A SAMPLE APPROXIMATION APPROACH FOR OPTIMIZATION WITH PROBABILISTIC CONSTRAINTS

JAMES LUEDTKE* AND SHABBIR AHMED†

Abstract. We study approximations of optimization problems with probabilistic constraints in which the original distribution of the underlying random vector is replaced with an empirical distribution obtained from a random sample. We show that such a sample approximation problem with risk level larger than the required risk level will yield a lower bound to the true optimal value with probability approaching one exponentially fast. This leads to an a priori estimate of the sample size required to have high confidence that the sample approximation will yield a lower bound. We then provide conditions under which solving a sample approximation problem with a risk level smaller than the required risk level will yield feasible solutions to the original problem with high probability. Once again, we obtain a priori estimates on the sample size required to obtain high confidence that the sample approximation problem will yield a feasible solution to the original problem. Finally, we present numerical illustrations of how these results can be used to obtain feasible solutions and optimality bounds for optimization problems with probabilistic constraints.

Key words. probabilistic constraints, chance constraints, Monte Carlo, stochastic programming, large deviation

AMS subject classification. 90C15

1. Introduction.

We consider optimization problems with probabilistic constraints (also known as chance constraints) of the form

$$\text{(PCP)} \qquad \min\left\{f(x) : x \in X,\ \Pr\left\{G(x,\xi) \leq \mathbf{0}\right\} \geq 1-\epsilon\right\}$$

where $X \subset \mathbf{R}^n$ represents a deterministic feasible region, $f : \mathbf{R}^n \to \mathbf{R}$ represents the objective to be minimized, ξ is a random vector with suport $\Xi \subseteq \mathbf{R}^d$, $G : \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}^m$ is a given constraint mapping and ϵ is a risk parameter chosen by the decision maker, typically near zero, e.g., $\epsilon = 0.01$ or $\epsilon = 0.05$. Such problems are sometimes called probabilistic programs. In PCP a single probabilistic constraint is enforced over *all* rows in the constraints $G(x,\xi) \leq \mathbf{0}$, rather than requiring that each row independently be satisfied with high probability. Such a constraint is known as a *joint probabilistic constraint*, and is appropriate in a context in which it is important to have all constraints satisfied simultaneously and there may be dependence between random variables in different rows.

Problems with joint probabilistic constraints have been extensively studied; see [25] for background and an extensive list of references. Probabilistic constraints have been used in various applications including supply chain management [17], production planning [21], optimization of chemical processes [13, 14] and surface water quality management [30].

Unfortunately, probabilistic programs are still largely intractable except for a few special cases. There are two primary reasons for this intractability. First, in general, for a given $x \in X$, the quantity $\Pr\left\{G(x,\xi) \leq \mathbf{0}\right\}$ is hard to compute, as it requires multi-dimensional integration, and hence just checking feasibility of a solution is difficult. Second, the feasible region defined by a probabilistic constraint generally

*Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706 (jrluedt1@wisc.edu).

†H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (sahmed@isye.gatech.edu). This research has been supported in part by the National Science Foundation under grants DMI-0133943 and DMI-0522485.

is not convex. In this paper, we study how the difficulty in checking feasibility can be addressed by solving a *sample approximation* problem based on a Monte Carlo sample of ξ. In particular, we study how this approximation can be used to generate feasible solutions and optimality bounds for general probabilistic programs.

The sample approximation we study is a probabilistic program in which the original distribution of the random vector ξ is replaced with the empirical distribution obtained from the random sample. We show that such a sample approximation problem with risk level larger than the nominal risk level ϵ will yield a lower bound to the true optimal value with probability approaching one exponentially fast. This leads to an a priori estimate of the sample size required to have high confidence that the sample approximation will yield a lower bound. We also discuss alternative means of generating lower bounds, which can be used regardless of the sample size used. We then provide conditions under which solving a sample approximation problem with a risk level smaller than ϵ will yield feasible solutions to the original problem with high probability. Once again, we obtain a priori estimates on the sample size required to obtain high confidence that the sample approximation problem will yield a feasible solution to the original problem.

Recently, a number of approaches have been proposed to find approximate solutions to probabilistic programs; the common theme among these is that they all seek "safe" or conservative approximations which can be solved efficiently. That is, they propose approximation problems which are convex and yield solutions which are feasible, or at least highly likely to be feasible, to the original probabilistic program. Approaches of this type include: the scenario approximation method studied by Calafiore and Campi [7, 8] and extended by Nemirovski and Shapiro [22]; the Bernstein approximation scheme of Nemirovski and Shapiro [23]; and robust optimization e.g., [4, 6, 11]. The conservative approximations, when applicable, are attractive because they allow efficient generation of feasible solutions. In particular, they can yield feasible solutions when the probabilistic constraint is "hard," that is, with ϵ very small, such as $\epsilon = 10^{-6}$ or even $\epsilon = 10^{-12}$. However, in a context in which ϵ is not so small, such as $\epsilon = 0.05$ or $\epsilon = 0.01$, the probabilistic constraint is more likely to represent a "soft" constraint, one which the decision-maker would like to have satisfied, but is willing to allow a nontrivial chance that it will be violated if doing so would sufficiently decrease the cost of the implemented solution. In this latter context, it would be desirable to obtain solutions which are feasible to the probabilistic constraint *along with* an assurance that the solutions are not much more costly than the lowest cost solution attaining the same risk level. In this way, the decision-maker can be confident they are choosing from solutions on the efficient frontier between the competing objectives of cost and risk. Unfortunately, the recently proposed conservative approximations say very little in terms of how conservative the solutions are. In particular, it is generally not possible to make a statement about how much worse the objective is relative to the optimal value at a fixed risk level ϵ.

The scenario approximation methods are most similar to the sample approach we study in that they solve an approximation problem based on an independent Monte Carlo sample of the random vector. For example, the scenario approximation of [7, 8] takes a sample $\xi^1, \ldots, \xi^N$ and solves the problem

$$\min_{x \in X} \big\{ f(x) : G(x, \xi^i) \leq \mathbf{0} \quad i = 1, \ldots, N \big\}. \tag{1}$$

That is, the scenario approximation enforces *all* of the constraints corresponding to the sample taken. When the nominal problem is convex (that is, $X \subseteq \mathbf{R}^n$ is a

convex set, f is convex and G is convex in x for each ξ), they show that the scenario approximation problem will yield a feasible solution to PCP with probability at least $1-\delta$ for

$$N \geq \frac{2}{\epsilon}\log\left(\frac{1}{\delta}\right) + 2n + \frac{2n}{\epsilon}\log\left(\frac{2}{\epsilon}\right). \tag{2}$$

In addition, under the stated convexity assumptions, the scenario approximation problem remains a convex program. An advantage of this approach relative to the approximations [4, 6, 11, 23] is that the only assumption that is made on the distribution of ξ is that it can be sampled from.

The key difference between the sample approximation we study and scenario approximation is that we allow the risk level in the sample approximation problem to be positive, that is, we do not require that all sampled constraint sets be satisfied. Instead, the constraint sets which will be satisfied can be chosen optimally. The disadvantage of this scheme is that the sample approximation problem with positive risk level has a non-convex feasible region, and hence may be difficult to solve despite having a simplified probabilistic structure. Specifically, if we allow k of the N sampled constraint sets to be violated, then we must choose a set of k constraint sets which will not be enforced, and there are $\binom{N}{k}$ possible sets to choose from. Choosing the optimal set is an NP-hard problem even in a very special case [20]. However, in some special cases, such as when randomness appears only in the right-hand side of the constraints, the sample approximation problem may be relatively tractable to solve with integer programming techniques, see [20, 19]. In addition, for generating feasible solutions to PCP, our analysis indicates that with appropriately chosen parameters *any* feasible solution to the sample approximation problem will be feasible to the original problem with high probability, so that it is sufficient to generate heuristic solutions. Similarly, to obtain a lower bound for PCP, it is sufficient to obtain a lower bound for the appropriate sample approximation problem.

In the context of generating feasible solutions for PCP, our sample approximation scheme includes as a special case the scenario approximation of [7, 8] in which the constraints corresponding to all sampled vectors ξ^i are enforced. In this special case, we obtain results very similar to those in [8] in terms of how many samples should be used to yield a solution feasible to PCP with high probability. However, our analysis is quite different from the analysis of [8], and in particular, requires a significantly different set of assumptions. In some cases our assumptions are more stringent, but there are also a number of cases in which our assumptions apply and those of [8] do not, most notably if the feasible region X is not convex, as in the case of a mixed-integer program. Thus, our results complement those of [8] in two ways: first we show that sample approximations with positive risk levels can be used to yield feasible solutions to PCP, and second we relax the convexity assumptions. Another closely related work is [9], in which the authors consider a sample approximation problem in which some of the sampled constraints are allowed to be violated. When the nominal problem is convex and a nondegeneracy assumption holds, they present an estimate on the sample size needed to obtain a feasible solution with high probability when a fixed number of sampled constraint sets are discarded optimally. Under these assumptions, their results for generating feasible solutions are very similar to the results we present. The unique contributions of the present paper are: (1) we use assumptions which are significantly different from the convexity and nondegeneracy assumptions used in [9] (neither set of assumptions implies the other), (2) we analyze a method for generating

lower bounds on the optimal value (which is useful for validating the quality of a given solution), (3) we prove that the sample approximation yields an *exact* optimal solution with high probability when X is finite (as in the case of an integer program) and (4) we conduct extensive numerical experiments on practical size problems indicating the potential of the approach.

The sample approximation problem we study can be thought of as a variation of the well studied *Sample Average Approximation* (SAA) approach, see e.g. [1, 10, 16, 29]. The difference is that the approximation we study enforces a sample average constraint involving expectations of indicator functions, whereas the SAA approach typically optimizes a sample average objective. Shapiro [28] and Wang [32] have considered SAA approximation for expected value constraints. However, in these works, the function taken under expectation in the constraints is assumed to be continuous, and hence these results cannot be directly applied because of the discontinuity of indicator functions. In [2] a model with expected value constraints in which the function taken under expectation is not necessarily continuous is considered, and hence their analysis does apply to the case of probabilistic constraints. However, they consider only the case in which the feasible region is finite, and they only discuss the theoretical rate of convergence. In contrast, we begin with a similar analysis for the finite feasible region case, but then extend the analysis to a number of significantly more general settings. In addition, we separate the analysis of when the sample approximation will be likely to yield a lower bound, and when it will be likely to yield feasible solutions. This separate analysis allows for the development of methods which yield optimality statements which hold with high probability.

Finally, we mention the work of Vogel [31], which considers convergence properties of the sample approximation we use for probabilistic programs. When only the right-hand side is random with continuous distribution, it is shown that the probability that the distance between the sample feasible region and true feasible region is larger than any positive threshold decreases exponentially fast with the size of the sample. However, the convergence rate has poor dependence on the dimension of the random vector, implying that the number of samples required to yield a reasonable approximation would have to grow exponentially in this dimension. Better convergence is demonstrated for the case of random right-hand side with discrete distribution. For the general case, linear convergence is demonstrated in the case of continuous distributions. Our analysis of the sample approximation problem extends these results by improving on the convergence rates and by analyzing what happens when the sample approximation problem is allowed to have different risk level than the nominal risk level ϵ. This allows the sample approximation problem to be used to generate feasible solutions and optimality bounds.

The remainder of this paper is organized as follows. In §2 we present and analyze the sample approximation scheme. We present results of a preliminary computational study of the use of the sample approximation scheme in §3. We close with concluding remarks and directions for future research in §4.

2. Analysis of Sample Approximation. We now study how Monte Carlo sampling can be used to generate probabilistically constrained problems with finite distribution which can be used to approximate problems with general distributions. Let us restate PCP as

$$(\mathrm{P}_\epsilon) \qquad z_\epsilon^* = \min\{ f(x) : x \in X_\epsilon \}$$

where

$$X_\epsilon = \Big\{ x \in X : \Pr\big\{G(x,\xi) \le \mathbf{0}\big\} \ge 1-\epsilon \Big\}.$$

We assume z_ϵ^* exists and is finite. For example, if X is compact and $G(x,\xi)$ is affine in x for each $\xi \in \Xi$, then X_ϵ is closed [12] and hence compact, and so if $f(x)$ is continuous then an optimal solution exists whenever $X_\epsilon \neq \emptyset$. Furthermore, we take as an assumption the measurability of any event S taken under probability, such as the event $\{G(x,\xi) \le \mathbf{0}\}$ for each $x \in X$.

If X is a polyhedron, $f(x) = cx$, $G(x,\xi) = \xi - Tx$ $(d = m)$ then we obtain the probabilistically constrained linear program with random right-hand side

$$\min\Big\{ cx : x \in X, \Pr\big\{Tx \ge \xi\big\} \ge 1-\epsilon \Big\}.$$

We can also model a two-stage problem in which we make a decision x and wish to guarantee that with probability at least $1-\epsilon$ there is a feasible recourse decision y satisfying $Wy \ge H(x,\xi)$, where W is an m by l matrix, and $H : \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}^m$. This is accomplished by letting $G : \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}$ be defined by

$$G(x,\xi) = \min_{\mu,y} \{\mu : Wy + \mu\mathbf{e} \ge H(x,\xi), \mu \ge -1\}$$

where $\mathbf{e} \in \mathbf{R}^m$ is a vector of all ones. Indeed, $G(x,\xi) \le \mathbf{0}$ if and only if there exists $y \in \mathbf{R}^l$ and $\mu \le 0$ such that $Wy + \mu\mathbf{e} \ge H(x,\xi)$, which occurs if and only if there exists $y \in \mathbf{R}^l$ such that $Wy \ge H(x,\xi)$.

Due to the general difficulty in calculating $\Pr\big\{G(x,\xi) \le \mathbf{0}\big\}$ for a given $x \in X$, we seek to approximate P_ϵ by solving a sample approximation problem. We let $\xi^1, \ldots, \xi^N$ be an independent Monte Carlo sample of the random vector ξ. Then, for fixed $\alpha \in [0,1)$ the sample approximation problem is defined to be

$$(\mathrm{P}_\alpha^N) \qquad \hat{z}_\alpha^N = \min\big\{ f(x) : x \in X_\alpha^N \big\}$$

where

$$X_\alpha^N = \Big\{ x \in X : \frac{1}{N} \sum_{i=1}^N \mathbb{I}\big(G(x,\xi^i) \le \mathbf{0}\big) \ge 1-\alpha \Big\}$$

where $\mathbb{I}(\cdot)$ is the indicator function which takes values one when $\cdot$ is true and zero otherwise. We adopt the convention that if $X_\alpha^N = \emptyset$ then $\hat{z}_\alpha^N = +\infty$, whereas if P_α^N is unbounded, we take $\hat{z}_\alpha^N = -\infty$. We assume that, except for these two cases, P_α^N has an optimal solution. This assumption is satisfied, for example, if X is compact, $f(x)$ is continuous and $G(x,\xi)$ is continuous in x for each $\xi \in \Xi$, since then X_α^N is the union of finitely many compact sets (in this case $\hat{z}_\alpha^N = -\infty$ is also not possible). If $\alpha = 0$, the sample approximation problem P_0^N corresponds to the *scenario approximation* of probabilistic constraints, studied in [8] and [22]. Our goal is to establish statistical relationships between problems P_ϵ and P_α^N for $\alpha \ge 0$. We first consider when P_α^N yields lower bounds for P_ϵ, then consider when P_α^N yields feasible solutions for P_ϵ.

2.1. Lower Bounds. We now establish a bound on the probability that P_α^N yields a lower bound for P_ϵ. Let

$$\rho(\alpha,\epsilon,N) = \sum_{i=0}^{\lfloor \alpha N \rfloor} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}.$$

$\rho(\alpha, \epsilon, N)$ represents the probability of having at most $\lfloor \alpha N \rfloor$ "successes" in N independent trials, in which the probability of a success in each trial is ϵ.

LEMMA 1. *Assume* P_ϵ *has an optimal solution. Then,*

$$\Pr\{\hat{z}_\alpha^N \leq z_\epsilon^*\} \geq \rho(\alpha, \epsilon, N).$$

Proof. Let $x^* \in X_\epsilon$ be an optimal solution to P_ϵ. Then, $\Pr\{G(x^*, \xi^i) \not\leq \mathbf{0}\} \leq \epsilon$ for each i. Hence, if we call the event $\{G(x^*, \xi^i) \not\leq \mathbf{0}\}$ a success, then the probability of a success in trial i is $\bar{\phi}(x^*) := \Pr\{G(x^*, \xi^i) \not\leq \mathbf{0}\} \leq \epsilon$. By definition of X_α^N, $x^* \in X_\alpha^N$ if and only if

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N \mathbb{I}\big(G(x^*, \xi^i) \leq \mathbf{0}\big) \geq 1 - \alpha \quad &\Leftrightarrow \quad \frac{1}{N}\sum_{i=1}^N \mathbb{I}\big(G(x^*, \xi^i) \not\leq \mathbf{0}\big) \leq \alpha \\ &\Leftrightarrow \quad \sum_{i=1}^N \mathbb{I}\big(G(x^*, \xi^i) \not\leq \mathbf{0}\big) \leq \lfloor \alpha N \rfloor. \end{aligned}$$

Hence, $\Pr\{x^* \in X_\alpha^N\}$ is the probability of having at most $\lfloor \alpha N \rfloor$ successes in N trials. Also, if $x^* \in X_\alpha^N$ then $\hat{z}_\alpha^N \leq z_\epsilon^*$. Thus,

$$\Pr\{\hat{z}_\alpha^N \leq z_\epsilon^*\} \geq \Pr\{x^* \in X_\alpha^N\} = \rho(\alpha, \bar{\phi}(x^*), N) \geq \rho(\alpha, \epsilon, N)$$

since $\rho(\alpha, \epsilon, N)$ is decreasing in ϵ. ∎

For example, if $\alpha = 0$ as in previously studied scenario approximation [8, 22], then we obtain $\Pr\{\hat{z}_\alpha^N \leq z_\epsilon^*\} \geq \rho(0, \epsilon, N) = (1-\epsilon)^N$. For this choice of α, it becomes very unlikely that the sample approximation P_α^N will yield a lower bound as N gets large. For $\alpha > \epsilon$ we see different behavior: the sample approximation yields a lower bound with probability approaching one exponentially fast as N increases. The proof is based on Hoeffding's inequality.

THEOREM 2 (Hoeffding's Inequality [15]). *Let* $Y_1, \ldots, Y_N$ *be independent random variables with* $\Pr\{Y_i \in [a_i, b_i]\} = 1$ *where* $a_i \leq b_i$ *for* $i = 1, \ldots, N$. *Then, if* $t > 0$

$$\Pr\Big\{\sum_{i=1}^N (Y_i - \mathbb{E}[Y_i]) \geq tN\Big\} \leq \exp\Big\{-\frac{2N^2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\Big\}.$$

THEOREM 3. *Let* $\alpha > \epsilon$ *and assume* P_ϵ *has an optimal solution. Then,*

$$\Pr\{\hat{z}_\alpha^N \leq z_\epsilon^*\} \geq 1 - \exp\{-2N(\alpha - \epsilon)^2\}.$$

Proof. Let x^* be an optimal solution to P_ϵ. As in the proof of Lemma 1, if $x^* \in X_\alpha^N$ then $\hat{z}_\alpha^N \leq z_\epsilon^*$. For $i = 1, \ldots, N$ let Y_i be a random variable taking value 1 if $G(x^*, \xi^i) \not\leq \mathbf{0}$ and 0 otherwise. Then, $\Pr\{Y_i \in [0, 1]\} = 1$ and $E[Y_i] \leq \epsilon$. Hence,

$$\begin{aligned} \Pr\{\hat{z}_\alpha^N > z_\epsilon^*\} \leq \Pr\{x^* \notin X_\alpha^N\} &= \Pr\Big\{\frac{1}{N}\sum_{i=1}^N Y_i > \alpha\Big\} \\ &\leq \Pr\Big\{\frac{1}{N}\sum_{i=1}^N (Y_i - E[Y_i]) > \alpha - \epsilon\Big\} \\ &\leq \exp\Big\{-\frac{2N^2(\alpha-\epsilon)^2}{N}\Big\} = \exp\{-2N(\alpha-\epsilon)^2\} \end{aligned}$$

where the first inequality follows since $E[Y_i] \leq \epsilon$ and the second inequality follows from Hoeffding's inequality. □

Theorem 3 states that by taking a risk parameter $\alpha > \epsilon$ in our sample approximation problem, we will obtain a lower bound to the true optimal value with probability approaching one exponentially fast as N increases. Stated another way, suppose we solve a sample approximation problem P_α^N with $\alpha = \epsilon$. Then for any $\gamma > 0$ such that $\gamma < \epsilon$, the optimal value of this problem, $\hat{z}_\epsilon^N$ will be a lower bound to the optimal value of $\mathrm{P}_{\epsilon-\gamma}$ with probability approaching one exponentially fast with N. If γ is small this states that the optimal solution to the sample problem will have cost no worse than any solution that is "slightly less risky" than the nominal risk level ϵ.

Theorem 3 immediately yields a method for generating lower bounds with specified confidence $1 - \delta$, where $\delta \in (0, 1)$. If we select $\alpha > \epsilon$ and

$$N \geq \frac{1}{2(\alpha - \epsilon)^2} \log\Big(\frac{1}{\delta}\Big)$$

then Theorem 3 ensures that $\hat{z}_\alpha^N \leq z_\epsilon^*$ with probability at least $1 - \delta$. Indeed, with this choice of α and N, we have

$$\Pr\big\{\hat{z}_\alpha^N > z_\epsilon^*\big\} \leq \exp\big\{-2N(\alpha - \epsilon)^2\big\} \leq \exp\Big\{-\log\Big(\frac{1}{\delta}\Big)\Big\} = \delta.$$

Because $1/\delta$ is taken under logarithm, we can obtain a lower bound with high confidence, i.e. with δ very small, without increasing the sample size N too large. On the other hand, the required sample size grows quadratically with $1/(\alpha - \epsilon)$ and hence will be large for α very close to ϵ.

Lemma 1 can also be used to obtain lower bounds with specified confidence, using the bounding procedure proposed by Nemirovski and Shapiro [23]. They restrict $\alpha = 0$ in the sample approximation, but the technique can be applied in exactly the same way when $\alpha > 0$, and it is likely this can make the bounding technique significantly more powerful. The idea is as follows. Take M sets of N independent samples of ξ, given by $\xi^{i,j}$ for $j = 1, \ldots, M$ and $i = 1, \ldots, N$ and for each j solve the associated sample approximation problem

$$\hat{z}_{\alpha,j}^N = \min\big\{f(x) : x \in X_{\alpha,j}^N\big\}$$

where

$$X_{\alpha,j}^N = \Big\{x \in X : \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\big(G(x, \xi^{i,j}) \leq \mathbf{0}\big) \geq 1 - \alpha\Big\}.$$

We then rearrange the values $\{\hat{z}_{\alpha,j}^N\}_{j=1}^M$ to obtain the *order statistics* $\hat{z}_{\alpha,[j]}^N$ for $j = 1, \ldots, M$ satisfying $\hat{z}_{\alpha,[1]}^N \leq \cdots \leq \hat{z}_{\alpha,[M]}^N$. Then, a lower bound which is valid with specified confidence $1 - \delta$ can be obtained as follows.

THEOREM 4. *Let $\delta \in (0, 1)$, $\alpha \in [0, 1)$, and N, L and M be positive integers such that $L \leq M$ and*

$$\sum_{i=0}^{L-1} \binom{M}{i} \rho(\alpha, \epsilon, N)^i \big(1 - \rho(\alpha, \epsilon, N)\big)^{M-i} \leq \delta. \tag{3}$$

Then,

$$\Pr\big\{\hat{z}_{\alpha,[L]}^N \leq z_\epsilon^*\big\} \geq 1 - \delta.$$

Proof. We show $\Pr\{\hat{z}^N_{\alpha,[L]} > z^*_\epsilon\} \leq \delta$. Note that $\hat{z}^N_{\alpha,[L]} > z^*_\epsilon$ if and only if less than L of the values $\hat{z}^N_{\alpha,j}$ satisfy $\hat{z}^N_{\alpha,j} \leq z^*_\epsilon$. Thus, calling the event $\{\hat{z}^N_{\alpha,j} \leq z^*_\epsilon\}$ a success, the event $\hat{z}^N_{\alpha,[L]} > z^*_\epsilon$ occurs if and only if there are less than L successes in M trials, in which the probability of a success is $\eta := \Pr\{\hat{z}^N_{\alpha,j} \leq z^*_\epsilon\}$. The result then follows since $\eta \geq \rho(\alpha, \epsilon, N)$ by Lemma 1 and so

$$\sum_{i=0}^{L-1} \binom{M}{i} \eta^i (1-\eta)^{M-i} \leq \sum_{i=0}^{L-1} \binom{M}{i} \rho(\alpha,\epsilon,N)^i \big(1-\rho(\alpha,\epsilon,N)\big)^{M-i} \leq \delta$$

by (3). □

An interesting special case of Theorem 4 is obtained by taking $L = 1$. In this case, we are taking as our lower bound the minimum of the optimal values obtained from solving the M sample approximation problems. To have confidence $1-\delta$ that the lower bound is truly a lower bound, we should choose M such that

$$\big(1-\rho(\alpha,\epsilon,N)\big)^M \leq \delta. \tag{4}$$

With the choice of $L = 1$, let us consider how large M should be with $\alpha = 0$ and with $\alpha = \epsilon$. With $\alpha = 0$, we obtain $\rho(0, \epsilon, N) = (1-\epsilon)^N$. Hence, to have confidence $1-\delta$ to obtain a lower bound, we should take

$$M \geq \log\left(\frac{1}{\delta}\right) / \log\left(\frac{1}{1-(1-\epsilon)^N}\right). \tag{5}$$

Using the inequality $\log(1+x) \leq x$ for $x > 0$ we have

$$\log\left(\frac{1}{1-(1-\epsilon)^N}\right) = \log\left(1 + \frac{(1-\epsilon)^N}{1-(1-\epsilon)^N}\right) \leq \frac{(1-\epsilon)^N}{1-(1-\epsilon)^N}.$$

Hence, when $\alpha = 0$, we should take

$$M \geq \log\left(\frac{1}{\delta}\right) \frac{1-(1-\epsilon)^N}{(1-\epsilon)^N}.$$

Thus, for fixed $\epsilon \in (0,1)$, the required M grows exponentially in N. For example, using (5), if $\delta = 0.001$ and $\epsilon = 0.01$, then for $N = 250$ we need $M \geq 82$, for $N = 500$ we need $M \geq 1,048$, and for $N = 750$ we need $M \geq 12,967$. If $\delta = 0.001$ and $\epsilon = 0.05$, then for $N = 50$ we should take $M \geq 87$, for $N = 100$ we should take $M \geq 1,160$, and for $N = 150$ we must already have $M \geq 15,157$! Thus, to keep M reasonably small, we must keep N small, but this will weaken the lower bound obtained in each sample.

Now suppose we take $L = 1$ and $\alpha = \epsilon$. Then, for N "large enough" (e.g. $N\epsilon \geq 10$), we have $\rho(\epsilon, \epsilon, N) \approx 1/2$. Indeed, $\rho(\epsilon, \epsilon, N)$ is the probability that a binomial random variable with success probability ϵ and N trials is at most $\lfloor \epsilon N \rfloor$. With N large enough relative to ϵ, this probability can be approximated by the probability that a random variable with Normal distribution having mean ϵN does not exceed $\lfloor \epsilon N \rfloor$. Because the median of the normal distribution equals the mean, we obtain $\rho(\epsilon, \epsilon, N) \gtrsim 1/2$. Thus, with $L = 1$ and $\alpha = \epsilon$, we should choose M such that $(1/2)^M \leq \delta$, or

$$M \geq \log_2\left(\frac{1}{\delta}\right).$$

Note that this bound is *independent of* N *and* ϵ. For example, for $\delta = 0.001$, we should take $M \geq 10$. The independence of N has the advantage that we can take N to be as large as is computationally tractable, which will tend to make each of the optimal values $\hat{z}^N_{\epsilon,j}$ closer to the true optimal z^*_ϵ, and hence make the lower bound $\min_j\{\hat{z}^N_{\epsilon,j}\}$ tighter.

We close this section by commenting that although our results have been stated in terms of the *exact* optimal solution $\hat{z}^N_\alpha$ of the sample approximation problem, it is not necessary to calculate this value exactly to use the results. All the results about lower bounds for z^*_ϵ will be valid if $\hat{z}^N_\alpha$ is replaced with a lower bound of $\hat{z}^N_\alpha$, at the expense, of course, of weakening the lower bound.

2.2. Feasible Solutions. We now consider conditions under which an optimal solution to P^N_α, if one exists, is feasible to P_ϵ. The idea is that if we take the risk parameter α in P^N_α to be smaller than ϵ, then for N large enough the feasible region of P^N_α will be a subset of the feasible region of P_ϵ, so that any optimal solution to P^N_α must be feasible to P_ϵ. Unlike the case for lower bounds, we will need to make additional assumptions to assure P^N_α yields a feasible solution with high probability.

We begin by assuming that the feasible region X is finite. Note, however, that $|X|$ may be exponentially large, for example X could be the feasible region of a bounded integer program. We then show how this assumption can be relaxed and replaced with some milder assumptions.

2.2.1. Finite X. THEOREM 5. *Suppose X is finite and $\alpha \in [0, \epsilon)$. Then,*

$$\Pr\{X^N_\alpha \subseteq X_\epsilon\} \geq 1 - |X \setminus X_\epsilon| \exp\{-2N(\epsilon - \alpha)^2\}.$$

Proof. Consider any $x \in X \setminus X_\epsilon$, i.e. $x \in X$ with $\Pr\{G(x,\xi) \leq \mathbf{0}\} < 1 - \epsilon$. We want to estimate the probability that $x \in X^N_\alpha$. For $i = 1, \ldots, N$ define the random variable Y_i by $Y_i = 1$ if $G(x, \xi^i) \leq \mathbf{0}$ and $Y_i = 0$ otherwise. Then, $E[Y_i] = \Pr\{G(x,\xi^i) \leq \mathbf{0}\} < 1 - \epsilon$ and $\Pr\{Y_i \in [0,1]\} = 1$. Observing that $x \in X^N_\alpha$ if and only if $(1/N)\sum_{i=1}^N Y_i \geq 1 - \alpha$ and applying Hoeffding's inequality, we obtain

$$\begin{aligned}\Pr\{x \in X^N_\alpha\} = \Pr\Big\{\frac{1}{N}\sum_{i=1}^N Y_i \geq 1 - \alpha\Big\} &\leq \Pr\Big\{\sum_{i=1}^N (Y_i - E[Y_i]) \geq N(\epsilon - \alpha)\Big\} \\ &\leq \exp\{-2N(\epsilon - \alpha)^2\}.\end{aligned}$$

Then,

$$\begin{aligned}\Pr\{X^N_\alpha \not\subseteq X_\epsilon\} &= \Pr\{\exists x \in X^N_\alpha \text{ s.t. } \Pr\{G(x,\xi) \leq \mathbf{0}\} < 1 - \epsilon\} \\ &\leq \sum_{x \in X \setminus X_\epsilon} \Pr\{x \in X^N_\alpha\} \\ &\leq |X \setminus X_\epsilon| \exp\{-2N(\epsilon - \alpha)^2\}. \qquad \square\end{aligned}$$

For fixed $\alpha < \epsilon$ and $\delta \in (0,1)$, Theorem 5 shows that if we take

$$N \geq \frac{1}{2(\epsilon - \alpha)^2} \log\left(\frac{|X \setminus X_\epsilon|}{\delta}\right)$$

then if P^N_α is feasible, it will yield a feasible solution to P_ϵ with probability at least $1 - \delta$. If $|X| \leq U^n$, we can take

$$N \geq \frac{1}{2(\epsilon - \alpha)^2} \log\left(\frac{1}{\delta}\right) + \frac{n}{2(\epsilon - \alpha)^2} \log(U). \tag{6}$$

Note that N grows linearly with the dimension n of the feasible region, and logarithmically with $1/\delta$, so that the confidence of generating a feasible solution can be made large without requiring N to be too large. However, the quadratic dependence on $\epsilon - \alpha$ implies that this a priori estimate of how large N should be will grow quite large for α near ϵ.

Theorem 5 states that for $\alpha < \epsilon$, *every* feasible solution to the sample approximation problem will be feasible to the original problem with risk level ϵ with high probability as N gets large. This is in contrast to the results of scenario approximation presented in [8] in which $\alpha = 0.0$ is required, and the result is that the *optimal* solution to the sample approximation problem will be feasible to the original problem with high probability. The advantage of our approach is that one need not solve the sample approximation problem to optimality to obtain a solution to the original problem. Simple heuristics which select which sampled constraints to be satisfied, e.g. greedily or by local search, can be used to yield feasible solutions for the approximation problem, which by virtue of Theorem 5 will have high probability of being feasible to the original problem. This comment also applies to subsequent feasibility results in which we relax the assumption that the feasible region X is finite.

In this case of finite X, we can combine Theorem 5 with Theorem 3 to demonstrate that solving a sample approximation with $\alpha = \epsilon$ will yield an exact optimal solution with probability approaching one exponentially fast with N. Let X_ϵ^* be the set of optimal solutions to P_ϵ and define $\underline{\alpha} = \max\{\Pr\{G(x,\xi) \not\leq \mathbf{0}\} : x \in X_\epsilon^*\}$. By definition, we have $z_{\underline{\alpha}}^* = z_\epsilon^*$. Next, let $\overline{\alpha} = \min\{\Pr\{G(x,\xi) \not\leq \mathbf{0}\} : x \in X \setminus X_\epsilon\}$. By definition, we have $\overline{\alpha} > \epsilon$. Finally, define $\kappa = \min\{\epsilon - \underline{\alpha}, \overline{\alpha} - \epsilon\}$.

COROLLARY 6. *Assume $\underline{\alpha} < \epsilon$. Then,*

$$\Pr\{\hat{z}_\epsilon^N = z_\epsilon^*\} \geq 1 - (|X| + 1)\exp\{-2N\kappa^2\}.$$

Proof. First observe that $\kappa > 0$ when $\underline{\alpha} < \epsilon$. Next, we apply Theorem 3 with $\underline{\alpha}$ in place of ϵ and ϵ in place of α to obtain $\Pr\{\hat{z}_\epsilon^N \leq z_{\underline{\alpha}}^*\} \geq 1 - \exp\{-2N(\epsilon - \underline{\alpha})^2\}$. Because $z_{\underline{\alpha}}^* = z_\epsilon^*$ this implies $\Pr\{\hat{z}_\epsilon^N > z_\epsilon^*\} \leq \exp\{-2N(\epsilon - \underline{\alpha})^2\}$.

We next observe that the proof of Theorem 5 can be modified to show the slightly stronger result that

$$\Pr\{X_\alpha^N \subseteq X_\epsilon'\} \geq 1 - |X \setminus X_\epsilon'|\exp\{-2N(\epsilon - \alpha)^2\}$$

where $X_\epsilon' = \{x \in X : \Pr\{G(x,\xi) \leq \mathbf{0}\} > 1 - \epsilon\}$. (In the proof, we consider each $x \in X \setminus X_\epsilon'$ and observe that the defined random variable Y_i satisfies $E[Y_i] \leq 1 - \epsilon$. The remainder of the proof is identical with X_ϵ replaced by X_ϵ'.) Applying this result, we obtain

$$\Pr\{X_\epsilon^N \subseteq X_{\overline{\alpha}}'\} \geq 1 - |X \setminus X_{\overline{\alpha}}'|\exp\{-2N(\overline{\alpha} - \epsilon)^2\}.$$

However, if $x \in X_{\overline{\alpha}}'$ then $\Pr\{G(x,\xi) \not\leq \mathbf{0}\} < \overline{\alpha}$ and by definition of $\overline{\alpha}$ this implies $\Pr\{G(x,\xi) \not\leq \mathbf{0}\} \leq \epsilon$ and thus $X_{\overline{\alpha}}' \subseteq X_\epsilon$. It follows that

$$\Pr\{\hat{z}_\epsilon^N < z_\epsilon^*\} \leq \Pr\{X_\epsilon^N \not\subseteq X_\epsilon\} \leq |X|\exp\{-2N(\overline{\alpha} - \epsilon)^2\}.$$

Therefore,

$$\begin{aligned}\Pr\{\hat{z}_\epsilon^N \neq z_\epsilon^*\} &\leq \Pr\{\hat{z}_\epsilon^N > z_\epsilon^*\} + \Pr\{\hat{z}_\epsilon^N < z_\epsilon^*\} \\ &\leq \exp\{-2N(\epsilon - \underline{\alpha})^2\} + |X|\exp\{-2N(\overline{\alpha} - \epsilon)^2\} \\ &\leq (1 + |X|)\exp\{-2N\kappa^2\}. \qquad \square\end{aligned}$$

The assumption that $\underline{\alpha} < \epsilon$ is mild since, because X is finite, there are only finitely many values of $\epsilon \in [0,1]$ for which it is possible to have $\epsilon = \underline{\alpha}$. Stated another way, if we add a random perturbation uniformly distributed in $(-\gamma, \gamma)$ to ϵ, where γ can be arbitrarily small, then the assumption will hold with probability one. On the other hand, the number of scenarios required to guarantee reasonably high probability of obtaining the optimal solution will be at least proportional to $(\epsilon - \underline{\alpha})^{-2}$, and hence may be very large. Thus, Corollary 6 illustrates the *qualitative* behavior of the sample approximation with $\alpha = \epsilon$ in the finite feasible region case but may not be useful for estimating the required sample size.

If we take $\alpha = 0$ in Theorem 5, we obtain improved dependence of N on ϵ.

THEOREM 7. *Suppose X is finite and $\alpha = 0$. Then,*

$$\Pr\{X_0^N \subseteq X_\epsilon\} \geq 1 - |X \setminus X_\epsilon|(1-\epsilon)^N.$$

Proof. With $\alpha = 0$, if $x \in X$ satisfies $\Pr\{G(x,\xi) \leq \mathbf{0}\} < 1 - \epsilon$, then $x \in X_0^N$ if and only if $G(x,\xi^i) \leq \mathbf{0}$ for each $i = 1, \ldots, N$, and hence $\Pr\{x \in X_0^N\} < (1-\epsilon)^N$. The claim then follows just as in the proof of Theorem 5. □

When $\alpha = 0$, to obtain confidence $1 - \delta$ that P_α^N will yield a feasible solution to P_ϵ whenever P_α^N is feasible, we should take

$$N \geq \log^{-1}\left(\frac{1}{1-\epsilon}\right) \log\left(\frac{|X \setminus X_\epsilon|}{\delta}\right).$$

If $|X| \leq U^n$, then it is sufficient to take

$$N \geq \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right) + \frac{n}{\epsilon} \log U \tag{7}$$

where we have used the inequality $\log(1/(1-\epsilon)) \geq \epsilon$. Hence, with $\alpha = 0$, the required sample size again grows linearly in n, but now also linearly with $1/\epsilon$. Note the similarity between the bound (7) and the bound of Campi and Calafiore [8],

$$N \geq \frac{2}{\epsilon} \log\left(\frac{1}{\delta}\right) + 2n + \frac{2n}{\epsilon} \log\left(\frac{2}{\epsilon}\right)$$

which also exhibits linear dependence in n and (nearly) linear dependence in $1/\epsilon$. This is interesting considering the significantly different assumptions used for the analysis. In [8] it is assumed that X is a convex set and $G(x,\xi)$ is a convex function of x for every possibly value of ξ. In contrast, we make the strong assumption that X is finite, but require no other assumptions on the form of the random constraint $G(x,\xi) \leq \mathbf{0}$.

2.2.2. Random right-hand side. We now show how the assumption that X is finite can be relaxed when the probabilistic constraint involves randomness only in the right-hand side. Thus, in this section we assume $G(x,\xi) = \xi - g(x)$ where $g : \mathbf{R}^n \to \mathbf{R}^m$, and $\Xi \subseteq \mathbf{R}^m$. Let the cumulative distribution function of ξ be $F(y) = \Pr\{\xi \leq y\}$ for $y \in \mathbf{R}^m$. Then, the feasible region of the probabilistically constrained problem with random right-hand side is

$$\bar{X}_\epsilon = \Big\{x \in X : F(g(x)) \geq 1 - \epsilon\Big\}.$$

The feasible region of the sample approximation problem for $\alpha \in [0,1)$ is

$$\bar{X}_\alpha^N = \Big\{ x \in X : \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\big(g(x) \geq \xi^i\big) \geq 1 - \alpha \Big\}.$$

We first consider the case that ξ has a finite distribution, that is, $\Xi = \{\xi^1, \ldots, \xi^K\}$. Note that K may be very large, for example $K = U^m$ for a positive integer U. Next, for $j = 1, \ldots, m$ define $\Xi_j = \{\xi_j^k : k = 1, \ldots, K\}$ and finally let $C = \prod_{j=1}^m \Xi_j$.

THEOREM 8. *Suppose ξ has a finite distribution and let $\alpha \in [0, \epsilon)$. Then,*

$$\Pr\big\{ \bar{X}_\alpha^N \subseteq \bar{X}_\epsilon \big\} \geq 1 - |C| \exp\big\{ -2N(\epsilon - \alpha)^2 \big\}.$$

Proof. Let $C_\epsilon = \{y \in C : F(y) \geq 1 - \epsilon\}$ and

$$C_\alpha^N = \Big\{ y \in C : \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\big(y \geq \xi^i\big) \geq 1 - \alpha \Big\}.$$

Because C is a finite set, we can apply Theorem 5 to obtain

$$\Pr\big\{ C_\alpha^N \subseteq C_\epsilon \big\} \geq 1 - |C| \exp\big\{ -2N(\epsilon - \alpha)^2 \big\}. \tag{8}$$

Now, let $x \in \bar{X}_\alpha^N$, so that $x \in X$ and $\sum_{i=1}^N \mathbb{I}\big(g(x) \geq \xi^i\big) \geq N(1-\alpha)$. Define $\bar{y} \in C$ by

$$\bar{y}_j = \max\{y_j \in \Xi_j : y_j \leq g_j(x)\} \qquad j = 1, \ldots, m$$

so that by definition, $\bar{y} \leq g(x)$. Next, note that if $g(x) \geq \xi^i$ for some i, then also $\bar{y} \geq \xi^i$ since $\xi^i \in C$. Hence, $\sum_{i=1}^N \mathbb{I}\big(\bar{y} \geq \xi^i\big) \geq N(1-\alpha)$ and so $\bar{y} \in C_\alpha^N$. Hence, when $C_\alpha^N \subseteq C_\epsilon$, $F(\bar{y}) \geq 1 - \epsilon$ and because $\bar{y} \leq g(x)$, also $F(g(x)) \geq 1 - \epsilon$ and so $x \in \bar{X}_\epsilon$. Since $x \in \bar{X}_\alpha^N$ was arbitrary, this shows that when $C_\alpha^N \subseteq C_\epsilon$, $\bar{X}_\alpha^N \subseteq \bar{X}_\epsilon$ and the result follows from (8). ∎

If, for example, $|\Xi_j| \leq U$ for each j, then $|C| \leq U^m$ so to obtain confidence $1 - \delta$ that $\bar{X}_\alpha^N \subseteq \bar{X}_\epsilon$ it is sufficient to take

$$N \geq \frac{1}{2(\epsilon - \alpha)^2} \log\left(\frac{1}{\delta}\right) + \frac{m}{2(\epsilon - \alpha)^2} \log U. \tag{9}$$

The difference between this bound and (6) is that (9) depends linearly on m, the dimension of ξ, whereas (6) depends linearly on n, the dimension of x.

Similar to the case of finite feasible region X, when ξ has a finite distribution, it can be shown that the sample approximation problem with $\epsilon = \alpha$ will yield an exact optimal solution with probability approaching one as N increases. The statement and proof of this result is completely analogous to that of Corollary 6 and is omitted for the sake of brevity.

As in the case of Theorem 7, if we take $\alpha = 0$, we can obtain the stronger convergence result

$$\Pr\big\{ \bar{X}_0^N \subseteq \bar{X}_\epsilon \big\} \geq 1 - |C|(1 - \epsilon)^N.$$

The assumption in Theorem 8 that Ξ is finite can be relaxed if we assume $\bar{X}_\epsilon \subseteq \bar{X}(l,u) := \{x \in X : l \leq g(x) \leq u\}$ for some $l, u \in \mathbf{R}^m$. This assumption is not very strict. Indeed, if we define $l \in \mathbf{R}^m$ by

$$l_j = \min\{l \in \mathbf{R} : F_j(l) \geq 1 - \epsilon\}$$

where F_j is the marginal distribution of ξ_j for $j = 1, \ldots, m$ then $g(x) \geq l$ for any $x \in \bar{X}_\epsilon$. This holds because if $g_j(x) < l_j$ for some j, then $\Pr\{g(x) \geq \xi\} \leq \Pr\{g_j(x) \geq \xi_j\} = F_j(g_j(x)) < 1 - \epsilon$ by definition of l_j and hence $x \notin \bar{X}_\epsilon$. Furthermore, if X is compact and $g(x)$ is continuous in x, then if we define $u \in \mathbf{R}^m$ by

$$u_j = \max\{g_j(x) : x \in X\} \qquad j = 1, \ldots, m$$

then each u_j is finite, and by definition, $g(x) \leq u$ for any $x \in \bar{X}$. Under the assumption that $\bar{X}_\epsilon \subseteq \bar{X}(l, u)$ the assumption that Ξ is finite can be replaced by the assumption that $\Xi \cap \{y \in \mathbf{R}^m : l \leq y \leq u\}$ is finite, leading to a result similar to Theorem 8, with a nearly identical proof.

Alternatively, when $\bar{X}_\epsilon \subseteq \bar{X}(l, u)$, we can obtain a similar result if ξ has a Lipschitz continuous cumulative distribution function F on $[l, u] = \{y \in \mathbf{R}^m : l \leq y \leq u\}$. That is, we assume there exists $L > 0$ such that

$$|F(y) - F(y')| \leq L\|y - y'\|_\infty \qquad \forall y, y' \in [l, u]$$

where $\|y\|_\infty = \max\{|y_j| : j = 1, \ldots, m\}$. Under the assumption that $\bar{X}_\epsilon \subseteq \bar{X}(l, u)$ we add the constraints $l \leq g(x) \leq u$ to the sample approximation problem to obtain

$$\bar{X}^N_\alpha(l, u) = \Big\{x \in \bar{X}(l, u) : \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\big(g(x) \geq \xi^i\big) \geq 1 - \alpha\Big\}.$$

We define $D = \max\{u_j - l_j : j = 1, \ldots, m\}$. Then we have

THEOREM 9. *Suppose $\bar{X}_\epsilon \subseteq \bar{X}(l, u)$ and F is Lipschitz continuous with constant L. Let $\alpha \in [0, \epsilon)$ and $\beta \in (0, \epsilon - \alpha)$. Then,*

$$\Pr\big\{\bar{X}^N_\alpha(l, u) \subseteq \bar{X}_\epsilon\big\} \geq 1 - \lceil DL/\beta \rceil^m \exp\big\{-2N(\epsilon - \alpha - \beta)^2\big\}.$$

Proof. Let $K = \lceil DL/\beta \rceil$ and define $Y_j = \{l_j + (u_j - l_j)i/K : i = 1, \ldots, K\}$ for $j = 1, \ldots, m$ and $Y = \prod_{j=1}^m Y_j$, so that $|Y| = K^m$ and that for any $y \in [l, u]$ there exists $y' \in Y$ such that $y' \geq y$ and $\|y - y'\|_\infty \leq \beta/L$. Indeed, for a given $y \in [l, u]$ such a y' can be obtained by letting

$$y'_j = \min\{w \in Y_j : w \geq y_j\} \qquad j = 1, \ldots, m.$$

With this definition of y', we have $y' \geq y$ and

$$|y'_j - y_j| = y'_j - y_j \leq (u_j - l_j)/K \leq D/K \leq \beta/L \qquad j = 1, \ldots, m.$$

Next, let $Y_{\epsilon-\beta} = \{y \in Y : F(y) \geq 1 - \epsilon + \beta\}$ and

$$Y^N_\alpha = \Big\{y \in Y : \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\big(y \geq \xi^i\big) \geq 1 - \alpha\Big\}. \tag{10}$$

Since Y is finite and $\alpha < \epsilon - \beta$, we can apply Theorem 5 to obtain

$$\Pr\big\{Y^N_\alpha \subseteq Y_{\epsilon-\beta}\big\} \geq 1 - |Y| \exp\big\{-2N(\epsilon - \alpha - \beta)^2\big\}.$$

Now, let $x \in \bar{X}^N_\alpha(l, u)$ and let $y' \in Y$ be such that $y' \geq g(x)$ and $\|y' - g(x)\|_\infty \leq \beta/L$. By Lipschitz continuity of F, this implies

$$F(y') - F(g(x)) \leq L\|y' - g(x)\|_\infty \leq \beta. \tag{11}$$

Because x satisfies $\sum_{i=1}^N \mathbb{I}\big(g(x) \geq \xi^i\big) \geq N(1-\alpha)$ and $y' \geq g(x)$, we have $\sum_{i=1}^N \mathbb{I}\big(y' \geq \xi^i\big) \geq N(1-\alpha)$ and hence $y' \in Y_\alpha^N$. Thus, using (11), when $Y_\alpha^N \subseteq Y_{\epsilon-\beta}$ occurs,

$$F(g(x)) \geq F(y') - \beta \geq (1 - \epsilon + \beta) - \beta = 1 - \epsilon.$$

Since $x \in \bar{X}_\alpha^N(l,u)$ was arbitrary, $Y_\alpha^N \subseteq Y_{\epsilon-\beta}$ implies $\bar{X}_\alpha^N(l,u) \subseteq \bar{X}_\epsilon$ and the result follows from (10). □

To obtain confidence at least $1-\delta$ that $\bar{X}_\alpha^N(l,u) \subseteq \bar{X}_\epsilon$ it is sufficient to take

$$N \geq \frac{1}{2(\epsilon-\alpha-\beta)^2} \log\left(\frac{1}{\delta}\right) + \frac{m}{2(\epsilon-\alpha-\beta)^2} \log\Big\lceil \frac{DL}{\beta} \Big\rceil.$$

Note that for fixed $\epsilon > 0$ and $\alpha \in [0, \epsilon)$, β is a free parameter which can be chosen in $(0, \epsilon - \alpha)$. If, for example, we take $\beta = (\epsilon - \alpha)/2$ we obtain

$$N \geq \frac{2}{(\epsilon-\alpha)^2} \log\left(\frac{1}{\delta}\right) + \frac{2m}{(\epsilon-\alpha)^2} \log\Big\lceil \frac{2DL}{\epsilon-\alpha} \Big\rceil.$$

Once again, if $\alpha = 0$, similar arguments can be used to conclude that if

$$N \geq \frac{2}{\epsilon} \log\left(\frac{1}{\delta}\right) + \frac{2m}{\epsilon} \log\Big\lceil \frac{2DL}{\epsilon} \Big\rceil$$

then $\Pr\big\{\bar{X}_0^N(l,u) \subseteq \bar{X}_\epsilon\big\} \geq 1-\delta$.

2.2.3. Lipschitz continuous G. We now turn to the problem of using a sample approximation problem to generate feasible solutions to P_ϵ when X is not necessarily finite, and $G(x,\xi)$ does not necessarily have the form $G(x,\xi) = g(x) - \xi$. In this section, we assume for simplicity of exposition that G takes values in $\mathbf{R}$. This is without loss of generality, since if $\bar{G} : \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}^m$ we can define $G : \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}$ by $G(x,\xi) = \max\{\bar{G}_j(x,\xi) : j = 1,\ldots,m\}$ and the constraints $G(x,\xi) \leq 0$ and $\bar{G}(x,\xi) \leq \mathbf{0}$ are equivalent. In this section, we shall make the following Lipschitz continuity assumption on G.

ASSUMPTION 1. *There exists $L > 0$ such that*

$$|G(x,\xi) - G(x',\xi)| \leq L\|x - x'\|_\infty \qquad \forall x, x' \in X \text{ and } \forall \xi \in \Xi.$$

It is important that the Lipschitz constant L is independent of $\xi \in \Xi$, and this condition may make Assumption 1 appear rather stringent. There are, however, interesting cases in which the assumption does hold. For example, if Ξ is finite (with possibly huge cardinality) and $G(x,\xi)$ is Lipschitz continuous with Lipschitz constant $L(\xi)$ for each $\xi \in \Xi$, then Assumption 1 holds with $L = \max\{L(\xi) : \xi \in \Xi\}$. Alternatively, if Ξ is compact and $G(x,\xi) = \max\{T_j(\xi)x : j = 1,\ldots,m\}$ and $T_j : \Xi \to \mathbf{R}^n$ is continuous in ξ for each j, then Assumption 1 holds with

$$L = \sup_{\xi \in \Xi}\big\{\max\{\|T_j(\xi)\|_\infty : j = 1,\ldots,m\}\big\}.$$

To generate feasible solutions for this general case, we will also need to modify the sample approximation problem somewhat. In addition to taking a risk level α less than the nominal risk level ϵ, we will require that at least $(1-\alpha)N$ of the constraints be satisfied *strictly*. That is, for a fixed $\gamma > 0$, we define the sample approximation feasible region to be

$$X_{\alpha,\gamma}^N = \Big\{x \in X : \frac{1}{N}\sum_{i=1}^N \mathbb{I}\big(G(x,\xi) + \gamma \leq 0\big) \geq 1 - \alpha\Big\}.$$

Finally, we will assume that X is bounded, and let $D = \sup\{\|x - x'\|_\infty : x, x' \in X\}$ be the diameter of X.

THEOREM 10. *Suppose X is bounded with diameter D and Assumption 1 holds. Let $\alpha \in [0, \epsilon)$, $\beta \in (0, \epsilon - \alpha)$ and $\gamma > 0$. Then,*

$$\Pr\big\{X^N_{\alpha,\gamma} \subseteq X_\epsilon\big\} \geq 1 - \lceil 1/\beta \rceil \lceil 2LD/\gamma \rceil^n \exp\big\{-2N(\epsilon - \alpha - \beta)^2\big\}.$$

Proof. For $x \in X$, let $\phi(x) = \Pr\big\{G(x, \xi) \leq 0\big\}$. Let $J = \lceil 1/\beta \rceil$ and for $j = 1, \ldots, J-1$, define

$$X_j = \Big\{x \in X : \frac{j-1}{J} \leq \phi(x) < \frac{j}{J}\Big\}$$

and let $X_J = \{x \in X : (J-1)/J \leq \phi(x) \leq 1\}$. Next, we claim that for each j there exists a finite set $Z^\gamma_j \subseteq X_j$ such that $|Z^\gamma_j| \leq \lceil 2LD/\gamma \rceil^n$ and for all $x \in X_j$ there exists $z \in Z^\gamma_j$ such that $\|x - z\|_\infty \leq \gamma/L$. Indeed, because $X_j \subseteq X$ and X is bounded with diameter D, there exists a finite set $Y \subseteq \mathbf{R}^n$ with $|Y| \leq \lceil 2LD/\gamma \rceil^n$ such that for all $x \in X$ there exists $y \in Y$ such that $\|x - y\|_\infty \leq \gamma/2L$. For any $y \in \mathbf{R}^n$ and $\eta > 0$, define $B(y, \eta) = \{x \in \mathbf{R}^N : \|y - x\|_\infty \leq \eta\}$. Now, let $Y'_j = \{y \in Y : X_j \cap B(y, \gamma/2L) \neq \emptyset\}$, and for $y \in Y'_j$ select an arbitrary $x_y \in X_j \cap B(y, \gamma/2L)$. Then, let $Z^\gamma_j = \bigcup_{y \in Y'_j} x_y$. By definition, $Z^\gamma_j \subseteq X_j$ and $|Z^\gamma_j| \leq \lceil 2LD/\gamma \rceil^n$. In addition, for any $x \in X_j$, there exists y such that $x \in B(y, \gamma/2L)$ and because for this y, $X_j \cap B(y, \gamma/2L) \neq \emptyset$ there exists $x_y \in Z^\gamma_j$ such that $\|x_y - y\|_\infty \leq \gamma/2L$. Hence,

$$\|x_y - x\|_\infty \leq \|x_y - y\|_\infty + \|y - x\|_\infty \leq \gamma/L.$$

Now, define $Z^\gamma = \bigcup_{j=1}^J Z^\gamma_j$ and observe that $|Z^\gamma| \leq J\lceil 2LD/\gamma \rceil^n$. Next, define $Z^\gamma_{\epsilon-\beta} = \big\{x \in Z^\gamma : \Pr\big\{G(x, \xi) \leq 0\big\} \geq 1 - \epsilon + \beta\big\}$ and

$$Z^{\gamma,N}_\alpha = \Big\{x \in Z^\gamma : \frac{1}{N}\sum_{i=1}^N \mathbb{I}\big(G(x, \xi^i) \leq 0\big)\Big\}.$$

Since Z^γ is finite and $\alpha < \epsilon - \beta$ we can apply Theorem 5 to obtain

$$\Pr\big\{Z^{\gamma,N}_\alpha \subseteq Z^\gamma_{\epsilon-\beta}\big\} \geq 1 - \lceil 1/\beta \rceil \lceil 2LD/\gamma \rceil^n \exp\big\{-2N(\epsilon - \alpha - \beta)^2\big\}. \tag{12}$$

Now consider an arbitrary $x \in X^N_{\alpha,\gamma}$. Let $j \in \{1, \ldots, J\}$ be such that $x \in X_j$. By definition of Z^γ_j there exists $z \in Z^\gamma_j$ such that $\|x - z\|_\infty \leq \gamma/L$. By definition of X_j and because $Z^\gamma_j \subseteq X_j$, we have $|\phi(x) - \phi(z)| \leq \beta$. In addition, Assumption 1 implies $|G(x, \xi^i) - G(z, \xi^i)| \leq \gamma$. Hence, if $G(x, \xi^i) + \gamma \leq 0$ then $G(z, \xi^i) \leq 0$ and, because x satisfies $\sum_{i=1}^N \mathbb{I}\big(G(x, \xi^i) + \gamma \leq 0\big) \geq N(1 - \alpha)$, it follows that z satisfies $\sum_{i=1}^N \mathbb{I}\big(G(z, \xi^i) \leq 0\big) \geq N(1 - \alpha)$. Thus $z \in Z^{\gamma,N}_\alpha$ and so if $Z^{\gamma,N}_\alpha \subseteq Z^\gamma_{\epsilon-\beta}$ then $\phi(z) \geq 1 - \epsilon + \beta$. Thus, $\phi(x) \geq \phi(z) - \beta \geq 1 - \epsilon$ when $Z^{\gamma,N}_\alpha \subseteq Z^\gamma_{\epsilon-\beta}$. Since $x \in X^N_{\alpha,\gamma}$ was arbitrary, $Z^{\gamma,N}_\alpha \subseteq Z^\gamma_{\epsilon-\beta}$ implies $X^N_{\alpha,\gamma} \subseteq X_\epsilon$ and the result follows from (12). ∎

Once again, for fixed ϵ and $\alpha < \epsilon$, β is a free parameter to be chosen in $(0, \epsilon - \alpha)$. If we choose, for example $\beta = (\epsilon - \alpha)/2$, then we can assure $X^N_{\alpha,\gamma} \subseteq X_\epsilon$ with confidence at least $1 - \delta$ by taking

$$N \geq \frac{2}{(\epsilon - \alpha)^2}\Big[\log\Big(\frac{1}{\delta}\Big) + n\log\Big\lceil\frac{2LD}{\gamma}\Big\rceil + \log\Big\lceil\frac{2}{\epsilon - \alpha}\Big\rceil\Big].$$

Additionally, if $\alpha = 0$, similar arguments show that $X^N_{0,\gamma} \subseteq X_\epsilon$ occurs with probability at least $1 - \delta$ if

$$N \geq \frac{2}{\epsilon}\Big[\log\left(\frac{1}{\delta}\right) + n\log\Big\lceil\frac{2LD}{\gamma}\Big\rceil + \log\Big\lceil\frac{2}{\epsilon}\Big\rceil\Big].$$

Regardless of whether $\alpha = 0$ or $\alpha > 0$ the term $1/\gamma$ is taken under log, and hence γ can be made very small without significantly increasing the required sample size, suggesting that modifying the sample approximation problem to require at least $(1-\alpha)N$ of the sampled constraints to be satisfied with slack at least γ need not significantly alter the feasible region.

2.2.4. A Posteriori Feasibility Checking. The results of §2.2.1 - 2.2.3 demonstrate that with appropriately constructed sample approximation problems, the probability that the resulting feasible region will be a subset of the true feasible region X_ϵ approaches one exponentially fast. This gives strong theoretical support for using these sample approximations to yield solutions feasible to X_ϵ. These results yield *a priori* estimates on how large the sample size N should be to have high confidence the sample approximation feasible region will be a subset of X_ϵ. However, these a priori estimates are likely to yield required sample sizes which are very large, and hence the sample approximation problems will still be impractical to solve. This is particularly true if $\alpha > 0$ and $\epsilon - \alpha$ is small. However, typically in sampling approximation results such as these, the a priori estimates of the required sample size are very conservative, and in fact much smaller sample sizes are sufficient. See [18] for a computational demonstration of this phenomenon for the case of Sample Average Approximation applied to two-stage stochastic linear programs. Thus, a natural alternative to using the sample size suggested by the a priori estimates is to solve a sample approximation problem with a smaller sample to yield a candidate solution $\hat{x} \in X$, and then conduct an *a posteriori* check to see whether $\Pr\big\{G(\hat{x},\xi) \leq \mathbf{0}\big\} \geq 1 - \epsilon$. A simple method for conducting an a posteriori analysis of the risk of a candidate solution is to take a single very large Monte Carlo sample $\xi^1, \ldots, \xi^{N'}$ and count how many times $G(\hat{x},\xi^i) \leq \mathbf{0}$ holds. Bounds on the true risk $\Pr\big\{G(\hat{x},\xi) \leq \mathbf{0}\big\}$ which hold with high confidence can then be constructed, and if N' is very large, these bounds should be tight. This approach will not work well if the allowed risk ϵ is extremely small, but on the other hand, we do not expect the sample approximation approach to be practical in this case anyway. Of course, if good estimates of $\Pr\big\{G(\hat{x},\xi) \leq \mathbf{0}\big\}$ can be obtained efficiently by some other method, then this other method should be used for a posteriori feasibility checking. For example, if $G(x,\xi) = \xi - g(x)$ and the components of ξ are independent, then $\Pr\big\{g(x) \geq \xi\big\}$ can be calculated as $\prod_i \Pr\big\{g_i(x) \geq \xi_i\big\}$.

3. Numerical Experiments. We conducted experiments to test the effectiveness of the sample approximation approach for yielding good feasible solutions and lower bounds. In particular, our aim is to determine whether using $\alpha > 0$ in the sample approximation can yield better solutions than when using $\alpha = 0$ as in the scenario approximation approach of [7, 22]. In addition, we test whether reasonable lower bounds which are valid with high probability can be obtained. We first conducted tests on a probabilistic version of the classical set covering problem, which has been studied recently in [5, 26, 27]. This problem has both finite feasible region and finite distribution (although both are exponentially large) so that for generating feasible solutions, the stronger Theorems 5 and 8 apply. These results are given in §3.1. We also conducted tests on a probabilistic version of the transportation problem. For this

problem, the feasible region is continuous and we also use a joint normal distribution for the right-hand side vector, so that Theorem 9 applies. These results are presented in §3.2.

Note that although Theorem 3 provides support for using the sample approximation scheme to generate lower bounds, we will use Theorem 4 to actually obtain lower bounds which are valid with high confidence, because it can be used regardless of how large the sample size N is (with the possible drawback that using smaller N will yield weaker lower bounds). Similarly, Theorems 5, 8, and 9 support the use of sample approximation to yield feasible solutions, but we do not use these Theorems to guide our choice of α and N. Indeed, the bounds implied by these theorems would suggest using N which is far too large to be able to solve the approximation problem. Instead, we experiment with different values of α and N, and perform an *a posteriori* test on each solution generated to determine whether it is feasible (with high confidence).

3.1. Probabilistic Set Cover Problem. The Probabilistic Set Cover Problem is given by

$$\text{(PSC)} \qquad \min\{cx : \Pr\{Ax \geq \xi\} \geq 1 - \epsilon, x \in \{0,1\}^n\}$$

where $c \in \mathbf{R}^n$ is the cost vector, A is an $m \times n$ zero-one matrix and ξ is a random vector taking values in $\{0,1\}^m$. We conducted tests on a single instance of PSC, with two values of ϵ, 0.05 and 0.1.

3.1.1. Test Instance. Following [5], we based our tests on a deterministic set covering instance, scp41, of the OR library [3], which has $m = 200$ rows and $n = 1000$ columns. Also following [5], the random vector ξ is assumed to consist of 20 independent sub-vectors, with each sub-vector having size $k = 10$ following the *circular* distribution. The circular distribution is defined by parameters $\lambda_j \in [0,1]$ for $j = 1, \ldots, k$. First, Bernoulli random variables Y_j for $j = 1, \ldots, k$ are generated independently with $\Pr\{Y_j = 1\} = \lambda_i$. Then, the random sub-vector is defined by $\xi_j = \max\{Y_j, Y_{j+1}\}$ for $j < k$ and by $\xi_k = \max\{Y_1, Y_k\}$. Because of the simple form of this distribution, given a solution x, it is possible to calculate exactly $\Pr\{Ax \geq \xi\}$. Thus, when a solution is obtained from a sample approximation problem, we test a posteriori whether it is feasible at a given risk level by exactly calculating $\Pr\{Ax \geq \xi\}$. To illustrate this calculation, we show how to calculate the probability for a single sub-vector, that is $\Pr\{\xi_j \leq y_j,\ j = 1, \ldots, k\}$. Then, with $y = Ax$, the overall probability $\Pr\{Ax \geq \xi\}$ is calculated as the product of the probabilities for each sub-vector. Let $J = \{1 \leq j \leq k : y_j = 0\}$. Then,

$$\Pr\{\xi_j \leq y_j,\ j = 1, \ldots, k\} = \Pr\{\xi_j = 0, j \in J\} = \Pr\{Y_j = 0, j \in J^+\} = \prod_{j \in J^+} (1 - \lambda_j)$$

where $J^+ = \cup_{j \in J}\{j, (j+1) \mod k\}$. Although in this test, calculation of the distribution function is easy, we stress that this is not a necessary condition to use the sample approximation, it is only necessary that sampling from the distribution can be done efficiently.

3.1.2. Solving the Sample Approximation. To solve the sample approximation problem of the PSC, we used a MIP formulation which is equivalent to an extended formulation studied in [20] (see also [19]). The formulation is not exactly the same, since because the random right-hand side can take on only two values, it can be simplified somewhat. Let the scenarios obtained in the sample of size N be

denoted by ξ^i for $i = 1, \ldots, N$, where each $\xi^i \in \{0,1\}^m$. Then, the formulation we use is

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & Ax \geq y \\ & y_j + z_i \geq 1 \qquad \forall i, j \text{ s.t. } \xi^i_j = 1 \qquad (13) \\ & \sum_{i=1}^{N} z_i \leq p \qquad (14) \\ & x \in \{0,1\}^n,\ z \in \{0,1\}^N,\ y \in \{0,1\}^m \end{aligned}$$

where $p = \lfloor \alpha N \rfloor$. We could relax the intregrality restriction on the y variables, but we found that leaving this restriction and also placing higher branching priority on these variables significantly improved performance when solving with CPLEX 9.0. The intuition behind this is that if we fix $y_j = 1$, then we are enforcing the constraint $A^j x \geq 1$, and on the other hand, if we fix $y_j = 0$, then any scenario i for which $\xi^i_j = 1$ will be fixed to 1, and constraint (14) will quickly become binding. We also found that some simple preprocessing of the formulation significantly helped solution times. If, for a row j, $\sum_i \xi^i_j > p$, then we cannot have $y_j = 0$, and so we fixed $y_j = 1$, and the corresponding inequalities (13) for j were not included. After this preprocessing, for each j there will be at most p inequalities in (13), so that these inequalities add at most mp rows and $O(mp)$ nonzeros to the formulation. Using this formulation, we found the sample approximation problems could be solved quickly, in all cases in less than ten seconds, and usually much less. However, this may be due to the particular distribution used (and the simplicity of the underlying set cover instance), and thus this should not be taken as a study of the effectiveness of this formulation in general. Rather, we are interested here only in the properties of the solutions generated by the sample approximation problems.

3.1.3. Feasible Solutions. We first tested the effectiveness of the sample approximation approach for generating feasible solutions. To do so, we varied the risk level of the approximation problem, α, and sample size, N. For each combination of α and N we generated and solved 10 sample approximation problems. Table 1 gives statistics of the solutions generated for the PSC instance with $\epsilon = 0.05$, and Table 2 gives the same for the PSC instance with $\epsilon = 0.1$. For each combination of α and

TABLE 1
Solution results for PSC sample problems with $\epsilon = 0.05$.

		Solution Risk				Feasible Solutions Cost				
α	N	Ave	Min	Max	σ	#	Ave	Min	Max	σ
0.00	100	0.107	0.048	0.185	0.042	1	425.0	425	425	***
	110	0.071	0.013	0.100	0.029	3	425.7	424	429	2.9
	120	0.069	0.013	0.152	0.049	4	424.8	424	427	1.5
	130	0.062	0.020	0.124	0.036	5	424.8	420	429	4.3
	140	0.042	0.018	0.080	0.017	8	425.6	421	429	2.8
	150	0.041	0.005	0.080	0.026	6	427.3	421	429	3.1
0.05	1000	0.056	0.041	0.072	0.009	2	414.0	414	414	0.0
	3000	0.044	0.041	0.055	0.005	8	414.0	414	414	0.0
	5000	0.044	0.041	0.060	0.006	8	414.0	414	414	0.0
	7500	0.041	0.041	0.041	0.000	10	414.0	414	414	0.0
	10000	0.044	0.041	0.054	0.005	8	414.0	414	414	0.0

N, we report statistics on the risk of the generated solutions, where for a solution x,

the risk is $\Pr\{Ax \not\geq \xi\}$, as well as on the costs of the *feasible* solutions generated, i.e. those solutions which have risk less than 0.05 and 0.1 respectively. For the risk of the solutions, we report the average, minimum, maximum and sample standard deviation over the 10 solutions. For the solutions costs, we report first how many solutions were feasible, then report the average, minimum, maximum and sample standard deviation of the cost taken over these solutions.

TABLE 2
Solution results for PSC sample problems with $\epsilon = 0.1$.

		Solution Risk				Feasible Solutions Cost				
α	N	Ave	Min	Max	σ	#	Ave	Min	Max	σ
0.0	80	0.203	0.095	0.311	0.076	1	420.0	420	420	***
	90	0.169	0.084	0.239	0.051	1	428.0	428	428	***
	100	0.107	0.048	0.185	0.042	4	426.0	423	428	500.7
	110	0.071	0.013	0.100	0.029	9	425.4	421	429	499.8
	120	0.069	0.013	0.152	0.049	7	424.6	419	428	534.3
	130	0.062	0.020	0.124	0.036	7	425.3	420	429	488.8
0.1	1000	0.111	0.095	0.141	0.015	4	401.3	400	403	1.5
	3000	0.101	0.092	0.115	0.006	6	401.0	400	402	1.1
	5000	0.101	0.092	0.108	0.005	5	401.2	400	402	1.1
	7500	0.099	0.092	0.105	0.004	7	401.1	400	402	1.1
	10000	0.097	0.088	0.103	0.004	8	401.8	400	404	1.3

We first discuss results for the case of nominal risk level $\epsilon = 0.05$. When using $\alpha = 0$, the best results were obtained with N in the range of 100-150, and these are the results we report. With $\alpha = 0$, as N increases more constraints are being enforced, which leads to smaller feasible region of the approximation and higher likelihood that the optimal solution of the approximation is feasible at the nominal risk level. However, the smaller feasible region also causes the cost to increase, so that increasing N more would yield overly conservative solutions. We also conducted tests with $\alpha = 0.05$, and for this value of α we used significantly larger sample sizes. The best feasible solution found using $\alpha = 0$ had cost 420, and the average cost of the feasible solutions found was significantly greater than this. When $\alpha = 0.05$, every sample size N yielded at least one feasible solution in the ten runs, and *every* feasible solution found had cost 414. Thus, using $\alpha = 0.05$ consistently yields solutions which are closer to the efficient frontier between the objectives of risk and cost.

For $\epsilon = 0.1$, we observed similar results. In this case, when using $\alpha = 0$, the best results were obtained with N in the range of 80-130. The best solution found using $\alpha = 0$ had cost 419, whereas the best solution found using $\alpha = 0.1$ was 400, which was obtained by one of the ten runs for *every* sample size N. In addition, observe from Table 1 that using $\alpha = 0.05$ yields solutions with risk not exceeding 0.05 and cost of 414, which is also less than the cost of the best solution found that had risk not exceeding 0.1 when using $\alpha = 0$. Thus, using $\alpha > 0$ we are able to get solutions with lower risk *and* lower cost as compared to those obtained when using $\alpha = 0$.

In terms of the variability of the risks and costs of the solutions generated, using $\alpha > 0$ and a much larger sample size yielded solutions with much lower variability than when using $\alpha = 0$ and small sample size. This is not surprising since using a larger sample size naturally should reduce variability. On the other hand, constraining the sample approximation to have $\alpha = 0$ prohibits the use of a larger sample size, as the solutions produced then become overly conservative.

3.1.4. Lower Bounds. We next discuss the results for obtaining lower bounds for the PSC. We used the procedure of Theorem 4 with $\alpha = \epsilon$ and $M = 10$. We use

the same 10 sample approximation problems as when generating feasible solutions. As argued after Theorem 4, with $\alpha = \epsilon$, we have $\rho(\alpha, \epsilon, N) = \rho(\epsilon, \epsilon, N) \gtrsim 1/2$. Then, if we take $L = 1$ the test of Theorem 4 yields a lower bound with confidence 0.999. Taking $L = 1$ corresponds to taking the minimum optimal value over all the $M = 10$ runs (not just over the ones which yielded feasible solutions). More generally, we can take $L \in \{1, \ldots, 10\}$ yielding a lower bound with confidence at least

$$1 - \sum_{i=0}^{L-1} \binom{10}{i} \rho(\epsilon, \epsilon, N)^i \big(1 - \rho(\epsilon, \epsilon, N)\big)^{10-i} \gtrsim 1 - \sum_{i=0}^{L-1} \binom{10}{i} (1/2)^{10}$$

to obtain possibly "tighter" lower bounds of which we are less confident.

TABLE 3
Lower bounds for PSC sample problems with $\alpha = \epsilon = 0.05$.

	LB with confidence at least:				Gap with confidence at least:			
N	0.999	0.989	0.945	0.828	0.999	0.989	0.945	0.828
1000	412	414	414	414	0.5%	0.0%	0.0%	0.0%
3000	412	414	414	414	0.5%	0.0%	0.0%	0.0%
5000	412	414	414	414	0.5%	0.0%	0.0%	0.0%
7500	414	414	414	414	0.0%	0.0%	0.0%	0.0%
10000	413	414	414	414	0.2%	0.0%	0.0%	0.0%

The results obtained using varying values of N and $\epsilon = \alpha = 0.05$ are given in Table 3. The gaps reported are the percent by which the lower bound is below the best feasible solution (414, obtained with $\alpha = 0.05$ and any if the tested sample sizes N). Thus, for example, by solving 10 problems with sample size $N = 1000$ we obtained a feasible solution of cost 414, and lower bound of 412, which is valid with probability at least 0.999. In addition, we obtain a lower bound of 414 which is valid with probability at least 0.989. Thus, we have confidence at least 0.989 that 414 is the optimal value. Similar results were obtained with larger sample sizes.

Table 4 yields the lower bound results obtained with $\epsilon = \alpha = 0.1$ and varying sample size N. Solving 10 sample problems with $N = 1000$ we obtained a feasible solution of cost 400, and can say with confidence 0.999 that the optimal solution is at most 0.8% less costly than this solution. Using $N = 5000$ (or greater), we obtain a feasible solution of the same cost, but a lower bound which states that with confidence at least 0.999 this feasible solution is optimal.

TABLE 4
Lower bounds for PSC sample problems with $\alpha = \epsilon = 0.1$.

	LB with confidence at least:				Gap with confidence at least:			
N	0.999	0.989	0.945	0.828	0.999	0.989	0.945	0.828
1000	397	397	398	398	0.8%	0.8%	0.5%	0.5%
3000	399	400	400	400	0.3%	0.0%	0.0%	0.0%
5000	400	400	400	400	0.0%	0.0%	0.0%	0.0%
7500	400	400	400	400	0.0%	0.0%	0.0%	0.0%
10000	400	400	400	400	0.0%	0.0%	0.0%	0.0%

3.2. Probabilistic Transportation Problem. We next tested the sampling approach on a probabilistic version of the classical transportation problem, which we call the Probabilistic Transportation Problem (PTP). In this problem, we have a set of suppliers I and a set of customers D with $|D| = m$. The suppliers have limited

capacity M_i for $i \in I$. There is a transportation cost c_{ij} for shipping a unit of product from supplier $i \in I$ to customer $j \in D$. The customer demands are random and are represented by a random vector $\tilde{d}$ taking values in $\mathbf{R}^m$. We assume we must choose the shipment quantities before the customer demands are known. We enforce the probabilistic constraint

$$\Pr\Big\{\sum_{i\in I} x_{ij} \geq \tilde{d}_j, j = 1, \ldots, m\Big\} \geq 1 - \epsilon \tag{15}$$

where $x_{ij} \geq 0$ is the amount shipped from supplier $i \in I$ to customer $j \in D$. The objective is to minimize distribution costs subject to (15), and the supply capacity constraints

$$\sum_{j\in D} x_{ij} \leq M_i, \quad \forall i \in I.$$

3.2.1. Test Instances. We conducted our tests on an instance with 40 suppliers and 50 customers. The supply capacities and cost coefficients were randomly generated using normal and uniform distributions respectively. The demand is assumed to have a joint normal distribution. The mean vector and covariance matrix were randomly generated. We considered two cases for the covariance matrix: a low variance and a high variance case. In the low variance case, the standard deviation of the one dimensional marginal random demands is 10% of the mean on average. In the high variance case, the covariance matrix of the low variance case is multiplied by 25, yielding standard deviations of the one dimensional marginal random demands being 50% of the mean on average. In both cases, we consider a single risk level $\epsilon = 0.05$.

We remark that for this particular choice of distribution, the feasible region defined by the probabilistic constraint is convex [24]. However, the dimension of the random vector $\tilde{d}$ is $m = 50$, and so evaluating $\Pr\{y \geq \tilde{d}\}$ for a single vector $y \in \mathbf{R}^m$ would present a computational challenge, whereas in our approach we merely need to generate random samples from the joint normal distribution, which is relatively easy. On the other hand, we have not conducted experiments using the convex programming approach, so we cannot comment on whether our approach works better than this. This would be an interesting future experiment. Our intention here is merely to test our approach on a problem with continuous feasible region and distribution.

Once a sample approximation is solved yielding solution $\hat{x}$, we use a single very large sample ($N' = 250,000$), to estimate $\Pr\{\hat{y} \geq \tilde{d}\}$ where $\hat{y} \in \mathbf{R}^m$ is the vector given by $\hat{y}_j = \sum_{i\in I} \hat{x}_{ij}$ for $j \in D$. Letting $d^1, \ldots, d^{N'}$ be the realizations of this large sample, we calculate $\sum_{i=1}^{N'} \mathbb{I}(\hat{y} \geq d^i)$ and use the normal approximation to the binomial distribution to construct an upper bound $\hat{\alpha}$ on the true solution risk $\Pr\{\hat{y} \geq \tilde{d}\}$, which is valid with confidence 0.999. Henceforth for this experiment, if we say a solution is feasible at risk level ϵ, we mean $\hat{\alpha} \leq \epsilon$, and so it is feasible at this risk level with confidence 0.999. We used such a large sample to get a good estimate of the true risk of the solutions generated, but we note that because this sample was so large, generating this sample and calculating $\sum_{i=1}^{N'} \mathbb{I}(\hat{y} \geq d^i)$ often took longer than solving the sample approximation itself.

3.2.2. Solving the Sample Approximation. We solved the sample approximation problem using a mixed-integer programming formulation, augmented with a class of strong valid inequalities. We refer the reader to [20, 19] for details of this formulation and the valid inequalities, as well as detailed computational results for

solving the sample approximation problems. However, we mention that in contrast to the probabilistic set cover problem, solving the sample approximation problem with the largest sample size we consider ($N = 10000$) and largest α (0.05) takes a non-trivial amount of time, in some cases as long as 30 minutes. On the other hand, for $N = 5000$, the worst case was again $\alpha = 0.05$ and usually took less than 4 minutes to solve.

TABLE 5
Solution results for low variance PTP sample problems with $\epsilon = 0.05$.

		Solution Risk				Feasible Solutions Cost				
α	N	Ave	Min	Max	σ	#	Ave	Min	Max	σ
0.000	900	0.048	0.036	0.066	0.011	7	2.0266	2.0199	2.0320	0.0045
	950	0.047	0.039	0.055	0.005	6	2.0244	**2.0185**	2.0291	0.0041
	1000	0.045	0.040	0.051	0.004	8	2.0253	2.0185	2.0300	0.0039
	1500	0.033	0.025	0.043	0.005	10	2.0336	2.0245	2.0406	0.0053
0.030	5000	0.049	0.045	0.050	0.002	6	2.0098	2.0075	2.0114	0.0013
	7500	0.045	0.041	0.047	0.002	10	2.0112	2.0094	2.0136	0.0015
	10000	0.042	0.041	0.044	0.001	10	2.0129	2.0112	2.0145	0.0010
0.033	5000	0.052	0.049	0.054	0.002	2	2.0080	2.0073	2.0088	0.0011
	7500	0.048	0.045	0.051	0.002	7	2.0092	2.0075	2.0107	0.0012
	10000	0.045	0.044	0.047	0.001	10	2.0103	2.0089	2.0118	0.0009
0.036	5000	0.055	0.053	0.057	0.002	0	***	***	***	***
	7500	0.052	0.049	0.054	0.002	2	2.0079	2.0077	2.0080	0.0002
	10000	0.049	0.047	0.051	0.001	8	2.0080	**2.0066**	2.0093	0.0008

3.2.3. Low Variance Instance. We begin by presenting results for the instance in which the distribution of demand has relatively low variance. For generating feasible solutions, we tested $\alpha = 0$ with various sample size N and report the results for the sample sizes which yielded the best results. Once again, this means we use a relatively small sample size for the case $\alpha = 0$, as compared to the cases with $\alpha > 0$. We tested several values of $\alpha > 0$ and varying sample size. In contrast to the PSC case, we found that taking $\alpha = \epsilon$ or even α close to ϵ did not yield feasible solutions, even with a large sample size. Thus, we report results for several different values of α in the range 0.03 to 0.036. The reason we report results for this many different values of α is to illustrate that within this range, the results are not extremely sensitive to the choice of α (results for more values of α can be found in [19]).

TABLE 6
Lower bounds for low variance PTP sample problems with $\alpha = \epsilon = 0.05$.

	LB with confidence at least:				Gap with confidence at least:			
N	0.999	0.989	0.945	0.828	0.999	0.989	0.945	0.828
1000	1.9755	1.9757	1.9775	1.9782	1.55%	1.54%	1.45%	1.42%
3000	1.9879	1.9892	1.9892	1.9910	0.93%	0.87%	0.87%	0.78%
5000	1.9940	1.9943	1.9948	1.9951	0.63%	0.62%	0.59%	0.57%
7500	1.9954	1.9956	1.9959	1.9963	0.56%	0.55%	0.54%	0.52%
10000	1.9974	1.9977	1.9980	1.9981	0.46%	0.45%	0.43%	0.42%

Table 5 gives the characteristics of the solutions generated for the different values of α and N. We observe that as in the case of the PSC, the *average* cost of the feasible solutions obtained using $\alpha > 0$ is always less than the *minimum* cost of the feasible solutions obtained with $\alpha = 0$. However, for this instance, the minimum cost solution obtained using $\alpha = 0$ is not so significantly worse than the minimum cost solutions using different values of $\alpha > 0$, being between 0.40% and 0.58% more costly. As in

the case of the PSC, using $\alpha > 0$ and large N significantly reduced the variability of the risk and cost of the solutions generated.

We next investigated the quality of the lower bounds that can be obtained for PTP by solving sample approximation problems. As in the case of the PSC, we obtained lower bounds by generating and solving 10 sample approximation problems with $\alpha = \epsilon = 0.05$. By taking the lowest value of all the optimal values we obtain a lower bound valid with confidence 0.999, taking the second smallest yields a lower bound which is valid with confidence 0.989, etc. The results for different values of N are given in Table 6. For reference, the percentage gap between these lower bounds and the best feasible solution found (with cost 2.0066) is also given. Using $N \geq 3000$ we obtain lower bounds that are valid with confidence 0.999 and are within one percent of the best feasible solution, indicating that for this low variance instance, the lower bounding scheme yields good evidence that the solutions we have found are good quality.

TABLE 7
Solution results for high variance PTP sample problems with $\epsilon = 0.05$.

		Solution Risk				Feasible Solutions Cost				
α	N	Ave	Min	Max	σ	#	Ave	Min	Max	σ
0.000	900	0.050	0.035	0.066	0.010	4	3.5068	3.4672	3.5488	0.0334
	950	0.050	0.041	0.058	0.006	6	3.4688	**3.4403**	3.4917	0.0191
	1000	0.045	0.041	0.052	0.004	9	3.4895	3.4569	3.5167	0.0234
	1500	0.030	0.022	0.035	0.005	10	3.5494	3.5205	3.6341	0.0368
0.030	5000	0.050	0.045	0.053	0.002	4	3.4014	3.3897	3.4144	0.0101
	7500	0.046	0.043	0.050	0.002	9	3.4060	3.3920	3.4235	0.0098
	10000	0.043	0.041	0.046	0.001	10	3.4139	3.4001	3.4181	0.0055
0.033	5000	0.053	0.046	0.057	0.003	1	3.4107	3.4107	3.4107	***
	7500	0.049	0.046	0.054	0.002	7	3.3928	**3.3865**	3.4020	0.0062
	10000	0.046	0.042	0.049	0.002	10	3.3982	3.3885	3.4139	0.0086
0.036	5000	0.057	0.049	0.060	0.003	1	3.3979	3.3979	3.3979	***
	7500	0.053	0.050	0.057	0.002	0	***	***	***	***
	10000	0.050	0.046	0.053	0.002	4	3.3927	3.3859	3.3986	0.0054

3.2.4. High Variance Instance. Table 7 gives the characteristics of the solutions generated for the high variance instance. In this case, the *maximum* cost of a feasible solution generated using any combination of $\alpha > 0$ and N was less than the *minimum* cost of any feasible solution generated using $\alpha = 0$. The minimum cost feasible solution generated with $\alpha = 0$ was between 0.87% and 1.6% more costly than the best feasible solution generated for the different combinations of $\alpha > 0$ and N. Thus, it appears that for the high variance instance, using $\alpha > 0$ in a sample approximation is more important for generating good feasible solutions than for the low variance instance.

Table 8 gives the lower bounds for different confidence levels and sample sizes, as well as the gaps between these lower bounds and the best feasible solution found. In this case, solving 10 instances with sample size $N = 1000$ yields a lower bound that is not very tight, 5.11% from the best solution cost at confidence level 0.999. Increasing the sample size improves the lower bound, but even with $N = 10000$, the gap between the lower bound at confidence 0.999 and the best solution found is 1.83%. Thus, it appears that for the high variance instance, the sample approximation scheme exhibits considerably slower convergence, in terms of the lower bounds, the feasible solutions generated, or both.

TABLE 8
Lower bounds for high variance PTP sample problems with $\alpha = \epsilon = 0.05$.

	LB with confidence at least:				Gap with confidence at least:			
N	0.999	0.989	0.945	0.828	0.999	0.989	0.945	0.828
1000	3.2089	3.2158	3.2178	3.2264	5.11%	4.91%	4.85%	4.59%
3000	3.2761	3.2775	3.2909	3.2912	3.12%	3.08%	2.69%	2.68%
5000	3.3060	3.3075	3.3077	3.3094	2.24%	2.19%	2.19%	2.14%
7500	3.3083	3.3159	3.3165	3.3169	2.17%	1.95%	1.93%	1.92%
10000	3.3200	3.3242	3.3284	3.3299	1.83%	1.70%	1.58%	1.53%

4. Concluding Remarks. We have studied a sample approximation scheme for probabilistically constrained optimization problems and demonstrated how this scheme can be used to generate optimality bounds and feasible solutions for very general optimization problems with probabilistic constraints. We have also conducted a preliminary computational study of this approach. This study demonstrates that using sample approximation problems that allow a choice of which sampled constraints to satisfy can yield good quality feasible solutions. In addition, the sample approximation scheme can be used to obtain lower bounds which are valid with high confidence. We found that good lower bounds could be found in the case of finite (but possibly exponential) feasible region and distribution, and also in the case of continuous feasible region and distribution, provided the distribution has reasonably low variance. With continuous feasible region and distribution, if the distribution has high variance the lower bounds were relatively weak. Future work in this area will include conducting more extensive computational tests, and also extending the theory to allow generation of samples which are not necessarily independent and identically distributed. For example, the use of variance reduction techniques such as Latin hypercube sampling or Quasi-Monte Carlo sampling may yield significantly faster convergence. In addition, to apply the results of this paper to more general probabilistic programs, such as mixed-integer programming with a random constraint matrix, it will be necessary to study how to solve the non-convex sample approximation problem in these cases.

Acknowledgments. The authors express thanks to Alexander Shapiro for his helpful comments and suggestions related to this work.

REFERENCES

[1] S. AHMED AND A. SHAPIRO, *The sample average approximation method for stochastic programs with integer recourse*. Preprint available at www.optimization-online.org, February 2002.
[2] J. ATLASON, M.A. EPELMAN, AND S.G. HENDERSON, *Call center staffing with simulation and cutting plane methods*, Ann. Oper. Res., 127 (2004), pp. 333–358.
[3] J.E. BEASLEY, *OR-library: Distributing test problems by electronic mail*, Journal of the Operational Research Society, 41 (1990), pp. 1069–1072.
[4] A. BEN-TAL AND A. NEMIROVSKI, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805.
[5] P. BERALDI AND A. RUSZCZYŃSKI, *The probabilistic set-covering problem*, Oper. Res., 50 (2002), pp. 956–967.
[6] D. BERTSIMAS AND M. SIM, *The price of robustness*, Oper. Res., 52 (2004), pp. 35–53.
[7] G.C. CALAFIORE AND M.C. CAMPI, *Uncertain convex programs: randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25–46.
[8] ———, *The scenario approach to robust control design*, IEEE Trans. Automat. Control, 51 (2006), pp. 742–753.
[9] M.C. CAMPI, G. CALAFIORE, AND S. GARATTI, *New results on the identification of interval predictor models*, in Proc. 16th IFAC World Congress, Prague, 2005.

[10] L. Dai, H. Chen, and J.R. Birge, *Convergence properties of two-stage stochastic programming*, J. Optim. Theory Appl., 106 (2000), pp. 489–509.
[11] L. El Ghaoui and H. Lebret, *Robust solutions to least-squares problems with uncertain data*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1035–1064.
[12] W.-R. Heilmann, *A note on chance-constrained programming*, Journal of the Operational Research Society, 34 (1983), pp. 533–537.
[13] R. Henrion, P. Li, A. Möller, M.C. Steinbach, M. Wendt, and G. Wozny, *Stochastic optimization for operating chemical processes under uncertainty*, in Online Optimization of Large Scale Systems, M. Grötschel, S.O. Krunke, and J. Rambau, eds., Springer, 2001, pp. 457–478.
[14] R. Henrion and A. Möller, *Optimization of a continuous distillation process under random inflow rate*, Comput. Math. Appl., 45 (2003), pp. 247–262.
[15] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30.
[16] A.J. Kleywegt, A. Shapiro, and T. Homem de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2001), pp. 479–502.
[17] M.A. Lejeune and A. Ruszczyński, *An efficient trajectory method for probabilistic inventory-production-distribution problems*, Oper. Res., 55 (2007), pp. 378–394.
[18] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behavior of sampling methods for stochastic programming*, Ann. Oper. Res., 142 (2006), pp. 215–241.
[19] J. Luedtke, *Integer Programming Approaches to Some Non-convex and Stochatic Optimization Problems*, PhD thesis, Georgia Institute of Technology, Atlanta, GA, 2007.
[20] J. Luedtke, S. Ahmed, and G. Nemhauser, *An integer programming approach for linear programs with probabilistic constraints*, in IPCO 2007, M. Fischetti and D.P. Williamson, eds., Lecture Notes in Comput. Sci., Berlin, 2007, Springer-Verlag, pp. 410–423. Full version available at www.optimization-online.org.
[21] M.R. Murr and A. Prékopa, *Solution of a product substitution problem using stochastic programming*, in Probabilistic Constrained Optimization: Methodology and Applications, S.P. Uryasev, ed., Kluwer Academic Publishers, 2000, pp. 252–271.
[22] A. Nemirovski and A. Shapiro, *Scenario approximation of chance constraints*, in Probabilistic and Randomized Methods for Design Under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, London, 2005, pp. 3–48.
[23] ———, *Convex approximations of chance constrained programs*, SIAM J. Optim., 17 (2006), pp. 969–996.
[24] A. Prékopa, *On probabilistic constrained programmming*, in Proceedings of the Princeton Symposium on Mathematical Programming, H.W. Kuhn, ed., Princeton, N.J., 1970, Princeton University Press, pp. 113–138.
[25] ———, *Probabilistic programming*, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds., vol. 10 of Handbooks Oper. Res. Management Sci., Elsevier, 2003, pp. 267–351.
[26] A. Saxena, *A short note on the probabilistic set covering problem*. Available online at www.optimization-online.org, March 2007.
[27] A. Saxena, V. Goyal, and M. Lejeune, *MIP reformulations of the probabilistic set covering problem*. Available online at www.optimization-online.org, February 2007.
[28] A. Shapiro, *Asymptotic behavior of optimal solutions in stochastic programming*, Math. Oper. Res., 18 (1993), pp. 829–845.
[29] A. Shapiro and T. Homem-de-Mello, *On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim., 11 (2000), pp. 70–86.
[30] A.K. Takyi and B.J. Lence, *Surface water quality management using a multiple-realization chance constraint method*, Water Resources Research, 35 (1999), pp. 1657–1670.
[31] S. Vogel, *Stability results for stochastic programming problems*, Optimization, 19 (1988), pp. 269–288.
[32] W. Wang, *Sample Average Approximation of Risk-Averse Stochastic Programs*, PhD thesis, Georgia Institute of Technology, Atlanta, GA, 2007.