilon\}$. Let $\mathbf{H}^0_{1,N} = \{\mathbf{H}^0_1,\ldots,\mathbf{H}^0_N\}$ denote N IID samples drawn according to the central probability measure $\mathbb{Q}_0$. Let $\mathcal{Y}[N,\beta]$ denote the set

$$\mathcal{Y}[N,\beta] = \left\{\mathbf{x} \in \mathcal{X} : f(\mathbf{x},\mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}^0_i\| \leq \beta, i = 1,\ldots,N\right\}. \tag{5.4}$$

As in Chapter 3, the set $\mathcal{Y}[N,\beta]$ is defined using a *particular* sequence of IID samples $\mathbf{H}^0_{1,N}$ of length N drawn according to the measure $\mathbb{Q}_0$.

Theorem 9 *Fix $\epsilon > 0$. Suppose $\bar{\mathbf{x}} \in \mathcal{X}$ with $\mathrm{err}(\bar{\mathbf{x}}) > \epsilon$. Then, for all $N \geq 1$,*

$$\mathbb{Q}_0^N\left(\mathbf{H}^0_{1,N} : \bar{\mathbf{x}} \in \mathcal{Y}[N,\beta]\right) \leq e^{-(\epsilon-\beta)N}. \tag{5.5}$$

Remark 1 *The probability $\mathbb{Q}_0^N(\mathbf{H}^0_{1,N} : \bar{\mathbf{x}} \in \mathcal{Y}[N,\beta])$ decays exponentially with N only if $\epsilon > \beta$. Thus, uncertainty in the measure manifests itself as a lower bound on the achievable error probability. See Appendix A for more detailed discussion on this issue.*

Proof: Fix $0 < \eta \leq \epsilon$. Since $\mathrm{err}(\bar{\mathbf{x}}) > \epsilon$ we can select $\mathbb{Q}_1 \in \mathcal{Q}$ such that $\mathbb{Q}_1(\mathbf{H} : f(\bar{\mathbf{x}}, \mathbf{H}) > 0) > \epsilon - \eta$. Let $\mathbf{H}^i \sim \mathbb{Q}_i$, $i = 0, 1$. Since $\rho_p(\mathbb{Q}_1, \mathbb{Q}_0) \leq \beta$, the Strassen-Dudley Representation Theorem (Theorem 7) implies that there exists a coupling $(\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0)$ of the pair $(\mathbf{H}^1, \mathbf{H}^0)$ such that (4.4) holds, i.e. $\mathbb{P}\big(\|\tilde{\mathbf{H}}^1 - \tilde{\mathbf{H}}^0\| > \beta\big) \leq \beta$. Let $\{(\tilde{\mathbf{H}}^1_1, \tilde{\mathbf{H}}^0_1), \ldots, (\tilde{\mathbf{H}}^1_N, \tilde{\mathbf{H}}^0_N)\}$ denote N IID samples of the jointly distributed pair of random vectors $(\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0)$. Let $\tilde{\mathcal{Y}}[N, \beta]$ denote the set

$$\tilde{\mathcal{Y}}[N, \beta] = \left\{ \mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0_i\| \leq \beta, i = 1, \ldots, N \right\}. \tag{5.6}$$

Since $\mathbf{H}^0 \stackrel{\mathrm{D}}{=} \tilde{\mathbf{H}}^0$ and $\bar{\mathbf{x}}$ is fixed, we have that

$$\mathbb{Q}_0^N \left(\mathbf{H}^0_{1,N} : \bar{\mathbf{x}} \in \mathcal{Y}[N, \beta] \right) = \mathbb{P}^N \left(\tilde{\mathbf{H}}^0_{1,N} : \bar{\mathbf{x}} \in \tilde{\mathcal{Y}}[N, \beta] \right). \tag{5.7}$$

Moreover,

$$\begin{aligned} \mathbb{P}^N \left(\tilde{\mathbf{H}}^0_{1,N} : \bar{\mathbf{x}} \in \tilde{\mathcal{Y}}[N, \beta] \right) &= \prod_{i=1}^{N} \mathbb{P} \left(\tilde{\mathbf{H}}^0_i : f(\bar{\mathbf{x}}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0_i\| \leq \beta \right), \\ &= \left(\mathbb{P} \left(\tilde{\mathbf{H}}^0 : f(\bar{\mathbf{x}}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0\| \leq \beta \right) \right)^N. \end{aligned} \tag{5.8}$$

Each term in (5.8) can be bounded as follows.

$$\begin{aligned} &\mathbb{P} \left(\tilde{\mathbf{H}}^0 : f(\bar{\mathbf{x}}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0\| \leq \beta \right) \\ &= \mathbb{P} \left((\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0) : f(\bar{\mathbf{x}}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0\| \leq \beta, \|\tilde{\mathbf{H}}^1 - \tilde{\mathbf{H}}^0\| \leq \beta \right) \\ &\quad + \mathbb{P} \left((\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0) : f(\bar{\mathbf{x}}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}^0\| \leq \beta, \|\tilde{\mathbf{H}}^1 - \tilde{\mathbf{H}}^0\| > \beta \right), \\ &\leq \mathbb{P} \left(\tilde{\mathbf{H}}^1 : f(\bar{\mathbf{x}}, \tilde{\mathbf{H}}^1) \leq 0 \right) + \mathbb{P} \left((\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0) : \|\tilde{\mathbf{H}}^1 - \tilde{\mathbf{H}}^0\| > \beta \right), \qquad (5.9) \\ &\leq (1 - \epsilon + \eta) + \beta, \qquad (5.10) \end{aligned}$$

where (5.9) follows from the fact that the probability only increases if one removes restrictions, and (5.10) follows from the bound (4.4) and the fact that $\tilde{\mathbf{H}}^1 \stackrel{\mathrm{D}}{=} \mathbf{H}^1$. From (5.7), (5.10) and (5.8), we have

$$\begin{aligned} &\mathbb{Q}_0^N \left(\mathbf{H}^0_{1,N} : \bar{\mathbf{x}} \in \mathcal{Y}[N, \beta] \right) \\ &\quad = \mathbb{P}^N \left(\tilde{\mathbf{H}}^0_{1,N} : \bar{\mathbf{x}} \in \tilde{\mathcal{Y}}[N, \beta] \right) \leq (1 - \epsilon + \beta + \eta)^N \leq e^{-N(\epsilon - \beta - \eta)}. \end{aligned}$$

Since $\eta \leq \epsilon$ was arbitrary, the result follows. ■

Note that we only generate samples according to the central measure $\mathbb{Q}_0$. The coupling is a construct needed to translate the bound on extremal measure that achieves the supremum in the definition of $\mathrm{err}(\mathbf{x})$ to the measure $\mathbb{Q}_0$.

Suppose $\mathbb{Q}(\mathbf{H} : \|\mathbf{H}\| > R) = 0$ for all $\mathbb{Q} \in \mathcal{Q}$. Define the set $\mathcal{Y}_\lambda[N]$ as follows.

$$\mathcal{Y}_\lambda[N] = \left\{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \lambda\beta, i = 1, \ldots, N\right\}. \tag{5.11}$$

Then Theorem 8 in Chapter 4 and Markov's inequality implies the following corollary.

Corollary 4 *Fix $\epsilon > 0$ and $\bar{\mathbf{x}} \in \mathcal{X}$ with $\mathrm{err}(\bar{\mathbf{x}}) > \epsilon$. Suppose $\mathbb{Q}(\mathbf{H} : \|\mathbf{H}\| > R) = 0$ for all $\mathbb{Q} \in \mathcal{Q}$. Then*

$$\mathbb{P}\left(\bar{\mathbf{x}} \in \mathcal{Y}_\lambda[N]\right) \leq \left(1 - \epsilon + \frac{1 + 2R}{\lambda}\right)^N. \tag{5.12}$$

Unlike in Theorem 9, here we have a parameter λ that can be controlled to achieve any desired probability of error ϵ.

Next, we establish a robust analog of Lemma 2 in Chapter 3. We show that if the VC dimension d_f of the function $f(\cdot, \cdot)$ is finite, $\rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta$, and the number of samples $N = O(\frac{d_f}{\epsilon - \beta})$ (a precise bound is given in Theorem 10),

$$\mathbb{Q}_0^N \left(\mathbf{H}_{1,N}^0 : \mathcal{Y}[N, \beta] \subseteq \mathcal{X}_\epsilon(\mathbb{Q})\right) \geq 1 - \delta.$$

This result should be interpreted as follows. The distribution of the parameters $\mathbf{H}$ is uncertain and is only known to lie in the uncertainty set $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$ and we want to characterize the set of decisions $\mathbf{x}$ that satisfy $\mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0) \leq \epsilon$ no matter which probability measure $\mathbb{Q}$ is selected from the uncertainty set $\mathcal{Q}$. The bound above shows that for $N = \mathcal{O}(\frac{d_f}{\epsilon - \beta})$ the set $\mathcal{Y}[N, \beta]$ is a conservative approximation for $\mathcal{X}_\epsilon(\mathbb{Q})$ given in (2.1) for any fixed $\mathbb{Q}$ with a high probability.

Theorem 10 *Fix $\delta > 0$, $\epsilon > \beta$ and $\mathbb{Q}_1 \in \mathcal{Q}$. Suppose the VC dimension d_f of the function $f(\cdot, \cdot)$ is $3 < d_f < \infty$ and $\beta + 2^{-\epsilon/2} < 1$. Then*
$\mathbb{Q}_0^N \left(\mathbf{H}_{1,N}^0 : \mathcal{Y}[N, \beta] \not\subseteq \mathcal{X}_\epsilon(\mathbb{Q}_1)\right) \leq \delta$, for all N satisfying

$$\begin{array}{rcl} N & \geq & \max\Big\{ \quad d_f, \quad \frac{8}{\epsilon}, \quad \frac{2d_f}{e(1-\beta)} \ln\left(\frac{e}{1-\beta}\right) + \frac{2}{1-\beta} \ln\left(\frac{e}{(e-1)\delta}\right) + 1, \\ & & \qquad \frac{4d_f}{\epsilon - \mu} \log\left(\frac{12}{\epsilon - \mu}\right) + \frac{4}{\epsilon - \mu} \log\left(\frac{2}{\delta(1-\beta)}\right) \Big\}, \end{array} \tag{5.13}$$

where $\mu = 2\left(\frac{\epsilon}{2} + \log(\beta + 2^{-\epsilon/2})\right)$ and e is the base of natural logarithm.

Remark 2 *Since $\beta = 0$ implies $\mu = 0$, we recover the non-robust result in Lemma 2 when $\beta = 0$.*

Proof: Since the measure $\mathbb{Q}_1 \in \mathcal{Q}$ is fixed, we will abbreviate $\mathcal{X}_\epsilon(\mathbb{Q}_1)$ by $\mathcal{X}_\epsilon$. Let $\mathcal{X}_\epsilon^c$ denote the complement of the set $\mathcal{X}_\epsilon$. As in the proof of Theorem 9, let $\mathbf{H}^i \sim \mathbb{Q}_i$, $i = 0, 1$ and let $(\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0)$ denote a coupling of the pair $(\mathbf{H}^1, \mathbf{H}^0)$ such that (4.4) holds, i.e. $\mathbb{P}\big(\|\tilde{\mathbf{H}}^1 - \tilde{\mathbf{H}}^0\| > \beta\big) \leq \beta$. Let $\big\{(\tilde{\mathbf{H}}_1^1, \tilde{\mathbf{H}}_1^0), \ldots, (\tilde{\mathbf{H}}_N^1, \tilde{\mathbf{H}}_N^0)\big\}$ denote N IID samples of the jointly distributed pair of random vectors $(\tilde{\mathbf{H}}^1, \tilde{\mathbf{H}}^0)$. Then

$$\begin{aligned}
&\mathbb{Q}_0^N \left(\mathbf{H}_{1,N}^0 : \mathcal{Y}[N, \beta] \not\subseteq \mathcal{X}_\epsilon\right) \\
&\quad = \quad \mathbb{P}^N \left(\tilde{\mathbf{H}}_{1,N}^0 : \tilde{\mathcal{Y}}[N, \beta] \cap \mathcal{X}_\epsilon^c \neq \emptyset\right), \\
&\quad = \quad \sum_{j=0}^{N} \mathbb{P}^N \left((\tilde{\mathbf{H}}_i^0, \tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N} : \tilde{\mathcal{Y}}[N, \beta] \cap \mathcal{X}_\epsilon^c \neq \emptyset, |\mathcal{I}| = j\right),
\end{aligned}$$

where $\mathcal{I} = \big\{i \in \{1, ..., N\} : \|\tilde{\mathbf{H}}_i^1 - \tilde{\mathbf{H}}_i^0\| \leq \beta\big\}$. For a set $\mathcal{I} \subseteq \{1, \ldots, N\}$ let $\mathcal{A}(\mathcal{I})$ denote the event

$$\mathcal{A}(\mathcal{I}) = \left\{(\tilde{\mathbf{H}}_i^0, \tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N} : \|\tilde{\mathbf{H}}_i^1 - \tilde{\mathbf{H}}_i^0\| \leq \beta, \forall i \in \mathcal{I}, \|\tilde{\mathbf{H}}_i^1 - \tilde{\mathbf{H}}_i^0\| > \beta, \forall i \notin \mathcal{I}\right\}$$

and let $\mathcal{Y}[\mathcal{I}, \beta] = \big\{\mathbf{x} \in \mathcal{X} : f(\mathbf{x}, \mathbf{z}) \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \tilde{\mathbf{H}}_i^0\| \leq \beta, i \in \mathcal{I}\big\}$. Fix $\mathcal{I}_1, \mathcal{I}_2 \subseteq \{1, ..., N\}$ with $|\mathcal{I}_1| = |\mathcal{I}_2|$. Since $\{(\tilde{\mathbf{H}}_i^1, \tilde{\mathbf{H}}_i^0)\}$, $i = 1, ..., N$ are IID, it is clear that

$$\begin{aligned}
&\mathbb{P}^N \left((\tilde{\mathbf{H}}_i^0, \tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N} : \tilde{\mathcal{Y}}[N, \beta] \cap \mathcal{X}_\epsilon^c \neq \emptyset, \mathcal{A}(\mathcal{I}_1)\right) \\
&\quad = \quad \mathbb{P}^N \left((\tilde{\mathbf{H}}_i^0, \tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N} : \tilde{\mathcal{Y}}[N, \beta] \cap \mathcal{X}_\epsilon^c \neq \emptyset, \mathcal{A}(\mathcal{I}_2)\right). \qquad (5.14)
\end{aligned}$$

Set $\mathcal{I}_0 = \emptyset$, and $\mathcal{I}_j = \{1, \ldots, j\}$, $j = 1, \ldots, N$. Since there are $\binom{N}{j}$ possible selections for a

set $\mathcal{I}$ of cardinality j, (5.14) implies that

$$\begin{aligned}
&\mathbb{P}^N\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N}:\tilde{\mathcal{Y}}[N,\beta]\cap\mathcal{X}_\epsilon^c\neq\emptyset,|\mathcal{I}|=j\right)\\
&=\binom{N}{j}\mathbb{P}^N\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N}:\tilde{\mathcal{Y}}[N,\beta]\cap\mathcal{X}_\epsilon^c\neq\emptyset,\mathcal{A}(\mathcal{I}_j)\right),\\
&\leq\binom{N}{j}\mathbb{P}^N\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i=1,\ldots,N}:\tilde{\mathcal{Y}}[\mathcal{I}_j,\beta]\cap\mathcal{X}_\epsilon^c\neq\emptyset,\mathcal{A}(\mathcal{I}_j)\right), \qquad (5.15)\\
&=\binom{N}{j}\mathbb{P}^j\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i\in\mathcal{I}_j}:\tilde{\mathcal{Y}}[\mathcal{I}_j,\beta]\cap\mathcal{X}_\epsilon^c\neq\emptyset,\|\tilde{\mathbf{H}}_i^1-\tilde{\mathbf{H}}_i^0\|\leq\beta,\forall i\in\mathcal{I}_j\right)\times\\
&\qquad\qquad\mathbb{P}^{N-j}\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i\notin\mathcal{I}_j}:\|\tilde{\mathbf{H}}_i^1-\tilde{\mathbf{H}}_i^0\|>\beta,\forall i\notin\mathcal{I}_j\right), \qquad (5.16)\\
&\leq\binom{N}{j}\beta^{N-j}\times\\
&\qquad\mathbb{P}^j\left((\tilde{\mathbf{H}}_i^0,\tilde{\mathbf{H}}_i^1)_{i=1,\ldots,j}:\tilde{\mathcal{Y}}[\mathcal{I}_j,\beta]\cap\mathcal{X}_\epsilon^c\neq\emptyset,\|\tilde{\mathbf{H}}_i^1-\tilde{\mathbf{H}}_i^0\|\leq\beta,\forall i\in\mathcal{I}_j\right), \qquad (5.17)\\
&\leq\binom{N}{j}\beta^{N-j}\mathbb{P}^j\left(\tilde{\mathbf{H}}_{1,j}^1:\exists\mathbf{x}\in\mathcal{X}_\epsilon^c\text{ s.t. }f(\mathbf{x},\tilde{\mathbf{H}}_i^1)\leq 0,\forall i\in\mathcal{I}_j\right), \qquad (5.18)
\end{aligned}$$

where (5.15) and (5.18) follows from the fact that the probability only increases if one removes restrictions, (5.16) follows from the fact that $\{(\tilde{\mathbf{H}}_i^1,\tilde{\mathbf{H}}_i^0)\}$, $i=1,...,N$ are IID, and (5.17) follows from the bound (4.4). Note that the bound (5.18) only involves the random vector $\tilde{\mathbf{H}}^1$, or equivalently the (unknown) true measure $\mathbb{Q}_1$. Thus, once again we have used coupling to translate a bound in terms of the central measure $\mathbb{Q}_0$ to one involving the measure $\mathbb{Q}_1$. We do not need coupling beyond this stage of the proof. In the rest of this proof we bound (5.18) using Lemma 2 applied to the (unknown) measure $\mathbb{Q}_1$. Let $N_0=\lfloor\frac{8}{\epsilon}\rfloor$.

Then

$$\begin{aligned}
&\sum_{j=0}^{N}\binom{N}{j}\beta^{N-j}\mathbb{P}^{j}\left(\tilde{\mathbf{H}}_{1,j}^{1}:\exists\mathbf{x}\in\mathcal{X}_{\epsilon}^{c}\text{ s.t. }f(\mathbf{x},\tilde{\mathbf{H}}_{i}^{1})\leq 0,\forall i\in\mathcal{I}_{j}\right)\\
&=\sum_{j=0}^{N_0}\binom{N}{j}\beta^{N-j}\mathbb{P}^{j}\left(\tilde{\mathbf{H}}_{1,j}^{1}:\exists\mathbf{x}\in\mathcal{X}_{\epsilon}^{c}\text{ s.t. }f(\mathbf{x},\tilde{\mathbf{H}}_{i}^{1})\leq 0,\forall i\in\mathcal{I}_{j}\right)\\
&\quad+\sum_{j=N_0+1}^{N}\binom{N}{j}\beta^{N-j}\mathbb{P}^{j}\left(\tilde{\mathbf{H}}_{1,j}^{1}:\exists\mathbf{x}\in\mathcal{X}_{\epsilon}^{c}\text{ s.t. }f(\mathbf{x},\tilde{\mathbf{H}}_{i}^{1})\leq 0,\forall i\in\mathcal{I}_{j}\right),\\
&\leq\sum_{j=0}^{N_0}\binom{N}{j}\beta^{N-j}\\
&\quad+\sum_{j=N_0+1}^{N}\binom{N}{j}\beta^{N-j}\mathbb{P}^{j}\left(\tilde{\mathbf{H}}_{1,j}^{1}:\exists\mathbf{x}\in\mathcal{X}_{\epsilon}^{c}\text{ s.t. }f(\mathbf{x},\tilde{\mathbf{H}}_{i}^{1})\leq 0,\forall i\in\mathcal{I}_{j}\right),\\
&\leq\sum_{j=0}^{N_0}\binom{N}{j}\beta^{N-j}+\sum_{j=N_0+1}^{N}\binom{N}{j}\beta^{N-j}\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2},
\end{aligned}\tag{5.19}$$

where (5.19) follows from Lemma 2 and the bound (3.5). The rest of this proof is tedious algebra to prove a "nice" bound on (5.19).

$$\begin{aligned}
&\sum_{j=0}^{N_0}\binom{N}{j}\beta^{N-j}+\sum_{j=N_0+1}^{N}\binom{N}{j}\beta^{N-j}\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2}\\
&=\underbrace{\sum_{j=0}^{N_0}\binom{N}{j}\beta^{N-j}\left(1-\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2}\right)}_{\tau_1}+\\
&\qquad\underbrace{\sum_{j=0}^{N}\binom{N}{j}\beta^{N-j}\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2}}_{\tau_2}.
\end{aligned}\tag{5.20}$$

To complete the proof we show that if N is large enough the terms τ_1 and τ_2 are bounded by $\tau_1\leq\delta\beta$ and $\tau_2\leq\delta(1-\beta)$, which implies that $\tau_1+\tau_2\leq\delta$. We can bound τ_1 as follows. Let $d_0=\lfloor\frac{d_f}{e}\rfloor$ where e is the base of natural logarithm. Then

$$\begin{aligned}
\tau_1&=\sum_{j=0}^{d_0}\binom{N}{j}\beta^{N-j}\left(1-\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2}\right)\\
&\qquad+\sum_{j=d_0+1}^{N_0}\binom{N}{j}\beta^{N-j}\left(1-\left(\frac{2ej}{d_f}\right)^{d_f}2^{1-\epsilon j/2}\right).
\end{aligned}\tag{5.21}$$

Note that for $\frac{d_f}{e} \leq d_0 + 1 \leq j \leq N_0 \leq \frac{8}{\epsilon}$. Thus, we have

$$\begin{aligned}
1 - \left(\frac{2ej}{d_f}\right)^{d_f} 2^{1-\epsilon j/2} &\leq 1 - \left(\frac{2ej}{d_f}\right)^{d_f} 2^{1-\epsilon N_0/2}, \\
&\leq 1 - \left(\frac{2ej}{d_f}\right)^{d_f} 2^{1-\epsilon\frac{8}{2\epsilon}}, \\
&= 1 - \left(\frac{2ej}{d_f}\right)^{d_f} 2^{-3}, \\
&\leq 1 - \left(\frac{2ed_f}{d_f e}\right)^{d_f} 2^{-3}, \\
&= 1 - 2^{d_f - 3} \leq 0.
\end{aligned}$$

The last inequality follows from the assumption that $d_f > 3$. Therefore,

$$\begin{aligned}
\tau_1 &\leq \sum_{j=0}^{d_0} \binom{N}{j} \beta^{N-j} \left(1 - \left(\frac{2ej}{d_f}\right)^{d_f} 2^{1-\epsilon j/2}\right), \\
&\leq \sum_{j=0}^{d_0} \binom{N}{j} \beta^{N-j}, \\
&\leq \frac{\beta N}{N - d_0} \sum_{j=0}^{d_0} \binom{N-1}{j} \beta^{N-1-j}, \\
&\leq \frac{N\beta(1-\beta)^{-d_0}}{N - d_0} \sum_{j=0}^{d_0} \binom{N-1}{j} \beta^{N-1-j}(1-\beta)^j, \\
&= \left(\frac{N\beta(1-\beta)^{-d_0}}{N - d_0}\right) P(1-\beta, N-1, d_0), \qquad (5.22)
\end{aligned}$$

where $P(p, N, s)$ denotes the probability of at most s successes in N IID Bernoulli trials, each with a success probability p. Then, Chernoff bound implies that

$$\tau_1 \leq \frac{N\beta(1-\beta)^{-d_0}}{N - d_0} \exp\left\{-\frac{(N-1)(1-\beta)}{2} + d_0\right\}.$$

For $N \geq d_f \geq ed_0$ we have $\frac{N}{N-d_0} \leq \frac{e}{e-1}$. Therefore,

$$\tau_1 \leq \frac{e\beta(1-\beta)^{-d_f/e}}{e-1} \exp\left\{-\frac{(N-1)(1-\beta)}{2} + \frac{d_f}{e}\right\}. \qquad (5.23)$$

Thus, $\tau_1 \leq \delta\beta$ for all

$$N \geq \frac{2d_f}{e(1-\beta)} \ln\left(\frac{e}{1-\beta}\right) + \frac{2}{1-\beta} \ln\left(\frac{e}{(e-1)\delta}\right) + 1. \qquad (5.24)$$

Next, we bound τ_2 as follows.

$$\begin{aligned}
\tau_2 &= \sum_{j=0}^{N} \binom{N}{j} \beta^{N-j} \left(\frac{2ej}{d_f}\right)^{d_f} 2^{1-\epsilon j/2}, \\
&= 2\left(\frac{2e}{d_f}\right)^{d_f} \sum_{j=0}^{N} \binom{N}{j} j^{d_f} \beta^{N-j} 2^{-\epsilon j/2}, \\
&\leq 2\left(\frac{2e}{d_f}\right)^{d_f} N^{d_f} \sum_{j=0}^{N} \binom{N}{j} \beta^{N-j} 2^{-\epsilon j/2}, \\
&= 2\left(\frac{2e}{d_f}\right)^{d_f} N^{d_f} (\beta + 2^{-\epsilon/2})^N. \qquad (5.25)
\end{aligned}$$

Since $0 < \beta + 2^{-\epsilon/2} < 1$, $\mu = 2(\frac{\epsilon}{2} + \log(\beta + 2^{-\epsilon/2}))$ is well defined and

$$\tau_2 \leq 2\left(\frac{2e}{d_f}\right)^{d_f} N^{d_f} (\beta + 2^{-\epsilon/2})^N = 2\left(\frac{2e}{d_f}\right)^{d_f} N^{d_f} (2^{-(\epsilon-\mu)/2})^N.$$

Then an analysis similar to the one given in [12] (see also [1] pg. 95 for details) shows that $\tau_2 \leq (1-\delta)\beta$ for all

$$N \geq \frac{4d_f}{\epsilon - \mu} \log\left(\frac{12}{\epsilon - \mu}\right) + \frac{4}{\epsilon - \mu} \log\left(\frac{2}{\delta(1-\beta)}\right) \qquad (5.26)$$

The result follows from (5.20), (5.24), and (5.26). ■

Since Theorem 10 is the robust analog of Corollary 1, it, too, provides only a *sufficient* condition to ensure that, with a high probability, $\mathcal{Y}[N, \beta] \subseteq \mathcal{X}_\epsilon(\mathbb{Q})$, for any fixed $\mathbb{Q} \in \mathcal{Q}$. Erdoğan and Iyengar [36] have recently extended the results in [61] to provide a *necessary* condition when $\mathbb{Q}_0$ is symmetric with a concentration of measure property and $f(\mathbf{x}, \mathbf{h})$ is Lipschitz continuous. The approach in [36] is slightly different. For details see § 3.2 and Chapter 6.

The last result in this section is the robust analog of Lemma 4 in Chapter 3.

Theorem 11 *Fix $\epsilon > \beta$. Let $\widehat{\mathbf{x}}$ denote the optimal solution of the robust sampled problem (1.8). Then $\mathbb{Q}_0^N(\mathbf{H}_{1,N}^0 : \widehat{\mathbf{x}} \notin \bar{\mathcal{X}}_\epsilon) \leq \left(\frac{eN}{n}\right)^n e^{-(\epsilon-\beta)(N-n)}$.*

Proof: The robust sampled problem (1.8) has constraints of the form

$$f(\mathbf{x}, \mathbf{z}) \leq 0, \ \forall \mathbf{z}, \ \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, \quad i = 1, \ldots, N.$$

Suppose a constraint of the form $f(\mathbf{x}, \bar{\mathbf{z}}) \leq 0$ is a support constraint for the robust sampled problem (1.8). We will associate this support constraint with
$k = \text{argmin}\left\{i : \|\bar{\mathbf{z}} - \mathbf{H}_i^0\| \leq \beta\right\}$.

Let $\mathcal{I} \subseteq \{1, \ldots, N\}$ with $|\mathcal{I}| \leq n$ and let

$$\mathcal{H}_{\mathcal{I}}^N = \{(\mathbf{h}_1, \ldots, \mathbf{h}_N) : \text{all support constraints are associated with some } i \in \mathcal{I}\}.$$

Then Theorem 3 implies that the cartesian product $\mathcal{H}^N (\equiv \bigotimes_{i=1}^N \mathcal{H}) = \cup_{\{\mathcal{I} \subseteq \{1,\ldots,N\} : |\mathcal{I}| \leq n\}} \mathcal{H}_{\mathcal{I}}^N$. Thus,

$$\begin{aligned}
&\mathbb{Q}_0^N(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \bar{\mathcal{X}}_\epsilon) \\
&\quad \leq \sum_{\{\mathcal{I} \subseteq \{1,\ldots,N\} : |\mathcal{I}| \leq n\}} \mathbb{Q}_0^N(\mathbf{H}_{1,N} \in \mathcal{H}_{\mathcal{I}}^N : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \bar{\mathcal{X}}_\epsilon) \\
&\quad = \sum_{\{\mathcal{I} \subseteq \{1,\ldots,N\} : |\mathcal{I}| \leq n\}} \Big(\mathbb{Q}_0^n(\mathbf{H}_{i \in \mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \bar{\mathcal{X}}_\epsilon) \\
&\qquad\qquad \prod_{i \notin \mathcal{I}} \mathbb{Q}_0(\mathbf{H}_i : f(\widehat{\mathbf{x}}_{\mathcal{I}}, \mathbf{z}) \leq 0,\ \forall \|\mathbf{z} - \mathbf{H}_i\| \leq \beta | \mathcal{A}_{\mathcal{I}}) \Big),
\end{aligned}$$

where $\widehat{\mathbf{x}}_{\mathcal{I}}$ denotes the optimal solution of the robust sampled problem (2.6) with only the robust constraints corresponding to the samples $i \in \mathcal{I}$ present, $\mathcal{A}_{\mathcal{I}}$ is the event $\mathcal{A}_{\mathcal{I}} = \{\mathbf{H}_{i \in \mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \bar{\mathcal{X}}_\epsilon\}$ and each term in the sum can be written as the product because $\mathbf{H}_{1,N}^0$ are IID samples. Theorem 9 implies that $\mathbb{Q}_0(\mathbf{H} : f(\widehat{\mathbf{x}}_{\mathcal{I}}, \mathbf{z}) \leq 0,\ \forall \|\mathbf{z} - \mathbf{H}\| \leq \beta | \mathcal{A}_{\mathcal{I}}) \leq e^{-(\epsilon - \beta)}$. Thus,

$$\begin{aligned}
\mathbb{Q}_0^N(\mathbf{H}_{1,N} : \widehat{\mathbf{x}} \notin \bar{\mathcal{X}}_\epsilon) &\leq e^{-(\epsilon-\beta)(N-n)} \sum_{\{\mathcal{I} \subseteq \{1,\ldots,N\} : |\mathcal{I}| \leq n\}} \mathbb{Q}_0^n(\mathbf{H}_{i \in \mathcal{I}} : \widehat{\mathbf{x}}_{\mathcal{I}} \notin \bar{\mathcal{X}}_\epsilon), \\
&\leq e^{-(\epsilon-\beta)(N-n)} \Big(\sum_{k=0}^{n} \binom{N}{k} \Big) \leq \Big(\frac{eN}{n} \Big)^n e^{-(\epsilon-\beta)(N-n)},
\end{aligned}$$

where the last inequality follows from the bound (3.5). ■

5.2 Tractability of the robust sampled problem

In Chapter 1 we introduced the robust sampled problem (1.8), as an approximation for the ambiguous chance constrained problem (5.1) and in this chapter we established bounds on the number of samples N required to approximate the robust feasible set $\bar{\mathcal{X}}_\epsilon$ and also on the number of samples required to only ensure that the optimal solution $\widehat{\mathbf{x}}$ of the robust problem (1.8) is feasible for (5.1). All along we have implicitly assumed that the robust

sampled problem (1.8) is efficiently solvable. In this section, we characterize the functions $f(\cdot,\cdot)$, the probability metric ρ and the norm $\|\cdot\|$ on the parameter space $\mathcal{H}$ for which the robust sampled problem (1.8) is tractable both in theory and in practice. The results in this section are motivated by [9].

We restrict attention to the following two classes of constraint functions.

(a) Affine functions: $\mathcal{X} = \mathbf{R}^n$, $\mathcal{H} = \{\mathbf{h} = (h_0, \bar{\mathbf{h}}) : h_0 \in \mathbf{R}, \bar{\mathbf{h}} \in \mathbf{R}^n\} = \mathbf{R}^{n+1}$, and $f(\mathbf{x}, \mathbf{h}) = h_0 + \bar{\mathbf{h}}^T\mathbf{x}$.

(b) Second-order cone functions: $\mathbf{x} \in \mathbf{R}^n$, $\mathcal{H} = \{\mathbf{h} = (\mathbf{A}, \mathbf{b}, \mathbf{u}, v) : \mathbf{A} \in \mathbf{R}^{p\times n}, \mathbf{b} \in \mathbf{R}^p, \mathbf{u} \in \mathbf{R}^n, v \in \mathbf{R}\}$, and $f(\mathbf{x}, \mathbf{h}) = \sqrt{(\mathbf{A}\mathbf{x}+\mathbf{b})^T(\mathbf{A}\mathbf{x}+\mathbf{b})} - \mathbf{u}^T\mathbf{x} - v$.

The uncertainty set $\mathcal{Q}$ considered in this paper is given by $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \le \beta\}$ where ρ_p denotes the Prohorov metric. Since the Prohorov metric is defined in terms of the norm $\|\cdot\|$ on the space $\mathcal{H}$, we first select this norm. We restrict attention to norms that satisfy

$$\|\mathbf{u}\| = \| \, |\mathbf{u}| \, \|,$$

where $|\mathbf{u}|$ denotes the vector obtained by taking the absolute value of each of the components. For a given norm $\|\cdot\|$, the constant β defining $\mathcal{Q}$ is set by the desired level of confidence. Note that β can also be set in terms of any distance measure that is an upper bound for the Prohorov metric. See Section 4 for details.

First we consider the case of affine constraint functions $f(\mathbf{x}, \mathbf{h}) = f(\mathbf{x}, (h_0, \bar{\mathbf{h}})) = h_0 + \bar{\mathbf{h}}^T\mathbf{x}$. Let $\mathbf{e}_j$, $j = 1, \ldots, n+1$ denote the j-th basis vector in $\mathbf{R}^{n+1}$. Define $\mathcal{U}(\mathbf{h}) = \big\{\mathbf{z} : \mathbf{z} = \mathbf{h} + \sum_{i=1}^{n+1} w_j\mathbf{e}_j, \|\mathbf{w}\| \le \beta\big\}$. Then the robust sampled problem (1.8) is given by

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & z_0 + \bar{\mathbf{z}}^T\mathbf{x} \le 0, \quad \forall (z_0, \bar{\mathbf{z}}) \in \mathcal{U}(\mathbf{H}_i^0), \quad i = 1, \ldots, N, \\ & \mathbf{x} \in \mathcal{X}. \end{array} \tag{5.27}$$

Results in [9] show that (5.27) can be reformulated as follows.

$$\begin{array}{lll} \min & \mathbf{c}^T\mathbf{x} & \\ \text{s.t.} & f(\mathbf{x}, \mathbf{H}_i^0) \le -\beta y_i, & i = 1, \ldots, N, \\ & |x_j| \le t_j^i, & j = 1, \ldots, n, \quad i = 1, \ldots, N, \\ & 1 \le t_{n+1}^i, & i = 1, \ldots, N, \\ & \|\mathbf{t}^i\|_* \le y_i, & i = 1, ..., N, \\ & \mathbf{y} \in \mathbf{R}^N, \ \mathbf{t}^i \in \mathbf{R}^{n+1}, & i = 1, \ldots, N, \\ & \mathbf{x} \in \mathcal{X}. & \end{array} \tag{5.28}$$

where $\|\mathbf{s}\|_* = \max_{\{\|\mathbf{r}\|\leq 1\}}\{\mathbf{s}^T\mathbf{r}\}$ denotes the dual norm of $\|\cdot\|$. For the $\mathcal{L}_1$ or $\mathcal{L}_\infty$ norms (5.28) reduces to a linear program; whereas when the norm $\|\cdot\|$ is an $\mathcal{L}_p$-norm, $p \neq \{1, \infty\}$, (5.28) is equivalent to a second-order cone program.

Next, consider the second-order cone constraints. Let $\mathbf{e}_j \in \mathbf{R}^{(p+1)(n+1)}$ denote the j-th standard basis vector in $\mathbf{R}^{(p+1)(n+1)}$. For $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbf{R}^{p\times n}$ let

$$\text{vec}(\mathbf{A}) = \begin{bmatrix}\mathbf{a}_1^T & \cdots & \mathbf{a}_n^T\end{bmatrix}^T \in \mathbf{R}^{pn},$$

and, for $\mathbf{h} = [\text{vec}(\mathbf{A})^T, \mathbf{b}^T, \mathbf{u}^T, v]^T \in \mathbf{R}^{(p+1)(n+1)}$, define $\mathcal{U}(\mathbf{h}) = \left\{\mathbf{z} : \mathbf{z} = \mathbf{h} + \sum_{j=1}^{(p+1)(n+1)} w_j\mathbf{e}_j,\ \|\mathbf{w}\| \leq \beta\right\}$. It is shown in [9] that any feasible solution to the problem (5.29) below is also feasible for the robust sampled problem (1.8), i.e. (5.29) is a conservative approximation of (1.8).

$$\begin{array}{lll}
\min & \mathbf{c}^T\mathbf{x} & \\
\text{s.t.} & f(\mathbf{x}, \mathbf{H}_i^0) \leq -\beta y_i, & i = 1, \ldots, N, \\
& g_j^i(\mathbf{x}) \leq t_j^i, & j = 1, \ldots, (p+1)(n+1), \quad i = 1, \ldots, N, \\
& \|\mathbf{t}^i\|_* \leq y_i, & i = 1, \ldots, N, \\
& \mathbf{y} \in \mathbf{R}^N, \mathbf{t}^i \in \mathbf{R}^{(p+1)(n+1)}, & i = 1, \ldots, N, \\
& \mathbf{x} \in \mathcal{X}, &
\end{array} \tag{5.29}$$

where

$$g_j^i(\mathbf{x}) = \begin{cases} |x_l| & j = p(l-1)+k,\ k = 1, \ldots, p, \quad l = 1, \ldots, n, \\ 1 & j = pn + k, \quad k = 1, \ldots, p \\ |x_l| & j = (p+1)n + l, \quad l = 1, \ldots, n \\ 1 & j = (p+1)(n+1). \end{cases}$$

The problem (5.29) is a second-order cone program for all $\mathcal{L}_p$ norms.

Chapter 6

Two-stage ambiguous chance constraints

In this chapter, we extend the approximation results to ambiguous two-stage chance constrained problems. We again assume that the limited knowledge about the distribution is characterized by the uncertainty set

$$\mathcal{Q} = \big\{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\big\}. \tag{6.1}$$

where the metric $\rho_p(\cdot, \cdot)$ denotes the Prohorov metric which is defined in Chapter 4.

We make the following additional assumptions on the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$.

Assumption 2

(a) The function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ *is bi-affine, i.e.* $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h}$*, where the vector* $\mathbf{a}_0(\mathbf{x})$ *and the matrix* $\mathbf{A}_1(\mathbf{x})$ *are affine functions of* $\mathbf{x}$*.*

(b) The extreme points $\{\bar{\boldsymbol{\lambda}}^{(i)} : i = 1, \ldots, p\}$ *of the polytope* $\{\boldsymbol{\lambda} : \mathbf{W}^T\boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq 0\}$ *are explicitly known.*

These constraints are quite restrictive and we will comment on them in § 8.2.

6.1 A sample approximation algorithm

The aim of this chapter is to establish that Algorithm SolveAmbChance displayed in Figure 6.1 produces a solution $\widehat{\mathbf{x}}$ that performs "well" for all $\mathbb{Q} \in \mathcal{Q}$. The structure of

Input: $\epsilon > 0$, $\delta \in (0,1)$, separation oracles $S_{\mathcal{X}}$ and $S_{\mathcal{R}_\beta}$
Output: $\widehat{\mathbf{x}}$
set $\mathcal{P} \leftarrow \emptyset$, $N = \lceil \frac{1}{\epsilon-\beta} \ln\left(\frac{1}{\delta}\right) \rceil$, $M = \lceil 2n^2 \ln(\frac{nR_{\mathcal{X}}^2 \|\mathbf{c}\|_2}{r\omega} + 2) \rceil$
set $\mathbf{x}_0 \leftarrow \mathbf{0}$, $\mathbf{E}_0 \leftarrow R_{\mathcal{X}}\mathbf{I}$
for t = 1, ..., M **do**
    Construct a direction vector $\mathbf{d}_t$
        $s_t \leftarrow S_{\mathcal{X}}(\mathbf{x}_{t-1})$
        if $\mathbf{x}_{t-1} \notin \mathcal{X}$, **set** $\mathbf{d}_t \leftarrow \nabla s_t$
        else
            generate $\mathbf{H}_{1,N}$ IID $\mathbb{Q}_0$
            $L_{N(t-1)+i} \leftarrow S_{\mathcal{R}_\beta}(\mathbf{x}_{t-1}; \mathbf{H}_i)$, $i = 1, \ldots, N$.
        if $\exists\ j \in \{1, \ldots, Nt\}$ such that $L_j(\mathbf{x}_{t-1}) > 0$, **set** $\mathbf{d}_t \leftarrow \nabla L_j$
        else set $\mathbf{d}_t = \mathbf{c}$ and $\mathcal{P} \leftarrow \mathcal{P} \cup \mathbf{x}_{t-1}$
    Given $(\mathbf{x}_{t-1}, \mathbf{E}_{t-1})$ and $\mathbf{d}_t$, set $(\mathbf{x}_t, \mathbf{E}_t)$ by the Ellipsoid method update
if $\mathcal{P} = \emptyset$ **return** $\widehat{\mathbf{x}} \leftarrow \emptyset$; **else return** $\widehat{\mathbf{x}} \leftarrow \text{argmin}_{\mathbf{x} \in \mathcal{P}}\{\mathbf{c}^T\mathbf{x}\}$

Figure 6.1: Algorithm SolveAmbChance

SolveAmbChance is very similar to SolveChance of § 3.2, with the following two distinctions. First, the number of samples N per iteration is given by $N = \lceil \frac{1}{\epsilon-\beta} \ln\left(\frac{1}{\delta}\right) \rceil$ (contrast with $N = \lceil \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right) \rceil$); thus ambiguity requires us to generate more samples per iteration. And second, instead of $S_{\mathcal{R}}$, Algorithm SolveAmbChance employs the oracle $S_{\mathcal{R}_\beta}$ that is a separation oracle for the set (for a fixed $\mathbf{h}$)

$$\begin{aligned} \mathcal{R}_\beta &= \Big\{\mathbf{x} \mid \mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C},\ \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta\Big\}, \\ &= \Big\{\mathbf{x} \mid\ \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta,\ \exists \mathbf{v} \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{z})\Big\}. \end{aligned} \tag{6.2}$$

The set $\mathcal{R}_\beta$ has the same structure as the feasible set of an adjustably robust linear program [4]. It is well-known that checking membership in such a set is NP-Complete when the underlying polytope is described by a set of inequalities [4]. Since we assume (see Assumption 2-(b)) that the extreme points of the underlying polytope are explicitly available, membership in $\mathcal{R}_\beta$ can be checked efficiently. From (6.2), it follows $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if, for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$,

$$\begin{aligned} 0 \quad \leq \quad P_{xz} \ = \ \max \quad & \theta \\ \text{s.t.} \quad & \mathbf{W}\mathbf{v} - \mathbf{f}(\mathbf{x}, \mathbf{z}) \geq \mathbf{1}\theta. \end{aligned} \tag{6.3}$$

It is easy to check that P_{xz} is always feasible. In order to construct the separating hyperplane, we consider the following two cases.

(i) There exists $\mathbf{y}$ such that $\mathbf{W}\mathbf{y} > \mathbf{0}$. In this case, $\mathbf{x} \in \mathcal{R}_\beta$ for all $\mathbf{h}$. Thus, $S_{\mathcal{R}_\beta}(\mathbf{x}) = \mathbf{0}$.

(ii) There does not exist $\mathbf{y}$ such that $\mathbf{W}\mathbf{y} > \mathbf{0}$. Then P_{xz} is bounded; thus, by strong duality, there is no duality gap, i.e.

$$\begin{aligned} P_{xz} &= \min\{-\big(\mathbf{f}(\mathbf{x},\mathbf{z})\big)^T\boldsymbol{\lambda} : \mathbf{W}^T\boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}, \\ &= \min\{-\big(\mathbf{f}(\mathbf{x},\mathbf{z})\big)^T\bar{\boldsymbol{\lambda}}^{(i)} : i = 1,\ldots,p\}, \end{aligned}$$

where $\{\bar{\boldsymbol{\lambda}}^{(i)} : i = 1,\ldots,p\}$ are the extreme points of the dual polytope. Recall that we have assumed that $\{\bar{\boldsymbol{\lambda}}^{(i)}\}$ are explicitly known.

Thus, $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if,

$$\begin{aligned} 0 &\leq \min\{P_{xz} : \|\mathbf{z} - \mathbf{h}\| \leq \beta\}, \\ &= \min_{1\leq i\leq p} \left\{-(\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T\bar{\boldsymbol{\lambda}}^{(i)} - \beta\|\mathbf{A}_1(\mathbf{x})^T\bar{\boldsymbol{\lambda}}^{(i)}\|\right\}. \end{aligned}$$

Suppose the minimum above is strictly negative. Let $k = \text{argmin}_{1\leq i\leq p}\big\{-(\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T\bar{\boldsymbol{\lambda}}^{(i)} - \beta\|\mathbf{A}_1(\mathbf{x})^T\bar{\boldsymbol{\lambda}}^{(i)}\|\big\}$. For a fixed $\mathbf{h}$ consider the convex function $g(\mathbf{u};\mathbf{h}) : \mathbf{R}^n \mapsto \mathbf{R}$ defined as follows

$$g(\mathbf{u};\mathbf{h}) = (\mathbf{a}_0(\mathbf{u}) + \mathbf{A}_1(\mathbf{u})\mathbf{h})^T\bar{\boldsymbol{\lambda}}^{(k)} + \beta\|\mathbf{A}_1(\mathbf{u})^T\bar{\boldsymbol{\lambda}}^{(k)}\|.$$

Then any sub-gradient of the function $g(\cdot;\mathbf{h})$ at $\mathbf{u} = \mathbf{x}$ serves as the separating hyperplane.

6.2 Analysis of the algorithm

The following result extends Theorem 4 to the ambiguous setting. Note that for all the results in this chapter the relevant probability measure is the product measure $\mathbb{Q}_0^{MN}$, since all the samples are drawn independently from the central measure $\mathbb{Q}_0$.

Theorem 12 *Suppose the output $\widehat{\mathbf{x}}$ of* SolveAmbChance *is well-defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$. Then, for every fixed $\mathbb{Q} \in \mathcal{Q}$, we have that $\mathbb{Q}(\mathbf{H} : \mathbf{f}(\widehat{\mathbf{x}},\mathbf{H}) \notin \mathcal{C}) > \epsilon$ with a probability at most $M\delta$.*

Proof: Fix a measure $\mathbb{Q} \in \mathcal{Q}$ and let $\widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \{\mathbf{x} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{x}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon\}$. By construction, the events $\{\widehat{\mathbf{x}} \neq \emptyset\} = \cup_{t=0}^{M-1}\{\mathbf{x}_t \in \mathcal{P}\}$. Consequently,

$$\{\widehat{\mathbf{x}} \neq \emptyset\} \cap \{\widehat{\mathbf{x}} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\} \subseteq \cup_{t=0}^{M-1}\big(\{\mathbf{x}_t \in \mathcal{P}\} \cap \{\mathbf{x}_t \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\}\big). \tag{6.4}$$

Fix t. Let B_t denote the event that $\mathbf{x}_{t-1}$ satisfies all the N inequalities generated by the oracle S_{R_β} at iteration t. Define

$$\mathcal{Y}_t[N, \beta] = \left\{\mathbf{x} \in \mathcal{X} \mid \mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C},\ \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, i = N(t-1)+1, \ldots, Nt\right\}, \tag{6.5}$$

where $\mathbf{H}_{N(t-1)+1,Nt}^0$ denote N IID samples drawn according to the central probability measure $\mathbb{Q}_0$ at iteration t. Then it is clear that the event $B_t = \{\mathbf{x}_{t-1} \in \mathcal{Y}_t[N, \beta]\}$.

Let A_t denote the event that the iterate $\mathbf{x}_{t-1}$ satisfies all the $N(t-1)$ inequalities generated by the oracle $S_{\mathcal{R}_\beta}$ *before* iteration t. Then it is clear that $\{\mathbf{x}_{t-1} \in \mathcal{P}\} = A_t \cap B_t$. Thus,

$$\begin{aligned}
\mathbb{Q}_0^{Nt}\big(A_t \cap B_t \cap \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\}\big) &\leq \mathbb{Q}_0^{Nt}\big(B_t \cap \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\}\big), \\
&\leq \mathbb{Q}_0^{Nt}\big(B_t \mid \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\big), \\
&\leq \delta, \qquad (6.6)
\end{aligned}$$

where the bound (6.6) follows from Theorem 6 in [34]. The result follows applying the union bound to the expression in (6.4). ■

As before, define

$$\mathcal{X}_\mathcal{I} = \{\mathbf{x} \in \mathcal{X} : L_j(\mathbf{x}) \leq 0, j = 1, \ldots, MN\},$$

where L_j, $j = 1, \ldots, MN$, denote the set of linear inequalities generated by the oracle $S_{\mathcal{R}_\beta}$ over the course of Algorithm SolveAmbChance. Then a simple extension of the technique used to prove Lemma 5 establishes the following.

Lemma 7 *Suppose the set $\mathcal{X}_\mathcal{I}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r. Then the solution $\widehat{\mathbf{x}}$ returned by the* SolveAmbChance *is well defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$.*

For $\alpha > 1$, $r > 0$, and a fixed measure $\mathbb{Q} \in \mathcal{Q}$, let

$$\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) = \left\{\mathbf{x} \in \mathcal{X}^{-r} \,\middle|\, \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{x}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}) > 1 - \epsilon\right\}, \tag{6.7}$$

where

$$\mu_r = (\alpha + 1)\beta K_f^\mathcal{X}\big(R_\mathcal{X}\big) + rK_f^\mathcal{H}\big(R_\mathcal{H} + \beta\big). \tag{6.8}$$

The set $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r)$ denotes the set of points that are robustly feasible for the chance constraint corresponding to the measure $\mathbb{Q}$.

Theorem 13 *Fix $\mathbf{y} \in \mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r)$. Then the Euclidean ball $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$ with a probability at least $1 - MNe^{-\psi(\alpha, 1-\epsilon-\beta)}$.*

Proof: Let $\mu_r = (\alpha+1)\beta K_f^\mathcal{X}\big(R_\mathcal{X}\big) + rK_f^\mathcal{H}\big(R_\mathcal{H} + \beta\big)$. Then, we have that

$$\begin{aligned}
1-\epsilon &< \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), && (6.9)\\
&\leq \beta + \mathbb{Q}_0(\mathbf{H} + \mathbf{u} : \|\mathbf{u}\| \leq \beta, \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), && (6.10)\\
&\leq \beta + \mathbb{Q}_0(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^\mathcal{X}(R_\mathcal{X})}), && (6.11)\\
&= \beta + \mathbb{Q}_0(\alpha^{-1}\mathbf{H} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{H}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^\mathcal{X}(R_\mathcal{X})}), && (6.12)
\end{aligned}$$

where the inequalities (6.9) and (6.10), respectively, follow from the definitions of $\mathbf{y}$ and the Prohorov metric, and the inequality (6.11) follows from the Lipschitz continuity of the function f.

Let $\mathcal{H}_\mathbf{y} = \{\mathbf{h} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{h}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^\mathcal{X}(R_\mathcal{X})}\}$. Then (6.12) and the concentration property of $\mathbb{Q}_0$ imply that $\mathbb{Q}_0(\mathcal{H}_\mathbf{y}) \geq 1 - e^{-\psi(\alpha, 1-\epsilon-\beta)}$ provided $1-\epsilon-\beta > \bar{\theta}$. By Assumption 1 we have

$$\mathcal{H}_\mathbf{y} \subseteq \Big\{\mathbf{h} : \mathbf{f}(\mathbf{y}, \mathbf{z}) \in \mathcal{C}^{-\mu_r + (\alpha+1)\beta K_f^\mathcal{X}(R_\mathcal{X})}, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta\Big\}. \tag{6.13}$$

Thus, for all $\mathbf{h} \in \mathcal{H}_\mathbf{y}$ and $\mathbf{x} \in \mathcal{B}_r(\mathbf{y})$, Assumption 1 implies that for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$, we have that

$$\begin{aligned}
\|\mathbf{f}(\mathbf{x}, \mathbf{z}) - \mathbf{f}(\mathbf{y}, \mathbf{z})\| &\leq K_f^\mathcal{H}(\|\mathbf{z}\|)\|\mathbf{x} - \mathbf{y}\|,\\
&\leq rK_f^\mathcal{H}(\|\mathbf{h}\| + \beta),\\
&\leq rK_f^\mathcal{H}(R_\mathcal{H} + \beta). && (6.14)
\end{aligned}$$

Since $\mathbf{h} \in \mathcal{H}_\mathbf{y}$, (6.13) and (6.14) imply that for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$,

$$\mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C}^{-\mu_r + (\alpha+1)\beta K_f^\mathcal{X}(R_\mathcal{X}) + rK_f^\mathcal{H}(R_\mathcal{H} + \beta)} = \mathcal{C}^0 = \mathcal{C}.$$

Consequently,

$$\begin{aligned}
&\mathbb{Q}_0^{MN}(\mathbf{H}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \not\subseteq \mathcal{X}_\mathcal{I})\\
&\quad\leq \mathbb{Q}_0^{MN}((\mathbf{H}_{1,MN} : \mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C}, \forall \mathbf{x} \in \mathcal{B}_r(\mathbf{y}), \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i\| \leq \beta, i = 1, \ldots, MN)^c),\\
&\quad\leq MN\mathbb{Q}_0(\mathcal{H}_\mathbf{y}^c) \leq MNe^{-\psi(\alpha, 1-\epsilon-\beta)}.
\end{aligned}$$

■

The following corollary establishes that the output of SolveAmbChance is well-defined with a high probability.

Corollary 5 *Suppose* $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) \neq \emptyset$. *Then the output* $\widehat{\mathbf{x}}$ *of* SolveAmbChance *is well-defined (i.e.* $\widehat{\mathbf{x}} \neq \emptyset$*) with a probability at least* $1 - MNe^{-\psi(\alpha, 1-\epsilon-\beta)}$.

Proof: Lemma 7 implies that the solution $\widehat{\mathbf{x}}$ is well defined if the set $\mathcal{X}_{\mathcal{I}}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r and the probability of such event is, by Theorem 13, at least $1 - MNe^{-\psi(\alpha, 1-\epsilon-\beta)}$. ■

Remark 3 *By setting* $\beta = 0$, *we recover the corresponding "unambiguous" versions of Theorem 12 and Corollary 5, namely Theorem 4 and Corollary 3.*

We are now in position to state the main result of this chapter. Consider the chance constrained problem,

$$\begin{array}{ll} \min & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \Big\{ \mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon \Big\}, \end{array} \tag{6.15}$$

corresponding to a measure $\mathbb{Q} \in \mathcal{Q}$. We will call $\widehat{\mathbf{x}}$ an $(\mathbb{Q}, \alpha, r, \omega)$-approximate solution of (6.15) if

$$\mathbf{c}^T \widehat{\mathbf{x}} \leq \inf_{\mathbf{x} \in \mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r)} \mathbf{c}^T \mathbf{x} + \omega. \tag{6.16}$$

Then the following result holds.

Theorem 14 *For all* $\mathbb{Q} \in \mathcal{Q}$ *such that* $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) \neq \emptyset$, SolveAmbChance *returns an* $(\mathbb{Q}, \alpha, r, \omega)$*-approximate solution with a probability at least* $(1 - MNe^{-\psi(\alpha, 1-\epsilon-\beta)})(1 - M\delta)$.

The proof of this result proceeds along the lines of Theorem 6.

Chapter 7

Computational Experiments

In this chapter we illustrate the techniques proposed in this dissertation and compare the ambiguous chance constrained programs with their non-ambiguous versions. We will consider two of the applications introduced in § 2.1: the single stage portfolio selection problem with Value at Risk constraints and the two-stage network design problem.

7.1 Portfolio Selection with VaR Constraints

Let us consider the portfolio selection problem with the VaR constraint, which was explained in detail in § 2.1.1:

$$\begin{array}{ll} \max & \boldsymbol{\mu}^T \mathbf{x} \\ \text{s.t.} & \mathbb{Q}(\mathbf{H} : \mathbf{H}^T \mathbf{x} \geq \eta) \geq 1 - \epsilon, \\ & \sum_{j=1}^{n} \mathbf{x}_j = 1, \end{array} \tag{7.1}$$

where $\mathbf{x}$ is the portfolio, $\mathbf{H} \sim \mathbb{Q}$ is the random variable denoting assets returns, and $\mathbf{E}[\mathbf{H}] = \boldsymbol{\mu}$. For simplicity, we will assume that the mean returns $\boldsymbol{\mu}$ are known and fixed.

As explained in Chapter 3 the problem (7.1) can be solved efficiently only if the distribution $\mathbb{Q}$ has some very specific properties. For general distribution functions, the only reasonable way to approximate the problem (7.1) is to use the sampling methods explained in Chapter 3.

Furthermore, although the bounds on the sample size N given in Chapter 3 are distribution-free, one must be able to sample from $\mathbb{Q}$ in order to construct the sample approximation to problem (7.1). Also, there is an implicit assumption that the distribution $\mathbb{Q}$ of the random returns is fixed. However, in real life the probability distributions are never known

exactly! Therefore, we are in a position to use the ambiguous chance constrained programs to compute an optimal portfolio.

The goal in our computational experiments is to compare performance of the ambiguous chance constrained portfolio selection with that of the classical (non-ambiguous) chance constrained program. Throughout these experiments we assumed that our investment universe consists of $n = 30$ stocks and we set the parameters $\eta = -0.4$, $\epsilon = 0.1$, and $\delta = 0.05$. The true return distribution $\mathbb{Q}$ is assumed to belong to the uncertainty set

$$\mathcal{Q} = \big\{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\big\}, \tag{7.2}$$

where ρ_p is the Prohorov metric (see Chapter 4) and $\beta = 0.05$. The central distribution $\mathbb{Q}_0$, an estimate of the underlying distribution, is assumed to be the following multivariate Laplace distribution:

$$\mathbb{Q}_0 = \boldsymbol{\mu} + \sqrt{\mathbf{Y}}\mathbf{Z},$$

where $\mathbf{Y} \sim Expo(1)$ is the exponential distribution with mean 1 and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. The classical chance constrained program assumes that the distribution is known perfectly and uses the distribution $\mathbb{Q}_0$ in the chance constrained problem:

$$\begin{array}{rl} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t.} & \mathbb{Q}_0(\mathbf{H} : \mathbf{H}^T\mathbf{x} \geq \eta) \geq 1 - \epsilon, \\ & \sum_{j=1}^n \mathbf{x}_j = 1, \end{array} \tag{7.3}$$

and approximates the problem (7.3) by the following sampled problem:

$$\begin{array}{rl} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t.} & (\mathbf{H}_i^0)^T\mathbf{x} \geq \eta, \quad i = 1, \ldots, N_c \\ & \sum_{j=1}^n \mathbf{x}_j = 1, \end{array} \tag{7.4}$$

where $\mathbf{H}_i^0$, $i = 1, \ldots, N_c$, are N_c IID samples from $\mathbb{Q}_0$. The ambiguous chance constrained version of (7.3) is

$$\begin{array}{rl} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t.} & \mathbb{Q}(\mathbf{H} : \mathbf{H}^T\mathbf{x} \geq \eta) \geq 1 - \epsilon, \forall \mathbb{Q} \in \mathcal{Q} \\ & \sum_{j=1}^n \mathbf{x}_j = 1. \end{array} \tag{7.5}$$

The robust sampled problem that approximates (7.5) is:

$$\begin{array}{rl} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t.} & \mathbf{z}^T\mathbf{x} \geq \eta, \quad \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, \quad i = 1, \ldots, N_a \\ & \sum_{j=1}^n \mathbf{x}_j = 1, \end{array} \tag{7.6}$$

where $\mathbf{H}_i^0$, $i = 1, \ldots, N_a$, are N_a IID samples from $\mathbb{Q}_0$ and $\|\cdot\|$ is the Euclidean norm. The choice of the norm $\|\cdot\|$ and the parameter are closely related (see Chapter 4.)

Note that both of the sampled problems (7.4) and (7.6) use samples from the central distribution $\mathbb{Q}_0$. Differences between these problems are the robust sampled problem has parameter uncertainty sets around each sample and it uses more samples $N_a \geq N_c$, the equality holds if and only if $\beta = 0$.

We set the number of samples used by (7.4), N_c, according to Corollary 2 and the number of samples in (7.6), N_a, according to Theorem 11. This theorem is the robust analog of Corollary 2. The number of samples in this experiment were $N_c = 1918$ and $N_a = 4607$.

In our computational experiments we compute optimal solutions to the problems (7.4) and (7.6), denoted by $\mathbf{x}_c$ and $\mathbf{x}_a$, respectively, using samples from $\mathbb{Q}_0$ but test performances of these solutions under another measure $\mathbb{S}$. In particular, we assume that the unknown true distribution is $\mathbb{S}$ and $\mathbb{Q}_0$ is an estimate of this unknown distribution. Then we estimate the expected returns and the "infeasibility" (or error) of the solutions. The error of a portfolio $\mathbf{x}$ under a measure $\mathbb{S}$ is defined as

$$\text{err}(\mathbf{x}) = \mathbb{S}(\mathbf{H} : \mathbf{H}^T\mathbf{x} < \eta)$$

The error $\text{err}(\mathbf{x})$ and the expected return $\text{Ret}(\mathbf{x})$ of a portfolio $\mathbf{x}$ under a measure $\mathbb{S}$ is estimated by Monte Carlo methods using samples from $\mathbb{S}$. As one can see a portfolio $\mathbf{x}$ has a better performance than a portfolio $\mathbf{y}$ if $\mathbf{x}$ has a smaller error and a higher expected return, i.e., $\text{err}(\mathbf{x}) < \text{err}(\mathbf{y})$ and $\text{Ret}(\mathbf{x}) > \text{Ret}(\mathbf{y})$.

To compare performances of the portfolios $\mathbf{x}_c$ and $\mathbf{x}_a$, we assumed that the true distribution $\mathbb{S}$ has a negatively skewed Multivariate Laplace distribution given by

$$\mathbb{S} = \boldsymbol{\mu} - \zeta|\boldsymbol{\mu}|\mathbf{Y} + \sqrt{\mathbf{Y}}\mathbf{Z}$$

where $\mathbf{Y} \sim Expo(1)$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $\zeta \geq 0$ is a constant that controls the skewness of the distribution $\mathbb{S}$. In addition to the negative skewness of $\mathbb{S}$, it also has heavier tails than the normal distribution. Therefore, the assumption that the asset returns have negatively skewed Multivariate Laplace distribution is reasonable in the context of finance [52]. It is easy to see that as ζ goes to 0, the evaluation distribution $\mathbb{S}$ converges to the central distribution $\mathbb{Q}_0$ and as the value of ζ increases skewness of $\mathbb{S}$ increases. In Table 7.1, we present the errors and the expected returns of the solutions $\mathbf{x}_c$ and $\mathbf{x}_a$ for different values of the skewness parameter ζ. As we mentioned in Chapter 4 that the Prohorov metric is very

ζ	err($\mathbf{x}_a$)	Ret($\mathbf{x}_a$)	err($\mathbf{x}_c$)	Ret($\mathbf{x}_c$)
0	0.0026	0.0793	0.0034	0.0984
1.0	0.0217	0.0001	0.0317	0.0004
1.5	**0.0444**	**- 0.0397**	**0.0637**	**- 0.0490**
2.0	**0.0753**	**- 0.0799**	**0.1052**	**- 0.0987**

Table 7.1: Performance of the solutions of sampled problems (7.4), $\mathbf{x}_c$, and (7.6), $\mathbf{x}_a$.

hard to work with. Unfortunately we were not able to prove that whether the measure $\mathbb{S}$ belongs to $\mathcal{Q}$ or not for the values of ζ.

The results in Table 7.1 imply that when ζ is small, i.e., the true distribution $\mathbb{S}$ is close to its estimate $\mathbb{Q}_0$, both of the solutions $\mathbf{x}_c$ and $\mathbf{x}_a$ have error less than $\epsilon = 0.1$. Therefore these solutions are feasible to the chance constraint. Also, the portfolio $\mathbf{x}_c$ has higher error and higher expected return. Therefore, we can conclude that when the evaluation distributions are close to the central distribution, the ambiguous chance constrained programs produce more conservative solutions, as one might expect. However, contributions of the ambiguous chance constrained programs to the area of optimization under uncertainty becomes clear when the true distribution $\mathbb{S}$ is significantly different than $\mathbb{Q}_0$, i.e., when the value of ζ is high. As seen in Table 7.1, when $\zeta > 1$ then $\mathbf{x}_a$ has lower error and higher expected return. Also, when $\zeta = 2$ the portfolio $\mathbf{x}_c$ has error higher than $\epsilon = 0.1$ and therefore it is infeasible to the chance constraint.

7.2 Network Design

In this section we illustrate the techniques proposed for the two-stage ambiguous chance constrained programs by solving a two-stage network design on the simple network displayed in Figure 7.1. The node s is a source node with an infinite capacity and the nodes 1,2, and 3 are sink nodes with demands given by the vector $\mathbf{d} = (d_1, d_2, d_3)^T \geq \mathbf{0}$. See also § 2.1.4.

For a given demand vector $\mathbf{d}$ the network design problem is given by

$$\begin{array}{lrcl} \min & \mathbf{c}^T\mathbf{u} & & \\ \text{s.t.} & \mathbf{A}\mathbf{v} & \geq & \mathbf{d}, \\ & \mathbf{u} - \mathbf{v} & \geq & \mathbf{0}, \\ & \mathbf{u}, \mathbf{v} \geq \mathbf{0}, & & \end{array} \tag{7.7}$$

where the cost vector $\mathbf{c}$ is assumed to be strictly positive, $-\mathbf{A}$ denotes the node-arc incidence

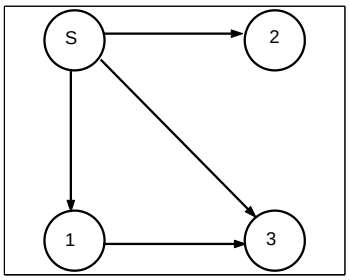

Figure 7.1: Chance Constrained Network Design

matrix of the network, the variable $\mathbf{u}$ denotes the capacity on the arcs, and the variable $\mathbf{v}$ denotes the flow on the network.

Once the network is constructed, i.e. the capacities $\mathbf{u}$ are installed on the arc, we assume that it will be used over a reasonably long period over which the demand $\mathbf{d}$ can change. We model variation in demand by assuming that it is random. In particular we assume that the random demand $\mathbf{D} \sim \mathbb{Q}$ where $\mathbb{Q} \in \mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}_0, \mathbb{Q}) \leq \beta\}$, $\mathbb{Q}_0 = \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$ is a multivariate normal distribution with mean $\bar{\mathbf{d}}$ and covariance matrix $\sigma^2\mathbf{I}$, and $\mathbf{I}$ denotes the Identity matrix. Note that the flow conservation constraints $\mathbf{Av} \geq \mathbf{d}$ in (7.7) are formulated as inequalities instead of equalities as is usually the case – this is necessary to accommodate random demands $\mathbf{D}$.

Although the capacity cannot, typically, be altered over the life of the network, the flow $\mathbf{v}$ is chosen *after* observing the realization of the demand. Thus, the network design problem is a two-stage optimization problem: the capacities $\mathbf{u}$ are the first stage variables, and the flows $\mathbf{v} \in \mathbf{R}^n$ are the second stage variables. The objective of the two-stage optimization problem is to find a minimal cost capacity allocation $\mathbf{u}$ that guarantees that at least $(1-\epsilon)$-fraction of the random demand $\mathbf{D}$ can be feasibly routed in the resulting network. One can transform this network design problem into a chance constrained problem of the form (3.8) by setting

$$\mathbf{x} = \mathbf{u}, \quad \mathbf{H} = \mathbf{D}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{A} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}, \quad f(\mathbf{x}, \mathbf{H}) = \begin{bmatrix} \mathbf{H} \\ -\mathbf{x} \\ \mathbf{0} \end{bmatrix}, \quad \mathcal{X} = \mathbf{R}^n_+.$$

In our computational experiments, we compare performances of three different solution strategies:

(a) Deterministic solution: $\mathbf{x}_{det}$ denotes a solution of the deterministic optimization prob-

lem (7.7) with the demand vector $\mathbf{D}$ is set equal to its mean value $\bar{\mathbf{d}}$.

(b) Chance constrained solution: $\mathbf{x}_{sc}$ is an output of Algorithm SolveChance given in Figure 3.1.

(c) Ambiguous chance constrained solution: $\mathbf{x}_{sac}$ is an output of Algorithm SolveAmbChance given in Figure 6.1.

Instead of simply verifying the theoretical results presented in the previous sections, we have attempted to investigate issues that we were not able to settle theoretically. For example, we test the hypothesis that the ambiguous chance constrained solution is very conservative (and protects against a set of measures much larger than $\mathcal{Q}$) by using different test distributions $\mathbb{Q}$.

7.2.1 Algorithmic details

Oracles and Sampling SolveChance uses two oracles, $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$, and SolveAmbChance exploits $S_{\mathcal{X}}$ and $S_{\mathcal{R}_\beta}$. Since $\mathcal{X} = \mathbf{R}^n_+$ and the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine, (3.11) implies that $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$ are polynomial-time oracles. The oracle $S_{\mathcal{R}_\beta}$ requires explicit characterization of extreme points of a polytope, therefore is not polynomial for some cases. However, for a network design problem, Atamtürk and Zhang [3] establish that $S_{\mathcal{R}_\beta}$ is a polynomial-time oracle for some special networks.

The algorithms SolveChance and SolveAmbChance use samples from $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. Since SolveAmbChance uses more samples than SolveChance, to avoid generating too many samples, we first run SolveAmbChance and use the samples generated during its run to calculate an output of SolveChance.

Stopping criterion The number of iterations M in SolveChance and SolveAmbChance (see Figure 3.1 or Figure 6.1) was chosen large enough to guarantee convergence in the underlying Ellipsoid-like algorithm. In fact, any stopping rule that guarantees convergence in the Ellipsoid algorithm is sufficient for the results in this paper to hold.

In our experiments we use a new stopping rule proposed in [8]. Let M denote the upper bound defined in Figure 3.1. For $t = 1, \ldots, M$, define $v_t = R_{\mathcal{X}}\|\mathbf{d}_t\| - \mathbf{d}_t^T\mathbf{x}_{t-1}$, $V_t = \max\{V_{t-1}, v_t\}$, with $V_0 = 0$, and $\rho_t = |\det(\mathbf{E}_t)|^{1/n}$. Then, we terminate the algorithm at any iteration $t \leq M$ when

$$\frac{\rho_t}{r} < \frac{\omega}{V_t + \omega}.$$

β	$N_{asc}(\beta)$	$M \times N_{asc}(\beta)$	$\widehat{M} \times N_{asc}(\beta)$	$(M - \widehat{M}) \times N_{asc}(\beta)$
0.0300	150	78150	69600	8550
0.0200	100	52100	46400	5700
0.0150	86	44806	39904	4902
0.0100	75	39075	34800	4275
0.0050	67	34907	31088	3819
0	60	31260	27840	3420

Table 7.2: Number of iterations and samples as a function of the stopping rule

In our numerical experiments $M = 521$ whereas the average number of iterations $\widehat{M}$ with the new stopping rule was approximately $\widehat{M} = 464$. This reduction in the number of iterations significantly reduces the number of samples generated. In Table 7.2, the column labeled $N_{asc}(\beta)$ lists the number of samples required per iteration of SolveAmbChance for a given $\beta > 0$ ($N_{asc}(0)$ corresponds to SolveChance), the columns labeled $M \times N_{asc}(\beta)$ and $\widehat{M} \times N_{asc}(\beta)$ list the total number of samples needed with the old and the new stopping rule, respectively, and the last column summarizes the decrease in the number of samples.

Performance evaluation We assume that the distribution $\mathbb{Q}$ of the random demand $\mathbf{D}$ is uncertain and is only known to belong to the uncertainty set $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}_0, \mathbb{Q}) \leq \beta, \mathbb{Q}_0 \sim \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})\}$. Since checking whether a measure $\mathbb{Q} \in \mathcal{Q}$ is hard, we test the performance of the solutions using measures that are "similar" to $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. In Section 7.2.2 we discuss performance of the solutions when the test measure $\mathbb{Q}$ is another Normal distribution with different parameters and in Section 7.2.3 we discuss the results for Cauchy and Pareto distributions.

The performance of the solutions was estimated by out-of-sample testing. We generated 50 batches of 1000 samples each (i.e. a total of 50000 samples). The error of a solution $\mathbf{x}$ with respect to a batch k, $k = 1, \ldots, 50$, is given by

$$\text{err}_k(\mathbf{x}) = \frac{1}{1000} \sum_{j=1}^{1000} (1 - \mathbf{1}_{\mathcal{R}(\mathbf{H}_j)}(\mathbf{x}))$$

where $\mathbf{1}_{\mathcal{A}}$ is the indicator function of set $\mathcal{A}$ and $\mathcal{R}(\mathbf{H}_j) = \{\mathbf{x} \mid \exists \mathbf{v} \in \mathbf{R}^k \text{ s.t. } \mathbf{W}\mathbf{v} \geq f(\mathbf{x}, \mathbf{H}_j)\}$. The error of $\mathbf{x}$, $\text{err}(\mathbf{x})$, is defined as the average of $\text{err}_k(\mathbf{x})$ over the 50 batches.

Problem parameters The mean demand vector $\bar{\mathbf{d}}$ was set to $\bar{\mathbf{d}} = (3, 2, 2)^T$ and standard deviation σ was set to $\sigma = 0.5 \times 10^{-2}$. The unit capacity costs are taken as $c_{s1} = 2$, $c_{s2} = 0.5$,

	$\mathbf{x}_{asc}$			$\mathbf{x}_{sc}$			$\mathbf{x}_{det}$		
$\widehat{\beta}$	err	stdev	c	err	stdev	c	err	stdev	c
0.0475	0	0	11.3888	**0.2131**	0.0019	11.1810	**1.0000**	0.0000	11.0000
0.0450	0	0	11.3790	**0.2623**	0.0014	11.1814	**0.9999**	0.0001	11.0000
0.0400	0.0000	0.0000	11.3277	**0.0955**	0.0014	11.1873	**0.9997**	0.0001	11.0000
0.0300	0.0000	0.0000	11.2912	**0.0707**	0.0014	11.1707	**0.9983**	0.0002	11.0000
0.0250	0.0000	0.0000	11.2798	0.0137	0.0006	11.1910	**0.9954**	0.0003	11.0000
0.0200	0.0001	0.0001	11.2509	0.0130	0.0005	11.1799	**0.9894**	0.0005	11.0000
0.0150	0.0001	0.0001	11.2358	0.0087	0.0003	11.2074	**0.9775**	0.0006	11.0000
0.0100	0.0002	0.0001	11.2162	0.0045	0.0003	11.1834	**0.9577**	0.0010	11.0000
0.0050	0.0005	0.0001	11.1922	0.0015	0.0002	11.1790	**0.9233**	0.0010	11.0000
0.0025	0.0003	0.0001	11.1916	0.0010	0.0001	11.1875	**0.9033**	0.0013	11.0000
0.0005	0.0005	0.0001	11.1759	0.0027	0.0002	11.1680	**0.8782**	0.0014	11.0000

Table 7.3: Performance when the test distribution $\mathbb{Q} = \mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$.

$c_{s3} = 2$, and $c_{13} = 0.1$. The target violation probability ϵ was set to $\epsilon = 0.05$. The other parameters $\delta = 0.05$, $r = 10^{-2}$, and $\omega = 10^{-2}$.

7.2.2 Experiments with Normal distributions

In this section we report the performances of the three solutions when the test samples are drawn from $\mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$, where $\mathbf{e}$ is the vector of all ones and $\widehat{\beta} = 0.0475$, i.e. a Normal distribution with a shifted mean. As indicated in the previous section, the training distribution $\mathbb{Q}_0 = \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. The solutions $\mathbf{x}_{sc}$ and $\mathbf{x}_{asc}$ were computed from one run of Algorithms SolveChance and SolveAmbChance , respectively. We will comment on this later in this section.

Table 7.3 displays the performance of the three solutions strategies as a function of β. The columns labeled "err" list the empirical estimate of the violation probability, the column labeled "stdev" lists the standard deviation of the empirical estimate, and the column labeled "c" lists the cost of the solution. We highlight the instances where empirical estimate "err" violates the bound $\epsilon = 0.05$ by listing the value in bold face. The constant β defining the uncertainty set $\mathcal{Q}$ (see (4.1)) for the ambiguous chance constraint problem was set equal to $\widehat{\beta}$. We can draw the following conclusions from the results displayed in Table 7.3.

	$\mathbf{x}_{asc}(\beta)$		
β	err($\mathbf{x}_{asc}(\beta)$)	stdev($\mathbf{x}_{asc}(\beta)$)	$c(\mathbf{x}_{asc}(\beta))$
0.0475	0	0	11.3799
0.0450	0	0	11.3589
0.0400	0	0	11.3393
0.0300	0.0020	0.0002	11.2867
0.0250	0.0015	0.0002	11.2935
0.0200	0.0078	0.0004	11.2570
0.0150	0.0184	0.0006	11.2406
0.0100	**0.0632**	0.0013	11.2154
0.0050	**0.0776**	0.0011	11.2109
0.0025	**0.2176**	0.0018	11.1806
0.0005	**0.2319**	0.0018	11.1790

Table 7.4: Performance when the test distribution $\mathbb{Q} = \mathcal{N}\big(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I}\big)$, with $\widehat{\beta} = 0.0475$.

(i) The deterministic solution $\mathbf{x}_{det}$ does not have any tolerance for variance in demand. This is not surprising since it completely ignores the distributional information.

(ii) The chance constrained solution $\mathbf{x}_{sc}$ has more tolerance for random variations in demand as well as changes in the underlying distribution – the chance constrained solution meets bound on the violation probability $\epsilon = 0.05$ for all values of $\beta \leq 0.025$. However, this tolerance come with a higher capacity cost. As β is increased beyond the threshold value of 0.025, the performance of the chance constrained solution deteriorates very sharply.

(iii) The ambiguous chance constrained solution never violates the bound $\epsilon = 0.05$. In fact, setting the constant β defining $\mathcal{Q}$ equal to $\widehat{\beta}$ of the test distribution results in very conservative solutions – the violation probability estimates are considerably smaller than the allowable bound of ϵ. This also results in a correspondingly higher capacity cost.

Since setting $\beta = \widehat{\beta}$ resulted in very conservative solutions, we next investigated whether setting $\beta << \widehat{\beta}$ results in feasible solutions with costs comparable to that of the chance constrained solution $\mathbf{x}_{sc}$. Table 7.4 displays the performance of the ambiguous chance constrained solutions for different values of β when $\widehat{\beta} = 0.0475$. We see that $\mathbf{x}_{asc}(\beta)$ for all $\beta \geq 0.015$ is feasible with respect to $\widehat{\beta}$. Moreover, the cost of $\mathbf{x}_{asc}(0.015)$ is comparable

$\widehat{\mathbf{x}}_{sc}$		
err	stdev	c
0.1853	0.0017	11.1851

	$\widehat{\mathbf{x}}_{asc}(\beta)$		
β	err	stdev	c
0.0300	0.0005	0.0001	11.3006
0.0200	0.0071	0.0004	11.2599
0.0150	0.0220	0.0007	11.2413
0.0100	0.0462	0.0010	11.2223
0.0050	**0.1150**	0.0015	11.1997

Table 7.5: Performances of $\widehat{\mathbf{x}}_{asc}(\beta)$ and $\widehat{\mathbf{x}}_{ac}$ when $\mathbb{Q} = \mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$, $\widehat{\beta} = 0.0475$.

to that of the chance constrained solution $\mathbf{x}_{sc}$ while at the same time providing insurance against all test distributions with $0 \leq \widehat{\beta} \leq 0.0475$. These results reiterate that $\mathbf{x}_{asc}(\beta)$ is very conservative for a given level β; therefore, the designer has to carefully select β to protect against over-designing.

The chance constraint is a non-convex constraint and the output of Algorithms SolveChance and SolveAmbChance are both random. Erdoğan and Iyengar [35] show that taking a convex combination of samples of a random solution significantly improves the performance in such a situation. We test this result by studying the performance of the solutions $\widehat{\mathbf{x}}_{sc}$ and $\widehat{\mathbf{x}}_{asc}$ that are generated by taking the empirical average over $N = 50$ independent runs. Table 7.5 presents the results related to $\widehat{\mathbf{x}}_{asc}(\beta)$ and $\widehat{\mathbf{x}}_{sc}$. It is clear from the results in Table 7.5 that averaging improves the performance. In particular, the $\widehat{\mathbf{x}}_{asc}(\beta)$ is feasible for $\widehat{\beta} = 0.0475$ even when one sets $\beta = 0.01$.

7.2.3 Experiments with other distributions

Table 7.6 displays the performance of the three solution strategies when the test distribution $\mathbb{Q}$ is a multivariate Cauchy distribution with median $\bar{\mathbf{d}}$, i.e. each $\mathbf{D}_i$ is a Cauchy random variable with median $\mathbf{d}_i$, $i = 1, 2, 3$. We assumed that $\mathbf{D}_i$, $i = 1, 2, 3$ are independent. Once again we see that the deterministic solution $\mathbf{x}_{det}$ and the chance constrained solution $\mathbf{x}_{sc}$ are both infeasible. The ambiguous chance constrained solution is feasible for all values of $\beta \geq 0.01$ and the additional cost for the robustness offered by the ambiguous chance constrained solution is at most 1.6%. Note that we have no guarantee that the Cauchy distribution belongs to the uncertainty set $\mathcal{Q}$ for any value of β.

Table 7.7 displays the performances of the three solutions when the test distribution is

$\mathbf{x}_{sc}$		
err	stdev	c
0.0681	0.0081	11.1930

$\mathbf{x}_{det}$		
err	stdev	c
0.8722	0.0109	11.0000

	$\mathbf{x}_{asc}(\beta)$		
β	err	stdev	c
0.0475	0.0286	0.0044	11.3710
0.0400	0.0378	0.0058	11.3384
0.0300	0.0414	0.0067	11.2827
0.0250	0.0395	0.0057	11.2688
0.0200	0.0495	0.0063	11.2482
0.0150	0.0489	0.0069	11.2445
0.0100	0.0481	0.0066	11.2238
0.0050	**0.0631**	0.0064	11.2115
0.0025	**0.0712**	0.0074	11.1998

Table 7.6: Performance when test distribution $\mathbb{Q}$ is multivariate Cauchy with median $\bar{\mathbf{d}}$.

Pareto with variance $\sigma^2\mathbf{I}$ and a new mean $\widehat{\mathbf{d}}$ such that CDF of $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$ is greater than the CDF of the Pareto for all $\mathbf{d}$, i.e. each component of the chosen Pareto distribution stochastically dominates $\mathcal{N}(\bar{\mathbf{d}}_i, \sigma^2)$ for $i = 1, 2, 3$. The performance of the three solutions is very similar to the case when the test distribution was Cauchy – the ambiguous chance constrained solution provides robustness to perturbation in the distribution at a modest increase in cost.

$\mathbf{x}_{sc}$		
err	stdev	c
0.0554	0.0076	11.1846

$\mathbf{x}_{det}$		
err	stdev	c
0.9809	0.0045	11.0000

	$\mathbf{x}_{asc}(\beta)$		
β	err	stdev	c
0.0475	0.0013	0.0011	11.3842
0.0450	0.0016	0.0014	11.3626
0.0400	0.0036	0.0016	11.3297
0.0300	0.0080	0.0030	11.3004
0.0250	0.0082	0.0028	11.2753
0.0200	0.0108	0.0032	11.2443
0.0150	0.0241	0.0053	11.2402
0.0100	0.0255	0.0046	11.2230
0.0050	**0.0513**	0.0067	11.2221
0.0025	**0.0513**	0.0070	11.1911
0.0005	**0.0666**	0.0073	11.2259

Table 7.7: Performance when $\mathbb{Q}$ is multivariate Pareto with a shifted mean and variance $\sigma^2\mathbf{I}$.

Chapter 8

Conclusion and Future Work

8.1 Contributions

In this dissertation we study single and two stage ambiguous chance constrained programs where the uncertainty set for the ambiguous measure is given by a ball defined in terms of the Prohorov metric. In this formulation, the random parameter $\mathbf{H}$ is known to have a fixed distribution $\mathbb{Q}$. However, the decision maker is only able to estimate that $\mathbb{Q}$ belongs to an uncertainty set $\mathcal{Q}$ of the form $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$, where ρ_p denotes the Prohorov metric and β is an exogenously defined constant.

In the case of single stage optimization problems we approximate the ambiguous chance constrained program by a robust sampled problem where each constraint is a robust constraint centered at a sample drawn according to the "central" distribution in the uncertainty set of measures. On of the main contributions of this dissertation is to show that all the sample complexity results known in the context of usual single stage chance constrained problems extend to the ambiguous setting. Our extensions are based on the Strassen-Dudley Representation Theorem, which states that when the *distributions* of two random variables are close in the Prohorov metric one can construct a *coupling* of the random variables such that the *samples* are close with a high probability. Coupling is just a construct needed to prove the results; it is never used in computing the solution to the robust sampled problem.

We also consider the two-stage convex chance constrained problems that was formulated by Nemirovski and Shapiro [61]. In [61], the authors propose an ellipsoid-like iterative solution algorithm for the special case where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ (see (3.8)) is bi-affine.

We show that the Algorithm SolveChance (see Figure 3.1) extends the results in [61] to bi-convex $\mathbf{f}(\mathbf{x}, \mathbf{h})$ in a fairly straightforward fashion. The computational complexity of SolveChance as well as the quality of its output $\widehat{\mathbf{x}}$ depend on the radius r of the largest Euclidean ball that can be inscribed in the random set $\mathcal{X}_{\mathcal{I}}$ (see (3.12)) that is defined by the random set of linear inequalities generated during one run of SolveChance. Since the set $\mathcal{X}_{\mathcal{I}}$ is random, selecting r is difficult; yet SolveChance requires r as an input. In this dissertation we provide some guidance for selecting r. We show that the largest value of r is related to the degree of robustness of the two-stage chance constrained problem – the more robust the problem, the higher one can set the parameter r. This is reminiscent of results relating the condition number of optimization problems to their computational complexity [65, 66, 38].

Next, we formulate ambiguous two-stage chance constrained problems. We construct an algorithm SolveAmbChance that solves the ambiguous two-stage chance constrained problem under some technical assumptions.

8.2 Future work

In Chapter 6 we assumed that the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine and the extreme points of the dual polytope corresponding to the set $\mathcal{C}$ are explicitly known. The bi-affine assumption can be defended on the grounds that it still allows one to model a wide variety of applications; however, the latter assumption is much more serious and cannot be relaxed in general. We are currently exploring the possibility of replacing the adjustable robust characterization (5.4) by a chance constrained characterization. Let $\mathcal{L} = \{\boldsymbol{\lambda} : \mathbf{W}^T\boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}$ and let $\mathbb{P}$ denote any probability measure on $\mathcal{L}$. Let $g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) = (\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T\boldsymbol{\lambda} + \beta\|\mathbf{A}_1^T(\mathbf{x})\boldsymbol{\lambda}\|$. Then $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if, $g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0$, for all $\boldsymbol{\lambda} \in \mathcal{L}$, i.e. $\mathbb{P}(\boldsymbol{\lambda} : g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0) = 1$. Currently, we are investigating how one can relax this constraint to $\mathbb{P}(\boldsymbol{\lambda} : g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0) > 1 - \epsilon$.

In Chapter 4 we introduced several probability metrics and show that uncertainty set $\mathcal{Q} = \{\mathbb{Q} : \rho(\mathbb{Q}, \mathbb{Q}_0) \leq \gamma\}$ can be conservatively approximated by an uncertainty set $\tilde{\mathcal{Q}} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta(\gamma)\}$ defined in terms of the Prohorov metric. However, we have no way of measuring the "blow-up" of the uncertainty set that occurs in changing the metrics. This issue can be resolved by either establishing tight lower bounds on the Prohorov metric or by constructing Representation results for the other metrics.

Also, we only considered uncertainty sets that are norm balls defined in terms of the Prohorov metric. One could consider "tiling" a more general uncertainty set by norm balls of a given radius and construct a robust sampled problem by drawing samples according to the centers of the balls (a simplified version of this idea appears in [15]). Since the constant ϵ that controls the violation probability in the ambiguous chance constrained problem has to be greater than the radius β of the norm ball, such an approach is attractive even when the uncertainty is a norm ball. However, it is not clear how to efficiently select the centers of the balls to "tile" the uncertainty set.

Finally, there is the issue of proving worst-case lower bounds on the number of samples required to learn the solution of an ambiguous chance constrained problem, i.e. a refinement of Lemma 3 that accounts for ambiguity in the measure.

Bibliography

[1] M. Anthony and N. Biggs. *Computational Learning Theory*. Cambridge University Press, 1992.

[2] P. Artzner, F. Delbaen, J-P. Eber, and D. Heath. Coherent risk measures. *Math. Finance*, 9(203-228), 1999.

[3] A. Atamtürk and M. Zhang. Two-stage robust network flow and design for demand uncertainty. Manuscript, 2005.

[4] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Math. Program.*, 99(2, Ser. A):351–376, 2004.

[5] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. *SIAM J. Optim.*, 7(4):991–1016, 1997.

[6] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[7] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999.

[8] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, Philadelphia, PA, 2001.

[9] D. Bertsimas and M. Sim. Robust conic optimization. Under review in *Math. Prog.*, May 2004.

[10] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

[11] J. R. Birge and J. R. B. Wets. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Math. Prog. Study*, 27(54-102), 1986.

[12] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, Oct 1989.

[13] J. Bukszár. Upper bounds for the probability of a union by multitrees. *Adv. in Appl. Probab.*, 33:437–452, 2001.

[14] G. Calafiore and M. C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. To appear in *Math. Prog.*, 2003.

[15] G. Calafiore and M. C. Campi. Decision making in an uncertain environment: the scenario-based optimization approach. Working paper, 2004.

[16] A. Charnes and W. W. Cooper. Uncertain convex programs: randomized solutions and confidence levels. *Mgmt. Sc.*, 6:73–79, 1959.

[17] Z. Chen and L.G. Epstein. Ambiguity, risk and asset returns in continuous time. Mimeo, 2000.

[18] G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, USA, 1963.

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

[20] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. To appear in *Math. Oper. Res.*, 2001.

[21] M. Dempster and G. Consigli. *World Wide Asset and Liability Modeling*, chapter The CALM stochastic programming model for dynamic asset-liability management, pages 464–500. Cambridge University Press, 1998.

[22] M. Dempster, N.H. Pedron, E Modeva, J. Scott, and A. Sembos. Planning logistic operations in the oil industry. *Journal of Operational Research Society*, 51(11):1271–1288, 2000.

[23] Darinka Dentcheva, András Prékopa, and Andrzej Ruszczyński. Concavity and efficient points of discrete distributions in probabilistic programming. *Math. Program.*, 89(1, Ser. A):55–77, 2000.

[24] C.L. Dert. *Asset Liability Management for Pension Funds, A Multistage Chance Constrained Programming Approach*. PhD thesis, Erasmus University, Rotterdam, The Netherlands, 1995.

[25] R. M. Dudley. Distance of probability measures and random variables. *Ann. Math. Stat.*, 39:1563–1572, 1968.

[26] J. Dupačová. The minimax approach to stochastic program and illustrative application. *Stochastics*, 20:73–88, 1987.

[27] J. Dupačová. Stochastic programming: minimax approach. In *Encyclopedia of Optimization*. Kluwer, 2001.

[28] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, 18(4):1035–1064, 1997.

[29] L. El Ghaoui and A. Nilim. The curse of uncertainty in dynamic programming and how to fix it. To appear in *Oper. Res.*. UC Berkeley Tech Report UCB-ERL-M02/31, November 2002.

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

[31] G. Eppen, R. Martin, and L. Schrage. A scenario approach to capacity planning. *Operational Research*, 37(4):517–525, 1989.

[32] L. G. Epstein and M. Schneider. Recursive multiple priors. Technical Report 485, Rochester Center for Economic Research, August 2001. Available at http://rcer.econ.rochester.edu. To appear in *J. Econ. Theory*.

[33] L. G. Epstein and M. Schneider. Learning under Ambiguity. Technical Report 497, Rochester Center for Economic Research, October 2002. Available at http://rcer.econ.rochester.edu.

[34] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. To appear in *Math. Prog.*, 2004.

[35] E. Erdoğan and G. Iyengar. Boosting, importance sampling, and convex chance constrained problem. Manuscript, 2005.

[36] E. Erdoğan and G. Iyengar. On two-stage convex chance constrained problems. To appear in *Mathematical Methods of Operations Research*, 2006.

[37] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Fin. and Stoch.*, 6:429–447, 2002.

[38] R. M. Freund and J. R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J. Opt.*, 10:155–176, 1999.

[39] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *Intl. Stat. Rev.*, 7(3):419–435, 2002.

[40] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *J. Math. Econ.*, 18(2):141–153, 1989.

[41] D. Goldfarb and Iyengar G. Robust convex quadratically constrained programs. *Mathematical Programming Ser B.*, 97(3):495–515, 2003.

[42] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Math. Oper. Res.*, 28(1):1–38, 2003.

[43] F. R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The approach based on influence functions.* John Wiley & Sons, 1986.

[44] L. P. Hansen and T. J. Sargent. Robust control and model uncertainty. *American Economic Review*, 91:60–66, 2001.

[45] R. Henrion. Structural properties of linear probabilistic constraints. *Stochastic programming E-print series (SPEPS)*, 13, 2005.

[46] G. Iyengar. Robust dynamic programming. To appear in *Math. Oper. Res.*. Available at http://www.corc.ieor.columbia.edu/reports/techreports/tr-2002-07.pdf, 2002.

[47] R. Jagannathan. Minimax procedure for a class of linear programs under uncertainty. *Oper. Res.*, 25:173–177, 1977.

[48] Yuri Kan. Application of the quantile optimization to bond portfolio selection. In *Stochastic optimization techniques (Neubiberg/Munich, 2000)*, volume 513 of *Lecture Notes in Econom. and Math. Systems*, pages 285–308. Springer, Berlin, 2002.

[49] M. J. Kearns and U. V. Vazirani. *An introduction to computational learning theory.* MIT Press, Cambridge, MA, 1997.

[50] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244(S):1093–1096, 1979. Translated in Soviet Mathematics Doklady 20:1 (1979).

[51] C. M. Lagoa, X. Li, and M. Sznaier. Probabilistically constrained linear programs and risk-adjusted controller design. *SIAM Journal of Optimization*, 15:938–951, 2005.

[52] M. Linden. A model for stock return distributions. *International Journal of Finance and Economics*, 6:159–169, 2001.

[53] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra Appl.*, 284(1-3):193–228, 1998.

[54] P. M. Long. The complexity of learning according to two models of a drifting environment. *Machine Learning*, 37(3):337–354, 1999.

[55] H. M. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952.

[56] H. M. Markowitz. *Portfolio Selection.* Wiley, New York, 1959.

[57] R. O. Michaud. *Efficient Asset Management: a practical guide to stock portfolio optimization and asset allocation.* Harward Business School Press, Boston, 1998.

[58] F.H. Murphy, S. Sen, and A.L. Soyster. Electric utility capacity expansion planning with uncertain load forecasts. *AIIE Transaction 14:52-59*, 1982.

[59] A. Nemirovski. On tractable approximations of randomly perturbed convex constraints. In *Proc. 42nd IEEE Conf. Dec. Contr. (CDC)*, volume 3, pages 2419– 2422, 2003.

[60] A. Nemirovski and A. Shapiro. On complexity of stochastic programming problems. Published electronically in: *Optimization Online.*

[61] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. To appear in *Probabilistic and randomized methods for design under uncertainty*, 2004.

[62] C.A. Poojari and B. Varghese. Genetic algorithms based technique for solving chance constrained problems. Technical report, CARISMA, Brunel University, 2006.

[63] A. Prekopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

[64] S. T. Rachev. *Probability metrics and the stability of stochastic models*. John Wiley & Sons, 1991.

[65] J. Renegar. Some perturbation theory for linear programming. *Math. Prog.*, 65:73–91, 1994.

[66] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Prog.*, 70:279–351, 1995.

[67] R. T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, New Jersey, 1997.

[68] A. Ruszczynski and A. Shapiro, editors. *Stochastic Programming*. Handbook in Operations Research and Management Science. Elsevier, 2003.

[69] A. Ruszczynski and A. Shapiro. Optimization of risk measures. Available at http://www.optimization-online.org/DB_HTML/2004/02/822.html, 2004.

[70] A. Ruszczynski and A. Shapiro. Optimization of risk measures. Available at http://ideas.repec.org/p/wpa/wuwpri/0407002.html, 2004.

[71] S. Sen, R.D. Doverspike, and S. Cosares. Network planning with random demand. *Telecommunication Systems 3:11-30*, 1994.

[72] A. Shapiro. Some recent developments in stochastic programming. *ORB Newsletter*, 13, March 2004. Available at http://www.ballarat.edu.au/ard/itms/CIAO/ORBNewsletter/issue13.shtml#11.

[73] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. To appear in *SIAM J. Opt.*, 2004.

[74] A. Shapiro and A. J. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17:523–542, 2002.

[75] N. Z. Shor. Cut-off method with space extension in convex programming problems. *Cybernetics*, 13:94–96, 1977.

[76] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.*, 36:423–439, 1965.

[77] H. Thorisson. *Coupling, Stationary, and Regeneration.* Probability and its Applications. Springer-Verlag, 2000.

[78] A. Tomasgard, J.A. Audestad, S. Dye, L. Stougie, M.H. van der Vlerk, and S.W. Wallace. Modelling aspects of distributed processing in telecommunication networks. *Annals of Operations Research*, 82:161–184, 1998.

[79] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, NY, 1995.

[80] J. Žáčková. On minimax solutions of stochastic linear programs. *Čas. Pěst. Mat.*, pages 423–430, 1966.

Appendix A

Ambiguity with Total Variation Norm

In this section, we consider uncertainty sets defined in terms of the Total Variation metric and argue that the condition $\beta \leq \epsilon$ is necessary for the results in Chapter 5 to hold. The analysis presented in this section was suggested by Prof. Alexander Shapiro.

We will assume that the sets $\mathcal{X}$ and $\mathcal{H}$ are bounded. Fix $\mathbf{x} \in \mathcal{X}$ and let $\mathcal{B}_x = \{\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0\}$ denote the "bad" set of parameter values. Suppose the uncertainty set $\mathcal{Q} = \{\mathbb{Q} : \rho_{tv}(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$, where ρ_{tv} denotes the total variation metric defined in (4.3). Then, it follows that

$$\text{err}(\mathbf{x}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(\mathcal{B}_x) \leq \min\{\mathbb{Q}_0(\mathcal{B}_x) + \beta, 1\}, \tag{A.1}$$

where the first term in the bound follows from the definition of the Total variation metric. We show below that the bound is achieved.

Define the measure $\bar{\mathbb{Q}}$ as follows

$$\bar{\mathbb{Q}}(\mathcal{B}) = \begin{cases} \beta \mathbb{U}_{\mathcal{B}_x}(\mathcal{B} \cap \mathcal{B}_x) + (1-\beta)\mathbb{Q}_0(\mathcal{B} \cap \mathcal{B}_x^c), & \mathbb{Q}_0(\mathcal{B}_x) = 0, \\ \alpha_1 \mathbb{Q}_0(\mathcal{B} \cap \mathcal{B}_x) + \alpha_2 \mathbb{Q}_0(\mathcal{B} \cap \mathcal{B}_x^c), & \min\{\mathbb{Q}_0(\mathcal{B}_x), \mathbb{Q}_0(\mathcal{B}_x^c)\} > 0, \\ \mathbb{Q}_0(\mathcal{B}), & \mathbb{Q}_0(\mathcal{B}_x) = 1, \end{cases}$$

where $\mathbb{U}_{\mathcal{B}_x}$ denotes the uniform measure on the set $\mathcal{B}_x$, and

$$\alpha_1 = \frac{\min\{\mathbb{Q}_0(\mathcal{B}_x) + \beta, 1\}}{\mathbb{Q}_0(\mathcal{B}_x)}, \qquad \alpha_2 = \frac{1 - \min\{\mathbb{Q}_0(\mathcal{B}_x) + \beta, 1\}}{\mathbb{Q}_0(\mathcal{B}_x^c)}.$$

Then it is easy to verify that $\bar{\mathbb{Q}}$ is a probability measure that achieves the upper bound in (A.1) and satisfies $\rho_{tv}(\bar{\mathbb{Q}}, \mathbb{Q}_0) \leq \beta$. Thus, the inequality in (A.1) is, in fact, an equality. The bound (A.1) leads to the following conclusions.

(a) The ambiguous chance constraint $\sup\{\mathbb{Q}(\mathcal{B}_x) : \rho_{tv}(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\} \leq \epsilon$ is feasible only if $\epsilon \geq \beta + \mathbb{Q}_0(\mathcal{B}_x) \geq \beta$. Since the Prohorov metric $\rho_p \leq \rho_{tv}$, it follows that $\epsilon > \beta$ is *necessary* for the feasibility of the robust chance constraint $\sup\{\mathbb{Q}(\mathcal{B}_x) : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\} \leq \epsilon$.

(b) For $\epsilon \geq \beta$, the ambiguous chance constraint $\sup\{\mathbb{Q}(\mathcal{B}_x) : \rho_{tv}(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\} \leq \epsilon$ is equivalent to the (nominal) chance constraint $\mathbb{Q}_0(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0) \leq \epsilon - \beta$, i.e. a chance constraint with tighter requirement. One interpretation of this result is that by introducing ambiguity via ρ_{tv} one does *not* protect against a set of measures; one only increases the probabilistic guarantee with respect to the central measure $\mathbb{Q}_0$.

We have not been able to establish a similar result for the Prohorov metric. Since $\rho_p \leq \rho_{tv}$, a neighborhood defined in terms of the Prohorov metric is more conservative and is likely to protect against measures other than the central measure. At the very least the bound (A.1) alerts us that the power of the ambiguous chance constraint can be quite sensitive to the metric used to define the neighborhood of measures.