The Dissertation Committee for Tara Rengarajan
certifies that this is the approved version of the following dissertation:

Probabilistic Bicriteria Models: Sampling Methodologies and Solution Strategies

Committee:

David P. Morton, Supervisor

John J. Hasenbein

Erhan Kutanoglu

Kumar Muthuraman

Elmira Popova

Probabilistic Bicriteria Models: Sampling Methodologies and Solution Strategies

by

Tara Rengarajan, M.Sc., M.A., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2010

To my mother.

Acknowledgments

The four and half years I have spent in the Operations Research program at the University of Texas have been a tremendous experience. I am deeply grateful to my adviser David Morton for his encouragement and patience that never fail to stimulate. I have also benefited enormously from numerous interactions with other faculty, as well as friends at UT OR/IE, particularly Ali Koç, Mike Nehme and Yufen Shao. I owe many thanks to them. Finally, whatever I have brought to fruition has been largely due to the boundless love and support of my parents, brother and husband. Thank you all.

Probabilistic Bicriteria Models: Sampling Methodologies and Solution Strategies

Publication No. _____

Tara Rengarajan, Ph.D.
The University of Texas at Austin, 2010

Supervisor: David P. Morton

Many complex systems involve simultaneous optimization of two or more criteria, with uncertainty of system parameters being a key driver in decision making. In this thesis, we consider probabilistic bicriteria models in which we seek to operate a system reliably, keeping operating costs low at the same time. High reliability translates into low risk of uncertain events that can adversely impact the system. In bicriteria decision making, a good solution must, at the very least, have the property that the criteria cannot both be improved relative to it. The problem of identifying a broad spectrum of such solutions can be highly involved with no analytical or robust numerical techniques readily available, particularly when the system involves nontrivial stochastics. This thesis serves as a step in the direction of addressing this issue. We show how to construct approximate solutions using Monte Carlo sampling, that are sufficiently close to optimal, easily calculable and subject to a low margin of error. Our approximations can be used in bicriteria decision

making across several domains that involve significant risk such as finance, logistics and revenue management.

As a first approach, we place a premium on a low risk threshold, and examine the effects of a sampling technique that guarantees a prespecified upper bound on risk. Our model incorporates a novel construct in the form of an uncertain disrupting event whose time and magnitude of occurrence are both random. We show that stratifying the sample observations in an optimal way can yield savings of a high order. We also demonstrate the existence of generalized stratification techniques which enjoy this property, and which can be used without full distributional knowledge of the parameters that govern the time of disruption. Our work thus provides a computationally tractable approach for solving a wide range of bicriteria models via sampling with a probabilistic guarantee on risk. Improved proximity to the efficient frontier is illustrated in the context of a perishable inventory problem.

In contrast to this approach, we next aim to solve a bicriteria facility sizing model, in which risk is the probability the system fails to jointly satisfy a vector-valued random demand. Here, instead of seeking a probabilistic guarantee on risk, we instead seek to approximate well the efficient frontier for a range of risk levels of interest. Replacing the risk measure with an empirical measure induced by a random sample, we proceed to solve a family of parametric chance-constrained and cost-constrained models. These two sampling-based approximations differ substantially in terms of what is known regarding their asymptotic behavior, their computational tractability,

and even their feasibility as compared to the underlying "true" family of models. We establish however, that in the bicriteria setting we have the freedom to employ either the chance-constrained or cost-constrained family of models, improving our ability to characterize the quality of the efficient frontiers arising from these sampling-based approximations, and improving our ability to solve the approximating model itself. Our computational results reinforce the need for such flexibility, and enable us to understand the behavior of confidence bounds for the efficient frontier.

As a final step, we further study the efficient frontier in the cost versus risk tradeoff for the facility sizing model in the special case in which the (cumulative) distribution function of the underlying demand vector is concave in a region defined by a highly-reliable system. In this case, the "true" efficient frontier is convex. We show that the convex hull of the efficient frontier of a sampling-based approximation: (i) can be computed in strongly polynomial time by relying on a reformulation as a max-flow problem via the well-studied selection problem; and, (ii) converges uniformly to the true efficient frontier, when the latter is convex. We conclude with numerical studies that demonstrate the aforementioned properties.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Bicriteria Optimization

Most decision making challenges today involve the need to simultaneously minimize multiple objectives. Even with just two objectives, it is very rarely the case that both objectives can reach their respective minima simultaneously; hence, a tradeoff is required. Bicriteria minimization is the study of optimal tradeoffs in a dual objective decision making environment. Its relevance in decision analysis cannot be overstated; indeed, it forms the basic firmament of optimizing asset allocation in the financial services and investment industry.

In this thesis, we deal with bicriteria minimization problems in which one of the two objectives is *risk*, as defined by the probability of an undesirable event, and the other is the cost of system design. The goal of our work is to devise efficient methods to solve bicriteria risk models, and assess their merits in a theoretical and computational context. Our problems are motivated by common application domains in the operations research literature, such as inventory control and facility sizing.

In a bicriteria setting, the standard notion of optimality needs to be replaced by a concept that is aligned with the existence of more than one ob-

jective function. A natural extension from the single objective case is *Pareto* optimality, which states that a candidate solution is optimal if neither of the objectives can be improved without degrading the other. The set of cost-risk pairs determined by the set of Pareto optimal solutions is called the *efficient frontier* (EF). Since there may be multiple Pareto optimal solutions to a bi-criteria model, and since there is usually no ordering among Pareto optimal solutions, the decision maker must turn to other considerations to select the final solution from the set of Pareto optimal solutions. From an optimization perspective, we aim to generate the entire set of Pareto optimal solutions within a practical range of risk levels, or at least, a minimal subset of it that determines the EF in this range. Determining the EF is, in general, a hard problem, and more so, with a risk term that typically involves multivariate probability distributions that may be analytically intractable. In this thesis, we focus on sampling methods in order to construct approximations to the EF. Our efforts are aimed at using Monte Carlo sampling to form approximations that are computationally inexpensive, and enjoy good convergence properties to the true EF.

Bicriteria models are a special case of multiple criteria optimization models, which have been widely studied. Tracing back the origins of the subject leads us to economics, particularly, welfare theory and utility theory, from which multiple criteria optimization is believed to have originated. At least as far back as the nineteenth century, economists have considered the notion of deriving utility value from their decisions, and looked at the notion

of equilibrium in a multi-utility setting. While Pareto characterized a welfare equilibrium in the early twentieth century, it was not until some fifty years later that a formal definition of Pareto optimality was laid down by [9]. The term *multiple criteria decision making* came into being around this time, when the underlying theory was being established. A number of articles and books on multiple criteria optimization have appeared since, with application to a broad range of areas, prominently finance. An extensive collection of works by Geoffrion (see for example, [17, 18]) and more recently, Steuer (see [6, 50]) serves as excellent technical reference, while survey articles such as [15, 16] trace the development of the subject in detail.

Multiple criteria optimization is also referred to as vector optimization. The terminology stems from the understanding that there is a vector of objectives and we seek to optimize a suitably defined function of this vector. For the rest of this work, we assume that the objectives must be minimized. A vector minimization problem with two objectives can be expressed generally as

$$\begin{aligned} \underset{x}{\text{vmin}} \quad & [f_1(x), f_2(x)] \\ \text{s.t.} \quad & x \in D. \end{aligned} \tag{1.1}$$

Here, the real-valued functions $f_1(\cdot)$ and $f_2(\cdot)$ denote the objectives to be minimized over decision vectors x in the set D defined by systemic constraints. A solution $x \in D$ is Pareto optimal if there does not exist $y \in D$ such that $f_i(y) \leq f_i(x)$, $i = 1, 2$, and at least one of these inequalities is strict. The set $\{(f_1(x), f_2(x)) \mid x \text{ is Pareto optimal}\}$ is the associated efficient frontier.

The definition of Pareto optimality may be strengthened or weakened as follows: $x \in D$ is *strictly* Pareto optimal if there does not exist $y \in D$, $y \neq x$ such that $f_i(y) \leq f_i(x)$, $i = 1, 2$; $x \in D$ is *weakly* Pareto optimal if there does not exist $y \in D$ such that $f_i(y) < f_i(x)$, $i = 1, 2$. While a Pareto optimal solution is the most desirable from a realistic perspective, these variants are often useful since most models used to solve multiple criteria optimization problems are guaranteed to produce optimal solutions that fall into one of these three categories under general conditions. For the purposes of our work in the following chapters however, we are concerned solely with Pareto optimal solutions.

While the decision vector x presents the decision maker with a plan for system design, the value derived from it is captured by $f_i(x)$, $i = 1, 2$. It is useful to consider the image space of values of the vector valued map $f(\cdot)$ given by $f(x) = (f_1(x), f_2(x))$ for better understanding of the trade-off between the objectives. The pair $(f_1(x), f_2(x))$ is on the EF when there are no solutions $y \in D$ that yield $f(y) = (f_1(y), f_2(y))$ that lies in the negative orthant, relative to $f(x)$. Thus, the EF forms a boundary for the image space and represents the set of all optimal trade-offs. The EF is exactly that part of the lower boundary which is strictly decreasing; Figure 1.1 below illustrates the EF for a bicriteria model involving risk and cost of system design.

In order to obtain Pareto optimal solutions, the objectives to be minimized may be combined to obtain a real-valued function, the *value function*, that can be used in a single objective minimization model. In other words,

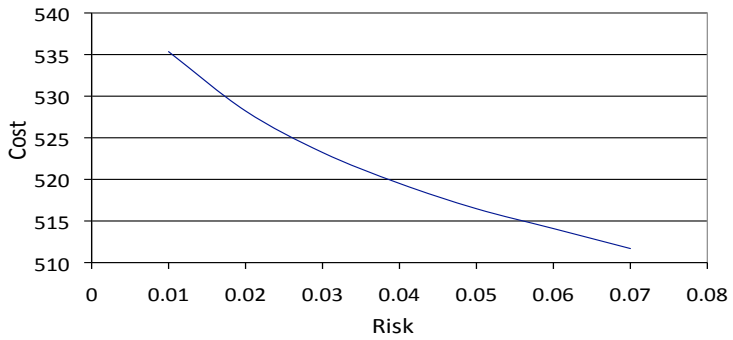

Figure 1.1: Efficient frontier for a model trading off cost and risk

a value function $v(\cdot)$ is a real-valued function over a superset of the image space with the property that if $f(x)$ and $f(y)$ lie in the image space and $f_i(x) \leq f_i(y)$, $i = 1, 2$, then $v(f(x)) \leq v(f(y))$. Solvability of the associated single objective minimization program is an important factor in determining its applicability. In the following section, we present an overview of some common approaches to forming and minimizing value functions.

1.2 Single Objective Model Formulations

There exist two primary approaches to solving multicriteria optimization problems in general — non-interactive and interactive. In the non-interactive approach, the value function and other parameters of the problem are specified beforehand, and the problem solved to optimality. An interactive approach receives information progressively, and alternates between solves and input from the decision maker. We refer the interested reader to [15] for details on interactive methods and focus our study on the non-interactive approach,

based largely on the material in [14]. For a more comprehensive study, see the books [39, 54, 55].

The choice of value function is often dictated by the preferences of the decision maker. The value functions studied in this section are driven by factors such as the existence of a scalarizing utility function, ranking among the objectives, worst-case minimization and performance relative to an "ideal" solution. Furthermore, as noted earlier, tractability and optimality properties also matter greatly. In what follows, we briefly motivate the form that each value function takes, and furnish the main results relating optimal solutions to the model and Pareto optimality to (1.1). We frame our presentation in the context of two objectives; however, all methods, along with their properties in relation to Pareto optimality, extend to a more general multiple criteria setting.

a) *Weighted Sum*

Consider a scalarized problem of the form

$$\begin{aligned} \min_{x} \quad & \lambda_1 f_1(x) + \lambda_2 f_2(x) \\ \text{s.t.} \quad & x \in D, \end{aligned} \tag{1.2}$$

where $\lambda_i \geq 0$, $i = 1, 2$. Without loss of generality, we may assume that $\lambda_1 + \lambda_2 = 1$ so that program (1.2) minimizes a convex combination of the objectives. Minimizing a convex combination suggests a balanced approach towards trade-off, with the weights λ_i determining the importance of the associated objective.

It is easily established that any optimal solution to (1.2) is weakly Pareto optimal. This can be strengthened to Pareto optimality when all weights λ_i are strictly positive. To see if the converse is true, consider a problem in which the image space is convex (see Figure 1.2). Suppose x is Pareto optimal so that the point $\mathrm{P}_x = (f_1(x), f_2(x))$ lies on the EF. Translate the origin to P_x and let $\mathbb{R}^2_-$ represent the negative orthant under this translation. Pareto optimality of x implies that the translated image space has empty intersection with $\mathbb{R}^2_-$. A fundamental result from convex analysis states that when the relative interiors of two convex sets have nonempty intersection, they can be separated by a hyperplane. These facts can be used to show that x minimizes a linear function of the objectives with non-negative weights, i.e., an objective function of the form in (1.2) (see [14] for a proof).

Figure 1.3 suggests that the requirement of convexity of the entire image space is too restrictive. Indeed, convexity "around" the EF seems to suffice. This weaker notion of convexity holds, for example, when the objective functions and the set D are convex. Under such an assumption, the set of all optimal solutions to the parametric family (1.2) with non-negative weights λ_i, $i = 1, 2$, is a superset of the set of Pareto optimal solutions.

b) *ϵ–Constraint Method*

As the name suggests, the ϵ-constraint method enforces a hard constraint on all but one of the objectives, parametrized by the vector $\epsilon \in \mathbb{R}^2_+$. We

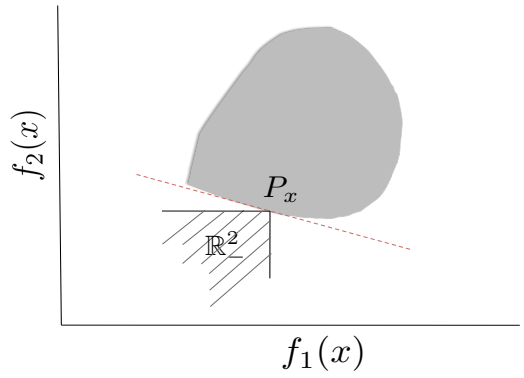

Figure 1.2: A convex image space for a bicriteria problem. The red dotted line is a separating hyperplane between the EF and the translated negative orthant.

solve the family of programs

$$\begin{aligned} \min_x \quad & f_k(x) \\ \text{s.t.} \quad & f_i(x) \leq \epsilon_i,\ i = 1, 2,\ i \neq k, \\ & x \in D, \end{aligned} \tag{1.3}$$

for $k = 1, 2$. As in the weighted sums method, an optimal solution to (1.3) is weakly Pareto optimal. To derive the analog of the positive weights condition for the method of weighted sum, observe that positivity of the weights ensures that the following condition holds: at any optimal solution, each objective is minimized when the other is fixed. This, in turn, implies Pareto optimality. In (1.3), this property holds only for the objective being minimized. Thus, a sufficient condition for Pareto

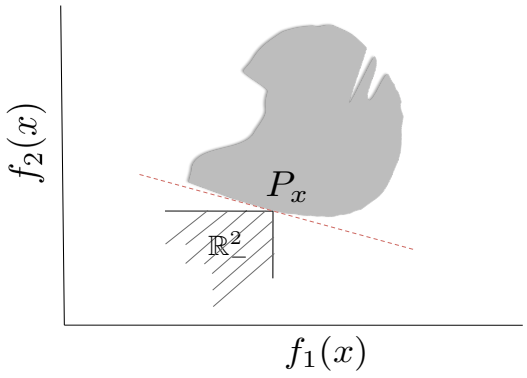

Figure 1.3: An image space for a bicriteria problem. Convexity around the EF indicates existence of a separating hyperplane between the EF and the translated negative orthant.

optimality of a solution x is that it is optimal to (1.3) for some vector ϵ and $k = 1, 2$. It can also be seen that a Pareto optimal solution x is optimal to (1.3) with ϵ_i replaced by $f_i(x)$.

Observe that if the inequality constraints in (1.3) are tight at every optimal solution for every non-negative vector ϵ and for every k, then every optimal solution to (1.3) is Pareto optimal. Hence, the entire EF can be generated by solving the above parametric family of programs, a fact that we use in our work.

c) *Lexicographic Ordering*

Suppose there exists an order among the objectives that is given by $f_1 \succeq$

f_2, where $f_i \succeq f_j$ means that f_i is ranked higher than f_j. This induces a total ordering in the image space given by the relation: $(f_1(x), f_2(x)) \leq (f_1(y), f_2(y))$ if $f_k(x) \leq f_k(y)$ where $k = \min\{i \mid f_i(x) \neq f_i(y)\}$. The lexicographic ordering method finds a solution x that minimizes the image vector $f(x) = (f_1(x), , f_2(x))$. This means that the single objective $f_1(x)$ is first minimized, and if there exists more than one optimal solution, $f_2(x)$ is minimized over the set of optimal solutions. We thus solve two minimization problems of the form

$$\begin{aligned} z_k^* = \min_x \quad & f_k(x) \\ \text{s.t.} \quad & f_i(x) = z_i^*, \; i = 1, \ldots, k-1, \\ & x \in D, \end{aligned} \tag{1.4}$$

for $k = 1, 2$ with the equality constraints being vacuous for $k = 1$.

Note that the method simply minimizes $f_1(\cdot)$, with $f_2(\cdot)$ acting as a tie-breaking rule. It follows immediately that any optimal solution obtained at termination of this algorithm is Pareto optimal; indeed, at an optimal solution, each objective attains its minimum value with the other objective held fixed.

d) *Worst-Case Minimization*

If the goal of the decision maker is to safeguard against the worst performing objective, the adversarial model below is useful:

$$\begin{aligned} \min_x \quad & \max_{i \in \{1,2\}} \; f_i(x) \\ \text{s.t.} \quad & x \in D. \end{aligned} \tag{1.5}$$

This is essentially a robust optimization model and can be difficult to solve. Moreover, it suffers from the disadvantage that it penalizes high values per se without taking into account intrinsic differences between the objectives such as units of measurement or the lowest possible value. For instance, a better single objective for (1.5) is given by the function $\max_{1\leq i\leq m} (f_i(x) - z_i^*)$, where $z_i^* = \min_{x\in D} f_i(x)$. This variant, of course, requires two single objective minimizations prior to solving the adversarial problem.

An optimal solution to (1.5) or the above variant is weakly Pareto optimal. Note that the objective function for both models is simply the $\|\cdot\|_\infty$ norm of the distance between the image vector and a "reference vector", such as the origin or (z_1^*, z_2^*). Other norms such as the $\|\cdot\|_p$ norm, with $1 \leq p < \infty$ may be used as well. When the distance metric is given by $\max_{i\in\{1,2\}} \lambda_i(f_i(x) - z_i^*)$, it is called a *weighted Tchebycheff* scalarizing function, and the metric is called the weighted Tchebycheff metric. The Tchebycheff scalarizing function is well-studied and used by several authors such as [44, 51]. More on distance minimizing value functions may be found in [14].

Other methods exist, such as lexicographic max ordering, which combines lexicographic ordering with worst-case minimization. Goal programming addresses the setting in which there exist target values b_i for the functions f_i, $i = 1, 2$, and we aim to minimize simultaneously the objectives

$f_i(\cdot) - b_i$, $i = 1, 2$. Obtaining tight bounds on the EF is also a problem of significant research interest; see [14] for details on these and related topics. Finally, when the decision maker's preferences cannot be entirely specified prior to optimization, interactive optimization is used. The Tchebycheff interactive approach serves as a good example and is widely in use, since it has the property that a solution is optimal if and only if it is Pareto optimal [15, 50, 51]. For recent enhancements and applications of the method, see for example, [24] and [25].

1.3 Bicriteria Models with Risk

When the number of criteria to be minimized is two, the problem takes on interpretations that are simpler to understand than general multicriteria problems. Given two objectives $h(\cdot)$ and $p(\cdot)$, the *efficient frontier* denoted by $EF(h, p)$ is the set of all pairs $(p(x), h(x))$ ranging over Pareto optimal solutions x. We thus have a plot of $h(\cdot)$ versus $p(\cdot)$ which is a strictly decreasing function. In portfolio allocation for instance, $h(x)$ may be the negative expected return while $p(x)$ may be the variance of that return, and x is the allocation design.

In this thesis, we are concerned with bicriteria models in which one of the objectives, $p(\cdot)$, is risk, and the other, $h(\cdot)$, is the cost of system design. Models of this nature occur widely in finance, and other areas such as telecommunications. Ruszczyński and Vanderbei [47] construct the EF of a mean-risk model, while Pagnoncelli et al. [38] make a computational study of

the efficient frontier of a sampling-based approximation to an asset-allocation model. We also point to [12, 23, 46, 52] for other recent work on bicriteria optimization with risk. The trade-off in this setting is apparent — a highly reliable design will cost more and lead to lower risk than a less reliable design. A desirable feature of bicriteria models is that the problem of forming the EF can be recast as a parametric family of probabilistically-constrained programs with a single individual chance constraint. Observe that this is simply the ϵ-constraint method, with a upper-bounding constraint on risk. There is a rich literature on models with probabilistic constraints and objectives; among them, [41] and [43].

Note that $EF(h, p)$ may have discontinuities or even be a set of disconnected points. This happens when it is not possible to express $h(\cdot)$ as a continuous function of $p(\cdot)$ (or vice versa) on the EF. However, if $EF(h, p)$ is connected, then $h(\cdot)$ can be expressed as a strictly decreasing continuous function of $p(\cdot)$ on it. In this case, any optimal solution to the parametric family of ϵ-constrained programs

$$\begin{aligned} \min_{x} \quad & h(x) \\ \text{s.t.} \quad & p(x) \leq \epsilon, \\ & x \in D, \end{aligned} \tag{1.6}$$

is Pareto optimal. Combining this with our earlier observations, we see that the set of optimal solutions to the family (1.6) coincides with the set of Pareto optimal solutions to the associated bicriteria problem, at least when the above mentioned continuity holds.

The risk objective $p(x)$ represents the probability of an undesirable event. Specifically, let $\tilde{\xi}$ denote a random vector that forms the uncertainty component of the problem. We assume that the event $G(x, \tilde{\xi}) \notin C$ is an undesirable event, where $G(\cdot, \cdot)$ is a function taking values in a set containing C. As a simple example, let C be the positive orthant, $\mathbb{R}^n_+$, and $G(x, \tilde{\xi}) = x - \tilde{\xi}$. The risk, in this case, is the probability that the design vector x fails to dominate component-wise, the random vector $\tilde{\xi}$, i.e., $p(x) = \mathbb{P}(x \not\geq \tilde{\xi})$.

Our models can be accommodated within the framework of stochastic programming with recourse in the sense that both cost and risk can involve second stage decisions and beyond. Cost can include an expected cost from recourse decisions, and risk can represent the probability that the cost of recourse decisions is small. See, for example, the work of Schultz and Tiedemann [49], who propose and study stochastic programs with recourse that include a risk term in the objective.

1.4 Overview of Contents

The object of this dissertation is to estimate the efficient frontier of probabilistic bicriteria models. Our overall approach is to approximate the risk objective through the use of Monte Carlo sampling. This said, the single objective model used to solve for the EF varies between Chapters 2, 3 and 4. We assess the merits of our approach via numerical experiments on instances of a model typical of our setting. Each chapter details the assumptions in the model, the sampling procedure and formation of the single objective program,

and includes a computational study.

Chapter 2 considers models with a modest number of disruptions in the planning horizon, where we use the notion of a disruption to refer to an undesirable event of so severe a magnitude that it occurs only a modest number of times to the planning horizon. We assume, in contrast to standard stochastic programming models, that while the number of disruptions is known, their times of occurrence are also random. We approximate the bicriteria model with a convex approximation derived from constraint sampling. Our work builds on the results of Calafiore and Campi [7] who establish a bound on the expected probability of violation of the chance-constraint from naive sampling. We derive an improved bound by using stratified sampling, and establish savings in sampling costs up to an order of magnitude in the number of time periods, over naive sampling. For a single disruption model, we establish the surprising result that even in the absence of exact knowledge of the probability distribution of the time of disruption, a polynomial sampling strategy is order optimal for most well-behaved distributions. Using a perishable inventory model, we computationally demonstrate both that stratification gives savings in sampling costs, and that stratification yields solutions closer on average to the efficient frontier of the bicriteria model than solutions derived from naive sampling.

We then consider a setting in which the probability distribution on the time of the disruption is chosen by an adversary, subject to box constraints, whose goal is to make our improved bound as bad as possible. This leads to a

two-person game with a convex-concave objective function. We use min-max theory to establish that the order of play is immaterial, and that the model can be solved by an intuitive water-filling algorithm. Finally, we extend our improved bound to a setting in which there are two disruptions, and show that the order improvement over naive sampling realized in the single-disruption case holds in this setting as well. Our argument is easily seen to carry over to the general case of a modest number of disruptions.

In Chapter 3, we adopt the ϵ–constraint method to estimate the EF. That is, we sample the randomness and form a parametric single objective optimization model in which the distribution of $\tilde{\xi}$ is replaced by its sampling distribution in order to form $p(x)$. We first establish that for the purposes of estimating the EF in a meaningful range of risk or cost values, it is immaterial which of the two objectives appears in the constraint. We assume that at all optimal solutions to the single objective program, the constraint is tight, thus ensuring that all optimal solutions are Pareto optimal for the approximation. We thus form an *approximate EF* and analyze its convergence to the true EF as the sample size grows large. We assess the quality of the approximate EF by forming confidence bounds on the optimality gap for each single objective program, and establish from our experiments with a facility-sizing model, that the confidence intervals have modest width.

In Chapter 4, we focus on fast algorithms to approximate the EF. We consider the setting in which the EF is convex in the desirable range of risk values, and establish that a sequence of convex approximations to the

EF, obtained from sampling, converges uniformly to the EF on any compact subinterval contained in this range of convexity. To illustrate this, we consider the facility-sizing model from Chapter 3, and implement a polynomial time algorithm for forming the convex approximations. We show how this algorithm can be modified for enhanced performance by size reduction. Our results on convergence and savings in computational costs are backed by the outcome of our numerical studies. Finally, we summarize in Chapter 5.

Chapter 2

Convex Approximations of a Probabilistic Bicriteria Model with Disruptions

2.1 Introduction

In this chapter, we consider a probabilistic bicriteria program spanning finitely many time periods, involving random disruptions. In a significant departure from standard approaches, we model both the *time* and *magnitude* of the disruption as random. This is typical of events such as fires, storms and market crashes, which can seriously disrupt a system. The enormity of a disruption suggests that it occurs infrequently, possibly at most once to our finite planning horizon. Under such random-time disruptions, we study problems in which decisions are made in each time period subject to system dynamics and a bicriteria objective involving cost and risk. Another distinguishing feature of our model is that when a disruption occurs in a particular time period, the subsequent good or bad event only involves decisions corresponding to that time period.

The following perishable inventory management problem is a motivating example for models incorporating random disruptions. (See the reviews [32] and [20] for a review of perishable inventory theory.)

Example 1. Consider a perishable inventory management problem over finitely many planning periods. In each time period, a firm manufactures a perishable product. That product may be made available for immediate use, i.e., "placed on the shelf" or may be stored for future use. Unconsumed product on the shelf perishes at the end of each period. A known nominal demand occurs in each time period. The random disruption appears in the form of excess demand, and occurs in at most one time period. Only product placed on the shelf can be used to meet nominal and random demand. Costs incurred by the firm include production costs, storage costs and penalty costs incurred due to unmet nominal demand. The firm hedges against the disruption by ensuring that excess demand is satisfied with high probability. Decisions are made prior to the start of the first period and under the assumption that the probability distribution governing the magnitude and time of the disruption is known. (This assumption is relaxed in Section 2.4.)

We now turn to a review of literature related to the problem we consider.

2.1.1 Related Work

The efficient frontier of a probabilistic bicriteria model can be determined by means of a parametric probabilistically-constrained programs. Much of the computationally-oriented research in probabilistically-constrained programs has focused on tractable special cases and approximate solution methods. We outline below such work most closely related to ours.

Luedtke [26] formulates a probabilistically-constrained program as a mixed-integer program and develops techniques to strengthen the formulation by means of valid inequalities. Nemirovski and Shapiro [36] use constraint sampling from an importance-sampling distribution, to construct a tractable approximation under the assumption that the random constraints are bilinear in the decision variables and the random parameters. In later work, Nemirovski and Shapiro [35] develop a large deviation-type approximation that is convex and efficiently solvable. Luedtke and Ahmed [27] approximate the probability distribution in a probabilistically-constrained program with the empirical distribution obtained from a sample, and obtain lower and upper confidence bounds on the model's optimal value. Calafiore and Campi [7, 8] provide probabilistic feasibility guarantees for probabilistically-constrained programs with convex objective and constraint functions by the use of constraint sampling. Our work focuses on constraint sampling under a sampling budget and builds on the results of Calafiore and Campi. For our random-time disruption model, we present sampling strategies that reduce the sample size requirement, as compared to direct application of the work in [7].

There has been little work in stochastic programming on the notion of the time of disruption being random. Salmerón et al. [48] study the problem of transporting military cargo between seaports subject to a biological attack by an enemy. In contrast to the probabilistic model we consider, their model is a stochastic mixed-integer program with recourse. A substantial body of work has also been devoted to the area of "disruption management" [53]. The

modeling approach in our work differs from disruption management in that decisions in our model are made initially to hedge against future disruption, while disruption management focuses on decisions made after the disruption.

2.1.2 Main Contributions

We approximately solve our probabilistic bicriteria model via constraint sampling subject to a sampling budget. In constraint sampling, we draw observations of the stochastic parameters and force a set of constraints generated by these observations to hold simultaneously. Thus, the original probabilistic program is replaced by a random, finitely constrained, sampled program.

In our bicriteria model, we seek to keep both cost and probabilistic risk low. The constraint sampling approximation minimizes cost subject to system dynamic constraints plus a set of constraints that ensures the bad event does not occur at the sampled observations. The constraint sampling approximation is a random program and hence, when optimized, yields a random decision vector. We call the expected value of the probabilistic risk function with respect to this random solution, the *expected probability of violation*.

Calafiore and Campi [7] bound the expected probability of violation. Furthermore, they use this bound and Markov's inequality to bound the probability that the random solution is infeasible to an associated probabilistically-constrained program. An improved bound on this probability is derived in [8]. When using constraint sampling to approximately solve a probabilistically-constrained program, the probability of obtaining a feasible solution is of fore-

most interest. In contrast, our goal is to approximately solve a probabilistic bicriteria model. Instead of focusing on the probability the chance constraint is violated for a specific level of risk tolerance, we study the trade-off between the expected probability of violation and cost. The bulk of this paper focuses on sampling in a manner that exploits the structure of the single-disruption model to tighten the bound on the expected probability of violation, tightened over that achieved by naive sampling. Apart from the resultant savings in sampling costs, the value of the tightened bound lies in potential improvement in cost for a given risk level.

The remainder of the chapter is organized as follows. In Section 2.2, we develop our model for a single random-time disruption problem. Then, we give an overview of the naive sampling result of Calafiore and Campi adapted to our random-time disruption model. In Section 2.3.1, we derive our improved bound under stratified sampling. We discuss optimal and order-optimal sampling strategies in Section 2.3.2, highlighting the simultaneous improvement in expected probability of violation and cost. In Section 2.4, we present an adversarial model for hedging against unknown distributions for the time of disruption. Section 2.5 extends our results to handle multiple disruptions.

2.2 The Model with a Single Disruption

2.2.1 Formulation and Assumptions

Consider the bicriteria minimization problem

$$\mathcal{BP}: \quad \underset{x}{\text{vmin}} \quad \left[\sum_{t=1}^{T} h_t(x_t),\ \mathbb{P}(f(x,\tilde{\xi}) > 0)\right] \tag{2.1a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} g_t(x_t) \leq b, \tag{2.1b}$$

$$x_t \in D_t,\ t = 1, \ldots, T, \tag{2.1c}$$

where $x = (x_1, \ldots, x_T)$. Model $\mathcal{BP}$ has two objective functions, a cost function $\sum_{t=1}^{T} h_t(x_t)$ and a risk function $\mathbb{P}(f(x,\tilde{\xi}) > 0)$. The vector minimization in (2.1a) means that the set of optimal solutions to $\mathcal{BP}$ is the set of all *Pareto optimal* points for the two functions. In other words, we seek the efficient frontier of the bicriteria problem, namely, the set of all points $(\sum_{t=1}^{T} h_t(x_t),\ \mathbb{P}(f(x,\tilde{\xi}) > 0))$ with the property that there does not exist $y = (y_1, \ldots, y_T)$ satisfying (2.1b) and (2.1c) such that $\sum_{t=1}^{T} h_t(y_t) \leq \sum_{t=1}^{T} h_t(x_t),\ \mathbb{P}(f(y,\tilde{\xi}) > 0) \leq \mathbb{P}(f(x,\tilde{\xi}) > 0)$ and at least one of these inequalities is strict. We typically restrict either the cost function or the risk function to some reasonable range rather than forming the entire efficient frontier of solutions.

We endow $\mathcal{BP}$ with the following structure. For every t, $h_t(\cdot)$ and $g_t(\cdot)$ are convex functions on D_t, a closed, convex set. The random vector $\tilde{\xi}$ has support Γ. We assume $f(\cdot,\xi)$ is convex on $\Pi_{t=1}^{T} D_t$ for every ξ in Γ. We interpret x_t as the decision vector for the t^{th} time period and denote the dimension of x_t by $\dim(x_t)$.

We assume that the random vector $\tilde{\xi}$ has the form $\tilde{\xi} = (\tilde{I}, \tilde{\delta})$, where the binary vector $\tilde{I} = (\tilde{I}_1, \ldots, \tilde{I}_T)$ indicates the time of disruption via $\tilde{I}_t = 1$, and $\tilde{\delta} = (\tilde{\delta}_1, \ldots, \tilde{\delta}_T)$ with $\tilde{\delta}_t$ being a random vector with support Γ_t. The notion that $\mathcal{BP}$ is a model with at most one disruption is captured by interpreting T as an artificial time period, in conjunction with $\sum_{t=1}^{T} \tilde{I}_t = 1$. The random vector $\tilde{\delta}_t$ represents the magnitude of the disruption given that it occurs in time period t. So, given a realization of $\tilde{\xi}$, we know when the disruption occurs as well as its magnitude. We let p_t denote the probability that the disruption occurs in time period t, i.e., $p_t = \mathbb{P}(\tilde{I}_t = 1),\ t = 1, \ldots, T$.

Assumption 1. The function f separates via $f(x, \tilde{\xi}) = \sum_{t=1}^{T} \tilde{I}_t\, f_t(x_t, \tilde{\delta}_t)$.

Given that the realizations of $\tilde{I}$ are unit vectors, Assumption 1 implies that convexity of $f(\cdot, \xi)$ for every $\xi \in \Gamma$ is equivalent to convexity of $f_t(\cdot, \delta_t)$ for every $\delta_t \in \Gamma_t,\ t = 1, \ldots, T$. Assumption 1 endows our model with additional structure that we exploit in deriving our main result in Section 2.3.1.

To illustrate the above model, we formulate the perishable inventory problem described in Example 1.

> **Example 1 (contd.).** Let h_t and l_t respectively denote unit holding and penalty costs for unsatisfied nominal demand of the product, and let d_t denote this nominal demand, all in period $t,\ t = 1, \ldots, T$. Let $\tilde{\delta}_t$ denote the random demand given that the disruption occurs in period t. The production cost in each time period is a convex piecewise linear function with two pieces and the break point at threshold level $c_t,\ t = 1 \ldots, T$. Production

in period t incurs a unit cost r_t up to c_t units and a unit cost $r'_t > r_t$ beyond. We now have the following model.

$$\begin{aligned}
\underset{x,y,u,s,w}{\text{vmin}} \quad & \left[\sum_{t=1}^{T}(r_t u_t^1 + r'_t u_t^2 + h_t s_t + l_t w_t),\ \mathbb{P}\Big(\sum_{t=1}^{T}\tilde{I}_t(\tilde{\delta}_t - x_t) > 0\Big)\right] \\
\text{s.t.} \quad & s_{t-1} + u_t^1 + u_t^2 = y_t + x_t + s_t,\ t = 1, \ldots, T, \\
& s_0 = 0,\ s_T = 0, \\
& w_t \geq d_t - y_t,\ t = 1, \ldots, T, \\
& u_t^1 \leq c_t,\ t = 1, \ldots, T, \\
& x_t, y_t, u_t^1, u_t^2, s_t, w_t \geq 0,\ t = 1, \ldots, T.
\end{aligned} \tag{2.2}$$

Here, the primary decision variables are u_t^1 and u_t^2, respectively denoting production amounts capped by, and in excess of, c_t, y_t denoting the amount used to meet nominal demand, x_t denoting the amount used to meet random demand, and surplus s_t stored for use in period $t+1$. We start with zero stock and end with zero stock. The amount of unsatisfied nominal demand is denoted by w_t. We revisit this model in the context of computational results in Section 2.3.4.

Model $\mathcal{BP}$ can be used to describe stochastic programs with, and without, recourse. In both cases, the decisions x_t, $t = 1, \ldots, T$, are *static*, i.e., they are made at the beginning before the disruption occurs. When there is no recourse, the decisions x_t, $t = 1, \ldots, T$, are implemented in successive time periods regardless of the occurrence of a disruption. In this sense, the disruption is an event that does not affect the dynamics of decision making but only

affects whether $f(x, \tilde{\xi}) \leq 0$ holds. When there is recourse, the decisions dictated by x are implemented until the time of disruption, say t. The function $f_t(\cdot, \cdot)$ can then represent a recourse function with the following form:

$$
\begin{aligned}
f_t(x_t, \delta_t) = \quad & \min_{y_t} \quad q_t(y_t, \delta_t), && \text{(2.3a)} \\
& \text{s.t.} \quad r_t(y_t, \delta_t) \leq -w_t(x_t, \delta_t), && \text{(2.3b)} \\
& \qquad\quad y_t \in B_t(\delta_t), && \text{(2.3c)}
\end{aligned}
$$

where $q_t(\cdot, \delta_t)$ and $r_t(\cdot, \delta_t)$ are convex functions on the closed convex set $B_t(\delta_t)$, and $w_t(\cdot, \delta_t)$ is convex on D_t, for every $\delta_t \in \Gamma_t$. The decision vector y_t can, for example, have dimension $\dim(x_{t+1}) + \cdots + \dim(x_T)$ and represent a recourse decision from period t to the horizon. When the disruption occurs in time period t, decisions $x_{t+1}, \ldots, x_T$ become irrelevant and this can be captured by introducing probability masses p_t appropriately in $\mathcal{BP}$. The cost objective function in (2.1a) becomes $\sum_{t=1}^{T} \left(\sum_{s=t}^{T} p_s\right) h_t(x_t)$, to capture the expected cost of the non-recourse decisions. If the constraint in (2.1b) represents a budget constraint, it can be a "hard" constraint, as it is currently stated, or it can be similarly expressed to capture an expected value budget constraint.

2.2.2 A Convex Approximation by Constraint Sampling

Assume we can draw independent and identically distributed (i.i.d.) observations from the distributions of $\tilde{\xi}$ and $\tilde{\delta}_t,\ t = 1, \ldots, T$. An observation drawn from the distribution of $\tilde{\delta}_t$ is simply conditioned on $\tilde{I}_t = 1$. Drawing M i.i.d. observations of $\tilde{\xi}$, denoted by $\tilde{\xi}^j, j = 1, \ldots, M$, we can form the convex pro-

gram

$$\begin{aligned} \mathcal{RP}^M: \quad & \min_x \quad \sum_{t=1}^{T} h_t(x_t) \\ & \text{s.t.} \quad \sum_{t=1}^{T} g_t(x_t) \leq b, \\ & \qquad f(x, \tilde{\xi}^j) \leq 0,\ j = 1, \ldots, M, \\ & \qquad x_t \in D_t,\ t = 1, \ldots, T. \end{aligned} \tag{2.4}$$

The convex program $\mathcal{RP}^M$ replaces the risk function in $\mathcal{BP}$ with the sampled constraints (2.4). Under the j^{th} observation $\tilde{\xi}^j = (\tilde{I}^j, \tilde{\delta}^j)$, the sampled constraint (2.4) has the form $f_t(x_t, \tilde{\delta}_t^j) \leq 0$, where $t = \sum_{t'=1}^{T} t' \tilde{I}_{t'}^j$ is the time period in which the disruption occurs. We assume that $\mathcal{RP}^M$ is feasible and attains its optimal value w.p.1. (See [7] for a discussion of the case in which $\mathcal{RP}^M$ violates these conditions.) Note that when $f_t(\cdot, \delta_t^j)$ is a recourse function as in (2.3), constraint (2.4) can be replaced by constraints (2.3b), (2.3c) and the constraint $q_t(y_t, \delta_t^j) \leq 0$ where $t = \sum_{t'=1}^{T} t' \tilde{I}_{t'}^j$.

Calafiore and Campi establish the following result, which we express in our setting.

Theorem 2.2.1. (Calafiore and Campi [7]) *Let M be a non-negative integer. And, let $\tilde{\xi}^1, \ldots, \tilde{\xi}^M$, i.i.d. from the distribution of $\tilde{\xi}$, be used to define $\mathcal{RP}^M$. Assume that $\mathcal{RP}^M$ has a unique optimal solution $\tilde{x}^M$, or that an appropriate tie-breaking rule is applied to obtain $\tilde{x}^M$. Then,*

$$\mathbb{E}\left[\mathbb{P}(f(\tilde{x}^M, \tilde{\xi}) > 0)\right] \leq \frac{1}{M+1} \sum_{t=1}^{T} \dim(x_t), \tag{2.5}$$

where the expectation is with respect to $(\tilde{\xi}^1, \ldots, \tilde{\xi}^M)$.

Theorem 2.2.1 specifies the sense in which we may view $\mathcal{RP}^M$ as an approximation of $\mathcal{BP}$: The bicriteria model has two objective functions, cost and risk which are simultaneously minimized in the Pareto sense. As the number of observations, M, grows large, greater emphasis is placed on minimizing risk. $\mathcal{RP}^M$ enables us to study the trade-off in cost in such an approach while enjoying the advantage of being computationally tractable.

We refer to $\tilde{x}_M$ as a *candidate solution* for $\mathcal{BP}$. For any candidate solution, $\tilde{x}_M$, we refer to $\mathbb{P}\big(f(\tilde{x}_M, \tilde{\xi}) > 0\big)$ as the *probability of violation* for $\tilde{x}_M$. Theorem 2.2.1 quantifies the expected probability of violation for the solution of $\mathcal{RP}^M$. By ranging the value of M in $\mathcal{RP}^M$, we can produce an approximation of $\mathcal{BP}$'s efficient frontier. Of course, $\mathcal{BP}$ is a nonconvex optimization model because $\mathbb{P}\big(f(x, \tilde{\xi}) > 0\big)$ is in general, not a convex function. So, there are limits to what one can expect from the convex approximating model $\mathcal{RP}^M$.

Drawing observation $\tilde{\xi}^j$ leads to a constraint that only involves the time period for which $\tilde{I}_t = 1$. So, with appropriate reindexing, constraints (2.4) can be rewritten as

$$f_t(x_t, \tilde{\delta}_t^j) \;\; \leq \;\; 0, \; t = 1, \ldots, T, \; j \in J_t,$$

where $J_t = \{j \mid \tilde{I}_t^j = 1\}$. Let N_t be the cardinality of J_t for each t so that $\sum_{t=1}^{T} N_t = M$. Each N_t is a random variable that counts the number of observations of $\tilde{\xi}$ for which $\tilde{I}_t = 1$. We focus on the following question in the

next section: Given the special structure of the single disruption model, can we improve on the bound (2.5) of Theorem 2.2.1 by controlling the sizes N_t? Or, equivalently, if we want the expected probability of violation to not exceed ϵ, can we achieve this with fewer observations than that prescribed by (2.5)?

We approach this issue as follows: Suppose that we draw N_t i.i.d. observations of $\tilde{\delta}_t,\ t = 1, \ldots, T$, with $\sum_{t=1}^{T} N_t = M$. In contrast to Theorem 2.2.1's setting, here we draw observations in a stratified manner so that N_t is deterministic. We show in Section 2.3.1 that the structure of the single disruption model $\mathcal{BP}$ enables derivation of an improved bound via stratification.

Before turning to our bound, we review the technical lemma that leads to inequality (2.5), and provide an alternate proof of this result. This lemma is proved in [7], with a shorter proof appearing in [8]. Both proofs work by contradiction and are based on the same idea. This idea can be used to construct a direct proof that is perhaps a little more concise.

We begin by stating the classical Helly's theorem from convex analysis, which can, for example, be found in [5].

Theorem 2.2.2. (Helly) *Consider a finite collection of convex subsets of $\mathbb{R}^n$, and assume that the intersection of every subcollection of $n+1$ (or fewer) sets has a nonempty intersection. Then, the entire collection has a nonempty intersection.*

Consider the convex program

$$\begin{aligned} \mathcal{P}: \quad z^* = \min_{x \in D} \quad & h(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \ i = 1, \ldots, m, \end{aligned}$$

where $h(\cdot)$ and $f_i(\cdot)$, $i = 1, \ldots, m$, are convex functions on D, a closed, convex subset of $\mathbb{R}^n$. Define $\mathcal{P}_k$ as the problem obtained by dropping the k^{th} constraint in $\mathcal{P}$:

$$\begin{aligned} \mathcal{P}_k: \quad z_k^* = \min_{x \in D} \quad & h(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \ i = 1, \ldots, k-1, k+1, \ldots, m. \end{aligned}$$

When solving the above models, we adopt the convention that if there exist multiple optimal solutions, we choose the one with the smallest $\| \cdot \|_2$ norm. Since all sets and functions are convex, this two-norm tie-breaking rule extracts a unique optimal solution.

Definition 2.2.1. The k^{th} constraint of $\mathcal{P}$ is called a *support constraint* for $\mathcal{P}$ if the optimal solutions to $\mathcal{P}_k$ and $\mathcal{P}$ are different under the two-norm tie-breaking rule.

We use the following lemma to derive our improved bound on the expected probability of violation.

Lemma 2.2.3. *The number of support constraints for $\mathcal{P}$ is at most n.*

Proof. Let x^* be the optimal solution to $\mathcal{P}$, and x_k^* be the optimal solution to $\mathcal{P}_k$, under the two-norm tie-breaking rule. Let $S \neq \emptyset$ be the index set of

support constraints. Define $\mathcal{X}' = \text{co}(\{x_k^*\}_{k\in S})$ and $\mathcal{X}_i = \{x \mid f_i(x) \leq 0\}$, $i = 1, \ldots, m$. If the k^{th} constraint is a support constraint for $\mathcal{P}$, then we have $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i\neq k}}^{m} \mathcal{X}_i \neq \emptyset$.

We claim that $\mathcal{X}' \cap \bigcap_{i=1}^{m} \mathcal{X}_i = \emptyset$. Suppose, by way of contradiction, that $w \in \mathcal{X}' \cap \bigcap_{i=1}^{m} \mathcal{X}_i$. Then, w is feasible for $\mathcal{P}$. Since dropping the k^{th} constraint gives a relaxation of $\mathcal{P}$, we have $z_k^* \leq z^*$ for all k in S. Let S' be a minimal subset of indices of S such that $w \in \text{co}(\{x_k^*\}_{k\in S'})$. If $z_k^* < z^*$ for some k in S', then, by the definition of S' and the convexity of h, it follows that $h(w) < z^*$, contradicting the optimality of x^*. Hence, $z_k^* = z^*$ for all k in S'. This implies $\|x_k^*\|_2 < \|x^*\|_2$ for all k in S' which in turn, by the convexity of $\|\cdot\|_2$, implies that $\|w\|_2 < \|x^*\|_2$, a contradiction to the tie-breaking rule. Thus, the claim is established.

Applying the contrapositive of Helly's theorem to the collection of sets $\{\mathcal{X}', \mathcal{X}_1, \ldots, \mathcal{X}_m\}$, there exists a subcollection of at most $n+1$ sets that has empty intersection. Since problem $\mathcal{P}$ is feasible, i.e., $\bigcap_{i=1}^{m} \mathcal{X}_i$ is nonempty, the collection of sets with empty intersection includes $\mathcal{X}'$. Reindexing the sets $\mathcal{X}_i$ if necessary, let the collection of sets with empty intersection be given by $\{\mathcal{X}', \mathcal{X}_1, \ldots, \mathcal{X}_\ell\}$, where $\ell \leq n$. By the definitions of S and support constraint, the k^{th} constraint is a support constraint if and only $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i\neq k}}^{m} \mathcal{X}_i \neq \emptyset$. For $k \notin \{1, \ldots, \ell\}$, we have $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i\neq k}}^{m} \mathcal{X}_i \subseteq \mathcal{X}' \cap \bigcap_{i=1}^{\ell} \mathcal{X}_i = \emptyset$, and hence for such k, the k^{th} constraint is not a support constraint. Therefore, the k^{th} constraint is a support constraint only if $k \in \{1, \ldots, \ell\}$, i.e., the number of support constraints is at most ℓ, which, in turn, is at most n. $\square$

2.3 Stratified Sampling

2.3.1 An Improved Bound

The following result provides an analog of Theorem 2.2.1 when we draw samples in a stratified manner. This new bound is used in Section 2.3.2 to develop alternative sampling strategies.

Theorem 2.3.1. *Let* $\mathbf{N} = (N_1, \ldots, N_T)$ *be a vector of non-negative integers satisfying* $\sum_{t=1}^{T} N_t = M$. *And, let* $\tilde{\delta}_t^1, \ldots, \tilde{\delta}_t^{N_t}$, *i.i.d. from the distribution of* $\tilde{\delta}_t$, $t = 1, \ldots, T$, *be used to define the convex program*

$$\begin{aligned} \mathcal{RP}^{\mathbf{N}}: \quad \min_{x} \quad & \sum_{t=1}^{T} h_t(x_t) \\ \text{s.t.} \quad & \sum_{t=1}^{T} g_t(x_t) \leq b, \\ & f_t(x_t, \tilde{\delta}_t^j) \leq 0,\ t = 1, \ldots, T,\ j = 1, \ldots, N_t, \\ & x_t \in D_t,\ t = 1, \ldots, T. \end{aligned}$$

Assume that $\mathcal{RP}^{\mathbf{N}}$ *has a unique optimal solution* $\tilde{x}^{\mathbf{N}}$, *or that the two-norm tie-breaking rule is applied to obtain* $\tilde{x}^{\mathbf{N}}$. *Then,*

$$\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0\big)\right] \leq \sum_{t=1}^{T} p_t \frac{\dim(x_t)}{N_t + 1}, \tag{2.6}$$

where the expectation is with respect to $(\tilde{\delta}_1^1, \ldots, \tilde{\delta}_1^{N_1}, \ldots, \tilde{\delta}_T^1, \ldots, \tilde{\delta}_T^{N_T})$.

Proof. Let $\tilde{\delta}^{\mathbf{N}} = (\tilde{\delta}_1^1, \ldots, \tilde{\delta}_1^{N_1}, \ldots, \tilde{\delta}_T^1, \ldots, \tilde{\delta}_T^{N_T})$. The expected probability of

violation is given by

$$\begin{aligned}
\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}},\tilde{\xi})>0\big)\right] &= \mathbb{E}\left[\mathbb{E}_{\tilde{\xi}}\left[\mathbb{I}\big(f(\tilde{x}^{\mathbf{N}},\tilde{\xi})>0\big)\,|\,\tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} p_t \mathbb{E}_{\tilde{\xi}}\left[\mathbb{I}\big(f(\tilde{x}^{\mathbf{N}},\tilde{\xi})>0\big)\,|\,\tilde{I}_t=1,\,\tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} p_t \mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}\big(f_t(\tilde{x}_t^{\mathbf{N}},\tilde{\delta}_t)>0\big)\,|\,\tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \sum_{t=1}^{T} p_t \underbrace{\mathbb{E}\left[\mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}\big(f_t(\tilde{x}_t^{\mathbf{N}},\tilde{\delta}_t)>0\big)\,|\,\tilde{\delta}^{\mathbf{N}}\right]\right]}_{V(\mathbf{N},t)}, \qquad (2.7)
\end{aligned}$$

where $\mathbb{I}(A)$ is an indicator random variable on the event A. Observe that the term $\mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}(f_t(\tilde{x}_t^{\mathbf{N}},\tilde{\delta}_t)>0)\,|\,\tilde{\delta}^{\mathbf{N}}\right]$ is the conditional probability of violation, given that the disruption occurs in period t, and given $\tilde{\delta}^{\mathbf{N}}$ which determines $\tilde{x}_t^{\mathbf{N}}$. Defining $V(\mathbf{N},t)$ as in (2.7), $V(\mathbf{N},t)$ is then the conditional expected probability of violation given that the disruption occurs in period t.

To derive the upper bound for $V(\mathbf{N},t)$, we fix the time period t and argue as follows: Let

$$\mathbf{N}^{+} = (N_1,\ldots,N_{t-1},N_t+1,N_{t+1},\ldots,N_T)$$

and define

$$\begin{aligned}
\mathcal{RP}^{\mathbf{N}^{+}}: \quad \min_{x} \quad & \sum_{s=1}^{T} h_s(x_s) \\
\text{s.t.} \quad & \sum_{s=1}^{T} g_s(x_s) \le b, \\
& f_s(x_s,\tilde{\delta}_s^{j}) \le 0,\ s=1,\ldots,T,\ j=1,\ldots,\mathbf{N}_s^{+}, \qquad (2.8) \\
& x_s \in D_s,\ s=1,\ldots,T.
\end{aligned}$$

Problem $\mathcal{RP}^{\mathbf{N}^+}$ is identical to $\mathcal{RP}^{\mathbf{N}}$ with an additional constraint in period t, from an i.i.d. observation $\tilde{\delta}_t^{N_t+1}$ from the distribution of $\tilde{\delta}_t$. For $j \in \{1, \ldots, N_t+1\}$, let $\mathcal{RP}_j^{\mathbf{N}^+}$ be identical to $\mathcal{RP}^{\mathbf{N}^+}$ except that the constraint associated with observation $\tilde{\delta}_t^j$ in (2.8) is dropped, and let $\tilde{x}_j^{\mathbf{N}^+}$ be the optimal solution to $\mathcal{RP}_j^{\mathbf{N}^+}$ under the two-norm tie-breaking rule. Let $\tilde{\delta}^{\mathbf{N}^+}(j)$ be the sample used to define $\mathcal{RP}_j^{\mathbf{N}^+}$ and $\tilde{\delta}^{\mathbf{N}^+}$ be that for $\mathcal{RP}^{\mathbf{N}^+}$, where we continue to suppress dependence on t. Then,

$$\begin{aligned} V(\mathbf{N},t) &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}(N_t+1)}\left[\mathbb{E}_{\tilde{\delta}_t^{N_t+1}}\left[\mathbb{I}(f_t(\tilde{x}_{N_t+1,t}^{\mathbf{N}^+}, \tilde{\delta}_t^{N_t+1}) > 0) \,|\, \tilde{\delta}^{\mathbf{N}^+}(N_t+1)\right]\right] \\ &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}_{N_t+1,t}^{\mathbf{N}^+}, \tilde{\delta}_t^{N_t+1}) > 0)\right], \end{aligned}$$

where $\tilde{x}_{N_t+1,t}^{\mathbf{N}^+}$ is the period t subvector of $\tilde{x}_{N_t+1}^{\mathbf{N}^+}$. Furthermore, for $j = 1, \ldots, N_t+1$,

$$\begin{aligned} \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}(j)}\left[\mathbb{E}_{\tilde{\delta}_t^j}\left[\mathbb{I}(f_t(\tilde{x}_{j,t}^{\mathbf{N}^+}, \tilde{\delta}_t^j) > 0) \,|\, \tilde{\delta}^{\mathbf{N}^+}(j)\right]\right] &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}_{j,t}^{\mathbf{N}^+}, \tilde{\delta}_t^j) > 0)\right] \\ &= V(\mathbf{N},t), \end{aligned}$$

since $(\tilde{\delta}_t^1, \ldots, \tilde{\delta}_t^{N_t+1})$ are i.i.d. Hence,

$$\begin{aligned} V(\mathbf{N},t) &= \frac{1}{N_t+1}\sum_{j=1}^{N_t+1} \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}_{j,t}^{\mathbf{N}^+}, \tilde{\delta}_t^j) > 0)\right] \\ &= \frac{1}{N_t+1}\mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\Big[\underbrace{\sum_{j=1}^{N_t+1} \mathbb{I}(f_t(\tilde{x}_{j,t}^{\mathbf{N}^+}, \tilde{\delta}_t^j) > 0)}_{\tilde{u}_t^{\mathbf{N}}}\Big]. \end{aligned} \tag{2.9}$$

Defining $\tilde{u}_t^{\mathbf{N}}$ as in (2.9), we see that $\tilde{u}_t^{\mathbf{N}}$ counts the number of observations in the sample $\tilde{\delta}^{\mathbf{N}^+}$, which are support constraints for the problem $\mathcal{RP}^{\mathbf{N}^+}$.

We now rewrite $\mathcal{RP}^{\mathbf{N}^+}$ in the form

$$\begin{aligned} \min_{x} \quad & h_t(x_t) + G^{\mathbf{N}^+}(x_t) \\ \text{s.t.} \quad & f_t(x_t, \tilde{\delta}_t^j) \leq 0,\ j = 1, \ldots, N_t + 1, \\ & x_t \in H^{\mathbf{N}^+}, \end{aligned} \tag{2.10}$$

where $H^{\mathbf{N}^+} = \{x_t \in D_t \mid \exists\, x_s \in D_s, s = 1, \ldots, T, s \neq t, \text{ satisfying } \sum_{s=1}^{T} g_s(x_s) \leq b\}$, and

$$\begin{aligned} G^{\mathbf{N}^+}(x_t) = \min_{\substack{x_s, s=1,\ldots,T \\ s \neq t}} \quad & \sum_{\substack{s=1 \\ s \neq t}}^{T} h_s(x_s) \\ \text{s.t.} \quad & \sum_{\substack{s=1 \\ s \neq t}}^{T} g_s(x_s) \leq b - g_t(x_t), \\ & f_s(x_s, \tilde{\delta}_s^j) \leq 0,\ s = 1, \ldots, T,\ s \neq t,\ j = 1, \ldots, N_s, \\ & x_s \in D_s,\ s = 1, \ldots, T,\ s \neq t. \end{aligned}$$

Convexity of $f_s(\cdot, \tilde{\delta}_s^j), g_s(\cdot), h_s(\cdot)$ and $D_s,\ s = 1, \ldots, T$, implies $H^{\mathbf{N}^+}$ is convex, and $G^{\mathbf{N}^+}(\cdot)$ is convex on $H^{\mathbf{N}^+}$. By Lemma 2.2.3, the number of support constraints for (2.10), or equivalently $\tilde{u}_t^{\mathbf{N}}$, is at most $\dim(x_t)$. Thus,

$$V(\mathbf{N}, t) \leq \frac{\dim(x_t)}{N_t + 1},$$

and using this in equation (2.7) yields

$$\mathbb{E}[\mathbb{P}(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0)] \leq \sum_{t=1}^{T} p_t \frac{\dim(x_t)}{N_t + 1},$$

as desired. □

The potential value of Theorem 2.3.1 lies in choosing the sample sizes N_t in a stratified manner so as to tighten the bound of (2.6) over that achieved by (2.5). This is discussed in the next section.

2.3.2 Optimal Sampling Strategies

In this section, we design sampling allocations, $N_1, \ldots, N_T$, using bound (2.6) on the expected probability of violation from Theorem 2.3.1 to guide our sample allocation. Given a sampling budget M, we let $M = NT$ to indicate that individual time periods receive an average allocation of N observations. In this setting, we seek to solve

$$\mathbb{SS}_{\text{int}}: \quad \min_{N_1,\ldots,N_T} \quad \sum_{t=1}^{T} \frac{p_t n_t}{N_t + 1} \tag{2.11a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} N_t = NT,$$

$$N_t \in \mathbb{Z}_+, \;\; t = 1, \ldots, T, \tag{2.11b}$$

where $\mathbb{Z}_+$ is the set of non-negative integers, and we denote $\dim(x_t)$ by n_t, $t = 1, \ldots, T$. For small $M = NT$, the objective function in (2.11a) would involve terms $p_t \min(1, n_t/(N_t + 1))$ but for large M, (2.11a) suffices since the optimal value shrinks to zero as M grows. We denote by $\mathbb{SS}$ the continuous relaxation of $\mathbb{SS}_{\text{int}}$, i.e., $\mathbb{SS}$ is identical to model $\mathbb{SS}_{\text{int}}$ except that constraints (2.11b) are replaced by $N_t \geq 0$, $t = 1, \ldots, T$. The objective function for $\mathbb{SS}$ is strictly convex in $(N_1, \ldots, N_T)$ and its feasible region is convex and compact. The

unique optimal solution to $\mathcal{SS}$ and its optimal value $V^*(N)$ are given by

$$N_t^* = \left(\frac{(p_t n_t)^{1/2}}{\sum_{t=1}^{T} (p_t n_t)^{1/2}} \right) (N+1)T - 1, \; t = 1, \ldots, T, \tag{2.12}$$

$$V^*(N) = \frac{\left(\sum_{t=1}^{T} (p_t n_t)^{1/2} \right)^2}{(N+1)T}. \tag{2.13}$$

For practical implementation, we assume that the sample sizes are sufficiently large that we may round the values in (2.12) to achieve integer-valued allocations. For the analysis that follows, we neglect such issues and simply consider the continuous allocation.

2.3.3 Improvement over Naive Sampling

The proportion $\alpha_t = (p_t n_t)^{1/2} / \sum_{t=1}^{T} (p_t n_t)^{1/2}$ that appears in (2.12) is effectively the fraction of our sample budget that is allocated to period t. If we instead perform naive sampling and construct model $\mathcal{RP}^M$, then a random number of observations is allocated to each time period. However, for large M, the proportion of observations in $\mathcal{RP}^M$ allocated to period t converges (by the law of large numbers) to p_t. Hence, it is natural to consider what happens when we replace the optimal proportion in (2.12) with $\alpha_t = p_t, \; t = 1, \ldots, T$. Theorem 2.3.2 compares the bound so obtained with the bound in Theorem 2.2.1 with $M = NT$, and characterizes the improvement attained by the optimal value bound (2.13).

Theorem 2.3.2. *Let $V^1(N) = \sum_{t=1}^{T} n_t/(NT+1)$ denote the bound (2.5) on the expected probability of violation with $M = NT$. Let $V^2(N)$ denote the*

bound (2.6) of Theorem 2.3.1 with $N_t = p_t(N+1)T - 1$. *And, let* $V^*(N)$ *denote the bound (2.13) of Theorem 2.3.1 under sample size allocation (2.12). Then,*

$$(i) \quad V^*(N) \le V^2(N) \le V^1(N),$$

$$(ii) \quad \frac{V^2(N)}{V^1(N)} \to 1 \ \text{ as } N \to \infty, \tag{2.14a}$$

$$(iii) \quad \frac{V^*(N)}{V^2(N)} = \frac{\left(\sum_{t=1}^T (p_t n_t)^{1/2}\right)^2}{\sum_{t=1}^T n_t}. \tag{2.14b}$$

Moreover, if $n_t = n, \ t = 1, \ldots, T,$

$$(iv) \quad \frac{1}{T} \le \frac{V^*(N)}{V^2(N)} \le 1, \tag{2.14c}$$

and these bounds are tight, i.e., they can be achieved.

Proof. Substituting $N_t = p_t(N+1)T - 1$ into bound (2.6) yields

$$V^2(N) = \frac{1}{(N+1)T} \sum_{t=1}^T n_t, \tag{2.15}$$

and so $V^2(N) \le V^1(N)$ is immediate. $V^*(N) \le V^2(N)$ follows from the fact that $N_t = p_t(N+1)T - 1$ is a feasible solution to $\mathcal{SS}$ and $V^*(N)$ is its optimal value. Given $V^1(N) = \sum_{t=1}^T n_t/(NT+1)$ and $V^2(N)$ in equation (2.15), result (2.14a) follows, and equation (2.14b) is simply the ratio of (2.13) to (2.15). With $n_t = n, \ t = 1, \ldots, T$, equation (2.14b) reduces to $(\sum_{t=1}^T p_t^{1/2})^2/T$. If we minimize $(\sum_{t=1}^T p_t^{1/2})^2/T$ subject to $\sum_{t=1}^T p_t = 1, \ p_t \ge 0, \ t = 1, \ldots, T$, then the optimal p is a degenerate distribution, with all the probability mass in a single time period, and maximizing $(\sum_{t=1}^T p_t^{1/2})^2/T$ subject to the same

constraints yields the equal allocation solution $p = (1/T, \ldots, 1/T)$. These two distributions achieve the lower and upper bounds, respectively, in (2.14c). $\square$

Result (2.14a) of Theorem 2.3.2 shows that the bound of Theorem 2.2.1 and that of Theorem 2.3.1 under the proportional allocation, $N_t = p_t(N+1)T-1$, are essentially identical. Results (2.14b) and (2.14c) of the theorem characterize the improvement over these bounds provided by the optimal allocation (2.12). The factor of improvement can be T^{-1}, i.e., the improvement can be an order of magnitude in the number of time periods. In other words, if $n_t = n,\ t = 1, \ldots, T$, then $V^1(N)$ and $V^2(N)$ shrink to zero with N^{-1}, but $V^*(N)$ can shrink to zero as quickly as $N^{-1}T^{-1}$.

More generally, if $n_t = n,\ t = 1, \ldots, T$, then $V^2(N) = \Theta(1/N)$ as $T \to \infty$ and $V^*(N) = \Theta\left((\sum_{t=1}^{T} p_t^{1/2})^2/NT\right)$ as $T \to \infty$. For $V^*(N)$ to achieve the rate $N^{-1}T^{-1}$, we require $(\sum_{t=1}^{T} p_t^{1/2})^2 = \Theta(1)$ as $T \to \infty$. The degenerate distribution which achieves the rate in the proof of Theorem 2.3.2 is extreme and arguably not one that would arise in practice. When modeling the time until a disruption occurs, a natural distribution that arises is a geometric distribution, i.e., $p_t = (1-q)q^{t-1},\ t \in \mathbb{Z}_+$, is the probability that the disruption occurs in period t, for some $q \in (0, 1)$. Of course, this distribution only applies for an infinite horizon while our model has finite horizon T. One adaptation under a finite horizon is to introduce an additional artificial period to the model, $T + 1$. With $p_t = (1 - q)q^{t-1},\ t = 1, \ldots, T$, and $p_{T+1} = q^T$, we retain the memoryless property of the geometric distribution. Another natural adaptation is the truncated geometric distribution given by $p_t = (1 -$

$q)q^{t-1}/(1-q^T),\ t=1,\ldots,T$. The following corollary establishes that the $N^{-1}T^{-1}$ rate is achieved by these variants of the geometric distribution.

Corollary 2.3.3. *Let $V^*(N)/V^2(N)$ be given by equation (2.14b), assume $n_t = n,\ t = 1,\ldots,T$, and let $q \in (0,1)$. If the probability mass function governing the time of the disruption satisfies*

$$p_t = (1-q)q^{t-1},\ t = 1,\ldots,T,\ p_{T+1} = q^T \tag{2.16a}$$

or

$$p_t = \frac{(1-q)q^{t-1}}{1-q^T},\ t = 1,\ldots,T, \tag{2.16b}$$

then $V^*(N)/V^2(N) = \Theta(1/T)$ as $T \to \infty$.

Proof. Under (2.16a) and $n_t = n,\ t = 1,\ldots,T$,

$$\frac{V^*(N)}{V^2(N)} = \frac{\left((1-q)^{1/2}\sum_{t=1}^{T} q^{(t-1)/2} + q^{T/2}\right)^2}{T} = \Theta\left(\frac{1}{T}\right).$$

Under (2.16b) and $n_t = n,\ t = 1,\ldots,T$,

$$\frac{V^*(N)}{V^2(N)} = \frac{(1-q)\left(\sum_{t=1}^{T} q^{(t-1)/2}\right)^2}{T(1-q^T)} = \Theta\left(\frac{1}{T}\right).$$

□

2.3.4 Proximity to Efficient Frontier

Bounds (2.5) and (2.13) can be used to assess the proximity to the efficient frontier under naive and optimal stratified sampling. Setting the naive sampling bound (2.5) equal to a risk level ϵ in the range of interest, with $M = NT$,

we obtain

$$\frac{1}{NT+1}\sum_{t=1}^{T} n_t = \epsilon.$$

Similarly, setting the optimal allocation bound (2.13) to ϵ yields

$$\frac{\left(\sum_{t=1}^{T}(p_t n_t)^{1/2}\right)^2}{(N+1)T} = \epsilon.$$

In order to ensure that the expected probability of violation is bounded by ϵ, we require therefore that

$$N \geq \left\lceil \frac{1}{T}\left(\frac{1}{\epsilon}\sum_{t=1}^{T} n_t - 1\right)\right\rceil \tag{2.17}$$

for the naive strategy and

$$N \geq \left\lceil \frac{1}{\epsilon T}\left(\sum_{t=1}^{T}(p_t n_t)^{1/2}\right)^2 - 1\right\rceil \tag{2.18}$$

for the optimal allocation strategy. The convex approximations $\mathcal{RP}^{NT}$ and $\mathcal{RP}^{\mathbf{N}}$ can be solved with N equal to the lower bounds of (2.17) and (2.18), to yield candidate solutions. By repeating this procedure and averaging, we obtain estimates of the associated expected cost. These values can then be used to assess the relative proximity to the efficient frontier for the two strategies.

We consider an instance of the inventory control model in Section 2.2.1 with $T = 20$. The time of disruption is assumed to obey a truncated geometric distribution as given in (2.16b), with $q = .8$. We assume that the random variables $\tilde{\delta}_t$ obey a Pareto distribution with a shape parameter, k, equal to 20, and a threshold ψ_0, equal to 1. This implies we have $\mathbb{P}(\tilde{\delta}_t > x) = (\psi_0/x)^{k-1}$ for

Table 2.1: Comparison of naive and stratified sampling schemes for the single product inventory control problem of model (2.2).

Risk	Opt. Cost	Naive		Stratified		$\Delta\overline{z}^*$
		N	$\mathbb{E}$[Opt. Cost]	N	$\mathbb{E}$[Opt. Cost]	
0.010	15.498	100	15.994	72	15.888	21.38%
0.015	15.263	67	15.751	48	15.615	27.99%
0.020	15.100	50	15.552	36	15.465	19.36%
0.025	14.975	40	15.450	28	15.311	29.29%
0.030	14.874	34	15.354	24	15.213	29.36%
0.035	14.789	29	15.250	20	15.129	26.24%
0.040	14.716	25	15.180	18	15.073	23.16%
0.045	14.653	23	15.127	16	15.024	21.62%
0.050	14.596	20	15.059	14	15.915	31.11%
0.055	14.545	19	15.003	13	14.903	21.79%

$x \geq \psi_0$ and 1 otherwise, for $t = 1, \ldots, T$. Given this distribution, we impose the condition that the allocation x_t must be at least the threshold value of ψ_0. Thus, the model includes the additional constraints $x_t \geq \psi_0,\ t = 1, \ldots, T$. The nominal demand in each time period is set equal to .5. Costs decay geometrically over time; we set $h_1 = 3, l_1 = 1.1, r_1 = 1,\ r_1' = 1.5$, with a common decay rate of .9. The cut-off level c_t is chosen to be 3 for all t. We emphasize that the actual values of these parameters are unimportant as long as the following conditions hold: (1) T is large enough to capture the asymptotic behavior discussed in Section 2.3.3, and (2) costs decay at an appreciable rate over time in order to take advantage of the slower decay of the allocations N_t^* with time, in comparison with p_t.

Table 2.2: Parameter estimates and confidence interval half-widths for Table 2.1. EOC denotes expected optimal cost.

Risk	Naive				Stratified			
	EOC	CI Width	$\mathbb{E}$[PoV]	CI Width	EOC	CI Width	$\mathbb{E}$[PoV]	CI Width
0.010	15.994	.0293	.0099	.0003	15.888	.0290	.0098	.0003
0.015	15.751	.0286	.0148	.0005	15.615	.0280	.0149	.0005
0.02	15.552	.0277	.0202	.0007	15.465	.0284	.0197	.0007
0.025	15.450	.0271	.0245	.0010	15.311	.0280	.0252	.0010
0.030	15.354	.0277	.0291	.0009	15.213	.0251	.0295	.0010
0.035	15.250	.0269	.0348	.0012	15.129	.0272	.0345	.0010
0.040	15.180	.0303	.0384	.0014	15.073	.0280	.0379	.0014
0.045	15.127	.0274	.0418	.0014	15.024	.0262	.0418	.0015
0.050	15.059	.0251	.0471	.0018	14.915	.0279	.0491	.0018
0.055	15.003	.0267	.0501	.0019	14.903	.0282	.0525	.0019

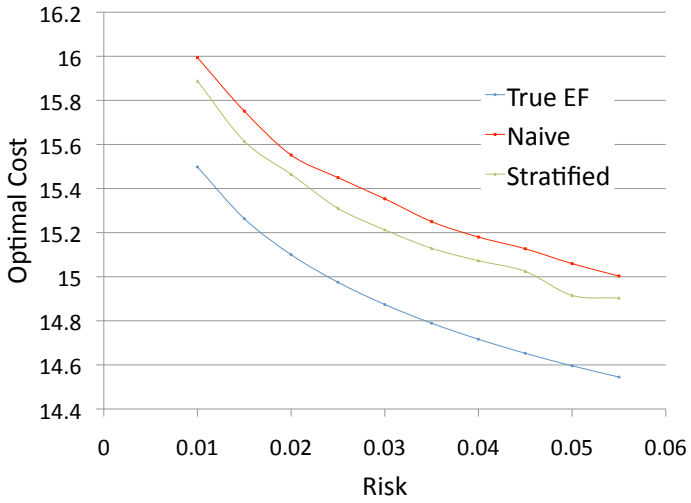

Figure 2.1: Plot of optimal cost against risk threshold showing improved proximity to the true efficient frontier with stratified sampling, compared to naive sampling. We use model (2.2) with parameter values given in Section 2.3.4.

Both the nonlinear bicriteria model and the convex approximations were solved in the GAMS environment, with calls to MINOS and CPLEX. Observe that the risk and cost functions are both convex, and hence the true

efficient frontier is also convex. To approximate the expected optimal cost from sampling, we replicate the sampling experiment 200 times. We report the results in Table 2.1, including the values of risk, optimal cost, and estimated expected optimal cost using the naive and optimal allocation strategies. We also report the values of the lower bounds in (2.17) and (2.18) with $n_t = 1,\ t = 1, \ldots, T$. (Note that this is the effective number of decision variables in each time period affected by the sampled constraints, since the rest can be subsumed into an appropriate recourse function as in the proof of Theorem 2.3.1.) Finally, we also report $\Delta\overline{z}^*$, the improvement in cost for stratified sampling over naive sampling, as a percentage of the difference between the true cost and expected cost due to naive sampling. In Table 2.2, we report the half-widths of 95% confidence intervals associated with the estimates for expected cost and probability of violation.

The smaller values of N in Table 2.1 under stratified sampling, relative to naive sampling, demonstrate savings in sampling costs. The cost values indicate a significant improvement with using optimal stratification. We see that the expected optimal value of the stratified approximating model improves on that of the naive model by at least 19%, and as much as 31%, for the range of risk values considered. The numbers in Table 2.2 show that sampling error is no more than 4%. Figure 2.1 plots the true efficient frontier as well as expected optimal costs from both strategies and illustrates the increased proximity to the efficient frontier due to optimal stratification.

2.3.5 Comparing Allocation Strategies

In this section, we study how other allocation strategies compare with the optimal allocation (2.12) as T grows large. We consider sample allocations of the form $N_t = \alpha_t(N+1)T - 1,\ t = 1, \ldots, T$, satisfying $\sum_{t=1}^{T} \alpha_t = 1,\ \alpha_t \geq 0,\ t = 1, \ldots, T$. Under the assumption $n_t = n,\ t = 1, \ldots, T$, the bound (2.6) under such an allocation reduces to $\frac{n}{(N+1)T} \sum_{t=1}^{T} p_t/\alpha_t$. Letting $p = (p_1, \ldots, p_T)$ and defining $w(\alpha;\, p) = \sum_{t=1}^{T} p_t/\alpha_t$, the bound becomes $n\, w(\alpha; p)/((N+1)T)$. Observe that

a) The optimal allocation vector α^* is unique, α_t^* is proportional to the square root of p_t, and we have

$$\min_{\alpha \in \Delta} \ w(\alpha; p) = \left(\sum_{t=1}^{T} p_t^{1/2} \right)^2, \tag{2.19}$$

where $\Delta = \{\alpha \in \mathbb{R}_+^T \mid \sum_{t=1}^{T} \alpha_t = 1\}$.

b) The uniform allocation vector $\alpha_{\text{unif}} = (1/T, \ldots, 1/T)$ and the proportional allocation vector $\alpha_{\text{prop}} = (p_1, \ldots, p_T)$ behave identically with respect to $w(\cdot\,;\, p)$, i.e., $w(\alpha_{\text{unif}};\, p) = w(\alpha_{\text{prop}};\, p)$. Thus, there is no difference between allocating samples proportionally (as is done in naive sampling) and allocating samples uniformly.

c) The inequality $w(\alpha_{\text{unif}}\,;\, p) = T \geq w(\alpha^*\,;\, p) = (\sum_{t=1}^{T} {p_t}^{1/2})^2$ is tight if and only if $p_t = 1/T,\ t = 1, \ldots, T$. In other words, the uniform allocation is optimal if and only if the distribution governing the time of occurrence of the disruption is also uniform.

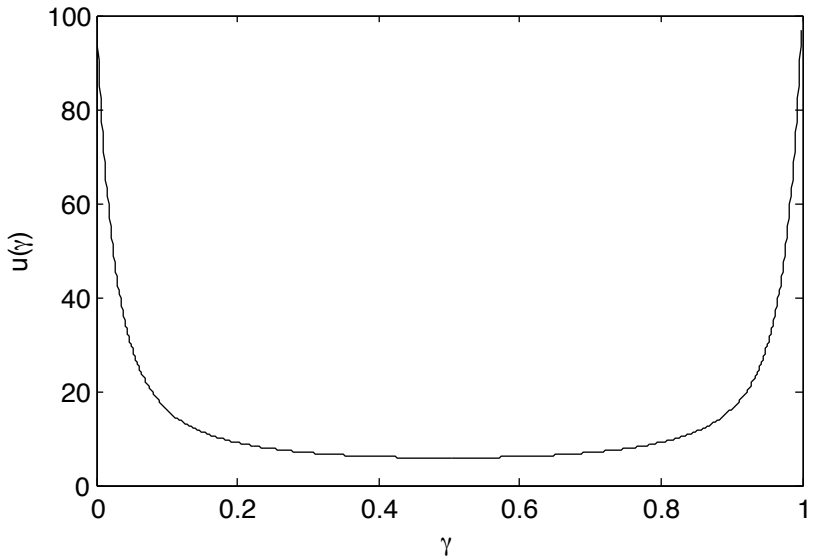

Figure 2.2: Plot of $u(\gamma)$ showing dependence of bound on the expected probability of violation on allocations N_t that are proportional to the γ^{th} power of the masses p_t, when $p_t = (1-q)q^{t-1}/(1-q^T),\ t = 1\ldots, T$, with $T = 100$ and $q = .5$.

The observation $w(\alpha_{\text{unif}};\, p) = w(\alpha_{\text{prop}};\, p)$ suggests $w(\cdot\,;\, p)$ is symmetric when α is proportional to a power of p, i.e., when $\alpha = \alpha(\gamma)$ with $\alpha_t(\gamma) = p_t^\gamma / \sum_{t=1}^T p_t^\gamma,\ t = 1, \ldots, T$, where $\gamma \in \mathbb{R}$. Let

$$\begin{aligned} u(\gamma) \,= w(\alpha(\gamma)\,;\, p) \;\; &= \;\; \sum_{t=1}^T (p_t)^\gamma \sum_{t=1}^T \frac{p_t}{(p_t)^\gamma} \\ &= \;\; \sum_{t=1}^T (p_t)^\gamma \sum_{t=1}^T (p_t)^{1-\gamma}. \end{aligned}$$

Thus, $u(\cdot)$ is symmetric about $\gamma = 1/2$. Further, as T grows large, we restrict attention to non-degenerate choices of p and γ such that (a) $(p_t)^\gamma$ decays fast enough to be summable in T, and (b) $(p_t)^\gamma$ decays slow enough that $p_t^{1-\gamma}$ is

summable in T. Note that if conditions (a) and (b) hold for some vectors p and α, then $u(\gamma) = \Theta(1)$ as $T \to \infty$, and we achieve the rate of $N^{-1}T^{-1}$ achieved by α^* in bound (2.6). Recalling that $(\sum_{t=1}^{T} p_t^{1/2})^2 \geq 1$ for all choices of p, it follows that these values of γ are order-optimal in the sense of the optimization in (2.19). Now, if $\gamma \leq 0$, condition (a) is not satisfied, while if $\gamma \geq 1$, condition (b) is not satisfied. However, if the masses p_t decay geometrically, then, for any value of $\gamma \in (0, 1)$, conditions (a) and (b) both hold. Figure 2.2 illustrates this behavior. However, if the masses p_t decay polynomially, then the range of γ values for which conditions (a) and (b) hold simultaneously is more restrictive. Theorem 2.3.4 formalizes this observation.

Theorem 2.3.4. *Assume $n_t = n$, $t = 1, \ldots, T$. Further, assume that one of the following holds:*

(i) $p_t \propto t^{-r}$, $t = 1, \ldots, T$ for some $r > 2$, and $\gamma \in (1/r, 1 - 1/r)$

(ii) $p_t \propto q^t$, $t = 1 \ldots, T$ for some $q \in (0, 1)$, and $\gamma \in (0, 1)$.

Let $\alpha_t = p_t^{\gamma} / \sum_{t=1}^{T} p_t^{\gamma}$, $t = 1, \ldots, T$. Then,

$$\frac{w(\alpha\,;\, p)/\big((N+1)T\big)}{V^2(N)} \;\; = \;\; \Theta(1/T) \;\; \text{as} \;\; T \to \infty.$$

Proof. It suffices to establish that $u(\gamma) = \Theta(1)$ as $T \to \infty$. Suppose (i) holds. Then,

$$\begin{aligned} u(\gamma) &= \sum_{t=1}^{T} 1/t^{\gamma r} \sum_{t=1}^{T} 1/t^{(1-\gamma)r} / \sum_{t=1}^{T} 1/t^{r} \\ &= \Theta(1) \quad \text{as } T \to \infty, \end{aligned}$$

since $\gamma r > 1$, $(1-\gamma)r > 1$ and $r > 2$. Next, suppose (ii) holds. Then,

$$\begin{aligned} u(\gamma) &= \sum_{t=1}^{T} q^{\gamma t} \sum_{t=1}^{T} q^{(1-\gamma)t} / \sum_{t=1}^{T} q^{t} \\ &= \Theta(1) \quad \text{as } T \to \infty, \end{aligned}$$

since $0 < q$, $\gamma < 1$. □

Theorem 2.3.4 suggests that for a more general probability mass function p, there may be conditions under which a polynomial allocation, α, yields similar results. In other words, we pose the question: When are the (generalized) conditions (a) $\sum_{t=1}^{T} \alpha_t$ is summable in T and (b) $\sum_{t=1}^{T} p_t/\alpha_t$ is summable in T, simultaneously true as T grows large? For example, suppose the distribution of the time of disruption has finite second moment in the limit $T \to \infty$, i.e., $\sum_{t=1}^{T} t^2 p_t$ is uniformly bounded. Under this assumption, letting $\alpha_t \propto t^{-2}$ satisfies conditions (a) and (b). More generally, both conditions hold when p has uniformly bounded k^{th} moment, and $\alpha_t \propto t^{-k}$, for some $k > 1$. Theorem 2.3.5 furnishes the details of this argument.

Theorem 2.3.5. *(Bounded Moment Condition) Let $k > 1$ and assume $n_t = n,\ t = 1, \ldots, T$. Further, assume*

$$\limsup_{T \to \infty} \sum_{t=1}^{T} t^k p_t < \infty. \tag{2.20}$$

Let $\alpha_t = t^{-k} / \sum_{t=1}^{T} t^{-k},\ t = 1, \ldots, T$. Then,

$$\frac{w(\alpha; p) / \big((N+1)T\big)}{V^2(N)} = \Theta(1/T) \text{ as } T \to \infty.$$

Proof. As in Theorem 2.3.4, it suffices to show that $w(\alpha; p) = \Theta(1)$ as $T \to \infty$. From (2.20), there exists a finite constant c with $\sum_{t=1}^{T} t^k p_t < c$ for all T; we also have $\sum_{t=1}^{T} t^{-k} < \zeta(k) = \sum_{t=1}^{\infty} t^{-k}$, where $\zeta(k)$ is finite since $k > 1$. Thus,

$$1 \leq (\sum_{t=1}^{T} p_t^{1/2})^2 \leq w(\alpha; p) = \sum_{t=1}^{T} \frac{p_t}{\alpha_t} = \sum_{t=1}^{T} t^{-k} \sum_{t=1}^{T} t^k p_t < c\,\zeta(k) \;\; \forall T$$

establishing that $w(\alpha; p) = \Theta(1)$ as desired. □

The import of Theorem 2.3.5 is the following: If the distribution governing the time of disruption is unknown, we cannot use the sample size allocation of (2.12). However, if that distribution can be assumed to satisfy the bounded moment condition (2.20), then the theorem specifies a sample-size allocation that is both order-optimal and improves on that of naive sampling by an order of magnitude in T.

Furthermore, from Theorems 2.3.4 and 2.3.5, if p_t decays geometrically, any geometric allocation that decays more slowly than p, as well as any summable polynomially-decaying allocation, is order optimal. Surprisingly, among such order-optimal allocations, a suitable polynomial allocation outperforms a geometric allocation that is far from optimal. For example, let p_t have a truncated geometric distribution, i.e., $p_t = (1-q)q^{t-1}/(1-q^T),\ t = 1, \ldots, T$. Letting $q = 1/6$, $T = 100$ and $\gamma = 1/4$ gives $u(\gamma) \approxeq 3.123$, while letting $\alpha_t = t^{-2}/\sum_{t=1}^{T} t^{-2},\ t = 1, \ldots, T$, gives $w(\alpha; p) \approxeq 2.747$. This dominance holds for all $\gamma \in (0, 1/4] \cup [3/4, 1)$. This is summarized in the remark below.

Remark. There exists a time horizon T and a truncated geometric distribution p over this horizon for which, under $n_t = n,\ t = 1, \ldots, T$, a suitably chosen

polynomial allocation outperforms a non-trivial set of geometric allocations in the sense of bound (2.6).

In summary, we observe: If p is unknown, or subject to perturbation, but decays fast enough to have a bounded k^{th} moment, Theorem 2.3.5 suggests a polynomial allocation. The next section considers the case when we cannot make this assumption, and the distribution of the time of disruption is selected to make our choice of α as bad as possible.

2.4 An Adversarial Problem

In this section, we consider perturbations of p that are more severe than those considered above. In particular, we assume an adversary selects the distribution governing the time of disruption. We model this by considering a game involving two players, the sampler and the adversary. The sampler picks the allocation vector α first, and then the adversary picks the distribution p. The sampler seeks to minimize the expected probability of violation while the adversary's goal is to maximize the same. For simplicity, we assume $n_t = n,\ t = 1, \ldots, T$. If the adversary is allowed to choose any probability mass function p_t, $t = 1, \ldots, T$, then the problem is devoid of interest since the adversary simply chooses $t' \in \operatorname{argmin}_{1 \le t \le T} \alpha_t$ and sets $p_{t'} = 1$. However, we assume we have some information concerning the distribution of the time of disruption that allows us to restrict the probability mass function to lie in a proper subset of the simplex Δ. Specifically, we consider the case in which in-

tervals $[p_t^l, p_t^u]$ are specified, where $0 < p_t^l \leq p_t^u < 1,\ t = 1, \ldots, T$. We therefore have the adversarial min-max formulation

$$\mathcal{AP}: \quad \min_{\alpha \in \Delta} \Big(\max_{p \in \Delta \cap B} \ \sum_{t=1}^{T} \frac{p_t}{\alpha_t} \Big),$$

where $B = [p_1^l, p_1^u] \times \cdots \times [p_T^l, p_T^u]$ and where we assume that $\Delta \cap B \neq \emptyset$.

The following proposition is based on the theory of min-max optimization and saddle points; see, for example [5].

Theorem 2.4.1. *Consider problem $\mathcal{AP}$, a two-person zero-sum game involving a sampler who chooses α, and an adversary who chooses p. Then,*

(a) $\mathcal{AP}$ is equivalent to the problem

$$\max_{p \in \Delta \cap B} \Big(\min_{\alpha \in \Delta} \ \sum_{t=1}^{T} \frac{p_t}{\alpha_t} \Big) = \max_{p \in \Delta \cap B} \Big(\sum_{t=1}^{T} {p_t}^{1/2} \Big)^2, \tag{2.21}$$

i.e., the order of optimization is immaterial.

(b) The sampler has a unique optimal strategy no matter who plays first. The adversary has a unique optimal solution when he plays first, but may have multiple optimal strategies when the sampler plays first.

(c) The uniform probability vector $p_{\text{unif}} = (1/T, \cdots, 1/T)$ lies in B and is optimal for the adversary if and only if α_{unif} is optimal for the sampler. Further, p_{unif} is optimal for the adversary if and only if all feasible solutions $p \in \Delta \cap B$ are optimal for the adversary when he plays second.

Proof. (a) The function $w(\cdot; p)$ is convex and closed for each $p \in \Delta \cap B$, and $w(\alpha; \cdot)$ is concave and closed for each $\alpha \in \Delta$. These facts in conjunction

with the compactness of Δ and B imply (for example, see [5] - Prop. 2.6.4) that the minmax equality holds, i.e.,

$$\min_{\alpha\in\Delta}\Big(\max_{p\in\Delta\cap B}\ \sum_{t=1}^{T}\frac{p_t}{\alpha_t}\Big)\ =\ \max_{p\in\Delta\cap B}\Big(\min_{\alpha\in\Delta}\ \sum_{t=1}^{T}\frac{p_t}{\alpha_t}\Big).$$

From (2.19), the right-hand side of (2.21) follows.

(b) Strict concavity of $\sum_{t=1}^{T} p_t^{1/2}$ along with convexity and compactness of $\Delta \cap B$ imply that a unique optimal solution, p^*, exists for the problem on the right-hand side of (2.21). This shows that the adversary has a unique strategy when he plays first. Further, the inner minimization on the left-hand side of (2.21) is simply (2.19) with a unique optimal solution, α^*. Hence, the sampler has a unique optimal solution as the second player.

Consider the inner maximization over p in $\mathcal{AP}$. This is a simply a linear program subject to a convexity constraint and upper and lower bounds on the variables. Since the lower bounds are all nonzero, using translation, the objective of the linear program can be rewritten as $\sum_{t=1}^{T} p_t^l/\alpha_t + \sum_{t=1}^{T} p_t'/\alpha_t$, where $p_t' = p_t - p_t^l$. Thus, the outer minimization over α has an objective function that is the sum of a strictly convex function and a convex function, and is therefore strictly convex. Hence, there exists a unique optimal solution, say $\hat{\alpha}$, to the minimization problem in $\mathcal{AP}$, i.e., the sampler has a unique optimal strategy, $\hat{\alpha}$, as the first player.

It now follows that $(\hat{\alpha}, p^*)$ is a saddle point for $w(\cdot,\cdot)$. This implies that $\hat{\alpha} \in \operatorname{argmin}_{\alpha\in\Delta}\ w(\alpha, p^*)$. By uniqueness of the sampler's strategy as

the second player, we have $\hat{\alpha} = \alpha^*$. The sampler's strategy is therefore identical, regardless of the order of play.

Now, suppose $\alpha^*_{t_1} = \alpha^*_{t_2}$ for some $t_1 \neq t_2$. Then, $p^*_{t_1} = p^*_{t_2}$, where p^* is again the unique optimal solution when the adversary plays first. Any vector $q \in \Delta \cap B$ such that $q_t = p^*_t, t \notin \{t_1, t_2\}$ and $q_{t_1} + q_{t_2} = p^*_{t_1} + p^*_{t_2}$, is also optimal for the adversary after the sampler has chosen α^*. Thus, in the event of a tie in the components of α^*, there may exist multiple optimal strategies for the adversary as the second player.

(c) We have

$$\begin{aligned} \max_{p \in \Delta \cap B} \; w(\alpha_{\text{unif}}, p) &= w(\alpha_{\text{unif}}, p_{\text{unif}}) \\ &\leq w(\alpha, p_{\text{unif}}), \end{aligned}$$

where the inequality follows from our optimal allocation results in Section 2.3.2. If $p_{\text{unif}} \in \Delta \cap B$, then we also have $w(\alpha, p_{\text{unif}}) \leq \max_{p \in \Delta \cap B} \; w(\alpha, p)$. Hence,

$$\max_{p \in \Delta \cap B} \; w(\alpha_{\text{unif}}, p) \leq \max_{p \in \Delta \cap B} \; w(\alpha, p),$$

which establishes that α_{unif} is optimal for the sampler. Conversely, if α_{unif} is optimal for the sampler, then the uniqueness of the sampler's strategy, together with (a), implies that $p_{\text{unif}} \in \Delta \cap B$ and is optimal for the adversary. The second statement in (c) follows from the above inequality.

□

Remark. Theorem 2.4.1 shows that the adversary has no advantage in playing second, i.e., we can solve the easier problem of optimal allocation after the adversary chooses p. It also shows that if the adversary can choose the uniform distribution, stratification is of no value to the sampler, at least in the sense of (2.6), for large sample sizes.

The concavity of the objective function on the right-hand side of (2.21) suggests an algorithm based on a "water-filling" type of argument. The algorithm, given below, initializes all p_t values at their lower bounds, p_t^l, then increases the values of the smallest p_t (increasing multiple components simultaneously if there is a tie) until the condition $\sum_{t=1}^T p_t = 1$ is satisfied, subject to the upper bounds, p_t^u.

Algorithm 1

Input: $p_t^l,\ p_t^u,\ t = 1, \ldots, T$.

Output: Optimal solution p^*.

Initial Step. Set $p_t = p_t^l,\ t = 1, \ldots, T,\ \ s = \sum_{t=1}^T p_t$.

If $s > 1$, the problem is infeasible.

If $s = 1$, set $p_t^* = p_t,\ t = 1, \ldots, T$, and terminate.

Iterative Step. Set $S = \text{argmin}_{1 \le t \le T}\, p_t$. Simultaneously increase p_t for $t \in S$, updating s until one or more of the following happens:

(i) $s = 1$, upon which set $p_t^* = p_t,\ t = 1, \ldots, T$, and terminate.

(ii) The set S increases in cardinality, i.e., the value of p_t for $t \in S$ coincides with $p^l_{t'}$ for some $t' \notin S$. Let $S = S \cup \{t' \mid t' \notin S,\, p^l_{t'} = p_t \,\forall\, t \in S\}$.

(iii) The set S decreases in cardinality, i.e., the value of p_t coincides with p^u_t for some $t \in S$. Let $S' = \{t \in S \mid p_t = p^u_t\}$. Let $p^*_t = p_t,\ t \in S'$ and $S = S \setminus S'$. Continue.

2.5 Multiple Disruptions

In this section, we allow multiple disruptions instead of just one, and show that the results of Section 2.2 extend to two or more disruptions. We derive bounds on the expected probability of violation when there are ℓ disruptions, where ℓ is small compared to T. We establish, analogous to the single disruption model, that stratified sampling yields a bound that is $\Theta(1/NT^{\ell})$, while naive sampling yields a bound that is $\Theta(1/NT^{\ell-1})$. Thus, when the number of disruptions is small when compared to the horizon T, stratification reduces the sample size requirement by an order of magnitude in T. For simplicity of presentation, our results below are for the case of two disruptions. Extensions to the case where there are more than two such events are straightforward.

2.5.1 A Two-Disruption Model

The model for two disruptions is identical to that of the $\mathcal{BP}$ model (2.1), except that instead of assuming $\sum_{t=1}^{T} \tilde{I}_t = 1$, we now assume $\sum_{t=1}^{T} \tilde{I}_t = 2$. The random vector $\tilde{\xi}$ has the form $\tilde{\xi} = (\tilde{I}, \tilde{\delta})$, with $\tilde{I} = (\tilde{I}_1, \ldots, \tilde{I}_T)$ and

$\tilde{\delta} = (\tilde{\delta}_1, \ldots, \tilde{\delta}_T)$. We define

$$S = \{(t, s) \mid t = 1, \ldots, T-1,\ s = t+1, \ldots, T\},$$

and let $p_{t,s} = \mathbb{P}(\tilde{I}_t = 1, \tilde{I}_s = 1)$, $(t, s) \in S$, where $\sum_{(t,s)\in S} p_{t,s} = 1$. And, given that $\tilde{I}_t = \tilde{I}_s = 1$, $\tilde{\delta}_t$ and $\tilde{\delta}_s$ denote respectively, the magnitude of the disruptions in periods t and s, for $(s, t) \in S$. In the single-disruption model, we allow for the possibility that no disruption occurs by having period T as an artificial time period. The same idea allows the two-disruption model to capture having at most two disruptions. To do so, we simply introduce two artificial time periods, to capture the cases in which there is exactly one disruption and no disruptions. We allow the vectors $\tilde{\delta}_t$ and $\tilde{\delta}_s$ to be dependent.

In place of Assumption 1, we instead place the following structure on $f(x, \tilde{\xi})$:

Assumption 2. The function f separates via

$$f(x, \tilde{\xi}) = \sum_{(t,s)\in S} \tilde{I}_t \tilde{I}_s \max\{\, f_t(x_t, \tilde{\delta}_t), f_s(x_s, \tilde{\delta}_s)\,\}.$$

We assume we can draw i.i.d. observations from the distribution of $\tilde{\xi}$ and the joint distribution of $(\tilde{\delta}_t, \tilde{\delta}_s)$, $(t, s) \in S$. When we sample $\tilde{\xi}^j = (\tilde{I}^j, \tilde{\delta}^j)$, $j = 1, \ldots, M$, under the single-disruption assumption, the j^{th} constraint reduces to $f_t(x_t, \tilde{\delta}_t^j) \leq 0$, where t is the single period indicated by $\tilde{I}^j$. Under the two-disruption assumption, that constraint becomes $\max(f_t(x_t, \tilde{\delta}_t^j), f_s(x_s, \tilde{\delta}_s^j)) \leq 0$, where t and s are the two periods of disruption indicated by $\tilde{I}^j$. Of course, we remove the "max" by enforcing $f_t(x_t, \tilde{\delta}_t^j) \leq 0$ and $f_s(x_s, \tilde{\delta}_s^j) \leq 0$. Theorem

2.2.1 of Calafiore and Campi applies directly in the two-disruption setting, i.e., by solving $\mathcal{RP}^M$, we obtain $\tilde{x}^M$ which satisfies the bound in (2.5).

The following result generalizes Theorem 2.3.1. The proof is along the same lines and is omitted.

Theorem 2.5.1. *Let* $\mathbf{N} = \big((N_{t,s})_{(t,s)\in S}\big)$ *be a vector of non-negative integers satisfying* $\sum_{(t,s)\in S} N_{t,s} = M$. *And, let* $(\tilde{\delta}_t^1, \tilde{\delta}_s^1), \ldots, (\tilde{\delta}_t^{N_{t,s}}, \tilde{\delta}_s^{N_{t,s}})$, *i.i.d. from the joint distribution of* $(\tilde{\delta}_t, \tilde{\delta}_s)$, $(t,s) \in S$, *be used to define the convex program*

$$\begin{aligned}
\mathcal{RP}^{\mathbf{N}}: \quad \min_{x} \quad & \sum_{t=1}^{T} h_t(x_t) \\
\text{s.t.} \quad & \sum_{t=1}^{T} g_t(x_t) \le b, \\
& \max(f_t(x_t, \tilde{\delta}_t^j), f_s(x_s, \tilde{\delta}_s^j)) \le 0,\ (t,s) \in S,\ j = 1, \ldots, N_{t,s}, \\
& x_t \in D_t,\ t = 1, \ldots, T.
\end{aligned}$$

Assume that $\mathcal{RP}^{\mathbf{N}}$ *has a unique optimal solution* $\tilde{x}^{\mathbf{N}}$, *or that the two-norm tie-breaking rule is applied to obtain* $\tilde{x}^{\mathbf{N}}$. *Then,*

$$\mathbb{E}\Big[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0\big)\Big] \le \sum_{(t,s)\in S} p_{t,s} \frac{\dim(x_t) + \dim(x_s)}{N_{t,s} + 1}, \tag{2.22}$$

where the expectation is with respect to the augmented random vector whose $(t,s)^{\text{th}}$ *component is* $(\tilde{\delta}_{t,s}^1, \ldots, \tilde{\delta}_{t,s}^{N_{t,s}})$, $(t,s) \in S$.

2.5.2 Optimal Allocation

We carry out the analysis for optimal sampling strategies analogous to that of Section 2.3.2. We have $|S| = T(T-1)/2$ and the role of p in the single-disruption case is played here by the vector $p = (p_{t,s})_{(t,s)\in S}$. Hence, we consider

a budget of $M = NT(T-1)/2$ for constraint sampling in the two-disruption setting. All decision vectors x_t are assumed to have dimension n. The optimal allocation and optimal expected probability of violation are

$$\begin{aligned} N^*_{t,s} &= \frac{(p_{t,s})^{1/2}}{\sum_{(t,s)\in s} (p_{t,s})^{1/2}} (N+1)\Big(\frac{T(T-1)}{2}\Big) - 1, \ (t,s) \in S, \\ V^*(N) &= \frac{2n\Big(\sum_{(t,s)\in S} (p_{t,s})^{1/2}\Big)^2}{(N+1)\Big(\frac{T(T-1)}{2}\Big)}. \end{aligned} \tag{2.23}$$

We observe that with $M = NT(T-1)/2$ and $\dim(x_t) = n$ for all decision vectors x_t, the bound (2.5) decays as $N^{-1}T^{-1}$. When p has a nested geometric distribution, i.e., the waiting time for the first disruption is geometric, and the waiting time for the second disruption given the time of the first is also geometric, it can be established that $V^*(N)$ decays as $N^{-1}T^{-2}$. Thus, we see that there can be an improvement of an order of magnitude in T with stratified sampling. In the case of ℓ disruptions where ℓ/T is $\Theta(1/T)$ as $T \to \infty$, bound (2.5) decays as $N^{-1}T^{-(\ell-1)}$, and when p has a nested geometric distribution, $V^*(N)$ decays as $\ell N^{-1}T^{-\ell}$. Thus, the improvement is again of an order of magnitude in T. We prove this below for the two-disruption case.

Theorem 2.5.2. *Assume* $\dim(x_t) = n, \ t = 1, \ldots, T$. *Let* $V^1(N)$ *denote the bound (2.5) on the expected probability of violation with* $M = NT(T-1)/2$.

And, let $V^*(N)$ denote the bound (2.23). Then,

$$(i) \quad \frac{V^*(N)}{V^1(N)} = \Theta\left(\frac{\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2}{T}\right) \quad \text{for large } T, \tag{2.24a}$$

$$(ii) \quad \frac{2N}{T(N+1)} + \Theta(1/T^3) \leq \frac{V^*(N)}{V^1(N)} \leq \frac{N(T-1)}{N+1} + \Theta(1/T). \tag{2.24b}$$

Furthermore, if $p_{t,s} = 2/T(T-1)$, $(t,s) \in S$, the upper bound in (2.24b) is tight; and if p is a nested geometric distribution, then $V^*(N)/V^1(N) = \Theta(1/T)$ as $T \to \infty$.

Proof. We have

$$V^1(N) = \frac{nT}{NT(T-1)/2+1}. \tag{2.25}$$

from which, using (2.23), we have

$$\frac{V^*(N)}{V^1(N)} = \frac{2(NT(T-1)+2)\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2}{(N+1)T^2(T-1)}. \tag{2.26}$$

The equality in (2.24a) now follows, as does (2.24b). Tightness of the upper bound is obtained by setting p to the discrete uniform distribution. To prove the final result, let $0 < q, w < 1$ and

$$p_{t,s} = \frac{(1-q)q^{t-1}}{1-q^{T-1}} \cdot \frac{(1-w)w^{s-t-1}}{1-w^{T-t}},$$

i.e., the components of p are probability masses corresponding to a "nested" geometric distribution. Letting $u = 1-w$ and $p = 1-q$ for notational

convenience, we have

$$\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2 = \left(\sum_{(t,s)\in S} \frac{u^{1/2}p^{1/2}w^{(s-t-1)/2}q^{(t-1)/2}}{(1-w^{T-t})^{1/2}(1-q^{T-1})^{1/2}}\right)^2$$
$$= \frac{u\,p}{(1-q^{T-1})}\left(\sum_{(t,s)\in S} \frac{w^{(s-t-1)/2}q^{(t-1)/2}}{(1-w^{T-t})^{1/2}}\right)^2.$$

After some simplification, we get

$$\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2 = \frac{(1-w)(1-q)}{(1-q^{T-1})(1-w^{1/2})^2}\left(\sum_{t=1}^{T-1} \frac{q^{(t-1)/2}(1-w^{(T-t)/2})}{(1-w^{T-t})^{1/2}}\right)^2. \quad (2.27)$$

The term on the right hand side of (2.27) is $\Theta(1)$ as $T \to \infty$, which completes the proof. □

2.6 Concluding Remarks

In this chapter, we consider convex approximations of a multiperiod bicriteria minimization model with cost and risk as objectives. Towards constructing the efficient frontier for our model, we resort to constraint sampling to ensure that the risk is low. Our model incorporates random disruptions and is endowed with special structure that suggests a stratification strategy in sampling. We show that optimal stratification can provide improvements in sampling cost up to an order of magnitude in the number of time periods over a naive strategy when the number of disruptions is small. We also illustrate using an example from perishable inventory theory that stratification produces better proximity

to the efficient frontier on average. We attribute this to greater tightness of the bound on the expected probability of violation with stratification than naive sampling. We also pursue order-optimal stratification strategies, motivated by the fact that in practice, waiting time distributions are often unknown. Assuming that the unknown distribution is "well-behaved," we demonstrate that polynomially decaying allocations yield order-wise improvement identical to optimal allocation.

While the advantage of the approach used here is that the convex approximation is tractable, it suffers from the drawback that there is no guarantee, even probabilistic, of solution quality; in other words, there is no guarantee regarding proximity to the efficient frontier. We address this in the following chapter, via a numerical study, and show that sample average approximations can provide confidence intervals on the distance to the efficient frontier, that are reasonably tight.

Chapter 3

Estimating the Efficient Frontier of a Probabilistic Bicriteria Model

3.1 Introduction

In Chapter 2, we emphasize a threshold on risk in forming a sampling approximation to the EF. There, we are conservative with respect to risk, which enables the use of a tractable optimization model in place of the, possibly nonconvex, probabilistic program associated with the bicriteria model. In many situations, the coarseness of this approximation may be unattractive, especially in cases where cost is highly sensitive to perturbations of the risk level. With the constraint sampling approach outlined in Chapter 2, it is difficult to estimate the sample size needed to exactly match the risk level; indeed, we do not know how loose bound (2.5) or (2.6) is. The risk measure in the model therefore needs to be evaluated exactly or near-exactly. Unless the probability distribution governing the randomness, along with the functions associated with the bad event, have a special form, we cannot expect to evaluate precisely the risk measure, particularly in the multivariate setting. Instead, we replace the risk measure with a Monte Carlo estimator, and use this estimator in our

bicriteria optimization model. Our focus in this chapter is on estimating the EF using a sampling-based approximation, and assessing the closeness of that frontier to the true EF.

One approach to assessing the quality of a sampling-based approximation to a stochastic program is to form a confidence interval on the optimality gap of a candidate solution [4, 29]. This type of approach has been applied when minimizing risk subject to a cost constraint [30, 31], but unfortunately it does not apply when minimizing cost subject to a probabilistic constraint on risk. An issue in the latter setting is that the probabilistically-constrained program may be infeasible but its sampling-based approximation feasible or vice versa. So, an alternate approach has been developed to deal with this subtlety [27, 38]. As we show in this chapter, the simpler approach of [4] can be applied to our bicriteria model when assessing the closeness of the approximate EF to the true frontier. Moreover, we can form the approximate EF by solving a sequence of sampling-based problems in which: (i) we minimize risk over a range of cost thresholds or (ii) we minimize cost over a range of risk thresholds. Even though analysis of the latter problem (for a fixed risk threshold) is more difficult, we show that we can employ either sampling-based approximation when constructing, and assessing the quality of, the approximate EF.

3.2 Towards Solving a Probabilistic Bicriteria Model

As seen in Chapter 1, a bicriteria model takes the form:

$$\underset{x \in X}{\text{vmin}} \quad [h(x), p(x)]\,. \tag{3.1}$$

For instance, consider the following facility-sizing problem: There are m facilities at which nonnegative capacities $x_i,\ i = 1, \ldots, m$, are to be installed. The random demand at facility i is denoted $\tilde{\xi}_i$, and the joint distribution of the random vector $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_m)$ is assumed to be known. A realization $\xi = (\xi_1, \ldots, \xi_m)$ of the demand is said to be satisfied by the decision vector x if $x_i \geq \xi_i$, for all $i = 1, \ldots, m$. Our goal comprises two objectives: keeping the cost of installation as well as the probability of violating demand low.

More generally, let $X \subset \mathbb{R}^d$ constrain system design, let $\Xi \subset \mathbb{R}^m$ denote the support of $\tilde{\xi}$, and let $h : X \mapsto \mathbb{R}$ denote cost. Let $G : X \times \Xi \mapsto \mathbb{R}^\ell$, $C \subset \mathbb{R}^\ell$ and let $p(x) = \mathbb{P}(G(x, \tilde{\xi}) \notin C)$ denote risk. Throughout, we assume $h(\cdot)$ and $p(\cdot)$ are lower semicontinuous, and that X is compact. These constructs define model (3.1), whose solution we approach via parametric families of single-objective optimization models.

By enforcing a risk threshold, we obtain the family of probabilistically-constrained programs:

$$\begin{aligned} z_{\mathrm{c}}^*(\epsilon) \ = \ \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & \mathbb{P}(G(x, \tilde{\xi}) \notin C) \leq \epsilon, \end{aligned} \tag{3.2}$$

where ϵ is a *risk-level* parameter lying in a range $[\,\underline{\epsilon}\,, \bar{\epsilon}\,] \subseteq [0, 1]$ of interest. Under fairly mild conditions—lower semicontinuity of h plus $X \cap \{x \mid \mathbb{P}(G(x, \tilde{\xi}) \notin$

$C) \leq \epsilon\}$ being nonempty and compact—there exists an optimal solution to model (3.2) that is efficient. Conversely, if x is efficient for model (3.1), then x is optimal for model (3.2) when $\epsilon = p(x)$. In general, there may be optimal solutions to (3.2) that are not efficient. This can occur when (3.2) has multiple optimal solutions that have distinct values of $p(\cdot)$.

In such cases, an additional step is required to minimize $p(x)$ subject to the cost being equal to $z_{\mathrm{c}}^*(\epsilon)$. However, we assume the probabilistic constraint in (3.2) is tight at all optimal solutions for all $\epsilon \in [\,\underline{\epsilon}\,, \bar{\epsilon}\,]$, eliminating the need for such postprocessing. So, in principle we can solve model (3.1) over the range of risk tolerances of interest by solving model (3.2) for all $\epsilon \in [\,\underline{\epsilon}\,, \bar{\epsilon}\,]$.

We can swap the objective function and constraint in the approach sketched above, and alternatively consider the parametric family of programs constrained by cost:

$$\begin{aligned} z_{\mathrm{p}}^*(t) \;=\; \min_{x \in X} \quad & \mathbb{P}(G(x, \tilde{\xi}) \notin C) \\ \text{s.t.} \quad & h(x) \leq t. \end{aligned} \tag{3.3}$$

Analogous to our initial approach, we assume here that a range, $[\,\underline{t}\,, \bar{t}\,]$, of cost thresholds of interest is known. (For consistency we should have $z_{\mathrm{p}}^*(\bar{t}) = \underline{\epsilon}$ and $z_{\mathrm{p}}^*(\underline{t}) = \bar{\epsilon}$.) As for (3.3), ranging $t \in [\,\underline{t}\,, \bar{t}\,]$ and solving model (3.3) yields efficient solutions, and the EF, provided the budget constraint in (3.3) is tight at all optimal solutions. Under the tightness assumption, an obvious correspondence exists between the two models: If $(h(x), p(x))$ is on the EF, it can be obtained by solving either model (3.2) with $\epsilon = p(x)$ or model (3.3) with $t = h(x)$.

When $\tilde{\xi}$ is multivariate, and the probability distribution and the event $\{G(x,\tilde{\xi}) \notin C\}$ lack special form, solving either model (3.2) or (3.3) exactly is impossible. We therefore resort to Monte Carlo sampling to approximate the two models via *sample average approximation*. In other words, we draw n independent and identically distributed (i.i.d.) observations from the distribution of $\tilde{\xi}$, denoted $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$, and approximate model (3.2) by

$$\begin{aligned} z_{\mathrm{c}}^*(n,\epsilon) \;=\; & \min_{x\in X} \quad h(x) && \text{(3.4a)} \\ & \text{s.t.} \quad \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(G(x,\tilde{\xi}^j) \notin C) \leq \epsilon, && \text{(3.4b)} \end{aligned}$$

and model (3.3) by

$$\begin{aligned} z_{\mathrm{p}}^*(n,t) \;=\; & \min_{x\in X} \quad \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(G(x,\tilde{\xi}^j) \notin C) && \text{(3.5a)} \\ & \text{s.t.} \quad h(x) \leq t, && \text{(3.5b)} \end{aligned}$$

where the indicator function $\mathbb{I}(\cdot)$ takes value one if its argument is true and zero otherwise. We assume G and C are such that the function $\mathbb{I}(G(\cdot,\tilde{\xi}) \notin C)$ is lower semicontinuous on X, w.p.1, (See, e.g., [38] for conditions under which lower semicontinuity is ensured. Moreover, under the hypotheses of Theorem 3.5.1 in a subsequent section, lower semicontinuity of $p(x)$ is ensured.) With the measure thus discretized, models (3.4) and (3.5) can be reformulated as mixed integer programs. This said, differences between the two models are evident. It is enough to solve model (3.4) for the subset of values of $\epsilon \in [\,\underline{\epsilon}\,, \bar{\epsilon}\,]$ such that ϵn is an integer. In model (3.5) however, the values of $t \in [\,\underline{t}\,, \bar{t}\,]$ for which constraint (3.5b) will be tight at an optimal solution are not known

a priori. On the other hand, the objective function of model (3.5) can only take on values $0, 1/n, 2/n, \ldots$ and that fact can be exploited in specifying the termination criterion for a branch-and-bound algorithm. Both models allow for tightening their linear-programming relaxations by means of valid inequalities and preprocessing and are amenable to solution via special-purpose branching schemes; see [28, 33, 34].

3.3 Related Work

Bicriteria risk models have been the subject of recent study by a number of authors. Most studies have been efforts to solve probabilistically-constrained programs, due to their general analytical intractability. Pagnoncelli et al. [38] consider a chance-constrained sampling approximation to a portfolio allocation model, along the lines of (3.4). Under the assumption of a normal and lognormal distribution on $\tilde{\xi}$, they reduce the true problem (3.2) to a form that lends itself to linear programming based methods after Monte Carlo sampling. Our work, on the other hand, deals with (3.2) in generality, performing the sampling prior to any assumptions on the distribution of $\tilde{\xi}$. Luedtke and Ahmed [27] derive upper and lower confidence bounds on the optimal value to (3.2) by using the approximation (3.4) with the risk level replaced by $\alpha \neq \epsilon$. Nemirovski and Shapiro [36] use a sampling method based on importance sampling to derive a tractable approximation to (3.2). Riis and Lodahl [46] use a type of L-shaped method to solve a mixed integer program formulation of (3.4), in the context of telecommunication networks. More recently, Ralphs

et al. [44] devise an efficient algorithm for solving for the EF of a bicriteria model when it has finite cardinality, using a variant of the weighted Tchebycheff model. Dellino et al. [10, 11] address the construction of the efficient frontier involving expected cost and variance of that cost from the simulation optimization perspective. Our work, while related to these studies, parallels most closely the work of [27], but while their approach forms one-sided confidence bounds on the optimal value to (3.2), we use the sampling approximation to form two-sided confidence bounds on the optimality gap associated with the optimal values to (3.3) and (3.5).

A note on the Tchebycheff approach is in order. In Chapter 2, we note that the Tchebycheff approach of Steuer has the ability to produce all points on the EF. A good illustration of the algorithm can be found in [21] wherein Greis et al. employ the approach; actual implementation discretizes the space of cost coefficients to obtain a sufficiently dense and representative subset of the EF. Our approach is similar in spirit, at least in the sense that it is only possible to vary ϵ or t over a finite subset during implementation. We guarantee convergence of the approximate EF to the true EF by means of a noninteractive procedure while the Tchebycheff approach instead aims to work interactively with the decision maker.

3.4 Estimating the Efficient Frontier: An Equivalence Result

If we could solve model (3.2) exactly for $\epsilon \in [\underline{\epsilon}, \overline{\epsilon}]$, we would obtain the corresponding points on $EF(h, p)$ via the pairs $(\epsilon, z_{\mathrm{c}}^*(\epsilon))$ for $\epsilon \in [\underline{\epsilon}, \overline{\epsilon}]$. Alternatively, if we could solve model (3.3) for $t \in [\underline{t}, \overline{t}]$, we would obtain $EF(h, p)$ via $(z_{\mathrm{p}}^*(t), t)$, $t \in [\underline{t}, \overline{t}]$. Figure 3.1 depicts such an EF for the facility-sizing problem sketched at the beginning of Section 3.2.

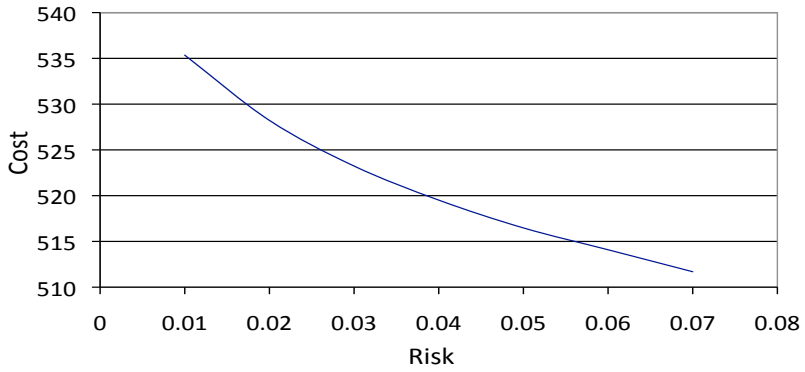

Figure 3.1: Efficient frontier for a capacity expansion problem

Let $\epsilon \in (\underline{\epsilon}, \overline{\epsilon})$, fix the i.i.d. sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ from the distribution of $\tilde{\xi}$ and consider model (3.4), where we may assume $\epsilon = r/n$ for some positive integer $r \leq n$. We assume that model (3.4) is feasible and that the expression $p_n(x) = n^{-1} \sum_{j=1}^{n} \mathbb{I}(G(x, \tilde{\xi}^j) \notin C)$ on the left-side of constraint (3.4b) is identical at all optimal solutions. By choosing r suitably, we may assume that (3.4b) is tight at all optimal solutions, which, in turn, implies that every optimal solution is on $EF(h, p_n)$.

In addition to $z_c^*(n, r/n)$ denoting the optimal value of model (3.4) with $\epsilon = r/n$, we use $X_c^*(n, r/n)$ to denote the set of optimal solutions to (3.4), i.e., $X_c^*(n, r/n)$ is the set of decisions x at which $h(x) = z_c^*(n, r/n)$ and $p_n(x) = r/n$.

Now, consider model (3.5) with $t = z_c^*(n, r/n)$, and for the *same sample*, $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$, used to define model (3.4). Every solution in $X_c^*(n, r/n)$ is feasible for model (3.5) and has the same objective function value of r/n. If model (3.5) were to have a feasible solution with an objective function value less than r/n, this would contradict our tightness assumption. Similarly, if model (3.5) has an optimal solution at which constraint (3.5b) is not tight, then the optimal objective function value for model (3.4) is less than $z_c^*(n, r/n)$, a contradiction. It follows that $X_c^*(n, r/n)$ is the set of optimal solutions, and r/n is the optimal value, for model (3.5).

The above argument establishes the following notion of equivalence between models (3.4) and (3.5): Varying ϵ in the set $E_n = \{\lfloor \underline{\epsilon} n \rfloor / n, (\lfloor \underline{\epsilon} n \rfloor + 1)/n, \ldots, \lfloor \bar{\epsilon} n \rfloor / n\}$ and solving model (3.4) is equivalent to varying t in the set $T_n = \{z_c^*(n, \epsilon) \mid \epsilon \in E_n\}$ and solving model (3.5). In what follows, we exploit this equivalence between models (3.4) and (3.5).

3.5 Asymptotic Analysis and Construction of Confidence Intervals

3.5.1 Asymptotics

We cannot construct $EF(h, p)$ exactly. Instead, we employ Monte Carlo approximations and hence we are interested in asymptotic results as the sample size n grows large for models (3.4) and (3.5). Both models are amenable to such analyses, but the analysis of model (3.5) is arguably simpler. When randomness appears in the constraints, we are unsure whether a solution to a Monte Carlo approximation will be feasible to the original problem. In the models we consider, feasibility is ensured for a solution from (3.5), and we can focus on whether such a solution is near-optimal. For these reasons, in this chapter, we restrict attention to asymptotic results for model (3.3), and refer the interested reader to [38] for asymptotic analysis of (3.4). We also use (3.4) in Chapter 4 for fast computation of a convex approximation to $EF(h, p_n)$.

The notion of epiconvergence of the approximating objective function to the true objective function, denoted $p_n(x) \overset{\text{epi}}{\to} p(x)$ in the context of models (3.3) and (3.5), is central to establishing convergence results. It is well-known (see [1], for example) that $p_n(x) \overset{\text{epi}}{\to} p(x)$, and in particular, on $\{x \in X \mid h(x) \leq t\}$ w.p.1. The importance of epiconvergence is evidenced in the following result.

Theorem 3.5.1. (Attouch [2]) *Suppose $f_n(x) \overset{\text{epi}}{\to} f(x)$ as $n \to \infty$. Then, we have $\limsup_{n\to\infty} \inf_x f_n(x) \leq \inf_x f(x)$. If $x_n \in \operatorname{argmin}_x f_n(x)$, $n = 1, \ldots,$ and x is a limit point of $\{x_n\}$, then $x \in \operatorname{argmin}_x f(x)$. Further, if there is a compact*

set C *such that* $\operatorname{argmin}_x f_n(x) \cap C \neq \emptyset$, $n = 1, \ldots,$ *then* $\inf_x f_n(x) \to \inf_x f(x)$ *as* $n \to \infty$.

To express the theorem in our setting, we set $f_n(x) = p_n(x)$ on $\{x \in X \mid h(x) \leq t\}$ and $f_n(x) = +\infty$ elsewhere (and similarly for $f(x)$ and $p(x)$), and set $C = \{x \in X \mid h(x) \leq t)\}$. Theorem 3.5.1 thus establishes consistency of minimizers to (3.5), and that the pseudo EF converges pointwise to $EF(h, p)$, both w.p.1. Theorem 3.5.1 asymptotically justifies approximating the EF via model (3.5) in two senses: (i) Solving (3.5) with sample size n and budget t, for which constraint (3.5b) is tight (for all n) yields a solution $x_{\mathrm{p}}^*(n, t)$. Theorem 3.5.1 shows that if $\hat{x}_{\mathrm{p}}(t)$ is a limit point of $\{x_{\mathrm{p}}^*(n, t)\}$ then $\hat{x}_{\mathrm{p}}(t)$ solves (3.3) and hence is efficient, i.e., $(p(\hat{x}_{\mathrm{p}}(t), h(\hat{x}_{\mathrm{p}}(t)),)$ is on $EF(h, p)$, with $h(\hat{x}_{\mathrm{p}}(t)) = t$, w.p.1; (ii) When solving (3.5), we obtain optimal value $z_{\mathrm{p}}^*(n, t)$ and Theorem 3.5.1 states that $z_{\mathrm{p}}^*(n, t)$ converges to $z_{\mathrm{p}}^*(t)$, w.p.1., where $(z_{\mathrm{p}}^*(t), t)$ is on the EF. Result (i) is of primary importance because of foremost concern is the decision we implement, but we see shortly the importance of (ii). Also, these results extend to the case where we replace t with $\{t_n\}$ that converges to t, and this is of interest for reasons sketched at the end of Section 3.4, where $t \in T_n$. When we solve model (3.5) for finite n and for a range of $t \in T_n$, we obtain $(p(x_{\mathrm{p}}^*(n, t), h(x_{\mathrm{p}}^*(n, t)))$ and $(z_{\mathrm{p}}^*(n, t), t)$. We refer to the former as the *approximate* EF and the latter as the *pseudo* EF. Next, we show how the pseudo and approximate EFs allow us to form a confidence interval (CI) about $EF(h, p)$. Finally, we know by the equivalence result of Section 3.4 that when we speak of solving model (3.5) for $t \in T_n$, we can

instead solve model (3.4) for $\epsilon \in E_n$.

3.5.2 Confidence Intervals

The asymptotic results of Theorem 3.5.1 justify approximating efficient solutions and the EF via our Monte Carlo approach, at least in the limit as n grows to infinity. However, this result provides little insight regarding the proximity of the approximate (and pseudo) EF to the true EF when we solve a family of Monte Carlo approximations with n finite. For a fixed t and fixed n, if the optimality gap of model (3.3), $p(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(t)$, is small, then $x_{\mathrm{p}}^*(n,t)$ is near efficient. To estimate this gap, we first estimate $p(x_{\mathrm{p}}^*(n,t))$ via the sample mean

$$p_{n'}(x_{\mathrm{p}}^*(n,t)) = (n')^{-1} \sum_{j=1}^{n'} \mathbb{I}(G(x_{\mathrm{p}}^*(n,t), \tilde{\xi}^j) \notin C).$$

Here, the sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^{n'}$ is i.i.d. from the distribution of $\tilde{\xi}$, and is independent of that used to find $x_{\mathrm{p}}^*(n,t)$. Moreover, we can choose $n' \gg n$ because forming $p_{n'}(x_{\mathrm{p}}^*(n,t))$, given $x_{\mathrm{p}}^*(n,t)$, simply requires function evaluations of G and the associated feasibility checks, i.e., it does not involve solving a mixed-integer program, as is required to form $x_{\mathrm{p}}^*(n,t)$. We use $z_{\mathrm{p}}^*(n,t)$ to estimate $z_{\mathrm{p}}^*(t)$. In addition to Theorem 3.5.1's asymptotic result, we know $\mathbb{E}[z_{\mathrm{p}}^*(n,t)] \leq z_{\mathrm{p}}^*(t)$ [29, 37]. So, our point estimate of the optimality gap is $[p_{n'}(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(n,t)]^+ \geq p_{n'}(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(n,t)$, with $\mathbb{E}[p_{n'}(x_{\mathrm{p}}^*(n,t))] - \mathbb{E}[z_{\mathrm{p}}^*(n,t)] \geq p(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(t)$.

We use this estimator to construct a one-sided CI, providing an upper

bound on the optimality gap. Towards this end, we write x_n^* for $x_{\rm p}^*(n,t)$, z_n^* for $z_{\rm p}^*(n,t)$ and z^* for $z_{\rm p}^*(t)$ for ease of notation. We have from the standard central limit theorem (CLT) for the sample mean of i.i.d. random variables that, as $n' \to \infty$,

$$\sqrt{n'}\left[p_{n'}(x_n^*) - p(x_n^*)\right] \;\Rightarrow\; \sigma N(0,1), \tag{3.6}$$

where $\sigma^2 = p(x_n^*)(1-p(x_n^*))$, $N(0,1)$ is a standard normal random variable and "$\Rightarrow$" denotes convergence in distribution. By the strong law of large numbers, $\lim_{n'\to\infty} p_{n'}(x_n^*) \to p(x_n^*)$, w.p.1, and so, by Slutsky's theorem, we can estimate the population variance and obtain

$$\lim_{n'\to\infty} \mathbb{P}\left(p(x_n^*) \le p_{n'}(x_n^*) + \underbrace{z_{\alpha/2}\,\left[p_{n'}(x_n^*)(1-p_{n'}(x_n^*))/n'\right]^{1/2}}_{\epsilon_u} \right) = 1 - \frac{\alpha}{2}, \tag{3.7}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.

Next, consider the limiting distribution of $n^{1/2}\,(z_n^* - z^*)$. Let x^* be an optimal solution to (3.3). Using the fact that x^* is feasible, but suboptimal, for model (3.5), we have $z_n^* \le p_n(x^*)$, where the same $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ are used to define these two estimators. Hence,

$$\begin{aligned} &\mathbb{P}\left(z^* \ge z_n^* - z_{\alpha/2}\left[z_n^*(1-z_n^*)/n\right]^{1/2} \right) \\ &\qquad \ge \mathbb{P}\left(z^* \ge p_n(x^*) - z_{\alpha/2}\left[z_n^*(1-z_n^*)/n\right]^{1/2} \right). \end{aligned} \tag{3.8}$$

Under the hypotheses of Theorem 3.5.1, $z_n^* \to z^*$, w.p.1, as $n \to \infty$. Again employing Slutsky's theorem and the standard CLT for the sample mean of

i.i.d. random variables, coupled with (3.8), we have

$$\lim_{n\to\infty} \mathbb{P}\left(z^* \geq z_n^* - \underbrace{z_{\alpha/2}\left[z_n^*(1-z_n^*)/n\right]^{1/2}}_{\epsilon_\ell} \right)$$
$$\geq \lim_{n\to\infty} \mathbb{P}\left(z^* \geq p_n(x^*) - z_{\alpha/2}\left[z_n^*(1-z_n^*)/n\right]^{1/2} \right) = 1 - \tfrac{\alpha}{2}. \quad (3.9)$$

We write "$= 1 - \frac{\alpha}{2}$" in both (3.7) and (3.9) and this is correct, provided $p(x_n^*) \in (0,1)$ and $z^* \in (0,1)$. Otherwise, we can replace "$=$" with "$\geq$" and have a valid statement. In any case, using the Boole-Bonferroni inequality we can infer from (3.7) and (3.9) that when n and n' are sufficiently large

$$\mathbb{P}\left(p(x_n^*) - z^* \leq [p_{n'}(x_n^*) - z_n^*]^+ + \epsilon_\ell + \epsilon_u\right) \gtrapprox 1 - \alpha, \quad (3.10)$$

where "$\gtrapprox$" is interpreted as "approximately greater than or equal to."

Summarizing in unabridged notation our procedure for forming a CI about the EF for a fixed t we have:

Input: Cost budget t, value α (e.g., $\alpha = 0.10$), sample size n for model (3.5) and sample size for upper bound n'.

Output: Solution $x_{\rm p}^*(n,t)$ with approximate $(1-\alpha)$ CI on its optimality gap.

Step 1. Sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ i.i.d. from the distribution of $\tilde{\xi}$. Form and solve (3.5) to obtain $x_{\rm p}^*(n,t)$ and $z_{\rm p}^*(n,t)$.

Step 2. Sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^{n'}$ i.i.d. from the distribution of $\tilde{\xi}$, and independent of that in step 1. Form

$$p_{n'}(x_{\rm p}^*(n,t)) = (n')^{-1} \sum_{j=1}^{n'} \mathbb{I}(G(x_{\rm p}^*(n,t), \tilde{\xi}^j) \notin C) \quad (3.11)$$

and

$$\begin{aligned}\epsilon_\ell &= z_{\alpha/2}\left[z^*_{\mathrm{p}}(n,t)(1-z^*_{\mathrm{p}}(n,t))/n\right]^{1/2}\\ \epsilon_u &= z_{\alpha/2}\left[p_{n'}(x^*_{\mathrm{p}}(n,t))(1-p_{n'}(x^*_{\mathrm{p}}(n,t)))/n'\right]^{1/2},\end{aligned} \tag{3.12}$$

and output $x^*_{\mathrm{p}}(n,t)$ and a one-sided CI on its optimality gap

$$[p_{n'}(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(n,t)]^+ + \epsilon_\ell + \epsilon_u.$$

Restating our main confidence interval result (3.10), in our unabridged notation, we have

$$\mathbb{P}\left(p(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(t) \le [p_{n'}(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(n,t)]^+ + \epsilon_\ell + \epsilon_u\right) \gtrapprox 1-\alpha. \tag{3.13}$$

We could use Student t quantiles in (3.12), but for the sample sizes we have in mind, they are practically identical to normal quantiles. We employ the above procedure in our computations to derive approximate confidence bounds on the proximity of our Monte Carlo approximation to the EF. Importantly, the effort to form the confidence interval, beyond solving the Monte Carlo approximation itself, is minimal. Specifically, the additional work simply involves forming estimator (3.11) and the sampling error estimates (3.12).

Finally, observe that by the equivalence result of Section 3.4, we can use model (3.4) to obtain $x^*_{\mathrm{p}}(n,t)$ and $z^*_{\mathrm{p}}(n,t)$ by varying $\epsilon \in E_n$. More importantly, the construction of the above confidence interval on the optimality gap $p(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(t)$ remains valid. This fact endows us with the flexibility to use either model in our computations.

3.6 Facility-Sizing Problem

We now return to the facility-sizing model sketched at the beginning of Section 3.2. The cost of installing capacities $x = (x_1, \ldots, x_m)$ is $h(x) = \sum_{i=1}^{m} c_i x_i$, the risk of failing to satisfy demand $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_m)$ is $p(x) = \mathbb{P}(\tilde{\xi} \not\leq x)$, and $X = \{x \mid x \geq 0\}$. We assume the unit costs coefficients c_i, $i = 1, \ldots, m$, are positive. So, our bicriteria model (3.1) specializes to

$$\operatorname*{vmin}_{x \geq 0} \quad \left[\sum_{i=1}^{m} c_i x_i,\ \mathbb{P}(\tilde{\xi} \not\leq x) \right].$$

Models (3.2) and (3.3) specialize to

$$\begin{aligned} z_{\mathrm{c}}^*(\epsilon) \ = \ & \min_{x \geq 0} \quad \sum_{i=1}^{m} c_i x_i \\ & \text{s.t.} \quad \mathbb{P}(\tilde{\xi} \not\leq x) \leq \epsilon, \end{aligned}$$

and

$$\begin{aligned} z_{\mathrm{p}}^*(t) \ = \ & \min_{x \geq 0} \quad \mathbb{P}(\tilde{\xi} \not\leq x) \\ & \text{s.t.} \quad \sum_{i=1}^{m} c_i x_i \leq t. \end{aligned}$$

The sampling-based approximations, models (3.4) and (3.5), become the following mixed-integer programs

$$\begin{aligned} z_{\mathrm{c}}^*(n, \epsilon) \ = & \min_{x,y} \quad \sum_{i=1}^{m} c_i x_i \\ & \text{s.t.} \quad \frac{1}{n} \sum_{j=1}^{n} y_j \leq \epsilon, \\ & \qquad x_i \geq \xi_i^j (1 - y_j), i = 1, \ldots, m, j = 1, \ldots, n, \\ & \qquad y_j \in \{0, 1\}, j = 1, \ldots, n, \end{aligned} \tag{3.14}$$

and

$$
\begin{aligned}
z_{\mathrm{p}}^*(n,t) \; = \min_{x,y} \quad & \frac{1}{n}\sum_{j=1}^{n} y_j \\
\text{s.t.} \quad & \sum_{i=1}^{m} c_i x_i \le t, \\
& x_i \ge \xi_i^j (1-y_j), i = 1,\ldots,m, j = 1,\ldots,n, \\
& y_j \in \{0,1\}, j = 1,\ldots,n,
\end{aligned}
\tag{3.15}
$$

where y_j indicates whether we satisfy ($y_j = 0$) or fail to satisfy ($y_j = 1$) demand realization $\xi^j = (\xi_1^j, \ldots, \xi_m^j)$ with facility sizes $x = (x_1, \ldots, x_m)$. We note that $p(x) = 1 - \mathbb{P}(\tilde{\xi} \le x)$ is a continuous function if the random vector $\tilde{\xi}$ has a continuous distribution. In this case, when using model (3.15) to approximate $EF(h,p)$, we have, in addition to the conclusions of Theorem 3.5.1, $\lim_{n\to\infty} p(x_{\mathrm{p}}^*(n,t)) = z_{\mathrm{p}}^*(t)$, w.p.1.

The concluding remarks of Section 3.4 show that models (3.4) and (3.5) have the same set of optimal solutions for pairs of $\epsilon \in E_n$ and $t \in T_n$. A key assumption in establishing this equivalence is that at any optimal solution solution to model (3.4) with $\epsilon \in E_n$, constraint (3.4b) is tight. We now show that this holds for the facility-sizing model (3.14), provided $\tilde{\xi}$ has a continuous distribution.

Theorem 3.6.1. *Consider model (3.14) with $c_i > 0$, $i = 1,\ldots,m$, and demand realizations $\xi^j = (\xi_1^1, \ldots, \xi_m^j)$, $j = 1,\ldots,n$. Assume that for each i, ξ_i^j, $j = 1,\ldots,n$, are distinct and positive. Let $\epsilon = r/n$, where $r \in \{0,1,\ldots,n\}$. Then, the constraint $1/n \sum_{j=1}^{n} y_j \le \epsilon$ is tight at every optimal solution to model (3.14).*

Proof. The claim is immediate if $r = 0$ and so assume $r \geq 1$. Let $J = \{1, \ldots, n\}$ index the samples, and for $x \geq 0$, let $J^+(x) = \{j \mid x_i \geq \xi_i^j, i = 1, \ldots, m\}$ index the demand realizations that are satisfied by x and let $J^-(x) = J \setminus J^+(x)$ index those x fails to satisfy. Let (x^*, y^*) be an optimal solution to (3.14). So, $y_j^* = 0$ for $j \in J^+(x^*)$, $y_j^* = 1$ for $j \in J^-(x^*)$, and $|J^-(x^*)| \leq r$. It is easily seen that $x_i^* = \max_{j \in J^+(x^*)} \xi_i^j$.

Constraint (4.7) being tight at x^* means $|J^-(x^*)| = r$. Suppose this does not hold, i.e., $|J^-(x^*)| < r$. This implies $|J^+(x^*)| > n - r$. Let $j(i) \in \operatorname{argmax}_{j \in J^+(x^*)} \xi_i^j$. Note that for each i, $j(i)$ is uniquely determined, given x^*, because ξ_i^j, $j = 1, \ldots, n$, are distinct. Now, we can select any i, say $i = 1$, and define

$$x' = \big(\max_{j \in J^+(x^*) \setminus j(1)} \xi_1^j, x_2^*, \ldots, x_m^* \big),$$

where $\max_{j \in \emptyset}(\cdot) \equiv 0$. Note $J^-(x') = J^-(x^*) \cup \{j(1)\}$, so that $|J^-(x')| = |J^-(x^*)|+1$, again because ξ_1^j, $j = 1, \ldots, n$ are distinct, So (x', y'), where $y_j' = 0$ for $j \in J^+(x')$ and $y_j^* = 1$ for $j \in J^-(x')$ is feasible to (3.14) and decreases the objective function by $c_1 (x_1^* - x_1') > 0$. This contradicts optimality of x^* and proves the claim. □

3.7 Computational Results

We consider the following instance of the facility-sizing model. There are $m = 40$ facilities with a per unit cost of installing capacity of $c_i = 1$ for each facility. The demand vector $\tilde{\xi}$ is assumed to be multivariate normal with all components having a mean 10, variance 1 and all pairwise correlation

coefficients equal to 0.8. The demand vector is truncated so that demand is bounded above. This implies that X may be assumed to be compact and the convergence results of Theorem 3.5.1 hold true.

While model (3.15), which is the specialization of model (3.5), provides the framework for the asymptotic analysis of Section 3.5, we observed computationally that model (3.14) solves more quickly than model (3.15), particularly for large values of n. The results of Section 3.4 and Section 3.5 justify employing either model and so we took advantage of model (3.14) in our computations. We use the tightened reformulation of this mixed integer program, along with valid inequalities, discussed in [34]. Finally, we use a sample size of $n' = 200,000$ to estimate $p(x^*_{\mathrm{p}}(n,t))$. We run the procedure outlined in Section 3.5 for different values of cost, and use the same set of n observations and the same independent set of n' observations for each instance.

Table 3.1 and Figure 3.2 show the values of $z^*_{\mathrm{p}}(n,t)$ and $p_{n'}(x^*_{\mathrm{p}}(n,t))$ for varying values of t. The numbers are indicative of a typical sample path of $(\tilde{\xi}^1, \tilde{\xi}^2, \ldots)$. Note that although the pseudo EF is a discontinuous step function, we plot it as a piecewise linear, continuous curve — this makes for better visualization of convergent behavior. For a given n, the trajectory of $(z^*_{\mathrm{p}}(n,t),t)$ lies on $EF(h,p_n)$ and shows the trade-off between the two objectives. For a fixed value of t, we see that $z^*_{\mathrm{p}}(n,t)$ increases with n. This trend is consistent with that suggested by $\mathbb{E}[z^*_{\mathrm{p}}(n,t)] \leq \mathbb{E}[z^*_{\mathrm{p}}(n+1,t)]$ [29, 37]. When n is small, the point estimate of the optimality gap estimate $[p_{n'}(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(n,t)]^+$ is relatively large, and this difference shrinks as n grows. Putting aside sam-

pling error for a moment, two factors contribute to this: The suboptimality of solution $x^*_{\mathrm{p}}(n,t)$ in model (3.15), $p(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(t)$, and the bias of the lower-bound estimator $z^*_{\mathrm{p}}(t) - \mathbb{E}[z^*_{\mathrm{p}}(n,t)]$. Of course, we do not know $z^*_{\mathrm{p}}(t)$ but Figure 3.3 shows how $p_{n'}(x^*_{\mathrm{p}}(n,t))$ tends to shrink and $z^*_{\mathrm{p}}(n,t)$ tends to grow with n for a specific value of t. This plot suggests that $z^*_{\mathrm{p}}(n,t)$ is growing at a faster rate than $p_{n'}(x^*_{\mathrm{p}}(n,t))$ is shrinking, and this indicates bias is likely the dominant contributor. This same effect can be seen in Figure 3.2, where there tends to be a larger gap between the pseudo EF plots than those of the approximate EF, as n grows. As indicated above, the pseudo EF tends to grow in n for fixed t, and because the same sample is used when n is fixed, it necessarily shrinks in t for fixed n. While a similar tendency is present, this type of monotonicity is not ensured for the approximate EF as this involves an "out-of-sample" assessment, $p_{n'}(x^*_{\mathrm{p}}(n,t))$.

Approximate 90% confidence bounds are reported in Table 3.2 for $n = 25,000$ and $n' = 200,000$. Sampling error associated with $p_{n'}(x^*_{\mathrm{p}}(n,t))$, ϵ_u, is about 0.001 while that for $z^*_{\mathrm{p}}(n,t)$, ϵ_ℓ, is between 0.002 and 0.003. For smaller values of n, the confidence interval on the optimality gap is dominated by the point estimate of the gap, $[p_{n'}(x^*_{\mathrm{p}}(n,t)) - z^*_{\mathrm{p}}(n,t)]^+$. However, for $n = 25,000$, i.e., for the values listed in Table 3.2 this point estimate of the optimality gap and the sampling error are of comparable magnitude.

Figure 3.3 shows the upper and lower confidence bounds for varying sample sizes and for a fixed cost. The size of the error term for the lower bound decreases proportional to $1/\sqrt{n}$, with the growing sample size n. The

Table 3.1: Values of $z_{\mathrm{p}}^*(n,t)$ and $p_{n'}(x_{\mathrm{p}}^*(n,t))$ for varying values of sample size n, and cost of installation t.

t	$n=500$		$n=1000$		$n=2000$		$n=5000$		$n=10,000$		$n=25,000$	
	L	U	L	U	L	U	L	U	L	U	L	U
486	0.062	0.111	0.072	0.108	0.081	0.106	0.089	0.102	0.095	0.100	0.096	0.098
488	0.052	0.101	0.063	0.100	0.072	0.095	0.081	0.094	0.086	0.092	0.089	0.091
490	0.044	0.095	0.056	0.090	0.064	0.087	0.073	0.085	0.078	0.083	0.082	0.083
492	0.038	0.086	0.050	0.082	0.057	0.079	0.065	0.077	0.069	0.075	0.073	0.074
494	0.034	0.082	0.044	0.078	0.051	0.074	0.058	0.070	0.062	0.068	0.066	0.067
496	0.028	0.075	0.038	0.070	0.045	0.068	0.051	0.063	0.055	0.062	0.059	0.060
498	0.024	0.070	0.033	0.063	0.040	0.059	0.046	0.058	0.050	0.056	0.053	0.055

L: $z_{\mathrm{p}}^*(n,t)$

U: $p_{n'}(x_{\mathrm{p}}^*(n,t))$

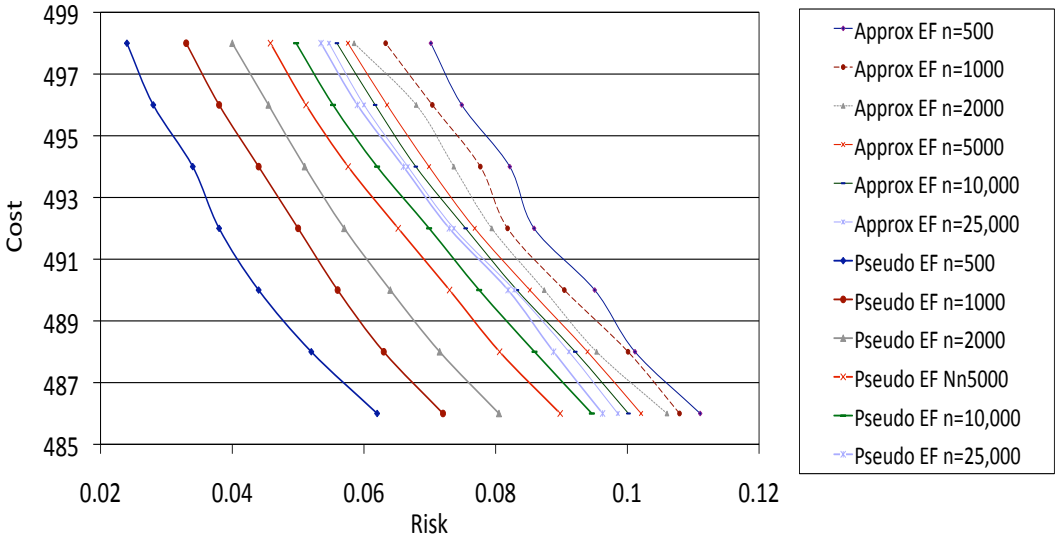

Figure 3.2: Approximate and pseudo efficient frontiers for varying sample sizes n.

Table 3.2: Confidence bounds for $p(x_{\mathrm{p}}^*(n,t))$, $z_{\mathrm{p}}^*(t)$ and the optimality gap $p(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(t)$, with $n = 25,000$.

t	$z_{\mathrm{p}}^*(n,t) - \epsilon_\ell$	$p_{n'}(x_{\mathrm{p}}^*(n,t)) + \epsilon_u$	Optimality Gap
486	0.093	0.099	0.006
488	0.086	0.092	0.006
490	0.079	0.084	0.005
492	0.070	0.075	0.005
494	0.063	0.068	0.005
496	0.057	0.061	0.004
498	0.051	0.055	0.004

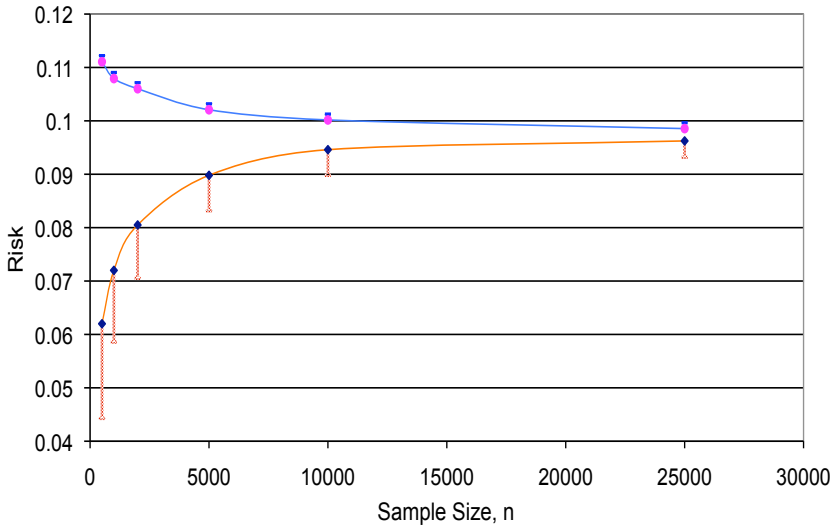

Figure 3.3: Point estimators and sampling error for $t = 486$. Lower line is $z_{\mathrm{p}}^*(n,t)$, upper line is $p_{n'}(x_{\mathrm{p}}^*(n,t))$, and the vertical line segments represent error bars, with $n' = 200,000$.

size of the error term for the upper bound is of course unaffected by increasing sample sizes and is much smaller than that for the lower bound due to the

large sample size n' used to evaluate $p_{n'}(x^*_{\mathrm{p}}(n,t))$.

Solve times in our experiments were fairly large for large sample sizes. Solving the mixed integer program (3.14) for 40 facilities and $n = 25,000$ realizations takes about 30 minutes. This is not surprising, considering we are solving a problem that is known to be hard, but there is clearly a need for a fast algorithm that perhaps does not produce the exact pseudo EF but approximates it well in the risk range of interest.

3.8 Concluding Remarks

Our studies in Chapters 1 and 2 provide interesting comparisons on the two approaches used therein. Constraint sampling induces an approximate EF that, on average, is much less optimal than the approximate EF produced by the sample average approximation approach employed in this chapter. Furthermore, sample average approximation offers a guarantee on the quality of the solution it produces, in the form of an upper bound on its optimality gap, that holds true with the degree of confidence specified by the decision maker. On the other hand, solving a chance-constrained program is not always desirable, as it may not be convex. Constraint sampling produces a convex program, which can be solved easily, at least for our model. Indeed, for the inventory model in Chapter 2, the sampled constraints reduce to simple lower bounds on the decision variables, thus transforming the model from a hard probabilistically-constrained program to a simple linear program.

In the following chapter, we consider problems in which the EF is convex, at least in the risk range of interest. We show how convexity ensures convergence of a sequence of convex approximations to the true EF, and also demonstrate computationally the working of a fast algorithm to form these approximations.

Chapter 4

A Fast Algorithm for Estimating the EF: Convergence and Computational Enhancements

4.1 Introduction

In many investment decisions, parameters that determine systemic constraints can range in such a broad array of values that solving for all possible settings of parameters is too expensive. A realistic approach is to solve for a meaningful subset of values, and use techniques such as interpolation to obtain operating decisions for the rest. In the facility sizing model, the risk threshold varies in a continuum $[\underline{\epsilon}, \overline{\epsilon}]$, where $0 < \underline{\epsilon} < \overline{\epsilon} < 1$. While solving for the true efficient frontier $EF(h, p)$ is often impossible, evaluating a sample approximation to it can still be expensive, such as solving a mixed integer program for all risk values that are jump points of the pseudo EF. A cheaper alternative can be to solve for the pseudo EF only at risk values that yield high return in an incremental sense. In other words, suppose we have obtained a point on the pseudo EF. Solving for the pseudo EF at risk values that yield the greatest

cost savings per additional violated scenario, in relation to the incumbent solution, provides value to the decision maker, provided there is a quick means of calculating these maximum relative benefit configurations. More importantly, we would like such a subset approximation to the pseudo EF to obey convergence properties near identical to the pseudo EF. This chapter shows how both these goals can be achieved, under the assumption that the true EF is convex. Furthermore, successive points on the subset approximation are nested, in the following sense: Capacity installed to hedge against a given risk level need only be augmented at each facility when the risk is decreased to another value in the subset. This can be a highly desirable feature when risk thresholds are subject to perturbation, and installing capacity is an expensive undertaking.

We begin by reviewing that the pseudo EF for the facility sizing model is the set of all pairs $(\epsilon, z_{\mathrm{c}}^*(n,\epsilon))$ obtained by solving the parametric family of chance-constrained programs below, where $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ is an i.i.d. sample of observations of the demand vector $\tilde{\xi}$.

$$\begin{aligned} z_{\mathrm{c}}^*(n,\epsilon) \; = \min_{x \in X} \quad & \sum_{i=1}^{m} c_i x_i \\ \text{s.t.} \quad & \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(\tilde{\xi}^j \not\leq x) \leq \epsilon. \end{aligned} \tag{4.1}$$

Model (4.1) approximates the underlying deterministic chance-constrained model

$$\begin{aligned} z_{\mathrm{c}}^*(\epsilon) \; = \min_{x \in X} \quad & \sum_{i=1}^{m} c_i x_i \\ \text{s.t.} \quad & \mathbb{P}(\tilde{\xi} \not\leq x) \leq \epsilon, \end{aligned} \tag{4.2}$$

which determines the true EF. The pseudo EF is a decreasing step function with a finite number of discontinuities. The convex envelope of the pseudo EF (defined formally later in this chapter) is a convex curve that lower bounds it and is the pointwise supremum of all curves with this property. The envelope is a piecewise linear function whose extreme points (where two pieces meet) are on the pseudo EF, i.e., they correspond to design solutions that cost the least, given the risk at that point. The envelope represents the "tightest" outer convex approximation to the pseudo EF, a property that assumes importance if we know that $EF(h, p)$ is convex.

Now, observe that by definition of the envelope, its epigraph, which represents a "convexification" of the image space for the sampled problem, is convex. As outlined in our remarks in the introductory chapter, any extreme point on the envelope can therefore be found by solving a single objective program that minimizes a weighted sum of cost and risk. The objective $\lambda_1 h(x) + \lambda_2 p(x)$ can be renormalized to include a single parameter λ that represents a dualizing parameter for the risk constraint in model (4.1). Thus, the parametric program

$$\min_{x \in X} \quad \sum_{i=1}^{m} c_i x_i + \lambda \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(\tilde{\xi}^j \not\leq x) \tag{4.3}$$

can be used to solve for the envelope completely by varying λ appropriately. Model (4.3) can be interpreted as the problem of choosing a set of scenarios such that there exists an installation design that violates exactly the scenarios contained in this set, and that minimizes the sum of the installation costs and

violation penalties, with a penalty of λ/n per violation. Equivalently, we seek to maximize the difference between cost savings, in relation to the zero risk setting, and the penalty term. This maximization model is a special case of the selection problem introduced by Balinski [3] and Rhys [45]. The selection problem is the following: Given are a set of items, and a collection of sets, each made up of one or more of these items. Each item has an associated cost, while each set carries a benefit if selected. The problem is to find which sets to select so as to maximize the profit, i.e., the objective is to maximize the sum of benefits of the selected sets minus the cost of selecting items that occur in any of the selected sets. Recognizing items as scenarios and sets as pairs (i, j), where i indexes facilities, and j indexes scenarios, and benefits as incremental savings for a facility i due to violating scenario j, and the cost of an item as the penalty λ/n, the similarity between the two models is apparent. Rhys and Balinski establish independently that the selection problem is solvable in strongly polynomial-time as a min-cut problem on a bipartite network with a source s and a terminal t. We later describe a variant of this network that we use in our computations.

Our work, or more generally, the selection problem, relates closely to recent work by Hochbaum [22] , who considers a selection problem in which facilities must be installed, with the budget made available over time, so that nestedness is a vital requirement. The author shows that solutions that are extreme points on the envelope are nested by means of a parametric max-flow argument; in other words, the solutions that can be obtained in polynomial

time happen to obey the nestedness property. Note that the number of extreme points could, in theory, be equal to the maximum value of the budget (assume the budget takes only integral values) in the model of Hochbaum, or the sample size in our model. Or, it could take the minimum value of 2, as happens when the only extreme points are the empty set and the entire set of facilities or scenarios. The so-called *pseudoflow algorithm*, proposed in this work, is a strongly polynomial-time algorithm that, interestingly, finds all the extreme points with the same complexity as a single s-t min-cut problem. Nehme establishes related results in [33], in the setting of a stochastic network interdiction model that aims to detect smuggling of nuclear material along the borders of a country. The model in this work is equivalent to the facility-sizing problem we consider, with facilities and demand scenarios corresponding to smuggler scenarios and detectors respectively. Nestedness is established on the envelope, using a different approach from that of [22], as is the polynomial time solvability of the parametric problem (4.3). The author also develops a sequential linear-programming-based approach to construct the envelope.

In our work in this chapter, we build on the above results, exploiting nestedness and convexity to realize our aim of constructing the true EF. Our contributions are summarized below:

a) We strengthen the pointwise convergence results of the pseudo EF, detailed in Chapter 3, to uniform convergence. We use this to establish uniform convergence of the convex envelope of the pseudo EF, to the

true EF, under the assumption that the latter is convex in the risk range of interest.

b) We use uniform convergence of the envelopes to establish a density result that shows that asymptotically, the envelope is almost identical to the EF in a certain risk range.

c) We use the upper bound on the risk range as well as the nestedness property of the extreme points on the envelope to significantly reduce the size of the parametric min-cut model, which leads to enhanced computational performance of the push-relabel algorithm.

d) We propose and test computational enhancements of the min-cut model, demonstrating improved solution times and fast convergence.

4.2 Convexity of the Efficient Frontier

Consider the chance-constrained parametric formulation

$$\begin{aligned} z_{\mathrm{c}}^*(\epsilon) \; = & \min_{x \in X} \quad \sum_{i=1}^{m} c_i x_i \\ & \text{s.t.} \quad \mathbb{P}(\tilde{\xi} \not\leq x) \leq \epsilon. \end{aligned} \tag{4.4}$$

If X is convex, and the cost function and probability of demand violation are both convex and lower semicontinuous in the decision vector x in the range $[\underline{\epsilon}, \overline{\epsilon}]$, then it follows that $z_{\mathrm{c}}^*(\cdot)$ is convex in this range, provided X is compact. This usually means that the chance-constrained variant of the bicriteria model is computationally tractable. However, this is not true of

model (4.4) because of the multivariate integration that $\mathbb{P}(\tilde{\xi} \not\leq x)$ involves. Unfortunately, the sampling-based approximation to (4.4) is NP-hard to solve. This is disappointing considering the underlying convexity of the true problem. We find a way around this shortcoming of the approximation by constructing a convex approximation to the sampling-based approximation, namely the convex envelope. The advantages of the envelope are that (i) we can derive uniform convergence results to the true EF in $[\underline{\epsilon}, \overline{\epsilon}]$ in an almost everywhere sense, and (ii) forming such an envelope only requires effort that is strongly polynomial in the size of the approximation, as shown in [22] and [33]. We discuss this in greater detail in the sections that follow.

It is natural, at this point, to look for conditions that guarantee that general risk and cost functions $p(\cdot)$ and $h(\cdot)$, are convex. The cost objective $h(\cdot)$ involves deterministic parameters of the model, and standard tools from analysis may be used to verify its convexity. Convexity of the risk function $p(\cdot)$ is, however, less understood in the literature. Recall that for the general bicriteria model, $p(x) = \mathbb{P}(G(x, \tilde{\xi}) \notin C)$, and $G(\cdot)$, C and the distribution of $\tilde{\xi}$ must be specified in order to study its properties. Consider the facility-sizing model in which $G(x, \xi) = \xi - x$, and the set C is the orthant of nonpositive components of appropriate dimension. In this case, $p(x)$ reduces to $\mathbb{P}(x \not\geq \tilde{\xi})$. Suppose the random vector $\tilde{\xi}$ has a concave cumulative distribution function (cdf) for x sufficiently large, i.e., suppose there exists $x_0 \in \mathbb{R}^m$ such that the function $1 - p(x) = \mathbb{P}(\tilde{\xi} \leq x)$ is concave in the region $A = \{x \mid x \geq x_0\}$. Then, in the same region, $p(\cdot)$ is convex. Since large values of the design vector x

correspond to low values of risk, this implies that $p(\cdot)$ is convex when it takes values close to zero. Assuming $[\underline{\epsilon}, \overline{\epsilon}]$ is contained in this risk interval, $EF(h, p)$ is convex in this high reliability range, provided $h(\cdot)$ is also convex, and X is compact.

We now pose the question: What are some probability distributions with the property that the cdf is concave for large values of the argument? Prékopa [41] establishes that the multivariate normal distribution enjoys this property. He also establishes such a result for the Dirichlet distribution, using a similar line of proof. Beyond this, concavity properties of cdfs of multivariate probability distributions do not seem to be, in general, well understood.

4.3 Convergence of Approximating Envelopes

For the rest of this chapter, we assume that $EF(h, p)$ is convex for $p(\cdot) \in [0, \epsilon^*]$, $0 < \epsilon^* \leq 1$. The function $z_{\mathrm{c}}^*(n, \cdot)$ is a monotone decreasing, right-continuous step function, and Theorem 4.3.1 below, from [38], provides conditions under which it converges pointwise to $z_{\mathrm{c}}^*(\cdot)$ as the sample size n grows large. We state the result in generality for the chance-constrained variant of the bicriteria problem and its sampled version given below.

$$\begin{aligned} z_{\mathrm{c}}^*(\epsilon) \;&=\; \min_{x \in X} \quad h(x) \\ &\qquad \text{s.t.} \quad \mathbb{P}(G(x, \tilde{\xi}) \leq 0) \geq 1 - \epsilon, \end{aligned} \tag{4.5}$$

$$\begin{aligned} z_{\mathrm{c}}^*(n, \epsilon) \;&=\; \min_{x \in X} \quad h(x) \\ &\qquad \text{s.t.} \quad \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(G(x, \tilde{\xi}^j) \leq 0) \geq 1 - \epsilon, \end{aligned} \tag{4.6}$$

where the observations $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ form an i.i.d. sample. Recall that $p(x) = \mathbb{P}(G(x, \tilde{\xi}) \not\leq 0)$.

Theorem 4.3.1. (Pagnoncelli et al. [38]) *Suppose X is compact, $h(\cdot)$ is continuous, $G(\cdot, \cdot)$ is a Cartheodory function on $X \times (0, 1)$, and condition (1) below is true:*

(1) There exists an optimal solution $\overline{x}$ to (4.5) such that for any $\delta > 0$, there exists $x \in X$ satisfying $\|x - \overline{x}\| \leq \delta$ and $p(x) < \epsilon$.

Then, $z_c^*(n, \epsilon) \to z_c^*(\epsilon)$ w.p.1 as $n \to \infty$.

The result states that the pseudo EF converges to the true EF for fixed values of risk in $(0, 1)$ as the sample size grows large, and that this holds w.p.1. Figure 4.1 shows a pseudo EF and the true EF.

Condition *(1)* exists for the sake of robustness to perturbations to the risk level ϵ, and is assumed to hold in our case. Continuity and measurability of $G(\cdot, \cdot)$ in its arguments is easily shown to hold for the function $G(x, \xi) = \xi - x$. The following result establishes conditions under which pointwise convergence may be strengthened to uniform convergence, denoted by $z_c^*(n, \epsilon) \overset{\text{unif}}{\to} z_c^*(\epsilon)$.

Theorem 4.3.2. (Doob [13]) *Suppose $\{F_n(\cdot)\}$ is a sequence of monotone increasing functions from $\mathbb{R}$ into $\mathbb{R}$ such that $F_n(x) \to F(x)$ at all points x at which $F(\cdot)$ is continuous, as $n \to \infty$. Then, $F_n(x) \overset{\text{unif}}{\to} F(x)$ on each compact interval of continuity of $F(\cdot)$ as $n \to \infty$.*

Setting $F = -z_c^*(\cdot)$, $F_n = -z_c^*(n, \cdot)$, we see that the hypotheses of Theorem 4.3.2 are satisfied. Thus, applying the theorem to our setting, we get that the

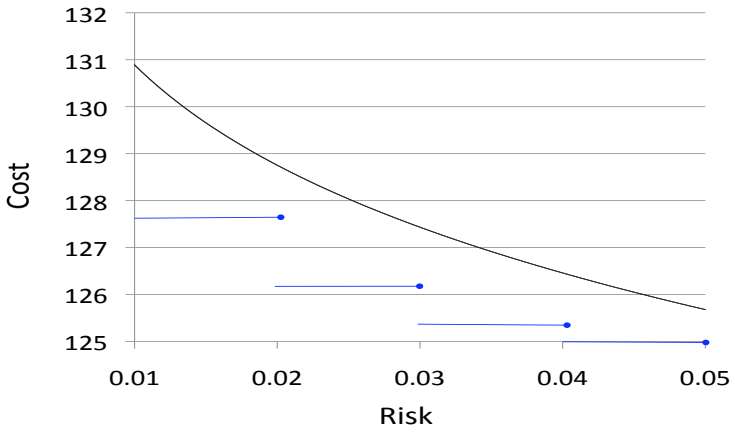

Figure 4.1: True and pseudo EFs for a model trading off cost and risk. The black line is the true EF, and the blue line, the pseudo EF, with the circles marking discontinuities. The pseudo EF is formed with a sample of $n = 100$ observations.

pseudo EF converges uniformly to $EF(h, p)$ w.p.1 on every compact subinterval of $[0, 1]$ on which $EF(h, p)$ is continuous.

Now, given that $EF(h, p)$ is convex, forming some "convexified" version of the pseudo EF seems indicated. Let $f(\cdot)$ be a function defined on a convex set X. Define the *convexification* of $f(x)$ (see [5]) as

$$\hat{f}(x) = \inf_{\substack{x_1, x_2 \in X \\ \alpha \in [0,1]}} \{\alpha f(x_1) + (1 - \alpha) f(x_2) \mid x = \alpha x_1 + (1 - \alpha) x_2\}.$$

We view $\hat{(\cdot)}$ as an operator that can be applied to a function. As discussed in [5], the convexification of a function is simply the pointwise supremum of all convex functions that bound the original function from below on X. Observe that the convexification is determined by the underlying set X. Furthermore,

we have $\text{epi}(\hat{f}) = \text{co}(\text{epi}(f))$, where $\text{epi}(f)$ denotes the epigraph of f, and $\text{co}(\text{epi}(f))$, its convex hull. We refer to the graph of $\hat{f}(\cdot)$ as the *convex envelope* of the graph of $f(\cdot)$. Hence, $EF(h,p)$ and the pseudo EF both have convex envelopes. Note that the convexification of $z_c^*(n,\cdot)$ is a piecewise linear function whose extreme points (where two adjacent pieces meet) are jump points of $z_c^*(n,\cdot)$.

Our main result is that the convex envelope of the pseudo EF converges uniformly to the convex envelope of $EF(h,p)$ on $[0,1]$. Before turning to the proof of this result, we consider its implication. If $EF(h,p)$ is convex in $[0,\epsilon^*]$, then its envelope is identical to it in a risk range $[0,\epsilon'] \subseteq [0,\epsilon^*]$. Thus, our result implies that the convex envelope of the pseudo EF converges uniformly to $EF(h,p)$ in $[0,\epsilon']$.

Theorem 4.3.3. *Suppose $EF(h,p)$ is continuous on $[0,1]$. Then $\hat{z}_c^*(n,\epsilon) \overset{\text{unif}}{\to} \hat{z}_c^*(\epsilon)$ on $[0,1]$ w.p.1 as $n \to \infty$. Further, suppose $EF(h,p)$ is convex on $[0,\epsilon^*]$. Then, $\hat{z}_c^*(n,\epsilon) \overset{\text{unif}}{\to} z_c^*(\epsilon)$ on $[0,\epsilon']$ w.p.1 as $n \to \infty$, where $\epsilon' = \sup\{\epsilon \in (0,\epsilon^*] \mid z_c^*(\delta) = \hat{z}_c^*(\delta)\, \forall \delta \in [0,\epsilon]\}$ is the leftmost point at which $z_c^*(\cdot)$ and $\hat{z}_c^*(\cdot)$ diverge.*

Proof. Fix $\epsilon \in [0,1]$ and consider the equivalent definition

$$\hat{z}_c^*(n,\epsilon) = \inf_{0 \le \epsilon_1 \le \epsilon \le \epsilon_2 \le 1} \{\alpha z_c^*(n,\epsilon_1) + (1-\alpha) z_c^*(n,\epsilon_2),\ \alpha = (\epsilon_2 - \epsilon)/(\epsilon_2 - \epsilon_1)\}.$$

For $\epsilon_2 = \epsilon_1 = \epsilon$, we define $\alpha z_c^*(n,\epsilon_1) + (1-\alpha) z_c^*(n,\epsilon_2) = z_c^*(n,\epsilon)$. Let $\delta > 0$ be given, and using uniform convergence of the pseudo EF to $EF(h,p)$ on $[0,1]$, let n_0 be a positive integer (independent of ϵ), such that for $n \ge n_0$, $\max_{\epsilon' \in [0,1]} |z_c^*(n,\epsilon') - z_c^*(\epsilon')| < \delta$ w.p.1. We denote the set $\{(\epsilon_1,\epsilon_2) \mid 0 \le$

$\epsilon_1 \leq \epsilon \leq \epsilon_2 \leq 1\}$ by $C(\epsilon)$ for ease of notation. For $(\epsilon_1, \epsilon_2) \in C(\epsilon)$ and $n \geq n_0$, we have

$$\begin{aligned}\alpha(z_c^*(\epsilon_1) - \delta) + (1-\alpha)(z_c^*(\epsilon_2) - \delta) &\leq \alpha z_c^*(n, \epsilon_1) + (1-\alpha) z_c^*(n, \epsilon_2) \\ &\leq \alpha(z_c^*(\epsilon_1) + \delta) + (1-\alpha)(z_c^*(\epsilon_2) + \delta),\end{aligned}$$

where α is given above. The inequalities can be simplified to

$$\begin{aligned}\alpha z_c^*(\epsilon_1) + (1-\alpha) z_c^*(\epsilon_2) - \delta &\leq \alpha z_c^*(n, \epsilon_1) + (1-\alpha) z_c^*(n, \epsilon_2) \\ &\leq \alpha z_c^*(\epsilon_1) + (1-\alpha) z_c^*(\epsilon_2) + \delta.\end{aligned}$$

Thus,

$$\begin{aligned}\inf_{(\epsilon_1,\epsilon_2)\in C(\epsilon)} \alpha z_c^*(\epsilon_1) + (1-\alpha) z_c^*(\epsilon_2) - \delta &\leq \hat{z}_c^*(n, \epsilon) \\ &\leq \inf_{(\epsilon_1,\epsilon_2)\in C(\epsilon)} \alpha z_c^*(\epsilon_1) + (1-\alpha) z_c^*(\epsilon_2) + \delta.\end{aligned}$$

In other words, we have

$$\hat{z}_c^*(\epsilon) - \delta \leq \hat{z}_c^*(n, \epsilon) \leq \hat{z}_c^*(\epsilon) + \delta,$$

for $n \geq n_0$ and arbitrary $\epsilon \in [0, 1]$, showing that $\hat{z}_c^*(n, \cdot)$ converges uniformly to $\hat{z}_c^*(\cdot)$ on $[0, 1]$ w.p.1. □

Uniform convergence of the approximating envelopes, while a strong result in itself, also has a significant implication: For most distributions, $EF(h, p)$ is a nonlinear, continuous, strictly decreasing function. By Theorem 4.3.3, the approximating envelope can be made arbitrarily and uniformly close to it in $[0, \epsilon']$ by choosing a sufficiently large sample size. Since the envelope is piecewise linear, this implies that asymptotically, there exist extreme

points of the approximation that are arbitrarily close to any point on the EF, in $[0, \epsilon']$. We formalize this intuitive argument in the corollary below.

Corollary 4.3.4. *Let $\epsilon_1 < \epsilon_2$ satisfy $[\epsilon_1, \epsilon_2] \subset [0, \epsilon']$ given in Theorem 4.3.3, and suppose $EF(h, p)$ is strictly convex on $[\epsilon_1, \epsilon_2]$. Let $X_{\epsilon \in [\epsilon_1, \epsilon_2]}[\hat{z}_c^*(n, \epsilon)]$ denote the number of extreme points of $\hat{z}_c^*(n, \cdot)$ on $[\epsilon_1, \epsilon_2]$. Then, $X_{\epsilon \in [\epsilon_1, \epsilon_2]}[\hat{z}_c^*(n, \epsilon)] \to \infty$ w.p.1 as $n \to \infty$.*

Proof. $\hat{z}_c^*(n, \cdot)$ is a piecewise linear function on $[\underline{\epsilon}, \overline{\epsilon}]$. If $X_{\epsilon \in [\epsilon_1, \epsilon_2]}[\hat{z}_c^*(n, \epsilon)]$ does not grow without bound, then there exists $\underline{\epsilon}' < \overline{\epsilon}'$ such that $\hat{z}_c^*(n, \cdot)$ is linear on $[\underline{\epsilon}', \overline{\epsilon}'] \subset [\epsilon_1, \epsilon_2]$ for all n. Thus, there is a nonzero probability that there exists $\delta > 0$ such that $\sup_{\epsilon \in [\underline{\epsilon}', \overline{\epsilon}']} |\hat{z}_c^*(n, \epsilon) - z_c^*(\epsilon)| > \delta$ for all n, by strict convexity of $z_c^*(\cdot)$. This contradicts Theorem 4.3.3 , i.e., it contradicts $\lim_{n \to \infty} \sup_{\epsilon \in [\underline{\epsilon}, \overline{\epsilon}]} |\hat{z}_c^*(n, \epsilon) - z_c^*(\epsilon)| = 0$, w.p.1. □

Corollary 4.3.4 may be viewed as a density result. Furthermore, it serves to distinguish the facility sizing problem with a concave cumulative distribution function for demand from more general versions of this model (and more general instances of the selection problem) in which it is possible to have an envelope consisting of just two points, corresponding to risk values of zero and one. The key assumption of convexity of $EF(h, p)$ close to zero risk is validated for some common distributions in Section 4.5. We next turn to construction of the envelope.

4.4 Construction of the Convex Envelope

We now return to the facility sizing model and detail how the convex envelope of the pseudo EF is constructed. For this, we turn to [22] and [33], and outline below, the results in these works that pertain to this construction.

Given a sample $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ from the distribution of $\tilde{\xi}$, let M denote the worst case cost, i.e., the cost incurred by the facility-sizing plan that ensures zero risk in model (4.6). Given a solution $x \in X$, we can then define the *cost savings*, $s(x)$, resulting from implementing x, as $s(x) = M - h(x)$. The pseudo EF can be interpreted as a function of $s(\cdot)$ versus $p(\cdot)$, and is now an increasing, step function of ϵ in $[0, 1]$. The concave envelope of the pseudo EF, defined analogously as the pointwise infimum of all upper-bounding concave functions, is a strictly concave, piecewise linear function. Adjacent extreme points on the envelope are separated by intervals that are multiples of $1/n$ on the risk axis. For the remainder of this chapter, we deal with this concave envelope, with the understanding that the convex envelope can be recovered by considering cost instead of savings as the objective.

Let K be a subset of $\mathcal{N} = \{1, \ldots, n\}$. Define the *gain function* $g(\cdot) : 2^{\mathcal{N}} \to \mathbb{R}$ by

$$g(K) = M - \sum_{i=1}^{m} c_i \max_{j \notin K}(\xi_i^j),$$

and the *violation function* $u(\cdot) : 2^{\mathcal{N}} \to \mathbb{R}$ by

$$u(K) = |K|.$$

The set K represents the set of scenarios for which demand is violated in the sense that the violated scenarios form the largest subset of K for which it is possible to violate exactly the members of that subset. This maximal subset of violated scenarios constitutes a design vector x for which we have $g(K) = s(x)$. Conversely, for every design vector x, there exists a unique minimal set $K \in 2^{\mathcal{N}}$ such that $g(K) = s(x)$. Thus, every set K corresponds to a point in the image space via the pair $(u(K)/n, g(K)$. We now proceed to characterize sets which correspond to points on the envelope, as well as extreme points on the envelope.

Given a set $K \in 2^{\mathcal{N}}$, the point $(u(K)/n, g(K))$ lies on the concave envelope if and only if for $C, D \subseteq \{1, \ldots, n\}$ with $u(C) \leq u(K) \leq u(D)$,

$$g(K) \geq \alpha g(C) + (1 - \alpha) g(D),$$

where $\alpha = \frac{u(D) - u(K)}{u(D) - u(C)}$. In other words, the chord joining $(u(C)/n, g(C))$ and $(u(D)/n, g(D))$ cannot lie above $(u(K)/n, g(K))$ for sets C and D satisfying the conditions above. Furthermore, $(u(K)/n, g(K))$ is an extreme point of the envelope if and only if for $C, D \subseteq \{1, \ldots, n\}$ with $u(C) < u(K) < u(D)$,

$$g(K) > \alpha g(C) + (1 - \alpha) g(D)$$

with α defined as above. As a matter of convention, we also define the two extremities $(0, 0)$ and $(1, M)$ as extreme points.

The gain function obeys the supermodularity property, i.e., if $A \subseteq B \subseteq \{1, \ldots, n\}$, and $j \in \{1, \ldots, n\} \setminus B$, then

$$g(B \cup \{j\}) - g(B) \geq g(A \cup \{j\}) - g(A).$$

Supermodularity of $g(\cdot)$ simply states that given two sets of violations, one contained in the other, greater savings can be obtained for the larger set when a new violation is added. It can also be seen that the violation function is submodular, i.e., if $A \subseteq B \subseteq \{1, \ldots, n\}$, and $j \in \{1, \ldots, n\} \setminus B$, then

$$u(B \cup \{j\}) - u(B) \leq u(A \cup \{j\}) - u(A).$$

For $A, B \subseteq \{1, \ldots, n\}$ with $u(A) \leq u(B)$, define the *gain-to-violation* ratio as

$$gv(A, B) = \frac{g(B) - g(A)}{u(B) - u(A)}.$$

Thus, $gv(A, B)$ represents the rate of increase in cost savings with the number of violations. The following result, proved in [33], establishes that when A (B) is fixed, points B (A) on the envelope maximize (minimize) this ratio.

Theorem 4.4.1. (Nehme [33]) *Let $A, B \subseteq \{1, \ldots, n\}$ be sets such that $u(A) \leq u(B)$.*

(i) Suppose B lies on the concave envelope of the pseudo EF. Then, $gv(A, B) = \max_{\substack{K \subseteq \{1,\ldots,n\} \\ u(K) \geq u(B)}} gv(A, K)$.

(ii) Suppose A lies on the concave envelope of the pseudo EF. Then, $gv(A, B) = \min_{\substack{K \subseteq \{1,\ldots,n\} \\ u(K) \leq u(A)}} gv(K, B)$.

The next result establishes the nestedness property, and can be proved by using supermodularity of the gain function and submodularity of the violation function [33, 49], by using properties of a parametric maximum-flow problem [22]

or by using properties of the linear programming relaxation of a reformulation of model (4.6) [33].

Theorem 4.4.2. (Nehme [33]) *Let $A, B \subseteq \{1, \ldots, n\}$ correspond to extreme points on the concave envelope of the pseudo EF, with $u(A) \leq u(B)$. Then, $A \subseteq B$.*

Recall that the chance-constrained sampled program (4.6) reduces to the mixed-integer program below:

$$\begin{aligned} z_{\mathrm{c}}^*(n,\epsilon) \;=\; & \min_{x,y} \quad \sum_{i=1}^{m} c_i x_i \\ & \text{s.t.} \quad \frac{1}{n}\sum_{j=1}^{n} y^j \leq \epsilon, \\ & \qquad x_i \geq \xi_i^j (1 - y^j), i = 1, \ldots, m, j = 1, \ldots, n, \\ & \qquad y^j \in \{0,1\}, j = 1, \ldots, n. \end{aligned}$$

We rewrite the model as a reward-collecting formulation in which we seek to maximize cost savings by failing to meet demand under a limited number of demand scenarios. Binary variables, w, indicate whether we collect these rewards in the following model:

$$\begin{aligned} z_{\mathrm{c}}^*(n,\epsilon) \;=& \max_{y,w} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} c_i \big(\xi_i^{t(i,j)} - \xi_i^{t(i,j+1)}\big) w_i^{t(i,j)} \\ & \text{s.t.} \quad \frac{1}{n}\sum_{j=1}^{n} y^j \leq \epsilon, \\ & \qquad y^j \geq w_i^j,\ i = 1, \ldots, m, j = 1, \ldots, n, \\ & \qquad w_i^{t(i,j)} \geq w_i^{t(i,j+1)},\ i = 1, \ldots, m, j = 1, \ldots, n-1, \\ & \qquad y^j, w_i^j \in \{0,1\},\ i = 1, \ldots, m, j = 1, \ldots, n, \end{aligned} \tag{4.7}$$

where $t(i,\cdot) : \{1,\ldots,n\} \mapsto \{1,\ldots,n\}$ is a map that expresses the demand realizations for facility i in decreasing order, i.e., $\xi_i^{t(i,j)} > \xi_i^{t(i,j+1)}$, $i = 1,\ldots,m, j = 1,\ldots,n-1$, and $\xi_i^{t(i,n+1)} = 0$, $i = 1,\ldots,n$. The following theorem completes the set-up for a linear programming algorithm to form the envelope.

Theorem 4.4.3. (Nehme [33]) *Let (y,w) be an extreme point solution of the LP relaxation of (4.7). Then, all the fractional components of y are equal.*

An LP-based algorithm to find extreme points of the concave envelope of the EF for model (4.6) is as follows: It iteratively finds successive extreme points from left ($\epsilon = 0$) to right ($\epsilon = \bar{\epsilon}$). The leftmost extreme point of the envelope is $(0,0)$ and corresponds to K being the empty set. At the iterative step, let K be the set of scenario indices violated at the last extreme point. Solve the LP-relaxation of (4.7) with $\epsilon = |K| + 1$. From Theorems 4.4.1 and 4.4.3, the optimal solution to this program corresponds to the next extreme point on the envelope, and the set of scenarios that are violated corresponds to the indices j for which $y^j > 0$ at the optimal solution. This procedure is then repeated after updating K, until all extreme points in $[\underline{\epsilon}, \bar{\epsilon}]$ are found. The algorithm solves at most $\lfloor \bar{\epsilon} n \rfloor$ linear programs, and hence runs in polynomial time in m and n.

Hochbaum uses a geometric idea to form the envelope. Consider the chord connecting the two extremities of the envelope $(0,0)$ and $(1,M)$. The slope λ of this chord is M. Note that the chord lies strictly below the envelope due to the fact that the envelope is comprised of more than one linear segment

(see Corollary 4.3.4.). Consider the maximization problem

$$\max_{\epsilon} \quad M - z_{\mathrm{c}}^{*}(n, \epsilon) - \lambda\, \epsilon \tag{4.8}$$

where $z_{\mathrm{c}}^{*}(n, \epsilon)$ is given by (4.6). Note that owing to the equivalence between choosing design vectors x and choosing sets $K \subseteq \{1, \ldots, n\}$, we can interpret the solution space of (4.8) as 2^n, with the decision variable being K, and the objective function and constraint modified suitably. There exists an optimal solution ϵ^* to the above program which is an extreme point of the envelope in $[0, 1]$, and at which the objective in (4.8) is positive. Note that if there exists a linear segment on the envelope with slope λ, model (4.8) has multiple optimal solutions corresponding to all points on the segment, of which we pick one of the two end points. This new extreme point determines two subintervals on the risk axis, on which the above procedure may be repeated, recursively, until all extreme points are found. Note that the equation for the chord joining extremities of the envelope on a subinterval involves a constant term, but it may dropped from the penalty term in the objective in (4.8) as it is independent of ϵ. This iterative procedure forms the envelope on $[0, 1]$.

Our aim is to form the envelope on $[\underline{\epsilon}, \overline{\epsilon}]$. Therefore, we replace (4.8) in the iterative step with the model

$$\begin{aligned} \max_{\epsilon} \quad & M - z_{\mathrm{c}}^{*}(n, \epsilon) - \lambda\, \epsilon \\ \text{s.t.} \quad & 0 \leq \epsilon \leq \overline{\epsilon}. \end{aligned} \tag{4.9}$$

Solving (4.9) produces the envelope on $[0, \overline{\epsilon}]$. While replacing the lower bound of zero with $\underline{\epsilon}$ would restrict the envelope to $[\underline{\epsilon}, \overline{\epsilon}]$, a lower bound on risk is not, in general, easily enforced in model (4.4) or its dualized version (4.8).

We conclude this section by outlining how the parametric model (4.9) can be solved as a parametric min-cut problem, enabling use of the push-relabel algorithm [19] or the pseudoflow algorithm [22]. The construction that follows is rooted in the ideas found in [3], [45] and [40] on solving the selection problem via a max-flow/min-cut problem. We construct a network with a source node s and a terminal node t. We (nominally) introduce n nodes, $v(j)$, $j = 1, \ldots, n$, and connect $v(j)$ to t by the arc $r(j)$, $j = 1, \ldots, n$. Connect $u(i, j)$ to $u(i, j-1)$ by arc $li(i, j), i = 1, \ldots, m$, $j = 2, \ldots, n$. We interpret the nodes as follows. For each i, the nodes $u(i, j)$, $j = 1, \ldots, n$ represent the demand realizations at the i^{th} facility, in decreasing order. For instance, $u(i, 1)$ represents the highest demand realization for facility i. The node $v(j)$ represents realization ξ^j, $j = 1, \ldots, n$. We complete the graph by connecting node $u(i, j)$ to node $v(t(i, j))$ using arc $lr(i, j)$.

We now endow the links with capacities as follows. The capacity of $l(i, j)$ equals $c_i(\xi_i^{t(i,j)} - \xi_i^{t(i,j+1)})$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. The capacities of the links $li(i, j)$ are all set equal to some big number $\overline{M}$. The capacities of the arcs $lr(i, j)$ are also set equal to $\overline{M}$. The capacities of the links $r(j)$ are all set equal to λ/n. Finally, we note that the number of nodes and edges in this network are linear in m and n. Specifically, we have $(m + 1)n + 2$ nodes and $m(3n - 1) + n$ arcs in our nominal network. We say that we *nominally* introduce m sets of n nodes because we can reduce this number as we show below and in Section 4.6.

It is helpful to consider an s-t min-cut problem on the above network

with capacities acting as costs of cutting edges. First, note that none of the arcs $li(i,j)$ or $lr(i,j)$ can occur in a min-cut due to the large cost. Consider the i^{th} facility for an arbitrary i. Let $j^*(i)$ be the smallest j such that $r(t(i,j^*(i)))$ is not in the cut. Then, it follows that $l(i,j)$ must be in the cut for $j \geq j^*$. The demand $\xi_i^{t(i,j^*(i))}$ equals the capacity installed at facility i. Repeating this argument for each facility, we see that the set of arcs $r(j)$ that are cut correspond to violated scenarios. The cost of this cut thus equals the sum of the installation cost and the term $\lambda|K|/n$, where K is the set of violated scenarios. Minimizing cost over all cuts is equivalent to maximizing $M-$ (cost of cut), i.e., maximizing $M - z_{\text{c}}^*(n,\epsilon) - \lambda\epsilon$. This is exactly the objective function in (4.9).

The constraint $0 \leq \epsilon \leq \bar{\epsilon}$ implies that for a design vector x and any facility i, the highest j for which $x_i < \xi_i^{t(i,j)}$ is $\lfloor \bar{\epsilon} n \rfloor$. This implies that the arcs $l(i,j)$ are in the min-cut for these values of j, thus eliminating the need for the associated nodes $u(i,j)$ in the network. Thus, the size of the network is reduced from $(m+1)n+2$ nodes to $m\lfloor \bar{\epsilon} n \rfloor + n + 2$ nodes, resulting in significant computational savings.

4.5 Concave Probability Distributions

Our discussion in Section 4.2 on the convexity of the EF shows that if the cdf $F(x) = \mathbb{P}(\tilde{\xi} \leq x)$ of $\tilde{\xi}$ is concave for large values of the argument, then the EF is convex for small values of risk, given certain other conditions. The

following result, due to Prékopa, establishes the concavity of $F(\cdot)$ in the case of a multivariate normal distribution.

Theorem 4.5.1. (Prékopa [42]) *Let $\tilde{\xi}$ be an m-variate normal random vector with mean zero and nonsingular correlation matrix R, and let $\Phi(x) = \mathbb{P}(\tilde{\xi} \leq x)$ be its cdf. Then, $\Phi(\cdot)$ is concave in the region $\{x \mid x_i \geq \sqrt{m-1}, i = 1, \ldots, m\}$.*

The result above can be generalized to an $m-$variate normal distribution with mean vector μ and covariance matrix Σ as follows.

Corollary 4.5.2. *Let $\tilde{\xi}$ be an m-variate normal random vector with mean μ and nonsingular covariance matrix Σ, and let $\Phi(x) = \mathbb{P}(\tilde{\xi} \leq x)$ be its cdf. Then, $\Phi(\cdot)$ is concave in the region $\{x \mid x_i \geq \mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'}, i = 1, \ldots, m\}$, where σ_i is the standard deviation of $\tilde{\xi}_i$, $i = 1, \ldots, m$.*

We can directly apply this result to the facility sizing model when demand is modeled as being multivariate normal. Furthermore, the values of risk for which the theorem can be applied, conform to our risk range $[\underline{\epsilon}, \bar{\epsilon}]$, assuming $\bar{\epsilon}$ is small. As observed in [42], in the case of the m-variate standard normal distribution with independent components, we have that $\mathbb{P}(\tilde{\xi}_i \leq \sqrt{m-1}, i = 1, \ldots, m) = \mathbb{P}(Z \leq \sqrt{m-1})^m$, where Z is a standard normal random variable, takes values close to 1 for modest values of m; for example, .93 for $m = 6$. In our experiments, we use, for instance, $m = 10$ for which it takes the value .986. This corresponds to a risk value of .014, which is an acceptable risk threshold for the design of many systems.

The following result guarantees a low risk region of convexity for the multivariate lognormal distribution on demand.

Corollary 4.5.3. *Let $\tilde{\xi}$ be an m-variate lognormal random vector with parameters μ and a nonsingular covariance matrix Σ, i.e., the component wise logarithm of $\tilde{\xi}$ is multivariate normal with mean μ and covariance matrix Σ. Let $\Phi(y) = \mathbb{P}(\tilde{\xi} \leq y)$ be its cdf. Then, $\Phi(\cdot)$ is concave in the region $\{y \mid y_i \geq e^{\mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'}},\ i = 1, \ldots, m\}$.*

Proof. We have

$$\mathbb{P}(\tilde{\xi} \leq y) = \mathbb{P}(\log \tilde{\xi} \leq \log y)$$

where the logarithm is taken component-wise. By Corollary 4.5.2, the function on the right side is concave in $\log y$ for $\log y_i \geq \mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'}$, $i = 1 \ldots, m$. Note that it is also increasing in $\log y$, which is a concave function. Since an increasing concave function of a concave function is concave, the result follows. □

The multivariate Dirichlet distribution is shown to have a similar property in [42]. The random vector $\tilde{\xi}$ given by $\tilde{\xi}_i = \frac{\tilde{Y}_i}{\tilde{Y}_1 + \cdots \tilde{Y}_m + \tilde{Y}_{m+1}}$, $i = 1, \ldots, m$, where each $\tilde{Y}_i$ is a gamma random variable with parameter a_i, follows a m-variate Dirichlet distribution with parameters $a_1, \ldots, a_{m+1}$. Prékopa establishes the following concavity result for its cdf.

Theorem 4.5.4. (Prékopa [42]) *Let $\tilde{\xi}$ obey an m-variate Dirichlet distribution with parameters $a_1, \ldots, a_{m+1}$, and let $\tilde{\phi}(x)$ be its cdf. Then, the following is true: (i) if $a_1 + \cdots + a_m \leq 1$ and $a_1 + \cdots + a_{m+1} \leq 2$, then $\tilde{\phi}(\cdot)$ is concave in*

the region $\{x \mid x \geq 0\}$*, and (ii) if* $a_1+\cdots+a_m \geq 1$ *and* $a_1+\cdots+a_{m+1} > 2$*, then* $\tilde{\phi}(\cdot)$ *is concave in the region* $\{x \mid x \geq x_0\}$ *where* $x_0 = \min(\frac{a_1+\cdots+a_m-1}{a_1+\cdots+a_{m+1}-2}, 1)$.

4.6 Computational Experience

We consider the facility sizing model of Chapter 3, and turn our attention to convergence of the concave envelope of the pseudo EF for cost savings and risk. We use the maxflow formulation idea of Hochbaum, in effect forming the concave envelope of the pseudo EF by forming the dualized version of the chance-constrained model (4.6) to find each extreme point on the envelope. In our experiments, we construct the min-cut network associated with (4.9), and then invoke an implementation of the push-relabel algorithm found in the COIN-OR online portal.

We take advantage of the nestedness property in Theorem 4.4.2. Consider the iterative step of the algorithm that solves (4.8) on a subinterval of risk values, say $[\epsilon_\ell, \epsilon_u]$. Since the optimal solution to (4.6) is available at the endpoints of this subinterval, the set of violated scenarios at the endpoints are known. Let $J_l(\cdot), J_u(\cdot) : \{1, \ldots, m\} \rightarrow \{1, \ldots, n\}$ be index-valued maps that map a facility to the rank of the highest demand satisfied at that facility, with subscripts l and u denoting solutions at lower and upper endpoints of the subinterval respectively. For instance, if the second highest demand equals the installed capacity at facility i at the lower endpoint, then $J_l(i) = 2$. Thus, $t(i, J_l(i)) = j^*(i)$ for all i (see Section 4.4 for definitions of maps $t(\cdot, \cdot)$ and

$j^*(\cdot)$). Now, for a point on the envelope that is strictly in the subinterval, consider the set of violated scenarios and fix a facility i. The set of nodes $u(i,j)$ for j such that $j < J_l(i)$, are unnecessary since the nestedness property implies that the arcs $l(i,j)$ are not in the min-cut, $j \leq J_l(i)$. Similarly, nodes $u(i,j)$ for $j \geq J_u(i)$, can also be eliminated, since $l(i,j)$ is in the min-cut for $j \geq J_u(i)$, using nestedness. Further, we set the capacity of arcs $r(j)$ to be zero for $j = t(i,j')$, $j' < J_l(i)$, since $r(j)$ is in the min-cut for such j. Setting the capacity to zero for these arcs is equivalent to incorporating the constraint that these arcs are in the cut. Thus, the size of the graph is reduced to $m(\lfloor \epsilon_u n \rfloor - \lfloor \epsilon_l n \rfloor) + n + 2$ nodes, from $m\lfloor \bar{\epsilon} n \rfloor + n + 2$ nodes in the original construction. This downsizing can prove to be vastly beneficial in running time considerations, as our experiments indicate.

We construct the envelopes for two distributions, namely, the multivariate normal and lognormal, with mean vector μ and nonsingular covariance matrix Σ. The convergence results of Section 4.3 use the hypothesis that the decision space X is compact, so that $EF(h,p)$ is well-defined on the risk interval $[0,1]$. Specifically, we must ensure that $EF(h,p)$ is real-valued for zero risk, and convex in an interval $[0,\epsilon^*]$ where $0 < \epsilon^* \leq 1$. In order to ensure that this condition holds, we may consider a truncated normal distribution as follows: Choose $\overline{\xi}_i \gg \mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'}$, $i = 1, \ldots, m$, and define the truncated probability density function (pdf) by

$$f_{\tilde{\eta}}(x) = \begin{cases} f_{\tilde{\xi}}(x)/C & \text{if } x_i \leq \overline{\xi}_i,\ i = 1, \ldots, m, \\ 0 & \text{otherwise.} \end{cases}$$

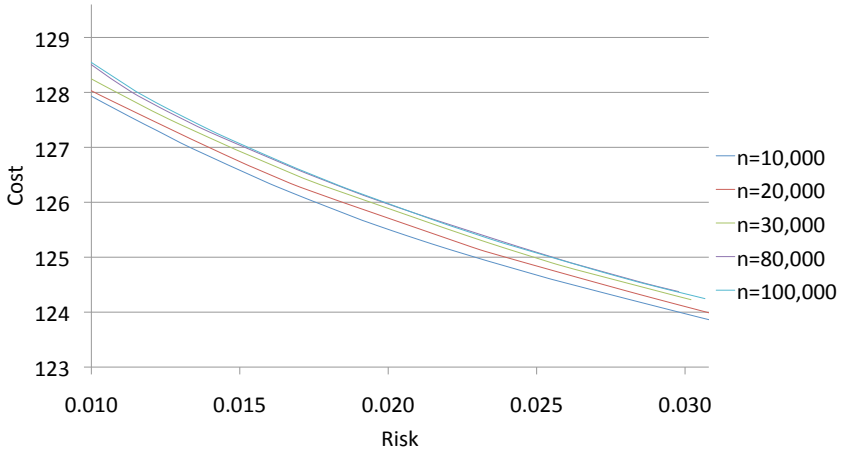

Figure 4.2: Plot showing trajectory of convex envelope of pseudo EF with increasing sample sizes for the facility sizing model. We use 10 facilities, with multivariate normal demand. Demand for each facility is assumed to have mean 10 and unit variance. The correlation coefficient for any pair of facilities is 0.8.

where $f_{\tilde{\xi}}(\cdot)$ is the pdf of the m-variate normal random vector $\tilde{\xi}$ with parameters μ and Σ, and $C = \mathbb{P}(\tilde{\xi}_i \leq \overline{\xi}_i)$, $i = 1, \ldots, m$. We say that $\tilde{\eta}$ obeys a truncated normal distribution. From Corollary 4.5.2, the cdf of $\tilde{\xi}$ is concave in the region $\{x \mid \mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'} \leq x_i \leq \overline{\xi}_i,\ i = 1, \ldots, m\}$. The cdf of $\tilde{\eta}$ in this region is simply a positive multiple of the cdf of $\tilde{\xi}$ in this region, and hence must also be concave. A similar argument holds for the lognormal distribution. For computational purposes however, choosing $\overline{\xi_i} \gg \mu_i + \sqrt{m-1}\, \max_{1 \leq i' \leq m} \sigma_{i'}$ is equivalent to not truncating the normal (or lognormal) distribution, and we dispense with any truncation in our experiments.

We first consider an instance of the problem for both the normal and lognormal distributions. We take $m = 10$ facilities, and the mean is chosen to

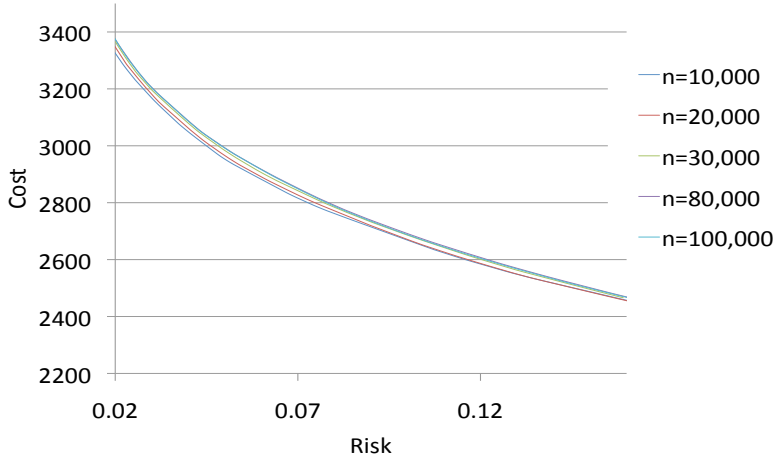

Figure 4.3: Plot showing trajectory of convex envelope of pseudo EF with increasing sample sizes for the facility sizing model. We use 10 facilities, with multivariate lognormal demand. Demand for each facility is assumed to have mean 10 and variance of 1. The correlation coefficient for any pair of facilities is 0.8.

be 10 for all facilities, with unit standard deviation. The correlation coefficient is chosen to be .8 for any distinct pair of facilities. Figures 4.2 and 4.3 show the trajectory of the envelopes for increasing sample sizes. For comparison of running times and confidence bounds with the MIP approach, we then let $m = 40$, with other parameters unchanged, with Figure 4.4 plotting the envelopes. For all cases, we observe that the risk value at the extreme point adjacent to $(1, M)$ is rather small — $\approxeq .03$ for normal, and $\approxeq .16$ for lognormal. For large sample sizes, these upper bounds are indicative of the parameter ϵ' at which the true EF and its convex envelope diverge.

We observe a substantial improvement in solve time over the MIP approach for the second set of computations with the normal distribution. For

a sample size of $n = 250,000$, the solve time was about 20 minutes for the range of epsilon $[0, .04]$, in comparison to 30 minutes to evaluate a single point on the efficient frontier with the MIP approach for $n = 25,000$ realizations. The acceleration of solves is further underlined by the observation that the number of extreme points found by the min-cut algorithm vary between 50 and 300. Overall, solve times were no more than 20 minutes, and no more than 4 minutes for most sample sizes considered.

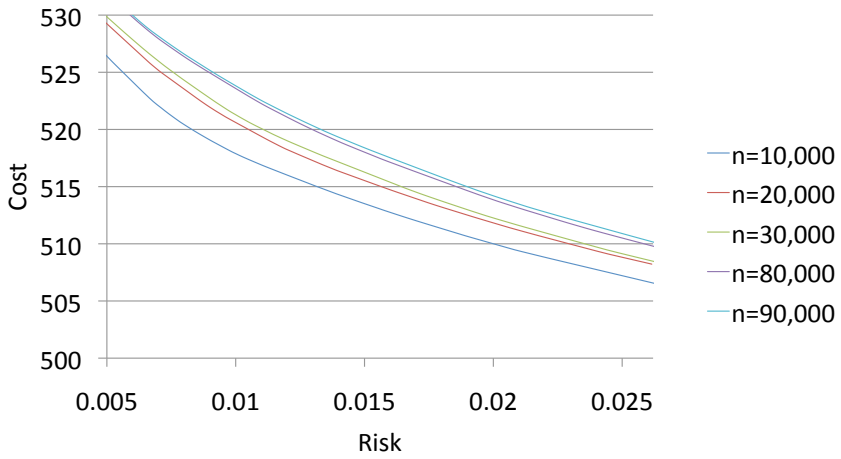

Figure 4.4: Plot showing trajectory of convex envelope of pseudo EF with increasing sample sizes for the facility sizing model. We use 40 facilities, with multivariate normal demand. Demand for each facility is assumed to have mean 10 and variance of 1. The correlation coefficient for any pair of facilities is 0.8.

Table 4.1 reports values of upper and lower confidence bounds on $z_c^*(n, \cdot)$, along with the width of the confidence interval on the optimality gap for the multivariate normal distribution with $m = 40$ facilities and $n = 90000$ realizations. We see that the large sample size makes it possible to lower

Table 4.1: Confidence bounds for $p(x_{\mathrm{p}}^*(n,t))$, $z_{\mathrm{p}}^*(t)$ and the optimality gap $p(x_{\mathrm{p}}^*(n,t)) - z_{\mathrm{p}}^*(t)$, for the model in Figure 4.4 with $n = 90,000$. Note that $z_{\mathrm{p}}^*(n,t) = \epsilon$, and $t = z_{\mathrm{c}}^*(n,\epsilon)$ since model (4.7) is used.

ϵ	$z_{\mathrm{c}}^*(n,\epsilon)$	$z_{\mathrm{p}}^*(n,t) - \epsilon_\ell$	$p_{n'}(x_{\mathrm{p}}^*(n,t)) + \epsilon_u$	Optimality Gap
0.001	546.4	0.001	0.002	0.0009
0.003	537.8	0.003	0.004	0.0011
0.005	531.9	0.005	0.006	0.0013
0.007	527.9	0.006	0.008	0.0015
0.009	524.9	0.008	0.009	0.0014
0.011	522.3	0.010	0.012	0.0014
0.013	520.0	0.012	0.014	0.0014
0.015	518.1	0.014	0.016	0.0015
0.018	515.4	0.017	0.019	0.0016
0.019	514.4	0.019	0.020	0.0016
0.021	512.9	0.021	0.022	0.0017
0.023	511.6	0.022	0.024	0.0017
0.025	510.5	0.024	0.026	0.0018
0.027	509.3	0.026	0.028	0.0018
0.029	508.3	0.028	0.030	0.0017

the magnitude of the width to much less than that obtained with the MIP approach ($\geq .004$).

4.7 Concluding Remarks

In this chapter, we build on work from Chapter 3, with the goal of obtaining fast and convergent approximations to the EF. We assume that the

EF is convex in the range of risk values of interest, in order to form the convex envelope of the pseudo EF as an approximation. We establish that under mild conditions, the sequence of envelopes converges uniformly to the EF as the sample size tends to infinity. We provide computational enhancements for a min-cut formulation of the envelope construction problem, which we show to yield significant savings in computational cost.

Chapter 5

Final Remarks

In this dissertation, we consider probabilistic bicriteria optimization models in which cost of design must be traded off against the risk of operating under uncertainty. Our goal is to form the efficient frontier of solutions, i.e., the set of risk-cost pairs which are Pareto optimal. Our risk objective typically involves a multivariate probability distribution that is difficult to handle analytically, in that closed-form expressions are not easy to derive, and approximations such as quadrature or sampling are required. The focus of our work lies in sampling approximations to the EF, with particular emphasis on sampling costs, convergence and computational efficiency of the approximating model obtained from sampling. When a hard constraint is imposed on risk, it is more commonly referred to as probabilistically-constrained programming, a class of difficult problems. Our work draws upon earlier work in the bicriteria, probabilistic-programming and stochastic programming domains.

In Chapter 2, we assume that a disruption in the model occurs at most once to the horizon, and adopt a scenario approximation approach to solving the model. We assume that achieving a low level of risk in the optimal

design decision is of foremost importance. We construct a convex approximation to the model from a random sample of observations of the uncertainty, and estimate the expected probability of violation for the solution to the approximation. We establish that the savings in sampling costs resulting from using an optimal stratified sampling strategy can be of the order of the time horizon of the model, over naive sampling. Our results are illustrated via a model instance from perishable inventory theory. We also demonstrate with our experiments that stratified sampling results in improved proximity to the efficient frontier on average, compared to naive sampling. One of our significant results in the stratified sampling setting is that for most well-behaved distributions, a polynomial allocation strategy is order-optimal. In contrast, we consider a model in which the distribution is ill-behaved, i.e., it is chosen to make the stratification strategy look as bad as possible. This yields a game-theoretic model which can be solved using an easily implementable algorithm. We extend our results to the case of multiple but a modest number of disruptions, in relation to the number of time periods. We establish savings in sampling costs that are identical to the single disruption model.

Chapter 3 is motivated by the need to evaluate the risk measure near-exactly in the context of a bicriteria probabilistic facility sizing problem. Replacing the risk measure by an empirical average, we form an approximation to the EF by solving a mixed integer program. We show that it is immaterial whether we limit the cost or risk to fall below a prespecified threshold – both models form exactly the same approximation to the EF. We exploit this

property in our experiments, taking advantage of the more obvious parameter settings for the chance-constrained program, as well as its better solvability through the use of valid inequalities and problem size reduction. In contrast to earlier works which provide one-sided bounds on the optimal value, we provide improved assessment of solution quality by forming a confidence interval on the optimality gap of the optimal solution to the sampled program. We see that the optimality gap arises mainly due to the bias associated with the optimal value for the approximation, rather than the suboptimality of its optimal solution. We conclude that the approximating program produces high-quality solutions for modest sample sizes, albeit at the cost of large solve times.

In Chapter 4, we propose enhancements of the earlier integer programming approach to construct the pseudo EF. Working under the assumption that the EF is convex in the range of risk values of interest, we build the convex envelope of the pseudo EF using a polynomial time algorithm from the literature. We prove that the sequence of envelopes so produced, converges uniformly to the EF for the bicriteria problem. While, convex envelopes may be poor approximations in general, we establish that in our setting, the set of risk values for the extreme points of the envelope becomes increasingly dense in the risk range of interest. Finally, we show how the algorithm may be tuned for better performance by reducing the size of the graph inherent to the algorithm. We illustrate our results by means of numerical studies on test problem instances and distributions for which the convexity property holds. Note that we are typically interested in highly reliable systems, i.e., systems in which the

risk value is low. In the work we have presented, we sample directly from the underlying demand distribution. However, as risk values shrink, we may need to alter our sampling strategy using, e.g., importance sampling, to ensure that we sample a sufficient number of demand observations with high values. This is a subject for future research.

An interesting question to consider is: what multivariate distributions can we sample from? Besides multivariate normal and lognormal, common multivariate distributions from which observations are drawn in practice, include the gamma, Student's t and Dirichlet distributions. The author is unaware of concavity properties of the cdf for the first two distributions in this list for large values of the argument. However, a numerical investigation of these distributions in the context of generating envelopes of the pseudo EF, could prove useful in establishing such a property — if the envelopes converge to a limiting function, then this limit function represents the envelope of the true EF, and hence its maximal nonlinear segment around zero risk, is a region of concavity of the cdf.

Bibliography

[1] Z. Artstein and R. J.-B. Wets. Consistency of minimizers and the slln for stochastic programs. *Journal of Convex Analysis*, 2:1–17, 1996.

[2] H. Attouch. *Variational Convergence for Functions and Operators*. Pitman, Boston, 1984.

[3] M. L. Balinski. On a selection problem. *European Journal of Operational Research*, 17:230–231, 1970.

[4] G. Bayraksan and D. P. Morton. Assessing solution quality in stochatic programs. *Mathematical Programming*, 108:495–514, 2006.

[5] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, 2003.

[6] R. Caballero, F. Ruiz, and R. E. Steuer. *Advances in Multiple Objective and Goal Programming*. Springer-Verlag, Berlin, 1997.

[7] G. Calafiore and M. C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005.

[8] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742–753, 2006.

[9] G. Debreu. Valuation equilibrium and Pareto optimum. In *Proceedings of the National Academy of Sciences*, pages 588–592, 1954.

[10] G. Dellino, J. P. C. Kleijnen, and C. Meloni. Robust optimization in simulation: Taguchi and krige combined. In *CentER Discussion Paper Series No. 2009-82*, pages 235–256. Tillburg University, 2009.

[11] G. Dellino, J. P. C. Kleijnen, and C. Meloni. Simulation-optimization under uncertainty through metamodeling and bootstrapping. In *Procedia - Social and Behavioral Sciences*, pages 7640–7641. Sixth International Conference on Sensitivity Analysis of Model Output, 2010.

[12] J. Dong and A. Nagurney. Bicriteria decision making and financial equilibrium: A variational inequality perspective. *Computational Economics*, 17:29–42, 2001.

[13] J. L. Doob. *Measure Theory*. Springer Verlag, New York, 1994.

[14] M. Ehrgott. *Multicriteria Optimization*. Springer Verlag, Berlin, 2000.

[15] M. Ehrgott and X. Gandibleux. *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*. Kluwer Academic Publishers, Norwell, 2002.

[16] M. Ehrgott and M. M. Wiecek. Multiobjective programming. In M. Ehrgott, J. Figueira, and S. Greco, editors, *Multiple Criteria Decision Analysis*, pages 667–722. Springer, Berlin, 2005.

[17] A. M. Geoffrion. Solving bicriterion mathematical programs. *Operations Research*, 15:39–54, 1967.

[18] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *ournal of Mathematical Analysis and Application*, 22:618–630, 1968.

[19] A. V. Goldberg and R. E. Tarjan. A new approach to the maximum flow problem. *Journal of the ACM*, 35:921–940, 1988.

[20] S. K. Goyal and B. C. Giri. Recent trends in modeling of deteriorating inventory. *European Journal of Operational Research*, 134:1–16, 2001.

[21] N. P. Greis, E. F. Wood, and R. E. Steuer. Multicriteria analysis of water allocation in a river basin: The Tchebycheff approach. *Water Resources Research*, 19:865–875, 1983.

[22] D. Hochbaum. Dynamic evolution of economically preferred facilities. *European Journal of Operational Research*, 193:649–659, 2009.

[23] R. Hochreiter. An evolutionary computation approach to scenario-based risk-return portfolio optimization for general risk measures. In *Lecture Notes in Computer Science, Volume 4448*, pages 199–207. Springer, Montpellier, 2007.

[24] I. Kaliszewsky. *Soft Computing for Complex Multiple Criteria Decision Making*. Springer, New York, 2006.

[25] J. Kim and S. Kim. A CHIM-based interactive Tchebycheff procedure for multiple objective decision making. *Computers and Operations Research*, 33:1557–1574, 2006.

[26] J. Luedtke. *Integer Programming Approaches for some Non-convex and Stochastic Optimization Problems*. Ph.D. thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, 2007.

[27] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19:674–699, 2008.

[28] J. Luedtke, S. Ahmed, and G. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 19:247–272, 2010.

[29] W. K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.

[30] D. P. Morton, E. Popova, and I. Popova. Efficient fund of hedge funds construction under downside risk measures. *Journal of Banking and Finance*, 30:503–518, 2006.

[31] D. P. Morton and R. K. Wood. On a stochastic knapsack problem and generalizations. In *Advances in Computational and Stochastic Optimization, Logic Programming and Heuristic Search: Interfaces in Computer*

Science and Operations Research, pages 149–168. Kluwer Academic Publishers, Boston, 1998.

[32] S. Nahmias. Perishable inventory theory: A review. *Operations Research*, 30:680–708, 1982.

[33] M. V. Nehme. *Two-person Games for Stochastic Network Interdiction: Models, Methods and Complexities*. Ph.D. thesis, The University of Texas at Austin, Austin, 2009.

[34] M. V. Nehme and D. P. Morton. Tightening a network interdiction model. In *Proceedings of the 2009 Industrial Engineering Research Conference*, Miami, 2009.

[35] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17:969–996, 2006.

[36] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In *Probabilistic and Randomized Methods for Design Under Uncertainty*, pages 3–48. Springer, London, 2006.

[37] V. I. Norkin, G. Ch. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.

[38] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Computational study of a chance constrained portfolio selection problem. `http://www.optimization-online.org`, 2008.

[39] P. M. Pardalos, Y. Siskos, and C. Zopounidis. *Advances in Multicriteria Analysis*. Kluwer Academic Publishers, Dordrecht, 1995.

[40] J. C. Picard. Maximal closure of a graph and applications to combinatorial problems. *Management Science*, 22:1268–1272, 1976.

[41] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, 1995.

[42] A. Prékopa. On the concavity of multivariate probability distribution functions. *Operations Research Letters*, 29:1–4, 2001.

[43] A. Prékopa. Probabilistic programming. In *Stochastic Programming*, pages 267–351. Elsevier, Amsterdam, 2003.

[44] T. K. Ralphs, M. J. Saltzman, and M. M. Wiecek. An improved algorithm for biobjective integer programming. *Annals of Operations Research*, 147:43–70, 2006.

[45] J. Rhys. Shared fixed costs and network flows. *Management Science*, 17:200–207, 1970.

[46] M. Riis and J. Lodahl. A bicriteria stochastic programming model for capacity expansion in telecommunications. *Mathematical Methods of Operations Research*, 56:83–100, 2002.

[47] A. Ruszczyński and R. J. Vanderbei. Frontiers of stochastically nondominated portfolios. *Econometrica*, 71:1287–1297, 2003.

[48] J. Salmerón, R. K. Wood, and D. P. Morton. A stochastic program for optimizing military sealift subject to attack. *Military Operations Research*, 14:19–39, 2009.

[49] R. Schultz and S. Tiedemann. Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. *SIAM Journal on Optimization*, 14:115–138, 2003.

[50] R. E. Steuer. *Multiple Criteria Optimization: Theory, Application, and Computation*. John Wiley, New York, 1986.

[51] R. E. Steuer and E. Choo. An interactive weighted Tchebycheff procedure for multiple objective programming. *Mathematical Programming*, 26:326–344, 1983.

[52] L. Yang and Y. Feng. A bicriteria solid transportation problem with fixed charge under stochastic environment. *Applied Mathematical Modelling*, 31:2668–2683, 2007.

[53] G. Yu and X. Qi. *Disruption Management: Framework, Models and Applications*. World Scientific, Singapore, 2004.

[54] P. L. Yu. *Multiple-Criteria Decision Making-Concepts, Techniques, and Extensions*. Plenum Press, New York, 1985.

[55] M. Zeleny. *Multiple Criteria Decision Making*. McGraw-Hill, New York, 1982.

Vita

Tara Rengarajan attended the Indian Institue of Technology, Chennai, securing an M.Sc in Mathematics prior to enrolling at the University of Texas at Austin in 1999. She received the M.A. degree in Mathematics from Texas, and then the M.S. degree in Industrial Mathematics from the University of Minnesota, Twin Cities. She moved back to Austin in 2005 and commenced doctoral work in Operations Research and Industrial Engineering at the University of Texas in the spring of 2006. She interned at United Airlines Inc. in 2007, where she worked on a gating assignment problem. She is slated to take up a permanent position in the revenue management industry starting in the fall of 2010. Tara is a trained vocalist and performs South Indian classical music (Carnatic music) in leisure hours. She lives in Austin with husband Sanjay Shakkottai and daughter Diya.

Permanent address: 10703 Keystone Bend
Austin, Texas 78750
trengan@yahoo.com

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.