Risk-Averse Stochastic Optimization: Probabilistically-Constrained Models and Algorithms for Black-Box Distributions

(Extended Abstract)

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 a 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, and facility location problems, in the risk-averse budget and robust models with *black-box distributions*. Our main contribution is to 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 simple rounding procedure that shows that one can exploit existing LP-based approximation algorithms for the 2-stage-stochastic and/or deterministic counterpart of the problem to round the fractional solution and obtain an approximation algorithm for the risk-averse problem. To the best of our knowledge, these are the *first approximation results* for problems involving *probabilistic constraints and black-box distributions*. A notable feature of our scheme is that it extends easily

to handle a significantly richer class of risk-averse problems, where we impose a *joint probabilistic budget constraint* on different components of the second-stage cost. Consequently, we also obtain approximation algorithms in the setting where we have a joint budget constraint on different portions 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*. An important and widely-used model 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, $c(x) + E_A[f_A(x, y_A)]$. 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 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 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 [21, 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 [21], 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 scenario-dependent budgets, 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; that is, one can use an approximation algorithm for the budget problem to obtain a near-optimal solution to the robust problem: scaling down the second-stage costs (and B) to make the second-stage contribution negligible) makes the objective functions of the budget and robust problems essentially identical

(modulo the constant term B). 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. But this seems rather specialized, 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 and the (distribution-based) robust model suffer from certain common drawbacks. A serious *algorithmic limitation* (see Section 7) is that for almost any (non-trivial) stochastic problem (e.g., *fractional* stochastic set cover with at most 3 {elements,sets,scenarios}), one cannot obtain *any* approximation guarantees in the black-box setting using any *bounded number of samples* (even allowing for a bounded budget inflation). Intuitively, this is because 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 obtain a low-cost solution in the robust model. Notice also that both 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. In contrast, risk-models considered in the finance and stochastic-optimization literature, such as the mean-risk model [29], value-at-risk (VaR) [31, 25, 33], conditional VaR [35], 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 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. We can now incorporate this solution concept into the model to arrive at the following new budget model, which we call the *risk-averse budget model*. We are now *given* a probability threshold

$\rho \in [0, 1]$ and a budget B, and we seek $(x, \{y_A\})$ so as to minimize $c(x) + E_A[f_A(x, y_A)]$ 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 $\{f_A(x, y_A)\}_{A \in \mathcal{A}}$ (i.e., the smallest B such that $\Pr_A[f_A(x) > B] \leq \rho$). Notice that ρ allows us to control the risk-aversion level and tradeoff risk-averseness against conservatism (as in [41]). Taking $\rho = 1$ in the risk-averse budget model gives the standard 2-stage recourse model, and $\rho = 0$, yields the (standard) budget- and robust models. 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 5) including the set cover, vertex cover, multicut on trees, 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 inflate the budget; 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 various more general settings, the most noteworthy one being where we have a joint budget constraint on different portions of the second-stage cost. We can also handle non-uniform scenario-budgets, and a generalization where the goal is to minimize $c(x)$ plus a weighted combination of $E_A[f_A(x, y_A)]$ and $Q_\rho[f_A(x, y_A)]$. We mainly consider risk-averse budgeted problems in the sequel, since (as in the case of the (standard) budget- and robust-problems) the risk-averse robust problem reduces to the risk-averse budgeted problem (see Sections 4.1 and 5).

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 4.3). We show that in the black-box setting, for a wide variety of 2-stage problems, for any $\epsilon, \kappa > 0$, in time $\mathsf{poly}\left(\frac{\lambda}{\epsilon \kappa \rho}\right)$, 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 7 that the dependence on $\frac{1}{\kappa \rho}$, is unavoidable in the black-box setting (as is the

dependence on λ [46]). One major difficulty faced in solving a probabilistic program such as ours, which one does not encounter for 2-stage problems, is that the feasible region of *even the fractional risk-averse problem* (i.e., where one can take fractional decisions) 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 fact, this is often *NP*-hard). We formulate an LP-relaxation (of even the fractional problem), where for every scenario A, we introduce a variable r_A, to indicate whether the budget is exceeded in that scenario, along with two sets of decision variables to denote the decisions taken in these two cases, and impose the constraint $\sum_A p_A r_A \leq \rho$ (in addition to various problem-specific constraints). This constraint however *couples* the different scenarios (notice again the contrast with the standard 2-stage recourse model), and to get around this difficulty, we use a *Lagrangian-relaxation* approach, where we decouple the scenarios by Lagrangifying this coupling constraint; Section 2 gives a more detailed outline of our algorithm. A key notable feature of our scheme (and the Lagrangian-relaxation approach) is that it extends easily to handle a *richer class of risk-averse problems*, where we impose a *joint probabilistic budget constraint* on different components of the second-stage cost (e.g., facility- and assignment-costs in risk-averse facility location).

The second component is a simple, general rounding procedure (Theorem 4.1) complementing the above scheme. We round an LP-solution to a solution to the fractional risk-averse problem losing a certain factor in the solution cost, budget, and probability of budget-violation. This then allows us to use suitable LP-based algorithms for the deterministic or (non-risk-averse) 2-stage analogue of the problem to obtain a near-optimal solution to the (integer) risk-averse problem. For example, for various covering problems, given an LP-based c-approximation algorithm for the deterministic analogue, we obtain an $O(c)$-approximation for the risk-averse problem using the $2c$-approximation algorithm for the 2-stage problem in [38].

Our techniques, and in particular, our scheme, yield versatile tools that we believe will find application in the design of approximation algorithms for other risk-averse problems and probabilistic programs.

Related work. Stochastic optimization is a field with a vast amount of literature; (see, e.g., [3, 31, 36]), but these problems have only recently been studied from an approximation-algorithms perspective. We survey the work that is most relevant to ours. Various 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. As mentioned earlier, the (standard) budget model was first considered by Gupta et al. [21], who designed approximation algorithms for stochastic network design problems in this model, and Dhamdhere et al. [11] introduced the demand-robust model (which we call the robust model), obtaining algorithms for the robust versions of various combinatorial optimization problems (and [18] obtains certain improvements). All these works focus on the polynomial-scenario setting. Feige et al. [14], and subsequently [26], considered the robust model with exponentially many scenarios that are specified implicitly via a cardinality constraint, and derived approximation algorithms in this more general model.

There is a large body of work in the finance and stochastic-optimization literature, dating back to [29], that deals with risk-modeling and optimization; see e.g., [35, 37] 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 [25, 33]. Problems involving probabilistic constraints are called *probabilistic* or *chance-constrained programs* [8] in the stochastic-optimization literature, and have been extensively studied (see, e.g., [32]). Some recent work [5, 30, 13] has focused on replacing the probabilistic constraint by more tractable ones 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. Although some approximation results are obtained in [5, 30, 13], they are obtained under various restrictions on the random variables (continuous) and distribution (concentration-of-measure), which are not satisfied by discrete problems. To the best of our knowledge, there is no prior work in this literature on the design of *efficient algorithms* with *provable worst-case guarantees* for discrete risk-optimization or probabilistic-programming problems.

In the CS literature, [27, 16] consider stochastic bin packing and knapsack with probabilistic constraints and obtained novel approximation algorithms for these problems. These results are however obtained for *specialized distributions* where the item sizes are *independent* random variables, which is far from the black-box setting. So et al. [41] consider the problem of minimizing the first-stage cost plus a risk-measure called the *conditional VaR* (CVaR) [35] and obtain approximation algorithms for various problems in the black-box setting (using quite different methods). In their model, the fractional problem yields a *convex program*, and they are able to use a nice representation theorem in [35] to convert their problem into a 2-stage problem and then adapt the methods in [6]. In our case, the non-convexity inherent in the probabilistic constraint creates various difficulties and we consequently need to work harder to obtain our result. Two recent unpublished manuscripts—[17], which is independent of our work, and [1], which appeared after a preliminary version of our work [44] appeared on the arXiv (and cites [44])—also consider probabilistic constraints, but in (specialized) *non-black-box* settings. Their problems fall into our risk-averse models, so in various cases, our general results yield guarantees for their specific problems. Goyal and Ravi [17] observe that approximating even one-stage problems is "hard" even when scenarios consist of only two "elements", and proceed to consider various one-element-per-scenario (1-PS) problems (in the poly-scenario setting). One-stage problems can be cast as 2-stage risk-averse budgeted problems by setting $B = 0$ and negligible (but positive) second-stage costs, so our results in Section 6 for 1-PS problems also apply to their problems. Agrawal et al. [1] consider, in our terminology, the one-stage and two-stage stochastic versions of set-cover and k-center in the *independent-activation* (IA) model. By exploiting independence, [1] design algorithms for stochastic k-center and one-stage set-cover that do not inflate the budget or ρ. Section 6 shows that IA-problems (and *a priori* stochastic problems) can often be reduced to the 1-PS setting. (We obtained these results after [1] appeared.) We also note that their adaptive set-cover problem can be cast as risk-averse budgeted set cover and so in contrast to their negative result, our results imply a *bicriteria* decision algorithm that inflates the budget by $O(\ln n)$ and ρ by 2 (say) if the problem is feasible.

The first approximation result for 2-stage recourse problems appears to be due to Dye, Stougie, and Tomasgard [12]. Starting with the work of Ravi and Sinha [34] and Immorlica et al. [24], which gave approximation algorithms for various 2-stage problems in the polynomial scenario and IA settings, various approximation results for 2-stage problems have been obtained; see, e.g., the survey [45]. Approximation algorithms in the black-box setting were first obtained by Gupta et al. [19], and subsequently by Shmoys and Swamy [38]. Multistage recourse problems in the black-box model were considered by [20, 46]; both obtain approximation algorithms with guarantees that deteriorate with the number of stages. Srinivasan [42] obtained improved guarantees for set cover and vertex cover that do not depend on the number of stages.

Our approximation scheme makes use of the SAA

method, which is an appealing method often used to solve stochastic problems. The effectiveness of this method has been analyzed in [28, 6, 46]. Kleywegt et al. [28] prove a non-polynomial bound on the sample size required for general 2-stage problems. Subsequently [6, 46] obtained improved polynomial bounds for a large class of structured 2-stage problems. The proof in [46], which also works for multistage programs, leverages approximate subgradients; our proof uses portions of their analysis. The proof in [6] applies only to 2-stage programs, but shows that even approximate solutions to the SAA problem translate to approximate solutions to the original problem.

2   Overview of our approach

Let (RA-P): min $\big(h(x) := c(x) + \mathrm{E}_A\big[f_A(x)\big]\big)$ s.t. $x \in \mathcal{F}$, $\mathrm{Pr}_A[f_A(x) > B] \le \rho$, denote the (discrete) risk-averse problem, where $\mathcal{F}$ is the finite feasible region of first-stage decisions, $f_A(x)$ is the minimum value of $f_A(x, y_A)$ over all feasible recourse actions y_A. A natural sampling-based approach for attacking (RA-P), which we call the "direct sample average approximation (SAA) approach", is to sample a certain number of scenarios to estimate the scenario probabilities, and then solve the following SAA analogue of the problem (SA-P): min $\big(\widehat{h}(x) := c(x) + \sum_A \widehat{p}_A f_A(x)\big)$ s.t. $x \in \mathcal{F}$, $\widehat{\mathrm{Pr}}_A[f_A(x) > B] \le \widehat{\rho} := \rho(1 + \kappa)$. Here $\widehat{p}_A$ is the frequency of scenario A, and $\widehat{\mathrm{Pr}}$ denotes the probability wrt. $\widehat{p}$. One can generalize the arguments in [6], to show (see Theorem B.1) that if one constructs (SA-P) using $\mathsf{poly}\big(\mathcal{I}, \frac{\lambda}{\epsilon\kappa\rho}\big)$ samples and can compute an α-approximate solution $\widehat{x}$ to (SA-P), then, with high probability, $h(\widehat{x}) \le \big(\alpha + O(\epsilon)\big) \cdot OPT_{\text{(RA-P)}}$ and $\mathrm{Pr}_A[f_A(\widehat{x}) > B] \le \rho\big(1 + O(\kappa)\big)$. (Throughout, "with high probability" means that we can ensure a failure probability of δ with $\mathsf{poly}\big(\ln(\frac{1}{\delta})\big)$ samples.) Notice however that this only says that $\widehat{x}$ in conjunction with the *optimal recourse solutions to each scenario A* yields a solution of cost at most $\big(\alpha + O(\epsilon)\big) \cdot OPT_{\text{(RA-P)}}$. The recourse problem is often *NP*-hard, and so using a β-approximation algorithm for the recourse problem yields only a worse $(\alpha\beta + \epsilon)$-approximation to the black-box problem. This introduces an undesirable approximation gap between the poly-scenario SAA problem and the black-box true problem; e.g., for risk-averse budgeted set cover, this only yields an $O(\ln^2 n)$-guarantee, whereas one can obtain an $O(\ln n)$-guarantee. Also, we still need to solve the poly-scenario risk-averse problem (SA-P), which is a challenging task, and remains challenging even if one moves to a fractional version of the problem (which would be one way of avoiding the issue of approximation gap, since the fractional recourse

problem is easily solvable).

To circumvent these difficulties, we adopt a different approach where we directly attack the black-box problem (instead of approximating it via an SAA problem), and hence obtain matching performance guarantees for both black-box and poly-scenario problems. Our approach also has the significant benefit that it also leads to approximation algorithms for more general risk-averse problems (see below). Since even the fractional version of the problem is a *non-convex* optimization problem, we formulate an LP-relaxation of (even) the (fractional) black-box problem, where we introduce a variable r_A for every scenario A intended to indicate if the budget is exceeded in scenario A, and impose the constraint $\sum_A p_A r_A \le \rho$ to capture the probabilistic budget constraint. This LP may have an exponential number of both variables and constraints (since both are indexed by scenarios), and moreover, the scenarios are now coupled by the above constraint. To get around the difficulty posed by coupling, we Lagrangify the above constraint using a dual variable $\Delta \ge 0$ to obtain a Lagrangian relaxation (LD), which has a structured *2-stage LP* (2St-P) embedded in it. Our goal now is to perform a search for the "right" value of Δ. That is, we consider different values of Δ, computing, for each Δ, a near-optimal solution to (2St-P) (using say the SAA method), and then return the solution whose $\sum_A p_A r_A$ is closest to ρ. However, it turns out that the strong near-optimality guarantees obtained in [38, 46, 6] for the classes of 2-stage programs considered therein, where one obtains an FPTAS, do not apply to (2St-P); in particular, (2St-P) does not fall into the class of programs considered in [38, 46], and the analysis in [38, 46, 6] only yields a super-polynomial sample size for obtaining a $(1 + \epsilon)$-optimal solution to (2St-P) (see the discussion following Lemma 4.1). The key insight here is to realize that aiming for a $(1 + \epsilon)$-approximation is *not* the right notion of near-optimality for (P). We make the crucial observation that one can instead obtain a (weak) "near-optimality" guarantee for (P) (see Lemma 4.1) that (a) is weak enough that one can prove such a guarantee with polynomial sample size using the SAA method by leveraging the approximate-subgradients based analysis in [46], and (b) yet is strong enough that it can be exploited in the search for the "right" Δ. Rounding this fractional solution yields (matching) approximation guarantees for various (poly-scenario and) black-box risk-averse problems.

A notable benefit of the Lagrangian-relaxation approach is that it is flexible enough to yield an approximation scheme for solving the LP-relaxation of a richer class of risk-averse problems, where there is a joint probabilistic budget constraint on different components

of the second-stage cost. For example, in risk-averse budgeted facility location, one can incorporate a constraint like $\Pr_A[$(total cost of $A)> B$, or (facility-cost of $A)> B_F$, or (assignment-cost of $A)> B_C] \leq \rho$ (see "Facility location" in Section 5). Obtaining an approximation result for this general risk-averse problem is significantly more challenging, due to the fact that the recourse action in a scenario is no longer determined solely by the first-stage decision (unlike before). The "direct SAA approach" is ill-equipped to deal with this complication as its analysis crucially relies on the fact that given a first-stage decision x, *both* the SAA and true problems will take the *same recourse action in a scenario*, which is no longer true (and we do not have a good handle on the various choices available to the SAA and true problems). In fact, an approach that decouples scenarios appears necessary to make any headway here.

3 Preliminaries

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 3.1. *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 [38] of f at the point $u \in \mathcal{P}$ if for every $v \in \mathcal{P}$, we have $f(v) - f(u) \geq \hat{d} \cdot (v - u) - \omega f(v) - \omega f(u) - \xi$.*

One can infer that, letting d_x denote a subgradient of f at x, the Lipschitz constant of f is at most $\max_x \|d_x\|$. Let $K > 0$, and $\tau, \varrho > 0$ be parameters with $\tau < 1$. Let $N = \log\left(\frac{2KR}{\tau}\right)$. 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 $G_\tau = G'_\tau \cup \{x + t(y - x), y + t(x - y) : x, y \in G'_\tau, \ t = 2^{-i}, \ i = 1, \ldots, N\}$. We call G_τ and G'_τ, an $\frac{\tau}{KN}$-*net* and an *extended $\frac{\tau}{KN}$-net* respectively of $\mathcal{P}$. 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 $|G_\tau| = \mathsf{poly}\left(\log\left(\frac{KR}{V\tau}\right)\right)$ [46]. The following result from [46], which we have adapted to our setting, will be our main tool for analyzing the SAA method.

LEMMA 3.1. *([46]) Let $\widehat{f}$ 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 $\widehat{f}(.)$ and an $\left(\frac{\varrho}{8N}, \frac{\xi}{2N}\right)$-subgradient of $f(.)$ at x. Let $\hat{x} = \operatorname{argmin}_{x \in \mathcal{P}} \widehat{f}(x)$. Then, $f(\hat{x}) \leq (1 + \varrho) \min_{x \in \mathcal{P}} f(x) + 6\tau + \xi$.*

LEMMA 3.2. (CHERNOFF-HOEFFDING BOUND [23]) *Let $X_1, \ldots, X_N$ be iid random variables with $X_i \in [0, 1]$ and $\mu = \mathrm{E}[X_i]$. Then, $\Pr\left[\left|\frac{1}{N} \sum_i X_i - \mu\right| > \epsilon\right] \leq 2e^{-2\epsilon^2 N}$.*

4 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 (RSC), 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^{I}_S, and is w^A_S 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^A_S are only revealed when we sample scenario A; thus, the "input size", denoted by $\mathcal{I}$ is $O(m + n + \sum_S \log w_S + \log B)$.

Let $\mathcal{P} = [0, 1]^m$. For a point $x \in \mathcal{P}$, 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. 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 (even) the (fractional) problem. Throughout we use A to index the scenarios in $\mathcal{A}$, and S to index the sets in $\mathcal{S}$.

$$\min \quad \sum_S w^{\mathrm{I}}_S x_S + \sum_{A,S} p_A\left(w^A_S y_{A,S} + w^A_S z_{A,S}\right) \quad \text{(RSCP)}$$

$$\text{s.t.} \qquad \sum_A p_A r_A \leq \rho \qquad (4.1)$$

$$\sum_{S:e \in S}\left(x_S + y_{A,S}\right) + r_A \geq 1 \quad \forall A, e \in A, \quad (4.2)$$

$$\sum_{S:e \in S}\left(x_S + y_{A,S} + z_{A,S}\right) \geq 1 \quad \forall A, e \in A, \quad (4.3)$$

$$\sum_S w^A_S y_{A,S} \leq B \quad \forall A \qquad (4.4)$$

$$x_S, y_{A,S}, z_{A,S}, r_A \geq 0 \quad \forall A, S. \quad (4.5)$$

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 re-

spectively 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, (4.4) ensures that the cost of the y_A decisions does not exceed the budget B, and (4.1) 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 (RSCP), which is at most the optimum value of fractional RSC. We show (Theorem 4.3) that for any $\epsilon, \kappa > 0$, one can efficiently compute a first-stage solution x and solutions (y_A, z_A, r_A) in every scenario A satisfying (4.2)–(4.5) such that $w^{\mathrm{I}} \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq (1 + 2\epsilon)OPT$, and $\sum_A p_A r_A \leq \rho(1 + \kappa)$. Complementing this, we give a simple rounding procedure (Theorem 4.1) based on the rounding theorem in [38] to convert a fractional solution to (RSCP) 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 state our rounding theorem first, in order to better motivate our goal of solving (RSCP).

THEOREM 4.1. (ROUNDING) *Let $(x, \{(y_A, z_A, r_A)\})$ be a solution satisfying (4.2)–(4.5) 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 $w^{\mathrm{I}} \cdot \hat{x} + \sum_A p_A f_A(\hat{x}) \leq (1 + \frac{1}{\varepsilon})C$ 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, we need only x to compute $\hat{x}$ and $\tilde{x}$, and can compute $\tilde{y}_A$ given only $\tilde{x}$ (or $\hat{x}$).

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 $w^{\mathrm{I}} \cdot x + \sum_A p_A f_A(\hat{x}) \leq (1 + \frac{1}{\varepsilon})C$ and $\Pr_A[f_A(\hat{x}) > (1 + \frac{1}{\varepsilon})B] \leq \sum_{A:r_A \geq \frac{1}{1+\varepsilon}} p_A \leq (1 + \varepsilon)P$.

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). ∎

Simple examples show that the $(1 + \frac{1}{\varepsilon}, 1 + \varepsilon)$-tradeoff above is unavoidable (i.e., given our LP-relaxation (RSCP)); see Appendix A. Notice that a blow-up in

the (cost and) budget is unavoidable, since the recourse problem for a scenario is an *NP*-hard problem. So the main question is whether one can obtain a tradeoff-free guarantee to the (fractional or integer) risk-averse problem in polytime, where the probability threshold violated by an arbitrarily small $(1 + \varepsilon)$-factor, and the cost and budget are inflated by a bounded (small) factor that is *independent* of ε. We call $(x, \{y_A\})$ a (c_1, c_2, c_3)-solution if its cost is at most $c_1(\text{optimum})$ and $\Pr_A[f_A(x, y_A) > c_2 B] \leq c_3 \rho$. A (c_1, c_2)-scheme is an algorithm that for any $\varepsilon > 0$, returns a $(c_1, c_2, 1+\varepsilon)$-solution in time $\mathsf{poly}(\mathcal{I}, \frac{1}{\varepsilon})$. Theorem 4.2 (proved in Appendix A) shows that, even for fractional RSC in the polynomial-scenario setting, a tradeoff-free guarantee such as a (c_1, c_2)-scheme, would yield guarantees for the densest k-subgraph (DkS) problem and its minimization version MinDkS; this strengthens a result of [17].

THEOREM 4.2. *A (c_1, c_2)-scheme for integer/fractional RSC (even in the polynomial-scenario setting) yields a c_1-approximation algorithm for MinDkS, and hence, a $2c_1^2$-approximation algorithm for DkS.*

Solving the risk-averse problem (RSCP). We now describe and analyze the procedure used to solve (RSCP). First, we get around the difficulty posed by the coupling constraint (4.1) in formulation (RSCP) by taking the Lagrangian dual of (4.1) introducing a dual variable Δ. This yields the following formulation, whose optimal value is also equal to OPT (this is easy to show using LP-duality).

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

where $g_A(\Delta; x) := \min \sum_S w_S^A (y_{A,S} + z_{A,S}) + \Delta r_A$

s.t. (4.2) − (4.4), $\quad y_{A,S}, z_{A,S}, r_A \geq 0 \quad$ for all S. (P)

Let $OPT(\Delta) = \min_{x \in \mathcal{P}} h(\Delta; x)$. So $OPT = \max_{\Delta \geq 0}(OPT(\Delta) - \Delta\rho)$. Let $\lambda = \max\{1, \max_{A,S}(w_S^A/w_S^{\mathrm{I}})\}$, which we assume is known. The main result of this section is as follows.

THEOREM 4.3. *For any $\epsilon, \gamma, \kappa > 0$, RiskAlg (see Fig. 1) runs in time $\mathsf{poly}(\mathcal{I}, \frac{\lambda}{\epsilon\kappa\rho}, \log(\frac{1}{\gamma}))$, and returns with high probability a first-stage solution x, and a solution (y_A, z_A, r_A) in each scenario A, such that $(x, \{(y_A, z_A, r_A)\})$ satisfy (4.2)–(4.5) and (i) $w^{\mathrm{I}} \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq (1 + \epsilon)OPT + \gamma$; and (ii) $\sum_A p_A r_A \leq \rho(1 + \kappa)$. Under the very mild assumption (∗) that $w^{\mathrm{I}} \cdot x + f_A(x) \geq 1$ for every $A \neq \emptyset$, $x \in \mathcal{P}$, we can convert this guarantee into a $(1 + 2\epsilon)$-multiplicative guarantee in the cost in time $\mathsf{poly}(\mathcal{I}, \frac{\lambda}{\epsilon\kappa})$.*

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

R1. Fix $\varepsilon = \epsilon/6$, $\zeta = \gamma/4$, $\eta = \rho\kappa/16$, $\sigma = \epsilon/6$. Consider the Δ values $\Delta_0, \Delta_1, \ldots, \Delta_k$, where $\Delta_0 = \gamma/4$, $\Delta_{i+1} = \Delta_i(1+\sigma)$ and k is the smallest value such that $\Delta_0(1+\sigma)^k \geq \mathsf{UB}$. Note that $k = O\big(\log(\frac{\mathsf{UB}}{\gamma})/\sigma\big)$.

R2. For each Δ_i, construct the SAA problem $\min_{x \in \mathcal{P}} \big(\widehat{h}(\Delta; x) := w^{\mathrm{I}} \cdot x + \sum_A \widehat{p}_A g_A(\Delta; x)\big)$ using $\mathsf{poly}\big(\mathcal{I}, \frac{\lambda}{\varepsilon\eta}, \log(\frac{\Delta_i}{\zeta})\big)$ samples (where $\widehat{p}_A$ is the frequency of scenario A in the sampled set), and compute its optimal solution $\big(x^{(i)}, \{(y_A^{(i)}, z_A^{(i)}, r_A^{(i)})\}\big)$ (where $(y_A^{(i)}, z_A^{(i)}, r_A^{(i)})$ is implicitly given). By sampling $n = \frac{1}{2(\kappa/8)^2\rho^2} \ln\big(\frac{4k}{\delta}\big)$ scenarios, for each $i = 0, \ldots, k$, compute an estimate $p'^{(i)} = \sum_A \widehat{q}_A r_A^{(i)}$ of $\sum_A p_A r_A^{(i)}$, where $\widehat{q}_A$ is the frequency of scenario A in the sampled set. Let $\rho' = \rho(1 + 3\kappa/4)$.

R3. If $p'^{(0)} \leq \rho'$ then return $x^{(0)}$ as the first-stage solution. [In scenario A, return $(y_A, z_A, r_A) = (y_A^{(0)}, z_A^{(0)}, r_A^{(0)})$].

R4. 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 $(y_A, z_A, r_A) = a(y_A^{(i)}, z_A^{(i)}, r_A^{(i)}) + (1-a)(y_A^{(i+1)}, z_A^{(i+1)}, r_A^{(i+1)})$.]

Figure 1: Procedure RiskAlg.

We show in Section 7 that the dependence on $\frac{1}{\kappa\rho}$ is unavoidable in the black-box model. The "greedy algorithm" for deterministic set cover [10] is an LP-based $\ln n$-approximation algorithm. So using Theorems 4.1 and 4.3, for any $\epsilon, \kappa, \varepsilon > 0$, we can efficiently compute a $(c, c, 1 + \kappa + \varepsilon)$-solution, where $c = 2\ln n\big(1 + \epsilon + \frac{1}{\varepsilon}\big)$.

Algorithm RiskAlg is essentially a search procedure for the "right" value of Δ, wrapped around the SAA method, which is used in step (R2) to compute a "near-optimal" solution to $\min_{x \in \mathcal{P}} h(\Delta; x)$ for any given $\Delta \geq 0$. As mentioned in Section 2, a key ingredient of the algorithm and analysis is to figure out a suitable notion of near-optimality to apply to the 2-stage problem $\min_{x \in \mathcal{P}} h(\Delta; x)$. In particular, as discussed below, the results in [38, 46, 6] for structured 2-stage programs do not quite apply here, and obtaining an FPTAS for the problem is too strong a guarantee to aim for.

The proofs in [38, 46] require that one be able to compute an (ω, ξ)-subgradient of the objective function $h(\Delta; .)$ at any given point x for *sufficiently small* ω, ξ (see Lemma 3.1). This can be done for their class of 2-stage programs since each component of the subgradient lies in an interval of the form $[-w\lambda, w]$ and can hence be estimated up to an additive error of ωw using $\mathsf{poly}\big(\frac{\lambda}{\omega}\big)$ samples, which then yields an $(\omega, 0)$-subgradient. However, $\min_{x \in \mathcal{P}} h(\Delta; x)$ does not fall into the class of prob-

lems considered in [38, 46]: for a subgradient $d = (d_S)$ of $h(\Delta; .)$, we can only say that $d_S \in [-w_S^A - \Delta, w_S^{\mathrm{I}}]$, which prevents us from obtaining an (ω, ξ)-subgradient using polynomial sample size for a suitably small (ω, ξ). (In particular, estimating d_S within an additive error of ωw_S^{I} to obtain an (ω, ξ)-subgradient would require $\mathsf{poly}\big(\frac{\lambda}{\epsilon} + \frac{\Delta}{\omega w_S^{\mathrm{I}}}\big)$ samples, and Δ/w_S^{I} need not be polynomially bounded.) The proof in [6] shows that if Λ is such that $g_A(\Delta; x) - g_A(\Delta; \mathbf{0}) \leq \Lambda w^{\mathrm{I}} \cdot x$ for every A and $x \in \mathcal{P}$, then $\mathsf{poly}\big(\mathcal{I}, \frac{\Lambda}{\epsilon}\big)$ samples suffice to construct a suitable SAA problem. But for our problem, we can only obtain the bound $g_A(\Delta; x) - g_A(\Delta; \mathbf{0}) \leq \lambda w^{\mathrm{I}} \cdot x + \Delta$, and Δ might be large compared to $w^{\mathrm{I}} \cdot x$.

Lemma 4.1 states the precise (weak) approximation guarantee satisfied by the solution returned in step (R2). The key insight underlying its proof (as elaborated in the analysis) is that since we allow for an *additive error measured relative to* Δ, it suffices to approximate each component d_S of the subgradient within an additive error proportional to $(w_S^{\mathrm{I}} + \Delta)$, and this requires only $\mathsf{poly}(\lambda)$ samples.

Given Lemma 4.1, 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 $\big(x, \{(y_A, z_A, r_A)\}\big)$ returned for Δ is such that $\sum_A p_A r_A$ is "close" to ρ. However, this search procedure is complicated by the fact that we have two sources of error whose magnitudes we need to control: first, we only have an approximate solution $\big(x, \{(y_A, z_A, r_A)\}\big)$ 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 Lemma 4.1, so we cannot actually compute or use $\sum_A p_A r_A$ in our search, but will need to estimate it via sampling. We set $\mathsf{UB} = 16(\sum_S w_S^{\mathrm{I}})/\rho$, so $\log \mathsf{UB}$ is polynomially bounded.

Analysis. For the rest of this section, ϵ, γ, κ are fixed values given by Theorem 4.3. We may assume $\epsilon \leq \kappa < 1$.

LEMMA 4.1. *Using* $\mathsf{poly}\big(\mathcal{I}, \frac{\lambda}{\varepsilon\eta}, \log(\frac{\Delta_i}{\zeta})\big)$ *samples one can construct an SAA problem in step (R2) of RiskAlg, so that with high probability, $x^{(i)}$ satisfies $h(\Delta_i; x^{(i)}) \leq (1 + \varepsilon)OPT(\Delta_i) + \eta\Delta_i + \zeta$.*

We defer the proof of Lemma 4.1 to the end of the analysis. Let $\gamma' = \gamma/4$, $\beta = \kappa/8$. Define $p^{(i)} = \sum_A p_A r_A^{(i)}$ and $cost^{(i)} = h(\Delta_i; x^{(i)}) = w^{\mathrm{I}} \cdot x^{(i)} + \sum_A p_A g_A(\Delta_i; x^{(i)})$. By Lemma 3.2, $\Pr[\forall i, |p'^{(i)} - p^{(i)}| \leq \beta\rho] \geq 1 - \delta$. Given Lemma 4.1, we assume that the high probability event "$\forall i$, $cost^{(i)} \leq (1+\varepsilon)OPT(\Delta_i) + \eta\Delta_i + \zeta$ and $|p'^{(i)} - p^{(i)}| \leq \beta\rho$" happens.

CLAIM 4.1. *We have $p^{(k)} < \rho/2$ and $p'^{(k)} < \rho/2$.*

Proof of Theorem 4.3. 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 (4.2)–(4.5) 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 $w^I \cdot x^{(0)} + \sum_A p_A w^A \cdot (y_A^{(0)} + z_A^{(0)})$

$$\leq h(\gamma'; x) \leq (1 + \varepsilon)OPT(\gamma') + \eta\gamma' + \zeta$$
$$\leq (1 + \varepsilon)OPT + \gamma'(1 + \varepsilon + \eta) + \zeta.$$

The last inequality is 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 4.1), 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 $w^I \cdot x + \sum_A p_A w^A \cdot (y_A + z_A) \leq a \cdot cost^{(i)} + (1 - a) \cdot cost^{(i+1)} - \Delta_i(a \cdot p^{(i)} + (1-a) \cdot p^{(i+1)})$, which is at most

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

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, $\rho' - \beta\rho - \eta(1 + \sigma) \geq (1 + \varepsilon + 2\sigma)\rho$. So the above quantity is at most $(1 + \varepsilon + 2\sigma)(OPT(\Delta_i) - \Delta_i\rho) + \zeta \leq (1 + \epsilon)OPT + \gamma$.

The running time is at most $(k + 1) \cdot \text{poly}\big(\mathcal{I}, \frac{\lambda}{\varepsilon\eta}, \log(\frac{\Delta_k}{\zeta})\big) + O\big(\frac{\ln k}{\beta^2\rho^2}\big)$, which is $\text{poly}\big(\mathcal{I}, \frac{\lambda}{\epsilon\kappa}, \log(\frac{1}{\gamma})\big)$ (plugging in $\varepsilon, \eta, \zeta, k$). $\blacksquare$

Proof of 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 $\big(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\}\big)$ is an optimal solution. It is clear that the solution is feasible since $\sum_A p_A \bar{r}_A = q \leq \rho$. To prove optimality, suppose $\big(x^*, \{(y_A^*, z_A^*, r_A^*)\}\big)$ 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 $\sum_{A:A \neq \emptyset} p_A w^A \cdot x^* - w^I \cdot x^*$, which is nonpositive since $w^A \leq \lambda w^I$ and $q \leq 1/\lambda$. It follows that setting $z_A = \hat{z}_A$ for a non-null scenario also gives 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 $OPT \geq \mathsf{LB} = \frac{\delta}{\ln(1/\delta)} \cdot \alpha$; in this case, we return the solution RiskAlg$(\epsilon, \epsilon\mathsf{LB}, \kappa)$ to obtain the desired guarantee. Otherwise, if $X = 0$, we return $\big(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\}\big)$ 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 $OPT \geq \mathsf{LB}$, which is true since $OPT \geq q \geq \alpha$. If $q \leq \delta/M$, then $\Pr[X = 0] \geq 1 - \delta$ and we return $\big(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\}\big)$ 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 $OPT \geq q > \mathsf{LB}$, and that $\big(\mathbf{0}, \{(\bar{y}_A, \bar{z}_A, \bar{r}_A)\}\big)$ is an optimal solution. $\blacksquare$

Proof of Lemma 4.1. Throughout, ε, η, ζ are fixed values given by Lemma 4.1. Fix $\Delta \geq 0$, and let (BSC-P) denote the problem $\min_{x \in \mathcal{P}} h(\Delta; x)$. The proof proceeds by analyzing the subgradients of $h(\Delta; .)$ and $\hat{h}(\Delta; .)$ and showing that they are component-wise close, and thereby arguing that Lemma 3.1 can be applied here. Let $R = \sqrt{m}$, $V = \frac{1}{2}$, so $\mathcal{P} \subseteq B(\mathbf{0}, R)$ and contains a ball of radius V. Let $\tau = \zeta/6$.

The proof has three parts. First, we obtain an expression for the subgradients of $h(\Delta; .)$ and $\hat{h}(\Delta; .)$ at x and prove the bound on the Lipschitz constant. The subgradient of $h(\Delta; .)$ and $\hat{h}(\Delta; .)$ at x is obtained from the optimal solutions to $g_A(\Delta; x)$ for every scenario A. The dual of $g_A(\Delta; x)$ is given by

$$\max \quad \sum_e (\alpha_{A,e} + \beta_{A,e})\big(1 - \sum_{S:e \in S} x_S\big) - B \cdot \theta_A \qquad \text{(D)}$$

$$\text{s.t.} \quad \sum_{e \in S}(\alpha_{A,e} + \beta_{A,e}) \leq w_S^A(1 + \theta_A) \quad \forall S,$$

$$\sum_{e \in S} \beta_{A,e} \leq w_S^A \quad \forall S, \quad \sum_e \alpha_{A,e} \leq \Delta,$$

$$\alpha_{A,e}, \beta_{A,e} \geq 0 \quad \forall e, \quad \alpha_{A,e} = \beta_{A,e} = 0 \quad \forall e \notin A.$$

Here $\alpha_{A,e}$ and $\beta_{A,e}$ are respectively the dual variables corresponding to the covering constraints (4.2) and (4.3), and θ_A is the dual variable corresponding to (4.4). Let $(\alpha_A^*, \beta_A^*, \theta_A^*)$ be an optimal dual solution to $g_A(\Delta; x)$. As in [38], we then have that the vectors d_x and $\hat{d}_x$ with components $d_{x,S} = w_S^I - \sum_A p_A \sum_{e \in S}(\alpha_{A,e}^* + \beta_{A,e}^*)$ $\hat{d}_{x,S} = w_S^I -$

$\sum_A \widehat{p}_A \sum_{e \in S} (\alpha^*_{A,e} + \beta^*_{A,e})$, are respectively subgradients of $h(\Delta; .)$ and $\widehat{h}(\Delta; .)$ at x. Since $\hat{d}_x$ and d_x both have ℓ_2 norm at most $\lambda \|w^{\mathrm{I}}\| + |\Delta|$, $\widehat{h}(\Delta; .)$ and $h(\Delta; .)$ have Lipschitz constant at most $K = \lambda \|w^{\mathrm{I}}\| + |\Delta|$.

Next, we argue that if d is a subgradient of $h(\Delta; .)$ at some point $x \in \mathcal{P}$, and $\hat{d}$ is a vector such that $|\hat{d}_S - d_S| \leq \omega w^{\mathrm{I}}_S + \xi/2m$ for all S, then $\hat{d}$ is an (ω, ξ)-subgradient of $h(\Delta; .)$ at x. 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 $\sum_{S : d_S \leq \hat{d}_S} (d_S - \hat{d}_S) y_S + \sum_{S : \hat{d}_S > d_S} (\hat{d}_S - d_S) x_S \geq -\omega \sum_S (w^{\mathrm{I}}_S y_S + w^{\mathrm{I}}_S x_S) - \xi \geq -\omega h(\Delta; y) - \omega h(\Delta; x) - \xi$.

Recall from Section 3 that $G_\tau \subseteq \mathcal{P}$ is an extended $\frac{\tau}{KN}$-net of $\mathcal{P}$, where $N = \log\left(\frac{2KR}{\tau}\right)$. We use $\mathcal{N} = 8N^2 \left(\frac{4\lambda}{\varepsilon} + \frac{m}{\eta}\right)^2 \ln\left(\frac{2|G_\tau|m}{\delta}\right)$ samples, which is $\mathsf{poly}\left(\mathcal{I}, \frac{\lambda}{\varepsilon \eta}, \log(\frac{\Delta}{\zeta})\right)$. In the sequel, we set $\omega = \varepsilon/8N$, $\xi = \eta \Delta/2N$. Finally, we argue that with probability at least $1 - \delta$, at every point $x \in G_\tau$, the vectors $\hat{d}_x$ and d_x defined above are component-wise close; in particular, they satisfy $|\hat{d}_{x,S} - d_{x,S}| \leq \omega w^{\mathrm{I}}_S + \xi/2m$ for all S and hence, $\hat{d}_x$ is an (ω, ξ)-subgradient of $h(\Delta; .)$ at x. So by Lemma 3.1, if $\hat{x}$ is a minimizer of $\widehat{h}(\Delta; .)$ over $\mathcal{P}$, then $h(\Delta; \hat{x}) \leq (1 + \varepsilon) OPT(\Delta) + \eta \Delta + \zeta$, which completes the proof.

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 $w^{\mathrm{I}}_S - \sum_{e \in S} (\alpha^*_{A,e} + \beta^*_{A,e})$ averaged over the scenarios sampled independently to construct the SAA problem $\widehat{h}(\Delta; .)$, and $\mathrm{E}[\hat{d}_{x,S}] = d_{x,S}$. The sample size $\mathcal{N}$ is specifically chosen so that the Chernoff bound (Lemma 3.2) implies the claim about component-wise closeness with probability at least $1 - \delta$. ∎

4.1 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 $w^{\mathrm{I}} \cdot x + Q_\rho[w^A \cdot y_A]$. Recall that $Q_\rho[w^A \cdot y_A]$ is the smallest B such that $\Pr_A[w^A \cdot y_A > B] \leq \rho$. As mentioned in the Introduction, this problem can be essentially reduced to RSC by simply "guessing" $B = Q_\rho[w^A \cdot y_A]$ for an optimal solution. We briefly describe this reduction here. Let OPT_{Rob} denote the optimum value of the *fractional* risk-averse robust problem $\min_{x \in \mathcal{P}} (w^{\mathrm{I}} \cdot x + Q_\rho[f_A(x)])$. For a given $B \geq 0$, we scale all the second-stage costs w^A_S and B by $\mu = \frac{\gamma}{\lambda \sum_S w^{\mathrm{I}}_S}$. So the contribution from the second-stage cost to the objective function is now at most γ. (Note that the "λ" for the resulting scaled problem is 1, so now the number of samples does not in fact depend on λ.) Let $OPT_{Rob}(B)$ denote the optimum value of the resulting

(RSCP) problem, which is a decreasing function of B. For any guess $B \geq 0$, and any $\epsilon, \gamma, \kappa > 0$, we can use $\mathsf{RiskAlg}$ to compute (nonnegative) $(x, \{y_A, z_A, r_A\})$ in time $\mathsf{poly}\left(\mathcal{I}, \frac{1}{\epsilon \kappa \rho}, \log(\frac{1}{\gamma})\right)$ satisfying (4.2)–(4.4) such that $w^{\mathrm{I}} \cdot x \leq cost\left(x, \{(y_A, z_A, r_A)\}\right) \leq (1 + \epsilon) OPT_{Rob}(B) + \gamma$ and $\sum_A p_A r_A \leq \rho(1 + \kappa)$. Let W be an upper bound on the optimum such that $\log W$ is polynomially bounded, e.g., $W = \sum_S w^{\mathrm{I}}_S$. We enumerate values of B in powers of $(1 + \epsilon)$, starting at γ and ending at the smallest value that is at least W. We use $\mathsf{RiskAlg}$ to compute a solution for each B, and return the one that minimizes $w^{\mathrm{I}} \cdot x + B$. Let $(\bar{x}, \{\bar{y}_A, \bar{z}_A, \bar{r}_A\})$, computed for $\bar{B}$, denote this solution. Let B^* be the "correct" guess. Note that $OPT_{Rob}(B^*) \leq OPT_{Rob} - B^* + \gamma$. We are guaranteed to enumerate $B' \in [B^*, (1 + \epsilon) B^* + \gamma]$. Let $(x', \{y'_A, z'_A, r'_A\})$ be the solution computed for B'. Then we have $w^{\mathrm{I}} \cdot \bar{x} + \bar{B} \leq w^{\mathrm{I}} \cdot x' + B' \leq (1 + \epsilon) OPT_{Rob}(B') + (1 + \epsilon) B^* + 2\gamma \leq (1 + \epsilon) OPT_{Rob} + 4\gamma$. We remark that the same ideas yield a similar guarantee for the LP-relaxation of a generalization of the problem, where we wish to minimize $w^{\mathrm{I}} \cdot x$ plus a weighted combination of $\mathrm{E}_A[w^A \cdot y_A]$ 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 4.3. Let $q = \sum_{A \neq \emptyset} p_A$. Note that if $q \leq \rho$, then $OPT_{Rob} = 0$ and $x = \mathbf{0}$ is an optimal solution; 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 (where ρ' is set in $\mathsf{RiskAlg}$) 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 any feasible solution otherwise. Note that $w^{\mathrm{I}} \cdot x + Q_{\rho'}[w^A \cdot y_A] = 0$. In the latter case, we set $\gamma = \epsilon$, and execute the procedure detailed above to obtain a $(1 + 5\epsilon)$-multiplicative guarantee.

Now one can use Theorem 4.1 as is to convert the obtained fractional solution $(x, \{y_A, z_A, r_A\})$ into an integer solution, or a solution to the fractional risk-averse robust problem. The budget-inflation can now be absorbed into the approximation ratio. For any $\epsilon, \kappa, \varepsilon > 0$, we obtain a fractional solution $\hat{x}$ such that $w^{\mathrm{I}} \cdot \hat{x} + Q_{\rho(1+\kappa+\varepsilon)}[f_A(\hat{x})] \leq \left(1 + \epsilon + \frac{1}{\varepsilon}\right) OPT_{Rob}$, and an integer solution $(\tilde{x}, \{\tilde{y}_A\})$ such that $w^{\mathrm{I}} \cdot \tilde{x} + Q_{\rho(1+\kappa+\varepsilon)}[w^A \cdot \tilde{y}_A] \leq 2c\left(1 + \epsilon + \frac{1}{\varepsilon}\right) OPT_{Rob}$ using an LP-based c-approximation algorithm for deterministic set cover.

5 Applications to combinatorial optimization problems

We now show that the techniques developed in Section 4 for risk-averse budgeted set cover 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—and facility location. This includes many of the problems considered in [19, 38, 11] in the standard 2-stage and demand-robust models. In all the applications, the first step is to prove an analogue of Theorem 4.3, that is, argue that RiskAlg can be used to obtain a near-optimal solution to a suitable LP-relaxation of the problem. (For facility location, we need to adapt 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 4.1, we first obtain 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.

Our results are intended to illustrate that approximation guarantees developed for the deterministic or 2-stage version of the problem can be converted to analogous guarantees for the risk-averse budgeted problem once we have a near-optimal solution to the LP-relaxation of the risk-averse problem, and we have not sought to optimize the approximation factors. Our approximation results also hold for non-uniform budgets, and translate to the risk-averse robust versions of our applications: an algorithm that returns a (c_1, c_2, c_3)-solution $(x, \{y_A\})$ for the budgeted problem can be sued to obtain a solution to the robust problem where $c(x) + Q_{\rho(1+c_3)}[f_A(x, y_A)] \leq \max\{c_1, c_2\} \cdot OPT_{Rob}$. We also achieve guarantees for the problem of minimizing $c(x)$ plus a weighted combination of $\mathrm{E}_A[f_A(x, y_A)]$ and $Q_\rho[f_A(x, y_A)]$.

5.1 Vertex cover and multicut on trees In the stochastic vertex cover problem, we are given a graph whose edges need to covered by vertices. The edge-set is random and determined by a distribution; one needs to pick vertices in stage I and in each scenario so that their union forms a vertex cover for the edges revealed in the scenario. In the stochastic multicut on trees problem, we are given a tree, and a (black-box) distribution over sets of s_i-t_i pairs; a feasible solution needs 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. In the risk-averse budgeted versions of these problems we are given a budget B and threshold ρ, and the goal is to compute a minimum-cost feasible solution such that Pr[second-stage cost > B] $\leq \rho$. Both 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 (RSCP) and by Theorem 4.3, obtain a near-optimal solution to the relaxation. Since there is an LP-based 2-approximation algorithm for the deterministic versions of both problems, applying Theorem 4.1 yields the following guarantees.

THEOREM 5.1. *For any $\epsilon, \kappa, \varepsilon > 0$, one can compute in polynomial time a $\left(4(1+\epsilon+\frac{1}{\varepsilon}), 4(1+\epsilon+\frac{1}{\varepsilon}), 1+\kappa+\varepsilon\right)$-solution for the risk-averse budgeted vertex cover and multicut on trees problems.*

5.2 Facility location In the risk-averse budgeted facility location problem (RFL), we have a set of m facilities $\mathcal{F}$, a client-set $\mathcal{D}$, and a distribution over client-demands. 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 may open facilities in stage I or in a given scenario, and in each scenario A we must assign each client $j \in A$ 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 to a facility i is 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 total facility-opening and client-assignment cost incurred in that scenario. The goal is to minimize the total expected cost subject to the usual condition that Pr[second-stage cost > B] $\leq \rho$. We formulate the following LP-relaxation of the problem. Throughout, i indexes the facilities in $\mathcal{F}$ and j the clients in $\mathcal{D}$.

$$\min \quad \sum_i f_i^\mathrm{I} y_i + \sum_{A \subseteq \mathcal{D}} p_A \Big(\sum_i f_i^A \big(y_{A,i} + v_{A,i} \big)$$
$$+ \sum_{j \in A, i} c_{ij} \big(x_{A,ij} + u_{A,ij} \big) \Big) \qquad \text{(RFLP)}$$

$$\text{s.t.} \qquad \sum_A p_A r_A \leq \rho \qquad (5.6)$$

$$\sum_i x_{A,ij} + r_A \geq 1 \quad \forall j \in A \quad (5.7)$$

$$\sum_i \big(x_{A,ij} + u_{A,ij} \big) \geq 1 \quad \forall j \in A \quad (5.8)$$

$$x_{A,ij} \leq y_i + y_{A,i} \quad \forall j \in A, i \qquad (5.9)$$

$$x_{A,ij} + u_{A,ij} \leq y_i + y_{A,i} + v_{A,i} \quad \forall j \in A, i$$
$$(5.10)$$

$$\sum_i f_i^A y_{A,i} + \sum_{j \in A, i} c_{ij} x_{A,ij} \leq B \quad \forall A \qquad (5.11)$$

$$y_i, y_{A,i}, v_{A,i}, x_{A,ij}, u_{A,ij}, r_A \geq 0 \quad \forall A, i, j. \ (5.12)$$

Here y_i denotes the first-stage decisions. Decisions $(x_{A,ij}, y_{A,i})$ and $(u_{A,ij}, v_{A,i})$ represent respectively the

actions taken in scenario A when does not exceed the budget ($r_A = 0$), and does exceed the budget ($r_A = 1$). Constraints (5.7)–(5.10) enforce that every client is assigned to an open facility (in both cases), and (5.11) is the budget constraint for a scenario. Let OPT be the optimal value of (RFLP). Given first-stage decisions $y \in \mathcal{P} := [0,1]^m$, let $\ell_A(y)$ be the minimum facility-location cost, over fractional solutions, incurred in scenario A to satisfy the clients in A.

THEOREM 5.2. *For any* $\epsilon, \gamma, \kappa > 0$, *in time* $\mathsf{poly}\big(\mathcal{I}, \frac{\lambda}{\epsilon \kappa \rho}, \log(\frac{1}{\gamma})\big)$, *one can compute* $\big(y, \{(x_A, y_A, u_A, v_A, r_A)\}\big)$ *that satisfies* (5.7)–(5.12) *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* $f^{\mathrm{I}} \cdot y + \ell_A(y) \geq 1$ *for every* $y \in [0,1]^m$, $A \neq \emptyset$.

Proof Sketch. As in Section 4, we Lagrangify (5.6) using a dual variable $\Delta \geq 0$ to obtain the problem $\max_{\Delta \geq 0}\big(-\Delta \rho + OPT(\Delta)\big)$ where $OPT(\Delta) = \min_{y \in \mathcal{P}} h(\Delta; y)$, $h(\Delta; y) = f^{\mathrm{I}} \cdot y + \sum_A p_A g_A(\Delta; y)$, 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 (5.7)–(5.12) (where the y_i's are fixed now). We argue briefly that RiskAlg can be used to compute the desired near-optimal solution; given this, the proof of the multiplicative guarantee is as in Theorem 4.3. Proving this involves two things: (a) coming up with a bound UB such that $\log$ UB is polynomially bounded so that one can restrict the search for the right value of Δ in RiskAlg; and (b) showing that for any $\Delta \geq 0$, an optimal solution to the SAA-problem $\min_{y \in \mathcal{P}} \widehat{h}(\Delta; y)$ (constructed in step (R2) of RiskAlg) yields a solution to $\min_{y \in \mathcal{P}} h(\Delta; y)$ that satisfies the approximation guarantee in Lemma 4.1.

There are two notable aspects in which 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 (RFLP) 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 (5.6)) to prove the closeness-in-subgradients property for the SAA objective function constructed in step (R2) and the true objective function.

Assume first that we have shown (b). Define $C_A = \sum_{j \in A}(\min_i c_{ij})$ and $C = \sum_j (\min_i c_{ij})$. Note that C_A is the minimum possible assignment cost that one can incur in scenario A. We may determine with high

probability using $O\big(\frac{1}{\rho \kappa}\big)$ samples if $\Pr_A[C_A > B] > \rho$ or $\Pr_A[C_A > B] \leq \rho\big(1 + \frac{5\kappa}{28}\big)$. In the former case, we can conclude that the problem is infeasible. In the latter case, we set $\widehat{\rho} = \rho\big(1 + \frac{5\kappa}{28}\big)$ and $\widehat{\kappa}$ such that $\widehat{\rho}(1 + \widehat{\kappa}) = \rho(1 + \kappa)$, and call procedure RiskAlg with these values of $\widehat{\rho}$ and $\widehat{\kappa}$ (and the given ϵ, γ), taking $\mathsf{UB} = \frac{32(1+\varepsilon)(\sum_i f_i^{\mathrm{I}} + C)}{3 \rho \kappa}$. It is not hard to see that with this upper bound, we have $p^{(k)}$, $p'^{(k)} < \rho' = \widehat{\rho}(1 + 3\widehat{\kappa}/4)$, and (as in the proof of Theorem 4.3) this suffices for the search for Δ in RiskAlg to go through.

Task (b) boils down to showing that the objective function $\widehat{h}(\Delta; .)$ of the SAA-problem (in step (R2)) and the true problem $h(\Delta; .)$ satisfy the conditions of Lemma 3.1. Again, with $R = \sqrt{m}$ and $V = \frac{1}{2}$, we have that $\mathcal{P} \subseteq B(\mathbf{0}, R)$ and contains a ball of radius V. Lemma 5.1 proves that this holds with high probability, with $K = \lambda \|f^{\mathrm{I}}\| + |\Delta|$, $\varrho = \varepsilon$, $\xi = \eta \Delta$ and $\tau = \zeta/6$ (and $N = \log\big(\frac{2KR}{\tau}\big)$ as in Section 3). 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. So by Lemma 3.1, the solution $\widehat{y}^{(i)} = \operatorname{argmin}_{y \in \mathcal{P}} \widehat{h}(\Delta_i; y)$ obtained in step (R2) for each Δ_i satisfies the requirements of Lemma 4.1. ∎

LEMMA 5.1. *With probability at least* $1 - \delta$, $\widehat{h}(\Delta; .)$ *and* $h(\Delta; .)$ *satisfy the conditions of Lemma 3.1 with* $K = \lambda \|f^{\mathrm{I}}\| + |\Delta|$, $\varrho = \varepsilon$ *and* $\xi = \eta \Delta$, *and* $\tau = \zeta/6$ *(and* $N = \log\big(\frac{2KR}{\tau}\big)$*)*.

Proof. The proof dovetails the proof of Lemma 4.1. We use $\mathcal{N} = 8N^2 \big(\frac{4\lambda}{\varepsilon} + \frac{m}{\eta}\big)^2 \ln\big(\frac{2|G_\tau|m}{\delta}\big)$ samples (where $G_\tau \subseteq \mathcal{P}$ is an extended $\frac{\tau}{KN}$-net of $\mathcal{P}$), which is $\mathsf{poly}\big(\mathcal{I}, \frac{\lambda}{\varepsilon \eta}, \log(\frac{\Delta}{\zeta})\big)$. Consider any $y \in \mathcal{P}$. Let $\alpha^*_{A,j}, \psi^*_{A,j}, \beta^*_{A,ij}, \Gamma^*_{A,ij}, \theta^*_A$ be the values of the dual variables corresponding to (5.7)–(5.11) respectively in an optimal dual solution to $g_A(\Delta; y)$. We choose an optimal dual solution that minimizes $\sum_{i,j} \beta^*_{A,ij}$. It is easy to show that the vectors $\widehat{d}_y = (\widehat{d}_{y,i})$ and $d_y = (d_{y,i})$ given by $\widehat{d}_{y,i} = f_i^{\mathrm{I}} - \sum_A \widehat{p}_A \sum_{j \in A}\big(\beta^*_{A,ij} + \Gamma^*_{A,ij}\big)$ and $d_{y,i} = f_i^{\mathrm{I}} - \sum_A p_A \sum_{j \in A}\big(\beta^*_{A,ij} + \Gamma^*_{A,ij}\big)$ are respectively subgradients of $\widehat{h}(\Delta; .)$ and $h(\Delta; .)$ 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, $\|\widehat{d}_y\|, \|d_y\| \leq K$ where $K = \lambda \|f^{\mathrm{I}}\| + \Delta$ for any $y \in \mathcal{P}$, so K is an upper bound on the Lipschitz constant of $h(\Delta; .)$ and $\widehat{h}(\Delta; .)$.

The second inequality is a constraint of the dual. Suppose $\beta^*_{A,ij} > 0$ for some j. The dual enforces the constraint $\alpha^*_{A,j} + \psi^*_{A,j} \leq c_{ij}(1 + \theta^*_A) + \beta^*_{A,ij} + \Gamma^*_{A,ij}$. 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 $\psi_{A,j}^* \le c_{ij} + \Gamma_{A,ij}^*$, we have that $\beta_{A,ij}^* \le \alpha_{A,j}^*$, so $\sum_j \beta_{A,ij}^* \le \sum_j \alpha_{A,j}^* \le \Delta$ (the last inequality follows from the dual constraint corresponding to r_A).

As in Lemma 4.1, if d is a subgradient of $h(\Delta; .)$ at y and $\hat{d}$ is a vector such that $|\hat{d}_i - d_i| \le \omega f_i^{\mathrm{I}} + \frac{\xi}{2m}$, then $\hat{d}$ is an (ω, ξ)-subgradient of $h(\Delta; .)$ at y.

Since $\mathrm{E}[\hat{d}_{y,i}] = d_{y,i}$ for every y and i, plugging in the sample size $\mathcal{N}$ and using the Chernoff bound (Lemma 3.2), we obtain with probability at least $1 - \delta$, $|\hat{d}_{y,i} - d_{y,i}| \le \frac{\varepsilon}{8N} f_i^{\mathrm{I}} + \frac{\eta\Delta}{4mN}$ for all i, for every point y in the extended $\frac{\tau}{KN}$-net G_τ of $\mathcal{P}$. Thus, with probability at least $1 - \delta$, $\hat{d}_y$ is an $\left(\frac{\varepsilon}{8N}, \frac{\eta\Delta}{2N}\right)$-subgradient of $h(\Delta; .)$ at y for every $y \in G_\tau$. ■

To round the LP-solution, as in part (i) of Theorem 4.1, we observe that if $\left(y, \{(x_A, y_A, u_A, v_A, r_A)\}\right)$ is a solution satisfying (5.7)–(5.12) of objective value C, then for any $\varepsilon > 0$, taking $\hat{y} = y\left(1 + \frac{1}{\varepsilon}\right)$ gives $\sum_i f_i \hat{y}_i + \sum_A p_A \ell_A(\hat{y}) \le \left(1 + \frac{1}{\varepsilon}\right)C$ and $\Pr_A[\ell_A(\hat{y}) > \left(1 + \frac{1}{\varepsilon}\right)B] \le (1 + \varepsilon)\sum_A p_A r_A$. 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 $\min\{2c, c + 1.5\}$-approximation algorithm for SUFL, by using it in conjunction with the 1.5-approximation algorithm for DUFL in [4]. "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. Second, 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. Using such a randomized local γ-approximation algorithm for SUFL, either the one in [38] or its improvement in [42], would yield a random solution such that $\Pr_A[expected\ cost\ of\ scenario\ A > \gamma B] \le \rho(1 + \kappa + \varepsilon)$, where the expectation is over the random choices of the algorithm. But we want to make the stronger claim that, *with high probability over the*

random choices of the algorithm, we return a solution where $\Pr_A[\text{cost of } A > \gamma B] \le \rho(1 + \kappa + \varepsilon)$. We take care of both these 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]): with probability 1, the algorithm should return a solution where each client's assignment cost is within some factor of its cost in the fractional solution. One can use the the deterministic Shmoys-Tardos-Aardal (STA) algorithm [40], or the randomized approximation algorithm of Swamy [43], both of which satisfy this condition (and are demand-oblivious). In particular, the STA-algorithm [40] returns a 4-approximate solution, where a client's assignment cost is blown up by a factor of at most 4. Combining this and the algorithm of [4] in the rounding procedure of [38] yields the following theorem.

THEOREM 5.3. *For any $\epsilon, \kappa, \varepsilon > 0$, one can compute a $\left(5.5(1 + \epsilon + \frac{1}{\varepsilon}), 5.5(1 + \epsilon + \frac{1}{\varepsilon}), 1 + \kappa + \varepsilon\right)$-solution to risk-averse budgeted facility location in polynomial time.*

Proof Sketch. Let $\left(y, \{(x_A, y_A, u_A, v_A)\}\right)$ be the solution given by Theorem 5.2. Let $\hat{y} = y\left(1 + \frac{1}{\varepsilon}\right)$, so that $\sum_i f_i \hat{y}_i + \sum_A p_A \ell_A(\hat{y}) \le \left(1 + \frac{1}{\varepsilon}\right)C$ and $\Pr_A[\ell_A(\hat{y}) > \left(1 + \frac{1}{\varepsilon}\right)B] \le (1 + \varepsilon)\sum_A p_A r_A$. Suppose we have a demand-oblivious LP-based α-approximation algorithm such that with probability 1, the algorithm returns an integer solution where each client's assignment cost is at α times its cost in the fractional solution. We utilize the rounding procedure in [38], which we sketch below for completeness, and also to demonstrate how the demand-obliviousness and "distance-preservation" properties allow one to (1) obtain a local approximation algorithm for SUFL), and (2) obtain the recourse action for scenario A given only $\hat{y}$ and the rounded first-stage solution.

For the first-stage decisions, we round $\min\{1, \hat{y}_i/\theta\}$, where $\theta = \frac{\alpha}{\alpha + 1.5}$, using the α-approximation algorithm to obtain the integer vector $\tilde{y}$, which gives the set of facilities opened in stage I. Let $a(j)$ denote the open facility that is nearest to j. Let $\bar{C}_j$ denote the minimum cost of assigning j fractionally to an extent of 1 to the facility-opening vector $\left(\min\{\hat{y}_i/\theta, 1\}\right)_i$. By demand-obliviousness, for any client-demands $(d_j)_{j \in \mathcal{D}}$, we have $f_i^{\mathrm{I}} \cdot \tilde{y} + \sum_j d_j c_{a(j)j} \le \alpha\left(f^{\mathrm{I}} \cdot \frac{\hat{y}}{\theta} + \sum_j d_j \bar{C}_j\right)$; and by distance preservation, we have $c_{a(j)j} \le \alpha \bar{C}_j$ for all j.

In a scenario A, we first compute the solution $(\hat{x}_A, \hat{y}_A)$ that determines $\ell_A(\hat{y})$ (which can be done efficiently). For every client $j \in A$, one can write $\hat{x}_{A,ij} = \hat{x}_{A,ij}^{\mathrm{I}} + \hat{x}_{A,ij}^{\mathrm{II}}$ such that $\hat{x}_{A,ij}^{\mathrm{I}} \le \hat{y}_i$ and $\hat{x}_{A,ij}^{\mathrm{II}} \le \hat{y}_{A,i}$. Let $\mathcal{D}_A = \{j \in A : \sum_i \hat{x}_{A,ij}^{\mathrm{I}} \ge \theta\}$. We assign each $j \in \mathcal{D}_A$ to $a(j)$; note that $c_{a(j)j} \le \frac{\alpha}{\theta} \cdot \sum_i c_{ij} \hat{x}_{A,ij}^{\mathrm{I}}$. Next, we run the LP-based 1.5-approximation algorithm for

DUFL on the instance with client-set $A \setminus \mathcal{D}_A$. This determines the facilities to open in scenario A $(\tilde{y}_A)$ and the assignment $(\tilde{x}_{A,ij})_i$ of clients j in $A \setminus \mathcal{D}_A$. We have $f^A \cdot \tilde{y} + \sum_{j \in A \setminus \mathcal{D}_A} \sum_i c_{ij} \tilde{x}, ij \leq \frac{1.5}{1-\theta} \cdot (f^A \cdot \hat{y}_A + \sum_{j \in A \setminus \mathcal{D}_A} \sum_i c_{ij} \hat{x}_{A,ij}^{\mathrm{II}})$, Also, note that $f^{\mathrm{I}} \cdot \tilde{y} + \sum_{A, j \in \mathcal{D}_A} p_A c_{a(j)j} \leq \frac{\alpha}{\theta} \cdot (f^{\mathrm{I}} \cdot \hat{y} + \sum_{A, j \in \mathcal{D}_{A,i}} p_A c_{ij} \hat{x}_{A,ij}^{\mathrm{I}})$. So we obtain a solution of cost at most $(\alpha + 1.5)C$ and the cost of each scenario A is at most $(\alpha + 1.5)\ell_A(\hat{y})$. Thus, with $\alpha = 4$, we obtain a $(5.5(1 + \epsilon + \frac{1}{\varepsilon}), 5.5(1 + \epsilon + \frac{1}{\varepsilon}), 1 + \kappa + \varepsilon)$-solution. $\blacksquare$

Budget constraints on individual components of the second-stage cost. As mentioned in Section 2, our techniques can be used to devise approximation algorithms for a fairly general risk-averse version of facility location (and other problems), where we impose the *joint probabilistic budget constraint* $\Pr_A[$ (total cost of A)$> B$, or (facility-cost of A)$> B_F$, or (assignment-cost of A)$> B_C] \leq \rho$. This has the effect of augmenting (RFLP) with the constraints $\sum_i f_i^A y_{A,i} \leq B_F$, and $\sum_{j,i} c_{ij} x_{A,ij} \leq B_C$, for each scenario A. (Note that by setting a 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, setting $B_F = 0$, $B = \infty$ means that we seek a minimum-cost solution where the facilities opened in stage I are such that with probability at least $1 - \rho$, we can assign the clients in a scenario A to the stage I facilities while incurring assignment cost at most B_C. Algorithm RiskAlg can be applied to solve even this more general LP, and Theorem 5.2 continues to hold here. Note that to describe a solution, even for the fractional risk-averse problem, an algorithm must now return not just the first-stage solution y but also specify how to compute the recourse action in a scenario, and RiskAlg *does indeed* do this.

Rounding procedure. The rounding procedure is similar to that in Theorem 5.3 and we highlight the main changes here. Let $(y, \{(x_A, y_A, u_A, v_A)\})$ be the solution given by RiskAlg for the general risk-averse problem. Note that in each scenario A, we can compute $(x_A, y_A, u_A, v_A, r_A)$ efficiently. Say that a scenario A is c-violated if at least one of its budget constraints is violated by more than a c-factor. We assume initially that we have a rounding algorithm for DUFL that given a fractional solution, returns an integer solution whose facility-opening, client-assignment, and total- cost are at most β times the corresponding quantity in the fractional solution. We prove later that any LP-based algorithm for DUFL can be morphed into such an algorithm.

Let $\theta = \frac{\alpha}{\alpha + \beta}$. As before, we set $\hat{y} = y(1 + \frac{1}{\varepsilon})$ and use the demand-oblivious distance-preserving LP-based α approximation algorithm to round $(\min\{\hat{y}_i/\theta, 1\})_i$ and

obtain an integer vector $\tilde{y}$ specifying which facilities to open in stage I. Let $a(j)$ be the open facility nearest to j.

In a scenario A, we first obtain $(x_A, y_A, u_A, v_A, r_A)$. We then extract a fractional solution $(\hat{x}_A, \hat{y}_A)$ for scenario A that is feasible given the first-stage decision $\hat{y}$. If $r_A \geq \frac{1}{1+\varepsilon}$ (indicating that we may "violate" scenario A), we set $\hat{x}_A = x_A + u_A$ and $\hat{y}_A = y_A + v_A$. Otherwise, we set $\hat{x}_A = (1 + \frac{1}{\varepsilon})x_A$, $\hat{y}_A = (1 + \frac{1}{\varepsilon})y_A$; note that in this case we have $f^A \cdot \hat{y}_A \leq (1 + \frac{1}{\varepsilon})B_F$, $\sum_{j \in A,i} c_{ij} \hat{x}_{A,ij} \leq (1 + \frac{1}{\varepsilon})B_C$, and $f^A \cdot \hat{y}_A + \sum_{j \in A,i} c_{ij} \hat{x}_{A,ij} \leq (1 + \frac{1}{\varepsilon})B$. We now round $(\hat{x}_A, \hat{y}_A)$ using the rounding procedure of [38] using the above value of θ, and the β-approximation algorithm to determine the solution for scenario A. That is, we split $\hat{x}_{A,ij}$ as $\hat{x}_{A,ij}^{\mathrm{I}} + \hat{x}_{A,ij}^{\mathrm{II}}$ where $\hat{x}_{A,ij}^{\mathrm{I}} \leq \hat{y}_i$ and $\hat{x}_{A,ij}^{\mathrm{II}} \leq \hat{y}_{A,i}$. Clients in $\mathcal{D}_A = \{j \in A : \sum_i \hat{x}_{A,ij}^{\mathrm{I}} \geq \theta\}$ are assigned to their nearest stage-I facilities, and we use the β-approximation algorithm to round $\frac{1}{1-\theta}((\hat{x}_{A,\bullet j}^{\mathrm{II}})_{j \in A \setminus \mathcal{D}_A}, \hat{y}_A)$ to obtain the facilities to open in scenario A $(\tilde{y}_A)$ and the assignment $(\tilde{x}_{A,ij})_i$ of clients $j \in A \setminus \mathcal{D}_A$. The properties of the α- and β-approximation algorithms yield the following bounds.

$$f^{\mathrm{I}} \cdot \tilde{y} + \sum_{A, j \in \mathcal{D}_A} p_A c_{a(j)j} \leq \frac{\alpha}{\theta}\Big(f^{\mathrm{I}} \cdot \hat{y} + \sum_{A, j \in \mathcal{D}_{A,i}} p_A c_{ij} \hat{x}_{A,ij}^{\mathrm{I}}\Big)$$

$$c_{a(j)j} \leq \frac{\alpha}{\theta}\Big(\sum_i c_{ij} \hat{x}_{A,ij}^{\mathrm{I}}\Big) \qquad \forall A, j \in \mathcal{D}_A$$

$$f^A \cdot \tilde{y}_A \leq \frac{\beta}{1-\theta}(f^A \cdot \hat{y}_A) \qquad \forall A$$

$$\sum_{j \in A \setminus \mathcal{D}_A,i} c_{ij} \tilde{x}_{A,ij} \leq \frac{\beta}{1-\theta}\Big(\sum_{j \in A \setminus \mathcal{D}_A,i} c_{ij} \hat{x}_{A,ij}^{\mathrm{II}}\Big) \qquad \forall A$$

and $f^A \cdot \tilde{y}_A + \sum_{j \in A \setminus \mathcal{D}_A,i} c_{ij} \tilde{x}_{A,ij} \leq \frac{\beta}{1-\theta}(f^A \cdot \hat{y}_A + \sum_{j \in A \setminus \mathcal{D}_A,i} c_{ij} \hat{x}_{A,ij}^{\mathrm{II}})$ $\forall A$.

Combining these bounds, we see that the total cost of the solution is at most $(\alpha + \beta)(1 + \frac{1}{\varepsilon})C$ and in each scenario A, the facility-opening, client-assignment, and total-cost are all at most $(\alpha + \beta)$ times the corresponding quantity in $(\hat{x}_A, \hat{y}_A)$. Thus, if $r_A < \frac{1}{1+\varepsilon}$ then these quantities are at most $(\alpha + \beta)(1 + \frac{1}{\varepsilon})\{B_F, B_C, B\}$ respectively. So we obtain that $\Pr_A[A \text{ is } (\alpha + \beta)(1 + \epsilon + \frac{1}{\varepsilon})\text{-violated}] \leq \rho(1 + \kappa + \varepsilon)$.

Finally, we note that any LP-based γ-approximation algorithm for DUFL can be used to obtain the desired β-approximation algorithm above with $\beta = 2\gamma$. Suppose (X, Y) is the solution to a DUFL instance with facility-costs $\{f_i\}$ and client-assignment costs $\{D_{ij}\}$. Let $P = \sum_i f_i Y_i$, $Q = \sum_{j,i} c_{ij} X_{ij}$ and $R = P + Q$. We run the γ-approximation algorithm with facility costs $\frac{R}{P}\{f_i\}$ and client-assignment costs $\frac{R}{Q}\{D_{ij}\}$ to obtain an integer solution $(\tilde{X}, \tilde{Y})$. It

follows that $\sum_i f_i \tilde{Y}_i \leq \gamma \cdot 2P$, $\sum_{j,i} c_{ij} \tilde{X}_{ij} \leq \gamma \cdot 2Q$, so $\sum_i f_i \tilde{Y}_i + \sum_{j,i} c_{ij} \tilde{X}_{ij} \leq \gamma \cdot 2R$. So we can obtain $\beta \leq 3$.

Taking $\alpha = 4$ and $\beta = 2 \times 1.5$, we obtain a solution of cost at most $7\left(1 + \epsilon + \frac{1}{\varepsilon}\right) OPT$ where $\Pr_A[A$ is $7\left(1 + \epsilon + \frac{1}{\varepsilon}\right)$-violated$] \leq \rho(1 + \kappa + \varepsilon)$. All of our arguments generalize to the setting with scenario-dependent budgets $\{(B^A, B_F^A, B_C^A)\}$.

6 Refinements

The RSC instances constructed in the proof of Theorem 4.2 to show the difficulty of obtaining a (c_1, c_2)-scheme, can be easily cast as instances of other risk-averse budgeted problems—e.g., *all* the problems in Section 5—where each scenario consists of (at most) two "elements" (e.g., clients in RFL). We show here that if all scenarios contain (at most) one element, then one *can* obtain an approximation that does not violate the probability threshold. (Note that the examples in Appendix A showing that the guarantees of Theorem 4.1 are tight are one-element instances.) Although the one-element-per-scenario setting appears rather restrictive, it has (surprisingly) rich modeling power. We uncover a close connection between the *independent activation* (IA) model, where each "element" is independently "activated" with probability p_j, and the one-element-per-scenario setting. The IA model is a popular model in Computer Science that has been considered in various stochastic contexts [27, 16, 24, 15, 39], and our results suggest that an understanding of risk-averse problems in the one-element-per-scenario setting may yield significant dividends for stochastic problems in the IA model. We focus on RFL for concreteness, but the same reductions apply to other problems.

Let $\mathcal{C}$ denote the class of all one-client-per-scenario RFL instances. (Note that clients may have non-uniform demands). We can show that various stochastic problems in the IA model *reduce to RFL problems over $\mathcal{C}$*. For example, consider *a priori facility location* (APFL)—we have a distribution over client-sets and the goal is to find an assignment of (all) clients to (open) facilities with minimum expected cost—and its risk-averse version, where we impose also the constraint $\Pr^{\text{IA}}[$assignment cost of an activated client $> B] \leq \rho$. *A priori* stochastic problems (see [2]) in the IA model have very recently been considered from the perspective of approximation algorithms in [15, 39]. Despite the contrast between (risk-averse) APFL and RFL restricted to class $\mathcal{C}$—the former is a one-stage problem where we choose the entire solution in advance and pay only for facilities used by the (random) set of activated clients; in the latter problem, we pay for all facilities opened in stage I and can augment our solution in stage II—we can

reduce risk-averse APFL to a RFL problem over class $\mathcal{C}$. Another example is *one-stage FL*: minimize the facility-opening + expected client-assignment cost, subject to $\Pr^{\text{IA}}[$assignment cost of a client $> B] \leq \rho$. ([1] give an approximation algorithm for a set-cover version of this problem: select a min-cost collection of sets so that $\Pr^{\text{IA}}[$element is uncovered$] \leq \rho$. This can be encoded as *non-metric* one-stage FL with $B = 0$, and $c_{ij} = 0$ or $\mu > 0$ (which is small) depending on whether or not i covers j.) In both reductions, we create a RFL instance in class $\mathcal{C}$ where: (a) each client (i.e., scenario) j has non-unit demand and budget B_j; and (b) the notion of risk-aversion is $\Pr_j[$second-stage assignment cost of $\{j\} > B_j] \leq \rho$; A $(c_1, c_2, 1)$-solution to RFL is a solution of cost at most c_1(optimum), where $\Pr_j[\{j\}$ is c_2-violated$] \leq \rho$. (Recall that a scenario $\{j\}$ is c-violated if (at least one of) its budget constraint(s) is violated by more than a c-factor.)

THEOREM 6.1. *Given a polytime algorithm for RFL (resp. non-metric RFL) that always returns a $(c_1, c_2, 1)$-solution, for both one-stage FL and risk-averse APFL (resp. non-metric $\{$one-stage FL, risk-averse APFL $\})$, one can compute an $O(c_1)$-approximate solution in polytime with $\Pr^{\text{IA}}[$assignment cost of an activated client $> c_2 B] \leq \rho$.*

Proof. The reduction from both problems to RFL is quite similar, and we point out the common ingredients first. Let $\left(\{f_i\}, \{c_{ij}\}, \{p_j\}, B, \rho\right)$ be an instance of one-stage FL or risk-averse APFL, where p_j is the activation probability of client j. We assume that the instance is feasible as this is easy to check. Let n be the number of clients. Let $\ell(t) = \ln\left(\frac{1}{1-t}\right)$ and $\ell_j = \ell(p_j)$. For a client-set S, define $\mathsf{ac}(S) = \Pr^{\text{IA}}[$some client in S is activated$] = 1 - \prod_{j \in S}(1 - p_j)$ and $\ell(S) = \sum_{j \in S} \ell_j$. We have $(1 - e^{-1}) \min\{1, \ell(S)\} \leq \mathsf{ac}(S) = 1 - e^{-\ell(S)} \leq \min\{1, \ell(S)\}$. Thus, $\mathsf{ac}(S) \leq t$ iff $\ell(S) \leq \ell(t)$. Let $M = \sum_j \ell_j$. In both reductions, scenario $\{j\}$ occurs with probability ℓ_j/M in our RFL problem, and we set the probability threshold to $\ell(\rho)/M$.

For one-stage FL, scenario $\{j\}$ has budget $\frac{B \cdot M p_j}{\ell_j}$ (on the assignment cost), and client j has demand $M p_j / \ell_j$. We set the first-stage facility costs to $\{f_i\}$, and the second-stage facility costs are set very high (e.g., $M(n \max_{i,j} c_{ij} + \sum_i f_i) / \min_j \ell_j$; note that we are in the poly-scenario model, so we need not worry about the inflation factor λ). An optimal one-stage-FL solution $\left(K^*, \{a^*(j)\right)$ yields a solution to the RFL instance, where we open the facilities in K^* in stage I and assign each client j to $a^*(j)$. This is feasible because the violated scenarios in the RFL instance correspond to the clients in $S := \{j : c_{a^*(j)j} > B\}$, and $\mathsf{ac}(S) \leq \rho$ im-

plies that $\ell(S)/M \leq \ell(\rho)/M$. Also, clearly the RFL-cost is equal to its one-stage-FL-cost. Now consider any $(c_1, c_2, 1)$-solution to the RFL problem. We may assume that no facilities are opened in stage II. We obtain a one-stage-FL-solution where we open all the (stage-I) facilities opened by the RFL-solution and assign clients as in the RFL-solution. The cost of the solution is unchanged. If $\{a(j)\}$ denotes the client assignment, then $\Pr^{\mathrm{IA}}[$assignment cost of an activated client $> c_2 B] = \mathsf{ac}\big(S := \{j : c_{ja(j)} > c_2 B\}\big)$, which is at most ρ since $\ell(S)/M \leq \ell(\rho)/M$.

For APFL, scenario $\{j\}$ has budget $\frac{B \cdot p_j}{\ell_j}$ and client j has demand p_j/ℓ_j. we set the first-stage cost of facility i to f_i/M, and its second-stage cost to f_i. An optimal APFL-solution $\big(K^*, \{a^*(j)\}\big)$ yields the following feasible solution to the RFL problem. For $i \in K^*$, we open i in stage I if $\ell(\{j : a^*(j) = i\}) \geq 1$, and otherwise we open i in every scenario $\{j\}$ for which $a^*(j) = i$; we assign each client j to $a^*(j)$. The RFL-cost of this solution is $\frac{1}{M} \cdot \big[\sum_{i \in K^*} f_i \min\{1, \ell(\{j : a^*(j) = i\})\} + \sum_j p_j c_{a^*(j)j}\big] \leq \frac{e/(e-1)}{M} \cdot OPT_{\mathsf{APFL}}$. The feasibility of the solution follows from the same calculations as in the one-stage-FL problem. Suppose we have a $(c_1, c_2, 1)$-solution to RFL. This translates to an APFL-solution where we open all the (stage-I and stage-II) facilities opened by the RFL-solution and assign clients as in the RFL-solution. A similar calculation as above shows that the APFL-cost is at most $M(\mathsf{RFL}\text{-cost})$; also, as in the one-stage-FL setting, if $\{a(j)\}$ denotes the client assignment and $S = \{j : c_{a(j)j} > c_2 B\}$, then we obtain that $\mathsf{ac}(S) \leq \rho$ since $\ell(S)/M \leq \ell(\rho)/M$.

Finally, note that we do not use "metricity" anywhere, so the same reductions apply to the set-cover versions of these problems as they can be cast as non-metric facility-location problems. ∎

The following result for metric RFL complements the above theorem. Its proof is deferred to the full version of this paper.

Theorem 6.2. *For any $\epsilon > 0$ and any RFL problem, one can compute a $\big(4 + \frac{6}{e} + \epsilon, 3, 1\big)$-solution in time $\mathsf{poly}\big(\text{input size}, \log(\frac{1}{\epsilon})\big)$.*

7 Sampling lower bounds

We now prove various lower bounds on the sample-size required in the black-box model to obtain a bounded approximation guarantee for the risk-averse budgeted and robust problems. Say that a solution is an (ϵ, γ)-optimal solution if its cost is at most $(1+\epsilon)OPT + \gamma$.

Theorem 7.1. *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 $\Omega\big(\frac{1}{\kappa}\big)$ samples if the probability-threshold is violated by at most an additive κ;*
- *requires $\Omega\big(\frac{1}{\kappa\rho}\big)$ samples if the probability-threshold is violated by at most a multiplicative $(1+\kappa)$-factor.*

The proof of Theorem 7.1 relies on 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 $q \in (\varrho, 2\varrho]$, 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 $\mathcal{N}(\delta; \varrho) = \ln\big(\frac{1}{\delta} - 1\big)/4\varrho$ coin tosses for each threshold ϱ.

Lemma 7.1. *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, $N(\delta; \varrho) \geq \mathcal{N}(\delta; \varrho) := \ln\big(\frac{1}{\delta} - 1\big)/4\varrho$ 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 $\Pr[X = 0] > (1 - 2\varrho)^{\mathcal{N}(\delta; \varrho)} \geq \frac{\delta}{1-\delta}$. So $\mathcal{A}$ will say "$q \leq \varrho$" (and hence, err) for $q = q_0$, with probability more than δ. ∎

Proof of Theorem 7.1. Given Lemma 7.1, our strategy is to construct a (very simple) RSC instance, where there is one key scenario A, whose probability determines whether or not one should take a certain first-stage decision to achieve a low cost solution with bounded budget inflation. We show that an algorithm that always returns an (ϵ, γ)-solution can be used to distinguish whether $p_A \leq \kappa$ or $p_A > 2\kappa$, and hence, the algorithm must draw a certain number of samples.

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 $w^{\mathrm{I}}_{S_i} = B$ for all i, and $w^A_{S_1} = 0$, $w^A_{S_2} = w^A_{S_3} = 2B/3$ 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 $p_{A_1} = \rho - \kappa$, $p_{A_0} = 1 - p_{A_1} - p_{A_2}$. 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 inflated 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-inflation.) So since $\mathcal{A}$ has failure probability at most δ, by Lemma 7.1, it must draw $\Omega\left(\frac{1}{\kappa}\right)$ 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 $\Omega\left(\frac{1}{\kappa\rho}\right)$ samples. Moreover, taking $\rho = 0$ shows that one cannot hope to achieve any approximation guarantees in the (standard) budget model with black-box distributions. ∎

To show the impossibility of approximation in the standard robust model with a bounded sample size, 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] S. Agrawal, A. Saberi, and Y. Ye. Stochastic combinatorial optimization under probabilistic constraints. *Unpublished manuscript. ArXiv e-prints*, 0809, September 2008.

[2] D. Bertsimas, P. Jaillet, and A. Odoni A priori optimization. *Oper. Research* 38(6):1019–1033, 1990.

[3] J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming.* Springer-Verlag, NY, 1997.

[4] J. Byrka. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. In *Proceedings, 10th APPROX*, pages 29–43, 2007.

[5] G. Calafiore and M. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[6] M. Charikar, C. Chekuri, and M. Pál. Sampling bounds for stochastic optimization. In *Proceedings, 9th RANDOM*, pages 257–269, 2005.

[7] M. Charikar, S. Khuller, D. Mount, and G. Narasimhan. Algorithms for facility location problems with outliers. In *Proceedings of the 12th ACM-SIAM SODA*, pages 642–651, 2001.

[8] A. Charnes and W. Cooper. Uncertain convex programs: randomized solutions and confidence levels. *Management Science*, 6:73–79,1959.

[9] F. Chudak and D. Shmoys. Improved approximation algorithms for the uncapacitated facility location problem. *SIAM Journal on Computing*, 33(1):1–25, 2003.

[10] V. Chvátal. A greedy heuristic for the set-covering problem. *Math. of Oper. Res.*, 4:233–235, 1979.

[11] K. Dhamdhere, V. Goyal, R. Ravi, and M. Singh. How to pay, come what may: approximation algorithms for demand-robust covering problems. In *Proc., 46th IEEE FOCS*, pages 367–378, 2005.

[12] S. Dye, L. Stougie, and A. Tomasgard. The stochastic single resource service-provision problem. *Naval Research Logistics*, 50(8):869–887, 2003.

[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. In *Proceedings of the 13th IPCO*, pages 439–453, 2007.

[15] N. Garg, A. Gupta, S. Leonardi, and P. Sankowski. Stochastic analyses for online combinatorial optimization problems. In *Proceedings of the 19th ACM-SIAM SODA*, pages 942–951, 2008.

[16] A. Goel and P. Indyk. Stochastic load balancing. In *Proceedings, 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 579–586, 1999.

[17] V. Goyal and R. Ravi. Approximation algorithms for robust covering problems with chance constraints. *Tepper WP 2008-E17*. Cited version: https://littlehurt.gsia.cmu.edu/gsiadoc/wp/2008-E17.pdf.

[18] 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. In *Proc., 23rd STACS*, pages 206–217, 2006.

[19] A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *Proceedings, 36th Annual ACM STOC*, pages 417–426, 2004.

[20] A. Gupta, M. Pál, R. Ravi, and A. Sinha. What about Wednesday? Approximation algorithms for multistage stochastic optimization. *Proceedings of 8th APPROX*, pages 86–98, 2005.

[21] A. Gupta, R. Ravi, and A. Sinha. An edge in time saves nine: LP rounding approximation algorithms for stochastic network design. In *Proceedings, 45th IEEE FOCS*, pages 218-227, 2004.

[22] M. Hajiaghayi, K. Jain The prize-collecting general-

ized Steiner tree problem via a new approach of primal-dual schema. In *Proceedings of the 17th ACM-SIAM SODA*, pages 631–640, 2006.

[23] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[24] N. Immorlica, 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.

[25] P. Jorion. *Value at Risk: A New Benchmark for Measuring Derivatives Risk*. Irwin Professional Publishers, New York, 1996.

[26] R. Khandekar, G. Kortsarz, V. Mirrokni, and M. Salavatipour. Two-stage robust network design with exponential scenarios. In *Proceedings, 16th ESA*, pages 589–600, 2008.

[27] J. Kleinberg, Y. Rabani, and É. Tardos. Allocating bandwidth for bursty connections. *SIAM Journal on Computing*, 30(1):191–217, 2000.

[28] 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.

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

[30] 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.

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

[32] A. Prékopa. Probabilistic programming. In A. Ruszczynski and A. Shapiro, eds., *Stochastic Programming*, vol. 10 of *Handbooks in Oper. Research and Mgmt. Sc.*, North-Holland, Amsterdam, 2003.

[33] M. Pritsker. Evaluating value at risk methodologies. *Journal of Financial Services Research*, 12(2/3):201–242, 1997.

[34] R. Ravi and A. Sinha. Hedging uncertainty: approximation algorithms for stochastic optimization problems. *Mathematical Programming, Series A*, 108:97–114, 2006.

[35] R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *J. Banking and Finance*, 26:1443–1471, 2002.

[36] A. Ruszczynski and A. Shapiro. Editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Mgmt. Sc.*, North-Holland, Amsterdam, 2003.

[37] A. Ruszczynski 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.

[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 K. Talwar. A constant approximation algorithm for the a priori Traveling Salesman Problem. In *Proceedings, 14th IPCO*, pages 331–343, 2008.

[40] D. B. Shmoys, É. Tardos, and K. I. Aardal. Approximation algorithms for facility location problems. In *Proceedings of the 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. In *Proceedings of the 9th APPROX*, pages 224–235, 2006.

[42] A. Srinivasan. Approximation algorithms for stochastic and risk-averse optimization. In *Proceedings of the 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. Algorithms for probabilistically-constrained models of risk-averse stochastic optimization with black-box distributions. *Unpublished manuscript. ArXiv e-prints, 805*, May 2008.

[45] C. Swamy and D. B. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. *ACM SIGACT News*, 37(1):33–46, March 2006.

[46] 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 *Proc., 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 357–366, 2005.

A Tightness of Theorem 4.1, and proof of Theorem 4.2

Tight examples for Theorem 4.1. Consider a RSC instance with k scenarios, each occurring with probability $\frac{1}{k}$. Each scenario A specifies a disjoint set of $1 + \frac{1}{\varepsilon}$ elements to be covered, and each of these elements is covered by a unique set. For notational simplicity, we assume that no set can be chosen in stage I; clearly, this can be simulated by setting exorbitant first-stage costs. We set the second-stage cost of all sets to 1, and set $B = 1$, $\rho = \frac{1}{1+\varepsilon}$. Then, the solution $y_{A,S} = \frac{\varepsilon}{1+\varepsilon}$ for all the sets S covering A, $r_A = \frac{1}{1+\varepsilon}$, for all A is feasible to (RSCP). But any solution to (RSCP) with $r_A \in \{0, 1\}$ must violate either (4.1) by a $(1+\varepsilon)$-factor or (4.4) by a $\left(1 + \frac{1}{\varepsilon}\right)$-factor; equivalently, if we require that $\Pr_A\left[f_A(\mathbf{0}) > \gamma B\right] \le \sigma\rho$, then either $\gamma \ge \left(1 + \frac{1}{\varepsilon}\right)$ or $\sigma \ge (1 + \varepsilon)$.

Now, modify the above instance so that there is

now just one set S that covers all the elements, whose first-stage cost is $1 + \frac{1}{\varepsilon}$ and the second-stage cost is negligible but non-zero. We set $B = 0$, so that the risk-averse budgeted problem becomes essentially a one-stage problem (of picking sets in stage I so as to cover "most" scenarios). In the following discussion We ignore the negligible second-stage cost incurred; all we need is that if some element of A is not covered in stage I, then the budget-constraint of scenario A is violated. The solution $x_S = \frac{\varepsilon}{1+\varepsilon}$, $y_{A,S} = 0$, $r_A = \frac{1}{1+\varepsilon}$ for all A is feasible to (RSCP) and has cost 1. But if x is a solution to the fractional risk-averse problem of cost $\gamma \cdot OPT$ with $\Pr_A[f_A(x) > 0] \leq \sigma\rho$, then either $\gamma \geq \left(1 + \frac{1}{\varepsilon}\right)$ or $\sigma \geq (1 + \varepsilon)$.

Proof of Theorem 4.2. Let $G = (V, E)$, k be the input to the MinDkS problem. Let $n = |V|$, $m = |E|$. We create a RSC instance with $m + 1$ scenarios. For each edge $e = (u, v)$, we create two elements (e, u) and (e, v) in our universe. For each node $u \in V$, we create a set $S_u := \{(e, u) : e \in \delta(u)\}$. For each edge $e = (u, v) \in E$, we create a scenario A_e, where we need to cover the two elements (e, u), (e, v); each such scenario A_e occurs with probability $\frac{1}{2m}$. We set the first-stage cost of each set S_u to 1, its second-stage cost to be negligible (e.g., $\frac{1}{n^3}$), but non-zero. Also, we set $B = 0$. So the budget constraint of a scenario A_{uv} is satisfied iff both S_u and S_v are picked (to an extent of 1) in stage I; otherwise, we incur a negligible second-stage cost for A_{uv}. In the sequel, we ignore this negligible second-stage cost incurred for "unsatisfied" scenarios. Observe that the RSC problem now essentially becomes a *one-stage* problem of picking suitable sets in stage I. If we now set $\rho = (m - k)/2m$, then it is easy to argue that any $(c_1, \cdot, 1)$-solution to RSC yields a c_1-approximate solution to MinDkS. This was shown by Goyal and Ravi [17]. To strengthen their result and prove that even obtaining a $(c_1, \cdot)$-scheme is difficult (modulo MinDkS), even when ρ is a constant, we do the following. We create a "filler" element f in our universe, a filler set $S_f = \{f\}$, and a filler scenario A_0 occurring with probability $\frac{k}{2m}$ where we have to cover f (and the null scenario $\emptyset$ occurs with the remaining probability). We give S_f a very high first-stage cost (e.g., n^3), and its second-stage cost to be negligible but nonzero. Note that any any $(c_1, \cdot, \cdot)$-solution to RSC with $c_1 \leq n$, will avoid picking S_f in stage I. We set $\rho = \frac{1}{2}$. It is clear that the size of the RSC instance is $\mathsf{poly}(m, n, k)$.

Any solution to MinDkS translates to a solution to RSC of the same cost where we pick the sets corresponding to the nodes in the solution in stage I. Now consider any $\left(c_1, \cdot, (1 + \frac{1}{2m})\right)$-solution to RSC (so S_f is not picked in stage I). Let k' be the number of edges in the subgraph induced by the nodes corresponding to the sets

picked in stage I. Then, $(m - k') \cdot \frac{1}{2m} + p_{A_0} \leq \frac{1}{2}\left(1 + \frac{1}{2m}\right)$, so $m - k' + k \leq m + \frac{1}{2}$; k' is an integer, so this implies that $k' \geq k$. So we obtain a c_1-approximate solution to MinDkS in polytime.

The same reduction works for fractional RSC, since as mentioned above, a scenario A_{uv} is satisfied iff both S_u and S_v are picked to an extent of 1 in the (fractional) solution.

Finally, Hajiaghayi and Jain [22] showed that a c_1-approximation algorithm for MinDkS yields a $2c_1^2$-approximation algorithm for DkS. ∎

B Proof of SAA result stated in Section 2

Recall that

$$\min \quad \left(h(x) := c(x) + \sum_A p_A f_A(x)\right) \tag{RA-P}$$
$$\text{s.t.} \quad x \in \mathcal{F}, \quad \Pr_A[f_A(x) > B] \leq \rho$$

denotes the (discrete) risk-averse problem, and

$$\min \quad \left(\widehat{h}(x) := c(x) + \sum_A \widehat{p}_A f_A(x)\right)$$
$$\text{s.t.} \quad x \in \mathcal{F}, \quad \widehat{\Pr}_A[f_A(x) > B] \leq \widehat{\rho} := \rho(1 + \kappa). \tag{SA-P}$$

is the corresponding SAA problem. Let $\mathcal{F}_\rho \subseteq \mathcal{F}$ be the feasible region of (RA-P). Let $x^* \in \mathcal{F}_\rho$ be an optimal solution to (RA-P), and $O^* = h(x^*)$ denote its value. We prove the following theorem.

THEOREM B.1. *Consider $k = \frac{2}{\epsilon}\log\left(\frac{1}{\delta}\right)$ SAA problems* (SA-P) *with objective functions $\widehat{h}^{(1)}, \ldots, \widehat{h}^{(k)}$, each constructed using $\mathcal{N} = \mathsf{poly}\left(\mathcal{I}, \frac{\lambda}{\epsilon + \kappa\rho}, \ln\left(\frac{1}{\delta}\right)\right)$ independent samples. Let $\mathcal{F}_{\widehat{\rho}}^{(i)}$ denote the (possibly empty) feasible region of* (SA-P). *Then, with probability at least $1 - 4\delta$,*
(i) if $\mathcal{F}_{\widehat{\rho}}^{(i)} = \emptyset$ for some i, then (RA-P) *is infeasible;*
(ii) Suppose for each $i = 1, \ldots, k$, $\left(x^{(i)}, \{y_A^{(i)}\}\right)$ is a (possibly infeasible) solution to the i-th SAA problem such that

$$c(x^{(i)}) + \sum_A \widehat{p}_A^{(i)} f_A(x^{(i)}, y_A^{(i)}) \leq \alpha \cdot \min_{x \in \mathcal{F}_{\widehat{\rho}}^{(i)}} \widehat{h}^{(i)}(x),$$
$$\widehat{\Pr}_A^{(i)}[f_A(x^{(i)}, y_A^{(i)}) > \gamma B] \leq \rho\left(1 + O(\kappa)\right) \tag{2.13}$$

where $\gamma \geq 1$. Let $j = \arg\min_{i=1}^k\left[c(x^{(i)}) + \sum_A \widehat{p}_A^{(i)} f_A(x^{(i)}, y_A^{(i)})\right]$ and $(\widehat{x}, \{\widehat{y}_A\}) = \left(x^{(j)}, \{y_A^{(j)}\}\right)$. Then, $h(\widehat{x}) \leq \left(\alpha + O(\epsilon)\right)O^*$ and $\Pr_A[f_A(\widehat{x}) > \gamma B] \leq \rho\left(1 + O(\kappa)\right)$.

Proof. The proof is along the lines of the proof of the SAA method in [6] and generalizes some of their arguments. We argue at various places that the sample

size implies that a certain event must happen with high probability, and proceed with the proof assuming that this high probability event happens. Note that $\ln|\mathcal{F}| = \mathsf{poly}(\mathcal{I})$. The sample size is chosen so that using standard Chernoff bounds, one can prove that $\left|\widehat{\mathrm{Pr}}_A^{(i)}[f_A(x) > B] - \mathrm{Pr}_A[f_A(x) > B]\right| \le \kappa\rho$, and $\left|\widehat{\mathrm{Pr}}_A^{(i)}[f_A(x) > \gamma B] - \mathrm{Pr}_A[f_A(x) > \gamma B]\right| \le \kappa\rho$, for every i and every $x \in \mathcal{F}$, with probability at least $1-\delta/(k|\mathcal{F}|)$. So these properties hold simultaneously for all i and all $x \in \mathcal{F}$ with probability at least $1 - \delta$. This means that for each i, we have $\mathcal{F}_\rho \subseteq \mathcal{F}_{\widehat{\rho}}^{(i)}$, so if $\mathcal{F}_{\widehat{\rho}}^{(i)} = \emptyset$, then (RA-P) is infeasible. Also, this shows that $\mathrm{Pr}_A[f_A(\hat{x}) > \gamma B] \le \widehat{\mathrm{Pr}}^{(j)}[f_A(\hat{x}) > \gamma B] + \kappa\rho \le \rho\big(1 + O(\kappa)\big)$. Now to bound $h(\hat{x})$, we observe that since $\mathcal{F}_\rho \subseteq \mathcal{F}_{\widehat{\rho}}^{(i)}$ for all i, (2.13) implies that

$$\widehat{h}^{(i)}(x^{(i)}) = c(x^{(i)}) + \sum_A \widehat{p}_A^{(i)} f_A(x^{(i)}, y_A^{(i)})$$
$$\le \alpha \cdot \min_{x \in \mathcal{F}_\rho} \widehat{h}^{(i)}(x) \quad \forall i = 1, \dots, k. \tag{2.14}$$

This says that for every i, $x^{(i)}$ is an α-approximate solution over $\mathcal{F}_\rho$. We now proceed in a way similar to [6], and adapt their arguments to show that the "best" of these $x^{(i)}$ α-approximate solutions, which we call $\hat{x}$, will be such that its $h(.)$-value will be an α-approximation to the minimum $h(.)$-value over $\mathcal{F}_\rho$

Let $M = \frac{2\lambda}{\epsilon} \cdot O^*$. Call a scenario A "high", if $f_A(\mathbf{0}) \ge M$, and "low" otherwise. Let $p^h = \sum_{A:A \text{ is high}} p_A$ We use $\mathrm{E}_A^l[.]$ (resp. $\mathrm{E}_A^h[.]$) to denote the expectation (wrt. the true, i.e., p- distribution) where non-low (resp. non-high) scenarios contribute 0 (so $\mathrm{E}_A[.] = \mathrm{E}_A^l[.] + \mathrm{E}_A^h[.]$). Let $\widehat{p}^{(i),h}$, $\widehat{\mathrm{E}}_A^{(i),l}[.]$, and $\widehat{\mathrm{E}}_A^{(i),h}[.]$ denote these quantities for the i-th SAA problem.

Since $h(x^*) \ge \mathrm{E}_A^h[f_A(x^*)] \ge p^h(M - \lambda c(x^*))$, we have $p^h \le \frac{\epsilon}{\lambda}$. The sample size $\mathcal{N}$ is chosen so that Chernoff bounds ensure that with probability at least $1 - \delta$, for every i, we have $\widehat{p}^{(i),h} \le \frac{2\epsilon}{\lambda}$. It follows that

$$\mathrm{E}_A^h[f_A(\mathbf{0})] - \mathrm{E}_A^h[f_A(x)] \le \epsilon c(x) \quad \forall x \in \mathcal{F} \tag{2.15}$$
$$\widehat{\mathrm{E}}_A^{(i),h}[f_A(\mathbf{0})] - \widehat{\mathrm{E}}_A^{(i),h}[f_A(x)] \le 2\epsilon c(x) \quad \forall i, \, y \in \mathcal{F}. \tag{2.16}$$

Since $f_A(x) \le f_A(\mathbf{0}) < M$ for all low scenarios A and all $x \in \mathcal{F}$, again using Chernoff bounds, one can verify that with probability $1 - \delta$, we have

$$\left|\widehat{\mathrm{E}}_A^{(i),l}[f_A(x)] - \mathrm{E}_A^l[f_A(x)]\right| \le \epsilon O^* \quad \forall i, \, x \in \mathcal{F}. \tag{2.17}$$

Finally, notice that for every i, the expected value of $\widehat{\mathrm{E}}_A^{(i),h}[f_A(\mathbf{0})] = \sum_{A:A \text{ is high}} \widehat{p}_A^{(i)} f_A(\mathbf{0})$ is precisely

$\mathrm{E}_A^h[f_A(\mathbf{0})]$. Therefore, by Markov's inequality, we have that $\widehat{\mathrm{E}}_A^{(i),h}[f_A(\mathbf{0})] > (1 + \epsilon)\mathrm{E}_A^h[f_A(\mathbf{0})]$ holds with probability at most $\frac{1}{1+\epsilon} \le 1 - \epsilon/2$. So the probability that this happens for *all* $i = 1, \dots, k$ is at most $(1 - \epsilon/2)^k \le \delta$. So we may assume that there is some index t such that $\widehat{\mathrm{E}}_A^{(t),h}[f_A(\mathbf{0})]$ is at most

$$(1 + \epsilon)\mathrm{E}_A^h[f_A(\mathbf{0})] \le \mathrm{E}_A^h[f_A(x^*)] + p^h \lambda c(x^*)$$
$$\le O^* + \epsilon c(x^*) \tag{2.18}$$

Finally, we combine these various inequalities to obtain the desired result. Applying (2.14) to j and t, we have $\widehat{h}^{(j)}(\hat{x}) \le \alpha \cdot \widehat{h}^{(j)}(x^*)$ and

$$\widehat{h}^{(j)}(\hat{x}) \le c(x^{(j)}) + \sum_A \widehat{p}_A^{(j)} f_A(x^{(j)}, y_A^{(j)})$$
$$\le c(x^{(t)}) + \sum_A \widehat{p}_A^{(t)} f_A(x^{(t)}, y_A^{(t)}) \le \alpha \widehat{h}^{(t)}(x^*).$$

Multiplying the first inequality by $\frac{1}{\alpha}$ and the second by $1 - \frac{1}{\alpha}$ and adding, we get $\widehat{h}^{(j)}(\hat{x}) \le \widehat{h}^{(j)}(x^*) + (\alpha - 1)\widehat{h}^{(t)}(x^*)$. Let $Y^j = \mathrm{E}_A^h[f_A(\mathbf{0})] - \widehat{\mathrm{E}}_A^{(j),h}[f_A(\mathbf{0})]$. By repeatedly using (2.15)–(2.17), we get

$$h(\hat{x}) = c(\hat{x}) + \mathrm{E}_A^l[f_A(\hat{x})] + \mathrm{E}_A^h[f_A(\hat{x})]$$
$$\le c(\hat{x}) + \left(\widehat{\mathrm{E}}_A^{(j),l}[f_A(\hat{x})] + \epsilon O^*\right)$$
$$+ \left(\widehat{\mathrm{E}}_A^{(j),h}[f_A(\hat{x})] + Y^j + 2\epsilon c(\hat{x})\right)$$
$$\le \left(\widehat{h}^{(j)}(x^*) + Y^j\right) + (\alpha - 1)\widehat{h}^{(t)}(x^*)$$
$$+ \epsilon O^* + 2\epsilon c(\hat{x}). \tag{2.19}$$

We bound $\widehat{h}^{(j)}(x^*) + Y^{(j)}$ by

$$c(x^*) + \left(\mathrm{E}_A^l[f_A(x^*)] + \epsilon O^*\right) + \left(\mathrm{E}_A^h[f_A(x^*)] + \epsilon c(x^*)\right)$$
$$\le (1 + \epsilon)O^* + \epsilon c(x^*). \tag{2.20}$$

Similarly, we have $\widehat{h}^{(t)}(x^*) \le c(x^*) + \big(\mathrm{E}_A^l[f_A(x^*)] + \epsilon O^*\big) + \big((\mathrm{E}_A^h[f_A(x^*)] + \epsilon c(x^*)) + (\widehat{\mathrm{E}}_A^{(t),h}[f_A(\mathbf{0})] - \mathrm{E}_A^h[f_A(\mathbf{0})])\big)$. Substituting $\widehat{\mathrm{E}}_A^{(t),h}[f_A(\mathbf{0})] - \mathrm{E}_A^h[f_A(\mathbf{0})] \le \epsilon \mathrm{E}_A^h[f_A(\mathbf{0})] \le \epsilon O^* + \epsilon^2 c(x^*)$ from (2.18), we can simplify the above bound to

$$\widehat{h}^{(t)}(x^*) \le (1 + 2\epsilon)O^* + 2\epsilon c(x^*).$$

Finally, substituting this bound and (2.20), in (2.19), we obtain

$$h(\hat{x}) \le (1 + \epsilon)O^* + \epsilon c(x^*) + (\alpha - 1)\big((1 + 2\epsilon)O^* + 2\epsilon c(x^*)\big)$$
$$+ \epsilon O^* + 2\epsilon c(\hat{x}).$$

This implies that $\widehat{h}(\hat{x}) \le \big(\alpha + O(\epsilon)\big)O^*$. The success probability is at least $1 - 4\delta$. ∎