2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Nonparametric Estimation of Uncertainty Sets for Robust Optimization

Polina Alexeenko† and Eilyan Bitar†

***Abstract*— We investigate a data-driven approach to constructing uncertainty sets for robust optimization problems, where the uncertain problem parameters are modeled as random variables whose joint probability distribution is not known. Relying only on independent samples drawn from this distribution, we provide a nonparametric method to estimate uncertainty sets whose probability mass is guaranteed to approximate a given target mass within a given tolerance with high confidence. The nonparametric estimators that we consider are also shown to obey distribution-free finite-sample performance bounds that imply their convergence in probability to the given target mass. In addition to being efficient to compute, the proposed estimators result in uncertainty sets that yield computationally tractable robust optimization problems for a large family of constraint functions.**

***Index Terms*— Chance-constrained optimization, robust optimization, data-driven optimization, nonparametric estimation.**

I. Introduction

In this paper, we consider a class of optimization problems whose feasible regions are defined in terms of *chance constraints* [1] of the form

$$\Pr\left(f(x, \boldsymbol{u}) \leq 0\right) \geq \alpha. \tag{1}$$

Here, $f : \mathbb{R}^\ell \times \mathbb{R}^d \to \mathbb{R}$ denotes the constraint function, $x \in \mathbb{R}^\ell$ denotes the decision variable, and $\boldsymbol{u}$ is an $\mathbb{R}^d$-valued random vector that reflects uncertainty in the constraint parameters. The chance constraint (1) requires that the decision variable satisfy the constraint $f(x, \boldsymbol{u}) \leq 0$ with probability no smaller than $\alpha \in [0, 1]$.

Chance constrained optimization problems are challenging to solve for a variety of reasons. First, their feasible regions are generally nonconvex [2]. For example, chance constrained problems have been shown to be NP-hard even in the most basic setting where the constraint function is affine in both the decision variable and uncertain parameters [3]. Furthermore, verifying the feasibility of a candidate solution to a chance constrained problem is difficult, because it involves the evaluation of a multivariate integral which can be computationally expensive in high dimensions [2]. To complicate matters further, the underlying distribution according to which the random vector is distributed may be unknown.

A variety of approaches to the solution of chance constrained problems have been explored in the literature. There are several works that develop exact convex reformulations of chance constrained problems [4]–[8] and their distributionally robust generalizations [9]–[11]. These reformulations rely on a variety of assumptions, including restrictions on the family of probability distributions according to which the random vector is distributed and the functional form of the chance constraints. Ultimately, many problems of practical interest may fail to satisfy these narrow structural and distributional assumptions.

Because of the rarity of problem instances amenable to exact convex reformulation, there is another line of research focused on the design of approximation methods for chance constrained problems. One approach that has been explored extensively involves the construction of explicit convex inner (conservative) approximations to chance constrained optimization problems [12]–[17]. Again, a potential drawback of these methods stems from their reliance (to varying degrees) on knowledge of certain features of the underlying distribution, e.g., support or moment information. Data-driven approximation methods seek to alleviate the reliance on distributional assumptions that may be overly stringent or difficult to verify in practice. Instead, they utilize data sampled from the underlying distribution. For instance, the sample average approximation method [2] involves selecting an optimal subset of the sampled data that has empirical probability mass no smaller than the target reliability level α. While this approach may be less conservative than other data-driven methods, it gives rise to approximations in the form of mixed-integer optimization problems—which become computationally intractable to solve in the large sample regime.

There is another stream of literature focusing on data-driven approximations that give rise to tractable convex optimization problems. A particular category of methods uses the given data to construct estimates of the unknown distribution or its lower-order moments [9], [18]–[21]. These estimates, together with suitably defined confidence regions, give rise to distributionally robust approximations to the original chance constrained program that, in some instances, admit tractable convex reformulations.

Another family of data-driven methods known as *scenario approximations* have also been widely studied as another tractable alternative to the approximation of chance constrained problems [22], [23] and their distributionally robust counterparts [24], [25]. Specifically, scenario-based methods approximate the chance constraint (1) with n sampled constraints of the form

$$f(x, \boldsymbol{u}_i) \leq 0, \; i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ is an i.i.d. sample drawn from the unknown distribution. In addition to being distribution-free, scenario-based methods do not impose restrictions on the

This work was supported in part by NSF grant ECCS-1351621, NSF grant IIP-1632124, and an NSF Graduate Research Fellowship.

† Polina Alexeenko and Eilyan Bitar are with the School of Electrical and Computer Engineering, Cornell Univ., Ithaca, NY, 14853, USA. Emails: {pa357, eyb5}@cornell.edu

978-1-7281-7447-1/20/$31.00 ©2020 IEEE

functional form of the constraint function beyond requiring its convexity in the decision variable. However, these methods require a sample size that is at least $O\left(\frac{\ell}{1-\alpha}\ln\left(\frac{1}{\delta}\right)\right)$ in order to yield a solution that is feasible for the original chance constrained problem with probability at least $1-\delta$. While the dependence of this sample size requirement on the problem dimension ℓ can be improved for problems that are not fully supported (e.g., [23], [26], [27]), the bound is tight in general. Therefore, when the number of decision variables is large or when the reliability needed is high, scenario-based approximations can become computationally challenging to solve due to the large number of constraints that must be enforced.

A. Contribution and Related Work

Our approach is predicated on the conservative approximation of (1) in the form of a robust constraint:

$$f(x,u) \leq 0 \quad \forall u \in \mathcal{U}, \tag{3}$$

where the *uncertainty set* $\mathcal{U} \subset \mathbb{R}^d$ is constructed from data in a manner that ensures its satisfaction of the mass requirement $\Pr(\boldsymbol{u} \in \mathcal{U}) \geq \alpha$ with high probability. In this paper, we provide a nonparametric method to construct such uncertainty sets from data in a manner that guarantees that their probability mass is within a given tolerance of the target mass α with high confidence. The uncertainty sets that we propose are trivial to compute, satisfy distribution-free finite-sample statistical guarantees, and give rise to robust constraints (3) that are computationally tractable for a large family of constraint functions. Additionally, the proposed methodology allows for the explicit representation of nonconvex uncertainty sets as finite unions of convex sets—e.g., as the union of $m \in \mathbb{N}$ balls centered at sampled data points. In this manner, the resulting robust constraint (3) can be equivalently reformulated as a finite intersection of m simpler robust constraints, where each robust constraint is defined in terms of a convex uncertainty set. Importantly, the geometry of the underlying convex sets can be tailored to accommodate the structure of the constraint function to ensure the tractability of the resulting robust constraints. As another degree of freedom in the parameterization of these sets, the user is free to specify the number of constitutive sets used in the representation of the uncertainty set, and, hence, the number of robust constraints that must be enforced in the approximation. This provides the user with some degree of control over the size of the resulting robust optimization problem, unlike scenario approximation methods.

It is important to mention that the probabilistic coverage guarantees accompanying the class of uncertainty sets proposed in this paper require that the random vector $\boldsymbol{u}$ have a continuous distribution (cf. Assumption 1). The guarantees accompanying scenario approximation methods do not require such assumptions. We also note that, while we suggest a particular class of uncertainty sets in Section II-B, the probabilistic coverage guarantees provided in Section III hold for a more general family of uncertainty sets—e.g., those that satisfy Assumption 1.

Our approach is closely related to a class of existing methods that utilize data to construct uncertainty sets that yield tractable robust approximations to chance-constrained optimization problems. For example, Margellos et al. [28] develop a data-driven approach to estimate intervals that cover each component of the random vector with high probability, resulting in uncertainty sets that take the form of hyperrectangles. Bertsimas et al. [29] utilize statistical hypothesis tests to construct convex uncertainty sets from data. However, due to the convex geometry of the uncertainty sets produced by these methods, they may be limited in terms of their ability to accurately describe nonconvex high probability regions that reflect multimodality in the underlying distribution. Because of the limited expressiveness of these uncertainty sets, they may result in overly conservative approximations to the original chance constraint. Closer to the approach adopted in this paper, Campbell and How [30] consider representations in the form of unions of ellipsoids. A potential drawback of their approach, however, stems from the reliance of their theoretical results on the restrictive assumption that the unknown distribution belong to a family of Dirichlet process Gaussian mixtures.

The problem of estimating high probability sets from data also has similarities to minimum-volume (MV) set estimation problems. MV-set estimators based on the estimation of density level sets [31]–[34] typically yield nonconvex sets that are not readily expressible as finite unions of convex sets—limiting their tractability from a robust optimization perspective. More broadly, MV-set estimation methods include those based on nonparametric set estimation [35], [36], empirical quantile-based estimation [37], [38], classification via support vector machines [39], [40], and network flow-based methods [41]. With the exception of [38], the methods presented in these papers lack explicit finite-sample guarantees on their performance. A practical limitation of the algorithms proposed by Scott et al. [38], however, is that the computation of their MV-set estimates is intractable in high dimensions.

B. Organization

The remainder of the paper is organized as follows. In Section II, we introduce a data-driven method to construct uncertainty sets with a given probability mass. In Section III, we provide finite-sample statistical guarantees on the performance of the proposed class of estimators. Section IV illustrates the behavior of the estimators with several numerical experiments, and Section V concludes the paper.

C. Notation

We employ the following notational conventions throughout the paper. Let $\mathbb{R}$ and $\mathbb{Z}$ denote the sets of real numbers and integers, respectively. Given a positive integer $n \in \mathbb{Z}$, we let $[n] := \{1,\ldots,n\}$ denote the set of the first n positive integers. Given a real number $x \in \mathbb{R}$, we denote its *ceiling* by $\lceil x \rceil := \min\{n \in \mathbb{Z} \mid n \geq x\}$. Throughout, we use boldface symbols to denote random variables, and non-boldface symbols to denote particular values in the range of

a random variable and other deterministic quantities. We use $\Pr(A)$ to denote the probability of an event A.

II. Uncertainty Set Estimation

In this section, we provide a data-driven method to construct uncertainty sets whose probability mass is both close to and no smaller than a given target mass $\alpha \in (0,1)$ with high probability, based only on a finite random sample drawn from the unknown distribution of interest. In addition to the mass requirement, we seek representations that result in computationally tractable robust constraints.

The remainder of this section is organized as follows. In Sec. II-A, we provide a general approach to learning α-probability sets via the estimation of sublevel sets for a general class of continuous functions. In Sec. II-B, we introduce a particular class of functions whose level sets (expressible as finite unions of convex sets) yield tractable robust constraints for several important families of constraint functions. We close with a discussion on the connections between the proposed family of approximations and scenario-based approximation methods.

A. Estimation via Level Sets

With the aforementioned objectives in mind, we consider a family of uncertainty sets defined as r-sublevel sets[1] of a given continuous function $\phi : \mathbb{R}^d \to \mathbb{R}$, i.e.,

$$\mathcal{U}_\phi(r) := \{u \in \mathbb{R}^d \mid \phi(u) \leq r\}. \tag{4}$$

The proposed family of uncertainty sets (4) is parameterized by the user-specified inputs: the *shape function* ϕ and *level* r. Intuitively, the shape function ϕ should be chosen so that it is smallest over those regions with the greatest concentration of probability mass, in order to limit the volume (Lebesgue measure) of the resulting sublevel set. While this might suggest an approach to specifying ϕ in terms of a density estimate based on an i.i.d. sample drawn from the unknown distribution, the particular functional form of ϕ must also yield uncertainty (level) sets that are computationally tractable from a robust optimization perspective—a condition that most nonparametric (kernel) density estimators fail to satisfy. In Sec. II-B, we suggest a particular functional form for the shape function ϕ that is expressive enough to represent compact sets of arbitrary form, while ensuring the tractability of the robust constraints that it yields for a large family of constraint functions. It is important to note that, in the interest of lightening notation, we have omitted the potential dependence that the function ϕ may have on data drawn from the unknown distribution of interest.

For the remainder of this section, we treat the shape function ϕ as being fixed (implicitly conditioning all probabilistic statements on the data on which it is based), and focus our discussion on the role of the level r in controlling the behavior of the sublevel sets of ϕ. Clearly, the level $r \in \mathbb{R}$ directly influences the volume and probability mass of the sublevel sets of ϕ, with larger levels resulting in sublevel sets of greater volume. Thus, with the aim of limiting the conservatism of the robust approximations (3) induced by the proposed family of uncertainty sets, we are interested in characterizing the smallest level r such that the target mass requirement is satisfied. We refer to this as the *α-covering level*, which we formally define as follows.

Definition 1 (α-covering level): Let $\alpha \in (0,1)$. The *α-covering level* is defined as

$$\rho(\alpha) := \inf\{r \in \mathbb{R} \mid \Pr(\boldsymbol{u} \in \mathcal{U}_\phi(r)) \geq \alpha\}.$$

As one of the objectives of this paper, we are interested in constructing an estimator for the α-covering level $\rho(\alpha)$ based only on an i.i.d. training sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ drawn from the unknown distribution. Before introducing the particular family of estimators that we consider in this paper, it will be informative to express the α-covering level in terms of the quantile

$$\rho(\alpha) = F^{-1}(\alpha),$$

where $F^{-1}(\alpha) := \inf\{z \in \mathbb{R} \mid \Pr(\boldsymbol{\zeta} \leq z) \geq \alpha\}$ denotes the α-quantile of the transformed random variable $\boldsymbol{\zeta} := \phi(\boldsymbol{u})$. This reformulation suggests a natural estimator for $\rho(\alpha)$ in the form of an empirical quantile

$$F_n^{-1}(\alpha_n) := \inf\{z \in \mathbb{R} \mid F_n(z) \geq \alpha_n\}. \tag{5}$$

Here, $\alpha_n \in (0,1)$ is a sequence of probability levels, and $F_n(z) := (1/n)\sum_{i=1}^n \mathbf{1}\{\boldsymbol{\zeta}_i \leq z\}$ denotes the empirical distribution function associated with the transformed training sample

$$\boldsymbol{\zeta}_i := \phi(\boldsymbol{u}_i)$$

for $i = 1, \ldots, n$.

In Section III, we investigate the role of α_n in controlling the limiting behavior and rate of convergence of the proposed empirical quantile to the α-covering level. In particular, we establish minimum training sample size requirements on n ensuring that the probability mass covered by the resulting uncertainty set $\mathcal{U}_\phi(F_n^{-1}(\alpha_n))$ is within a given tolerance of (and no smaller than) the given target mass α with high probability.

B. Shape Functions ϕ Compatible with Robust Optimization

The functional form of the shape function ϕ (and the geometry of its sublevel sets) will play a critical role in governing the tractability of the family of robust constraints that it gives rise to. With this in mind, we suggest a class of shape functions whose sublevel sets are rich enough to represent compact sets of arbitrary form, while being simple enough to ensure the tractability of the resulting robust constraints. Specifically, we propose a particular class of shape functions whose sublevel sets admit explicit representations as finite unions of regular convex sets, e.g., p-norm balls. In this manner, the resulting robust constraint (3) can be equivalently reformulated as a finite intersection of simpler robust constraints, where each robust constraint is defined in terms of a convex uncertainty set. Being expressed as norm balls, the particular geometry of the underlying convex sets (e.g., ℓ_1 versus ℓ_2) can be tailored to accommodate the

[1]The decision to specify (4) in terms of sublevel sets of ϕ, as opposed to superlevel sets, is made for notational convenience.

structure of the given constraint function to facilitate the tractability of the corresponding robust constraint. In addition to determining the number of constraints, the number of sets in the union will also influence the volume of the uncertainty sets that we learn, and hence the conservatism of the resulting robust constraint approximations. Therefore, in treating the number of sets in the approximation as a user-specified parameter, one can tradeoff the computational complexity of the resulting optimization problem against the quality (conservatism) of its solutions.

With these design criteria in mind, we now introduce the specific class of estimators we consider. Given an i.i.d. sample $\bar{\boldsymbol{u}}_1, \ldots, \bar{\boldsymbol{u}}_m$ drawn from the unknown distribution of interest, define

$$\phi(u) := \min_{i \in [m]} \|u - \bar{\boldsymbol{u}}_i\|, \tag{6}$$

where $\|\cdot\|$ denotes the p-norm ($p \geq 1$) on $\mathbb{R}^d$. The specification of the shape function according to (6) gives rise to a class of uncertainty sets that are defined as a union of balls with a common radius, where each ball is centered at a randomly sampled point

$$\mathcal{U}_\phi(r) = \bigcup_{i=1}^{m} B(\bar{\boldsymbol{u}}_i, r). \tag{7}$$

Here $B(u, r) := \{v \in \mathbb{R}^d \mid \|u - v\| \leq r\}$ denotes the closed ball of radius $r \geq 0$ centered at $u \in \mathbb{R}^d$. The parameter r plays an analogous role to that of bandwidth parameters in the context of nonparametric kernel density estimation.[2]

It is important to note that we have augmented the notation used to denote the sample $(\bar{\boldsymbol{u}}_1, \ldots, \bar{\boldsymbol{u}}_m)$ (which we refer to as the *shape sample*), in order to distinguish its specification from the *training sample* $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ that is subsequently used to estimate the α-covering level. The two samples, which are assumed to be independent from one another, play distinct roles in the construction of the uncertainty set. The shape sample determines the positions of the balls comprising the uncertainty set, while the training sample is used to calibrate the radius r of each ball in a manner guaranteeing that the probability mass covered by the resulting uncertainty set is within a given tolerance of the target probability mass with high confidence.

Given an uncertainty set $\mathcal{U}_\phi(r)$ defined according to (7), one can equivalently reformulate the robust constraint (3) as the intersection of m robust constraints of the form

$$f(x, u) \leq 0 \quad \forall u \in B(\bar{\boldsymbol{u}}_i, r), \quad i = 1, \ldots, m. \tag{8}$$

Robust constraints defined in terms of uncertainty sets of this form result in computationally tractable optimization problems for a large family of constraint functions. For example, if the given constraint function is a *bi-affine function* of the form $f(x, u) = x^\top u - b$, then it is straightforward to show that the collection of robust constraints specified in (8) can be equivalently reformulated as convex constraints given by

$$x^\top \bar{\boldsymbol{u}}_i + r\|x\|_* \leq b, \quad i = 1, \ldots, m, \tag{9}$$

where $\|x\|_* := \sup\{x^\top y \mid \|y\| \leq 1\}$ denotes the *dual norm* associated with a given p-norm $\|\cdot\|$.

We refer the reader to [42] and [43] for a more comprehensive discussion surrounding the different families of robust constraints that admit equivalent reformulations as tractable convex constraints.

C. Comparison to Scenario Approximation Methods

The approach considered here gives rise to optimization problems that are similar in structure to those based on scenario approximations of chance constrained problems and their distributionally robust generalizations. Specifically, when the radius r of the balls composing the uncertainty sets is equal to zero, we recover the standard scenario approximations to chance constrained problems such as those presented in [22], [23]. When $r > 0$, we recover the robust scenario approximation to distributionally robust chance constrained problems studied in [24], [25]. Interestingly, the latter connection suggests that the class of approximations considered in this paper also possesses an intrinsic distributional robustness—a point that merits further examination as part of future research.

Although the class of approximations that we propose is similar in structure to those based on scenario approximation methods, our method offers an important computational advantage in terms of the number of constraints that must be enforced in the resulting approximation to the original chance constraint. Specifically, while scenario approximation methods result in approximations based on a number of sampled constraints that is at least $O(\frac{\ell}{1-\alpha} \ln\left(\frac{1}{\delta}\right))$, the method proposed in this paper treats the number of constraints m in the resulting robust approximation as a parameter that can be set by the user. Importantly, our theoretical guarantees hold for any value of m, allowing the user to determine the number of constraints that must be enforced in the resulting approximation to the original chance constraint.

It is also important to note that the class of approximations proposed in this paper is accompanied by certain limitations, as compared to scenario approximation methods. In particular, while scenario approximations (2) only require that the constraint function be convex in the decision variable, the tractability of our approximations (8) requires additional restrictions on the class of constraint functions that can be efficiently optimized over—e.g., that the constraint functions are also concave with respect to the uncertain variable.

III. Finite-Sample Statistical Guarantees

We now establish a bound on the rate at which the probability mass of the proposed class of uncertainty sets converges to the target mass as a function of the training sample size n. It is important to note that the theoretical results presented in this section hold not only for the class of uncertainty sets introduced in Section II-B, but more

[2]Nonparametric estimators defined as the union of balls centered on the sampled data have been previously studied in the context of distribution support estimation [35], [36].

generally for any family of uncertainty sets induced by shape functions that satisfy Assumption 1. To lighten the notation in the following discussion, we denote the probability mass of the uncertainty set $\mathcal{U}_\phi(r)$ conditioned on a level r by

$$\pi_\phi(r) := \Pr\left(\boldsymbol{u} \in \mathcal{U}_\phi(r)\right). \tag{10}$$

Throughout this section, we treat the shape function ϕ as being fixed and given. Accordingly, all subsequent probabilities that we state in this section should be interpreted as being conditioned on the given shape function ϕ. Before stating the main result of this section, we introduce a technical assumption that is assumed to hold throughout the remainder of the paper.

Assumption 1: Given a fixed shape function $\phi : \mathbb{R}^d \to \mathbb{R}$, the transformed random variable $\boldsymbol{\zeta} = \phi(\boldsymbol{u})$ is assumed to have a continuous distribution.

In particular, Assumption 1 is satisfied by any shape function ϕ constructed according to (6) if the underlying random vector $\boldsymbol{u}$ is continuous. With Assumption 1 in hand, we have the following result.

Theorem 1: Let $\varepsilon \in (0, 1-\alpha)$ and $\alpha_n \in (\alpha, \alpha+\varepsilon)$ for all $n \geq 1$. It follows that

$$\Pr\left(\pi_\phi(F_n^{-1}(\alpha_n)) < \alpha\right) \leq \exp\left(\frac{-n(\alpha_n - \alpha)^2}{2(1-\alpha)}\right) \tag{11}$$

and

$$\Pr\left(\pi_\phi(F_n^{-1}(\alpha_n)) > \alpha + \varepsilon\right) \leq \exp\left(\frac{-n(\alpha + \varepsilon - \alpha_n)^2}{2(\alpha+\varepsilon)}\right). \tag{12}$$

It follows that, for any fixed tolerance $\varepsilon \in (0, 1-\alpha)$, the probability that the mass covered by the uncertainty set $\mathcal{U}_\phi(F_n^{-1}(\alpha_n))$ is within ε of (and no less than) the target mass α approaches one at an exponential rate in the sample size. One may also use this result to establish conditions on the sequence α_n guaranteeing that $\pi_\phi(F_n^{-1}(\alpha_n)) \to \alpha$ in probability as the sample size n tends to infinity.

We note that the proof of Theorem 1 utilizes a number of standard arguments that have been used to establish several related results in the literature [44]–[46]. The crux of the main argument involves the reformulation of the error probabilities (11) and (12) as binomial tail probabilities, which are then bounded from above using a variant of the Chernoff bound.

Proof: All probabilities stated in this proof are assumed to be conditioned on the given shape function ϕ. We first prove inequality (11). First notice that, conditioned on a level r, it holds that

$$\pi_\phi(r) = \Pr\left(\phi(\boldsymbol{u}) \leq r\right) = F(r),$$

where F denotes the cumulative distribution function of the transformed random variable $\boldsymbol{\zeta} := \phi(\boldsymbol{u})$. This identity, combined with the assumed continuity of the distribution F, implies that

$$\Pr\left(\pi_\phi(F_n^{-1}(\alpha_n)) < \alpha\right) = \Pr\left(F_n^{-1}(\alpha_n) < F^{-1}(\alpha)\right).$$

The empirical quantile function can be expressed in terms of the *order statistics* of the transformed training sample, which we denote by $\boldsymbol{\zeta}_{(1)} \leq \boldsymbol{\zeta}_{(2)} \leq \cdots \leq \boldsymbol{\zeta}_{(n)}$. Specifically, it holds that

$$F_n^{-1}(\gamma) = \boldsymbol{\zeta}_{(\lceil n\gamma \rceil)} \tag{13}$$

for all $\gamma \in (0,1)$. Importantly, the cumulative distribution function of the k-th order statistic can be expressed according to the upper tail of a binomial:

$$\Pr\left(\boldsymbol{\zeta}_{(k)} \leq z\right) = \sum_{i=k}^{n} \binom{n}{i} F(z)^i (1-F(z))^{n-i} \tag{14}$$

for each $k = 1, \ldots, n$ [47]. It follows that

$$\begin{aligned}
&\Pr\left(F_n^{-1}(\alpha_n) < F^{-1}(\alpha)\right) \\
&\quad = \Pr\left(\boldsymbol{\zeta}_{(\lceil n\alpha_n \rceil)} < F^{-1}(\alpha)\right) \\
&\quad = \sum_{i=\lceil n\alpha_n \rceil}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \qquad (15) \\
&\quad = \sum_{i=0}^{n-\lceil n\alpha_n \rceil} \binom{n}{i} (1-\alpha)^i \alpha^{n-i}. \qquad (16)
\end{aligned}$$

The first equality follows from (13). The second equality follows from (14) and Assumption 1, which implies that $F(F^{-1}(\alpha)) = \alpha$. The third equality stems from an equivalent reformulation of the upper binomial tail (15) as a lower binomial tail. One can bound (16) from above using the following well-known upper bound on the lower tail of a binomial [48][Theorem 4.5], which is a direct consequence of the Chernoff bound.

Theorem 2: Let $\boldsymbol{\xi}$ be a binomial random variable with parameters $n \in \mathbb{N}$ and $p \in [0,1]$. For $k \leq np$, it holds that

$$\Pr\left(\boldsymbol{\xi} \leq k\right) \leq \exp\left(\frac{-(np-k)^2}{2np}\right).$$

Inequality (11) follows from an application of Theorem 2 to (16), where we use the fact that $n - \lceil n\alpha_n \rceil \leq n(1-\alpha_n)$.

The proof of inequality (12) is analogous in nature to the proof of (11). Using similar arguments, it is possible to show that

$$\begin{aligned}
&\Pr\left(\pi_\phi(F_n^{-1}(\alpha_n)) > \alpha + \varepsilon\right) \\
&\quad = \sum_{i=0}^{\lceil n\alpha_n \rceil - 1} \binom{n}{i} (\alpha+\varepsilon)^i (1-\alpha-\varepsilon)^{n-i}. \qquad (17)
\end{aligned}$$

Inequality (12) follows from an application of Theorem 2 to (17), where we use the fact that $\lceil n\alpha_n \rceil - 1 \leq n\alpha_n$. ■

It is also possible to use Theorem 1 to characterize a distribution-free bound on the sample size requirement ensuring that the probability mass covered by the uncertainty set satisfies the given tolerance with a given confidence $1-\delta$. We state the following corollary without proof, as it is an immediate consequence of Theorem 1.

Corollary 1: Let $\delta \in (0,1)$, $\varepsilon \in (0, 1-\alpha)$, and $\lambda \in (0,1)$. Set $\alpha_n = \alpha + \lambda\varepsilon$ for all $n \geq 1$. If

$$n \geq c(\lambda, \alpha, \varepsilon) \left(\frac{2}{\varepsilon^2}\right) \ln\left(\frac{2}{\delta}\right), \tag{18}$$

where $c(\lambda, \alpha, \varepsilon) := \max\{(1-\alpha)/\lambda^2, (\alpha+\varepsilon)/(1-\lambda)^2\}$, then

$$\Pr\left(\alpha \leq \pi_\phi(F_n^{-1}(\alpha_n)) \leq \alpha + \varepsilon\right) \geq 1 - \delta.$$

It is important to note that the sample size requirement (18) is dimension-free in that it does not depend on the dimension of the random vector or that of the decision variable. Additionally, for $\alpha \in (1/2, 1)$, it is straightforward to show that the value of λ which minimizes the sample size requirement (18) is given by

$$\lambda^* = \frac{1 - \alpha - \sqrt{(1-\alpha)(\alpha+\varepsilon)}}{1 - 2\alpha - \varepsilon}. \tag{19}$$

We also remark that it is possible to improve upon the sample size requirement (18) through a refinement of the upper bounds on the lower binomial tails (16)-(17) using arguments analogous to those in [46]. In particular, the dependence on the tolerance parameter ε can be improved to $O(1/\varepsilon)$.

IV. Experiments

In this section, we present numerical experiments illustrating certain features and properties of the proposed uncertainty set estimator. Throughout these experiments, we consider three different random vectors, each of which is distributed according to a distinct Gaussian mixture.

A. Examining the Role of m

Figure 1 illustrates the impact of the number of balls m on the shape and volume of the uncertainty sets produced by our method for three different random vectors distributed according to Gaussian mixtures whose density functions are depicted in Figures 1(a), 1(f), and 1(k). In these experiments, we consider a fixed probability mass $\alpha = 0.9$, tolerance parameter $\varepsilon = 0.05$, and confidence parameter $\delta = 0.05$. In specifying the uncertainty set estimator, we set $\alpha_n = \alpha + \lambda^* \varepsilon$ for all n, where λ^* is specified according to (19). For each value of $m \in \{1, 10, 100, 1000\}$, we construct an uncertainty set $\mathcal{U}_\phi(F_n^{-1}(\alpha_n))$ according to (7) using a training sample with size satisfying the bound (18).

As expected, the plots in Figure 1 show that the proposed quantile estimate for the α-covering level, (i.e., the radius of each ball) appears to decrease with the number of balls used in the approximation of the α-probability region. And, on balance, the volume of the uncertainty sets appears to shrink with m, resulting in a potential decrease in the conservatism of the approximation to the original chance constrained problem. In particular, notice that the volume of the uncertainty set ($m = 1$) depicted in Figure 1(l) is much larger than the volume of the uncertainty set ($m = 1000$) depicted in Figure 1(o). From Figure 1(k), we see that the probability mass of the underlying distribution in this example is concentrated at four disconnected regions. Due to the separation between these high probability regions, uncertainty sets constructed according to a single ball are likely to have large volume. By contrast, notice that the difference in volume between the uncertainty sets in Figures 1(g) and 1(j) is smaller, because in that example the underlying distribution has a significant amount of probability mass concentrated in a small connected region. Furthermore, the shapes and volumes of the uncertainty sets in Figures 1(b)-1(e) do not vary significantly for different values of m, because in this case the vast majority of the probability mass is concentrated in a circular region. It is important to note, however, that as m increases, the volume of outliers also appears to increase, potentially resulting in increased conservatism of the approximation. Thus, from a practical perspective, the number of balls used in the approximation should be tuned to balance this tradeoff.

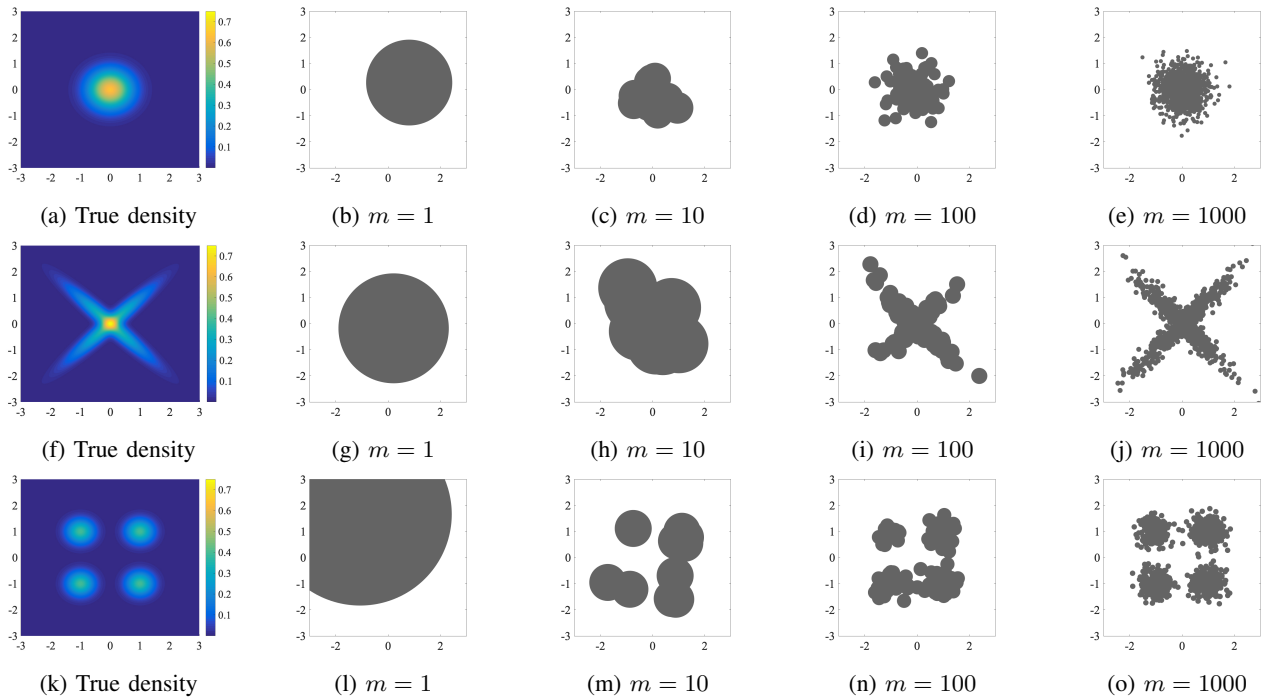
B. Illustrating Mass Consistency

In Figure 2, we conduct a Monte-Carlo analysis to illustrate the mass consistency of the proposed uncertainty set estimator by varying the tolerance parameter value. Throughout these experiments, we fix the number of balls to $m = 10$, consider a target probability mass of $\alpha = 0.9$, and consider a random vector distributed according to a Gaussian mixture whose density is depicted in Figure 1(f). We vary the tolerance parameter on a logarithmic scale between 0.005 and 0.05, and set $\alpha_n = \alpha + \lambda^* \varepsilon$, where λ^* is defined according to (19). For each value of the tolerance parameter, we chose the number of training samples to be the smallest integer satisfying the sample size requirement (18).

The numerical experiments are conducted as follows for each value of the tolerance parameter ε. First, we generate a random shape sample $\bar{\boldsymbol{u}}_1, \ldots, \bar{\boldsymbol{u}}_m$ that determines the centers of the balls composing the uncertainty set that we construct. For each experiment, we draw an i.i.d. training sample to evaluate $F_n^{-1}(\alpha_n)$ according to (5) and $\mathcal{U}_\phi(F_n^{-1}(\alpha_n))$ according to (7). We then estimate the probability mass covered by each uncertainty set $\pi_\phi\left(F_n^{-1}(\alpha_n)\right)$ using an empirical average based on one million independent samples of the random vector $\boldsymbol{u}$. We estimate empirical confidence intervals associated with $\pi_\phi\left(F_n^{-1}(\alpha_n)\right)$ using one thousand independent experiments. Figure 2 depicts the middle 90% empirical confidence interval associated with the probability mass covered by the uncertainty sets for each value of ε. Notice that, consistent with the result in Theorem 1, the fifth percentile of the probability mass covered by the uncertainty set remains above the target probability mass. Furthermore, the mass captured by the uncertainty set approaches the target probability mass with high confidence as the tolerance parameter value decreases to zero.

V. Conclusion

We provide a data-driven method to construct uncertainty sets for robust optimization problems. The estimators we consider are consistent, satisfy finite-sample performance guarantees, are efficient to compute, and give rise to tractable robust constraints for a large family of constraint functions. Furthermore, the proposed method provides the user with two mechanisms by which to control the complexity of the resulting robust optimization problem. First, the user can directly control the number of constraints in the resulting approximation while preserving the conservatism of the approximation with respect to the original chance constraint.



(a) True density (b) $m = 1$ (c) $m = 10$ (d) $m = 100$ (e) $m = 1000$

(f) True density (g) $m = 1$ (h) $m = 10$ (i) $m = 100$ (j) $m = 1000$

(k) True density (l) $m = 1$ (m) $m = 10$ (n) $m = 100$ (o) $m = 1000$

Fig. 1: The plots in the first column depict the probability density functions of three different $\mathbb{R}^2$-valued random vectors. The plots in columns two through five depict realizations of the uncertainty set $\mathcal{U}_\phi(F_n^{-1}(\alpha_n))$ for different values of m for each random vector.

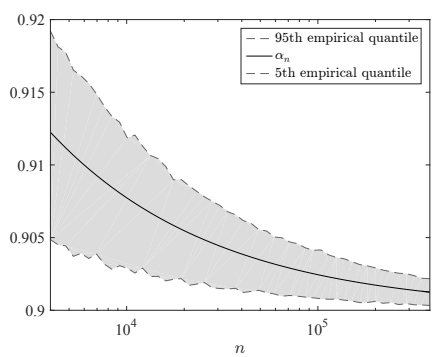


Fig. 2: The middle 90% empirical confidence interval associated with the coverage probability $\pi_\phi\left(F_n^{-1}(\alpha_n)\right)$ for different tolerance parameter values.

Second, the geometry of the uncertainty sets can be tailored to accommodate the structure of the given constraint function to ensure the computational tractability of the robust approximation. As a direction for future research, it would be interesting to refine our sample complexity results to explicitly reflect the role of the number of balls used in the approximation in determining the volume of the sets learned.

References

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

[2] S. Ahmed and A. Shapiro, "Solving chance-constrained stochastic programs via sampling and integer programming," in *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*. INFORMS, 2008, pp. 261–269.

[3] J. Luedtke, S. Ahmed, and G. L. Nemhauser, "An integer programming approach for linear programs with probabilistic constraints," *Mathematical Programming*, vol. 122, no. 2, pp. 247–272, 2010.

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

[5] A. Prekopa, "On probabilistic constrained programming," in *Proceedings of the Princeton Symposium on Mathematical Programming*, vol. 113. Princeton, NJ, 1970, p. 138.

[6] A. Prékopa, *Stochastic Programming*. Springer, 1995.

[7] C. M. Lagoa, "On the convexity of probabilistically constrained linear programs," in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 1. IEEE, 1999, pp. 516–521.

[8] C. M. Lagoa, X. Li, and M. Sznaier, "Probabilistically constrained linear programs and risk-adjusted controller design," *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 938–951, 2005.

[9] 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.

[10] W. Yang and H. Xu, "Distributionally robust chance constraints for non-linear uncertainties," *Mathematical Programming*, vol. 155, no. 1-2, pp. 231–265, 2016.

[11] 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.

[12] J. Pinter, "Deterministic approximations of probability inequalities," *Zeitschrift für Operations-Research*, vol. 33, no. 4, pp. 219–239, 1989.

[13] A. Ben-Tal and A. Nemirovski, "Robust solutions of linear programming problems contaminated with uncertain data," *Mathematical Programming*, vol. 88, no. 3, pp. 411–424, 2000.

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

[15] X. Chen, M. Sim, and P. Sun, "A robust optimization perspective on stochastic programming," *Operations Research*, vol. 55, no. 6, pp. 1058–1071, 2007.

[16] D. Bertsimas and D. B. Brown, "Constructing uncertainty sets for robust linear optimization," *Operations Research*, vol. 57, no. 6, pp. 1483–1495, 2009.

[17] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From cvar to uncertainty set: Implications in joint chance-constrained optimization," *Operations Research*, vol. 58, no. 2, pp. 470–485, 2010.

[18] 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.

[19] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen, "Robust solutions of optimization problems affected by uncertain probabilities," *Management Science*, vol. 59, no. 2, pp. 341–357, 2013.

[20] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, vol. 158, no. 1-2, pp. 291–327, 2016.

[21] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1-2, pp. 115–166, 2018.

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

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

[25] S.-H. Tseng, E. Bitar, and A. Tang, "Random convex approximations of ambiguous chance constrained programs," in *Proceedings of 55th IEEE Conference on Decision and Control*. IEEE, 2016, pp. 6210–6215.

[26] G. Schildbach, L. Fagiano, and M. Morari, "Randomized solutions to convex programs with multiple chance constraints," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2479–2501, 2013.

[27] X. Zhang, S. Grammatico, G. Schildbach, P. Goulart, and J. Lygeros, "On the sample size of random convex programs with structured dependence on the uncertainty," *Automatica*, vol. 60, pp. 182–188, 2015.

[28] K. Margellos, P. Goulart, and J. Lygeros, "On the road between robust optimization and the scenario approach for chance constrained optimization problems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2258–2263, 2014.

[29] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Mathematical Programming*, vol. 167, no. 2, pp. 235–292, 2018.

[30] T. Campbell and J. P. How, "Bayesian nonparametric set construction for robust optimization," in *Proceedings of the 2015 American Control Conference*. IEEE, 2015, pp. 4216–4221.

[31] B. Cadre, "Kernel estimation of density level sets," *Journal of Multivariate Analysis*, vol. 97, no. 4, pp. 999–1023, 2006.

[32] A. Cuevas, R. Fraiman *et al.*, "A plug-in approach to support estimation," *The Annals of Statistics*, vol. 25, no. 6, pp. 2300–2312, 1997.

[33] J. Lei, J. Robins, and L. Wasserman, "Distribution-free prediction sets," *Journal of the American Statistical Association*, vol. 108, no. 501, pp. 278–287, 2013.

[34] A. B. Tsybakov *et al.*, "On nonparametric estimation of density level sets," *The Annals of Statistics*, vol. 25, no. 3, pp. 948–969, 1997.

[35] L. Devroye and G. L. Wise, "Detection of abnormal behavior via nonparametric estimation of the support," *SIAM Journal on Applied Mathematics*, vol. 38, no. 3, pp. 480–488, 1980.

[36] A. Baíllo, A. Cuevas, and A. Justel, "Set estimation and nonparametric detection," *Canadian Journal of Statistics*, vol. 28, no. 4, pp. 765–782, 2000.

[37] W. Polonik, "Minimum volume sets and generalized quantile processes," *Stochastic Processes and their Applications*, vol. 69, no. 1, pp. 1–24, 1997.

[38] C. D. Scott and R. D. Nowak, "Learning minimum volume sets," *Journal of Machine Learning Research*, vol. 7, no. Apr, pp. 665–704, 2006.

[39] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[40] K. Muandet, K. Fukumizu, F. Dinuzzo, and B. Schölkopf, "Learning from distributions via support measure machines," in *Advances in Neural Information Processing Systems*, 2012, pp. 10–18.

[41] X. Huo and J.-C. Lu, "A network flow approach in finding maximum likelihood estimate of high concentration regions," *Computational Statistics & Data Analysis*, vol. 46, no. 1, pp. 33–56, 2004.

[42] D. Bertsimas and M. Sim, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, vol. 107, no. 1-2, pp. 5–36, 2006.

[43] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009, vol. 28.

[44] P. Khargonekar and A. Tikku, "Randomized algorithms for robust control analysis and synthesis have polynomial complexity," in *Proceedings of 35th IEEE Conference on Decision and Control*, vol. 3. IEEE, 1996, pp. 3470–3475.

[45] X. Chen and K. Zhou, "Order statistics and probabilistic robust control," *Systems & Control Letters*, vol. 35, no. 3, pp. 175–182, 1998.

[46] T. Alamo, R. Tempo, and A. Luque, "On the sample complexity of randomized approaches to the analysis and design under uncertainty," in *Proceedings of the 2010 American Control Conference*. IEEE, 2010, pp. 4671–4676.

[47] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Wiley, 2003.

[48] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge university press, 2005.