2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Random Convex Approximations of Ambiguous Chance Constrained Programs

Shih-Hao Tseng   Eilyan Bitar   Ao Tang

***Abstract*—We investigate an approach to the approximation of *ambiguous chance constrained programs* (ACCP) in which the underlying distribution describing the random parameters is itself uncertain. We model this uncertainty with the assumption that the unknown distribution belongs to a closed ball centered around a fixed and known distribution. Using only samples drawn from the central distribution, we approximate ACCP with a robust sampled convex program (RSCP), and establish an upper bound on the probability that a solution to the RSCP violates the original ambiguous chance constraint, when the uncertainty set is defined in terms of the Prokhorov metric. Our bound on the constraint violation probability improves upon the existing bounds for RSCPs in the literature. We also consider another approach to approximating ACCP by means of a sampled convex program (SCP), which is built on samples drawn from the central distribution. Again, we provide upper bounds on the probability that a solution to the SCP violates the original ambiguous chance constraint for uncertainty sets defined according to a variety of metrics.**

***Index Terms*—Stochastic optimization, chance constraints, randomized algorithms, sample complexity.**

I. Introduction

We consider a class of convex programs whose constraints are parameterized by an unknown vector $\delta \in \Delta \subseteq \mathbb{R}^m$.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & x \in \mathcal{X} \\ & f(x, \delta) \leq 0. \end{aligned} \tag{1}$$

Here $x \in \mathbb{R}^n$ is the decision variable, $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed and convex set, $c \in \mathbb{R}^n$ is fixed and known, and the function $f : \mathcal{X} \times \Delta \to \mathbb{R}$ is closed and convex in x for each $\delta \in \Delta$. We refer to (1) as an *uncertain convex program* (UCP).[1]

In the literature, there are two distinct approaches to the treatment of uncertainty in (1) – commonly referred to as robust and probabilistic. The *robust approach* prescribes the procurement of minimum cost solutions that respect the constraints for all $\delta \in \Delta$. Doing so, in general, requires the solution of a semi-infinite convex program, if the uncertainty set Δ has infinite cardinality. There is, however, a large family of constraint functions f and uncertainty sets Δ for which such semi-infinite programs admit exact or conservative reformulations as finite-dimensional convex programs that can be efficiently solved [2]–[4]. Nevertheless, an important drawback of such an approach stems from the potential for over-conservatism in the robust solutions it prescribes. Namely, there might be elements in the uncertainty set that occur very rarely in practice, but manifest in a large increase in cost by requiring that solutions obtained be feasible under their realization. Of interest, then, are applications in which the practioner might be willing to exchange a small reduction in robustness of a solution for a large reduction in the cost incurred by such solution.

Motivated by this, the *probabilistic approach* to the treatment of uncertainty in (1) models the unknown vector δ as a random vector and requires the problem constraints to hold with probability no less than a desired level. More formally, this amounts to solving the so called *chance constrained program* (CCP) associated with (1); it is of the form

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & x \in \mathcal{X} \\ & \mathbb{P}\left\{f(x, \delta) \leq 0\right\} \geq 1 - \epsilon, \end{aligned} \tag{2}$$

where $\mathbb{P}$ denotes the probability distribution according to which the random vector δ is defined, and $\epsilon \in [0, 1]$ represents an acceptable probability of constraint violation.[2] By allowing a slight chance of constraint violation, optimal solutions to (2) may achieve a significantly lower cost than their robust counterparts. Their procurement, however, stands as a challenging computational task in general. First, with the exception of a few cases, the chance constraint in (2) results in a nonconvex feasible set – rendering the procurement of globally optimal solutions a difficult task. Second, checking as to whether the chance constraint is satisfied at a given point requires the calculation of a multidimensional integral, which can be costly if high accuracy is required. We refer the reader to [9], [10] for a more nuanced discussion around such issues, and [2] for the development of a computationally tractable alternative to (2), which relies on the convex inner approximation of its feasible set by means of Bernstein inequalities.

Beyond the apparent computational difficulties inherent to solving chance constrained programs, a more fundamental limitation of the approach stems from the implicit assumption that the underlying distribution, according to which the chance constraint is defined, is fixed and exactly known.

Supported in part by NSF grants ECCS-1351621, CNS-1239178, CNS-1544761, US DoE under the CERTS initiative, and the Atkinson Center for a Sustainable Future.

Shih-Hao Tseng, Eilyan Bitar, and Ao Tang are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 14853, USA. Emails: {st688, eyb5, at422}@cornell.edu

[1] The reader may refer to [1] for a discussion on the generality of such formulation.

[2] Chance constrained programs, in a form similar to (2), were first introduced and studied by Charnes et al. [5], Charnes and Cooper [6], Miller and Wagner [7], and Prékopa [8].

978-1-5090-1837-6/16/$31.00 ©2016 IEEE

Such assumption is difficult to enforce in practice, where the underlying distribution is commonly inferred from data, and as a result, may only be known to within some accuracy. Such issue is exasperated in settings where the procurement of data (e.g., independent samples of δ) requires costly experimentation. This, in turn, serves to limit the amount of data available to calibrate an accurate estimate of the underlying distribution. In order to accommodate the potential for inaccuracy in the specification of the underlying distribution, we adopt an existing approach in the literature [2], [11]–[13], and consider a generalization of problem (2) in the form of an *ambiguous chance constrained program* (ACCP); it is defined as

$$\begin{aligned} &\text{minimize} \quad c^\top x \\ &\text{subject to} \quad x \in \mathcal{X} \\ &\qquad \mathbb{P}\{f(x,\delta) \le 0\} \ge 1-\epsilon, \quad \forall\, \mathbb{P} \in \mathcal{P}, \end{aligned} \tag{3}$$

where $\mathcal{P}$ is a fixed and known set of probability distributions, to which the underlying distribution is assumed to belong. We refer to $\mathcal{P}$ as the *ambiguity set.* The computational tractability of (3) depends on the way in which the ambiguity set $\mathcal{P}$ is specified; and there are a variety of ways in which to do so. The primary approach in the literature relies on moment-based specifications [13]–[16]. We adopt an alternative approach in line with [11], [17]–[20], which defines the ambiguity set as a closed ball of distributions centered around a fixed and known distribution $\mathbb{P}_0$, as $\mathcal{P} = \{\mathbb{P} : \rho(\mathbb{P}, \mathbb{P}_0) \le r\}$. Here, $\rho(\cdot,\cdot)$ denotes a suitable distance function between probability measures supported on Δ. The radius $r \ge 0$ of the ambiguity set captures one's confidence in the accuracy of the nominal distribution. For $r = 0$, the set $\mathcal{P}$ reduces to a singleton, and we recover the ambiguity-free formulation in (2).

A. Contribution and Organization

In Section II, we adopt the approach of Erdoğan and Iyengar [11], and approximate ACCP (3) by a *robust sampled convex program* (RSCP) defined in (5). As optimal solutions to RSCPs are themselves random variables, we establish in Proposition 1 a bound on the probability that such solutions violate the original ambiguous chance constraint when the ambiguity set $\mathcal{P}$ is defined in terms of the Prokhorov metric. Our bound on the constraint violation probability improves upon the best known bound for RSCPs in the literature (cf. Theorem 6 in [11]). Inspired by the work of [1], [21]–[23], we also consider another approach to approximating ACCP by means of a *sampled convex program* (SCP). An important difference of our approach is that the construction of the SCP does not require the ability to sample the true underlying distribution, but instead requires only the ability to sample the central distribution $\mathbb{P}_0$. In Section III, we provide bounds on the probability that solutions to such SCPs violate the original ambiguous chance constraint for ambiguity sets $\mathcal{P}$ defined according to a variety of metrics. We conclude the paper with directions for future research in Section IV.

B. Notation

Let $\mathbb{R}$ denote the set of real numbers and $\mathbb{N}$ the set of positive integers. Given a set $\Delta \subseteq \mathbb{R}^m$, we denote by $\mathcal{B}(\Delta)$ the Borel σ-algebra on Δ, and by $\mathcal{M}(\Delta)$ the set of all probability measures on the space $(\Delta, \mathcal{B}(\Delta))$. Given an integer N and a measure $\mathbb{P} \in \mathcal{M}(\Delta)$, we denote by $\mathbb{P}^N$ the product measure on the space $(\Delta^N, \mathcal{B}(\Delta)^N)$.

II. Robust Sampled Convex Programs

A. Ambiguity Set

As in [11], we consider ambiguity sets defined in terms of the Prokhorov metric ρ_p, which is defined as follows. Given two probability measures $\mathbb{P}$, $\mathbb{Q} \in \mathcal{M}(\Delta)$, the Prokhorov metric is defined as

$$\rho_p(\mathbb{P}, \mathbb{Q}) := \inf\{\gamma > 0 : \mathbb{P}\{A\} \le \mathbb{Q}\{A^\gamma\} + \gamma,\ \forall\, A \in \mathcal{B}(\Delta)\},$$

where $A^\gamma := \{y \in \Delta : \inf_{z\in A} \|y - z\| < \gamma\}$ denotes the γ-neighborhood of the set A. Here, $\|\cdot\|$ is a suitable norm on the space Δ. Essentially, the Prokhorov metric measures the minimum distance in probability between two random variables distributed according to $\mathbb{P}$ and $\mathbb{Q}$, respectively. Although the metric is difficult to compute in practice, there are inequalities relating it to a host of other metrics and distance functions, which are more easily calculated. We refer the reader to Billingsley [24] for more details on the Prokhorov metric and Gibbs and Su [25] for a survey.

With this metric in hand, we define our ambiguity set as a ball of distributions given by

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{M}(\Delta) : \rho_p(\mathbb{P}, \mathbb{P}_0) \le r\}. \tag{4}$$

The ambiguity set is parameterized by the central distribution $\mathbb{P}_0 \in \mathcal{M}(\Delta)$ and the radius of the ambiguity set $r \ge 0$. For the purpose of this paper, we assume that $\mathbb{P}_0$ and r are fixed and known. As to how they might be inferred from data, is an active area of research in and of itself [26], [27], and beyond the scope of this paper.

Taking the ambiguity set $\mathcal{P}$ to be defined according to (4), we describe the feasible set of the corresponding ACCP defined in (3) as

$$\mathcal{X}_\epsilon^r := \left\{ x \in \mathcal{X} : \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\{f(x,\delta) \le 0\} \ge 1-\epsilon \right\}.$$

Note that in the absence of distributional ambiguity (i.e., $r = 0$), the set $\mathcal{X}_\epsilon^0$ corresponds to the feasible set of the chance constrained program (2) defined in terms of the central distribution $\mathbb{P}_0$.

B. Robust Sampled Convex Programs (RSCP)

We adopt the approach of [11], and approximate the ambiguous chance constrained program by a *robust sampled convex program* (RSCP) defined as follows. Fix an integer $N \ge n$, and let $(\delta_1, \dots, \delta_N)$ be a collection of N independent and identically distributed random variables defined according to the central distribution $\mathbb{P}_0$. Given a realization of the random variables $(\delta_1, \ldots, \delta_N)$, the corresponding RSCP is defined as

$$\begin{aligned} &\text{minimize} \quad c^\top x \\ &\text{subject to} \quad x \in \mathcal{X} \\ &\qquad f(x,z) \leq 0, \quad \forall\, z \in \bigcup_{i=1}^{N} B_r(\delta_i) \cap \Delta, \end{aligned} \tag{5}$$

where $B_r(\delta_i) := \{z \in \mathbb{R}^m : \|z - \delta_i\| \leq r\}$ is a closed ball of radius r centered at the sample δ_i. For $r > 0$, the robust sampled convex program (5) amounts to a semi-infinite convex program, which can be efficiently solved, under suitable choice of norm, for certain families of constraint functions f and uncertainty sets Δ.[3] We make the following assumption, in a similar fashion to [22],

Assumption 1. For each collection of samples $(\delta_1, \ldots, \delta_N) \in \Delta^N$, the corresponding RSCP defined in (5) has a feasible set with a nonempty interior, and a unique optimal solution.

C. Probabilistic Guarantees for RSCP

We denote the optimal solution of the RSCP by $\hat{x}_N^r \in \mathcal{X}$. It is itself a random variable, as it depends implicitly on the collection of samples used to define the feasible region of the RSCP. It is natural then to ask, what is the probability that an optimal solution to RSCP is feasible for the original ACCP? Or, instead, what is the number of samples required in order that an optimal solution to RSCP be feasible for ACCP with high probability? To a large extent, these questions have been resolved by [21], [22] for the ambiguity-free case in which $r = 0$. We briefly summarize their results within the context of our formulation.

First notice that, for $r = 0$, the RSCP (5) simplifies to the so called sampled convex program (SCP) defined as:

$$\begin{aligned} &\text{minimize} \quad c^\top x \\ &\text{subject to} \quad x \in \mathcal{X} \\ &\qquad f(x,\delta_i) \leq 0, \quad i = 1, \ldots, N. \end{aligned} \tag{6}$$

We state the following result from [21], [22] for $r = 0$.

Theorem 1. Fix $0 \leq \epsilon \leq 1$. Let Assumption 1 hold. It follows that

$$\mathbb{P}_0^N \left\{ \hat{x}_N^0 \notin \mathcal{X}_\epsilon^0 \right\} \leq \Phi(\epsilon),$$

where

$$\Phi(\epsilon) := \begin{cases} 1, & \epsilon \in (-\infty, 0], \\ \sum_{i=1}^{n-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}, & \epsilon \in (0,1], \\ 0, & \epsilon \in (1, \infty). \end{cases}$$

Essentially, Theorem 1 provides an upper bound on the probability that an optimal solution to the SCP (6) violates the (ambiguity-free) chance constraint defined in terms of the central distribution $\mathbb{P}_0$. We prove the following Proposition, which builds upon Theorem 1 to establish an upper bound on the constraint violation probability of $\hat{x}_N^r$ when the radius of the ambiguity set is greater than or equal to zero. We defer its proof to Appendix A.

Proposition 1. Fix $0 \leq \epsilon \leq 1$ and $0 \leq r$. Let Assumption 1 hold. It follows that

$$\mathbb{P}_0^N \left\{ \hat{x}_N^r \notin \mathcal{X}_\epsilon^r \right\} \leq \Phi(\epsilon - r).$$

The upper bound $\Phi(\epsilon - r)$ on the probability that a solution to RSCP violates the ambiguous chance constraint has several interesting properties. First, notice that, for $r = 0$, we recover the ambiguity-free result of Theorem 1. For ambiguity sets with positive radii $r > 0$, the upper bound on the constraint violation probability is monotonically nondecreasing in r, and converges to one as $r \to \infty$. A particular weakness of Proposition 1 derives from the fact that $\Phi(\epsilon - r) = 1$ for $r \geq \epsilon$. In other words, the upper bound provides no useful information when the radius of the ambiguity set r (defined in terms of the Prokhorov metric) exceeds the allowable risk of constraint violation ϵ.

As a trivial corollary to Proposition 1, one can establish a lower bound on the number of samples required in order that constraint violation probability does not exceed a desired level.

Corollary 1. Fix $\beta \in (0,1)$, $\epsilon \in (0,1)$, and $r \in [0, \epsilon)$. Let Assumption 1 hold. Define

$$N(\epsilon - r, \beta) := \min \left\{ N \in \mathbb{N} : \Phi(\epsilon - r) \leq \beta \right\}. \tag{7}$$

If the number of samples satisfies $N \geq N(\epsilon - r, \beta)$, then $\mathbb{P}_0^N \left\{ \hat{x}_N^r \notin \mathcal{X}_\epsilon^r \right\} \leq \beta$. Moreover, it holds that

$$N(\epsilon - r, \beta) \leq \frac{2}{\epsilon - r} \left(\ln \beta^{-1} + n \right). \tag{8}$$

We omit the proof of Corollary 1, as the inequality (8) was first shown in [28].

D. Comparison with the Literature

We briefly discuss the relationship between the bound established in Proposition 1 and known bounds for RSCPs in the literature. In Theorem 6 of [11], the authors prove that, for $0 \leq r < \epsilon < 1$,

$$\mathbb{P}_0^N \left\{ \hat{x}_N^r \notin \mathcal{X}_\epsilon^r \right\} \leq \left(\frac{eN}{n} \right)^n e^{-(\epsilon - r)(N - n)}. \tag{9}$$

First, unlike the upper bound in Proposition 1, the bound in (9) does not recover the ambiguity-free result of Theorem 1 for $r = 0$. Second, the number of samples required by (9) to ensure that $\mathbb{P}_0^N \left\{ \hat{x}_N^r \notin \mathcal{X}_\epsilon^r \right\} \leq \beta$ can far exceed the sample size requirement defined in (7). In order to make a direct comparison, we define sample size requirement implied by the bound in (9) as

$$\overline{N}(\epsilon - r, \beta) := \min \left\{ N \in \mathbb{N} : \left(\frac{eN}{n} \right)^n e^{-(\epsilon - r)(N-n)} \leq \beta \right\}.$$

We include a comparison of $N(\epsilon - r, \beta)$ and $\overline{N}(\epsilon - r, \beta)$ in Table I. In many cases, the new sample size requirement improves upon the existing one by as much as an order of magnitude.

[3] We refer the reader to [3] for a comprehensive treatment of tractable robust convex programs.

TABLE I
SAMPLE SIZE REQUIREMENTS $N(\epsilon - r, \beta)$ VERSUS $\overline{N}(\epsilon - r, \beta)$ FOR DIFFERENT VALUES OF ϵ, GIVEN $n = 10$, $r = 0.1$ AND $\beta = 10^{-5}$.

ϵ	0.15	0.125	0.11	0.105	0.1025	0.101
$N(\epsilon - r, \beta)$	581	1171	2942	5895	11799	29513
$\overline{N}(\epsilon - r, \beta)$	1434	3175	8960	19460	41986	115027

III. SAMPLED CONVEX PROGRAMS

In this section, we explore an approach to approximating the ambiguous chance constrained program (ACCP) by means of a sampled convex program (SCP), defined previously in (6). Recall that we denote the optimal solution of the SCP by $\hat{x}_N^0 \in \mathcal{X}$. Our aim is to derive an upper bound on the probability that such a solution violates the ambiguous chance constraint when the ambiguity set has radius $r > 0$. More formally, we seek upper bounds on the probability $\mathbb{P}_0^N \left\{ \hat{x}_N^0 \notin \mathcal{X}_\epsilon^r \right\}$.

In pursuit of such bounds, we abandon the use of the Prokhorov metric, and consider ambiguity sets defined according to a variety of other metrics, which we presently define. In each of the following definitions, we let $\mathbb{P}, \mathbb{Q} \in \mathcal{M}(\Delta)$ and denote by p and q their respective probability densities with respect to the Lebesgue measure μ on Δ.

- *Total variation metric,* ρ_{tv}:
$$\rho_{tv}(\mathbb{P}, \mathbb{Q}) := \sup_{A \in \mathcal{B}(\Delta)} \left| \mathbb{P}\{A\} - \mathbb{Q}\{A\} \right|.$$
- *Hellinger metric,* ρ_h:
$$\rho_h(\mathbb{P}, \mathbb{Q}) := \left(\int_\Delta \left(\sqrt{p} - \sqrt{q} \right)^2 d\mu \right)^{\frac{1}{2}}.$$
- *Relative entropy distance,* ρ_e:
$$\rho_e(\mathbb{P}, \mathbb{Q}) := \int_\Delta p \ln \left(\frac{p}{q} \right) d\mu.$$
- χ^2*-distance,* ρ_{χ^2}:
$$\rho_{\chi^2}(\mathbb{P}, \mathbb{Q}) := \int_{S(\mathbb{P}) \cup S(\mathbb{Q})} \frac{(p-q)^2}{q} d\mu,$$
where $S(\cdot)$ represents the support of the probability distribution on Δ.

We consider ambiguity sets of the form $\mathcal{P} = \{\mathbb{P} \in \mathcal{M}(\Delta) : \rho_{(\cdot)}(\mathbb{P}, \mathbb{P}_0) \leq r\}$ for each of the previously defined metrics $\rho_{(\cdot)}$. As a matter of notational convenience, we leave the dependency of the ambiguity set on the specific metric unspecified, unless it is otherwise unclear from the context.

A. Probabilistic Guarantees for SCP

The crux of our approach centers on the conservative approximation of the ambiguous chance constraint in (3) as an ambiguity-free chance constraint defined in terms of the central distribution $\mathbb{P}_0$. In order to do so, we first define the *perturbed risk level* $\nu_\epsilon^r \in [0,1]$ associated with the ambiguity set $\mathcal{P}$ as

$$\nu_\epsilon^r := \sup\{\alpha : \mathbb{P}_0 \{A\} \leq \alpha \Rightarrow \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{A\} \leq \epsilon, \ \forall A \in \mathcal{B}(\Delta)\},$$

where we define $\nu_\epsilon^r = 0$ if the above problem is infeasible. Clearly, it holds that $\nu_\epsilon^r \leq \epsilon$. And, although left implicit, we emphasize that the perturbed risk level depends critically on the underlying distance function used to define the ambiguity set $\mathcal{P}$. Using the perturbed risk level of the ambiguity set, one can convert ACCP into an ambiguity-free chance constrained program under the central distribution $\mathbb{P}_0$. The resulting chance constrained program can then be directly approximated using SCP. Using the perturbed risk level ν_ϵ^r, we establish an upper bound on the probability that solutions to SCP violate the ambiguous chance constraint.

Lemma 1. Fix $0 \leq \epsilon \leq 1$ and $0 \leq r$. Let Assumption 1 hold for the corresponding SCP defined in (6). It follows that
$$\mathbb{P}_0^N \left\{ \hat{x}_N^0 \notin \mathcal{X}_\epsilon^r \right\} \leq \Phi(\nu_\epsilon^r).$$
Moreover, it holds that $\Phi(\nu_\epsilon^r) \leq \Phi(\nu)$ for all $\nu \leq \nu_\epsilon^r$.

We defer the proof of Lemma 1 to Appendix B. We remark that the calculation of perturbed risk level ν_ϵ^r can be computationally challenging. However, Lemma 1 suggests that a lower bound on ν_ϵ^r is sufficient to obtain an upper bound on the constraint violation probability of $\hat{x}_N^0$. In Proposition 2, we establish several such lower bounds for a variety of distance functions. We omit the proof of Proposition 2 due to space constraints.

Proposition 2. Fix $0 \leq \epsilon \leq 1$ and $0 \leq r$. For each of the following distance functions, the corresponding perturbed risk level ν_ϵ^r satisfies the lower bound:

(a) *Total variation metric,* ρ_{tv}:
$$\nu_\epsilon^r \geq \epsilon - r.$$
(b) *Hellinger metric,* ρ_h:
$$\nu_\epsilon^r \geq \max\left(\sqrt{\epsilon} - r, \ 0 \right)^2.$$
(c) *Relative entropy distance,* ρ_e:
$$\nu_\epsilon^r \geq \sup_{\lambda > 0} \frac{e^{-r}(\lambda+1)^\epsilon - 1}{\lambda}.$$
(d) χ^2*-distance,* ρ_{χ^2}:
$$\nu_\epsilon^r \geq \epsilon + \frac{r}{2} - \sqrt{r\epsilon + \frac{r^2}{4}}.$$

We remark that the lower bound in (c) was shown to hold with equality in [18]. In Figure 1, we plot the lower bounds on the perturbed risk level for each of the distance functions considered.

B. Sample Size Requirements

We briefly discuss the sample size requirements implied by Lemma 1 and Proposition 2. Given a desired confidence level $\beta \in (0,1)$, it follows from Lemma 1 that
$$N \geq N(\nu_\epsilon^r, \beta) \quad \Longrightarrow \quad \mathbb{P}_0^N \left\{ \hat{x}_N^0 \notin \mathcal{X}_\epsilon^r \right\} \leq \beta.$$
One can further bound $N(\nu_\epsilon^r, \beta)$ from above by using the lower bounds on the perturbed risk level ν_ϵ^r specified in Proposition 2. We include, in Table II, a list of the resulting

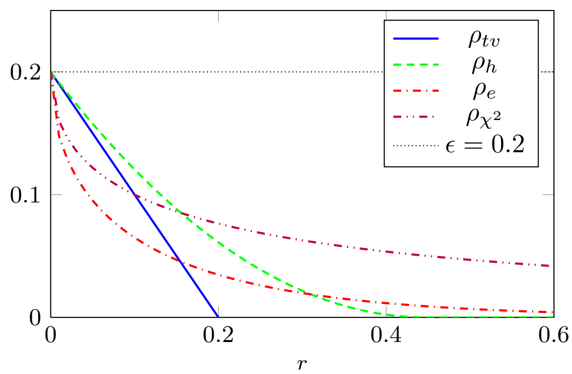

Fig. 1. Plot of lower bound from Proposition 2 on the perturbed risk level ν_ϵ^r versus r for $\epsilon = 0.2$. Each curve corresponds to a different distance function.

TABLE II

UPPER BOUNDS ON THE SAMPLE SIZE REQUIREMENT $N(\nu_\epsilon^r, \beta)$ VERSUS ϵ FOR AMBIGUITY SETS DEFINED IN TERMS THE TOTAL VARIATION METRIC (N_{tv}), HELLINGER METRIC (N_h), RELATIVE ENTROPY DISTANCE (N_e), AND χ^2-DISTANCE (N_{χ^2}). HERE, $n = 10$, $r = 0.1$ AND $\beta = 10^{-5}$.

ϵ	0.2	0.15	0.125	0.11	0.105	0.1025	0.101
N_{tv}	285	581	1171	2942	5895	11799	29513
N_h	235	348	449	540	578	599	612
N_e	444	762	1098	1438	1591	1678	1734
N_{χ^2}	285	426	552	664	711	736	752
N_0	137	187	226	258	271	278	282

sample size requirements versus ϵ, for $n = 10$, $r = 0.1$ and $\beta = 10^{-5}$. We denote by N_{tv}, N_h, N_e, and N_{χ^2} the sample size requirement associated with the total variation metric, Hellinger metric, relative entropy distance, and χ^2-distance, respectively. Finally, $N_0 := N(\epsilon, \beta)$ denotes the sample size required by the ambiguity-free setting ($r = 0$).

Table II reveals that, for a fixed radius r, the choice of metric used to define the ambiguity set can have a dramatic effect on the resulting sample size requirement. In particular, for $\epsilon = .101$, the sample size requirement implied by the total variation metric is two orders of magnitude larger than the sample size requirement associated the χ^2-distance or Hellinger metric.

IV. CONCLUSION

The results presented in this paper rely on the assumption of a fixed and known ambiguity set. In practice, such ambiguity sets would have to be inferred from data. Accordingly, it would be of interest to explore the extent to which the techniques and results presented in this paper might be extended to the setting in which the optimizer is required to construct an ambiguity set given access to only a *limited number* of IID samples drawn from the true underlying distribution. One natural approach might entail the specification of the ambiguity set as a ball of distributions centered around the empirical distribution associated with the given samples. The greater the sample size, the more accurate is the empirical distribution, and the smaller is the radius of the implied ambiguity set – probabilistically speaking that is. It is therefore natural to ask as to whether the results presented in this paper might be generalized to accommodate ambiguity sets to which the true distribution is known to belong with high probability.

APPENDIX

A. *Proof of Proposition 1*

The case of $\epsilon - r \le 0$ is trivial, as we can upper bound the constraint violation probability by $1 = \Phi(\epsilon - r)$. For the remainder of the proof, assume instead that $\epsilon - r > 0$.

Fix $x \in \mathcal{X}$. The constraint (5) is equivalent to

$$\sup_{z \in \bigcup_{i=1}^N B_r(\delta_i) \cap \Delta} f(x, z) \le 0,$$

which can also be represented as

$$\max_{i=1,\cdots,N} \left\{ \sup_{z \in B_r(\delta_i) \cap \Delta} f(x, z) \right\} \le 0.$$

Define $g(x, \delta) = \sup_{z \in B_r(\delta) \cap \Delta} f(x, z)$. Since $f(x, z)$ is closed and convex in x for each z, the function $g(x, \delta)$ is also convex and closed in x for each $\delta \in \Delta$. We can therefore equivalently reformulate the RSCP (5) as a SCP given by

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & x \in \mathcal{X} \\ & g(x, \delta_i) \le 0, \quad i = 1, \ldots, N, \end{aligned}$$

It follows from Assumption 1 that the resulting SCP has a feasible set with a nonempty interior, and a unique optimal solution $\hat{x}_N^r$.

Let $A = \{\delta \in \Delta : f(x, \delta) > 0\} \in \mathcal{B}(\Delta)$. In what follows, we show that $A^r \subseteq \{\delta \in \Delta : g(x, \delta) > 0\}$. By definition, we have that

$$y \in A^r \quad \Leftrightarrow \quad \exists\, z \in A, \ \|y - z\| < r.$$

This means that there exists $z \in B_r(y) \cap A \subseteq B_r(y) \cap \Delta$ such that $f(x, z) > 0$. It follows that $g(x, y) \ge f(x, z) > 0$ and $y \in \{\delta \in \Delta : g(x, \delta) > 0\}$.

According to the definitions of the Prokhorov metric and the ambiguity set, we have

$$\mathbb{P}_0 \{g(x, \delta) > 0\} + r \ge \mathbb{P} \{f(x, \delta) > 0\}$$

for all $\mathbb{P} \in \mathcal{P}$. It follows that

$$\begin{aligned} \hat{x}_N^r \notin \mathcal{X}_\epsilon^r &\Leftrightarrow \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \{f(\hat{x}_N^r, \delta) > 0\} > \epsilon \\ &\Rightarrow \mathbb{P}_0 \{g(\hat{x}_N^r, \delta) > 0\} + r > \epsilon. \end{aligned}$$

Hence,

$$\mathbb{P}_0^N \{\hat{x}_N^r \notin \mathcal{X}_\epsilon^r\} \le \mathbb{P}_0^N \{\mathbb{P}_0 \{g(\hat{x}_N^r, \delta) > 0\} > \epsilon - r\},$$

and the proposition follows from Theorem 1.

B. Proof of Lemma 1

To prove the lemma, we consider two cases separately: (i) $\nu_\epsilon^r = 0$, and (ii) $\nu_\epsilon^r > 0$.

(i) The case of $\nu_\epsilon^r = 0$ is trivial, as $\mathbb{P}_0^N\left\{\hat{x}_N^0 \notin \mathcal{X}_\epsilon^r\right\} \le 1 = \Phi\left(\nu_\epsilon^r\right)$.

(ii) We now prove the lemma for the case of $\nu_\epsilon^r > 0$. When $\hat{x}_N^0 \notin \mathcal{X}_\epsilon^r$, it holds that $\exists \mathbb{P} \in \mathcal{P}$ such that

$$\mathbb{P}\left\{f(\hat{x}_N^0, \delta) > 0\right\} > \epsilon.$$

Hence, it follows from the definition of the perturbed risk level ν_ϵ^r that

$$\mathbb{P}_0\left\{f(\hat{x}_N^0, \delta) > 0\right\} > \nu \qquad \forall\ \nu_\epsilon^r > \nu > 0$$

As a result,

$$\begin{aligned}\mathbb{P}_0^N\left\{\hat{x}_N^0 \notin \mathcal{X}_\epsilon^r\right\} \le&\ \mathbb{P}_0^N\left\{\mathbb{P}_0\left\{f(\hat{x}_N^0, \delta) > 0\right\} > \nu\right\}\\ =&\ \mathbb{P}_0^N\left\{\hat{x}_N^0 \notin \mathcal{X}_\nu^0\right\}\\ \le&\ \Phi(\nu)\end{aligned}$$

The last inequality follows from Theorem 1. Since the function Φ is continuous, we have that

$$\mathbb{P}_0^N\left\{\hat{x}_N^0 \notin \mathcal{X}_\epsilon^r\right\} \le \lim_{\nu \to \nu_\epsilon^r} \Phi(\nu) = \Phi(\nu_\epsilon^r).$$

Finally, since Φ is a monotonic non-increasing function, we have that $\Phi(\nu_\epsilon^r) \le \Phi(\nu)$ for all $\nu \le \nu_\epsilon^r$, thus completing the proof.

REFERENCES

[1] G. Calafiore and M. C. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, no. 1, pp. 25–46, 2005.

[2] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.

[4] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM review*, vol. 53, no. 3, pp. 464–501, 2011.

[5] A. Charnes, W. W. Cooper, and G. H. Symonds, "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil," *Management Science*, vol. 4, no. 3, pp. 235–263, 1958.

[6] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Management science*, vol. 6, no. 1, pp. 73–79, 1959.

[7] B. L. Miller and H. M. Wagner, "Chance constrained programming with joint constraints," *Operations Research*, vol. 13, no. 6, pp. 930–945, 1965.

[8] A. Prekopa, "On probabilistic constrained programming," in *Proceedings of the Princeton symposium on mathematical programming*. Princeton University Press Princeton, NJ, 1970, pp. 113–138.

[9] A. Prékopa, *Stochastic programming*. Springer Science & Business Media, 2013, vol. 324.

[10] A. P. Ruszczynski and A. Shapiro, *Stochastic programming*. Elsevier Amsterdam, 2003, vol. 10.

[11] E. Erdoğan and G. Iyengar, "Ambiguous chance constrained problems and robust optimization," *Mathematical Programming*, vol. 107, no. 1-2, pp. 37–61, 2006.

[12] I. Yanıkoğlu and D. den Hertog, "Safe approximations of ambiguous chance constraints using historical data," *INFORMS Journal on Computing*, vol. 25, no. 4, pp. 666–681, 2012.

[13] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, no. 1-2, pp. 167–198, 2013.

[14] G. C. Calafiore and L. El Ghaoui, "On distributionally robust chance-constrained linear programs," *Journal of Optimization Theory and Applications*, vol. 130, no. 1, pp. 1–22, 2006.

[15] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations research*, vol. 58, no. 3, pp. 595–612, 2010.

[16] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Operations research*, vol. 58, no. 4-part-1, pp. 902–917, 2010.

[17] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, "A distributionally robust perspective on uncertainty quantification and chance constrained programming," *Mathematical Programming*, vol. 151, no. 1, pp. 35–62, 2015.

[18] Z. Hu and L. J. Hong, "Kullback-leibler divergence constrained distributionally robust optimization," *Optimization Online*, 2013. [Online]. Available: http://www.optimization-online.org/DB_FILE/2012/3677.pdf

[19] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, pp. 1–37, 2012.

[20] G. C. Pflug, A. Pichler, and D. Wozabal, "The 1/n investment strategy is optimal under high model ambiguity," *Journal of Banking & Finance*, vol. 36, no. 2, pp. 410–417, 2012.

[21] G. C. Calafiore, "Random convex programs," *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3427–3464, 2010.

[22] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1211–1230, 2008.

[23] D. P. De Farias and B. Van Roy, "On constraint sampling in the linear programming approach to approximate dynamic programming," *Mathematics of operations research*, vol. 29, no. 3, pp. 462–478, 2004.

[24] P. Billingsley, *Convergence of probability measures*. John Wiley & Sons, 2013.

[25] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[26] E. Boissard, "Simple bounds for the convergence of empirical and occupation measures in 1-wasserstein distance," *Electronic Journal of Probability*, vol. 16, pp. 2296–2333, 2011.

[27] N. Fournier and A. Guillin, "On the rate of convergence in wasserstein distance of the empirical measure," *Probability Theory and Related Fields*, vol. 162, no. 3-4, pp. 707–738, 2015.

[28] G. C. Calafiore, "On the expected probability of constraint violation in sampled convex programs," *Journal of Optimization Theory and Applications*, vol. 143, no. 2, pp. 405–412, 2009.