Stochastic Combinatorial Optimization with Controllable Risk Aversion Level* (Extended Abstract)

Anthony Man–Cho So[1], Jiawei Zhang[2], and Yinyu Ye[3]

[1] Department of Computer Science, Stanford University, Stanford, CA 94305, USA.
manchoso@cs.stanford.edu

[2] Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York, NY 10012, USA.
jzhang@stern.nyu.edu

[3] Department of Management Science and Engineering and, by courtesy, Electrical Engineering, Stanford University, Stanford, CA 94305, USA.
yinyu-ye@stanford.edu

Abstract. Due to their wide applicability and versatile modeling power, stochastic programming problems have received a lot of attention in many communities. In particular, there has been substantial recent interest in 2–stage stochastic combinatorial optimization problems. Two objectives have been considered in recent work: one sought to minimize the expected cost, and the other sought to minimize the worst–case cost. These two objectives represent two extremes in handling risk — the first trusts the average, and the second is obsessed with the worst case. In this paper, we interpolate between these two extremes by introducing an one–parameter family of functionals. These functionals arise naturally from a change of the underlying probability measure and incorporate an intuitive notion of risk. Although such a family has been used in the mathematical finance [11] and stochastic programming [13] literature before, its use in the context of approximation algorithms seems new. We show that under standard assumptions, our risk–adjusted objective can be efficiently treated by the Sample Average Approximation (SAA) method [9]. In particular, our result generalizes a recent sampling theorem by Charikar et al. [2], and it shows that it is possible to incorporate some degree of robustness even when the underlying probability distribution can only be accessed in a black–box fashion. We also show that when combined with known techniques (e.g. [4, 14]), our result yields new approximation algorithms for many 2–stage stochastic combinatorial optimization problems under the risk–adjusted setting.

1 Introduction

A fundamental challenge that faces all decision–makers is the need to cope with an uncertain environment while trying to achieve some predetermined objectives.

* This research is supported in part by The Boeing Company.

One certainly does not need to go far to encounter such situations — for example, an office clerk trying to get to work as fast as possible while avoiding possibly congested roads; a customer in the supermarket trying to checkout while avoiding lines that may take a long time, and so on. From a decision–maker's perspective, it is then natural to ask whether one can determine the optimal decision given one's assessment of the uncertain environment. This is a motivating question in the field of stochastic optimization. To keep our discussion focused, we shall consider the class of 2–stage stochastic programs with recourse [1, 3], particularly in the context of combinatorial optimization problems. Roughly speaking, in the 2–stage recourse model, one commits to some initial (i.e. first stage) action x based on one's knowledge of the underlying probability distribution. The actions in the second stage cannot be determined in advance, since they depend on the actions of the first stage as well as the uncertain parameters of the problem. However, once those parameters are realized (according to the distribution), a recourse (i.e. second stage) action r can be taken so that, together with the first stage actions, all the requirements of the problem are satisfied. Naturally, one would seek for the action (x, r) that minimizes the "total cost". However, since the outcome is random, such an objective can have many possible interpretations. In this paper, we shall consider the problem of risk minimization. Specifically, let X be the set of permissible actions, and let $(\Omega, \mathscr{B}, \mathbb{P})$ be the underlying probability space. In accordance with the convention in the literature, we shall assume that the probability distribution is specified via one of the following models:

(a) **Scenario Model**: The set of scenarios $\mathcal{S}$ and their associated probabilities are explicitly given. Hence, under this model, an algorithm is allowed to take time polynomial in $|\mathcal{S}|$.
(b) **Black–Box Model**: The distribution of the scenarios is given as a black box. An algorithm can use this black box to draw independent samples from the distribution of scenarios.

We are interested in solving problems of the form:

$$\min_{x \in X} \{ g(x) \equiv c(x) + \Phi(q(x, \omega)) \} \tag{1}$$

where $c : X \to \mathbb{R}_+$ is a (deterministic) cost function, $q : X \times \Omega \to \mathbb{R}_+$ is another cost function that depends both on the decision $x \in X$ and some uncertain parameter $\omega \in \Omega$, and $\Phi : L^2(\Omega, \mathscr{B}, \mathbb{P}) \to \mathbb{R}$ is some *risk measure*. We shall refer to Problem (1) as a *risk–adjusted* 2*–stage stochastic program with recourse*. Two typical examples of Φ are the *expectation* operator and the *max* operator. The former gives rise to a risk–neutral objective, while the latter gives rise to an extremely risk–averse objective. Both of these risk measures have been studied in recent works on approximation algorithms for stochastic combinatorial optimization problems (see, e.g., [2, 4–8, 12, 14]). For the case where Φ is the expectation operator, it turns out that under the black–box model, one can obtain a near–optimal solution to Problem (1) with high probability by the so–called Sample Average Approximation (SAA) method [9]. Roughly speaking, the SAA method works as follows. Let $\omega^1, \ldots, \omega^N$ be N i.i.d. samples drawn from the underlying distribution, and consider the sampled problem:

$$\min_{x \in X} \frac{1}{N} \sum_{i=1}^{N} \left(c(x) + q(x, \omega^i) \right) \tag{2}$$

Under some mild assumptions, it has been shown [9] that the optimal value of (2) is a good approximation to that of (1) with high probability, and that the number of samples N can be bounded. Unfortunately, the bound on N depends on the maximum variance V (over all $x \in X$) of the random variables $q(x, \omega)$, which need not be polynomially bounded. However, in a recent breakthrough, Shmoys and Swamy [14] have been able to circumvent this problem for a large class of 2–stage stochastic linear programs. Specifically, by bounding the relative factor by which the second stage actions are more expensive than the first stage actions by a parameter λ (called the *inflation factor*), they are able to show that an adaptation of the ellipsoid method will yield an $(1+\epsilon)$–approximation with the number of samples (i.e. black–box accesses) bounded by a polynomial of the input size, λ and $1/\epsilon$. Subsequently, Charikar et al. [2] have established a similar but more general result using the SAA method. We should mention, however, that both of these results assume that the objective function is linear. Thus, in general, they do not apply to Problem (1).

On another front, motivated by robustness concerns, Dhamdhere et al. [4] have recently considered the case where Φ is the max operator and developed approximation algorithms for various 2–stage stochastic combinatorial optimization problems with recourse under that setting. Their framework works under the scenario model. In fact, since only the worst case matters, it is not even necessary to specify any probabilities in their framework. However, such a model can be too pessimistic. Also, as the worst–case scenario may occur only with an exponentially small probability, it seems unlikely that sampling techniques would apply to such problems.

From the above discussion, a natural question arises whether we can incorporate a certain degree of robustness (possibly with some other risk measures Φ) in the problem while still being able to solve it in polynomial time under the black–box model. If so, can we also develop approximation algorithms for some well–studied combinatorial optimization problems under the new robust setting?

Our Contribution. In this paper, we answer both of the above questions in the affirmative. Using techniques from the mathematical finance literature [11], we provide a unified framework for treating the aforementioned risk–adjusted stochastic optimization problems. Specifically, we use an one–parameter family of functionals $\{\varphi_\alpha\}_{0 \le \alpha < 1}$ to capture the degree of risk aversion, and we consider the problem $\min_{x \in X} \{c(x) + \varphi_\alpha(q(x, \omega))\}$. As we shall see, such a family arises naturally from a change of the underlying probability measure $\mathbb{P}$ and possesses many nice properties. In particular, it includes $\Phi = \mathbb{E}$ as a special case and $\Phi = \max$ as a limiting case. Thus, our framework provides a generalization of those in previous works. Moreover, our framework works under the most general black–box model, and we show that as long as one does not insist on considering the worst–case scenario, one can use sampling techniques to obtain near–optimal solutions to the problems discussed above efficiently. Our sampling theorem and its analysis can be viewed as a generalization of those by Charikar et al. [2]. Consequently, our result extends the class of problems that can be efficiently treated by the SAA method. Finally, by combining with techniques developed in earlier works [4, 6, 12, 14], we obtain new approximation algorithms for a large class of 2–stage stochastic combinatorial optimization problems under the new robust setting.

2 Motivation: Risk Aversion as Change of Probability Measure

We begin with the setup and some notations. Let $(\Omega, \mathscr{B}, \mathbb{P})$ be a probability space, and let $L^2(\Omega, \mathscr{B}, \mathbb{P})$ be the Hilbert space of square–integrable random variables with inner product $\langle \cdot, \cdot \rangle$ given by $\langle U, V \rangle = \int_\Omega UV \, d\mathbb{P}$. We shall assume that the second stage cost function q satisfies the following: (i) $q(x, \cdot)$ is measurable w.r.t. $\mathscr{B}$ for each $x \in X$, (ii) q is continuous w.r.t. x, and (iii) $\mathbb{E}[q(x, \omega)] < \infty$ for each $x \in X$. To motivate our approach, let us investigate how the following problems capture risk:

$$\min_{x \in X} \{c(x) + \mathbb{E}[q(x, \omega)]\} \tag{3}$$

$$\min_{x \in X} \left\{ c(x) + \sup_{\omega \in \Omega} q(x, \omega) \right\} \tag{4}$$

Problem (3) is a standard stochastic optimization problem, in which a first stage decision $x^* \in X$ is sought so that the sum of the first stage cost $c(x^*)$ and the expected second stage cost $\mathbb{E}[q(x^*, \omega)]$ is minimized. In particular, we do not consider any single scenario as particularly important, and hence we simply weigh them by their respective probabilities. On the other hand, Problem (4) is a pessimist's version of the problem, in which one considers the worst–case second stage cost over all scenarios. Thus, for each $x \in X$, we consider the scenario ω_x that gives the maximum second stage cost as most important, and we put a weight of 1 on ω_x and 0 on all $\omega \neq \omega_x$, regardless of what their respective probabilities are. These observations suggest the following approach for capturing risk. For each $x \in X$, let $f_x : \Omega \to \mathbb{R}_+$ be a measurable weighing function such that:

$$\int_\Omega f_x(\omega) \, d\mathbb{P}(\omega) = 1$$

Now, consider the problem:

$$\min_{x \in X} \{c(x) + \mathbb{E}[f_x(\omega) q(x, \omega)]\} \tag{5}$$

Observe that Problem (5) captures both Problems (3) and (4) as special cases. Indeed, if we set $f_x \equiv 1$, then we recover Problem (3). On the other hand, suppose that Ω is finite, with $\mathbb{P}(\omega) > 0$ for all $\omega \in \Omega$. Consider a fixed $x \in X$, and let $\omega' = \arg\max_{\omega\in\Omega} q(x,\omega)$. Then, by setting $f_x(\omega') = \frac{1}{\mathbb{P}(\omega')}$ and $f_x(\omega) = 0$ for all $\omega \neq \omega'$, we recover Problem (4).

From the above discussion, we see that one way of addressing risk is by changing the underlying probability measure $\mathbb{P}$ using a weighing function. Indeed, the new probability measure is given by:

$$\mathbb{Q}_x(\omega) \equiv f_x(\omega)\mathbb{P}(\omega) \tag{6}$$

and we may write $\mathbb{E}_{\mathbb{P}}\left[f_x(\omega)q(x,\omega)\right] = \mathbb{E}_{\mathbb{Q}_x}\left[q(x,\omega)\right]$. Alternatively, we can specify the probability measure $\mathbb{Q}_x$ directly without using weighing functions. As long as the new measure $\mathbb{Q}_x$ is absolutely continuous w.r.t. $\mathbb{P}$ for each $x \in X$ (i.e. $\mathbb{P}(\omega) = 0$ implies that $\mathbb{Q}_x(\omega) = 0$), there will be a corresponding weighing function f_x given precisely by (6). Thus, in this context, we see that f_x is simply the Radon–Nikodym derivative of $\mathbb{Q}_x$ w.r.t. $\mathbb{P}$.

Note that in the above formulation, we are allowed to choose a different weighing function f_x for each $x \in X$. Clearly, there are many possible choices for f_x. However, our goal is to choose the f_x's so that Problem (5) is computationally tractable. Towards that end, let us consider the following strategy. Let $\alpha \in [0,1)$ be a given parameter (the *risk–aversion level*), and define:

$$\mathcal{Q} = \left\{ f \in L^2(\Omega,\mathscr{B},\mathbb{P}) : 0 \le f(\omega) \le \frac{1}{1-\alpha} \text{ for all } \omega \in \Omega,\ \langle f, 1\rangle = 1 \right\} \tag{7}$$

For each $x \in X$, we take f_x to be the optimal solution to the following optimization problem:

$$f_x = \arg\max_{f\in\mathcal{Q}} \mathbb{E}_{\mathbb{P}}\left[f(\omega)q(x,\omega)\right] \tag{8}$$

Note that such an f_x always exists (i.e. the maximum is always attained), since the functional $f \mapsto \langle f, q(x,\cdot)\rangle$ is continuous, and the set $\mathcal{Q}$ is compact (in the weak*–topology) by the Banach–Alaoglu theorem (cf. p. 120 of [10]). Intuitively, the function f_x boosts the weights of those scenarios ω that have high second stage costs $q(x,\omega)$ by a factor of at most $(1-\alpha)^{-1}$, and zeroes out the weights of those scenarios that have low second stage costs. Note also that when $\alpha = 0$, we have $f_x \equiv 1$; and as $\alpha \nearrow 1$, f_x tends to a delta function at the scenario ω that has the highest cost $q(x,\omega)$. Thus, the definition of f_x in (8) captures the intuitive notion of risk as discussed earlier. We then define φ_α by $\varphi_\alpha(q(x,\omega)) \equiv \mathbb{E}_{\mathbb{P}}\left[f_x(\omega)q(x,\omega)\right]$, where f_x is given by (8).

At this point, it may seem that we need to perform the non–trivial task of computing f_x for many $x \in X$. However, it turns out that this can be circumvented by the following representation theorem of Rockafellar and Uryasev [11]. Such a theorem forms the basis for our sampling approach.

Fact 1. *(Rockafellar and Uryasev [11]) Let $\alpha \in (0,1)$, and for $x \in X$ and $\beta \in \mathbb{R}$, define:*

$$F_\alpha(x,\beta) = \beta + \frac{1}{1-\alpha}\mathbb{E}_{\mathbb{P}}\left[(q(x,\omega)-\beta)^+\right]$$

Then, $F_\alpha(x,\cdot)$ is finite and convex, with $\varphi_\alpha(q(x,\omega)) = \min_\beta F_\alpha(x,\beta)$. In particular, if q is convex w.r.t. x, then φ_α is convex w.r.t. x as well. Indeed, F_α is jointly convex in (x,β).

The power of the above representation theorem lies in the fact that it reduces the risk–adjusted stochastic optimization problem:

$$\min_{x\in X} \{c(x) + \varphi_\alpha(q(x,\omega))\} \tag{9}$$

to the well–studied problem of minimizing the expectation of a certain random function. Thus, it seems plausible that the machineries developed for solving the latter can be applied to Problem (9) as well. Moreover, when c, q are convex w.r.t. x and X is convex, Problem (9) is a convex optimization problem and hence can be solved (up to any prescribed accuracy) in polynomial time. In Section 3, we will show how the SAA method can be applied to obtain a near–optimal solution to (9).

3 Sampling Theorem for Risk–Adjusted Stochastic Optimization Problems

In this section, we show that for any fixed $\alpha \in [0,1)$, it suffices to have only a polynomial number of samples in order for the Sample Average Approximation (SAA) method [9] to yield a near–optimal solution to Problem (9). Our result and analysis generalize those in [2]. To begin, let X be a finite set, and let us assume that the functions $c : X \to \mathbb{R}$ and $q : X \times \Omega \to \mathbb{R}$ satisfy the following properties:

(a) **(Non–Negativity)** The functions c and q are non–negative for every first stage action $x \in X$ and every scenario $\omega \in \Omega$.
(b) **(Empty First Stage)** There exists a first stage action $\phi \in X$ such that $c(\phi) = 0$ and $q(x,\omega) \le q(\phi,\omega)$ for every $x \in X$ and $\omega \in \Omega$.
(c) **(Bounded Inflation Factor)** There exists an $\lambda \ge 1$ such that $q(\phi,\omega) - q(x,\omega) \le \lambda c(x)$ for every $x \in X$ and $\omega \in \Omega$.

We remark that the assumptions above are the same as those in [2] and capture those considered in recent work (see, e.g., [6, 8, 12, 14]). Now, let $g_\alpha(x) = c(x) + \varphi_\alpha(q(x,\omega))$. By Fact 1, we have $\min_{x\in X} g_\alpha(x) = \min_{(x,\beta)\in X\times\mathbb{R}} g'_\alpha(x,\beta)$, where $g'_\alpha(x,\beta) \equiv c(x) + \mathbb{E}_{\mathbb{P}}[q'(x,\beta,\omega)]$, and

$$q'(x,\beta,\omega) \equiv \beta + \frac{1}{1-\alpha}(q(x,\omega) - \beta)^+$$

Let $(x^*,\beta^*) \in X \times [0,\infty)$ be an exact minimizer of g'_α, and set $Z^* = g'_\alpha(x^*,\beta^*)$. It is easy to show that $\beta^* \in [0, Z^*]$. Furthermore, we have the following observation:

Lemma 1. *Let $\alpha \in [0,1)$, and let c, q and q' be as above.*

(a) *Let $\kappa \geq 1$ be fixed. For every $x \in X$, $\omega \in \Omega$ and $\beta \in [0, \kappa Z^*]$, we have:*

$$q'(x,\beta,\omega) \leq q'(\phi,\beta,\omega) \leq \max\left\{q'(\phi,0,\omega), q'(\phi,\kappa Z^*,\omega)\right\}$$

(b) *For every $x \in X$, $\omega \in \Omega$ and $\beta \in [0,\infty)$, we have:*

$$q'(\phi,\beta,\omega) - q'(x,\beta,\omega) \leq \frac{\lambda c(x)}{1-\alpha}$$

Before we proceed, let us first make a definition and state the version of the Chernoff bound that we will be using.

Definition 1. *We say that $x^* \in X$ is an exact (resp. γ–approximate) minimizer of a function f if we have $f(x^*) \leq f(x)$ (resp. $f(x^*) \leq \gamma f(x)$) for all $x \in X$.*

Lemma 2. (Chernoff Bound) *Let $V_1, \ldots, V_n$ be independent random variables with $V_i \in [0,1]$ for $i = 1, \ldots, n$. Set $V = \sum_{i=1}^n V_i$. Then, for any $\epsilon > 0$, we have $P\left(\left|V - \mathbb{E}[V]\right| > \epsilon n\right) \leq 2e^{-\epsilon^2 n}$.*

Here is our main sampling theorem.

Theorem 1. *Let $g'_\alpha(x,\beta) = c(x) + \mathbb{E}_{\mathbb{P}}[q'(x,\beta,\omega)]$, where c and q' satisfy the assumptions above, and $\alpha \in [0,1)$ is the risk–aversion level. Let $\epsilon \in (0, 1/3]$ and $\delta \in (0,1/2)$ be given. Set:*

$$\lambda_\alpha = \frac{\lambda}{1-\alpha}; \quad \eta = \max\left\{1, \frac{\alpha}{1-\alpha}\right\}$$

and define:

$$\hat{g}^N_\alpha(x,\beta) = c(x) + \beta + \frac{1}{N(1-\alpha)} \sum_{i=1}^N \left(q(x,\omega^i) - \beta\right)^+$$

to be the SAA of g'_α, where $\omega^1, \ldots, \omega^N$ are N i.i.d. samples from the underlying distribution, and

$$N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4(1-\alpha)^2} \log\left(\frac{\eta}{\epsilon} \cdot |X| \cdot \frac{1}{\delta}\right)\right)$$

Let $\kappa \geq 1$ be fixed, and suppose that $(\bar{x},\bar{\beta}) \in X \times [0,\kappa Z^*]$ is an exact minimizer of $\hat{g}^N_\alpha$ over the domain $X \times [0,\kappa Z^*]$. Then, with probability at least $1-2\delta$, the solution $(\bar{x},\bar{\beta})$ is an $(1+\Theta(\epsilon\kappa))$–approximate minimizer of g'_α.

Remarks:

(a) Note that $(\bar{x},\bar{\beta})$ needs not be a global minimizer of $\hat{g}^N_\alpha$ over $X \times [0,\infty)$, since such a global minimizer may have $\beta > \kappa Z^*$. In particular, the optimal solutions to the problems:

$$\min_{(x,\beta)\in X\times[0,\infty)} \hat{g}^N_\alpha(x,\beta) \tag{10}$$

and

$$\min_{(x,\beta)\in[0,\kappa Z^*]} \hat{g}^N_\alpha(x,\beta) \tag{11}$$

could be different. From a practitioner's point of view, it may be easier to solve (10) than (11), because in many applications, it is difficult to estimate Z^* without actually solving the problem. However, it can be shown (see Theorem 2) that by repeating the sampling sufficiently many times, we can obtain a sample average approximation $\hat{g}_\alpha^N$ whose exact minimizers $(\bar{x}^*, \bar{\beta}^*)$ over $X \times [0, \infty)$ satisfy $\bar{\beta}^* \le (1+\epsilon)Z^*$ with high probability. Thus, we can still apply the theorem even though we are solving Problem (10).

(b) Note that this theorem does not follow from a direct application of Theorem 3 of [2] for two reasons. First, the domain of our optimization problem is $X \times [0, \kappa Z^*]$, which is compact but not finite. However, this can be circumvented by using a suitably chosen grid on $[0, \kappa Z^*]$. A second, and perhaps more serious, problem is that there may not exist an $\beta_0 \in [0, \kappa Z^*]$ such that $q'(x, \beta, \omega) \le q'(\phi, \beta_0, \omega)$ for *all* $x \in X$ and $\omega \in \Omega$. Such an assumption is crucial in the analysis in [2]. On the other hand, we have the weaker statement of Lemma 1(a), and that turns out to be sufficient for establishing our theorem.

Proof. Let (x^*, β^*) be an exact minimizer of g'_α. Then, we have $Z^* = g'_\alpha(x^*, \beta^*)$. Our proof consists of the following three steps.

Step 1: Isolate the high–cost scenarios and bound their total probability mass. We divide the scenarios into two classes. We say that a scenario ω is *high* if $q(\phi, \omega)$ exceeds some threshold M; otherwise, we say that ω is *low*. Let $p = \mathbb{P}(\omega : \omega \text{ is high})$, and define:

$$\hat{l}_\alpha^N(x,\beta) = \frac{1}{N} \sum_{i:\, \omega^i \; low} q'(x, \beta, \omega^i); \quad \hat{h}_\alpha^N(x,\beta) = \frac{1}{N} \sum_{i:\, \omega^i \; high} q'(x, \beta, \omega^i)$$

Then, it is clear that $\hat{g}_\alpha^N(x,\beta) = c(x) + \hat{l}_\alpha^N(x,\beta) + \hat{h}_\alpha^N(x,\beta)$. Similarly, we define:

$$l'_\alpha(x,\beta) = \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega) \cdot \mathbf{1}_{\{\omega \text{ is low}\}}\right] = (1-p) \cdot \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega) \,|\, \omega \text{ is low}\right]$$
$$h'_\alpha(x,\beta) = \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega) \cdot \mathbf{1}_{\{\omega \text{ is high}\}}\right] = p \cdot \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega) \,|\, \omega \text{ is high}\right]$$

whence $g'_\alpha(x,\beta) = c(x) + l'_\alpha(x,\beta) + h'_\alpha(x,\beta)$. Now, using the arguments of [2], one can show that $p \le \frac{\epsilon}{\lambda_\alpha(1-\epsilon)}$. In particular, by the Chernoff bound (cf. Lemma 2), we have the following lemma:

Lemma 3. *Let N_h be the number of high scenarios in the samples $w^1, \ldots, w^N$. Then, with probability at least $1-\delta$, we have $N_h/N \le 2\epsilon/\lambda_\alpha$.*

Step 2: Establish the quality of the scenario partition. We claim that each of the following events occurs with probability at least $1-\delta$:

$$A_1 = \left\{ \left| l'_\alpha(x,\beta) - \hat{l}_\alpha^N(x,\beta) \right| \le 2\epsilon\kappa Z^* \;\text{ for every } (x,\beta) \in X \times [0, \kappa Z^*] \right\}$$
$$A_2 = \left\{ \hat{h}_\alpha^N(\phi,\beta) - \hat{h}_\alpha^N(x,\beta) \le 2\epsilon c(x) \;\text{ for every } (x,\beta) \in X \times [0,\infty) \right\}$$
$$A_3 = \left\{ h'_\alpha(\phi,\beta) - h'_\alpha(x,\beta) \le 2\epsilon c(x) \;\text{ for every } (x,\beta) \in X \times [0,\infty) \right\}$$

A crucial observation needed in the proof and in the sequel is the following:

Lemma 4. *For each $x \in X$ and $\omega \in \Omega$, the function $q'(x,\cdot,\omega)$ is η–Lipschitz (i.e. $\left|q'(x,\beta_1,\omega) - q'(x,\beta_2,\omega)\right| \le \eta\left|\beta_1 - \beta_2\right|$), where $\eta = \max\left\{1, \frac{\alpha}{1-\alpha}\right\}$.*

Due to space limitations, we defer the proofs to the full version of the paper.

Step 3: Establish the approximation guarantee.
With probability at least $1 - 2\delta$, we may assume that all of the above events occur. Then, for any $(x,\beta) \in X \times [0,\kappa Z^*]$, we have:

$$\begin{aligned} l'_\alpha(x,\beta) &\le \hat{l}^N_\alpha(x,\beta) + 2\epsilon\kappa Z^* && \text{(Event } A_1\text{)} \\ h'_\alpha(x,\beta) &\le h'_\alpha(\phi,\beta) && \text{(Lemma 1(a))} \\ 0 &\le \hat{h}^N_\alpha(x,\beta) + 2\epsilon c(x) - \hat{h}^N_\alpha(\phi,\beta) && \text{(Event } A_2\text{)} \end{aligned}$$

Upon summing the above inequalities, we obtain:

$$g'_\alpha(x,\beta) - \hat{g}^N_\alpha(x,\beta) \le 2\epsilon\kappa Z^* + 2\epsilon c(x) + h'_\alpha(\phi,\beta) - \hat{h}^N_\alpha(\phi,\beta) \tag{12}$$

By a similar maneuver, we can also obtain:

$$\hat{g}^N_\alpha(x,\beta) - g'_\alpha(x,\beta) \le 2\epsilon\kappa Z^* + 2\epsilon c(x) + \hat{h}^N_\alpha(\phi,\beta) - h'_\alpha(\phi,\beta) \tag{13}$$

Now, let $(\bar{x},\bar{\beta}) \in X \times [0,\kappa Z^*]$ be an exact minimizer of $\hat{g}^N_\alpha$ over $[0,\kappa Z^*]$. Upon instantiating (x,β) by $(\bar{x},\bar{\beta})$ in (12) and by (x^*,β^*) in (13) and summing, we have:

$$\begin{aligned} &g'_\alpha(\bar{x},\bar{\beta}) - g'_\alpha(x^*,\beta^*) + \hat{g}^N_\alpha(x^*,\beta^*) - \hat{g}^N_\alpha(\bar{x},\bar{\beta}) \\ &\le 4\epsilon\kappa Z^* + 2\epsilon c(\bar{x}) + 2\epsilon c(x^*) + h'_\alpha(\phi,\bar{\beta}) - h'_\alpha(\phi,\beta^*) + \hat{h}^N_\alpha(\phi,\beta^*) - \hat{h}^N_\alpha(\phi,\bar{\beta}) \end{aligned}$$

Using Lemma 4 and the fact that $p \le \frac{\epsilon}{\lambda_\alpha(1-\epsilon)}$, we bound:

$$\left|h'_\alpha(\phi,\bar{\beta}) - h'_\alpha(\phi,\beta^*)\right| \le p \cdot \eta\left|\bar{\beta} - \beta^*\right| \le \frac{\epsilon\eta\kappa Z^*}{\lambda_\alpha(1-\epsilon)} \le 2\epsilon\kappa Z^*$$

where the last inequality follows from the facts that $\alpha \in [0,1)$, $\epsilon \in (0,1/2]$ and $\lambda \ge 1$. Similarly, together with Lemma 3, we have:

$$\left|\hat{h}^N_\alpha(\phi,\beta^*) - \hat{h}^N_\alpha(\phi,\bar{\beta})\right| \le \frac{N_h}{N} \cdot \eta\left|\beta^* - \bar{\beta}\right| \le \frac{2\epsilon\eta\kappa Z^*}{\lambda_\alpha} \le 2\epsilon\kappa Z^*$$

Since we have $\hat{g}^N_\alpha(\bar{x},\bar{\beta}) \le \hat{g}^N_\alpha(x^*,\beta^*)$, we conclude that:

$$\begin{aligned} (1-2\epsilon)g'_\alpha(\bar{x},\bar{\beta}) &\le g'_\alpha(\bar{x},\bar{\beta}) - 2\epsilon c(\bar{x}) \\ &\le g'_\alpha(x^*,\beta^*) + 2\epsilon c(x^*) + 4\epsilon\kappa Z^* + 4\epsilon\kappa Z^* \\ &\le (1+10\epsilon\kappa)Z^* \end{aligned}$$

It follows that $g'_\alpha(\bar{x},\bar{\beta}) \le (1 + \Theta(\epsilon\kappa))Z^*$ as desired. □

The next theorem shows that by repeating the sampling sufficiently many times, we can obtain an SAA $\hat{g}^N_\alpha$ whose exact minimizers $(\bar{x}^*,\bar{\beta}^*)$ over $X \times [0,\infty)$ satisfy $\bar{\beta}^* \le (1+\epsilon)Z^*$ with high probability. Due to space limitations, we defer its proof to the full version of the paper.

Theorem 2. *Let* $\alpha \in [0,1)$, $\epsilon \in (0,1/3]$ *and* $\delta \in (0,1/3)$ *be given, and let*

$$k = \Theta\left(\left(1+\frac{1}{\epsilon}\right)\log\frac{1}{\delta}\right); \quad N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4(1-\alpha)^2}\log\left(\frac{\eta}{\epsilon}\cdot|X|\cdot\frac{1}{\delta}\cdot k\right)\right)$$

Consider a collection $\hat{g}_\alpha^{1,N},\dots,\hat{g}_\alpha^{k,N}$ *of independent SAAs of* g'_α*, where each* $\hat{g}_\alpha^{i,N}$ *uses* N *i.i.d. samples of the scenarios. For* $i=1,2,\dots,k$*, let* $(\bar{x}^i,\bar{\beta}^i)$ *be an exact minimizer of* $\hat{g}_\alpha^{i,N}$ *over* $X\times[0,\infty)$*. Set* $v=\arg\min_i \hat{g}_\alpha^{i,N}(\bar{x}^i,\bar{\beta}^i)$*. Then, with probability at least* $1-3\delta$*, the solution* $(\bar{x}^v,\bar{\beta}^v)$ *satisfies* $\bar{\beta}^v \le (1+\epsilon)Z^*$ *and is an* $(1+\Theta(\epsilon))$*–minimizer of* g'_α*.*

Note that in Theorems 1 and 2, we assume that the problem of minimizing $\hat{g}_\alpha^N$ can be solved exactly. In many cases of interest, however, we can only get an approximate minimizer of $\hat{g}_\alpha^N$. The following theorem shows that we can still guarantee a near–optimal solution in this case. Again, its proof can be found in the full version of the paper.

Theorem 3. *Let* $\alpha \in [0,1)$, $\epsilon \in (0,1/3]$ *and* $\delta \in (0,1/5)$ *be given. Let* $k = \Theta((1+\epsilon^{-1})\log\delta^{-1})$, $k' = \Theta((1+\epsilon^{-1})\log(k\delta^{-1}))$*, and set:*

$$N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4(1-\alpha)^2}\log\left(\frac{\eta}{\epsilon}\cdot|X|\cdot\frac{1}{\delta}\cdot kk'\right)\right)$$

Consider a collection $\left\{\hat{g}_\alpha^{(i,j),N}\right\}_{i=1,j=1}^{i=k,j=k'}$ *of independent SAAs of* g'_α*, where each* $\hat{g}_\alpha^{(i,j),N}$ *uses* N *i.i.d. samples of the scenarios. Then, with probability at least* $1-5\delta$*, one can find a pair of indices* (u,v) *such that any* γ*–approximate minimizer of* $\hat{g}_\alpha^{(u,v),N}$ *is an* $(1+\Theta(\epsilon))\gamma$*–minimizer of* g'_α*.*

As we shall see, Theorems 2 and 3 will allow us to obtain efficient approximation algorithms for a large class of risk–adjusted stochastic combinatorial optimization problems under the black–box model. Thus, we are able to generalize the recent results of [2, 4, 14].

4 Applications

In this section, we consider two stochastic combinatorial optimization problems that are special cases of Problem (9) and develop approximation algorithms for them. Due to space limitations, we refer those readers who are interested in more applications to the full version of the paper. Our techniques rely heavily on the following easily–checked properties of φ_α: for random variables $Z_1, Z_2 \in L^2(\Omega,\mathscr{B},\mathbb{P})$ and any $\alpha\in[0,1)$, we have (i) (*Translation Invariance*) $\varphi_\alpha(c+Z_1) = c+\varphi_\alpha(Z_1)$ for any constant c; (ii) (*Positive Homogeneity*) $\varphi_\alpha(cZ_1) = c\varphi_\alpha(Z_1)$ for any constant $c>0$; (iii) (*Monotonicity*) if $Z_1 \le Z_2$ a.e., then $\varphi_\alpha(Z_1)\le\varphi_\alpha(Z_2)$. We shall assume that the cost functions satisfy the properties in Section 3. In view of Theorems 2 and 3, we shall also assume that, for each of the problems under consideration, there is only a polynomial number of scenarios.

4.1 Covering Problems

The 2–stage stochastic set cover problem is defined as follows. We are given a universe U of elements $e_1, \ldots, e_n$ and a collection of subsets of U, say $S_1, \ldots, S_m$. There is a probability distribution over scenarios, and each scenario specifies a subset $A \subset U$ of elements to be covered by the sets $S_1, \ldots, S_m$. Each set S_i has an a priori weight w_i^I and an a posteriori weight w_i^{II}. In the first stage, one selects some of these sets, incurring a cost of w_S^I for choosing set S. Then, a scenario $A \subset U$ is drawn according to the underlying distribution, and additional sets may then be selected, thus incurring their a posteriori costs. Following [14], we formulate the 2–stage problem as follows:

$$\text{minimize} \quad \sum_S w_S^I x_S + \varphi_\alpha(q(x, A)) \quad \text{subject to } x_S \in \{0, 1\} \ \forall\, S$$

where $q(x, A) = \min\{\sum_S w_S^{II} r_{A,S} : r_A \in \mathcal{F}(x, A)\}$, and

$$\mathcal{F}(x, A) = \left\{ r_A : \sum_{S: e \in S} r_{A,S} \geq 1 - \sum_{S: e \in S} x_S \ \ \forall\, e \in A;\ r_{A,S} \in \{0, 1\} \ \forall\, S \right\}$$

By relaxing the binary constraints, we obtain a convex program that can be solved (up to any prescribed accuracy) in polynomial time. Now, using the properties of φ_α and the arguments in [14], one can show the following: if there exists a deterministic algorithm that finds an integer solution whose cost, for each scenario, is at most ρ times the cost of the solution of the relaxed problem, then one can obtain an 2ρ–approximation algorithm for the risk–adjusted stochastic covering problem. In particular, for the risk–adjusted stochastic vertex cover problem, we obtain an 4–approximation algorithm.

4.2 Facility Location Problem

In the 2–stage stochastic facility location problem, we are given a set of facilities F and a set of clients D. Each scenario $A \in \{1, 2, \ldots, N\}$ specifies a subset $D_A \subseteq D$ of clients to be served. The connection cost between client j and facility i is c_{ij}, and we assume that the c_{ij}'s satisfy the triangle inequality. Facility i has a first–stage opening cost of f_i^0 and a recourse cost of f_i^A in scenario A. The goal is to open a subset of facilities in F and assign each client to an open facility. In the full version of the paper, we show how to adapt an algorithm by Shmoys et al. [15] to obtain an 8–approximation algorithm for the risk–adjusted version of this problem.

5 Conclusion and Future Work

In this paper, we have motivated the use of a risk measure to capture robustness in stochastic combinatorial optimization problems. By generalizing the sampling theorem in [2], we have shown that the risk–adjusted objective can be efficiently treated by the SAA method. Furthermore, we have exhibited approximation algorithms for various stochastic combinatorial optimization problems under the risk–adjusted setting. Our work opens up several interesting directions for future research. For instance, it would be interesting to develop approximation algorithms for other stochastic combinatorial optimization problems under our risk–adjusted setting. Also, there are other risk measures that can be used to capture robustness (see [13]). Can theorems similar to those established in this paper be proven for those risk measures?

References

1. E. M. L. Beale, *On Minimizing a Convex Function Subject to Linear Inequalities*, J. Royal Stat. Soc., Ser. B (Methodological) 17(2):173–184, 1955.
2. M. Charikar, C. Chekuri, M. Pál, *Sampling Bounds for Stochastic Optimization*, Proc. 9th RANDOM, pp. 257–269, 2005.
3. G. B. Dantzig, *Linear Programming under Uncertainty*, Manag. Sci. 1(3/4):197–206, 1955.
4. K. Dhamdhere, V. Goyal, R. Ravi, M. Singh, *How to Pay, Come What May: Approximation Algorithms for Demand–Robust Covering Problems*, Proc. 46th FOCS, pp. 367–378, 2005.
5. S. Dye, L. Stougie, A. Tomasgard, *The Stochastic Single Resource Service–Provision Problem*, Naval Research Logistics 50(8):869–887, 2003.
6. A. Gupta, M. Pál, R. Ravi, A. Sinha, *Boosted Sampling: Approximation Algorithms for Stochastic Optimization*, Proc. 36th STOC, pp. 417–426, 2004.
7. A. Gupta, R. Ravi, A. Sinha, *An Edge in Time Saves Nine: LP Rounding Approximation Algorithms for Stochastic Network Design*, Proc. 45th FOCS, pp. 218–227, 2004.
8. N. Immorlica, D. Karger, M. Minkoff, V. Mirrokni, *On the Costs and Benefits of Procrastination: Approximation Algorithms for Stochastic Combinatorial Optimization Problems*, Proc. 15th SODA, pp. 691–700, 2004.
9. A. J. Kleywegt, A. Shapiro, T. Homem–De–Mello, *The Sample Average Approximation Method for Stochastic Discrete Optimization*, SIAM J. Opt. 12(2):479–502, 2001.
10. P. D. Lax, *Functional Analysis*, Wiley–Interscience, 2002.
11. R. T. Rockafellar, S. Uryasev, *Conditional Value–at–Risk for General Loss Distributions*, J. Banking and Finance 26:1443–1471, 2002.
12. R. Ravi, A. Sinha, *Hedging Uncertainty: Approximation Algorithms for Stochastic Optimization Problems*, Proc. 10th IPCO, pp. 101–115, 2004.
13. A. Ruszczyński, A. Shapiro, *Optimization of Risk Measures*, in *Probabilistic and Randomized Methods for Design under Uncertainty* (G. Calafiore and F. Dabbene eds.), Springer–Verlag, 2005.
14. D. B. Shmoys, C. Swamy, *Stochastic Optimization is (Almost) as Easy as Deterministic Optimization*, Proc. 45th FOCS, pp. 228–237, 2004.
15. D. B. Shmoys, É. Tardos and K. I. Aardal, *Approximation Algorithms for Facility Location Problems*, Proc. 29th STOC, pp. 265–274, 1997.