

# Selected Topics in Chance-Constrained Programming

Tara Rengarajan \*

April 03, 2009

## **Abstract**

We consider chance-constrained programs in which the probability distribution of the random parameters is deterministic and known. Two prominent approaches to deal with these programs are sampling approximations and robust approximations. In the last decade, there has been enormous interest in both these areas of research. This article aims to provide a brief summary of a select number of publications pertaining to these approaches that have gained significant attention. We focus mainly on theoretical performance of proposed approximation schemes.

---

\*This report was submitted in partial fulfilment of the requirements for a graduate course in Stochastic Combinatorial Optimization.

# 1 Introduction

Chance-constrained programs continue to be a subject of avid interest in the optimization community for two reasons: their applicability is widespread and their general intractability has motivated the need for “good” approximation solution schemes. By “good”, we mean not only that the approximation scheme in question provide a solution that behaves well with respect to the chance-constrained program but also that this solution be arrived at using reasonable computational effort. Further, the attractiveness of a scheme is determined by the class of problems that it can be applied to. When the program enjoys structural properties that allow for an equivalent deterministic formulation that can be solved either directly or by using well-known approximation schemes such as gradient search or outer linearization methods, this is indeed the recommended approach. However, if the program does not admit sufficient structure that can be exploited or the said approximation schemes do not exhibit good empirical performance, then sampling based approximation schemes are often the designer’s first alternative. Sampling, in the traditional sense, is a way to approximate the distribution in the chance-constrained program. Thus, the approximation is also a chance-constrained program, that uses the sampling distribution instead of the original distribution. Sampling is also sometimes the only viable approach, owing to the vast number of realizations of the random parameters in the problem, easy applicability of sampling in practice and the wealth of theoretical results on sampling such as the strong/weak law of large numbers, central limit theorem and law of the iterated logarithm. Due to the presence of indicator variables, such sampling approximations to the chance-constrained program tend to be intractable integer programs and must be solved using branch-and-bound, cutting plane methods and related schemes. These schemes are designed to get near optimal solutions to the approximation. Hence, if the sampled program approximates the chance-constrained program to within a desired accuracy with high probability, then we can expect to obtain near optimal solutions to the chance-constrained program with high probability.

In contrast to sampling approximations which constitute a randomized approach, robust optimization has emerged as a promising deterministic alternative for certain classes of chance-constrained programs. Robust programs are intractable in general; however, pseudo-polynomial time algorithms can often be designed. The master scheme is to then solve a sequence of progressively restrictive robust approximations using these pseudo-polynomial time algorithms. Such a scheme is expected to guarantee feasibility, and in some cases, optimality to the chance-constrained program after sufficiently many iterations. Tractable robust programs have also been established to exist in many cases; the pioneering work of Ben-Tal and Nemirovski [1] is excellent reference on this point. Apart from its deterministic nature, robust approximation is a fundamentally different paradigm from sampling in that the central idea is one of worst-case optimization. While sampling preserves the chance-constrained notion of “how many” of the realizations of the random parameters are favorable, robust approximation focusses on “which” realizations are favorable - a question of quantity versus quality.

During the last few years, a new notion of approximation to chance-constrained programs that lies between the sampling and robust approximation regimes has emerged. The idea is use the sample to construct a deterministic program, as opposed to a chance-constrained program. Some authors refer to this as a “robust” approximation which is perhaps a little confusing since the approximation, though not a chance-constrained program is still random. The philosophy underlying such an approach is this: The approximating program is no longer an integer program if the chance-constrained program is not itself an integer program; at the same time, increasing the size of the sample allows for the sampling distribution to converge to the true distribution thus maintaining proximity to the chance-constrained program.

In this article, we attempt to provide an overview of robust as well as sampling-based schemes that have attracted significant attention in recent years. Our focus will be on a select list of publications in these areas. We remark that our choice of papers is not to be treated as being wholly representative of research in robust and sampling approximations over their collective time-span; indeed, it is meant to serve

rather as a guideline to recognizing the challenges inherent in chance-constrained programs and designing approximation schemes for the same. The main results in each paper are presented with proofs appearing either in entirety or in the form of an outline or not at all. Where the proof uses ideas that yield technical or conceptual insight we provide complete details; where algebraic manipulation or simple calculus is all that is used, we omit them and refer the reader to the paper in question. In the latter part of the article, we provide results of numerical experiments to analyze the strength of the results presented in the summary; we also show how chance-constrained programs come into play in the context of bicriteria optimization.

We begin by considering a canonical form of a chance-constrained program in Section 2. We then present our summary in two sections: Section 3 is on sampling approximations and Section 4 is devoted to robust approximation schemes. Section 5 summarizes significant results in tabular form. Section ?? constitutes the computational component. Concluding remarks are presented in Section 6.

## 2 The Model

Consider the optimization problem

$$\begin{aligned} \mathcal{PCP}_{\text{gen}} : \quad & \min_{x \in X} h(x) \\ \text{s.t.} \quad & \mathbb{P}(G_i(x, \tilde{\xi}) \in C_i) \geq 1 - \epsilon_i, \quad i = 1, \dots, m, \end{aligned} \quad (1)$$

where  $X$  and  $C_i$ ,  $i = 1, \dots, m$  are well-defined sets and the parameter  $\epsilon_i$  lies in  $(0, 1)$  for each  $i$ .  $\tilde{\xi}$  is a random variable with known distribution  $\mathbb{P}$ . We use the symbol  $\mathcal{PCP}$  to denote “probabilistically-constrained program”, a synonym for chance-constrained program, in keeping with notational convention; the subscript “gen” denotes that the program is written in its most general form. We will also refer to a chance-constrained program as PCP hereafter.  $\mathcal{PCP}_{\text{gen}}$  seeks to minimize an objective  $h(\cdot)$  over a set  $X$  of design variables subject to the set of chance constraints (1). The event  $G_i(x, \tilde{\xi}) \in C_i$  is a desired event, and the *risk level*  $\epsilon_i$  indicates what degree of violation of this event we are willing to allow. The difficulty in solving  $\mathcal{PCP}_{\text{gen}}$  arises mainly from the fact that the set  $X_{\epsilon_i} = \{x \in X : \mathbb{P}(G_i(x, \tilde{\xi}) \in C_i) \geq 1 - \epsilon_i\}$  may not be convex.

**Example.** Consider the following PCP:

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & h(x) \\ \text{s.t.} \quad & \mathbb{P}(\tilde{\xi}|x| \geq b) \geq 1 - \epsilon, \end{aligned} \quad (2)$$

where  $\tilde{\xi}$  obeys a uniform distribution on the unit interval  $(0, 1)$ ,  $b \in \mathbb{R}_+$  and  $\epsilon \in (0, 1)$ . Constraint (2) is violated if  $|x| < b$  and for  $x$  such that  $|x| \geq b$ ,

$$\begin{aligned} \mathbb{P}(\tilde{\xi}|x| \geq b) &= \mathbb{P}\left(\tilde{\xi} \geq \frac{b}{|x|}\right) \\ &= 1 - \frac{b}{|x|}. \end{aligned}$$

The constraint (2) then defines the set  $\{x : |x| \geq \frac{b}{\epsilon}\}$  which is a union of disjoint closed intervals and hence nonconvex.

While it does not follow that nonconvex programs are intractable, the following result due to Luedtke [8] establishes that we do not expect all nonconvex PCP’s to be tractable unless  $P = NP$ :

**Result.**  $\mathcal{PCP}_{\text{gen}}$  is NP-hard in the special case wherein  $G(\cdot, \cdot)$  is linear in  $x$  and  $\tilde{\xi}$ ,  $\tilde{\xi}$  has a discrete uniform distribution and  $C$  is an interval on the real line.

### 3 Sampling Approximations

Sampling approximations are designed on the assumption that it is possible to draw observations from the distribution  $\mathbb{P}$ . Recall that  $\mathbb{P}$  is simply the distribution of the random parameter  $\tilde{\xi}$  and hence every observation corresponds to a realization of this random parameter. These observations are then used in some manner to construct an approximation to  $\mathcal{PCP}_{\text{gen}}$ . If solving the approximation were easy and the PCP hard, then we do not expect substantial results in terms of solution quality. If solving the approximation is again difficult but holds more promise than  $\mathcal{PCP}$ , it may be possible to solve the approximation to near optimality without too much effort and solution quality results of value can be derived. A third line of attack is to draw observations from a distribution that is related to  $\mathbb{P}$  in some fashion, but possesses attractive properties that make the approximation tractable. In this section, we outline some recent work along these lines and compare their relative merits. In particular, we summarize the contributions of Calafiore and Campi [5, 6] and Luedtke [8]. We also touch briefly on the contribution of Nemirovski and Shapiro [9].

A note on computational effort for sampling approximations is in order. While solving the approximation is one aspect of complexity, the size of the sample required to guarantee the quality of the approximation is another equally important consideration. Almost all the work we discuss here has been aimed partly at improving sample size requirements from previous work, and we note the performance of every approximation scheme in this regard as we progress through the literature.

#### 3.1 Overview

For the remainder of this section, we assume that  $m = 1$ ; in other words, the PCP has only one chance constraint. We therefore consider the program

$$\begin{aligned} \mathcal{PCP} : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } \mathbb{P}(G(x, \tilde{\xi}) \in C) \geq 1 - \epsilon. \end{aligned} \tag{3}$$

The standard procedure for designing and evaluating sampling approximation schemes is as follows. Given  $\mathcal{PCP}$  and a number  $\delta \in (0, 1)$ , we

- (i) specify the distribution from which to sample,
- (ii) derive expressions for one or more of the following:
  - (a) a lower bound on the probability that an optimal solution to the approximation is feasible to  $\mathcal{PCP}$
  - (b) a lower bound on the probability that the optimal objective value of the approximation lower bounds the optimal objective value of  $\mathcal{PCP}$

Note that the lower bound in (a) immediately gives a lower bound on the probability that the optimal objective value of the approximation upper bounds the optimal objective value of  $\mathcal{PCP}$ .

- (iii) quantify the sample size required to ensure that the events in (a) and/or (b) above hold at a desired probability level  $1 - \delta$ .

In (a), we refer to the optimal objective value of the approximation as an *upper confidence bound* and in (b) as a *lower confidence bound*. Such bounds are *one-sided confidence bounds* as they bound the optimal objective value of  $\mathcal{PCP}$  in one direction only. It is important to note that the probabilities in (a) and (b) are with respect to the distribution obeyed by the sample. The quantity  $1 - \delta$  is called the *reliability level*. In (a), the reliability level is called a *level of feasibility guarantee*, while in (b), it is called a *level of solution quality*

*guarantee*. The papers we study (and to the best of our knowledge, other work on sampling approximations) provide one or the other or both but do not provide two-sided confidence bounds for the optimal cost of  $\mathcal{PCP}$ . Nemirovski and Shapiro [9] design, subject to certain assumptions, a sampling approximation that provides a lower bound on the probability that the optimal objective value of the approximation is no more than the optimal objective value of the  $\mathcal{PCP}$ -like problem

$$\begin{aligned} \mathcal{PCP}_s : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } \mathbb{P}(\tilde{\xi} \mid G(x, s\tilde{\xi}) \in C) \geq 1 - \epsilon \end{aligned}$$

plus an additive constant. Here,  $s$  designed to be a “small” number strictly larger than 1. Such a guarantee is stronger than a lower confidence bound and usually difficult to derive.

Sampling approximation schemes for chance-constrained programs fall broadly into two categories - *scenario approximation* and *sample average approximation*. In scenario approximation, a finite number of observations  $\tilde{\xi}^1, \dots, \tilde{\xi}^N$  are drawn from either the distribution  $\mathbb{P}$  or a related distribution. The chance constraint (3) is replaced by the set of  $N$  constraints,

$$G(x, \tilde{\xi}^j) \in C, \quad j = 1, \dots, N$$

to get the approximation

$$\begin{aligned} \mathcal{SA} : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } G(x, \tilde{\xi}^j) \in C, \quad j = 1, \dots, N. \end{aligned} \quad (4)$$

An attractive feature of  $\mathcal{SA}$  is that constraints (4) are tractable if  $G(\cdot, \cdot)$  and  $C$  are “nice”. For example, if  $C$  is an interval of the form  $(-\infty, a)$  and  $G(\cdot, \xi)$  is convex for all  $\xi$ , then (4) defines a convex set. As the sample size  $N$  increases, constraints (4) become more restrictive, and  $\mathcal{SA}$  tends towards the robust counterpart of  $\mathcal{PCP}$  in which the event  $G(x, \tilde{\xi}) \in C$  is enforced with probability 1. Therefore, it is intuitive to expect that as  $N$  tends to  $\infty$ , the probability that the feasible set for  $\mathcal{SA}$  is contained in the feasible set for  $\mathcal{PCP}$  tend to 1. In particular, the probability that an optimal solution to  $\mathcal{SA}$  is feasible to  $\mathcal{PCP}$  tends to 1. However, if the distribution sampled from is  $\mathbb{P}$  itself, then (4) is more conservative than (3) and hence a positive optimality gap can be expected to persist in the limit as  $N$  tends to  $\infty$ .

Sample average approximation refers to replacing the distribution  $\mathbb{P}$  with another “easy-to-use” distribution, typically the empirical distribution determined from a sample drawn from  $\mathbb{P}$ . In other words, let  $\tilde{\xi}^1, \dots, \tilde{\xi}^N$  be independent and identically distributed (i.i.d.) observations from  $\mathbb{P}$  and consider the approximation

$$\begin{aligned} \mathcal{SAA} : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } \frac{1}{N} \sum_{j=1}^N \mathbb{I}_{G(x, \tilde{\xi}^j) \in C} \geq 1 - \alpha, \end{aligned} \quad (5)$$

where  $\alpha \in (0, 1)$  and  $\mathbb{I}_A$  denotes the indicator random variable for the event  $A$ . The approximation  $\mathcal{SAA}$  is again a chance-constrained program, hence it is in the same spirit as  $\mathcal{PCP}$ . The risk level  $\alpha$  is usually chosen to be different from  $\epsilon$  so as to provide some slack to compensate for the sampling error. In particular, if  $\alpha < \epsilon$ , then we expect that with high reliability, an optimal solution to  $\mathcal{SAA}$  is feasible to  $\mathcal{PCP}$  in the limit as  $N$  tends to  $\infty$ . On the other hand, if  $\alpha > \epsilon$ , then we expect  $\mathcal{SAA}$  to have a smaller optimal objective value than  $\mathcal{PCP}$  with high reliability in the limit. The obvious drawback in approximation  $\mathcal{SAA}$  is that it could be as difficult as  $\mathcal{PCP}$  to solve.

Choosing between sample average approximations and scenario approximations is as much a matter of taste as one of technical consideration. Some concerns that arise are the following:

1. Does it make sense to solve a PCP approximately using a program that tends to the robust counterpart of PCP as the sample size increases (scenario approximation)? In other words, is the notion of risk level a crucial feature of problem solving?
2. How does the sample size requirement scale with a parameter other than  $\epsilon$  and  $\delta$ ? For instance, how does it vary with the dimension of the problem? With the dimension of the random parameter?
3. How do the various approaches compare in terms of numerical performance? A finite convex program is theoretically efficiently solvable but might take long to solve on a computer when the number of constraints is inhibiting, while a mixed integer program that approximates a chance-constrained program might solve faster with preprocessing, addition of valid inequalities and intelligent branching/fathoming rules.
4. Is the feasibility guarantee true for an optimal solution to the approximation or for all feasible solutions to the approximation? If the latter, is such a strong guarantee necessary? If the former, is it meaningful in contexts where the approximation itself may have to be solved approximately?

The papers we discuss address some of the above points but none of them addresses all of them (see Table 1 for a comparison chart). Furthermore, approximations  $\mathcal{SA}$  and  $\mathcal{SA}_{CC}$  are by no means exhaustive of approximations to  $\mathcal{PCP}$ . The reader can, without difficulty, envision other schemes that might exhibit performance competitive with either of these schemes. For instance, noting that the difficulty in  $\mathcal{PCP}$  arises from the expectation (of the indicator function) being present in the constraints, one might try relaxing the chance-constraint (3) or swapping (3) into the objective and the true objective function into the constraint. This approach is motivated by a bicriteria optimization view of the problem; in other words we would like both  $h(x)$  and  $\mathbb{P}(G(x, \tilde{\xi}) \in C)$  to be small without enforcing that one or the other be smaller than some threshold value.

### 3.2 Scenario Approximation

Calafiore and Campi [5] consider problem  $\mathcal{PCP}$  when  $X$  is convex and closed,  $h(\cdot)$  is convex and (3) is of the form  $\mathbb{P}(f(x, \tilde{\xi}) \leq 0) \geq 1 - \epsilon$ , where  $f(\cdot, \cdot)$  is scalar-valued and  $f(\cdot, \tilde{\xi})$  is convex with probability 1. The observations in the sample are assumed to be i.i.d. Thus, the sample approximation takes the form

$$\begin{aligned} \mathcal{SA}_{CC}^N : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } f(x, \tilde{\xi}^j) \leq 0, \quad j = 1, \dots, N. \end{aligned}$$

For the sake of simplicity, suppose  $\mathcal{SA}_{CC}^N$  is feasible, bounded and attains its optimal value. Assume also that all sampled programs to be considered in this section have a unique optimal solution<sup>1</sup>. Let  $x_N^*$  denote the optimal solution to  $\mathcal{SA}_{CC}$ . Then, the *probability of violation* of  $x_N^*$  is defined as

$$\text{PoV}(x_N^*) = \mathbb{P}(f(x_N^*, \tilde{\xi}) > 0).$$

In fact, the probability of violation may be defined as above for any  $x \in X$ . Since  $x_N^*$  is a random vector,  $\text{PoV}(x_N^*)$  is a random variable. One way to assess the quality of the sampling scheme is to estimate the expected probability of violation  $\mathbb{E}[\text{PoV}(x_N^*)]$ , where the expectation is with respect to the sampling distribution. This quantity measures the average “badness” of the sampling approximation. Another metric is the

---

<sup>1</sup>This assumption can be dispensed with, as shown in Calafiore and Campi [5]. However, this is merely a technical issue that does not affect the results presented herein in any way and we ignore it.

probability that  $x_N^*$  is infeasible to  $\mathcal{PCP}$ , viz.,  $\mathbb{P}(\text{PoV}(x_N^*) > \epsilon)$  with the probability being with respect to the sampling distribution. The latter perhaps captures better the notion of a feasibility guarantee for  $\mathcal{PCP}$ . The main result derived by Calafiore and Campi [5] is the following:

**Theorem 1.** *Let  $\epsilon \in (0, 1)$  and  $\delta \in (0, 1)$  be given and let  $N \geq \frac{n}{\epsilon\delta} - 1$ , where  $n$  is the dimension of the decision vector  $x$  in  $\mathcal{PCP}$ . Then,*

$$\mathbb{P}(\text{PoV}(x_N^*) > \epsilon) \leq \delta.$$

*In other words, with probability  $1 - \delta$ , the sample approximation  $\mathcal{SA}_{CC}^N$  produces an optimal solution that is feasible to  $\mathcal{PCP}$ .*

Theorem 1 provides guidance on how to choose the sample size so as to guarantee feasibility to  $\mathcal{PCP}$  at a desired level. This result and the sampled program  $\mathcal{SA}_{CC}^N$  completely specify a randomized algorithm to produce a solution to  $\mathcal{PCP}$  that is feasible with probability  $1 - \delta$ .

A result stronger than Theorem 1 is also shown to hold. We state this result below.

**Theorem 2.** *The expected probability of violation satisfies*

$$\mathbb{E}[\text{PoV}(x_N^*)] \leq \frac{n}{N+1}. \quad (6)$$

*Therefore, we have*

$$\mathbb{P}(\text{PoV}(x_N^*) > \epsilon) \leq \frac{n}{(N+1)\epsilon}. \quad (7)$$

*Proof.* The expected probability of violation is given by

$$\mathbb{E}[\text{PoV}(x_N^*)] = \mathbb{E}[\mathbb{P}(f(x_N^*, \tilde{\xi}) > 0)]$$

where the expectation is taken with respect to the joint distribution of  $(\tilde{\xi}^1, \dots, \tilde{\xi}^N)$ . We have

$$\mathbb{E}[\mathbb{P}(f(x_N^*, \tilde{\xi}) > 0)] = \mathbb{E}_{(\tilde{\xi}^1, \dots, \tilde{\xi}^N)} \left[ \mathbb{E}_{\tilde{\xi}} \left[ \mathbb{I}_{f(x_N^*, \tilde{\xi}) > 0} \mid (\tilde{\xi}^1, \dots, \tilde{\xi}^N) \right] \right]. \quad (8)$$

Consider

$$\begin{aligned} \mathcal{SA}_{CC}^{N+1} : \quad & \min_{x \in X} h(x) \\ \text{s.t.} \quad & f(x, \tilde{\xi}^j) \leq 0, \quad j = 1, \dots, N+1, \end{aligned}$$

where we use the superscript  $N+1$  to denote the number of observations in the sample. For  $k \in \{1, \dots, N+1\}$ , define  $\mathcal{SA}_{CC}^{N+1,k}$  to be the program obtained from  $\mathcal{SA}_{CC}^{N+1}$  by dropping the  $k^{\text{th}}$  constraint:

$$\begin{aligned} \mathcal{SA}_{CC}^{N+1,k} : \quad & \min_{x \in X} h(x) \\ \text{s.t.} \quad & f(x, \tilde{\xi}^j) \leq 0, \quad j = 1, \dots, N+1, \quad j \neq k. \end{aligned}$$

Let  $x_{N+1,k}^*$  be the optimal solution to  $\mathcal{SA}_{CC}^{N+1,k}$ . Then, using the tower property of the expectation, we have from (8),

$$\mathbb{E}[\text{PoV}(x_N^*)] = \mathbb{E}_{(\tilde{\xi}^1, \dots, \tilde{\xi}^{N+1})} \left[ \mathbb{I}_{f(x_{N+1,N+1}^*, \tilde{\xi}^{N+1}) > 0} \right]$$

Since the observations are i.i.d., we have for any  $k \in \{1, \dots, N + 1\}$ ,

$$\begin{aligned}
\mathbb{E}[\text{PoV}(x_N^*)] &= \mathbb{E}_{(\tilde{\xi}^1, \dots, \tilde{\xi}^{N+1})} \left[ \mathbb{I}_{f(x_{N+1,k}^*, \tilde{\xi}^k) > 0} \right] \\
&= \frac{1}{N+1} \sum_{k=1}^{N+1} \mathbb{E}_{(\tilde{\xi}^1, \dots, \tilde{\xi}^{N+1})} \left[ \mathbb{I}_{f(x_{N+1,k}^*, \tilde{\xi}^k) > 0} \right] \\
&= \frac{1}{N+1} \mathbb{E}_{(\tilde{\xi}^1, \dots, \tilde{\xi}^{N+1})} \left[ \sum_{k=1}^{N+1} \mathbb{I}_{f(x_{N+1,k}^*, \tilde{\xi}^k) > 0} \right]. \tag{9}
\end{aligned}$$

The summation term in (9) counts the number of constraints in  $\mathcal{SA}_{CC}^{N+1}$  that “support” the optimal solution  $x_{N+1}^*$ . In other words, it counts the number of indices  $k \in \{1, \dots, N + 1\}$  such that the optimal solution to  $\mathcal{SA}_{CC}^{N+1,k}$  violates the  $k^{\text{th}}$  constraint in  $\mathcal{SA}_{CC}^{N+1}$ . Such constraints are termed *support constraints* for  $\mathcal{SA}_{CC}^{N+1}$ . It can be established (see Calafiore and Campi [5]) that the number of support constraints for a program with a convex objective and convex constraints is at most the dimension of the decision space. Therefore, we have from (9),

$$\mathbb{E}[\text{PoV}(x_N^*)] \leq \frac{n}{N+1}$$

which is simply (6). The bound (7) follows from (6) by a direct application of Markov’s inequality.  $\square$

Bound (6) is tight in an asymptotic sense, as the sample size  $N$  tends to infinity. A study of the above proof of (6) shows that the inequality arises solely from the following fact: The number of support constraints for  $\mathcal{SA}_{CC}^N$  is at most  $n$ . Now, consider the following chance-constrained program

$$\begin{aligned}
\min_{x \in \mathbb{R}_+^2} \quad & -x - y \\
\text{s.t.} \quad & \mathbb{P}(\tilde{\xi}_1 x + \tilde{\xi}_2 y - \tilde{\xi}_3 \leq 0) \geq 1 - \epsilon, \\
& 0 \leq x \leq 2, \\
& 0 \leq y \leq 2.
\end{aligned}$$

where the random vector  $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3)$  has support  $\{(1, 0, 2 - 1/2^k) : k \in \mathbb{Z}_+\} \cup \{(0, 1, 2 - 1/2^k) : k \in \mathbb{Z}_+\}$ .

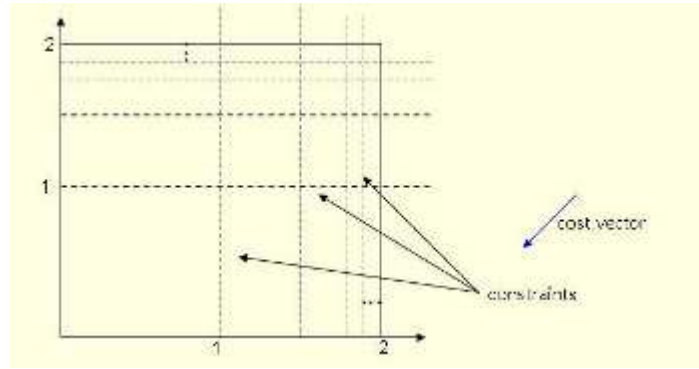


Figure 1: Tightness of bound on expected probability of violation

As  $N$  tends to infinity, the probability that all constraints in  $\mathcal{SA}_{CC}^N$  correspond to lines which are all either vertical or horizontal (see Figure 1) goes to zero. Thus, the probability that there exists at least one vertical and one horizontal constraint in  $\mathcal{SA}_{CC}^N$  goes to 1. But in such a case, the number of support constraints is exactly 2 which is the dimension of the decision space. Thus, the inequality in (6) tends to an equality as  $N \rightarrow \infty$ .

Recall that (7) is derived from (6) by using Markov's inequality and thus cannot be expected to be tight. In fact, a much tighter bound on  $\mathbb{P}(\text{PoV}(x_N^*) > \epsilon)$  is derived by Calafiore and Campi in a later work [6]. The relevant result is reproduced below.

**Theorem 3.** *Let  $N > n$ . The probability that  $x_N^*$  is infeasible to PCP satisfies*

$$\mathbb{P}(\text{PoV}(x_N^*) > \epsilon) \leq \binom{N}{n} (1 - \epsilon)^{N-n}. \quad (10)$$

Further, let  $\delta \in (0, 1)$  be given and let  $N \geq \frac{2}{\epsilon} \ln \frac{1}{\delta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon}$ . Then,

$$\mathbb{P}(\text{PoV}(x_N^*) > \epsilon) \leq \delta. \quad (11)$$

*Outline of Proof.* The central idea in the proof is based on the following claim:

*Claim.* There exists a set of  $n$  distinct indices  $k_1, \dots, k_n \in \{1, \dots, N\}$  such that the optimal solution to the program

$$\begin{aligned} \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & f(x, \xi^{k_j}) \leq 0, \quad j = 1, \dots, n \end{aligned}$$

is identical to  $x_N^*$ .

The claim is established using the classical Helly's theorem from convex analysis which we state below:

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

To see how the claim follows, consider the following two mutually exclusive and exhaustive cases.

*Case 1.* The optimal solution to the program

$$\min_{x \in X} h(x)$$

is identical to  $x_N^*$ . In this case, the claim holds for any choice of  $n$  indices in  $\{1, \dots, N\}$ .

*Case 2.* The optimal solution to the program

$$\min_{x \in X} h(x)$$

is not identical to  $x_N^*$ . Let  $\ell$  be the largest integer less than  $N$  with the following property: There exists a set  $I$  of  $\ell$  distinct indices in  $\{1, \dots, N\}$  such that the optimal solution to the program

$$\begin{aligned} \mathcal{SA}_{CC}^I : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } f(x, \tilde{\xi}^k) \leq 0, \quad k \in I \end{aligned}$$

is not identical to  $x_N^*$ . By hypothesis, there exists such an integer  $\ell$  that is at least 1. Equivalently, there exists a collection  $I$  of  $\ell$  indices such that the optimal objective value of  $\mathcal{SA}_{CC}^I$  is strictly less than the optimal objective value to  $\mathcal{SA}_{CC}^N$ . Let  $\mathcal{I}$  be the set of all collections  $I$  of indices with this property.

Further, let  $z_N^*$  denote the optimal value to  $\mathcal{SA}_{CC}^N$  and  $z_I^*$  denote the optimal value to  $\mathcal{SA}_{CC}^I$  for  $I \in \mathcal{I}$ . Define  $\mathcal{X} = \{x \in X \mid h(x) \leq \max_{I \in \mathcal{I}} z_I^*\}$ ; in other words  $\mathcal{X}$  contains solutions  $x$  in  $X$  that achieve an objective function value of at most  $\max_{I \in \mathcal{I}} z_I^*$ . Define  $X_k$  to be the set of solutions in  $X$  that satisfy the  $k^{\text{th}}$  constraint in  $\mathcal{SA}_{CC}^N$ . We note that  $\mathcal{X}$  and  $X_k$ ,  $k = 1, \dots, N$  are convex sets. Since  $\max_{I \in \mathcal{I}} z_I^* < z_N^*$ , we have that  $\mathcal{X} \cap \bigcap_{k=1}^N X_k$  is empty. Using the contrapositive of Helly's theorem, it follows that there exists a subcollection of  $n + 1$  sets in  $\{\mathcal{X}, X_1, \dots, X_N\}$  that has an empty intersection. Since  $\bigcap_{k=1}^N X_k$  is nonempty, it follows that this subcollection must include  $\mathcal{X}$ . Thus, there exists a set  $J$  of  $n$  distinct indices  $k_1, \dots, k_n$  such that  $\mathcal{X} \cap \bigcap_{k=1}^n X_k$  is empty. Equivalently, there exists a set  $J$  of  $n$  distinct indices  $k_1, \dots, k_n$  such that the optimal objective value of the program

$$\begin{aligned} \mathcal{SA}_{CC}^J : \quad & \min_{x \in X} h(x) \\ & \text{s.t. } f(x, \tilde{\xi}^{k_j}) \leq 0, \quad j = 1, \dots, n \end{aligned}$$

is strictly greater than  $\max_{I \in \mathcal{I}} z_I^*$ . This observation in conjunction with our choice of  $\ell$  implies that the optimal objective value of  $\mathcal{SA}_{CC}^J$  is identical to  $z_N^*$ . In other words, the optimal solutions to  $\mathcal{SA}_{CC}^J$  and  $\mathcal{SA}_{CC}^N$  are identical thus establishing the claim.

The remainder of the proof proceeds via the following argument. For any collection  $I$  of  $n$  distinct indices in  $\{1, \dots, N\}$ , let  $x_I^*$  denote the optimal solution to the program

$$\begin{aligned} & \min_{x \in X} h(x) \\ & \text{s.t. } f(x, \tilde{\xi}^k) \leq 0, \quad k \in I. \end{aligned}$$

Let  $\mathcal{I}$  denote the set of all collections of  $n$  distinct indices in  $\{1, \dots, N\}$ . Then, we have

$$\begin{aligned} \mathbb{P}(\text{PoV}(x_N^*) > \epsilon) &= \mathbb{P}\left(\text{PoV}(x_N^*) > \epsilon \cap \bigcup_{I \in \mathcal{I}} (x_I^* = x_N^*)\right) \quad (\text{using the claim}) \\ &= \mathbb{P}\left(\bigcup_{I \in \mathcal{I}} \left(\text{PoV}(x_N^*) > \epsilon \cap (x_I^* = x_N^*)\right)\right) \\ &= \mathbb{P}\left(\bigcup_{I \in \mathcal{I}} \left(\text{PoV}(x_I^*) > \epsilon \cap (x_I^* = x_N^*)\right)\right) \\ &\leq \sum_{I \in \mathcal{I}} \mathbb{P}\left(\text{PoV}(x_I^*) > \epsilon \cap (x_I^* = x_N^*)\right) \\ &= \sum_{I \in \mathcal{I}} \mathbb{P}\left(\text{PoV}(x_I^*) > \epsilon\right) \mathbb{P}\left(x_I^* = x_N^* \mid \text{PoV}(x_I^*) > \epsilon\right). \end{aligned}$$

The event  $x_I^* = x_N^*$  is simply the event that  $x_I^*$  does not violate any of the constraints in  $\mathcal{SA}_{CC}^N$  corresponding to indices not in  $I$ . Conditioned on the event  $\text{PoV}(x_I^*) > \epsilon$ , the probability that  $x_I^*$  does not violate the  $k^{\text{th}}$

constraint is at most  $1 - \epsilon$  for  $k \notin I$ . Since the constraints in  $\mathcal{SA}_{CC}^N$  are sampled in an i.i.d. manner, we obtain

$$\begin{aligned} \mathbb{P}(\text{PoV}(x_N^*) > \epsilon) &\leq \sum_{I \in \mathcal{I}} \mathbb{P}(\text{PoV}(x_I^*) > \epsilon) (1 - \epsilon)^{N-n} \\ &\leq \sum_{I \in \mathcal{I}} (1 - \epsilon)^{N-n} \\ &= \binom{N}{n} (1 - \epsilon)^{N-n} \end{aligned}$$

which proves (10).

A series of algebraic manipulations and facts from real analysis are then used to show that

$$\binom{N}{n} (1 - \epsilon)^{N-n} \leq \left(\frac{2}{\epsilon}\right)^n e^{-\epsilon(N/2-n)}$$

from which (11) follows. □

A comparison of Theorems 1 and 3 reveals the improvement in the sample size requirement for a feasibility guarantee at a desired level  $\delta$ : While  $N$  scales as  $\frac{1}{\epsilon}$  in Theorem 1, it scales as  $\frac{1}{\epsilon} \ln \frac{1}{\delta}$  in Theorem 3.

A one-sided confidence interval for the optimal objective value of  $\mathcal{PCP}$  can also be output by the scenario approximation scheme. Let  $z_N^*$  and  $z^*$  denote the optimal objective values to problems  $\mathcal{SA}_{CC}^N$  and  $\mathcal{PCP}$  respectively. We then have the following result.

**Theorem 4.** *Let  $\delta \in (0, 1)$  be given. Let  $N \geq \frac{2}{\epsilon} \ln \frac{1}{\delta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon}$ . Then,*

$$\mathbb{P}(z_N^* \geq z^*) \geq 1 - \delta. \tag{12}$$

*If  $\mathcal{PCP}$  is feasible with  $\epsilon = 1 - (1 - \delta)^{1/N}$ , then*

$$\mathbb{P}(z_N^* \leq z^*) \geq 1 - \delta. \tag{13}$$

*Proof.* Result (12) follows immediately from (11) by noting that if  $x_N^*$  is feasible to  $\mathcal{PCP}$ , then  $z_N^* \geq z^*$ . To prove (13), let  $x^*$  be the optimal solution to  $\mathcal{PCP}$ . Then, since  $x^*$  is feasible to  $\mathcal{PCP}$ ,

$$\text{PoV}(x^*) < \epsilon = 1 - (1 - \delta)^{1/N}.$$

Therefore, the probability that  $x^*$  satisfies the  $j^{\text{th}}$  constraint in  $\mathcal{SA}_{CC}^N$  is at least  $(1 - \delta)^{1/N}$ . Using independence of the observations, it follows that the probability that  $x^*$  is feasible to  $\mathcal{SA}_{CC}^N$  is at least  $1 - \delta$ . However, if  $x^*$  is feasible to  $\mathcal{SA}_{CC}^N$ , then  $z_N^* \leq z^*$ . Thus,

$$\mathbb{P}(z_N^* \leq z^*) \geq 1 - \delta,$$

proving (13). □

Since scenario approximation enforces all sampled constraints to hold, a scenario approximation procedure that samples from the distribution  $\mathbb{P}$  cannot be expected to guarantee solution quality. In other words, we do not expect the optimal objective value of the approximation to get “close” to the optimal objective value of  $\mathcal{PCP}$  as the sample size increases. In fact, we expect only an upper confidence bound in the limit as is clear from the non-asymptotic sample size requirement that yields (13). The results of Calafiore and Campi show that under convexity assumptions, we can get very good feasibility guarantees using scenario

approximation. Nemirovski and Shapiro [9] improve on these results in the following manner. Sampling is implemented using a “trial” distribution that is related to  $\mathbb{P}$ . Under the assumption of a bilinear function  $G(\cdot, \cdot)$  and certain “concentration” and “majorization” properties of the trial distribution and an intelligent choice of sampling parameters, they propose an iterative scheme based on the ellipsoid method that provides a feasibility guarantee as well as a guarantee of the form

$$\mathbb{P}(z_N^* \leq z_s^* + \omega) \geq 1 - \delta \quad (14)$$

where  $s \geq 1$  is an “amplification” factor,  $z_s^* \geq z^*$  is the optimal objective value of  $\mathcal{PCP}_s$  (see Section 3.1) and  $\omega > 0$  is the additive approximation factor. As noted earlier, (14) is stronger than a lower confidence bound. To give further details on the working of their scheme warrants a fair amount of setup and we omit further explanation.

### 3.3 Sample Average Approximation

Replacing expectations with a sample average is standard in stochastic programming, particularly in cases wherein the expectation appears in the objective function. However, when the expectation occurs in the constraints and we seek feasibility or optimality guarantees, some flexibility in the risk level is desired. Typically,  $\epsilon$  is perturbed toward a higher or lower value and the direction in which it is perturbed determines the nature of the guarantee. If  $\epsilon$  is increased, we expect a lower confidence bound, if it is decreased, we expect an upper confidence bound. Such guarantees apart, there remains the issue of how to solve the approximation. We do not dwell on this question of how to solve  $\mathcal{SAA}$  efficiently. The reader is referred to Luedtke [8] and the references therein for a comprehensive treatment of the same.

For the remainder of this section, we assume that the chance constraint takes the form  $\mathbb{P}(G(x, \tilde{\xi}) \leq 0) \geq 1 - \epsilon$ . Recall that the approximation  $\mathcal{SAA}$  depends on the risk level  $\alpha$ . Let  $z_\alpha^N$  denote the optimal objective value of  $\mathcal{SAA}$  and  $z^*$  denote the optimal objective value of  $\mathcal{PCP}$ . Also, let  $X_\epsilon$  denote the feasible set for  $\mathcal{PCP}$ . From the work of Luedtke [8], we have the following results:

**Theorem 5.** *Let  $\alpha > \epsilon$ ,  $\delta \in (0, 1)$  be given and let  $N \geq \frac{1}{2(\alpha - \epsilon)^2} \ln \frac{1}{\delta}$ . Then,*

$$\mathbb{P}(z_\alpha^N \leq z^*) \geq 1 - \delta.$$

*Let  $0 \leq \alpha < \epsilon$  and  $\delta \in (0, 1)$  be given. Suppose  $X$  is finite. If  $N \geq \frac{1}{2(\epsilon - \alpha)^2} \ln \frac{|X \setminus X_\epsilon|}{\delta}$ , then*

$$\mathbb{P}(z_\alpha^N \geq z^*) \geq 1 - \delta.$$

The proof is based on Hoeffding’s inequality and ideas similar to those employed in the proof of Theorem 4 and is omitted.

Comparing Theorems 4 and 5, we see that scenario approximation fares better than sample average approximation for the upper confidence bound. This agrees with our intuition somewhat in the sense that scenario approximation could quickly become more stringent than  $\mathcal{PCP}$  (assuming it stays feasible). We note that when  $\alpha = 0$ ,  $\mathcal{SAA}$  reduces to  $\mathcal{SA}$ . Moreover, in this case, the sample size requirement in Theorem 5 for an upper confidence bound can be improved to  $N \geq \ln^{-1} \left( \frac{1}{1 - \epsilon} \right) \ln \frac{|X \setminus X_\epsilon|}{\delta}$ . Using the inequality  $\ln \left( \frac{1}{1 - \epsilon} \right) \geq \epsilon$ , the requirement can be strengthened to  $N \geq \frac{1}{\epsilon} \ln \frac{1}{\delta} + \frac{1}{\epsilon} \ln |X \setminus X_\epsilon|$ . The similarity between this and the corresponding sample size requirement  $N \geq \frac{2}{\epsilon} \ln \frac{1}{\delta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon}$  for scenario approximation is striking. As the author notes, this is surprising given that the assumptions for the two schemes are rather different. If  $|X| \leq U^n$ , the term  $\ln(|X \setminus X_\epsilon|)$  in the sample size requirement can be strengthened to  $n \ln U$ .

The assumption in Theorem 5 that  $X$  is finite is restrictive. The author replaces this with a milder assumption, specifically that  $X$  is bounded and  $G(\cdot, \xi)$  is Lipschitz for every  $\xi$  with a Lipschitz constant that is

independent of  $\xi$ . This leads to a sample size requirement that scales similar to the case  $X$  finite.

Convergence results for sample average approximation require only mild assumptions. Pagnoncelli, Ahmed and Shapiro [10] establish the following result:

**Theorem 6.** *Suppose there exists an optimal solution  $x^*$  to the PCP such that for any  $\beta > 0$ , there exists  $x \in X_\epsilon$  such that  $\|x - x^*\| < \beta$ . Let  $\alpha = \epsilon$ ,  $X$  be compact,  $h(\cdot)$  and  $G(\cdot, \tilde{\epsilon})$  be continuous with probability 1. Let  $S^*$  and  $S_N^*$  denote the set of optimal solutions to PCP and  $\mathcal{SAA}$  respectively. Then,*

$$z_N^* \rightarrow z^* \text{ w.p. 1 as } N \rightarrow \infty$$

and

$$d(S_N^*, S) \rightarrow 0 \text{ w.p. 1 as } N \rightarrow \infty,$$

where we measure the distance between two sets  $A$  and  $B$  as  $d(A, B) = \sup_{x \in A} \inf_{y \in B} \|x - y\|$ .

The proof uses a standard sandwiching argument and the notion of epiconvergence of functions. The steps are straightforward and we refer the interested reader to Pagnoncelli, Ahmed and Shapiro [10].

The strongest assumption in the above result is perhaps the compactness of  $X$ . In many applications however, the decision vector  $x$  is naturally constrained to lie in a compact set, as when it represents an investment or procurement decision. The other assumptions only require that the PCP not be pathological. Thus, we are guaranteed that solving model  $\mathcal{SAA}$  for increasing sample sizes yields the optimal solution and the set of optimizers to the true chance-constrained problem in the limit. Of course, for such a result to hold, the risk level in the approximation must equal that of the PCP.

Since the conclusion of Theorem 6 is far stronger than the confidence bound statements in Theorem 5, one wonders wherein lies the advantage of perturbing the risk level in the sample average approximation, at least when  $X$  is compact. The answer probably lies in that the rate of convergence in Theorem 6 is poor in general and hence the sample size requirement implied by the result for a specified proximity guarantee of the optimal value and set of optimizers of  $\mathcal{SAA}$  to their counterparts in the PCP is far more conservative than that in Theorem 5.

## 4 Robust Approximations

Robust optimization is a distribution-free approach to solving PCPs. The idea here is to ignore the distribution of the random parameters and simply restrict them to lie in an uncertainty set that is contained in the support. The constraints are then enforced for all realizations of the parameters in the uncertainty set. This is often referred to as a robust approximation to a chance-constrained program. It is immediate that if the uncertainty set is large enough and the corresponding robust program is feasible, then any feasible solution to the robust program is feasible to the PCP. A number of authors have proposed such an approach for special cases of chance-constrained programs (see Bertsimas and Sim [2] and Klopfenstein and Nace [7] for instance) with a view to quantifying the extent of feasibility achieved as a function of the size of the uncertainty set as well as lay down conditions under which an optimal solution to the robust program is optimal to the PCP.

Tractability of the robust approximation depends on the properties of the PCP. In Bertsimas and Sim [2], the robust program is a linear program and hence tractable. In Klopfenstein and Nace [7], the robust program is a binary integer program with multiple knapsack constraints. For tractable approximations, the interesting question is how quickly the approximation yields a feasible solution to PCP as the measure of

the uncertainty set is increased. For intractable approximations, the same question arises, as also the design of pseudo-polynomial time algorithms or well-performing numerical schemes for solving the approximation. Just as in the case of scenario approximation, the robust approximation tends to the robust counterpart of the PCP as the measure of the uncertainty set tends to 1. Hence, we do not expect convergence of the optimal objective values to the true optimal objective value in general. However, there do exist special instances for which optimality results can be derived, though not necessarily in the limit. The PCP considered in Klopfenstein and Nace [7] is one such example. In this section, we discuss this work as well as that of Bertsimas and Sim [2]. The reader is also referred to the work of Ben-Tal and Nemirovski (1998) for further details on robust optimization.

Consider the single chance-constrained PCP

$$\begin{aligned} \mathcal{PCP} : \quad & \max_{x \in X} h(x) \\ & \text{s.t.} \quad \mathbb{P}(G(x, \tilde{\xi}) \in C) \geq 1 - \epsilon. \end{aligned}$$

Let  $\Omega$  be the support of  $\tilde{\xi}$  and  $\mathcal{F}$  be the  $\sigma$ -algebra on  $\Omega$  w.r.t. which  $\mathbb{P}$  is defined.  $\mathcal{PCP}$  is equivalent to the following model:

$$\mathcal{PCP}' : \quad \max_{V \in \mathcal{F}: \mathbb{P}(V) \geq 1 - \epsilon} \{ \max_{x \in X} h(x) \text{ s.t. } G(x, \xi) \in C \forall \xi \in V \}.$$

The equivalence of the two models arises from the fact that for  $x \in X$ ,  $x$  is feasible to  $\mathcal{PCP}$  if and only if there exists a  $V \in \mathcal{F}$  such that  $\mathbb{P}(V) \geq 1 - \epsilon$  and  $G(x, \xi) \in C$  for all  $\xi \in V$ . Now, suppose that the set  $V$  is restricted by some condition to lie in some strict subset  $\mathcal{H}$  of  $\mathcal{F}$ . Further, suppose we say that  $x$  is feasible if and only if  $G(x, \xi) \in C \forall \xi \in \cup_{V \in \mathcal{H}} V$ . Then, we get the alternate formulation

$$\begin{aligned} \mathcal{RP} : \quad & \max_{x \in X} h(x) \\ & \text{s.t.} \quad G(x, \xi) \in C, \xi \in V, V \in \mathcal{H}. \end{aligned} \tag{15}$$

Constraint (15) is referred to as a robustness constraint. This terminology arises from the view that the ‘‘uncertain’’ constraint  $G(x, \tilde{\xi}) \in C$  is enforced for *all* realizations of  $\tilde{\xi}$  lying in some uncertainty set, namely,  $\cup_{V \in \mathcal{H}} V$ .  $\mathcal{RP}$  approximates  $\mathcal{PCP}$  in the following sense: If the uncertainty set  $\cup_{V \in \mathcal{H}} V$  has large enough measure and  $\mathcal{RP}$  is feasible, then any  $x$  that is feasible to  $\mathcal{RP}$  is also feasible to  $\mathcal{PCP}$ . The underlying intuition is simple: as the uncertainty set gets larger, its measure increases. When the measure exceeds  $1 - \epsilon$ , we immediately have the result claimed. In particular, any optimal solution to  $\mathcal{RP}$  is feasible to  $\mathcal{PCP}$  for large enough uncertainty sets. Equivalently, the probability of violation of an optimal solution to  $\mathcal{RP}$  diminishes to zero as the measure of the uncertainty set approaches 1. This leads to the question: what is the decay rate of the probability of violation in this limit?

Consider the following linear chance-constrained model:

$$\begin{aligned} \mathcal{PCP}_{\text{lin}} : \quad & \max_x c^T x \\ & \text{s.t.} \quad \mathbb{P}(\tilde{A}x \leq b) \geq 1 - \epsilon, \\ & \quad \quad l \leq x \leq u. \end{aligned}$$

Let  $A = (\tilde{a}_{ij})$ ,  $b = (b_i)$  and  $x$  have dimension  $n$ . To obtain the robust approximation of Bertsimas and Sim [2], we assume that the entries in  $A$  have a nominal value, and that uncertainty arises from perturbing each entry about this value. In particular, suppose that  $\tilde{a}_{ij}$  has support  $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$  where  $a_{ij}$  is a nominal value and the non-negative number  $\hat{a}_{ij}$  is the maximum perturbation. Suppose further that the distribution

of  $\tilde{a}_{ij}$  is symmetric about  $a_{ij}$  and that the  $\tilde{a}_{ij}$ 's are pairwise independent. Define  $I = \{1, \dots, n\}$ . Let  $\Gamma = (\Gamma_i)$  be a vector of nonnegative integers such that  $\Gamma_i \leq n \forall i$ . Consider the following robust approximation of  $\mathcal{PCP}_{\text{lin}}$ :

$$\begin{aligned} \mathcal{RP}^\Gamma : \quad & \max_x c^T x \\ & \text{s.t.} \quad \sum_j a_{ij} x_j + \max_{S_i: S_i \subseteq I, |S_i| = \Gamma_i} \sum_{j \in S_i} \hat{a}_{ij} |x_j| \leq b_i \quad \forall i, \\ & \quad \quad \quad l \leq x \leq u. \end{aligned} \tag{16}$$

The interpretation of the robust constraints (16) in  $\mathcal{RP}^\Gamma$  is that we are allowed to perturb at most  $\Gamma_i$  of the  $n$  random parameters in the  $i^{\text{th}}$  row of  $\tilde{A}$ . It suffices to consider the worst perturbation on the coefficients in each constraint and our assumptions and the linearity of the problem yield (16). As  $\Gamma_i$  increases, the formulation becomes more restrictive since the space of perturbations grows. When  $\Gamma_i = n \forall i$ ,  $\mathcal{RP}^\Gamma$  is simply  $\mathcal{PCP}_{\text{lin}}$  with  $\epsilon = 0$ . The question posed above now reduces to: what is the decay rate of the probability of violation of the optimal solution to  $\mathcal{RP}^\Gamma$  as  $\Gamma \rightarrow (n, \dots, n)$ ? In other words, can we derive a good upper bound on  $\mathbb{P}(\tilde{A}x^\Gamma > b)$  where  $x^\Gamma$  is an optimal solution to  $\mathcal{RP}^\Gamma$ ?

We observe that  $\mathcal{RP}^\Gamma$  is nonlinear owing to the absolute value terms and the inner maximization appearing in the robust constraints. This can be partially linearized by replacing  $|x_j|$  with  $y_j$  and adding the constraints  $-y_j \leq x_j \leq y_j \forall j$  to yield the following model which we represent by the same notation  $\mathcal{RP}^\Gamma$ :

$$\begin{aligned} \mathcal{RP}^\Gamma : \quad & \max_{x,y} c^T x \\ & \text{s.t.} \quad \sum_j a_{ij} x_j + \max_{S_i: S_i \subseteq I, |S_i| = \Gamma_i} \sum_{j \in S_i} \hat{a}_{ij} y_j \leq b_i \quad \forall i, \\ & \quad \quad \quad -y_j \leq x_j \leq y_j \quad \forall j, \\ & \quad \quad \quad l \leq x \leq u, \\ & \quad \quad \quad y_j \geq 0 \quad \forall j. \end{aligned}$$

Bertsimas and Sim establish using a duality argument that  $\mathcal{RP}^\Gamma$  is equivalent to the following linear program:

$$\begin{aligned} \mathcal{RP}_{\text{lin}}^\Gamma : \quad & \max_{x,y,z,p} c^T x \\ & \text{s.t.} \quad \sum_j a_{ij} x_j + z_i \Gamma_i + \sum_{j \in I} p_{ij} \leq b_i \quad \forall i, \\ & \quad \quad \quad z_i + p_{ij} \geq \hat{a}_{ij} y_j \quad \forall i, j, \\ & \quad \quad \quad -y_j \leq x_j \leq y_j \quad \forall j, \\ & \quad \quad \quad l \leq x \leq u, \\ & \quad \quad \quad p_{ij} \geq 0 \quad \forall i, j, \\ & \quad \quad \quad y_j \geq 0 \quad \forall j, \\ & \quad \quad \quad z_i \geq 0 \quad \forall i. \end{aligned}$$

To see this, fix a row  $i$  and the  $y_j$ 's, and consider the problem

$$\mathcal{RP}_i^\Gamma(y) : \quad \max_{S_i: S_i \subseteq I, |S_i| = \Gamma_i} \sum_{j \in S_i} \hat{a}_{ij} y_j.$$

Now, introduce indicator variables  $v_{ij}$  where  $v_{ij}$  is 1 if and only if the index  $j$  belongs to the set (breaking ties between sets arbitrarily if necessary) that is the optimal solution to the above program. In other words,  $\mathcal{RP}_i^\Gamma(y)$  can be rewritten as

$$\begin{aligned} \hat{\mathcal{RP}}_i^\Gamma(y) : \quad & \max_v \sum_{j \in I} \hat{a}_{ij} y_j v_{ij} \\ & \text{s.t.} \quad \sum_{j \in I} v_{ij} \leq \Gamma_i, \\ & \quad \quad v_{ij} \in \{0, 1\} \quad \forall j. \end{aligned}$$

The binary integer constraints in  $\hat{\mathcal{RP}}_i^\Gamma(y)$  can be relaxed to the simple bounds  $0 \leq v_{ij} \leq 1 \quad \forall j$  without changing the optimal value. Thus, we have the equivalent linear model

$$\begin{aligned} \hat{\mathcal{RP}}_i^\Gamma(y) : \quad & \max_v \sum_{j \in I} \hat{a}_{ij} y_j v_{ij} \\ & \text{s.t.} \quad \sum_{j \in I} v_{ij} \leq \Gamma_i, \quad : z_i \\ & \quad \quad 0 \leq v_{ij} \leq 1 \quad \forall j, \quad : p_{ij} \end{aligned}$$

Taking the dual of  $\hat{\mathcal{RP}}_i^\Gamma(y)$  with dual variables as indicated gives

$$\begin{aligned} \mathcal{DRP}_i^\Gamma(y) : \quad & \min_{z,p} z_i \Gamma_i + \sum_{j \in I} p_{ij} \\ & \text{s.t.} \quad z_i + p_{ij} \geq \hat{a}_{ij} y_j \quad \forall j, \\ & \quad \quad p_{ij} \geq 0 \quad \forall j, \\ & \quad \quad z_i \geq 0. \end{aligned}$$

Substituting the above in  $\mathcal{RP}^\Gamma$  for each  $i$  and using strong duality yields the equivalent linear program of Bertsimas and Sim.

The authors then establish the following result to answer the question of decay rate:

**Theorem 7.** *Let  $x^\Gamma$  be an optimal solution to  $\mathcal{RP}^\Gamma$ . Then, the probability of violation of the  $i^{\text{th}}$  constraint, namely  $\mathbb{P}(\sum_j \tilde{a}_{ij} x_j^\Gamma > b_i)$ , satisfies*

$$\text{PoV}_i(x^\Gamma) = \mathbb{P}\left(\sum_{j \in I} \tilde{a}_{ij} x_j^\Gamma > b_i\right) \leq e^{-\frac{\Gamma_i^2}{2n}}. \quad (17)$$

*Outline of Proof.* Let  $S^*$  be the set that attains the maximum in (16) for  $x = x^\Gamma$ . For each  $j \in I$ , define the random variable

$$\tilde{\eta}_{ij} = \begin{cases} \frac{\tilde{a}_{ij} - a_{ij}}{\hat{a}_{ij}} & \text{if } \hat{a}_{ij} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

so that  $\tilde{a}_{ij} = a_{ij} + \hat{a}_{ij}\tilde{\eta}_{ij}$  for all  $j \in I$ . Consider

$$\begin{aligned} \mathbb{P}\left(\sum_{j \in I} \tilde{a}_{ij}x_j^\Gamma > b_i\right) &= \mathbb{P}\left(\sum_{j \in I} (a_{ij}x_j^\Gamma + \hat{a}_{ij}\tilde{\eta}_{ij}x_j^\Gamma) > b_i\right) \\ &= \mathbb{P}\left(\sum_{j \in I} (a_{ij}x_j^\Gamma + \hat{a}_{ij}\tilde{\eta}_{ij}|x_j^\Gamma|) > b_i\right) \end{aligned} \quad (18a)$$

$$\leq \mathbb{P}\left(\sum_{j \in I} (a_{ij}x_j^\Gamma + \hat{a}_{ij}\tilde{\eta}_{ij}|x_j^\Gamma|) > \sum_{j \in I} a_{ij}x_j^\Gamma + \sum_{j \in S^*} \hat{a}_{ij}|x_j^\Gamma|\right) \quad (18b)$$

$$= \mathbb{P}\left(\sum_{j \notin S^*} \hat{a}_{ij}\tilde{\eta}_{ij}|x_j^\Gamma| > \sum_{j \in S^*} \hat{a}_{ij}|x_j^\Gamma|(1 - \tilde{\eta}_{ij})\right) \quad (18c)$$

Equation (18a) follows from the independence and symmetry properties of the  $\tilde{\eta}_{ij}$ 's, and (18b) follows from the definition of  $S^*$ .

Let  $j^* = \operatorname{argmin}_{j \in S^*} \hat{a}_{ij}|x_j^\Gamma|$ . If  $\hat{a}_{ij^*}|x_{j^*}^\Gamma| = 0$ , then  $\hat{a}_{ij}|x_j^\Gamma| = 0$  for all  $j \notin S^*$ . Since  $\tilde{\eta}_{ij} \leq 1$  for all  $j \in I$ , this implies that the probability in (18c) is zero and the result (17) holds. Next, suppose  $\hat{a}_{ij^*}|x_{j^*}^\Gamma| \neq 0$ . Define

$$\gamma_{ij} = \begin{cases} 1 & \text{if } j \in S^*, \\ \frac{\hat{a}_{ij}|x_j^\Gamma|}{\hat{a}_{ij^*}|x_{j^*}^\Gamma|} & \text{if } j \notin S^*. \end{cases}$$

Then, we have from (18c),

$$\mathbb{P}\left(\sum_{j \in I} \tilde{a}_{ij}x_j^\Gamma > b_i\right) \leq \mathbb{P}\left(\sum_{j \notin S^*} \hat{a}_{ij}\tilde{\eta}_{ij}|x_j^\Gamma| > \hat{a}_{ij^*}|x_{j^*}^\Gamma| \sum_{j \in S^*} (1 - \tilde{\eta}_{ij})\right) \quad (19a)$$

$$= \mathbb{P}\left(\sum_{j \notin S^*} \gamma_{ij}\tilde{\eta}_{ij} > \sum_{j \in S^*} (1 - \tilde{\eta}_{ij})\right) \quad (19b)$$

$$= \mathbb{P}\left(\sum_{j \in I} \gamma_{ij}\tilde{\eta}_{ij} > \Gamma_i\right) \quad (19c)$$

$$\leq \mathbb{P}\left(\sum_{j \in I} \gamma_{ij}\tilde{\eta}_{ij} \geq \Gamma_i\right) \quad (19d)$$

The definitions of  $j^*$  and  $\gamma_{ij}$  yield (19a), (19b) and (19c).

Bound (19d) is already indicative of the control wielded by the parameter  $\Gamma_i$ ; the bound decreases as  $\Gamma_i$  is increased. However, it is not a useful bound as it depends on the  $\gamma_{ij}$ 's which in turn depend on  $x^\Gamma$  which is unknown. To derive a bound independent of  $x^\Gamma$ , Markov's inequality can be applied to (19d) in conjunction with some simple facts from real analysis to obtain the bound (17) in the theorem. As these calculations are routine, we refer the reader to Bertsimas and Sim [2] for details.  $\square$

While (17) is independent of  $x^\Gamma$ , it is potentially rather loose owing to the use of Markov's inequality in deriving it. Bertsimas and Sim establish a bound that is tighter with respect to the expression on the right hand side of (19d).

**Theorem 8.** *Let  $x^\Gamma$  be an optimal solution to  $\mathcal{RP}^\Gamma$ . Then, the probability of violation of the  $i^{\text{th}}$  constraint satisfies*

$$\text{PoV}_i(x^\Gamma) \leq \frac{B(n, \Gamma_i)}{16}, \quad (20)$$

where

$$\begin{aligned}
B(n, \Gamma_i) &= \frac{1}{2^n} \left( (1 - \mu) \sum_{l=\lfloor \nu \rfloor}^n \binom{n}{l} + \mu \sum_{l=\lfloor \nu \rfloor + 1}^n \binom{n}{l} \right) \\
&= \frac{1}{2^n} \left( (1 - \mu) \binom{n}{\lfloor \nu \rfloor} + \sum_{l=\lfloor \nu \rfloor + 1}^n \binom{n}{l} \right), \tag{21}
\end{aligned}$$

where  $\nu = \frac{\Gamma_i + n}{2}$  and  $\mu = \nu - \lfloor \nu \rfloor$ .

The proof proceeds via a sequence of algebraic identities and inequalities and is omitted.

We now indicate how the bound in Theorem 7 may be used to ensure the feasibility of  $x^\Gamma$  to  $\mathcal{PCP}_{\text{lin}}$ . Since the constraint coefficients are independent, the probability that the joint chance constraint in  $\mathcal{PCP}_{\text{lin}}$  is satisfied is simply the product of the individual probabilities of satisfying the constraints singly. Suppose  $A$  has  $m$  rows, let  $\delta$  be such that  $(1 - \delta)^m \geq 1 - \epsilon$  and suppose  $\Gamma_i$  satisfies  $\Gamma_i \geq \sqrt{2n \log \frac{1}{\delta}}$ ,  $i = 1, \dots, m$ . Then,

$$\begin{aligned}
\mathbb{P}(\tilde{A}x^\Gamma \leq b) &= \prod_{i=1}^m \mathbb{P}\left(\sum_{j \in I} \tilde{a}_{ij} x_j^\Gamma \leq b_i\right) \\
&= \prod_{i=1}^m (1 - \text{PoV}_i(x^\Gamma)) \\
&\geq \prod_{i=1}^m (1 - e^{-\frac{\Gamma_i^2}{2n}}) \quad (\text{from (17)}) \\
&\geq (1 - \delta)^m \\
&\geq 1 - \epsilon.
\end{aligned}$$

#### 4.1 Robust Approximation to a Chance-constrained Knapsack Problem

Motivated by the approximation in [2], Klopfenstein and Nace consider the chance-constrained knapsack problem

$$\begin{aligned}
\mathcal{PCP}_{\text{knapsack}} : \quad & \max_x \sum_{i=1}^n p_i x_i \\
& \text{s.t.} \quad \mathbb{P}\left(\sum_{i=1}^n \tilde{w}_i x_i \leq c\right) \geq 1 - \epsilon, \\
& x \in \{0, 1\}^n,
\end{aligned}$$

where  $\tilde{w} = (\tilde{w}_1, \dots, \tilde{w}_n)$  is a random vector of weights with support  $W$  and  $p = (p_1, \dots, p_n)$  is a vector of profits. The random variables  $\tilde{w}_i$ ,  $i = 1, \dots, n$  are assumed to be pairwise independent. As in Bertsimas and Sim [2],  $\tilde{w}_i$  is supported on the closed interval  $[\underline{w}_i, \bar{w}_i]$ . Let  $I = \{1, \dots, n\}$ . Analogous to  $\mathcal{RP}^\Gamma$ , we have the robust approximation:

$$\begin{aligned}
\mathcal{RP}_{\text{knapsack}}^\Gamma : \quad & \max_x \sum_{i \in I} p_i x_i \\
& \text{s.t.} \quad \sum_{i \in I} \underline{w}_i x_i + \max_{S: S \subseteq I, |S| = \Gamma} \sum_{j \in S} (\bar{w}_j - \underline{w}_j) x_j \leq c, \\
& x \in \{0, 1\}^n,
\end{aligned} \tag{22}$$

where  $\Gamma$  is an integer between 0 and  $n$ . Bounds analogous to (17) and (20) may be derived for  $\mathcal{RP}_{\text{knapsack}}^\Gamma$ . The paper notes a bound analogous to (17) (see Table 2) under the additional assumption of symmetry of the distribution of the  $\tilde{w}_i$ 's. This is stated as a proposition without proof by the authors and we believe the derivation closely parallels that in Bertsimas and Sim [2].

$\mathcal{RP}_{\text{knapsack}}^\Gamma$  is NP-hard since for  $\Gamma = 0$ , it reduces to the classical knapsack problem. The paper shows how the standard dynamic programming algorithm for a knapsack problem can be adapted to solve  $\mathcal{RP}_{\text{knapsack}}^\Gamma$ . This results in a pseudo-polynomial time algorithm for solving  $\mathcal{RP}_{\text{knapsack}}^\Gamma$ . While this is standard in some sense, the merit of the approximation emerges in the form of Theorem 9 below.

**Theorem 9.** *Suppose that*

- (i) *weights and profits can be sorted such that  $i < j \implies \underline{w}_i \leq \underline{w}_j, p_i \geq p_j$ ,*
- (ii)  *$\bar{w}_i - \underline{w}_i = \delta > 0 \forall i \in I$ ,*
- (iii)  *$\underline{w}_i, i \in I$  and  $c$  are integral multiples of  $\delta$ ,*
- (iv)  *$\tilde{\eta}_i = \frac{\bar{w}_i - \underline{w}_i}{\delta}, i \in I$  are identically distributed.*

*Then, there exists an integer  $\Gamma^* \in \{0, \dots, n\}$  and an optimal solution  $x^*$  to  $\mathcal{PCP}_{\text{knapsack}}$  such that*

- (a)  *$x^*$  is feasible to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$*
- (b) *Any optimal solution to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$  is feasible to  $\mathcal{PCP}_{\text{knapsack}}$ .*

*Proof.* Let  $x^*$  be an optimal solution to  $\mathcal{PCP}_{\text{knapsack}}$  and let  $I(x^*) = \{i \mid x_i^* = 1\}$ . Define

$$\Gamma^* = \min \left( \frac{c - \sum_{i \in I(x^*)} \underline{w}_i}{\delta}, |I(x^*)| \right).$$

From (ii) and (iii), it follows that  $\Gamma^*$  is a nonnegative integer. Further, the definition of  $\Gamma^*$  implies that  $x^*$  is feasible to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$ , thus proving (a).

To prove (b), the following claim is first established:

*Claim. There exists a positive integer  $m \leq n$  such that  $I(x^*) = \{1, \dots, m\}$ .*

Consider the solution  $x'$  specified by

$$x'_i = \begin{cases} 1 & \text{if } 1 \leq i \leq |I(x^*)|, \\ 0 & \text{otherwise.} \end{cases}$$

By (i), we have

$$\sum_{i \in I} p_i x'_i = \sum_{1 \leq i \leq |I(x^*)|} p_i \geq \sum_{i \in I(x^*)} p_i = \sum_{i \in I} p_i x_i^*, \quad \text{and} \quad (23a)$$

$$\sum_{i \in I} \underline{w}_i x'_i = \sum_{1 \leq i \leq |I(x^*)|} \underline{w}_i \leq \sum_{i \in I(x^*)} \underline{w}_i = \sum_{i \in I} \underline{w}_i x_i^*. \quad (23b)$$

Further,

$$\begin{aligned} \mathbb{P}\left(\sum_{i \in I} \tilde{w}_i x'_i \leq c\right) &= \mathbb{P}\left(\sum_{1 \leq i \leq |I(x^*)|} \tilde{w}_i \leq c\right) \\ &= \mathbb{P}\left(\delta \sum_{1 \leq i \leq |I(x^*)|} \tilde{\eta}_i \leq c - \sum_{1 \leq i \leq |I(x^*)|} \underline{w}_i\right) \quad (\text{using } \tilde{w}_i = \underline{w}_i + \delta \tilde{\eta}_i \text{ for each } i \in I) \\ &\geq \mathbb{P}\left(\delta \sum_{1 \leq i \leq |I(x^*)|} \tilde{\eta}_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i\right) \quad (\text{using (23b)}) \end{aligned} \quad (24)$$

$$= \mathbb{P}\left(\delta \sum_{i \in I(x^*)} \tilde{\eta}_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i\right) \quad (\text{using (iv)}) \quad (25)$$

$$= \mathbb{P}\left(\sum_{i \in I(x^*)} \tilde{w}_i \leq c\right)$$

$$= \mathbb{P}\left(\sum_{i \in I} \tilde{w}_i x_i^* \leq c\right)$$

$$\geq 1 - \epsilon.$$

Hence,  $x'$  is feasible to  $\mathcal{PCP}_{\text{knapsack}}$ . This, combined with (23a) implies that  $x'$  is an optimal solution to  $\mathcal{PCP}_{\text{knapsack}}$ , thus proving the claim.

Next, let  $\hat{x}$  be any optimal solution to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$ . Let  $I(\hat{x}) = \{i \mid \hat{x}_i = 1\}$ . If  $\Gamma^* = |I(x^*)|$ , then  $\mathbb{P}(\sum_{i \in I} \tilde{w}_i x_i^* \leq c) = 1$  and therefore,  $\hat{x}$  is feasible to  $\mathcal{PCP}_{\text{knapsack}}$ . If  $\Gamma^* < |I(x^*)|$ , then  $\sum_{i \in I(x^*)} \underline{w}_i = c - \delta \Gamma^*$ . Using that  $\hat{x}$  is feasible to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$ , we get

$$\sum_{i \in I(\hat{x})} \underline{w}_i \leq c - \delta \Gamma^* = \sum_{i \in I(x^*)} \underline{w}_i. \quad (26)$$

Since the  $\underline{w}_i$ 's are ordered in a nondecreasing sequence, (26) implies  $|I(\hat{x})| \leq |I(x^*)|$ . However, the optimality of  $\hat{x}$  to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$  implies  $\sum_{i \in I(\hat{x})} p_i \geq \sum_{i \in I(x^*)} p_i$ . The ordering on the  $p_i$ 's in turn implies  $|I(\hat{x})| \geq |I(x^*)|$ . Combining these two inequalities yields

$$|I(\hat{x})| = |I(x^*)|. \quad (27)$$

Using (26) and (27) and arguing exactly as in the case of  $x'$  earlier in the proof, we obtain that  $\hat{x}$  is feasible to  $\mathcal{PCP}_{\text{knapsack}}$ , proving (b).  $\square$

To see how an algorithm to solve  $\mathcal{PCP}_{\text{knapsack}}$  may be designed, let  $\text{Opt}(\mathcal{PCP}_{\text{knapsack}})$  and  $\text{Opt}(\mathcal{RP}_{\text{knapsack}}^{\Gamma^*})$  denote the optimal objective values to the two programs. Noting that the objective functions of  $\mathcal{PCP}_{\text{knapsack}}$  and  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$  are identical and using (a) from Theorem 9 gives

$$\text{Opt}(\mathcal{PCP}_{\text{knapsack}}) \leq \text{Opt}(\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}) \quad (28)$$

while (b) gives

$$\text{Opt}(\mathcal{PCP}_{\text{knapsack}}) \geq \text{Opt}(\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}). \quad (29)$$

Combining (28) and (29) yields that any optimal solution to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma^*}$  is optimal to  $\mathcal{PCP}_{\text{knapsack}}$ . Next, we note that the feasible region of  $\mathcal{RP}_{\text{knapsack}}^{\Gamma}$  shrinks as  $\Gamma$  increases so that from (a),  $x^*$  is feasible to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma}$  for  $\Gamma \leq \Gamma^*$ . Now, suppose we simply increase  $\Gamma$  in increments of 1 starting at 0 and solve the resulting robust knapsack problem  $\mathcal{RP}_{\text{knapsack}}^{\Gamma}$  to obtain an optimal solution  $\hat{x}^{\Gamma}$ . Then, by (b), in at most  $\Gamma^*$  steps, we are ensured that  $\hat{x}^{\Gamma}$  will be feasible to  $\mathcal{PCP}_{\text{knapsack}}$ . Using that  $x^*$  is feasible to  $\mathcal{RP}_{\text{knapsack}}^{\Gamma}$  for  $\Gamma \leq \Gamma^*$ , we stop when  $\hat{x}^{\Gamma}$  is feasible to  $\mathcal{PCP}_{\text{knapsack}}$  (we note that this condition can be checked if the probability in the chance constraint in  $\mathcal{PCP}_{\text{knapsack}}$  is easy to evaluate) and output  $\hat{x}^{\Gamma}$  as an optimal solution to  $\mathcal{PCP}_{\text{knapsack}}$ .

For completion, we give the dynamic programming algorithm that solves  $\mathcal{RP}_{\text{knapsack}}^{\Gamma}$ . First, suppose that the elements are sorted such that  $i < j \implies \bar{w}_i - \underline{w}_i \leq \bar{w}_j - \underline{w}_j$ . Let  $\text{Opt}(k, \Gamma', b)$  denote the optimal objective value of a robust knapsack problem with only the first  $k$  elements, a knapsack capacity of  $b$  and control parameter  $\Gamma'$ . In other words,  $\text{Opt}(k, \Gamma', b)$  denotes the optimal objective value of  $\mathcal{RP}^{\Gamma'}$  with only the first  $k$  elements and a knapsack capacity of  $b$ . We observe that  $\text{Opt}(\mathcal{RP}^{\Gamma}) = \text{Opt}(n, \Gamma, c)$ . The idea is to set up a forward recursion in  $k$ ,  $\Gamma'$  and  $b$  to obtain  $\text{Opt}(n, \Gamma, c)$  at termination. For  $k = 1$ , we have

$$\text{Opt}(1, 0, b) = \begin{cases} p_1, & \underline{w}_1 \leq b \\ 0 & \text{otherwise} \end{cases}$$

and for  $\Gamma' \geq 1$ ,

$$\text{Opt}(1, \Gamma', b) = \begin{cases} p_1, & \bar{w}_1 \leq b \\ 0 & \text{otherwise.} \end{cases}$$

For  $k > 1$ , we have

$$\text{Opt}(k, \Gamma', b) = \max(\text{Opt}(k-1, \Gamma', b), p_k + \text{Opt}(k-1, \Gamma' - 1, b - \bar{w}_k)). \quad (30)$$

To see (30), we note that the  $k^{\text{th}}$  element either gets picked or does not. The first term inside the maximum is the case when it is not picked. If it is picked, then suffices to consider perturbations in which  $\tilde{w}_k$  is perturbed to its highest value owing to the ordering on the weights. The running time of the algorithm is determined by the number of triples  $(k, \Gamma', b)$  at which  $\text{Opt}(k, \Gamma', b)$  is evaluated in the recursion which is pseudo-polynomial in the length of the input.

Condition (i) in Theorem 9 is perhaps overly restrictive. The proof of Theorem 9 shows that (i) is used only to prove (b). It is not clear if a weaker condition can yield the same conclusion. Condition (ii) and (iv) are used together and seem essential at this point. Condition (iii) however is needed only to ensure the integrality of the control parameter  $\Gamma$  and can be dispensed with in the sense that we can modify the robust approximations to allow  $\Gamma$  to be nonintegral and rework the entire argument for Theorem 9. Redesigning the algorithm is not clear however, since the increment to  $\Gamma$  can no longer be integral. However, from continuity arguments, one can conjecture that there exists an optimal increment value such that a straightforward extension of the algorithm above solves the chance-constrained knapsack within a desired accuracy in a finite number of steps.

## 5 Sampling and Robust Approximations at a Glance

Table 1: Comparison of Sampling Approximations

Reference	Type of Approx.	Feas. Guarantee	Opt. Guarantee	Sample Size Requirement (dominant terms)
Calafiore & Campi [5]	Scenario Approx.	Yes	Upper Bound	$\frac{n}{\epsilon\delta} - 1$
Calafiore & Campi [6]	Scenario Approx.	Yes	Upper Bound	$\frac{2}{\epsilon} \ln \frac{1}{\delta} + 2n + \frac{2n}{\epsilon} \ln \frac{2}{\epsilon}$
Luedtke [8]	Sample Avg. Approx.	Yes	Upper Bound	$\frac{1}{2(\epsilon-\alpha)^2} \ln \frac{ X \setminus X_\epsilon }{\delta}$
Luedtke [8]	Sample Avg. Approx. ( $\alpha = 0$ )	Yes	Upper Bound	$\frac{1}{\epsilon} \ln \frac{ X \setminus X_\epsilon }{\delta}$
Calafiore & Campi [5]	Scenario Approx.	–	Lower Bound	$\frac{\ln(1-\delta)}{\ln(1-\epsilon)}$ (exact)
Luedtke [8]	Sample Avg. Approx.	–	Lower Bound	$\frac{1}{2(\alpha-\epsilon)^2} \ln \frac{ X \setminus X_\epsilon }{\delta}$
Nemirovski & Shapiro [9]	Scenario Approx.	Yes	Upper Bound	See Nemirovski and Shapiro [9]
Nemirovski & Shapiro [9]	Scenario Approx.	Yes	Within constant factor up to additive constant	See Nemirovski and Shapiro [9]

Table 2: Comparison of Robust Approximations

Reference	Tractability of Approx.	Decay Rate of PoV of $i^{\text{th}}$ constraint	Opt. Guarantee	Solving the Approx.
Bertsimas & Sim [2]	Easy (linear program)	$e^{-\frac{\Gamma_i^2}{2n}}$	–	ellipsoid, simplex methods
Bertsimas & Sim [2]	Easy (linear program)	$B(n, \Gamma_i)$ (see (21))	–	ellipsoid, simplex methods
Klopfenstein & Nace [7]	Hard	$e^{-\frac{(2\Gamma-n)^2}{2n}}$	Yes	dynamic programming

## 6 Concluding Remarks

This article has summarized some recent and interesting results related to chance-constrained programs. In particular, we have focussed on sample size requirement results to guarantee lower and upper confidence bounds in the setting of scenario and sample average approximations. When feasibility and not solution quality is priority and computational savings are to be thought of, scenario approximation seems indicated. On the other hand, when the chance-constrained nature of the model is to be retained and solution quality not forsaken for feasibility, sample average approximation is the better option. Further, its convergence properties lend it additional merit when the PCP is to be solved near accurately.

We have also studied robust approximations to chance-constrained programs in special cases when the constraints are linear and the coefficients have independence and/or symmetry properties on their standardized supporting intervals. Robust and scenario approximations may be viewed in the same spirit as they replace feasibility to the PCP with the condition that the constraints hold for all realizations in a “suitably large” set. However, one is a randomized approach while the other is deterministic, hence we may only make probabilistic statements in the randomized approach. However, the robust approach fails to use the distribution in any way in setting up the approximation and some leverage may be lost due to this.

## References

- [1] A. Ben-Tal and A. Nemirovski, Robust Convex Optimization, *Math. Oper. Res.*, 23: 769-805, 1998.
- [2] D. Bertsimas and M. Sim, The Price of Robustness, *Oper. Res.*, 52(1): 35–53, 2004.
- [3] L. Dai, H. Chen and J. R. Birge, Convergence Properties of Two-stage Stochastic Programming, *J. of Opt. Theory and Applications*, 106: 489-509, 2000.
- [4] A. Prékopa, Probabilistic Programming, In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10, Handbook of Oper. Res. Management Sci., 267–351, Elsevier, 2003.
- [5] G. Calafiore and M. C. Campi, Uncertain Convex Programs: Randomized Solutions and Confidence Levels, *Math. Programming: Ser. A*, 102: 25–46, 2005.
- [6] G. C. Calafiore and M. C. Campi, The Scenario Approach to Robust Control Design, *IEEE Transactions on Automatic Control*, 51(5): 742–753, 2006.
- [7] O. Klopfenstein and D. Nace, A Robust Approach to the Chance-constrained Knapsack Problem, *Oper. Res. Letters*, 36(5): 628–632, 2008.
- [8] J. R. Luedtke, *Integer Programming Approaches for some Non-convex and Stochastic Optimization Problems*, PhD Thesis, Georgia Institute of Technology, December, 2007.
- [9] 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*, 3–48, Springer, London, 2005.
- [10] B. Pagnoncelli, S. Ahmed and A. Shapiro, Computational Study of a Chance-constrained Portfolio Selection Problem, *Preprint*, Optimization Online.