ESTIMATING THE EFFICIENT FRONTIER OF A PROBABILISTIC BICRITERIA MODEL

Tara Rengarajan
David P. Morton

Graduate Program in Operations Research
The University of Texas at Austin
Austin, TX 78712, U.S.A.

ABSTRACT

We consider a problem that trades off cost of system design with the risk of that design, where risk is measured by the probability of a bad event, such as system failure. Our interest lies in the problem class where we cannot evaluate this risk measure exactly, even for a given system design. We approach this problem via a bicriteria optimization model, replacing the risk measure by an Monte Carlo estimator and solving a parametric family of optimization models to produce an approximate efficient frontier. Optimizing system design with the risk estimator requires solution of a mixed integer program. We show that we can minimize risk over a range of cost thresholds or minimize cost over a range of risk thresholds and we examine associated asymptotics. The proximity of the approximate efficient frontier to the true efficient frontier is established via an asymptotically valid confidence interval with minimal additional work. Our approach is illustrated computationally using a facility-sizing problem.

1   INTRODUCTION

We consider an optimization model in which we must design a system under two conflicting objectives. One goal is to minimize the risk associated with the design, and the other goal is to minimize the design's cost. We cannot evaluate exactly the risk of a design, even when that design is fixed, and constrained minimization of risk is no easier. So, we replace the risk measure with an estimator and construct an approximate, i.e., estimated, efficient frontier, trading off risk and cost. Our risk measure is the probability of a "bad" event such as system failure or failure to have enough capacity to meet demand. Such models are often formulated as probabilistically-constrained stochastic programs, i.e., models in which we minimize cost subject to the probability of the bad event being at most ε, e.g., $\varepsilon = 0.05$. Alternatively, one can minimize risk subject to a constraint on system cost. There is a rich literature on models with probabilistic constraints and objectives; see Prékopa (1995).

We take the view that such problems are naturally bicriteria optimization models. A decision maker may wish to understand the tradeoff between risk and cost, instead of fixing a risk (or cost) threshold a priori, and minimizing cost (or risk). This perspective also has a rich history including Markowitz (1952) trading off risk, as measured by variance of a financial portfolio's return, with that portfolio's mean return. More recent bicriteria optimization in stochastic programming includes that of Ruszczyński and Vanderbei (2003) and Schultz and Tiedemann (2003). In our setting, 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. In this paper we replace the risk measure with a Monte Carlo estimator, and use this estimator in our bicriteria optimization model. Our focus is on constructing an approximate efficient frontier using a sampling-based approximation, and assessing the closeness of that frontier to the true efficient frontier.

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 (Mak, Morton, and Wood 1999, Bayraksan and Morton 2006). This type of approach has been applied when minimizing risk, as measured by probability of a bad event, subject to a cost constraint (Morton and Wood 1998, Morton, Popova, and Popova 2006), 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 (Pagnoncelli, Ahmed, and Shapiro 2008, Luedtke and Ahmed 2008). As we show in this paper, the simpler approach of Bayraksan and Morton (2006) can be applied to our bicriteria model when assessing the closeness of the approximate efficient frontier to the true frontier. Moreover, we can form the approximate efficient frontier 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 efficient frontier.

There is a large literature on bicriteria, and multicriteria, optimization that goes much beyond what we have cited above. See, for example, the books of Ehrgott and Gandibleux (2002), Pardalos, Siskos, and Zopounidis (1995), Steuer (1986), Yu (1985), Zeleny (1982). Central to bicriteria optimization is the well-studied notion of Pareto optimality: Let $h(x)$ and $p(x)$ be two objective functions that we seek to minimize simultaneously, subject to $x \in X$. This vector-minimization problem is written

$$\text{vmin}_{x \in X} \ [h(x), p(x)]. \tag{1}$$

A Pareto optimal solution, $x \in X$, is one for which it is impossible to decrease the value of $h(x)$ without increasing the value of $p(x)$ and vice versa. Solving model (1) amounts to finding a family of decision vectors $x \in X$ with the property that there does not exist a decision vector $x' \in X$ such that $h(x') \leq h(x)$, $p(x') \leq p(x)$ and at least one of these inequalities is strict. Hence, the optimal solution to the problem is a set of Pareto optimal decision vectors. A Pareto optimal solution is also referred to as an *efficient solution* and if x is efficient, the pair $[h(x), p(x)]$ is said to be on the *efficient frontier* of the bicriteria optimization model. Thus, solving a bicriteria model yields the associated efficient frontier. We denote the efficient frontier for the objective functions $h(\cdot)$ and $p(\cdot)$ by $EF(h, p)$. The notion of efficient solutions, and of the associated efficient frontier, extends to models with more than two objectives as detailed in the references cited above. That said, the focus of this paper is restricted to bicriteria optimization.

2 TOWARDS SOLVING A PROBABILISTIC BICRITERIA MODEL

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 ξ_i, and the joint distribution of the random vector $\xi = (\xi_1, \ldots, \xi_m)$ 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 ξ's support, and let $h : X \to \mathbb{R}$ denote cost. Let $G : X \times \Xi \to \mathbb{R}^\ell$, $C \subset \mathbb{R}^\ell$ and let $p(x) = \mathbb{P}(G(x, \xi) \notin C)$ denote risk, i.e., the probability of a bad event. Throughout, we assume h and p are lower semicontinuous. These constructs define model (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:

$$z_c^*(\varepsilon) = \min_{x \in X} \quad h(x) \tag{2a}$$

$$\text{s.t.} \quad \mathbb{P}(G(x, \xi) \notin C) \leq \varepsilon, \tag{2b}$$

where ε is a *risk-level* parameter lying in a range $[\underline{\varepsilon}, \overline{\varepsilon}] \subset [0, 1]$ of interest. Under fairly mild conditions—lower semicontinuity of h plus $X \cap \{x : \mathbb{P}(G(x, \xi) \notin C) \leq \varepsilon\}$ being nonempty and compact—there exists an optimal solution to model (2) that is efficient. Conversely, if x is efficient for model (1) then x is optimal for model (2) when $\varepsilon = p(x)$. In general, there may be optimal solutions to (2) that are not efficient. This can occur when (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 $z_c^*(\varepsilon)$. However, we assume the probabilistic constraint (2b) is tight at all optimal solutions for all $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$, eliminating the need for such postprocessing. So, in principle we can solve model (1) over the range of risk tolerances of interest by solving model (2) for all $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$.

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

$$z_p^*(t) = \min_{x \in X} \quad \mathbb{P}(G(x,\xi) \notin C) \tag{3a}$$

$$\text{s.t.} \quad h(x) \le t. \tag{3b}$$

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 $z_p^*(\bar{t}) = \underline{\varepsilon}$ and $z_p^*(\underline{t}) = \overline{\varepsilon}$.) As above, ranging $t \in [\underline{t}, \bar{t}]$ and solving model (3) yields efficient solutions, and the efficient frontier, provided constraint (3b) 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 efficient frontier, it can be obtained by solving either model (2) when $\varepsilon = p(x)$ or model (3) when $t = h(x)$.

When ξ is multivariate, and the probability distribution and the event $\{G(x,\xi) \notin C\}$ lack special form, solving either model (2) or (3) exactly is impossible. We therefore resort to Monte Carlo sampling to approximate the two models. In other words, we draw n independent and identically distributed (i.i.d.) observations from the distribution of ξ, denoted $\xi^1, \ldots, \xi^n$, and approximate model (2) by

$$z_c^*(n,\varepsilon) = \min_{x \in X} \quad h(x) \tag{4a}$$

$$\text{s.t.} \quad \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(G(x,\xi^j) \notin C) \le \varepsilon, \tag{4b}$$

and model (3) by

$$z_p^*(n,t) = \min_{x \in X} \quad \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(G(x,\xi^j) \notin C) \tag{5a}$$

$$\text{s.t.} \quad h(x) \le t, \tag{5b}$$

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 $\mathbb{I}(G(\cdot,\xi) \notin C)$ is lower semicontinuous on X, w.p.1. (See, e.g., Pagnoncelli, Ahmed, and Shapiro (2008) for conditions under which lower semicontinuity is ensured. Moreover, under the hypotheses of Theorem 1 in the next section, lower semicontinuity of $p(x)$ is ensured.) With the measure thus discretized, models (4) and (5) can be reformulated as mixed integer programs. This said, differences between the two models are evident. It is enough to solve model (4) for the subset of values of $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ such that εn is an integer. In model (5) however, the values of $t \in [\underline{t}, \bar{t}]$ for which constraint (5b) will be tight at an optimal solution are not known a priori. On the other hand, the objective function of model (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 Luedtke, Ahmed, and Nemhauser (2007), Nehme and Morton (2009).

Before proceeding, we briefly review other approaches to forming, or partially forming, the efficient frontier. Two such commonly-used approaches involve a composite weighted objective function and goal programming. In the former approach, we solve

$$\min_{x \in X} \quad \lambda h(x) + (1-\lambda)p(x), \tag{6}$$

for $\lambda \in (0,1)$. Any optimal solution to (6) is an efficient point, but the converse does not hold. There may be points on the efficient frontier that cannot be identified by solving (6). In goal programming, a threshold value is imposed on one of the objective functions. The violation is then penalized in the objective function of the proxy model as shown below

$$\min_{x \in X} \quad \lambda h(x) + (1-\lambda)(p(x) - \varepsilon)^+, \tag{7}$$

for $\lambda \in (0,1)$, where $(\cdot)^+ = \max\{\cdot, 0\}$. Thus, a deviation from the threshold of ε to higher values is penalized. When $p(x)$ measures risk in the probabilistic sense defined above and $\varepsilon = 0$, model (7) reduces to (6). In any case, the goal programming

approach suffers from the same inability to identify *all* points on the efficient frontier. This same issue arises when the deviation of $h(x)$ from a threshold is penalized instead.

Greis, Wood, and Steuer (1983) use the so-called Tchebycheff approach to multiple criteria optimization. Their algorithm uses a result due to Steuer and Choo (1983) that shows how to generate all points on the efficient frontier by varying the cost coefficients of a model that minimizes a Tchebycheff metric. Actual implementation discretizes the space of cost coefficients to obtain a sufficiently dense and representative subset of the efficient frontier. 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 efficient frontier to the true efficient frontier while the Tchebycheff approach instead aims to work interactively with the decision maker.

The use of a probabilistically-constrained program to find the efficient frontier of a related bicriteria optimization model is not new. Since probabilistically-constrained programs are relevant in a wide variety of applications, a sizeable volume of work has appeared along these lines in areas ranging from supply-chain management (Mishra et al. 2008) to portfolio allocation (Sharma, Jana, and Sharma 2008) to transportation (Yang and Feng 2007). In many such works, the probabilistic constraint is sufficiently simple that the approximations of the type we consider are unnecessary. However, we are concerned with the situation where the probabilistic constraint does not admit a closed form. Furthermore, our risk function $p(x)$ is, in general, nonconvex, and so standard nonlinear programming techniques will not ensure an optimal solution.

3   ESTIMATING THE EFFICIENT FRONTIER: AN EQUIVALENCE RESULT

If we could solve model (2) exactly for $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ we would obtain the corresponding points on the efficient frontier $EF(h, p)$ via the pairs $(z_c^*(\varepsilon), \varepsilon)$ for $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$. Alternatively, if we could solve model (3) for $t \in [\underline{t}, \overline{t}]$ we would obtain $EF(h, p)$ via $(t, z_p^*(t))$, $t \in [\underline{t}, \overline{t}]$. Figure 1 depicts such an efficient frontier for the facility-sizing problem sketched at the beginning of Section 2.
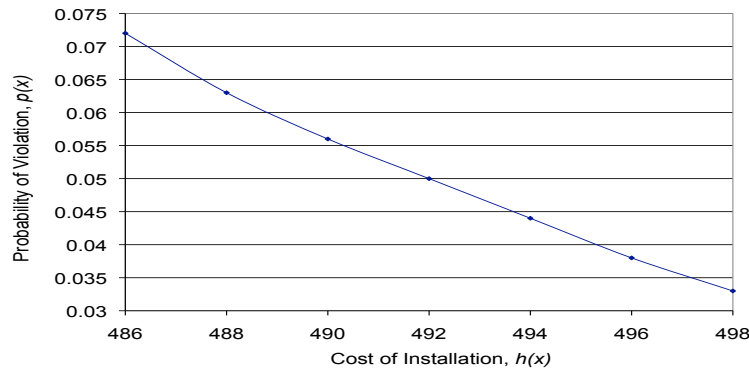

Figure 1: Efficient Frontier for a Capacity Expansion Problem

Let $\varepsilon \in (\underline{\varepsilon}, \overline{\varepsilon})$, fix the i.i.d. sample $\xi^1, \dots, \xi^n$ from the distribution of ξ and consider model (4), where we may assume $\varepsilon = r/n$ for some positive integer r. We assume that model (4) is feasible and that at any optimal solution, constraint (4b) is tight. In general, there may not exist an optimal solution at which (4b) is tight. If this is the case, then we replace r by the largest positive integer smaller than r such that there exists an optimal solution to model (4) with constraint (4b) being tight. Our assumption that (4b) is tight at every optimal solution implies that every optimal solution is on the efficient frontier $EF(h, p_n)$, where $p_n(x) = n^{-1} \sum_{j=1}^n \mathbb{I}(G(x, \xi^j) \notin C)$. In addition to $z_c^*(n, r/n)$ denoting the optimal value of model (4) with $\varepsilon = r/n$, we use $X_c^*(n, r/n)$ to denote model (4)'s set of optimal solutions, 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 (5) with $t = z_c^*(n, r/n)$ and for the *same sample* $\xi^1, \dots, \xi^n$ used to define model (4). Every solution in $X_c^*(n, r/n)$ is feasible for model (5) and has the same objective function value of r/n. If model (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 (5) has an optimal solution at which constraint (5b) is not tight, then the optimal objective function value for model (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 (5).

The above argument establishes the following notion of equivalence between models (4) and (5): Varying ε in the set $E_n = \{\lfloor \underline{\varepsilon} n \rfloor / n, (\lfloor \underline{\varepsilon} n \rfloor + 1)/n, \ldots, \lfloor \overline{\varepsilon} n \rfloor / n\}$ and solving model (4) is equivalent to varying t in the set $T_n = \{z_c^*(n, \varepsilon) \mid \varepsilon \in E_n\}$ and solving model (5). In what follows, we exploit this equivalence between models (4) and (5).

4 ASYMPTOTIC ANALYSIS AND CONSTRUCTION OF CONFIDENCE INTERVALS

4.1 Asymptotics

We cannot construct the efficient frontier $EF(h, p)$ exactly. Instead, we employ Monte Carlo approximations and hence we are interested in asymptotic results as the sample size n grows large for models (4) and (5). Both models are amenable to such analyses, but the analysis of model (5) is arguably simpler, both in the nature of the results we achieve and in the assumptions required to ensure those results. 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 (5), and we can focus on whether such a solution is near-optimal. For these reasons, we restrict attention to asymptotic results for model (5), and we point to Pagnoncelli, Ahmed, and Shapiro (2008) for asymptotic analysis of (4).

The notion of epiconvergence of the approximating objective function to the true objective function, denoted $p_n(x) \xrightarrow{\text{epi}} p(x)$ in the context of models (3) and (5), is central to establishing such results. The following result is a special case of a theorem due to Attouch and Wets (1991), adapted to our setting:

Theorem 1. *Let* $\xi^1, \ldots, \xi^n$ *be i.i.d. from the distribution of* ξ, *assume* $\{x \in X : h(x) \leq t\} \neq \emptyset$, *and assume* $\mathbb{I}(G(\cdot, \xi) \notin C)$ *is lower semicontinuous on X for all* $\xi \in \Xi$. *Then,*

$$p_n(x) \equiv n^{-1} \sum_{j=1}^{n} \mathbb{I}(G(x, \xi^j) \notin C) \xrightarrow{\text{epi}} \mathbb{P}(G(x, \xi) \notin C) \equiv p(x), \text{w.p.1.}$$

Moreover, if $x_p^*(n, t)$ *solves (5) then every limit point of* $\{x_p^*(n, t)\}$ *solves (3), w.p.1. Finally, if* $\{x \in X : h(x) \leq t\}$ *is compact then* $\lim_{n \to \infty} z_p^*(n, t) = z_p^*(t)$, *w.p.1.*

Theorem 1 asymptotically justifies approximating the efficient frontier via model (5) in two senses: (i) Solving (5) with sample size n and budget t, for which constraint (5b) is tight (for all n) yields a solution $x_p^*(n, t)$. Theorem 1 shows that if $\hat{x}_p(t)$ is a limit point of $\{x_p^*(n, t)\}$ then $\hat{x}_p(t)$ solves (3) and hence is efficient, i.e., $(h(\hat{x}_p(t)), p(\hat{x}_p(t)))$ is on $EF(h, p)$, with $h(\hat{x}_p(t)) = t$, w.p.1; (ii) When solving (5) we obtain optimal value $z_p^*(n, t)$ and Theorem 1 states that $z_p^*(n, t)$ converges to $z_p^*(t)$, w.p.1., where $(t, z_p^*(t))$ is on the efficient frontier. Result (i) is of primary importance because of foremost concern is the decision we implement, but we will 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, where $t \in T_n$. When we solve model (5) for finite n and for a range of $t \in T_n$ we obtain $(h(x_p^*(n, t)), p(x_p^*(n, t)))$ and $(t, z_p^*(n, t))$. We refer to the former as the *approximate* efficient frontier and the latter as the *pseudo* efficient frontier. Next, we show how the pseudo and approximate efficient frontiers allow us to form a confidence interval (CI) about $EF(h, p)$. Finally, we know by the equivalence result of Section 3 that when we speak of solving model (5) for $t \in T_n$, we can instead solve model (4) for $\varepsilon \in E_n$.

4.2 Confidence Intervals

The asymptotic results of Theorem 1 justify approximating efficient solutions and the efficient frontier 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) efficient frontier to the true efficient frontier 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), $p(x_p^*(n, t)) - z_p^*(t)$, is small then $x_p^*(n, t)$ is near efficient. To estimate this gap, we first estimate $p(x_p^*(n, t))$ via the sample mean

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

Here, the sample $\xi^1, \ldots, \xi^{n'}$ is i.i.d. from the distribution of ξ and is independent of that used to find $x_p^*(n,t)$. Moreover, we can choose $n' \gg n$ because forming $p_{n'}(x_p^*(n,t))$, given $x_p^*(n,t)$, simply requires a function evaluation of G and the associated feasibility check, i.e., it does not involve solving a mixed-integer program, as is required to form $x_p^*(n,t)$. We use $z_p^*(n,t)$ to estimate $z_p^*(t)$. In addition to Theorem 1's asymptotic result, we know $\mathbb{E}z_p^*(n,t) \leq z_p^*(t)$ (Mak, Morton, and Wood 1999, Norkin, Pflug, and Ruszczyński 1998). So, our point estimate of the optimality gap is $[p_{n'}(x_p^*(n,t)) - z_p^*(n,t)]^+ \geq p_{n'}(x_p^*(n,t)) - z_p^*(n,t)$, and $\mathbb{E}p_{n'}(x_p^*(n,t)) - \mathbb{E}z_p^*(n,t) \geq p(x_p^*(n,t)) - z_p^*(t)$.

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 $x_p^*(n,t)$, z_n^* for $z_p^*(n,t)$ and z^* for $z_p^*(t)$ 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 $n' \to \infty$,

$$\sqrt{n'}[p_{n'}(x_n^*) - p(x_n^*)] \;\Rightarrow\; \sigma N(0,1), \tag{8}$$

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

$$\lim_{n' \to \infty} \mathbb{P}\left(p(x_n^*) \leq p_{n'}(x_n^*) + \underbrace{z_{\alpha/2}\left[p_{n'}(x_n^*)(1 - p_{n'}(x_n^*))/n'\right]^{1/2}}_{\varepsilon_u} \right) = 1 - \frac{\alpha}{2}, \tag{9}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ 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). Using the fact that x^* is feasible, but suboptimal, for model (5), we have $z_n^* \leq p_n(x^*)$, where the same $\xi^1, \ldots, \xi^n$ are used to define these two estimators. Hence,

$$\mathbb{P}\left(z^* \geq z_n^* - z_{\alpha/2}[z_n^*(1 - z_n^*)/n]^{1/2}\right) \geq \mathbb{P}\left(z^* \geq p_n(x^*) - z_{\alpha/2}[z_n^*(1 - z_n^*)/n]^{1/2}\right). \tag{10}$$

Under the hypotheses of Theorem 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 (10), we have

$$\lim_{n \to \infty} \mathbb{P}\left(z^* \geq z_n^* - \underbrace{z_{\alpha/2}[z_n^*(1 - z_n^*)/n]^{1/2}}_{\varepsilon_\ell} \right) \geq \lim_{n \to \infty} \mathbb{P}\left(z^* \geq p_n(x^*) - z_{\alpha/2}[z_n^*(1 - z_n^*)/n]^{1/2}\right) = 1 - \frac{\alpha}{2}. \tag{11}$$

We write "$= 1 - \frac{\alpha}{2}$" in both (9) and (11) 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 (9) and (11) that when n and n' are sufficiently large

$$\mathbb{P}\left(p(x_n^*) - z^* \leq [p_{n'}(x_n^*) - z_n^*]^+ + \varepsilon_\ell + \varepsilon_u\right) \gtrsim 1 - \alpha, \tag{12}$$

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

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

Input: Cost budget t, value α (e.g., $\alpha = 0.10$), sample size n for model (5) and sample size for upper bound n'.

Output: Solution $x_p^*(n,t)$ with approximate $(1 - \alpha)$ CI on its optimality gap.

Step 1.   Sample $\xi^1, \ldots, \xi^n$ i.i.d. from the distribution of ξ. Form and solve (5) to obtain $x_p^*(n,t)$ and $z_p^*(n,t)$.

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

$$p_{n'}(x_p^*(n,t)) = (n')^{-1} \sum_{j=1}^{n'} \mathbb{I}(G(x_p^*(n,t), \xi^j) \notin C) \tag{13}$$

and

$$\varepsilon_\ell = z_{\alpha/2} \left[z_p^*(n,t)(1-z_p^*(n,t))/n \right]^{1/2} \tag{14a}$$

$$\varepsilon_u = z_{\alpha/2} \left[p_{n'}(x_p^*(n,t))(1-p_{n'}(x_p^*(n,t)))/n' \right]^{1/2}, \tag{14b}$$

and output $x_p^*(n,t)$ and a one-sided CI on its optimality gap

$$[p_{n'}(x_p^*(n,t)) - z_p^*(n,t)]^+ + \varepsilon_\ell + \varepsilon_u.$$

Restating our main confidence interval result (12), in our unabridged notation, we have

$$\mathbb{P} \left(p(x_p^*(n,t)) - z_p^*(t) \le [p_{n'}(x_p^*(n,t)) - z_p^*(n,t)]^+ + \varepsilon_\ell + \varepsilon_u \right) \gtrapprox 1 - \alpha. \tag{15}$$

We could use Student t quantiles in (14), 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 efficient frontier. 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 (13) and the sampling error estimates (14).

5 FACILITY-SIZING PROBLEM

We now return to the facility-sizing model sketched at the beginning of Section 2. The cost of installing capacities $x = (x_1, \ldots, x_m)$ is $h(x) = \sum_{i=1}^m c_i x_i$ and the risk of failing to satisfy demand $\xi = (\xi_1, \ldots, \xi_m)$ is $p(x) = \mathbb{P}(\xi \nleq x)$, and our additional constraints on x are simply $X = \{x : x \ge 0\}$. We assume the unit costs coefficients c_i, $i = 1, \ldots, m$, are positive. So, our bicriteria model (1) specializes to

$$\text{vmin}_{x \ge 0} \quad \left[\sum_{i=1}^m c_i x_i, \mathbb{P}(\xi \nleq x) \right]. \tag{16}$$

Models (2) and (3) specialize to

$$z_c^*(\varepsilon) = \min_{x \ge 0} \quad \sum_{i=1}^m c_i x_i \tag{17a}$$

$$\text{s.t.} \quad \mathbb{P}(\xi \nleq x) \le \varepsilon, \tag{17b}$$

and

$$z_p^*(t) = \min_{x \ge 0} \quad \mathbb{P}(\xi \nleq x) \tag{18a}$$

$$\text{s.t.} \quad \sum_{i=1}^m c_i x_i \le t. \tag{18b}$$

The sampling-based approximations, models (4) and (5), become the following mixed-integer programs

$$z_c^*(n,\varepsilon) = \min_{x,y} \quad \sum_{i=1}^m c_i x_i \tag{19a}$$

$$\text{s.t.} \quad \frac{1}{n} \sum_{j=1}^n y_j \le \varepsilon, \tag{19b}$$

$$x_i \ge \xi_i^j (1-y_j), i = 1, \ldots, m, j = 1, \ldots, n, \tag{19c}$$

$$y_j \in \{0,1\}, j = 1, \ldots, n, \tag{19d}$$

and

$$z_{\rm p}^*(n,t) \;=\; \min_{x,y} \quad \frac{1}{n}\sum_{j=1}^{n} y_j \tag{20a}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} c_i x_i \leq t, \tag{20b}$$

$$x_i \geq \xi_i^j(1-y_j), i=1,\ldots,m, j=1,\ldots,n, \tag{20c}$$

$$y_j \in \{0,1\}, j=1,\ldots,n, \tag{20d}$$

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 $x = (x_1,\ldots,x_m)$. Constraints (19c) and (20c) use the fact that demand is nonnegative and these constraints also absorb $x \geq 0$. We note that $p(x) = 1 - \mathbb{P}(\xi \leq x)$ is a continuous function if the random vector ξ has a continuous distribution. In this case, when using model (20) to approximate $EF(h,p)$, we have, in addition to the conclusions of Theorem 1, $\lim_{n\to\infty} p(x_{\rm p}^*(n,t)) = z_{\rm p}^*(t)$, w.p.1.

In Section 3, we showed that models (4) and (5) have the same set of optimal solutions for pairs of $\varepsilon \in E_n$ and $t \in T_n$. A key assumption in establishing this equivalence was that at any optimal solution solution to model (4), constraint (4b) is tight. We now show that this holds for the facility-sizing model (19), provided ξ has a continuous distribution.

Theorem 2. *Consider model (19) with $t > 0$, $c_i > 0$, $i = 1,\ldots,m$, 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 $\varepsilon = r/n$, where $r \in \{0,1,\ldots,n\}$. Then, constraint (19b) is tight at every optimal solution of model (19).*

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 $J^+(x) = \{j : x_i \geq \xi_i^j, i = 1,\ldots,m\}$ 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 (19). So, $y_j^* = 0$ for $j \in J^+(x^*)$ and $y_j^* = 1$ for $j \in J^-(x^*)$, and $|J^-(x^*)| \leq r$. Solution x^* satisfies $x_i^* = \max_{j\in J^+(x^*)} \xi_i^j$: We must have $x_i^* \geq \max_{j\in J^+(x^*)} \xi_i^j$ in order for realization j to be satisfied, and x^* is not optimal if for some i, $x_i^* > \max_{j\in J^+(x^*)} \xi_i^j$ because we can reduce x_i^* to $\max_{j\in J^+(x^*)} \xi_i^j$, remain feasible and decrease the objective function by $c_i\left(x_i^* - \max_{j\in J^+(x^*)} \xi_i^j\right) > 0$.

Constraint (19b) 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 $j(i) = \operatorname{argmax}_{j\in J^+(x^*)} \xi_i^j$. 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

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

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 (19) and decreases the objective function by $c_1(x_1^* - x_1') > 0$. This contradicts optimality of x^* and proves the claim. $\qquad\square$

6 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 ξ is assumed to be multivariate normal with all components having a mean 10, variance 1 and all pairwise correlation coefficients equal to 0.8. Formally, the demand vector is a truncated normal so that it is nonnegative.

While model (20), which is the specialization of model (5), provides the framework for the asymptotic analysis of Section 4, we observed computationally that model (19) solves more quickly than model (20), particularly for large values of n. The results of Section 3 justify employing either model and so we took advantage of model (19) in our computations. We use the tightened reformulation of this mixed integer program, along with valid inequalities, discussed in Nehme and Morton (2009). Finally, we use a sample size of $n' = 200000$ to estimate $p(x_{\rm p}^*(n,t))$. We run the procedure outlined in

Section 4 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 1 and Figure 2 show the values of $z_p^*(n,t)$ and $p_{n'}(x_p^*(n,t))$ for varying values of t. For a given n, the pair $(t, z_p^*(n,t))$ lies on $EF(h, p_n)$, and the tradeoff between the two objectives is observable. For a fixed value of t, we see that $z_p^*(n,t)$ increases with n in most cases. This trend is consistent with that suggested by $\mathbb{E}z_p^*(n,t) \leq \mathbb{E}z_p^*(n+1,t)$ (Mak, Morton, and Wood 1999, Norkin, Pflug, and Ruszczyński 1998). When n is small the point estimate of the optimality gap estimate $[p_{n'}(x_p^*(n,t)) - z_p^*(n,t)]^+$ is relatively large, and this difference shrinks as n grows. Putting aside sampling error for a moment, two factors contribute to this: The suboptimality of solution $x_p^*(n,t)$ in model (18), $p(x_p^*(n,t)) - z_p^*(t)$, and the bias of the lower-bound estimator $z_p^*(t) - \mathbb{E}z_p^*(n,t)$. Of course, we do not know $z_p^*(t)$ but Figure 3 shows how $p_{n'}(x_p^*(n,t))$ tends to shrink and $z_p^*(n,t)$ tends to grow with n for a specific value of t. This plot suggests that $z_p^*(n,t)$ is growing at a faster rate than $p_{n'}(x_p^*(n,t))$ is shrinking, and this indicates bias is likely the dominant contributor. This same effect can be seen in Figure 2, where there tends to be a larger gap between the pseudo efficient frontier plots than those of the approximate efficient frontier, as n grows. As indicated above, the pseudo efficient frontier 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 efficient frontier as this involves an "out-of-sample" assessment, $p_{n'}(x_p^*(n,t))$.

Table 1: Values of $z_p^*(n,t)$ and $p_{n'}(x_p^*(n,t))$ for varying values of cost of installation t

| t	$n=500$		$n=1000$		$n=2000$		$n=5000$		$n=10000$		$n=25000$	
	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$	$z_p^*(n,t)$	$p_{n'}(x_p^*(n,t))$
486	0.062	0.111	0.072	0.108	0.081	0.106	0.089	0.102	0.095	0.1	0.096	0.098
488	0.052	0.101	0.063	0.1	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
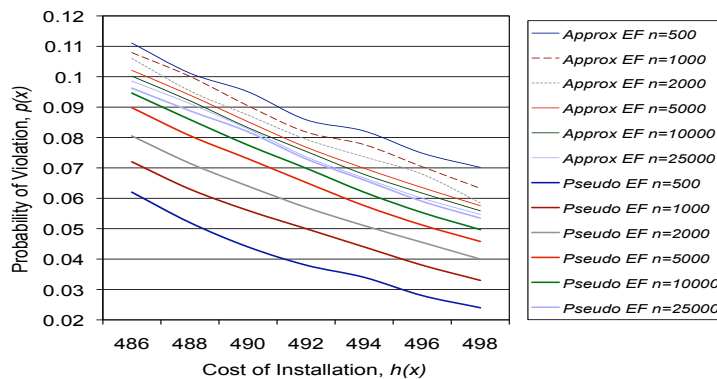
Figure 2: Approximate and Pseudo Efficient Frontiers for Varying Sample Sizes n

Approximate 90% confidence bounds are reported in Table 2 for $n = 25000$ and $n' = 200000$. Sampling error associated with $p_{n'}(x_p^*(n,t))$, ε_u, is about 0.001 while that for $z_p^*(n,t)$, 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, $[p_{n'}(x_p^*(n,t)) - z_p^*(n,t)]^+$. However, for $n = 25000$, i.e., for the values listed in Table 2 this point estimate of the optimality gap and the sampling error are of comparable magnitude.

Table 2: Confidence bounds for $p(x_p^*(n,t))$, $z_p^*(t)$ and the optimality gap $p(x_p^*(n,t)) - z_p^*(t)$, when $n = 25000$

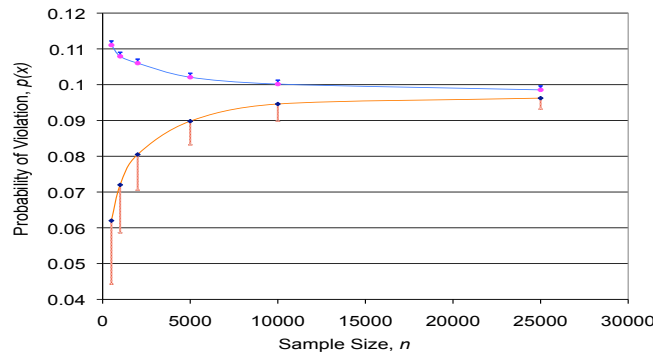| t	$z_p^*(n,t) - \varepsilon_\ell$	$p_{n'}(x_p^*(n,t)) + \varepsilon_u$	Optimality Gap
486	0.093	0.099	0.006
488	0.086	0.092	0.006
490	0.079	0.084	0.005
492	0.070	0.075	0.005
494	0.063	0.068	0.005
496	0.057	0.061	0.004
498	0.051	0.055	0.004

Figure 3: Approximation of Efficient Frontier for $t = 486$: Lower Line is $z_p^*(n,t)$ and Upper Line is $p_{n'}(x_p^*(n,t))$ with $n' = 200000$

Figure 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 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 $p_{n'}(x_p^*(n,t))$.

7 CONCLUDING REMARKS

We have examined a class of bicriteria optimization problems that trade-off risk and cost of system design. Risk is measured as the probability of a bad event, such as system failure or inability of the system to meet random demand. The paper has focused on the situation where we cannot evaluate this risk measure exactly even for a specified system design. Optimization of that risk measure is no easier. We estimated the risk measure via Monte Carlo sampling and proposed solving a parametric family of optimization models to produce an approximate efficient frontier. In general, asymptotics, and solution-quality assessment, associated with probabilistically-constrained models is more difficult than that for a model with a probabilistic objective function and deterministic constraints. In the bicriteria setting, we show an equivalence result between minimizing risk over a range of cost thresholds and minimizing cost over a range of risk thresholds. We establish the proximity of the approximate efficient frontier to the true efficient frontier via an asymptotically valid confidence interval with minimal additional work. We illustrated our approach using a facility-sizing problem.

REFERENCES

Attouch, H., and R. J.-B. Wets. 1991. Epigraphical processes: laws of large nunbers for random lsc functions. In *Séminaire d'Analyse Convexe*, 235–256. Montpellier.

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

Ehrgott, M., and X. Gandibleux. (Eds.) 2002. *Multiple criteria optimization: State of the art annotated bibliographic surveys*. Kluwer Academic Publishers, Norwell.

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

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

Luedtke, J., S. Ahmed, and G. Nemhauser. 2007. An integer programming approach for linear programs with probabilistic constraints. *http://www.optimization-online.org*. To appear in *Mathematical Programming*.

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

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.

Mishra, K., R. D. Gudi, S. C. Patwardhan, and G. Sardar. 2008. Midterm supply chain planning under uncertainty: a multiobjective chance constrained programming framework. *Industrial and Engineering Chemistry Research* 47:5501–5511.

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

Morton, D. P., and R. K. Wood. 1998. 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*, ed. D. Woodruff, 149–168. Kluwer Academic Publishers, Boston.

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

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

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

Pardalos, P. M., Y. Siskos, and C. Zopounidis. 1995. *Advances in multicriteria analysis*. Springer.

Prékopa, A. 1995. *Stochastic programming*. Kluwer Academic Publishers, Dordrecht.

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

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

Sharma, H. P., R. K. Jana, and D. K. Sharma. 2008. Chance constrained programming for security analysis and portfolio management. In *39th Annual Meeting of Decision Sciences Institute*. Baltimore, Maryland: Institute of Electrical and Electronics Engineers, Inc.

Steuer, R. E. 1986. *Multiple criteria optimization: Theory, application, and computation*. John Wiley, New York.

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

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

Yu, P. L. 1985. *Multiple-criteria decision making-concepts, techniques, and extensions*. Plenum Press, New York.

Zeleny, M. 1982. *Multiple criteria decision making*. McGraw-Hill, New York.

AUTHOR BIOGRAPHIES

TARA RENGARAJAN is a PhD student in the Graduate Program in Operations Research in the Mechanical Engineering Department at The University of Texas at Austin. Her research interests include simulation-based chance-constrained programming and mixed integer programming. Her email address is <tara_rengarajan@mail.utexas.edu>.

DAVID P. MORTON is Engineering Foundation Professor in the Graduate Program in Operations Research in the Mechanical Engineering Department at The University of Texas at Austin. His research interests include computational stochastic programming, including simulation-based approximations in stochastic programming. His email address is <morton@mail.utexas.edu>, and his web page is <http://www.me.utexas.edu/~orie/Morton.html>.