

Algorithms for Probabilistically-Constrained Models of Risk-Averse Stochastic Optimization with Black-Box Distributions

Chaitanya Swamy*

Abstract

We consider various stochastic models that incorporate the notion of risk-averseness into the standard 2-stage recourse model, and develop novel techniques for solving the algorithmic problems arising in these models. A key notable feature of our work that distinguishes it from work in some other related models, such as the (standard) budget model and the (demand-) robust model, is that we obtain results in the *black-box* setting, that is, where one is given *only* sampling access to the underlying distribution. Our first model, which we call the *risk-averse budget model*, incorporates the notion of risk-averseness via a *probabilistic constraint* that restricts the probability (according to the underlying distribution) with which the second-stage cost may exceed a given budget B to at most a given input threshold ρ . We also consider a closely-related model that we call the *risk-averse robust model*, where we seek to minimize the first-stage cost and the $(1 - \rho)$ -quantile (according to the distribution) of the second-stage cost.

We obtain approximation algorithms for a variety of combinatorial optimization problems including the set cover, vertex cover, multicut on trees, min cut, and facility location problems, in the risk-averse budget and robust models with *black-box distributions*. We first devise a *fully polynomial approximation scheme* for solving the *LP-relaxations* of a wide-variety of risk-averse budgeted problems. Complementing this, we give a rounding procedure that lets us use existing LP-based approximation algorithms for the 2-stage stochastic and/or deterministic counterpart of the problem to round the fractional solution. Thus, we obtain near-optimal solutions to risk-averse problems that preserve the budget approximately and incur a small blow-up of the probability threshold (both of which are unavoidable). To the best of our knowledge, these are the *first approximation results* for problems involving *probabilistic constraints and black-box distributions*. Our results extend to the setting with non-uniform scenario-budgets, and to a generalization of the risk-averse robust model, where the goal is to minimize the sum of the first-stage cost and a weighted combination of the expectation and the $(1 - \rho)$ -quantile of the second-stage cost.

1 Introduction

Stochastic optimization models provide a means to model uncertainty in the input data where the uncertainty is modeled by a probability distribution over the possible realizations of the actual data, called *scenarios*. Starting with the work of Dantzig [10] and Beale [2] in the 1950s, these models have found increasing application in a wide variety of areas; see, e.g., [4, 35] and the references therein. An important and widely-used model in stochastic programming is the *2-stage recourse model*: first, given the underlying distribution over scenarios, one may take some *first-stage* actions to construct an anticipatory part of the solution, x , incurring an associated cost $c(x)$. Then, a scenario A is realized according to the distribution, and one may take additional *second-stage recourse actions* y_A incurring a certain cost $f_A(x, y_A)$. The goal in the standard 2-stage model is to minimize the total expected cost, $c(x) + \mathbb{E}_A[f_A(x, y_A)]$. Many applications come under this setting. An oft-cited motivating example is the *2-stage stochastic facility location problem*. A company has to decide where to set up its facilities to serve client demands. The demand-pattern is not

*cswamy@math.uwaterloo.ca. Dept. of Combinatorics and Optimization, Univ. Waterloo, Waterloo, ON N2L 3G1. Supported in part by NSERC grant 32760-06.

known precisely at the outset, but one does have some statistical information about the demands. The first-stage decisions consist of deciding which facilities to open initially, given the distributional information about the demands; once the client demands are realized according to this distribution, we can extend the solution by opening more facilities, incurring their recourse costs. The recourse costs are usually higher than the original ones (e.g., because opening a facility later involves deploying resources with a small lead time), could be different for the different facilities, and could even depend on the realized scenario.

A common criticism of the standard 2-stage model is that the expectation measure fails to adequately measure the “*risk*” associated with the first-stage decisions: two solutions with the same expected cost are valued equally. But in realistic settings, one also considers the risk involved in the decision. For example, in the stochastic facility location problem, given two solutions with the same expected cost, one which incurs a moderate second-stage cost in all scenarios, and one where there is a non-negligible probability that a “disaster scenario” with a huge associated cost occurs, a company would naturally prefer the former solution.

Our models and results. We consider various stochastic models that incorporate the notion of risk-averseness into the standard 2-stage model and develop novel techniques for solving the algorithmic problems arising in these models. A key notable feature of our work that distinguishes it from work in some other related models [19, 11], is that we obtain results in the *black-box* setting, that is, where one is given *only* sampling access to the underlying distribution. To better motivate our models, we first give an overview of some related models considered in the approximation-algorithms literature that also embody the idea of risk-protection, and point out why these models are ill-suited to the design of algorithms in the black-box setting.

One simple and natural way of providing some assurance against the risk due to scenario-uncertainty is to provide bounds on the second-stage cost incurred in each scenario. Two closely related models in this vein are the *budget model*, considered by Gupta, Ravi and Sinha [19], and the (*demand-*) *robust model*, considered by Dhamdhere, Goyal, Ravi and Singh [11]. In the budget model, one seeks to minimize the expected total cost subject to the constraint that the second-stage cost $f_A(x, y_A)$ incurred in every scenario A be at most some input budget B . (In general, one could have a different budget B_A for each scenario A , but for simplicity we focus on the uniform-budget model.) Gupta et al. considered the budget model in the *polynomial scenario* setting, where one is given explicitly a list of all scenarios (with non-zero probability) and their probabilities, thereby restricting their attention to distributions with a polynomial-size support. In the robust model considered by Dhamdhere et al. [11], which is more in the spirit of robust optimization, the goal is to minimize $c(x) + \max_A f_A(x, y_A)$. It is easy to see how the two models are related: if one “guesses” the maximum second-stage cost B incurred by the optimum, then the robust problem essentially reduces to the budget problem with budget B , except that the second-stage cost term in the objective function is replaced by B (which is a constant). Notice that it is not clear how to even *specify* problems with exponentially many scenarios in the robust model. Feige et al. [14] expanded the model of [11] by considering exponentially many scenarios, where the scenarios are implicitly specified by a cardinality constraint. However, considering scenario-collections that are determined only by a cardinality constraint seems rather specialized and somewhat artificial, especially in the context of stochastic optimization; e.g., in facility location, it is rather stylized (and overly conservative) to assume that *every* set of k clients (for some k) may show up in the second-stage. We will consider a more general way of specifying (exponentially many) scenarios in robust problems, where the input specifies a black-box distribution and the collection of scenarios is then given by the support of this distribution. We shall call this model the *distribution-based* robust-model.

Both the budget model and the (distribution-based) robust model suffer from certain common drawbacks. A serious *algorithmic limitation* of both these models (see Section 5) is that for almost any (non-trivial) stochastic problem (such as *fractional* stochastic set cover with at most 3 elements, 3 sets, 3 scenarios), one cannot obtain *any* approximation guarantees in the black-box setting using any *bounded number of samples* (even allowing for a bounded violation of the budget in the budget model). Intuitively, the reason for this is that there could be scenarios that occur with vanishingly small probability that one will almost never

encounter in our samples, but which essentially force one to take certain first-stage actions in order to satisfy the budget constraints in the budget model, or to obtain a low-cost solution in the robust model. Notice also that both the budget and robust models adopt the conservative view that one needs to bound the second-stage cost in *every* scenario, regardless of *how likely* it is for the scenario to occur. (By the same token, they also provide the greatest amount of risk-aversion.) In contrast, many of the risk-models considered in the finance and stochastic-optimization literature, such as the mean-risk model [27], value-at-risk (VaR) constraints [30, 23, 32], conditional VaR [34], do factor in the probabilities of different scenarios.

Our models for risk-averse stochastic optimization address the above issues, and significantly refine and extend the budget and robust models. Our goal is to come up with a model that is sufficiently rich in modeling power to allow for black-box distributions, *and* in which one can obtain strong algorithmic results. Our models are motivated by the observation (see Appendix A) that it *is* possible to obtain approximation guarantees in the budget model with black-box distributions, if one allows the second-stage cost to exceed the budget with some “small” probability ρ (according to the underlying distribution). We can turn this solution concept around and incorporate it into the model to arrive at the following. We are now *given* a probability threshold $\rho \in [0, 1]$. In our new budget model, which we call the *risk-averse budget model*, given a budget B , we seek $(x, \{y_A\})$ so as to minimize $c(x) + \mathbb{E}_A[f_A(x, y_A)]$ subject to the *probabilistic constraint* $\Pr_A[f_A(x, y_A) > B] \leq \rho$. The corresponding *risk-averse (distribution-based) robust model* seeks to minimize $c(x) + Q_\rho[f_A(x, y_A)]$, where $Q_\rho[f_A(x, y_A)]$ is the $(1 - \rho)$ -quantile of $\{f_A(x, y_A)\}_{A \in \mathcal{A}}$, which is the smallest number B such that $\Pr_A[f_A(x) > B] \leq \rho$. Notice that the parameter ρ allows us to control the risk-aversion level and tradeoff risk-averseness against conservatism (in the spirit of [3, 41]). Taking $\rho = 1$ in the risk-averse budget model gives the standard 2-stage recourse model, whereas taking $\rho = 0$ in the risk-averse budget- or robust-models recovers the standard budget- and robust models respectively. In the sequel, we treat ρ as a constant that is not part of the input.

We obtain approximation algorithms for a variety of combinatorial optimization problems (Section 4) including the set cover, vertex cover, multicut on trees, min cut, and facility location problems, in the risk-averse budget and robust models with *black-box distributions*. We obtain near-optimal solutions that preserve the budget approximately and incur a small blow-up of the probability threshold. (One should expect to violate the budget here; otherwise, by setting very high first-stage costs, one would be able to solve the decision version of an *NP-hard* problem!) To the best of our knowledge, these are the *first approximation results* for problems with *probabilistic constraints and black-box distributions*. Our results extend to the setting with non-uniform scenario-budgets, and to a generalization of the risk-averse robust model, where the goal is to minimize $c(x)$ plus a weighted combination of $\mathbb{E}_A[f_A(x, y_A)]$ and $Q_\rho[f_A(x, y_A)]$. In the sequel, we focus primarily on the risk-averse budget model since results obtained this model essentially translate to the risk-averse robust model (the budget-violation can be absorbed into the approximation ratio).

Our results are built on two components. First, and this is the technically more difficult component and our main contribution, we devise a *fully polynomial approximation scheme* for solving the LP-relaxations of a wide-variety of risk-averse problems (Theorem 3.3). We show that in the black-box setting, for a wide variety of 2-stage problems, for any $\epsilon, \kappa > 0$, in time $\text{poly}\left(\frac{\lambda}{\epsilon\kappa\rho}\right)$, one can compute (with high probability) a solution to the LP-relaxation of the risk-averse budgeted problem, of cost at most $(1 + \epsilon)$ times the optimum where the probability that the second-stage cost exceeds the budget B is at most $\rho(1 + \kappa)$. Here λ is the maximum ratio between the costs of the same action in stage II and stage I (e.g., opening a facility or choosing a set). We show in Section 5 that the dependence on $\frac{1}{\kappa\rho}$, and hence, the violation of the probability-threshold, is unavoidable in the black-box setting. We believe that this is a general tool of independent interest that will find application in the design of approximation algorithms for other discrete risk-averse stochastic optimization problems, and that our techniques will find use in solving other probabilistic programs.

The second component is a simple rounding procedure (Theorem 3.2) that complements (and motivates) the above approximation scheme. As we mention below, our LP-relaxation is a relaxation of even the *fractional risk-averse problem* (i.e., where one is allowed to take fractional decisions). We give a general

rounding procedure to convert a solution to our LP-relaxation to a solution to the fractional risk-averse problem losing a certain factor in the solution cost, budget, and the probability of budget-violation. This allows us to then use an LP-based “local” approximation algorithm for the corresponding 2-stage problem to obtain an integer solution, where a local algorithm is one that approximately preserves the LP-cost of each scenario. In particular, for various covering problems, one can use the local $2c$ -approximation algorithm in [38], which is obtained using an LP-based c -approximation algorithm for the deterministic problem.

We need to overcome various obstacles to devise our approximation scheme. The first difficulty faced in solving a probabilistic program such as ours, is that the feasible region of *even the fractional problem* is a non-convex set. Thus, even in the *polynomial-scenario* setting, it is not clear how to solve (even) the fractional risk-averse problem. (In contrast, in the standard 2-stage recourse model, the fractional problem can be easily formulated and solved as a linear program (LP) in the polynomial-scenario setting.) We formulate an LP-relaxation (which is also a relaxation of the fractional problem), where we introduce a variable r_A for every scenario A that is supposed to indicate whether the budget is exceeded in scenario A . Correspondingly, we have two sets of decision variables to denote the decisions taken in scenario A in the two cases respectively where the budget is exceeded and where it is not exceeded. The constraints that enforce this semantics will of course be problem-specific, but a common constraint that figures in all these formulations is $\sum_A p_A r_A \leq \rho$, which captures our probabilistic constraint. This constraint, which *ouples* the different scenarios, creates significant challenges in solving the LP-relaxation. (Again, notice the contrast with the standard 2-stage recourse model.) We get around the difficulty posed by this coupling constraint by taking the *Lagrangian dual* with respect to this constraint, introducing a dual variable $\Delta \geq 0$. The resulting maximization problem (over Δ) has a 2-stage minimization LP embedded inside it. Although this 2-stage LP does not belong to the class of problems defined in [38, 45, 7], we prove that for any fixed Δ , this 2-stage LP can be solved to “near-optimality” using the *sample average approximation* (SAA) method. The crucial insight here is to realize that for the purpose of obtaining a near-optimal solution to the risk-averse LP, it suffices to obtain a rather *weak* guarantee for the 2-stage LP, where we allow for an additive error proportional to Δ . This guarantee is specifically tailored so that it is weak enough that one can prove such a guarantee by showing “closeness-in-subgradients” and the analysis in [45], and yet can be leveraged to obtain a near-optimal solution to (the relaxation of) our risk-averse problem. Given this guarantee, we show that one can efficiently find a suitable value for Δ such that the solution obtained for this Δ (via the SAA method) satisfies the desired guarantees.

Related work. Stochastic optimization is a field with a vast amount of literature; we direct the reader to [4, 30, 35] for more information on the subject. We survey the work that is most relevant to our work. Stochastic optimization problems have only recently been studied from an approximation-algorithms perspective. A variety of approximation results have been obtained in the 2-stage recourse model, but more general models, such as risk-optimization or probabilistic-programming models have received little or no attention.

The (standard) budget model was first considered by Gupta et al. [19], who designed approximation algorithms for stochastic network design problems in this model. Dhamdhere et al. [11] introduced the demand-robust model (which we call the robust model), and obtained algorithms for the robust versions of various combinatorial optimization problems; some of their guarantees were later improved by Golovin et al. [16]. All these works focus on the polynomial-scenario setting. Feige, Jain, Mahdian, and Mirrokni [14] considered the robust model with exponentially many scenarios that are specified implicitly via a cardinality constraint, and derived approximation algorithms for various covering problems in this more general model.

There is a large body of work in the finance and stochastic-optimization literature, dating back to Markowitz [27], that deals with risk-modeling and optimization; see e.g., [34, 1, 36] and the references therein. Our risk-averse models are related to some models in finance. In fact, the probabilistic constraint that we use is called a *value-at-risk* (VaR) constraint in the finance literature, and its use in risk-optimization is quite popular in finance models; it has even been written into some industry regulations [23, 32].

Problems involving probabilistic constraints are called *probabilistic* or *chance-constrained programs* [8, 29] in the stochastic-optimization literature, and have been extensively studied (see, e.g., Prékopa [31]). Recent work in this area [6, 28, 13] has focused on replacing the original probabilistic constraint by more tractable constraints so that any solution satisfying the new constraints also satisfies the original probabilistic constraint with high probability. Notice that this type of “relaxation” is *opposite* to what one aims for in the design of approximation algorithms, where we want that every solution to the original problem remains a solution to the relaxation (but most likely, not vice versa). Although some approximation results in the opposite direction are obtained in [6, 28, 13], they are obtained for very structured constraints of the type $\Pr_{\xi}[G(x, \xi) \notin C] \leq \rho$, where C is a convex set, ξ is a continuous random variable whose distribution satisfies a certain concentration-of-measure property, and $G(\cdot)$ is a bi-affine or convex mapping; also the bounds obtained involve a relatively large violation of the probability threshold (compared to our $(1 + \kappa)$ -factor). To the best of our knowledge, there is no prior work in the stochastic-optimization or finance literature on the design of *efficient algorithms* with *provable worst-case guarantees* for discrete risk-optimization or probabilistic-programming problems. In the Computer Science literature, [24] and [15] consider the stochastic bin packing and knapsack problems with probabilistic constraints that limit the overflow probability of a bin or the knapsack, and obtained novel approximation algorithms for these problems. Their results are however obtained for *specialized distributions* where the item sizes are *independent* random variables following Bernoulli, exponential, or Poisson distributions specified in the input. In the context of stochastic optimization, this constitutes a rather stylized setting that is far from the black-box setting.

The work closest in spirit to ours is that of So, Zhang, and Ye [41]. They consider the problem of minimizing the first-stage cost plus a risk-measure called the *conditional VaR* (CVaR) [34]. Their model interpolates between the 2-stage recourse model and the (standard) robust model (as opposed to the budget model in our case). They give an approximation scheme for solving the LP-relaxations of a broad class of problems in the black-box setting, using which they obtain approximation algorithms for certain discrete optimization problems. Our methods are however quite different from theirs. In their model, the fractional problem yields a *convex program* and moreover, they are able to use a nice representation theorem in [34] for the CVaR measure to convert their problem into a 2-stage problem and then adapt the methods in [7]. In our case, the non-convexity inherent in the probabilistic constraint creates various difficulties (first the non-convexity, then the coupling constraint) and we consequently need to work harder to obtain our result. We remark that our techniques can be used to solve a *generalization of their model*, where we have the same objective function but also include a probabilistic budget constraint as in our risk-averse budget model.

We now briefly survey the approximation results in recourse models. The first such approximation result appears to be due to Dye, Stougie, and Tomasgard [12]. The recent interest and flurry of algorithmic activity in this area can be traced to the work of Ravi and Sinha [33] and Immorlica, Karger, Minkoff and Mirrokni [22], which gave approximation algorithms for the 2-stage variants of various discrete optimization problems in the polynomial scenario [33, 22] and *independent-activation* [22] settings. Approximation algorithms for 2-stage problems with black-box distributions were first obtained by Gupta, Pál, Ravi and Sinha [17], and subsequently by Shmoys and Swamy [38] (see also preliminary version [39]). Various other approximation results for 2-stage problems have since been obtained; see, e.g., the survey [44]. Multistage recourse problems in the black-box model were considered by [18, 45]; both obtain approximation algorithms with guarantees that deteriorate with the number of stages, either exponentially [18] (except for multistage Steiner tree which was also considered in [20]), or linearly [45]; improved guarantees for set cover and vertex cover have been subsequently obtained [42].

Our approximation scheme makes use of the SAA method, which is a simple and appealing method for solving stochastic problems that is quite often used in practice. In the SAA method one samples a certain number of scenarios to estimate the scenario probabilities by their frequency of occurrence, and then solves the 2-stage problem determined by this approximate distribution. The effectiveness of this method depends on the sample size (ideally, polynomial) required to guarantee that an optimal solution to the SAA-problem

is a provably near-optimal solution to the original problem. Kleywegt et al. [25] (see also [37]) prove a bound that depends on the variance of a certain quantity that need not be polynomially bounded. Subsequently, Swamy and Shmoys [45], and Charikar et al. [7] obtained improved (polynomial) sample-bounds for a large class of structured 2-stage problems. The proof in [45], which also applies to multistage programs, is based on leveraging approximate subgradients, and our proof makes use of portions of their analysis. The proof of Charikar et al. [7] is quite different; it applies to 2-stage programs but proves the stronger theorem that even approximate solutions to the SAA problem translate to approximate solutions to the original problem.

2 Preliminaries

Let \mathbb{R}_+ denote $\mathbb{R}_{\geq 0}$. Let $\|u\|$ denote the ℓ_2 norm of u . The *Lipschitz constant* of a function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is the smallest K such that $|f(x) - f(y)| \leq K\|x - y\|$. We consider convex minimization problems $\min_{x \in \mathcal{P}} f(x)$, where $\mathcal{P} \subseteq \mathbb{R}_+^m$ with $\mathcal{P} \subseteq B(\mathbf{0}, R) = \{x : \|x\| \leq R\}$ for a suitable R , and f is convex.

Definition 2.1 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be a function. We say that $d \in \mathbb{R}^m$ is a subgradient of f at the point u if the inequality $f(v) - f(u) \geq d \cdot (v - u)$ holds for every $v \in \mathbb{R}^m$. We say that \hat{d} is an (ω, ξ) -subgradient of f at the point $u \in \mathcal{P}$ if for every $v \in \mathcal{P}$, we have $f(v) - f(u) \geq \hat{d} \cdot (v - u) - \omega f(v) - \omega f(u) - \xi$.*

The above definition of an (ω, ξ) -subgradient is slightly weaker than the notion of an ω -subgradient as defined in [38], where one requires that $f(v) - f(u) \geq d \cdot (v - u) - \omega f(u)$. But this difference is superficial; one could also implement the algorithm in [38] using the weaker notion of an (ω, ξ) -subgradient. It is well known (see [5]) that a convex function has a subgradient at every point. One can infer from Definition 2.1 that, letting d_x denote a subgradient of f at x , the Lipschitz constant of f is at most $\max_x \|d_x\|$.

Let K be a positive number, and τ, ϱ be two parameters with $\tau < 1$. Let $N = \log(\frac{2KR}{\tau})$. Let $G'_\tau \subseteq \mathcal{P}$ be a discrete set such that for any $x \in \mathcal{P}$, there exists $x' \in G'_\tau$ with $\|x - x'\| \leq \frac{\tau}{KN}$. Define $G_\tau = G'_\tau \cup \{x + t(y - x), y + t(x - y) : x, y \in G'_\tau, t = 2^{-i}, i = 1, \dots, N\}$. We call G_τ and G'_τ , an $\frac{\tau}{KN}$ -net and an *extended* $\frac{\tau}{KN}$ -net respectively of \mathcal{P} . As shown in [45], if \mathcal{P} contains a ball of radius V (where $V \leq 1$ without loss of generality), then one can construct G'_τ so that $|G'_\tau| = \text{poly}(\log(\frac{KR}{V\tau}))$. As mentioned earlier, our algorithms make use of the sample average approximation (SAA) method. The following result from Swamy and Shmoys [45], which we have adapted to our setting, will be our main tool for analyzing the SAA method.

Lemma 2.2 ([45]) *Let \hat{f} and f be two nonnegative convex functions with Lipschitz constant at most K such that at every point $x \in G_\tau$, there exists a vector $\hat{d}_x \in \mathbb{R}^m$ that is a subgradient of $\hat{f}(\cdot)$ and an $(\frac{\varrho}{8N}, \xi)$ -subgradient of $f(\cdot)$ at x . Let $\hat{x} = \text{argmin}_{x \in \mathcal{P}} \hat{f}(x)$. Then, $f(\hat{x}) \leq (1 + \varrho) \min_{x \in \mathcal{P}} f(x) + 6\tau + 2N\xi$.*

Lemma 2.3 (Chernoff-Hoeffding bound [21]) *Let X_1, \dots, X_N be iid random variables with each $X_i \in [0, 1]$ and $\mu = \mathbb{E}[X_i]$. Then, $\Pr[|\frac{1}{N} \sum_i X_i - \mu| > \epsilon] \leq 2e^{-2\epsilon^2 N}$.*

3 The risk-averse budgeted set cover problem: an illustrative example

Our techniques can be used to efficiently solve the risk-averse versions of a variety of 2-stage stochastic optimization problems, both in the risk-averse budget and robust models. In this section, we illustrate the main underlying ideas by focusing on the risk-averse budgeted set cover problem. In the risk averse budgeted set cover problem (RASCS), we are given a universe U of n elements and a collection \mathcal{S} of m subsets of U . The set of elements to be covered is uncertain: we are given a probability distribution $\{p_A\}_{A \in \mathcal{A}}$ of scenarios, where each scenario A specifies a subset of U to be covered. The cost of picking a set $S \in \mathcal{S}$ in the first-stage is w_S^1 , and is w_S^A in scenario A . The goal is to determine which sets to pick in stage I and which ones to pick

in each scenario so as to minimize the expected cost of picking sets, subject to $\Pr_A[\text{cost of scenario } A > B] \leq \rho$, where ρ is a constant that is not part of the input. Notice that the costs w_S^A are only revealed when we sample scenario A ; thus, the “input size”, denoted by \mathcal{I} is $O(m + n + \sum_S \log w_S + \log B)$.

For a given (fractional) point $x \in \mathbb{R}^m$ with $0 \leq x_S \leq 1$ for all S , define $f_A(x)$ to be the minimum value of $w^A \cdot y_A$ subject to $\sum_{S:e \in S} y_{A,S} \geq 1 - \sum_{S:e \in S} x_S$ for $e \in A$, and $y_{A,S} \geq 0$ for all S . Let $\mathcal{P} = [0, 1]^m$. As mentioned in the Introduction, the set of feasible solutions to even the *fractional risk-averse problem* (where one can buy sets fractionally) is not in general a convex set. We consider the following LP-relaxation of the problem, which is a relaxation of even the fractional risk-averse problem (Claim 3.1). Throughout we use A to index the scenarios in \mathcal{A} , and S to index the sets in \mathcal{S} .

$$\min \quad \sum_S w_S^I x_S + \sum_{A,S} p_A (w_S^A y_{A,S} + w_S^A z_{A,S}) \quad (\text{RASC-P})$$

$$\text{s.t.} \quad \sum_A p_A r_A \leq \rho \quad (1)$$

$$\sum_{S:e \in S} (x_S + y_{A,S}) + r_A \geq 1 \quad \text{for all } A, e \in A, \quad (2)$$

$$\sum_{S:e \in S} (x_S + y_{A,S} + z_{A,S}) \geq 1 \quad \text{for all } A, e \in A, \quad (3)$$

$$\sum_S w_S^A y_{A,S} \leq B \quad \text{for all } A \quad (4)$$

$$x_S, y_{A,S}, z_{A,S}, r_A \geq 0 \quad \text{for all } A, S. \quad (5)$$

Here x denotes the first-stage decisions. The variable r_A denotes whether one exceeds the budget of B for scenario A , and the variables $y_{A,S}$ and $z_{A,S}$ denote respectively the sets picked in scenario A in the situations where one does not exceed the budget (so $r_A = 0$) and where one does exceed the budget (so $r_A = 1$). Consequently, constraint (4) ensures that the cost of the y_A decisions does not exceed the budget B , and (1) ensures that the total probability mass of scenarios where one does exceed the budget is at most ρ . Let OPT denote the optimum value of (RASC-P).

A significant difficulty faced in solving (RASC-P) is that the scenarios are no longer *separable* given a first-stage solution, since constraint (1) couples the different scenarios. As a consequence, in order to specify a solution to (RASC-P) one needs to compute a first-stage solution and give an explicit procedure that computes (y_A, z_A, r_A) in any given scenario A . In our algorithms however, we can avoid this complication because, as we show below, given only the first-stage component of a solution to (RASC-P), one can round it to a first-stage solution to the fractional risk-averse problem (and then to an integer solution) losing a small factor in the solution cost and the probability-threshold. But observe that if we have a first-stage solution x to the fractional risk-averse problem with probability-threshold P such that there exist second-stage solutions yielding a total expected cost of C , then one can also easily compute second-stage solutions that yield no greater total cost (and where $\Pr[\text{second-stage cost} > B] \leq P$), by simply solving the LP $f_A(x)$ in each scenario A . This implies that our algorithm for solving (RASC-P) only needs to return a first-stage solution to (RASC-P) that can be extended to a near-optimum solution (without specifying an explicit procedure to compute the second-stage solutions).

We show (Theorem 3.3) that for any $\epsilon, \kappa > 0$, one can efficiently compute a first-stage solution x for which there exist solutions (y_A, z_A, r_A) in every scenario A satisfying (2)–(5) such that $w^1 \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq (1 + 2\epsilon) OPT$, and $\sum_A p_A r_A \leq \rho(1 + \kappa)$. Complementing this, we give a simple rounding procedure based on the rounding theorem in [38] to convert a fractional solution to (RASC-P) to an integer solution using an LP-based c -approximation algorithm for the *deterministic* set cover (DSC) problem, that is, an algorithm that returns a set cover of cost at most c times the optimum of the standard LP-relaxation for DSC. We prove this rounding theorem first, in order to better motivate our goal of solving (RASC-P).

Claim 3.1 *OPT is a lower bound on the optimum of the fractional risk-averse problem.*

Proof : We show that any solution \hat{x} to the fractional risk-averse problem can be mapped to a solution to (RASC-P) of no greater cost. Let \hat{y}_A be such that $f_A(\hat{x}) = w^A \cdot \hat{y}_A$, so $\Pr[f_A(\hat{x}) > B] \leq \rho$. We set $x = \hat{x}$. For scenario A , if $f_A(\hat{x}) \leq B$, we set $r_A = 0$, $y_A = \hat{y}_A$, $z_A = \mathbf{0}$. Otherwise, we set $r_A = 1$, $y_A = \mathbf{0}$, $z_A = \hat{y}_A$. It is easy to see that this yields a feasible solution to (RASC-P) of cost $w^I \cdot \hat{x} + \sum_A p_A f_A(\hat{x})$. ■

Theorem 3.2 (Rounding theorem) *Let $(x, \{(y_A, z_A, r_A)\})$ be a solution satisfying (2)–(5) of objective value $C = (w^A \cdot x + \sum_A p_A w^A \cdot (y_A + z_A))$, and let $P = \sum_A p_A r_A$. Given any $\varepsilon > 0$, one can obtain*

- (i) *a solution \hat{x} such that $w^I \cdot \hat{x} + \sum_A p_A f_A(\hat{x}) \leq (1 + \frac{1}{\varepsilon})C$ and $\Pr_A[f_A(\hat{x}) > (1 + \frac{1}{\varepsilon})B] \leq (1 + \varepsilon)P$;*
- (ii) *an integer solution $(\tilde{x}, \{\tilde{y}_A\})$ of cost at most $2c(1 + \frac{1}{\varepsilon})C$ such that $\Pr_A[w^A \cdot \tilde{y}_A > 2cB(1 + \frac{1}{\varepsilon})] \leq (1 + \varepsilon)P$ using an LP-based c -approximation algorithm for the deterministic set cover problem.*

Moreover, one only needs to know the first-stage solution x to obtain \hat{x} and \tilde{x} .

Proof : Set $\hat{x} = (1 + \frac{1}{\varepsilon})x$. Consider any scenario A . Observe that $(y_A + z_A)$ yields a feasible solution to the second-stage problem for scenario A . Also, if $r_A < \frac{1}{1+\varepsilon}$, then $(1 + \frac{1}{\varepsilon})y_A$ also yields a feasible solution. Thus, we have $f_A(\hat{x}) \leq w^A \cdot (y_A + z_A)$ and if $r_A < \frac{1}{1+\varepsilon}$ then we also have $f_A(\hat{x}) \leq (1 + \frac{1}{\varepsilon})B$. So $w^I \cdot x + \sum_A p_A f_A(\hat{x}) \leq (1 + \frac{1}{\varepsilon})C$ and $\Pr[f_A(\hat{x}) > (1 + \frac{1}{\varepsilon})B] \leq \sum_{A:r_A \geq \frac{1}{1+\varepsilon}} p_A \leq (1 + \varepsilon)P$.

We can now round \hat{x} to an integer solution $(\tilde{x}, \{\tilde{y}_A\})$ using the Shmoys-Swamy [38] rounding procedure (which only needs \hat{x}) losing a factor of $2c$ in the first- and second-stage costs. This proves part (ii). ■

3.1 Solving the risk-averse problem (RASC-P)

We now describe and analyze the procedure used to solve (RASC-P). First, we get around the difficulty posed by the coupling constraint (1) in formulation (RASC-P) by using the technique of Lagrangian relaxation. We take the Lagrangian dual of (1) introducing a dual variable Δ to obtain the following formulation.

$$\max_{\Delta \geq 0} -\Delta\rho + \left(\min_{x \in \mathcal{P}} h(\Delta; x) = w^I \cdot x + \sum_A p_A g_A(\Delta; x) \right) \quad (\text{LD1})$$

where $g_A(\Delta; x) = \min_S \sum w_S^A (y_{A,S} + z_{A,S}) + \Delta r_A$ s.t. (2)–(4), $y_{A,S}, z_{A,S}, r_A \geq 0$ for all S . (P)

It is straightforward to show via duality theory that (RASC-P) and (LD1) have the same optimal value, and moreover that if $(x^*, \{y_A^*\}, \{z_A^*\}, \{r_A^*\})$ is an optimal solution to (RASC-P) and Δ^* is the optimal value for the dual variable corresponding to (1) then $(\Delta^*; x^*, \{(y_A^*, z_A^*, r_A^*)\})$ is an optimal solution to (LD1). Recall that $\mathcal{P} = [0, 1]^m$. Let $OPT(\Delta) = \min_{x \in \mathcal{P}} h(\Delta; x)$. So $OPT = \max_{\Delta \geq 0} (OPT(\Delta) - \Delta\rho)$. Let $\lambda = \max\{1, \max_{A,S} (w_S^A / w_S^I)\}$, which we assume is known. The main result of this section is as follows. Throughout, when we say “with high probability”, we mean that a failure probability of δ can be ensured using $\text{poly}(\ln(\frac{1}{\delta}))$ -dependence on the sample size (or running time).

Theorem 3.3 *For any $\epsilon, \gamma, \kappa > 0$, RiskAlg (see Fig. 1) runs in time $\text{poly}(\mathcal{I}, \frac{\lambda}{\epsilon\kappa\rho}, \log(\frac{1}{\gamma}))$, and returns with high probability a first-stage solution x and solutions (y_A, z_A, r_A) for each scenario A that satisfy (2)–(5) and such that (i) $w^I \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq (1 + \epsilon)OPT + \gamma$; and (ii) $\sum_A p_A r_A \leq \rho(1 + \kappa)$. Under the very mild assumption (*) that $w^I \cdot x + f_A(x) \geq 1$ for every $A \neq \emptyset$, $x \in \mathcal{P}$,¹ we can convert this guarantee into a $(1 + 2\epsilon)$ -multiplicative guarantee in the cost in time $\text{poly}(\mathcal{I}, \frac{\lambda}{\epsilon\kappa\rho})$.*

¹A similar assumption is made in [38] to obtain a multiplicative guarantee.

RiskAlg (ϵ, γ, κ) [$\epsilon \leq \kappa < 1$; the quantities $p^{(i)}$, $cost^{(i)}$, (y_A, z_A, r_A) are used only in the analysis.]

- C1. Fix $\epsilon = \epsilon/6$, $\zeta = \gamma/4$, $\eta = \rho\kappa/16$. Also, set $\sigma = \epsilon/6$, $\gamma' = \gamma/4$, $\beta = \kappa/8$, and $\rho' = \rho(1 + 3\kappa/4)$. Consider the Δ values $\Delta_0, \Delta_1, \dots, \Delta_k$, where $\Delta_0 = \gamma'$, $\Delta_{i+1} = \Delta_i(1 + \sigma)$ and k is the smallest value such that $\Delta_0(1 + \sigma)^k \geq \text{UB}$. Note that $k = O(\log(\frac{\text{UB}}{\gamma'})/\sigma)$.
- C2. For each Δ_i , let $(x^{(i)}, \{(y_A^{(i)}, z_A^{(i)}, r_A^{(i)})\}) \leftarrow \text{SA-Alg}(\Delta; \epsilon, \eta, \zeta)$ (here $(y_A^{(i)}, z_A^{(i)}, r_A^{(i)})$ is an optimal solution to $g_A(\Delta_i; x^{(i)})$ and is implicitly given). Let $p^{(i)} = \sum_A p_A r_A^{(i)}$ and $cost^{(i)} = h(\Delta_i; x^{(i)}) = w^I \cdot x^{(i)} + \sum_A p_A (w^A \cdot (y_A^{(i)} + z_A^{(i)}) + \Delta_i r_A^{(i)})$.
- C3. By sampling $n = \frac{1}{2\beta^2\rho^2} \ln(\frac{4k}{\delta})$ scenarios, for each $i = 0, \dots, k$, compute an estimate $p'^{(i)} = \sum_A \hat{q}_A r_A^{(i)}$ of $p^{(i)}$, where \hat{q}_A is the frequency of scenario A in the sampled set.
- C4. If $p'^{(0)} \leq \rho'$ then return $x^{(0)}$ as the first-stage solution. [In scenario A , return $(y_A, z_A, r_A) = (y_A^{(0)}, z_A^{(0)}, r_A^{(0)})$].
- C5. Otherwise (i.e., $p'^{(0)} > \rho'$) find an index i such that $p'^{(i)} \geq \rho'$ and $p'^{(i+1)} \leq \rho'$ (we argue that such an i must exist). Let a be such that $a \cdot p'^{(i)} + (1-a)p'^{(i+1)} = \rho'$. Return the first-stage solution $x = a \cdot x^{(i)} + (1-a)x^{(i+1)}$. [In scenario A , return the solution $(y_A, z_A, r_A) = a(y_A^{(i)}, z_A^{(i)}, r_A^{(i)}) + (1-a)(y_A^{(i+1)}, z_A^{(i+1)}, r_A^{(i+1)})$].

SA-Alg ($\Delta; \epsilon, \eta, \zeta$) [K is (a bound on) the Lipschitz constant of $h(\Delta; \cdot)$; $\mathcal{P} \subseteq B(\mathbf{0}, R)$ and \mathcal{P} contains a ball of radius $V \leq 1$.]

- B1. Set $\tau = \zeta/6$, $N = \log(\frac{2KR}{V\tau})$. Let $G_\tau \subseteq \mathcal{P}$ be an extended $\frac{\tau}{KN}$ -net of \mathcal{P} as defined in Section 2, so that $|G_\tau| = \text{poly}(\log(\frac{KR}{V\tau}))$. Draw $\mathcal{N} = 8N^2(\frac{4\lambda}{\epsilon} + \frac{m}{\eta})^2 \ln(\frac{2|G_\tau|m}{\delta})$ samples and for each scenario A , set $\hat{p}_A = \mathcal{N}_A/\mathcal{N}$, where \mathcal{N}_A is the number of times scenario A is sampled.
- B2. Solve the SAA problem $\min_{x \in \mathcal{P}} \hat{h}(\Delta; x)$, where $\hat{h}(\Delta; x) = w^I \cdot x + \sum_A \hat{p}_A g_A(\Delta; x)$ to obtain a solution \hat{x} . Return \hat{x} and in scenario A , return the optimal solution to $g_A(\Delta; \hat{x})$.

Figure 1: The procedures RiskAlg and SA-Alg.

Procedure RiskAlg is described in Figure 1. In the procedure, we also specify the second-stage solutions for each scenario that can be used to extend the computed first-stage solution to a near-optimal solution to (RASC-P). We use these solutions only in the analysis.

We show in Section 5 that the dependence on $\frac{1}{\kappa\rho}$ is unavoidable in the black-box model. The “greedy algorithm” for deterministic set cover [9] is an LP-based $\ln n$ -approximation algorithm, so Theorem 3.3 combined with Theorem 3.2 shows that for any $\epsilon, \kappa, \epsilon > 0$ one can efficiently compute an integer solution $(\tilde{x}, \{\tilde{y}_A\})$ of cost at most $2 \ln n(1 + \epsilon + \frac{1}{\epsilon}) \cdot \text{OPT}$ such that $\Pr_A[w^A \cdot \tilde{y}_A > 2B \ln n(1 + \epsilon + \frac{1}{\epsilon})] \leq \rho(1 + \kappa + \epsilon)$.

Algorithm RiskAlg is essentially a search procedure for the “right” value of the Lagrangian multiplier Δ , wrapped around the SAA method, which is used in procedure SA-Alg to compute a near-optimal solution to the minimization problem $\min_{x \in \mathcal{P}} h(\Delta; x)$ for any given $\Delta \geq 0$. Theorem 3.4 states the precise approximation guarantee satisfied by the solution returned by SA-Alg. Given this, we argue that by considering polynomially many Δ values that increase geometrically up to some upper bound UB, one can find efficiently some Δ where the solution $(x, \{(y_A, z_A, r_A)\})$ returned by SA-Alg for Δ is such that $\sum_A p_A r_A$ is “close” to ρ . This will also imply that this solution is a near-optimal solution. We set $\text{UB} = 16(\sum_S w_S^I)/\rho$, so $\log \text{UB}$ is polynomially bounded. However, the search for the “right” value of Δ and our analysis are complicated by the fact that we have two sources of error whose magnitudes we need to control: first, we only have an approximate solution $(x, \{(y_A, z_A, r_A)\})$ for Δ , which also means that one cannot use any optimality conditions; second, for any Δ , we have only implicit access to the second-stage solutions $\{y_A, z_A, r_A\}$ computed by Theorem 3.4, so we cannot actually compute or use $\sum_A p_A r_A$ in our search, but will need to estimate it via sampling.

Theorem 3.4 *For any $\Delta \geq 0$, and any $\epsilon, \zeta, \eta > 0$, SA-Alg runs in time $\text{poly}(\mathcal{I}, \frac{\lambda}{\epsilon\eta}, \log(\frac{\Delta}{\zeta}))$ and returns, with high probability, a first-stage solution x such that $h(\Delta; x) \leq (1 + \epsilon)\text{OPT}(\Delta) + \eta\Delta + \zeta$.*

Analysis. For the rest of this section, ϵ, γ, κ are fixed values given by Theorem 3.3. We may assume without loss of generality that $\epsilon \leq \kappa < 1$. We prove Theorem 3.4 in Section 3.1.1. Here, we show how this leads to the proof of Theorem 3.3. Given Theorem 3.4 and Lemma 2.3, we assume that the high probability event “ $\forall i, \text{cost}^{(i)} \leq (1 + \epsilon)OPT(\Delta_i) + \eta\Delta_i + \zeta$ and $|p^{(i)} - p^{(i-1)}| \leq \beta\rho$ ” happens.

Claim 3.5 We have $p^{(k)} < \rho/2$ and $p'^{(k)} < \rho/2$.

Proof : If $p^{(k)} > \frac{\rho(1+\epsilon)}{4}$, then $\text{cost}^{(k)} - \eta\Delta_k > 2(1 + \epsilon)(\sum_S w_S^1) > (1 + \epsilon)OPT(\Delta_k) + \zeta$, which is a contradiction. The last inequality follows since $OPT(\Delta) \leq \sum_S w_S^1$ for any Δ . Therefore, $p^{(k)} < \rho/2$, and $p'^{(k)} \leq p^{(k)} + \beta\rho < \rho/2$. ■

Proof of Theorem 3.3 : Let x be the first-stage solution returned by RiskAlg, and (y_A, z_A, r_A) be the solution returned for scenario A . It is clear that (2)–(5) are satisfied. Suppose first that $p^{(0)} \leq \rho'$ (so $x = x^{(0)}$.) Part (ii) of the theorem follows since $p^{(0)} \leq p'^{(0)} + \beta\rho \leq \rho(1 + \kappa)$. Part (i) follows since

$$w^1 \cdot x^{(0)} + \sum_A p_A w^A \cdot (y_A^{(0)} + z_A^{(0)}) \leq h(\gamma'; x) \leq (1 + \epsilon)OPT(\gamma') + \eta\gamma' + \zeta \leq (1 + \epsilon)OPT + \gamma'(1 + \epsilon + \eta) + \zeta.$$

The penultimate inequality follows because for any Δ , we have $OPT(\Delta) \leq OPT(0) + \Delta \leq OPT + \Delta$.

Now suppose that $p^{(0)} > \rho'$. In this case, there must exist an i such that $p^{(i)} \geq \rho'$, and $p^{(i+1)} \leq \rho'$ because $p^{(0)} > \rho'$ and $p^{(k)} < \rho'$ (by Claim 3.5), so step C4 is well defined. We again prove part (ii) first. We have $\sum_A p_A r_A = a \cdot p^{(i)} + (1 - a)p^{(i+1)} \leq \rho' + \beta\rho \leq \rho(1 + \kappa)$. To prove part (i), observe that $w^1 \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq a \cdot \text{cost}^{(i)} + (1 - a) \cdot \text{cost}^{(i+1)} - \Delta_i(a \cdot p^{(i)} + (1 - a) \cdot p^{(i+1)})$, which is at most

$$(1 + \epsilon)\left(a \cdot OPT(\Delta_i) + (1 - a)OPT(\Delta_{i+1})\right) + \eta(a\Delta_i + (1 - a)\Delta_{i+1}) + \zeta - \Delta_i(\rho' - \beta\rho).$$

Now noting that $\Delta_{i+1} = (1 + \sigma)\Delta_i$, it is easy to see that $OPT(\Delta_{i+1}) \leq (1 + \sigma)OPT(\Delta_i)$. Also, $\rho' - \beta\rho - \eta(1 + \sigma) \geq (1 + \epsilon + 2\sigma)\rho$. So the above quantity is at most $(1 + \epsilon + 2\sigma)(OPT(\Delta_i) - \Delta_i\rho) + \zeta \leq (1 + \epsilon)OPT + \gamma$.

The running time is the time taken to obtain the solutions for all the Δ_i values plus the time taken to compute $p^{(i)}$ for each i . This is at most $(k + 1) \cdot \text{poly}\left(\mathcal{I}, \frac{\lambda}{\epsilon\eta}, \log\left(\frac{\Delta_k}{\zeta}\right)\right) + O\left(\frac{\ln k}{\beta^2 \rho^2}\right)$, using Theorem 3.4. Note that $\log(\Delta_k)$ is polynomially bounded. Plugging in $\epsilon, \eta, \zeta, \beta$, and k , we obtain the $\text{poly}\left(\mathcal{I}, \frac{\lambda}{\epsilon\kappa\rho}, \log\left(\frac{1}{\gamma}\right)\right)$ bound.

Proof of multiplicative guarantee. To obtain the multiplicative guarantee, we show that by initially sampling roughly $\max\{1/\rho, \lambda\}$ times, with high probability, one can either determine that $x = \mathbf{0}$ is an optimal first-stage solution, or obtain a lower bound on OPT and then set γ appropriately in RiskAlg to obtain the multiplicative bound. Recall that $f_A(x)$ is the minimum value of $w^A \cdot y_A$ over all $y_A \geq \mathbf{0}$ such that $\sum_{S:e \in S} y_{A,S} \geq 1 - \sum_{S:e \in S} x_S$ for $e \in A$. Call $A = \emptyset$ a null scenario. Let $q = \sum_{A:A \neq \emptyset} p_A$ and $\alpha = \min\{\rho, 1/\lambda\}$. Note that $OPT \geq q$. Let \hat{z}_A be an optimal solution to $f_A(\mathbf{0})$. Define a solution $(\bar{y}_A, \bar{z}_A, \bar{r}_A)$ for scenario A as follows. Set $(\bar{y}_A, \bar{z}_A, \bar{r}_A) = (\mathbf{0}, \mathbf{0}, 0)$ if $A = \emptyset$, and $(\mathbf{0}, \hat{z}_A, 1)$ if $A \neq \emptyset$. We first argue that if $q \leq \alpha$, then $(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\})$ is an optimal solution to (RASC-P). It is clear that the solution is feasible since $\sum_A p_A \bar{r}_A = q \leq \rho$. To prove optimality, suppose $(x^*, \{y_A^*, z_A^*, r_A^*\})$ is an optimal solution. Consider the solution where $x = \mathbf{0}$ and the solution for scenario A is $(\mathbf{0}, \mathbf{0}, 0)$ if $A = \emptyset$, and $(\mathbf{0}, z_A^* + y_A^* + x^*, 1)$ otherwise. This certainly gives a feasible solution. The difference between the cost of this solution and that of the optimal solution is at most $\sum_{A:A \neq \emptyset} p_A w^A \cdot x^* - w^1 \cdot x^*$, which is nonpositive since $w^A \leq \lambda w^1$ and $q \leq 1/\lambda$. Setting $z_A = \hat{z}_A$ for a non-null scenario can only decrease the cost, and hence, also yields an optimal solution.

Let δ be the desired failure probability, which we may assume to be less than $\frac{1}{2}$ without loss of generality. We determine with high probability if $q \geq \alpha$. We draw $M = \frac{\ln(1/\delta)}{\alpha}$ samples and compute $X =$ number of times a non-null scenario is sampled. We claim that with high probability, if $X > 0$ then $OPT \geq LB = \frac{\delta}{\ln(1/\delta)} \cdot \alpha$; in this case, we return the solution $\text{RiskAlg}(\epsilon, \epsilon LB, \kappa)$ to obtain the desired guarantee. Otherwise, if $X = 0$, we return $(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\})$ as the solution.

Let $r = \Pr[X = 0] = (1 - q)^M$. So $1 - qM \leq r \leq e^{-qM}$. If $q \geq \ln(\frac{1}{\delta})/M$, then $\Pr[X = 0] \leq \delta$, so with probability at least $1 - \delta$ we say that $OPT \geq LB$, which is true since $OPT \geq q \geq \alpha$. If $q \leq \delta/M$, then $\Pr[X = 0] \geq 1 - \delta$ and we return $(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\})$ as the solution, which is an optimal solution since $q \leq \alpha$. If $\delta/M < q < \ln(\frac{1}{\delta})/M$, then we always return a correct answer since it is both true that $OPT \geq q > LB$, and that $(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\})$ is an optimal solution. ■

3.1.1 Proof of Theorem 3.4

Throughout this section, ϵ, η, ζ are fixed at the values given in the statement of Theorem 3.4. Let (BSC-P) denote the problem $\min_{x \in \mathcal{P}} h(\Delta; x)$. The proof proceeds by analyzing the subgradients of $h(\Delta; \cdot)$ and $\hat{h}(\Delta; \cdot)$ and showing that Lemma 2.2 can be applied here.

We first note that the arguments given in [38, 45, 7] for 2-stage programs do not directly apply to (BSC-P) since it does not fall into the class of problems considered therein. Shmoys and Swamy [38] show (essentially) that if one can compute an (ω, ξ) -subgradient of the objective function $h(\Delta; \cdot)$ at any given point x for a *sufficiently small* ω, ξ , then one can use the ellipsoid method to obtain a near optimal solution to (BSC-P). They argue that for a large class of 2-stage LPs, one can efficiently compute an (ω, ξ) -subgradient using $\text{poly}(\frac{\lambda}{\omega})$ samples. Subsequently [45], they leveraged the proof of the ellipsoid-based algorithm to argue that the SAA method also yields an efficient approximation scheme for the same class of 2-stage LPs. These proofs rely on the fact that for their class of 2-stage programs, each component of the subgradient lies in a range bounded multiplicatively by a factor of λ and can be approximated additively using $\text{poly}(\lambda)$ samples. However, in the case of (BSC-P), for a subgradient $d = (d_S)$ of $h(\Delta; \cdot)$, we can only say that $d_S \in [-w_S^A - \Delta, w_S^I]$ (see Lemma 3.6), which makes it difficult to obtain an (ω, ξ) -subgradient using sampling for suitably small ω, ξ . Charikar, Chekuri and Pál [7] considered a similar class of 2-stage problems, and gave an alternate proof of efficiency of the SAA method showing that even approximate solutions to the SAA problem translate to approximate solutions to the original problem. Their proof shows that if Λ is such that $g_A(\Delta; x) - g_A(\Delta; \mathbf{0}) \leq \Lambda w^I \cdot x$ for every A and $x \in \mathcal{P}$, then $\text{poly}(\frac{\Lambda}{\epsilon})$ samples suffice to construct an SAA problem whose optimal solutions correspond to $(1 + \epsilon)$ -optimal solutions to the original problem. But for our problem, we can only obtain the bound $\Lambda \leq w^A \cdot x + \Delta(\sum_S x_S) \leq \lambda w^I \cdot x + \Delta \sum_S x_S$, and Δ might be large compared to $w^I \cdot x$.

The key insight that allows us to circumvent these difficulties is that in order to establish our (weak) guarantee, where we allow for an additive error measured relative to Δ , it suffices to be able to approximate each component d_S of the subgradient of $h(\Delta; \cdot)$ within an additive error proportional to $(w_S^I + \Delta)$, and this can be done by drawing $\text{poly}(\lambda)$ samples. This enables one to argue that functions $\hat{h}(\Delta; \cdot)$ and $h(\Delta; \cdot)$ satisfy the ‘‘closeness-in-subgradients’’ property stated in Lemma 2.2.

The subgradients of $h(\Delta; \cdot)$ and $\hat{h}(\Delta; \cdot)$ at x are obtained from the optimal dual solutions to $g_A(\Delta; x)$

for every A . The dual of $g_A(\Delta; x)$ is given by

$$\begin{aligned}
\max \quad & \sum_e (\alpha_{A,e} + \beta_{A,e}) \left(1 - \sum_{S:e \in S} x_S\right) - B\theta_A & (D) \\
\text{s.t.} \quad & \sum_{e \in S \cap A} (\alpha_{A,e} + \beta_{A,e}) \leq w_S^A (1 + \theta_A) & \text{for all } S \\
& \sum_{e \in S \cap A} \beta_{A,e} \leq w_S^A & \text{for all } S \\
& \sum_{e \in A} \alpha_{A,e} \leq \Delta \\
& \alpha_{A,e}, \beta_{A,e} \geq 0 & \text{for all } e \in A.
\end{aligned}$$

Here $\alpha_{A,e}$ and $\beta_{A,e}$ are respectively the dual variables corresponding to (2) and (3), and θ_A is the dual variable corresponding to (4). As in [38], we then have the following description of the subgradient of h .

Lemma 3.6 *Let $(\alpha_A^*, \beta_A^*, \theta_A^*)$ be an optimal dual solution to $g_A(\Delta; x)$. Then the vector d_x with components $d_{x,S} = w_S^I - \sum_A p_A \sum_{e \in S} (\alpha_{A,e}^* + \beta_{A,e}^*)$ is a subgradient of $h(\Delta; \cdot)$ at x .*

Since $\hat{h}(\Delta; \cdot)$ is of the same form as $h(\Delta; \cdot)$, we have similarly that $\hat{d}_x = (\hat{d}_{x,S})$, where $\hat{d}_{x,S} = w_S^I - \sum_A p_A \sum_{e \in S} (\alpha_{A,e}^* + \beta_{A,e}^*)$, is a subgradient of $\hat{h}(\Delta; \cdot)$ at x . Since \hat{d}_x and d_x both have ℓ_2 norm at most $\lambda \|w^I\| + |\Delta|$, $\hat{h}(\Delta; \cdot)$ and $h(\Delta; \cdot)$ have Lipschitz constant at most $K = \lambda \|w^I\| + |\Delta|$.

Lemma 3.7 *Let d be a subgradient of $h(\Delta; \cdot)$ at the point $x \in \mathcal{P}$, and suppose that \hat{d} is a vector such that $\hat{d}_S \in [d_S - \omega w_S^I - \xi/2m, d_S + \omega w_S^I + \xi/2m]$ for all S . Then \hat{d} is an (ω, ξ) -subgradient of $h(\Delta; \cdot)$ at x .*

Proof : Let y be any point in \mathcal{P} . We have $h(\Delta; y) - h(\Delta; x) \geq \hat{d} \cdot (y - x) + (d - \hat{d}) \cdot (y - x)$. The second term is at least

$$\sum_{S: d_S \leq \hat{d}_S} (d_S - \hat{d}_S) y_S + \sum_{S: \hat{d}_S > d_S} (\hat{d}_S - d_S) x_S \geq \sum_S (-\omega w_S^I y_S - \omega w_S^I x_S) - \xi \geq -\omega h(\Delta; y) - \omega h(\Delta; x) - \xi. \quad \blacksquare$$

In the sequel, we set $\omega = \varepsilon/8N$, $\xi = \eta\Delta/2N$. Let $(\alpha_A^*, \beta_A^*, \theta_A^*)$ be the optimal dual solution to $g_A(\Delta; x)$ used to define \hat{d}_x and d_x . Notice that $\hat{d}_{x,S}$ is simply $w_S^I - \sum_{e \in S} (\alpha_{A,e}^* + \beta_{A,e}^*)$ averaged over the scenarios sampled independently to construct the SAA problem $\hat{h}(\Delta; \cdot)$, and $\mathbb{E}[\hat{d}_{x,S}] = d_{x,S}$. The sample size \mathcal{N} in SA-Alg is specifically chosen so that the Chernoff bound (Lemma 2.3) implies that $|\hat{d}_{x,S} - d_{x,S}| \leq \omega w_S^I + \xi/2m$ for all S with probability at least $1 - \frac{\delta}{|G_\tau|}$ for every $x \in G_\tau$; hence, \hat{d}_x is an (ω, ξ) -subgradient of $h(\Delta; \cdot)$ at x (by Lemma 3.7). So taking the union bound shows that with probability at least $1 - \delta$, $\hat{h}(\Delta; \cdot)$ and $h(\Delta; \cdot)$ satisfy the conditions of Lemma 2.2 with $K = \lambda \|w^I\| + |\Delta|$, $\varrho = \varepsilon$ and ξ (as above), which yields the desired approximation guarantee.

We can take $R = \sqrt{m}$ and $V = \frac{1}{2}$ here, so the number of samples \mathcal{N} is $\text{poly}(\mathcal{I}, \frac{\lambda}{\varepsilon\eta}, \log(\frac{\Delta}{\xi}))$. \blacksquare

Remark 3.8 Notice that nowhere do we use the fact that the scenario-budgets are uniform, and thus, our results (Theorem 3.4 and hence, Theorem 3.3) extend to the setting where we have different budgets for the different scenarios. The scenario budgets $\{B^A\}$ are now not specified explicitly; we get to know B^A when we sample scenario A . (Notice that we may assume that $B^A \leq \lambda \sum_S w_S^I$ for all A .)

3.2 Risk-averse robust set cover

In the risk-averse robust set cover problem, the goal is to choose some sets x in stage I and some sets y_A in each scenario A so that their union covers A , so as to minimize $w^I \cdot x + Q_\rho[w^A \cdot y_A]$. Recall that $Q_\rho[w^A \cdot y_A]$ is the $(1 - \rho)$ -quantile of $\{w^A \cdot y_A\}_{A \in \mathcal{A}}$, that is, the smallest B such that $\Pr_A[w^A \cdot y_A > B] \leq \rho$. As mentioned in the Introduction, risk-averse robust problems can be essentially reduced to risk-averse budget problems. We briefly sketch this reduction here for the set cover problem. The same ideas can be used to obtain approximation algorithms for the risk-averse robust versions of all the applications considered in Section 4.

We use the common method of “guessing” $B = Q_\rho[w^A \cdot y_A]$ for an optimal solution. Given this guess, we need to find integral $(x, \{y_A\})$ so as to minimize $w^I \cdot x + B$ (and hence, $w^I \cdot x$) subject to the constraint that $x + y_A$ forms a set cover for A and $\Pr_A[w^A \cdot y_A > B] \leq \rho$. This looks very similar to the risk-averse budgeted set cover problem; the only difference is that the expected second-stage cost does not appear in the objective function. Thus, one can write an LP-relaxation for the (fractional) risk-averse robust problem that looks similar to (RASC-P) except that the objective function is now $w^I \cdot x$, and constraint (3) and the variables $z_{A,S}$ can be dropped. After Lagrangifying (1) using the dual variable Δ , we obtain the following problem

$$\max_{\Delta \geq 0} \quad -\Delta\rho + \left(\min \quad h'(\Delta; x) = w^I \cdot x + \sum_A p_A g'_A(\Delta; x) \right) \quad (\text{LD2})$$

where $g'_A(\Delta; x) = \min\{\Delta r_A : (2), (4), y_A \geq \mathbf{0}, r_A \geq 0\}$.

Let OPT_{Rob} denote the optimum value of the *fractional* risk-averse robust problem $\min_{x \in \mathcal{P}}(w^I \cdot x + Q_\rho[f_A(x)])$, and $OPT_{Rob}(B)$ denote the optimum value of (LD2) for a given $B \geq 0$. Note that $OPT_{Rob}(B)$ decreases with B . We prove that for any $B \geq 0$ and $\Delta \geq 0$, SA-Alg returns a solution to the inner minimization problem in (LD2) that satisfies the approximation guarantee stated in Theorem 3.4. Arguing as in the proof of Theorem 3.3, this implies that RiskAlg can be used to obtain a near-optimal solution to (LD2) while violating the probability threshold by a small factor.

The claimed approximation guarantee for SA-Alg follows because $h(\Delta; \cdot)$ and its sample-average approximation $\hat{h}'(\Delta; \cdot)$ constructed in SA-Alg satisfy the closeness-in-subgradients property of Lemma 2.2. Let $\alpha_{A,e}^*$ is the value of the dual variable corresponding to (2) in an optimal dual solution to $g'_A(\Delta; x)$. Note that $\sum_e \alpha_{A,e}^* \leq \Delta$ for all A . Similar to Lemma 3.6, we now have that the vectors $d_x = (d_{x,S})$ with $d_{x,S} = w_S^I - \sum_A p_A (\sum_{e \in S} \alpha_{A,e}^*)$ and $\hat{d}_x = (\hat{d}_{x,S})$ with $\hat{d}_{x,S} = w_S^I - \sum_A \hat{p}_A (\sum_{e \in S} \alpha_e^*)$ are respectively subgradients of $h'(\Delta; \cdot)$ and $\hat{h}'(\Delta; \cdot)$ at x . Let $N, \mathcal{N}, \tau, G_\tau$ be as defined in SA-Alg with $R = \sqrt{m}$, $V = \frac{1}{2}$ and $K = \|w^I\| + |\Delta|$. Using \mathcal{N} samples, for any $x \in G_\tau$, with very high probability we have that $|\hat{d}_{x,S} - d_{x,S}| \leq \eta\Delta/4mN$; thus, as in Lemma 3.7, \hat{d}_x is an $(0, \frac{\eta}{2N})$ -subgradient of $h'(\Delta; \cdot)$ at x . So Lemma 2.2 shows that SA-Alg returns a solution \hat{x} such that $h'(\Delta; \hat{x}) \leq OPT + \eta\Delta + \zeta$ with high probability. Notice that in fact, the approximation guarantee obtained via SA-Alg is purely additive. Also, *one can avoid the dependence of the sample-size on λ (and ε)* here since the modified form of the subgradient means that we can ensure that $|\hat{d}_{x,S} - d_{x,S}| \leq \eta\Delta/4mn$ for every $x \in G_\tau$ and component S using a number of samples that is independent of λ . This implies that for any $\epsilon, \gamma, \kappa > 0$, RiskAlg computes (nonnegative) $(x, \{y_A, r_A\})$ satisfying (2), (4) such that $w^I \cdot x \leq (1 + \epsilon)OPT_{Rob}(B) + \gamma$ and $\sum_A p_A r_A \leq \rho(1 + \kappa)$.

To complete the reduction, we describe how to guess B . Let $W = \sum_S w_S^I$, which is an upper bound on the optimum (with $\log W$ polynomially bounded). We use the standard method of enumerating values of B increasing geometrically by $(1 + \epsilon)$; we start at γ and end at the smallest value that is at least W . So if B^* is the “correct” guess, then we are guaranteed to enumerate $B' \in [B^*, (1 + \epsilon)B^* + \gamma]$. We use RiskAlg to compute the solution for each B , and return $(x, \{y_A, r_A\})$ that minimizes $w^I \cdot x + B$. Let $(x', \{y'_A, r'_A\})$ be the solution computed for B' . Then we have $w^I \cdot x + B \leq w^I \cdot x' + B' \leq (1 + \epsilon)OPT_{Rob}(B') + (1 + \epsilon)B^* + 2\gamma \leq (1 + \epsilon)OPT_{Rob} + 2\gamma$. We remark that the same techniques yield a similar guarantee

for the LP-relaxation of a generalization of the problem, where we wish to minimize $w^I \cdot x$ plus a weighted combination of $E_A[w^A \cdot y_A]$ and $Q_\rho[w^A \cdot y^A]$.

We can convert the above guarantee into a purely multiplicative one under the same assumption (*) stated in Theorem 3.3. Let $q = \sum_{A \neq \emptyset} p_A$. Notice that if $q \leq \rho$, then $OPT_{Rob} = 0$ and $x = \mathbf{0}$ is an optimal solution, and otherwise $OPT_{Rob} \geq 1$. Let δ be such that $(1 + \kappa) \frac{\delta}{\ln(1/\delta)} \leq 1$. Using $\frac{\ln(1/\delta)}{\rho'}$ samples we can determine with high probability if $q \leq \rho'$ or if $q > \rho$. In the former case, we return $x = \mathbf{0}$ and y_A in scenario A , where $y_A = \mathbf{0}$ if $A = \emptyset$ and is any feasible solution if $A \neq \emptyset$. Note that $w^I \cdot x + Q_{\rho'}[w^A \cdot y_A] = 0$. In the latter case, we set $\gamma = \epsilon$, and obtain a execute the procedure detailed above to obtain a $(1+3\epsilon)$ -multiplicative guarantee.

Finally, one can use Theorem 3.2 to round the fractional solution to an integer solution, or to a solution to the fractional risk-averse robust problem. (The violation of the budget B can now be absorbed into the approximation ratio.) For any $\epsilon, \kappa, \varepsilon > 0$, we obtain a fractional solution \hat{x} such that $w^I \cdot \hat{x} + Q_{\rho(1+\kappa+\varepsilon)}[f_A(\hat{x})] \leq (1 + \epsilon + \frac{1}{\varepsilon}) OPT_{Rob}$, and an integer solution $(\tilde{x}, \{\tilde{y}_A\})$ such that $w^I \cdot \tilde{x} + Q_{\rho(1+\kappa+\varepsilon)}[w^A \cdot \tilde{y}_A] \leq 2c(1 + \epsilon + \frac{1}{\varepsilon}) OPT_{Rob}$ using an LP-based c -approximation algorithm for deterministic set cover.

Setting $B = 0$ above yields a problem that is interesting in its own right. When $B = 0$, we seek a minimum-cost collection of sets x that are picked *only in stage I* such that $\Pr_A[x \text{ is not a set cover for } A] \leq \rho$. That is, we obtain a *chance-constrained problem* without recourse. As shown above (although $B = 0$ is not one of our “guesses”), we can solve this chance-constrained set cover problem to obtain a solution x such that $w^I \cdot x \leq (1 + \epsilon) OPT_{Rob}(0) + \gamma$ where $\Pr_A[x \text{ does not cover } A] \leq \rho(1 + \kappa)$.

4 Applications to combinatorial optimization problems

We now show that the techniques developed in Section 3 for the risk-averse budgeted set cover problem can be used to obtain approximation algorithms for the risk-averse versions of various combinatorial optimization problems such as covering problems—(set cover,) vertex cover, multicut on trees, min s - t cut—and facility location problems. This includes many of the problems considered in [17, 38, 11] in the standard 2-stage and demand-robust models.

In all the applications, the first step is to argue that procedure RiskAlg can be used to obtain a near-optimal solution to a suitable LP-relaxation of the problem while violating the probability threshold by a small factor. Theorem 3.3 proves this for covering problems; for multicommodity flow and facility location, we need to modify the arguments slightly. The second step, which is more problem-specific, is to round the LP-solution to an integer solution. Analogous to part (i) of Theorem 3.2, we first round the LP-solution to a solution to the fractional risk-averse problem. Given this, our task is now reduced to rounding a fractional solution to a standard 2-stage problem into an integral one. For this latter step, one can use any “local” LP-based approximation algorithm for the 2-stage problem, where a local algorithm is one that preserves approximately the cost of each scenario. (For set cover, vertex cover and multicut on trees, we may use part (ii) of Theorem 3.2 directly, which utilizes the local LP-rounding algorithm in [38] (which in turn is obtained using an LP-based approximation algorithm for the deterministic covering problem).) As in the case of risk-averse robust set cover, our results extend to the setting of non-uniform budgets.

We say that an algorithm is a (c_1, c_2, c_3) -approximation algorithm for the risk-averse problem with budget B and threshold ρ , if it returns a solution of cost at most c_1 times the optimum where the probability that the second-stage cost exceeds $c_2 \cdot B$ is at most $c_3 \cdot \rho$.

Our approximation results for the budgeted problem also translate to the risk-averse robust version of the problem. Specifically, a (c_1, c_2, c_3) -approximation algorithm for the budgeted problem implies that one can obtain an integer solution $(x, \{y_A\})$ to the robust problem such that $c(x) + Q_{\rho(1+c_3)}[f_A(x, y_A)] \leq \max\{c_1, c_2\} \cdot OPT_{Rob}$. As mentioned in Section 3.2, the robust problem with a guess of $Q_\rho[f_A(x, y_A)] = 0$

gives rise to a problem where one can take actions only in stage I and one seeks to “take care” of “most” second-stage scenarios; we can solve this chance-constrained problem approximately. We also achieve bicriteria approximation guarantees for the problem of minimizing $c(x)$ plus a weighted combination of $E_A[f_A(x, y_A)]$ and $Q_\rho[f_A(x, y_A)]$.

4.1 Covering problems

Vertex cover and multicut on trees. In the risk-averse budgeted vertex cover problem, we are given a graph whose edges need to be covered by vertices. The edge-set is random and determined by a distribution (on sets of edges). A vertex v may be picked in stage I or in a scenario A incurring a cost of w_v^I or w_v^A respectively. We are also given a budget B and a probability threshold ρ and require that the probability that the second-stage cost of picking vertices exceeds B be at most ρ . In the risk-averse version of multicut on trees, we are given a tree, a (black-box) distribution over sets of s_i - t_i pairs, a budget B , and a threshold ρ . The goal is to choose edges in stage I and in each scenario such that the union of edges picked in stage I and in scenario A forms a multicut for the s_i - t_i pairs that are revealed in scenario A . Moreover, the second-stage cost of picking edges may exceed B with probability at most ρ . The goal is to minimize the total expected cost.

Both these problems are structured cases of risk-averse budgeted set cover. So one can formulate an LP-relaxation of the risk-averse problem exactly as in (RASC-P) and by Theorem 3.3, obtain a near-optimal solution to the relaxation. We may then apply Theorem 3.2 directly to these problems to round the fractional solution. Since there is an LP-based 2-approximation algorithm for the deterministic versions of both problems, we obtain the following theorem.

Theorem 4.1 *For any $\epsilon, \kappa, \varepsilon > 0$, there is a $(4(1 + \epsilon + \frac{1}{\varepsilon}), 4(1 + \epsilon + \frac{1}{\varepsilon}), 1 + \kappa + \varepsilon)$ -approximation algorithm for the risk-averse budgeted versions of vertex cover and multicut on trees.*

Min s - t cut. In the stochastic min s - t cut problem, we are given an undirected graph $G = (V, E)$ and a source $s \in V$. The location of the sink t is random and given by a distribution. We may pick an edge e in stage I or in a scenario A incurring costs w_e and w_e^A respectively. The constraints are that in any scenario A with sink t_A , the edges picked in stage I and in that scenario induce an s - t_A cut, and the goal is to minimize the expected cost of choosing edges. In the risk-averse budgeted problem there is the additional constraint that the second-stage cost may exceed a given budget B with probability at most (a given value) ρ .

The LP-relaxation of the risk-averse problem based on a path-covering formulation is a special case of (RASC-P). The only additional observation needed to see that Theorem 3.3 can be applied here is that the covering problem (P) for a scenario A (and its dual) can be solved efficiently although there are an exponential number of constraints. Thus, procedures RiskAlg and SA-Alg can be implemented efficiently and we may obtain a near-optimal solution to the relaxation.

We use Theorem 3.2, part (i) to convert the solution to a near-optimal solution \hat{x} to the fractional risk-averse problem. We now use the algorithm in [11], which is a local LP-based $O(\log |V|)$ -approximation algorithm to round this solution to an integral one. Their algorithm requires that there exist multipliers λ_A in each scenario A such that $w_e^A = \lambda^A w_e$ for every e ; consequently we also need this for our result. A detail worth noting is that their algorithm requires access also to the second-stage fractional solutions (but not the scenario-probabilities). But this is not a problem since there are only polynomially many scenarios here corresponding to the different locations of the sink. So given the first-stage solution \hat{x} , one can simply compute the optimal fractional second-stage solution for each scenario for use in their algorithm.

Theorem 4.2 *For any $\epsilon, \kappa, \varepsilon > 0$, there is an $(O(\log |V|)(1 + \epsilon + \frac{1}{\varepsilon}), O(\log |V|)(1 + \epsilon + \frac{1}{\varepsilon}), 1 + \kappa + \varepsilon)$ -approximation algorithm for risk-averse budgeted min s - t cut.*

4.2 Facility location problems

In the risk-averse budgeted facility location problem (RAUFL), we have a set of m facilities \mathcal{F} , a client-set \mathcal{D} , and a distribution over client-demands. We may open facilities in stage I or in a given scenario, and in each scenario A , for every client j with non-zero demand d_j^A , we must assign its demand to a facility opened in stage I or in that scenario. The costs of opening a facility $i \in \mathcal{F}$ in stage I and in a scenario A are f_i^I and f_i^A respectively; the cost of assigning a client j 's demand in scenario A to a facility i is $d_j^A c_{ij}$, where the c_{ij} 's form a metric. The first-stage cost is the cost of opening facilities in stage I, and the cost of scenario A is the sum of all the facility-opening and client-assignment costs incurred in that scenario. The goal is to minimize the total expected cost subject to the usual condition that the probability that the second-stage cost exceeds B is at most some threshold ρ . For notational simplicity, we consider the case of $\{0, 1\}$ -demands, so a scenario $A \subseteq \mathcal{D}$ simply specifies the clients that need to be assigned in that scenario. We formulate the following LP-relaxation of the problem. Throughout, i indexes the facilities in \mathcal{F} and j the clients in \mathcal{D} .

$$\min \sum_i f_i^I y_i + \sum_{A \subseteq \mathcal{D}} p_A \left(\sum_i f_i^A (y_{A,i} + v_{A,i}) + \sum_{j \in A, i} c_{ij} (x_{A,ij} + u_{A,ij}) \right) \quad (\text{RAFL-P})$$

$$\text{s.t.} \quad \sum_A p_A r_A \leq \rho \quad (6)$$

$$\sum_i x_{A,ij} + r_A \geq 1 \quad \text{for all } j \in A \quad (7)$$

$$\sum_i (x_{A,ij} + u_{A,ij}) \geq 1 \quad \text{for all } j \in A \quad (8)$$

$$x_{A,ij} \leq y_i + y_{A,i} \quad \text{for all } j \in A, i \quad (9)$$

$$x_{A,ij} + u_{A,ij} \leq y_i + y_{A,i} + v_{A,i} \quad \text{for all } j \in A, i \quad (10)$$

$$\sum_i f_i^A y_{A,i} + \sum_{j \in A, i} c_{ij} x_{A,ij} \leq B \quad \text{for all } A \quad (11)$$

$$y_i, y_{A,i}, v_{A,i}, x_{A,ij}, u_{A,ij}, r_A \geq 0 \quad \text{for all } A, i, j. \quad (12)$$

Here y_i denotes the first-stage decisions. The variable r_A denotes if one exceeds the budget B in scenario A ; (6) limits the probability mass of such scenarios to at most ρ . The decisions $(x_{A,ij}, y_{A,i})$ and $(u_{A,ij}, v_{A,i})$ are intended to denote the decisions taken in scenario A in the two cases when does not exceed the budget, and when one does exceed the budget respectively. Correspondingly, (7) and (8) enforce that every client is assigned to a facility in these two cases, and (9) and (10) ensure that a client is only assigned to a facility opened in stage I or in that scenario in these two cases. Finally, (11) is the budget constraint for a scenario.

Let OPT be the optimal value of (RAFL-P). Given first-stage decisions $y \in [0, 1]^m$, let $\ell_A(y)$ denote the minimum cost of fractionally opening facilities and fractionally assigning clients in scenario A to open facilities (i.e., facilities opened to a combined extent of 1 in stage I and scenario A). Let $\mathcal{P} = [0, 1]^m$. As in Section 3, we Lagrangify (6) using a dual variable $\Delta \geq 0$ to obtain the problem $\max_{\Delta \geq 0} (-\Delta \rho + OPT(\Delta))$ where $OPT(\Delta) = \min_{y \in \mathcal{P}} h(\Delta; y)$, $h(\Delta; y) = f^I \cdot y + \sum_A p_A g_A(\Delta; y)$, and $g_A(\Delta; y)$ is the minimum value of $\sum_i f_i^A (y_{A,i} + v_{A,i}) + \sum_{j \in A, i} c_{ij} (x_{A,ij} + u_{A,ij}) + \Delta r_A$ subject to (7)–(12) (where the y_i 's are fixed now). As in Claim 3.1, it is easy to show that OPT is a lower bound on the optimal value of even the fractional risk-averse problem.

Theorem 4.3 *For any $\epsilon, \gamma, \kappa > 0$, in time $\text{poly}(\mathcal{I}, \frac{\lambda}{\epsilon \kappa \rho}, \log(\frac{1}{\gamma}))$, one can use *RiskAlg* to compute (with high probability) $(y, \{(x_A, y_A, u_A, v_A, r_A)\})$ that satisfies (7)–(12) with objective value $C \leq (1 + \epsilon)OPT + \gamma$ such that $\sum_A p_A r_A \leq \rho(1 + \kappa)$. This can be converted to a $(1 + 2\epsilon)$ -guarantee in the cost provided $f^I \cdot y + \ell_A(y) \geq 1$ for every $y \in [0, 1]^m$, $A \neq \emptyset$.*

Proof : Examining procedure RiskAlg, arguing that RiskAlg can be used to approximately solve (RAFL-P) involves two things: (a) coming up with a bound UB such that $\log \text{UB}$ is polynomially bounded so that one can restrict the search for the right value of Δ in RiskAlg; and (b) showing that an optimal solution to the SAA-version of the inner-minimization problem for any $\Delta \geq 0$ constructed in SA-Alg yields a solution to the true minimization problem that satisfies the approximation guarantee in Theorem 3.4.

There are two notable aspects in which the risk-averse facility location differs from risk-averse set cover. First, unlike in set cover, one cannot ensure that the cost incurred in a scenario is always 0 by choosing the first-stage decisions appropriately. Thus, the problem (RAFL-P) may in fact be infeasible. This creates some complications in coming up with an upper bound UB for use in RiskAlg. We show that one can detect by an initial sampling step that either the problem is infeasible, or come up with a suitable value for UB. Second, due to the non-covering nature of the problem, one needs to delve deeper into the structure of the dual LP for a scenario (after Lagrangifying (6)) to prove the closeness-in-subgradients property for SAA objective function constructed in SA-Alg and the true objective function.

Define $C_A = \sum_{j \in A} (\min_i c_{ij})$. This is the minimum possible assignment cost that one can incur in scenario A . We may determine with high probability using $O(\frac{1}{\rho\kappa})$ samples if $\Pr_A[C_A > B] > \rho$ or $\Pr_A[C_A > B] \leq \rho(1 + \frac{5\kappa}{28})$. In the former case, we can conclude that the problem is infeasible. In the latter case, we set $\hat{\rho} = \rho(1 + \frac{5\kappa}{28})$, $\hat{\kappa}$ such that $\hat{\rho}(1 + \hat{\kappa}) = \rho(1 + \kappa)$, and $\text{UB} = \frac{32(1+\varepsilon)(\sum_i f_i^1 + B)}{3\rho\kappa}$, and call procedure RiskAlg with these values of $\hat{\rho}$, $\hat{\kappa}$ and UB (and the given ε, γ). We prove in Claim 4.4 below that with this upper bound, $p^{(k)}, p'^{(k)} < \rho' = \hat{\rho}(1 + 3\hat{\kappa}/4)$; this is the only condition required for the search for Δ in RiskAlg.

Task (b) boils down to showing that the objective function $\hat{h}(\Delta; \cdot)$ of the SAA-problem in SA-Alg and the true problem $h(\Delta; \cdot)$ satisfy the conditions of Lemma 2.2. Due to the non-covering nature of the formulation, we need to derive additional insights about optimal dual solutions to $g_A(\Delta; y)$ to prove this. Lemma 4.5 proves that this holds with high probability, with $K = \lambda\|f^1\| + |\Delta|$, $\varrho = \varepsilon$ and $\xi = \frac{\eta\Delta}{2N}$. So by Lemma 2.2, the solution $\hat{y} = \operatorname{argmin}_{y \in \mathcal{P}} \hat{h}(\Delta; y)$ returned by SA-Alg satisfies the requirements of Theorem 3.4. As in the set cover problem, we may take $R = \sqrt{m}$, $V = \frac{1}{2}$, which ensures that the sample size is polynomially bounded. The proof of the conversion to a multiplicative guarantee is as in Theorem 3.3. ■

Recall that $\Delta_k \geq \text{UB}$ and $p^{(k)} = \sum_A p_A r_A^{(k)}$, where $(y, \{(x_A, y_A, u_A, v_A, r_A)\})$ is the solution returned by SA-Alg for Δ_k of cost $h(\Delta_k; y) \leq (1 + \varepsilon)OPT(\Delta_k) + \eta\Delta_k + \zeta$ with ε, η, ζ set as in RiskAlg.

Claim 4.4 We have $p^{(k)} < \rho'$ and $p'^{(k)} < \rho'$, where $\rho' = \hat{\rho}(1 + 3\hat{\kappa}/4)$.

Proof : Let $F = \sum_i f_i^1$ and $q = \Pr_A[C_A > B] \leq \rho(1 + 5\kappa/28)$. Consider the solution y with $y_i = 1$ for all i . For any $\Delta \geq 0$, we have $OPT(\Delta) \leq h(\Delta; y) \leq F + \sum_A p_A C_A + q\Delta \leq F + B + q\Delta$. Suppose $p^{(k)} \geq \rho' - \beta\hat{\rho}$. Then $\text{cost}^{(k)} - \eta\Delta_k \geq \Delta_k \hat{\rho}(1 + 9\hat{\kappa}/16) \geq \Delta_k \rho(1 + 9\kappa/16)$, where the last inequality follows since $\hat{\rho}(1 + \hat{\kappa}) = \rho(1 + \kappa)$ and $\hat{\rho} \geq \rho$. Also $(1 + \varepsilon)OPT(\Delta_k) + \zeta \leq 2(1 + \varepsilon)(F + B) + (1 + \varepsilon)q\Delta_k < 2(1 + \varepsilon)(F + B) + \Delta_k \rho(1 + 3\kappa/8)$ since $\varepsilon = \epsilon/6 \leq \hat{\kappa}/6 \leq \kappa/6$. But then $\text{cost}^{(k)} - \eta\Delta_k > (1 + \varepsilon)OPT(\Delta_k) + \zeta$ which gives a contradiction. So $p^{(k)} < \rho' - \beta\hat{\rho}$, which implies that $p'^{(k)} < \rho'$. ■

Lemma 4.5 With probability at least $1 - \delta$, $\hat{h}(\Delta; \cdot)$ and $h(\Delta; \cdot)$ satisfy the conditions of Lemma 2.2 with $K = \lambda\|f^1\| + |\Delta|$, $\varrho = \varepsilon$ and $\xi = \frac{\eta\Delta}{2N}$.

Proof : Consider a point $y \in \mathcal{P}$. Consider an optimal dual solution to $g_A(\Delta; y)$ where $\alpha_{A,j}^*, \psi_{A,j}^*, \beta_{A,ij}^*, \Gamma_{A,ij}^*, \theta_A^*$ are the optimal values of the dual variables corresponding to (7)–(11) respectively. Note that $g_A(\Delta; y)$ equals the objective value of this dual solution, which is given by

$$\sum_{j \in A} (\alpha_{A,j}^* + \psi_{A,j}^*) - \sum_i y_i \left(\sum_{j \in A} (\beta_{A,ij}^* + \Gamma_{A,ij}^*) \right) - B \cdot \theta_A^*.$$

We choose an optimal dual solution that minimizes $\sum_{i,j} \beta_{A,ij}^*$. As in Lemma 3.6, it is easy to show that the vectors $\hat{d}_y = (\hat{d}_{y,i})$ and $d_y = (d_{y,i})$ given by $\hat{d}_{y,i} = f_i^I - \sum_A \hat{p}_A \sum_{j \in A} (\beta_{A,ij}^* + \Gamma_{A,ij}^*)$ and $d_{y,i} = f_i^I - \sum_A p_A \sum_{j \in A} (\beta_{A,ij}^* + \Gamma_{A,ij}^*)$ are respectively subgradients of $\hat{h}(\Delta; \cdot)$ and $h(\Delta; \cdot)$ at y .

Now we claim that for every i , $\sum_j \beta_{A,ij}^* \leq \Delta$ and $\sum_j \Gamma_{A,ij}^* \leq f_i^A$. Given this, $\|\hat{d}_y\|, \|d_y\| \leq K$ where $K = \lambda \|f^I\| + \Delta$ for any $y \in \mathcal{P}$, so K is an upper bound on the Lipschitz constant of $h(\Delta; \cdot)$ and $\hat{h}(\Delta; \cdot)$.

The second inequality is a constraint of the dual, corresponding to variable $v_{A,i}$. Suppose $\beta_{A,ij}^* > 0$ for some j . The dual enforces the constraint $\alpha_{A,j}^* + \psi_{A,j}^* \leq c_{ij}(1 + \theta_A^*) + \beta_{A,ij}^* + \Gamma_{A,ij}^*$, corresponding to variable $x_{A,ij}$. We claim that this must hold at equality. By complementary slackness, we have $x_{A,ij}^* = y_i + y_{A,i}^*$ where $(x_A^*, y_A^*, u_A^*, v_A^*)$ is an optimal primal solution to $g_A(\Delta; y)$. So if $y_i > 0$ then $x_{A,ij}^* > 0$ and complementary slackness gives the desired equality. If $y_i = 0$ and the above inequality is strict, then we may decrease $\beta_{A,ij}^*$ while maintaining dual feasibility and optimality, which gives a contradiction to the choice of the dual solution. Thus, since the dual also imposes that $\psi_{A,j}^* \leq c_{ij} + \Gamma_{A,ij}^*$ (corresponding to $u_{A,ij}$), we have that $\beta_{A,ij}^* \leq \alpha_{A,j}^*$, so $\sum_j \beta_{A,ij}^* \leq \sum_j \alpha_{A,j}^* \leq \Delta$ (the last inequality follows from the dual constraint for r_A).

As in Lemma 3.7, if d is a subgradient of $h(\Delta; \cdot)$ at y and \hat{d} is a vector such that $|\hat{d}_i - d_i| \leq \omega f_i^I + \frac{\xi}{2m}$, then \hat{d} is an (ω, ξ) -subgradient of $h(\Delta; \cdot)$ at y .

Since $\mathbb{E}[\hat{d}_{y,i}] = d_{y,i}$ for every y and i , plugging in the sample size \mathcal{N} used in SA-Alg and using the Chernoff bound (Lemma 2.3), we obtain with probability at least $1 - \delta$, $|\hat{d}_{y,i} - d_{y,i}| \leq \frac{\varepsilon}{8\mathcal{N}} f_i^I + \frac{\eta\Delta}{4m\mathcal{N}}$ for all i , for every point y in the extended $\frac{\tau}{K\mathcal{N}}$ -net G_τ of \mathcal{P} . Thus, with probability at least $1 - \delta$, \hat{d}_y is an $(\frac{\varepsilon}{8\mathcal{N}}, \frac{\eta\Delta}{2\mathcal{N}})$ -subgradient of $h(\Delta; \cdot)$ at y for every $y \in G_\tau$. ■

We now discuss the rounding procedure. Analogous to part (i) of Theorem 3.2, it is not hard to see that if $(y, \{(x_A, y_A, u_A, v_A, r_A)\})$ is a solution satisfying (7)–(12) of objective value C with $P = \sum_A p_A r_A$, then for any $\varepsilon > 0$, taking $\hat{y} = y(1 + \frac{1}{\varepsilon})$ gives $\sum_i f_i \hat{y}_i + \sum_A \ell_A(\hat{y}) \leq (1 + \frac{1}{\varepsilon})C$ and $\Pr[\ell_A(\hat{y}) > (1 + \frac{1}{\varepsilon})B] \leq (1 + \varepsilon)P$. So now one can use a local approximation algorithm for 2-stage stochastic facility location (SUFL) to round \hat{y} .

Shmoys and Swamy [38] show that any LP-based c -approximation algorithm for the deterministic facility location problem (DUFL) that satisfies a certain “demand-obliviousness” property can be used to obtain a $\min\{2c, c + 1.52\}$ -approximation algorithm for SUFL, by using it in conjunction with the 1.52-approximation algorithm for DUFL in [26]. “Demand-obliviousness” means that the algorithm should round a fractional solution without having any knowledge about the client-demands, and is imposed to handle the fact that one does not have the second-stage solutions explicitly. There are some difficulties in applying this to our problem. First, the resulting algorithm for SUFL need not be local. Secondly, and more significantly, even if we do obtain a local approximation algorithm for SUFL by the conversion process in [38], the resulting algorithm may be *randomized*, if the c -approximation algorithm for DUFL is randomized. This is indeed the case in [38]; they obtain a randomized local 3.378-approximation algorithm using the demand-oblivious, randomized 1.858-approximation algorithm of Swamy [43]. (This was improved to a randomized local 3.25-approximation algorithm by Srinivasan [42], again using the algorithm in [43].) Using such a randomized local c' -approximation algorithm for SUFL would yield a random integer solution such that there is at least a $1 - \rho(1 + \kappa + \varepsilon)$ probability mass in scenarios for which the *expected cost* incurred, where the expectation is over the random choices of the algorithm, is at most $c'B(1 + \frac{1}{\varepsilon})$. But we would like to make the stronger claim that, *with high probability over the random choices of the algorithm*, we return a solution where the probability-mass of scenarios with cost at most $c'B(1 + \frac{1}{\varepsilon})$ is at least $1 - \rho(1 + \kappa + \varepsilon)$.

We can take care of both issues by imposing the following (sufficient) condition on the demand-oblivious algorithm for DUFL that is used to obtain an approximation algorithm for SUFL (via the conversion process in [38]): we require that with probability 1, the algorithm return an integer solution where *each client's assignment cost* is within some factor of its cost in the fractional solution. One can use the randomized

approximation algorithm of Swamy [43] or the deterministic Shmoys-Tardos-Aardal (STA) algorithm [40], both of which satisfy this condition. Given a fractional solution (x, y) to DUFL with facility cost F , for a parameter $\gamma \in (0, 1)$, the STA-algorithm returns an integer solution (\tilde{x}, \tilde{y}) with facility cost is at most F/γ , where for every j , $\sum_i c_{ij}\tilde{x}_{ij} \leq \frac{3}{1-\gamma} \cdot \sum_i c_{ij}x_{ij}$ (so for any demands, the total assignment cost is at most $\frac{3}{1-\gamma}$ times the fractional assignment cost). Taking $\gamma = \frac{1}{4}$ and applying the rounding procedure of [38] yields the following theorem.

Theorem 4.6 *For any $\epsilon, \kappa, \varepsilon > 0$, there is an $(5.52(1 + \epsilon + \frac{1}{\varepsilon}), 5.52(1 + \epsilon + \frac{1}{\varepsilon}), 1 + \kappa + \varepsilon)$ -approximation algorithm for risk-averse budgeted facility location.*

Remark 4.7 The local approximation algorithm for SUFL developed by [33] is unsuitable for our purposes, since this algorithm needs to know explicitly the second-stage fractional solution for each scenario, which is an exponential amount of information.

Budget constraints on individual components of the second-stage cost. Our techniques can be used to devise approximation algorithms for various fairly general risk-averse versions of facility location. Since the second-stage cost consists of two distinct components, the facility-opening cost and the client-assignment cost, one can consider risk-averse budgeted versions of the problem where we impose a *joint probabilistic budget constraint* on the total second-stage cost, and each component of the second-stage cost. That is, consider (RAFL-P) with the following additional constraints for each scenario A : $\sum_i f_i^A y_{A,i} \leq B_F$ and $\sum_{j,i} c_{ij}x_{A,ij} \leq B_C$. Here B_F and B_C are respectively budgets on the per-scenario facility-opening and client-assignment costs. To put it in words, (RAFL-P) augmented with the above constraints imposes the following joint probabilistic budget constraint:

$$\Pr_A \left[\text{total cost of scenario } A > B \text{ OR } \text{facility-cost of scenario } A > B_F \text{ OR } \text{assignment-cost of scenario } A > B_C \right] \leq \rho.$$

Note that by setting the appropriate budget to ∞ we can model the absence of a particular budget constraint. One can model various interesting situations by setting B, B_F, B_C suitably. For example, suppose we set $B_F = 0$ and $B = \infty$ (or equivalently $B = B_C$). Then we seek a minimum-cost solution where we want to choose facilities to open in stage I such that with probability at least $1 - \rho$, we can assign the clients in a scenario A to (a subset of) the stage I facilities while incurring assignment cost at most B_C . One can also consider risk-averse robust versions of the problem where we seek to minimize the first-stage cost plus the $(1 - \rho)$ -quantile of a certain component of the second-stage cost (i.e., the second-stage facility-opening, or assignment, or total cost). Employing the usual “guessing” trick, this gives rise to a budget problem where we have a budget constraint for a single component of the second-stage cost (that is, two of B, B_F and B_C are set to ∞). As before, the guarantees obtained for the budget problem (see below) translate to this risk-averse robust problem.

Our techniques can be used to solve this more general LP. Specifically, Theorem 4.3 continues to hold. But here we face the complication that even if we have a first-stage solution x to the fractional risk-averse problem for which we know that *there exist* second-stage feasible solutions that yield a solution of total expected cost C , it is not clear how to compute such feasible second-stage solutions. However, notice that RiskAlg not only returns a first-stage solution (with the above existence property) but also shows how to compute a suitable second-stage solution in each scenario A , which thus, allows us to specify completely a near-optimal solution to the LP-relaxation (where the RHS of (6) is $\rho(1 + \kappa)$). Whereas earlier we used these solutions only in the analysis, now they are part of the algorithm. In the rounding procedure, the first step, where we convert the solution to the LP-relaxation to a fractional solution to the risk-averse problem is unchanged. But we of course now need a stronger notion of “locality” from our approximation algorithm for

SUFL. We need an algorithm that approximately preserves (with probability 1) both the facility-opening and client-assignment components of the second-stage cost of each scenario. (Clearly, if the budget constraint is imposed on only one of the components then we only need the cost-preservation of that component.) Many LP-rounding algorithms for SUFL (such as the ones in [38, 42]) do in fact come with this stronger local guarantee. Thus, one can use these to obtain an approximation algorithm for the above risk-averse problem with multiple budget constraints.

Finally, we obtain the same approximation guarantees with non-uniform scenario budgets $\{(B^A, B_F^A, B_C^A)\}$. The only small detail here is that in order to obtain the upper bound UB for use in RiskAlg, we now determine if $\Pr[C_A > \min\{B_C^A, B^A\}]$ is greater than ρ or at most $\rho(1 + \frac{5\kappa}{28})$. In the former case, we conclude infeasibility, and in the latter, we set $\hat{\rho} = \rho(1 + \frac{5\kappa}{28})$, $\hat{\kappa}$ such that $\hat{\rho}(1 + \hat{\kappa}) = \rho(1 + \kappa)$, and $\text{UB} = \frac{32(1+\varepsilon)(\sum_i f_i^1 + \sum_j \max_i c_{ij})}{3\rho\kappa}$ and run RiskAlg with these values. (Note that we may assume that $B_C^A \leq \sum_j \max_i c_{ij}$ for all A .)

5 Sampling lower bounds

We now prove various lower bounds on the sample size required to obtain a bounded approximation guarantee for the risk-averse budgeted problem in the black-box model. We show that the dependence of the sample size on $\frac{1}{\kappa}$ for an *additive violation* of κ in the probability threshold is *unavoidable* in the black-box model even for the fractional risk-averse problem and even if we allow a bounded violation of the budget.

The crux of our lower bounds is the following observation. Consider the following problem. We are given as input a threshold $\varrho \in (0, \frac{1}{4})$ and a biased coin with probability q of landing heads, where the coin is given as a black-box; that is, we do not know q but may toss the coin as many times as necessary to “learn” q . The goal is to determine if $q \leq \varrho$ or $q > 2\varrho$; if $q \in (\varrho, 2\varrho]$ then the algorithm may answer anything. We prove that for any $\delta < \frac{1}{2}$, any algorithm that ensures error probability at most δ on every input must need at least $\mathcal{N}(\delta; \varrho) = \ln(\frac{1}{\delta} - 1)/4\varrho$ coin tosses for each threshold ϱ .

Lemma 5.1 *Let $\delta < \frac{1}{2}$ and $\mathcal{A}_{N(\delta; \varrho)}$ be an algorithm that has failure probability at most δ and uses at most $N(\delta; \varrho)$ coin tosses for threshold ϱ . Then, $N(\delta; \varrho) \geq \mathcal{N}(\delta; \varrho)$ for every $\varrho \in (0, \frac{1}{4})$.*

Proof : Suppose $N(\delta; \varrho) < \mathcal{N}(\delta; \varrho)$ for some $\varrho \in (0, \frac{1}{4})$. Let X be a random variable that denotes the number of times the coin lands heads. If $X = 0$ then the algorithm must say “ $q \leq \varrho$ ” with probability at least $1 - \delta$, otherwise the algorithm errs with probability more than δ on $q = 0$. But then for some $q_0 < \frac{1}{4}$ slightly greater than 2ϱ , we have $\Pr[X = 0] > (1 - 2\varrho)^{N(\delta; \varrho)} \geq \frac{\delta}{1 - \delta}$. So \mathcal{A} will say “ $q \leq \varrho$ ” (and hence, err) for $q = q_0$, with probability more than δ . ■

As a corollary we obtain that for any $\delta < \frac{1}{2}$, it is impossible to determine if $q = 0$ or if $q > 0$ with error probability at most δ using a bounded number of samples.

Now consider risk-averse budgeted set cover. We say that a solution is an (ϵ, γ) -optimal solution if its cost is at most $(1 + \epsilon)OPT + \gamma$. Suppose there is an algorithm \mathcal{A} for risk-averse budgeted set cover that on any input (with a black-box distribution) draws a bounded number of samples and returns an (ϵ, γ) -optimal solution with probability at least $1 - \delta$, $\delta < \frac{1}{2}$, where the probability-threshold is violated by at most κ . Consider the following risk-averse budgeted set-cover instance. There are three elements e_1, e_2, e_3 , three sets $S_i = \{e_i\}$, $i = 1, 2, 3$. The budget is $B \geq 6\gamma$ and the probability threshold is $\rho \leq \frac{1}{8(1+\epsilon)}$. The costs are $w_{S_i}^1 = B$ for all i , and $w_{S_1}^A = 0$, $w_{S_2}^A = w_{S_3}^A = 2B/3$ for every scenario A . Let $\kappa < \frac{1}{4}$. There are 3 scenarios: $A_0 = \emptyset$, $A_1 = \{e_1, e_2, e_3\}$, $A_2 = \{e_2, e_3\}$ with $p_{A_1} = \rho - \kappa$, $p_{A_2} = 1 - p_{A_1} - p_{A_0}$. Observe that if $p_{A_2} \leq \kappa$, then $OPT \leq \rho \cdot 4B/3$, and every (ϵ, γ) -optimal solution must have $x_{S_1} + x_{S_2} + x_{S_3} \leq \frac{1}{3}$. But if $p_{A_2} > 2\kappa$ (which is possible since $\rho < 1$) then any solution where the probability of exceeding the budget

is at most $\rho + \kappa$ must have $x_{S_2} + x_{S_3} \geq \frac{1}{2}$, otherwise the cost in both scenarios A_1 and A_2 will exceed B . Thus, algorithm \mathcal{A} can be used to determine if $p_{A_2} \leq \kappa$ or $p_{A_2} > 2\kappa$. This is true even if we allow the budget to be violated by a factor $c < \frac{10}{9}$ since we must still have $x_{S_2} + x_{S_3} > \frac{1}{3}$ if $p_{A_2} > 2\kappa$; choosing $B \gg 1$, $\rho \ll 1$, we can allow an arbitrarily large budget-violation. So since \mathcal{A} has failure probability at most δ , by Lemma 5.1, it must draw $\Omega(\frac{1}{\kappa})$ samples.

Taking $\kappa = 0$ shows that obtaining guarantees without violating the probability threshold is impossible with a bounded sample size, whereas taking $\kappa = \kappa\rho$ shows that a *multiplicative* $(1 + \kappa)$ -factor violation of the probability threshold requires $\Omega(\frac{1}{\kappa\rho})$ samples. Moreover, taking $\rho = 0$ shows that one cannot hope to achieve any approximation guarantees in the (standard) budget model with black-box distributions.

Theorem 5.2 *For any $\epsilon, \gamma > 0$, $\delta < \frac{1}{2}$, every algorithm for risk-averse budgeted set cover that returns an (ϵ, γ) -optimal solution with failure probability at most δ using a bounded number of samples*

- *must violate the probability threshold on some input;*
- *requires $\Omega(\frac{1}{\kappa})$ samples if the probability-threshold is violated by at most an additive κ ;*
- *requires $\Omega(\frac{1}{\kappa\rho})$ samples if the probability-threshold is violated by at most a multiplicative $(1 + \kappa)$ -factor.*

The proof of impossibility of approximation in the standard robust model with a bounded sample size is even simpler. Consider the following set cover instance. We have a single element e that gets “activated” with some probability p ; the cost of the set $S = \{e\}$ is 1 in stage I and some large number M in stage II. If $p = 0$ then $OPT = 0$, otherwise $OPT = 1$. Thus, it is easy to see that an algorithm returning an (ϵ, γ) -optimal solution can be used to distinguish between these two cases (it should set $x_S \leq \gamma$ in the former case, and x_S sufficiently large in the latter).

References

- [1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26,1487–1503, 2002.
- [2] E. M. L. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society, Series B*, 17:173–184; discussion 194–203, 1955.
- [3] D. Bertsimas and M. Sim. The price of robustness, *INFORMS Journal on Operations Research*, 52:35–38, 2004.
- [4] J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, NY, 1997.
- [5] J. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer-Verlag, NY, 2000.
- [6] G. Calafiore and M. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [7] M. Charikar, C. Chekuri, and M. Pál. Sampling bounds for stochastic optimization. *Proceedings, 9th RANDOM*, pages 257–269, 2005.
- [8] A. Charnes and W. Cooper. Uncertain convex programs: randomized solutions and confidence levels. *Management Science*, 6:73–79,1959.
- [9] V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4:233–235, 1979.

- [10] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.
- [11] K. Dhamdhere, V. Goyal, R. Ravi, and M. Singh. How to pay, come what may: approximation algorithms for demand-robust covering problems. *Proceedings, 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 367–378, 2005.
- [12] S. Dye, L. Stougie, and A. Tomaszgard. The stochastic single resource service-provision problem. *Naval Research Logistics*, 50(8):869–887, 2003. Also appeared as “The stochastic single node service provision problem”, COSOR-Memorandum 99-13, Dept. of Mathematics and Computer Science, Eindhoven Technical University, Eindhoven, 1999.
- [13] E. Erdoğan and G. Iyengar. On two-stage convex chance constrained problems. *Math. Methods of Operations Research*, 65(1):115–140, 2007.
- [14] U. Feige, K. Jain, M. Mahdian, and V. Mirrokni. Robust combinatorial optimization with exponential scenarios. *Proceedings, 13th IPCO*, pages 439–453, 2007.
- [15] A. Goel and P. Indyk. Stochastic load balancing. *Proceedings, 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 579–586, 1999.
- [16] D. Golovin, V. Goyal, and R. Ravi. Pay today for a rainy day: improved approximation algorithms for demand-robust min-cut and shortest path problems. *Proceedings, 23rd STACS*, pages 206–217, 2006.
- [17] A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. *Proceedings, 36th Annual ACM Symposium on Theory of Computing*, pages 417–426, 2004.
- [18] A. Gupta, M. Pál, R. Ravi, and A. Sinha. What about Wednesday? Approximation algorithms for multistage stochastic optimization. *Proceedings, 8th APPROX*, pages 86–98, 2005.
- [19] A. Gupta, R. Ravi, and A. Sinha. An edge in time saves nine: LP rounding approximation algorithms for stochastic network design. In *Proceedings, 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 218–227, 2004.
- [20] A. Hayrapetyan, C. Swamy, and É. Tardos. Network design for information networks. *Proceedings, 16th SODA*, pages 933–942, 2005.
- [21] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [22] N. Immerlica, D. Karger, M. Minkoff, and V. Mirrokni. On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems. *Proceedings, 15th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 684–693, 2004.
- [23] P. Jorion. *Value at Risk: A New Benchmark for Measuring Derivatives Risk*. Irwin Professional Publishers, New York, 1996.
- [24] J. Kleinberg, Y. Rabani, and É. Tardos. Allocating bandwidth for bursty connections. *SIAM Journal on Computing*, 30(1):191–217, 2000.
- [25] A. J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal of Optimization*, 12:479–502, 2001.

- [26] M. Mahdian, Y. Ye, and J. Zhang. Approximation algorithms for metric facility location problems. *SIAM Journal on Computing*, 36:411–432, 2006.
- [27] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- [28] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In G. Calafiore and F. Dabbene, editors. *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer-Verlag, 2005.
- [29] A. Prékopa. Contributions to the theory of stochastic programming. *Mathematical Programming*, 4:202–221, 1973.
- [30] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, 1995.
- [31] A. Prékopa. Probabilistic programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Mgmt. Sc.*, North-Holland, Amsterdam, 2003.
- [32] M. Pritsker. Evaluating value at risk methodologies. *Journal of Financial Services Research*, 12(2/3):201–242, 1997.
- [33] R. Ravi and A. Sinha. Hedging uncertainty: approximation algorithms for stochastic optimization problems. *Mathematical Programming, Series A*, 108:97–114, 2006.
- [34] R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.
- [35] A. Ruszczyński and A. Shapiro. Editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Mgmt. Sc.*, North-Holland, Amsterdam, 2003.
- [36] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors. *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer-Verlag, 2005.
- [37] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Mgmt. Sc.*, North-Holland, Amsterdam, 2003.
- [38] D. B. Shmoys and C. Swamy. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *Journal of the ACM*, 53(6):978–1012, 2006.
- [39] D. B. Shmoys and C. Swamy. Stochastic optimization is (almost) as easy as deterministic optimization. *Proceedings, 45th Annual IEEE FOCS*, pages 228–237, 2004.
- [40] D. B. Shmoys, É. Tardos, and K. I. Aardal. Approximation algorithms for facility location problems. *Proceedings, 29th Annual ACM Symposium on Theory of Computing*, pages 265–274, 1997.
- [41] A. M-C. So, J. Zhang, and Y. Ye. Stochastic combinatorial optimization with controllable risk aversion level. *Proceedings, 9th APPROX*, pages 224–235, 2006.
- [42] A. Srinivasan. Approximation algorithms for stochastic and risk-averse optimization. *Proceedings, 18th SODA*, pages 1305–1313, 2007.
- [43] C. Swamy. *Approximation Algorithms for Clustering Problems*. Ph.D. thesis, Cornell University, Ithaca, NY, 2004. <http://www.math.uwaterloo.ca/~cswamy/theses/master.pdf>.

- [44] C. Swamy and D. B. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. *ACM SIGACT News*, 37(1):33–46, March 2006. Also appeared in *Proceedings, 26th FSTTCS*, pages 5–19, 2006.
- [45] C. Swamy and D. B. Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. <http://www.math.uwaterloo.ca/~cswamy/papers/multistage-journ.pdf>. Preliminary version in *Proceedings, 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 357–366, 2005.

A A bicriteria approximation for the Shmoys-Swamy class of 2-stage stochastic LPs in the standard budget model ($\rho = 0$)

Here we sketch how one can obtain a bicriteria approximation algorithm for the class of 2-stage LPs introduced in [38] in the standard budget model (that is, where we have a *deterministic* budget constraint). We show that for any $\rho > 0$, in time inversely proportional to ρ , one can obtain a near-optimal solution where the total probability-mass of scenarios where the budget is violated is at most ρ . We consider the following class of 2-stage stochastic LPs [38]² in the standard budget model.

$$\begin{aligned} \min \quad & h(x) = w^1 \cdot x + \sum_A p_A f_A(x) \quad \text{subject to} \quad x \in \mathcal{P} \subseteq \mathbb{R}_+^m, \quad f_A(x) \leq B \quad \text{for all } A \quad (\text{Stoc-P}) \\ \text{where } f_A(x) = \quad & \min \quad w^A \cdot r_A + q^A \cdot s_A \\ \quad & \text{s.t.} \quad D^A s_A + T^A r_A \geq j^A - T^A x \\ \quad & r_A, s_A \geq \mathbf{0}, \quad r_A \in \mathbb{R}^m, s_A \in \mathbb{R}^\ell. \end{aligned}$$

Here (a) $T^A \geq \mathbf{0}$ for every scenario A , and (b) for every $x \in \mathcal{P}$, $\sum_{A \in \mathcal{A}} p_A f_A(x) \geq 0$ and the primal and dual problems corresponding to $f_A(x)$ are feasible for every scenario A . It is assumed that $\mathcal{P} \subseteq B(\mathbf{0}, R)$, and that \mathcal{P} contains a ball of radius V ($V \leq 1$) where $\ln(\frac{R}{V})$ is polynomially bounded. Define $\lambda = \max(1, \max_{A \in \mathcal{A}, S} \frac{w_S^A}{w_S^1})$; we assume that λ is known. Let OPT be the optimum value and \mathcal{I} denote the input size.

It is possible to adapt the proofs in [38, 7, 45] to obtain the bicriteria guarantee and one can also prove an SAA theorem in the style of [45, 7]. But perhaps, the simplest proof, which we now describe, is obtained using the ellipsoid-based algorithm in [38]. Let $\mathcal{P}' = \{x \in \mathcal{P} : f_A(x) \leq B \text{ for all } A\}$. Note that unlike in the case where we have a probabilistic budget constraint, \mathcal{P}' is a *convex set*.

Consider running the ellipsoid-based algorithm in [38] with the following modification. Suppose we wish to return a solution of value at most $(1 + \epsilon)OPT + \gamma$. Let $N = \text{poly}(m, \ln(\frac{KR}{V\gamma}))$ be a suitably large value that is equal to the number of iterations of the ellipsoid method. Let $\rho' = \rho/N$. Suppose the center of the current ellipsoid is $x \in \mathcal{P}$. Using $O(\frac{1}{\rho'})$ samples one can determine with high probability if $\Pr_A[f_A(x) > B] > \rho'/2$ or if $\Pr_A[f_A(x) > B] \leq \rho'$. In the former case, by sampling again $O(\frac{1}{\rho'})$ times, with very high probability, we can obtain a scenario A such that $f_A(x) > B$. Now we compute a subgradient $d_{A,x}$ of $f_A(\cdot)$ (which is obtained from an optimal dual solution to $f_A(x)$) at x , and use the inequality $d_{A,x} \cdot (y - x) \leq 0$ to cut the current ellipsoid. Notice that this is a *valid inequality* since for any $y \in \mathcal{P}'$, by the definition of a subgradient, we have $0 > f_A(y) - f_A(x) \geq d_{A,x} \cdot (y - x)$. In the latter case, where we detect that $\Pr_A[f_A(x) > B] \leq \rho'$, we continue as in the algorithm in [38]: we mark the current point x and use an approximate subgradient of $h(\cdot)$ at x to cut the current ellipsoid. Proceeding this

²This was stated in [39] with extra constraints $B^A s_A \geq h^A$, but this is equivalent to $\begin{pmatrix} B^A \\ D^A \end{pmatrix} s_A + \begin{pmatrix} \mathbf{0} \\ T^A \end{pmatrix} r_A \geq \begin{pmatrix} h^A \\ j^A \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ T^A \end{pmatrix} x$.

way we obtain a collection of marked points x_1, \dots, x_k , where $k \leq N$, such that with high probability, $\Pr_A[f_A(x_i) > B] \leq \rho'$ for each x_i , and by the analysis in [38] we have that $\min_i h(x_i)$ is “close” to OPT .

The next step in the algorithm in [38] is to find a point in the convex hull of x_1, \dots, x_k whose value is close to $\min_i h(x_i)$ (procedure FindMin). Notice that for any point y in the convex hull of x_1, \dots, x_k , we have $\Pr_A[f_A(y) > B] \leq k\rho' \leq \rho$: for any scenario A with $f_A(x_i) \leq B$ for all i , the convexity of $f_A(\cdot)$ implies that $f_A(y) \leq B$. Thus, although the set $\{x \in \mathcal{P} : \Pr_A[f_A(x) > B] \leq \rho\}$ is not convex, this does not present a problem for us. So one can use procedure FindMin in [38] to return a point y such that $h(y) \leq (1 + \epsilon)OPT + \gamma$ where $\Pr_A[f_A(y) > B] \leq \rho$.