

# Improved Compression-Based PAC Bounds for General Scenario Decision Making

Guillaume O. Berger

**Abstract**—Scenario decision making is a distribution-free approach to decision making in the presence of stochastic constraints, which consists in drawing samples of the random constraint (called “scenarios”) and making a decision that satisfies all—or a predefined fraction—of the sampled constraints. Several bounds have been proposed in the literature on the number of samples that is sufficient to guarantee with high probability that the risk of the scenario decision, i.e., its probability of violating a random constraint, is low. We focus on sample-complexity bounds that depend on the compression size of the decision-making problem. We propose new compression-based sample-complexity bounds that improve upon the existing ones without requiring stronger assumptions on the problem. Specifically, our bounds do not rely on the so-called “nondegeneracy” assumption (which is widely used in the scenario decision-making literature, but can be limiting in practice); our bounds are better than the existing ones not relying on the nondegeneracy assumption, and is very close to the best known bounds relying on the nondegeneracy assumption.

**Index Terms**—Optimization, Randomized Algorithms, Statistical Learning.

## I. INTRODUCTION

**R**ISK-AWARE decision making is an important problem in many applications of engineering, such as control, energy planning, healthcare, etc. In this context, the risk of a decision is defined as the probability that it violates some random constraint, e.g., the probability that the planned trajectory of an autonomous vehicle collides with some stochastically moving obstacle (other cars, pedestrians, etc.). In many applications of risk-aware decision making, the probability distribution of the random constraint is unknown (e.g., in trajectory planning, or energy management). This makes the problem of making decisions with a low risk challenging in general. The approach of *scenario decision making* [1]–[10] provides an effective way to address this challenge by leveraging the principle of sample-based methods. The approach consists in drawing  $N$  samples (called *scenarios*) of the random constraint, and making a decision that satisfies all, or a predefined fraction, of the sampled constraints. The decision made from the scenarios is called the *scenario decision*. A key aspect of this approach is that under some assumptions on the problem, one can provide guarantees on the risk of the scenario decision. In this work, we focus on PAC (*Probably Approximately Correct*) guarantees [11], [12]: namely, upper bounds on the probability (with respect to the sampling of the  $N$  scenarios) that the

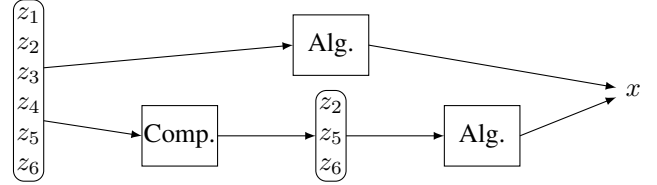


Fig. 1. Scenario decision algorithm (Alg.) that has compression size 3: any input set of scenarios (e.g.,  $z_1, \dots, z_6$ ) can be reduced or compressed (Comp.) to a subset of size 3 (e.g.,  $z_2, z_5, z_6$ ) and give the same decision ( $x$ ) when given as input to Alg.

risk of the scenario decision exceeds some tolerance  $\epsilon$ . These bounds, referred to as *confidence bounds*, depend on  $N$ ,  $\epsilon$ , and other problem-dependent quantities generally known as “problem complexity”. Common problem complexity measures include the VC dimension, the Rademacher complexity, and the compression size; see, e.g., [10] for a survey.

In this paper, we focus on the compression size, introduced in 1986 by Littlestone and Warmuth [13]. The compression size of a scenario decision-making algorithm is the smallest number of scenarios to which any finite set of scenarios can be reduced while providing the same decision; see Figure 1 for an illustration. In [13], a confidence bound is provided for scenario decision-making algorithms with bounded compression size. A better confidence bound is provided in [1] under additional assumptions on the problem (among others, each constraint must define a convex set of admissible decisions, and the set of admissible decisions with respect to all constraints must have nonempty interior). This bound is shown to be tight for certain instances of such problems (called *fully-supported*) [1]. A similar confidence bound is proposed in [14] when the algorithm and the probability distribution of the random constraint satisfy the so-called *nondegeneracy* assumption (essentially saying that the minimal-size compression set is unique). This assumption has been widely used in the scenario decision making literature [3], [6], [8], [15], [16]; however, it can be restrictive in practice (e.g., it excludes discrete probability distributions) and is by nature impossible to check when the probability distribution is unknown.

In the aforementioned works [1], [13], the scenario decision satisfies all the sampled constraints. Therefore, we call these algorithms *consistent*. By contrast, in *sample-and-discard* algorithms, introduced in [3], the scenario decision is allowed to violate some of the sampled constraints (which are “discarded”). A confidence bound for these algorithms under the nondegeneracy assumption and the so-called *conformity* assumption (essentially saying that the discarded constraints

This project has received funding from the Belgian National Fund for Scientific Research (F.R.S.–FNRS). GB is with ICTEAM institute, UCLouvain, Louvain-la-Neuve, Belgium. He is a FNRS Postdoctoral Researcher. (e-mails: guillaume.berger@uclouvain.be).

are violated by the scenario decision) is proposed in [3]. A better bound (reminiscent of the one in [1]) is proposed in [15] (see also [16]) when additionally the solution of the sample-based problem is computed using so-called *cascade optimization* (removing violated constraints sequentially, i.e., in “cascade”). This bound is shown to be tight for some instances (called *fully-supported*) [16]. However, the use of cascade optimization can be restrictive, sub-optimal, and computationally expensive. Furthermore, the analysis in [3], [15], [16] still requires the nondegeneracy assumption, which can be restrictive. Finally, in [17], a confidence bound that does not require the nondegeneracy assumption is proposed. However, this bound is looser than the ones in [3], [15], [16], and still assumes the conformity assumption.

The contribution of this paper is to provide new compression-based confidence bounds for consistent and sample-and-discard scenario decision-making algorithms, that do not require nondegeneracy, conformity, or similar assumptions on the problem and the algorithm. These bounds are to the best of our knowledge the best available bounds in the literature without those assumptions.

#### Related work

The initial derivation of the main result (Theorem 2) was strongly inspired by the recent work [18]. This work provides bounds on the risk of scenario decisions using the *wait-and-judge* approach (initially introduced in [6]). In the wait-and-judge approach, the compression size needs not be bounded for all sets of scenarios. However, given a set of  $N$  scenarios, one can compute its compression size (which can be any integer between 0 and  $N$ ), and from it, derive a probabilistic upper bound on the risk. Notably, the wait-and-judge bound in [18] does not require additional assumptions on the problem or the algorithm. When the compression size of every set of  $N$  scenarios is upper bounded, then the wait-and-judge bound still applies, and can be turned into a compression-based confidence bound (Theorem 1, item 3). Interestingly, the resulting confidence bound already improves upon the confidence bound proposed in [13]. By looking carefully at the derivation of the wait-and-judge bound in [18], one can observe that some parameters of the proof can be optimized to provide a better confidence bound in the case the compression size of every set of  $N$  scenarios is upper bounded. The best set of parameters that we were able to find for this purpose lead to the confidence bound presented in our main result (Theorem 2). The proof in [18] relies on advanced technical results related to duality in semi-infinite optimization. By contrast, the proof that we propose in this paper for our main result relies only on elementary probability theory.

*Notation.* We let  $\mathbb{N}$  be the set of nonnegative integers. For  $n \in \mathbb{N}$ , we let  $[n] = \{1, \dots, n\}$ . For a set  $A$ , we let  $A^* = \bigcup_{m=0}^{\infty} A^m$ . For  $m \in \mathbb{N}$  and a finite set  $A$ , the set of all subsets of  $A$  with  $m$  elements is denoted by  $\binom{A}{m}$ .

## II. PROBLEM STATEMENT AND BACKGROUND

We introduce the problem of risk-aware scenario decision making.<sup>1</sup> We assume given a set  $X$  of *decisions* and a set  $Z \subseteq 2^X$  of *constraints* on  $X$ . Given a decision  $x \in X$  and a constraint  $z \in Z$ , we say that  $x$  *satisfies*  $z$  if  $x \in z$ ; otherwise, we say that  $x$  *violates*  $z$ .

**Example 1.** In the following optimization problem:

$$\min_{x \in \mathbb{R}^2} \|x\| \quad \text{s.t.} \quad a^\top(x - c) \leq 1 \quad \forall a \in \mathbb{R}^2, \|a\| \leq 1,$$

where  $c \in \mathbb{R}^2$  is fixed, the decision space is  $X = \mathbb{R}^2$ , and the constraint space is  $Z = \{X_a \subseteq X : a \in \mathbb{R}^2, \|a\| \leq 1\}$ , where  $X_a = \{x \in \mathbb{R}^2 : a^\top(x - c) \leq 1\}$ .

Finding a decision that satisfies all constraints  $z$  in  $Z$  is often intractable if  $Z$  is large or unknown. An approach to circumvent this—in the case where  $Z$  can be sampled—is to sample  $N$  constraints  $z_1, \dots, z_N$  from  $Z$ , and find a decision  $x$  based on the sampled constraints. This process is called a *scenario decision algorithm*, which is thus a function from tuples of constraints to decisions:

**Definition 1.** A scenario decision algorithm is a function that given a tuple of constraints  $(z_1, \dots, z_N) \in Z^*$  returns a decision  $x \in X$ . Hence, it is a function  $\mathcal{A} : Z^* \rightarrow X$ .

**Remark 1.** We consider tuples of constraints, instead of sets of constraints, because in Section IV the multiplicity of the constraints will play a role. Furthermore, we need to describe the probability measure of the sets of i.i.d. sampled constraints, which is much easier to define when working with tuples of constraints (it is simply the product measure).

**Example 2.** Continuing Example 1, a scenario decision algorithm  $\mathcal{A}$  can be defined as follows: given  $z_i = X_{a_i}$ , define  $\mathcal{A}(z_1, \dots, z_N)$  as the optimal solution of

$$\min_{x \in \mathbb{R}^2} \|x\| \quad \text{s.t.} \quad a_i^\top(x - c) \leq 1 \quad \forall i \in [N].$$

This is an instance of scenario optimization [1].

The output of a scenario decision algorithm is sometimes called the *scenario decision*. Since the scenario decision is obtained from a subset of the constraints in  $Z$ , one can generally not hope that it satisfies *all* the constraints in  $Z$ . However, one can hope that it satisfies all constraints in  $Z$  except possibly those in a subset of small measure. This property is formalized by the notion of risk:

**Definition 2.** Given a probability measure  $P$  on  $Z$ , and a decision  $x \in X$ , the risk (also called violation probability) of  $x$  w.r.t.  $P$ , denoted by  $V_P(x)$ , is the probability that  $x$  violates a random constraint  $z \in Z$ , i.e.,  $V_P(x) \triangleq P[\{z \in Z : x \notin z\}]$ .

The risk of the scenario decision depends on the scenarios, which are sampled i.i.d. at random according to the distribution of the random constraint. Given a risk tolerance  $\epsilon$  and a number of scenarios  $N$ , we aim to upper bound the probability that the risk of the scenario decision is below  $\epsilon$ :

<sup>1</sup>The definitions and concepts presented in this section are classical, and can be found, e.g., in [2], [8], [9], [17].

**Definition 3.** Consider a scenario decision algorithm  $\mathcal{A}$ . A confidence bound for  $\mathcal{A}$  is a function  $q : (0, 1] \times \mathbb{N} \rightarrow [0, 1]$  such that for any tolerance  $\epsilon \in (0, 1)$ , sample size  $N \in \mathbb{N}$  and probability distribution  $P$  on  $Z$ , it holds with probability  $1 - q(\epsilon, N)$  that if one samples  $N$  scenarios  $(z_1, \dots, z_N) \in Z^N$  i.i.d. according to  $P$ , then the scenario decision returned by  $\mathcal{A}$  has risk below  $\epsilon$ , i.e.,

$$P^N[\{z \in Z^N : V_P(\mathcal{A}(z)) > \epsilon\}] \leq q(\epsilon, N),$$

where  $z$  is a shorthand notation for  $(z_1, \dots, z_N)$ .

**Remark 2.** If for all  $\epsilon \in (0, 1]$ ,  $\lim_{N \rightarrow \infty} q(\epsilon, N) = 0$ , then the scenario decision algorithm is said to be PAC (Probably Approximately Correct) [11], [12], [19].

Our goal in this paper is to provide confidence bounds for scenario decision algorithms whose compression size (see Section II-A below) is bounded. We make two assumptions on the scenario decision algorithms that we consider in this work: the algorithm is permutation invariant and stable:

**Assumption 1.** The scenario decision algorithm  $\mathcal{A}$  is

- permutation invariant: for all  $(z_1, \dots, z_N) \in Z^*$  and all permutations  $(i_1, \dots, i_N)$  of  $[N]$ ,  $\mathcal{A}(z_1, \dots, z_N) = \mathcal{A}(z_{i_1}, \dots, z_{i_N})$ ;
- stable: for all  $(z_1, \dots, z_{N+1}) \in Z^{\geq 1}$ ,  $\mathcal{A}(z_1, \dots, z_N) \in z_{N+1}$  implies that  $\mathcal{A}(z_1, \dots, z_{N+1}) = \mathcal{A}(z_1, \dots, z_N)$ .

Those assumption are very mild and standard in the literature [1]–[3], [9], [10], [14]–[17]. In Sections III and IV, we will consider separately two additional assumptions on the algorithm: (i) that the scenario decision satisfies all the sampled constraints (called *consistent*), and (ii) that the scenario decision satisfies all but a fixed number  $r$  of the sampled constraints (called *sample-and-discard*).

#### A. Compression Sets

Before ending this section, let us recall the notion of *compression*, which is instrumental in this work since we study confidence bounds based on the compression size. This notion was introduced in 1986 by Littlestone and Warmuth [13] in an unpublished manuscript cited in [20].

The notion of compression describes the property that the input tuple of scenarios  $z \in Z^*$  is compressible in the sense that a fixed-length subtuple  $z'$  of  $z$  leads to the same output:

**Definition 4** (Compression). Given a scenario decision algorithm  $\mathcal{A}$  and  $z := (z_1, \dots, z_N) \in Z^*$ , a subset  $\{i_1, \dots, i_M\} \subseteq [N]$ , is a compression set for  $z$  and  $\mathcal{A}$  if  $\mathcal{A}(z_{i_1}, \dots, z_{i_M}) = \mathcal{A}(z)$ . The set of all compression sets of  $z$  and  $\mathcal{A}$  is denoted by  $\kappa(z; \mathcal{A})$ .<sup>2</sup> The compression size of  $\mathcal{A}$ , denoted by  $\rho(\mathcal{A})$ , is the infimum of all  $d \in \mathbb{N}$  such that for every  $z \in Z^*$ , there is a compression set  $I \in \kappa(z; \mathcal{A})$  with  $|I| \leq d$ .

We review confidence bounds available in the literature for scenario decision algorithms whose compression size is bounded. We then propose new confidence bounds under Assumption 1. We do this for consistent algorithms (Section III) and sample-and-discard algorithms (Section IV).

<sup>2</sup>When clear from the context, we drop the dependence on  $\mathcal{A}$  and simply write  $\kappa(z)$ .

### III. CONSISTENT ALGORITHMS

In this section, we focus on consistent scenario decision algorithms, i.e., algorithms for which the scenario decision satisfies all the sampled constraints:

**Definition 5.** A scenario decision algorithm  $\mathcal{A}$  is consistent if for all  $z := (z_1, \dots, z_N) \in Z^*$ ,  $\mathcal{A}(z) \in \bigcap_{i=1}^N z_i$ .

The algorithm in Example 2 is consistent.

#### A. Previous Bounds in the Literature

We review confidence bounds available in the literature for consistent algorithms:

**Theorem 1.** Consider a consistent scenario decision algorithm  $\mathcal{A}$  satisfying Assumption 1. Let  $d \in \mathbb{N}$ , and assume that  $\rho(\mathcal{A}) \leq d$ . Then, confidence bounds for  $\mathcal{A}$  are given by

- 1)  $q(N, \epsilon) = \binom{N}{d}(1 - \epsilon)^{N-d}$ ,  $N \in \mathbb{N}_{\geq d}$  and  $\epsilon \in (0, 1]$ ; see [20, Theorem 6], [17, Theorem 2];
- 2)  $q(N, \epsilon) = \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i}$ ,  $N \in \mathbb{N}_{\geq d}$  and  $\epsilon \in (0, 1]$ , under the “nondegeneracy” assumption [14, Definition 2.7] on  $\mathcal{A}$  and  $P$ , or the “convex feasible set with nonempty interior” assumption [1, Assumption 1] on  $\mathcal{A}$ ; see [14, Theorem 3.3], [1, Theorem 1];
- 3)  $q(N, \epsilon) = \frac{N \binom{N}{d} (1 - \epsilon)^{N-d}}{\sum_{m=d}^{N-1} \binom{m}{d} (1 - \epsilon)^{m-d}}$ ,  $N \in \mathbb{N}_{\geq d}$  and  $\epsilon \in (0, 1]$ ; derived from [18, Theorem 1].

**Remark 3.** Note that  $q(N, \epsilon)$  in item 2) is not a confidence bound *stricto sensu* because it is not satisfied for every  $P$  (indeed,  $P$  needs to satisfy the nondegeneracy assumption); hence, it is not “distribution-free”. The nondegeneracy assumption is restrictive: it excludes distributions that have “concentrated masses” (aka. “atoms”).

**Remark 4.** The bound in item 3) is derived from the “wait-and-judge” bound in [18].<sup>3</sup> This provides an upper bound on the risk of the scenario decision, with predefined confidence, as a function of the confidence level, the number of samples, and the compression size of the tuple of sampled constraints. When the compression size of any tuple of sampled constraints is bounded by  $d$ , then a confidence bound in the sense of Definition 3 can be retrieved, which is presented in item 3).

As mentioned in the “Related work” section, our main result (Theorem 2 below), which improves upon the bounds in items 1) and 3), was initially obtained by optimizing parameters in the proof of [18, Theorem 1], to obtain better confidence bounds when the compression size of the algorithm is bounded. Nevertheless, for the best set of parameters that we found, a simpler, more intuitive proof was possible, which is presented in the next subsection.

Figure 2 compares the different bounds from Theorem 1. As we can see, the bound in item 2) is significantly better than the one in item 1); in fact, the former is tight for so-called “fully-supported” problems [14, Definition 2.5]. The bound in item

<sup>3</sup>To be complete, let us mention that wait-and-judge bounds were proposed, e.g., in [6], [8]. These bounds are better than the one in [18]; however, they require the nondegeneracy assumption. A key achievement of [18] is to remove this assumption and obtain a bound that is almost as good.

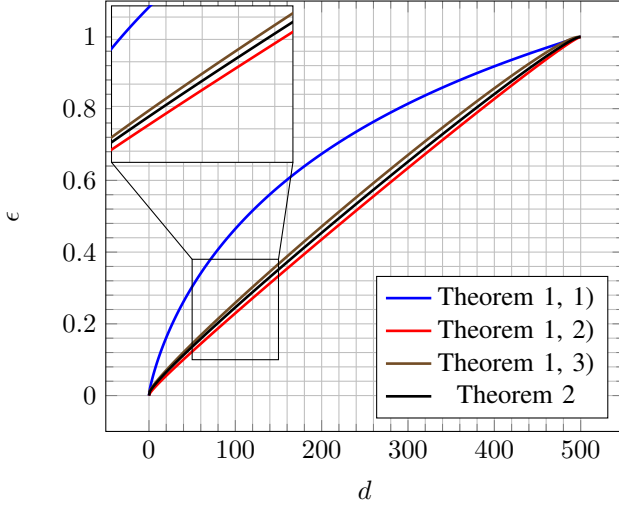


Fig. 2. The curves show the value of  $\epsilon$  so that  $q(N, \epsilon) = 0.05$  for  $N = 500$  for different values of  $d$ . Lower values of  $\epsilon$  indicate better performance. We see that our bound (Theorem 2) is very close to the bound in item 2) without requiring the nondegeneracy assumption. It is also slightly better than the one in item 3) and significantly better than the one in item 1), which all have the same assumptions.

3) is much closer to the one in item 2) and does not require the nondegeneracy assumption. In the next subsection, we provide a new compression-based confidence bound that is better than the one in item 3) and does not require the nondegeneracy assumption. The proof is also elementary.<sup>4</sup>

### B. Our Bound

We will prove the following result, which is the first main contribution of this paper:

**Theorem 2.** Consider a consistent scenario decision algorithm  $\mathcal{A}$  satisfying Assumption 1. Let  $d \in \mathbb{N}$ , and assume that  $\rho(\mathcal{A}) \leq d$ . A confidence bound for  $\mathcal{A}$  is given by

$$q(N, \epsilon) = \binom{N}{d} \min_{m=d, \dots, N} \binom{m}{d}^{-1} (1 - \epsilon)^{N-m}$$

for all  $N \in \mathbb{N}_{\geq d}$  and  $\epsilon \in (0, 1]$ .

Figure 2 shows the bound in Theorem 2. As we can see, this bound is better than the one in item 3) and requires the same assumptions. It is also very close to the one in item 2).

### C. Proof of Theorem 2

In the proof, we make the assumption that for each  $\mathbf{z} \in Z^{\geq d}$ , there is a *single* compression set of size  $d$  for  $\mathbf{z}$ . This assumption is made *without loss of generality* because one can always define an augmented decision set, augmented constraint set, and augmented scenario decision algorithm, for which the assumption holds and such that the augmented scenario decision matches with the scenario decision after projecting

<sup>4</sup>It does not rely on advanced mathematical tools such as Hausdorff moment problems as in [1, Theorem 1], or duality in semi-infinite optimization as in [18, Theorem 1].

on the original decision set. For such a construction, we refer the reader to [18, first paragraph of §2].<sup>5</sup>

**Assumption 2.** The scenario decision algorithm  $\mathcal{A}$  satisfies that for all  $\mathbf{z} \in Z^{\geq d}$ ,  $|\kappa(\mathbf{z})| = 1$ .

A consequence of Assumption 2 is the following:

**Lemma 1.** Let  $\mathcal{A}$  be as in Theorem 2 and satisfy Assumption 2. Let  $\mathbf{P}$  be a probability measure on  $Z$ . For any  $m \in \mathbb{N}_{\geq d}$  and  $I \in \binom{[m]}{d}$ , it holds that  $\mathbf{P}^m(\{\mathbf{z} \in Z^m : I \in \kappa(\mathbf{z})\}) = \binom{m}{d}^{-1}$ .

*Proof.* Fix  $m \in \mathbb{N}_{\geq d}$ . For each  $I \in \binom{[m]}{d}$ , define  $S_I = \{\mathbf{z} \in Z^m : I \in \kappa(\mathbf{z})\}$ . The permutation invariance of  $\mathcal{A}$  implies that  $\mathbf{P}^m(S_I)$  is independent of  $I$ . Furthermore, by Assumption 2, for any  $I_1 \neq I_2$ ,  $S_{I_1} \cap S_{I_2} = \emptyset$ . Hence,  $\mathbf{P}^m(S_I) = \binom{m}{d}^{-1}$ .  $\square$

Next, we leverage the bound in Theorem 1, item 1), and extend it by incorporating the probability that a given set of  $d$  elements is extracted from a set of  $m$  samples (Lemma 3):

**Lemma 2.** Let  $\mathcal{A}$  be as in Theorem 2 and satisfy Assumption 2. Let  $N \in \mathbb{N}_{\geq d}$ ,  $I \in \binom{[N]}{d}$  and  $m \in \mathbb{N}_{\geq d}$ . Let  $(j_1, \dots, j_m) \in [N]^m$  be such that there is  $I' := \{i_1, \dots, i_d\} \subseteq [m]$  satisfying  $I = \{j_{i_1}, \dots, j_{i_d}\}$ . It holds that for all  $\mathbf{z} := (z_1, \dots, z_N) \in Z^N$ , if  $I \in \kappa(\mathbf{z})$ , then  $I' \in \kappa(\mathbf{z}')$ , where  $\mathbf{z}' = (z_{j_1}, \dots, z_{j_m})$ .

*Proof.* Let  $\mathbf{z} := (z_1, \dots, z_N) \in Z^N$  be such that  $I \in \kappa(\mathbf{z})$ . Let  $I' := \{i_1, \dots, i_d\} \subseteq [m]$  be such that  $I = \{j_{i_1}, \dots, j_{i_d}\}$ . Let  $\mathbf{z}'' = (z_{j_{i_1}}, \dots, z_{j_{i_d}})$ . By definition of  $\kappa$ , we just need to show that  $\mathcal{A}(\mathbf{z}'') = \mathcal{A}(\mathbf{z}')$ . Therefore, first observe that  $\mathcal{A}(\mathbf{z}'') = \mathcal{A}(\mathbf{z})$  since  $I \in \kappa(\mathbf{z})$ . Then, since  $\mathcal{A}$  is consistent, it follows that  $\mathcal{A}(\mathbf{z}'') \in \bigcap_{k=1}^N z_k$ . Hence, since  $\mathcal{A}$  is stable and permutation invariant, we deduce that  $\mathcal{A}(\mathbf{z}') = \mathcal{A}(\mathbf{z}'')$ .  $\square$

**Lemma 3.** Let  $\mathcal{A}$  be as in Theorem 2 and satisfy Assumption 2. Let  $N \in \mathbb{N}_{\geq d}$ ,  $I \in \binom{[N]}{d}$ , and  $\epsilon \in (0, 1]$ . For all  $m \in \mathbb{N} \cap [d, N]$ , it holds that  $\mathbf{P}^N(\{\mathbf{z} \in Z^N : I \in \kappa(\mathbf{z}), V_{\mathbf{P}}(\mathcal{A}(\mathbf{z})) > \epsilon\}) \leq \binom{m}{d}^{-1} (1 - \epsilon)^{N-m}$ .

*Proof.* Let  $m \in \mathbb{N} \cap [d, N]$ ,  $(j_1, \dots, j_N)$  be a permutation of  $[N]$ , and  $I' := \{i_1, \dots, i_d\} \subseteq [m]$  be such that  $I = \{j_{i_1}, \dots, j_{i_d}\}$ . Let  $T' = \{\mathbf{z} \in Z^m : I' \in \kappa(\mathbf{z}), V_{\mathbf{P}}(\mathcal{A}(\mathbf{z})) > \epsilon\}$ . By Lemma 2 and the consistency of  $\mathcal{A}$ , it holds that for all  $\mathbf{z} \in Z^m$ , if  $I \in \kappa(\mathbf{z})$ , then  $\mathbf{z}' \in T'$  and  $\mathcal{A}(\mathbf{z}') \in \bigcap_{k=m+1}^N z_k$ , where  $\mathbf{z}$  is a shorthand notation for  $(z_1, \dots, z_N)$ , and  $\mathbf{z}'$  for  $(z_{j_1}, \dots, z_{j_m})$ .

Given  $x \in X$ , let  $R(x) = \{(z_1, \dots, z_{N-m}) \in Z^{N-m} : x \in \bigcap_{k=1}^{N-m} z_k\}$ . Note that  $\mathbf{P}^{N-m}(R(x)) = (1 - V_{\mathbf{P}}(x))^{N-m}$ . Let  $T = \{\mathbf{z} \in Z^N : I \in \kappa(\mathbf{z}), V_{\mathbf{P}}(\mathcal{A}(\mathbf{z})) > \epsilon\}$ . It holds that

$$\mathbf{P}^N(T) = \mathbf{E}_{\mathbf{z} \sim \mathbf{P}^N}[\mathbf{1}_{R(\mathcal{A}(\mathbf{z}'))}(\mathbf{z}'') \mathbf{1}_{T'}(\mathbf{z}')],$$

where  $\mathbf{z}''$  is a shorthand for  $(z_{j_{m+1}}, \dots, z_{j_N})$ . It follows that

$$\mathbf{P}^N(T) \leq (1 - \epsilon)^{N-m} \mathbf{P}^m(T') \leq (1 - \epsilon)^{N-m} \binom{m}{d}^{-1},$$

where we used Lemma 1 to obtain the second inequality.  $\square$

We are now able to conclude the proof of Theorem 2:

<sup>5</sup>This construction requires only the stability property in Assumption 1.

*Proof of Theorem 2.* Fix  $N \in \mathbb{N}_{\geq d}$ ,  $P$  a probability distribution on  $Z$ , and  $\epsilon \in (0, 1]$ . Let  $T = \{z \in Z^N : V_P(\mathcal{A}(z)) > \epsilon\}$ . For each  $I \in \binom{[N]}{d}$ , let  $T_I = \{z \in T : I \in \kappa(z)\}$ . It holds that  $T = \bigcup_{I \in \binom{[N]}{d}} T_I$ . Hence, by Lemma 3 and the union bound, we obtain that for every  $m \in \mathbb{N} \cap [d, N]$ ,

$$P^N(T) \leq \sum_{I \in \binom{[N]}{d}} P^N(T_I) \leq \binom{N}{d} \binom{m}{d}^{-1} (1 - \epsilon)^{N-m}.$$

This concludes the proof of the theorem.  $\square$

**Remark 5.** It is interesting to note that the proof of Lemma 3 uses the same idea as the proof of [20, Theorem 6], or [17, Theorem 2]. It is Assumption 2, which can be made without loss of generality for permutation invariant algorithms, and the stability property that allow (via Lemmas 1 and 2) to obtain the bound in Theorem 2, which improves upon them.

#### IV. RELAXING THE CONSISTENCY ASSUMPTION

In some applications, consistency can be too restrictive: it is sometimes preferable to discard a few sampled constraints before passing them to a consistent scenario decision algorithm if this can lead to scenario decisions with better performances (e.g., lower cost), especially when a lot of constraints are sampled. The algorithm that discards a fixed number of constraints and keeps the other ones is called a selection algorithm:

**Definition 6.** A function  $\sigma : Z^* \rightarrow Z^*$  is a selection algorithm with discarding size  $r \in \mathbb{N}$  if for every  $z := (z_1, \dots, z_N) \in Z^*$ , there are integers  $1 \leq i_1 < \dots < i_M \leq N$ , with  $M \geq N - r$ , such that  $\sigma(z) = (z_{i_1}, \dots, z_{i_M})$ .

The resulting scenario decision algorithm is called a sample-and-discard decision algorithm:

**Definition 7.** A scenario decision algorithm  $\mathcal{B}$  is a sample-and-discard decision algorithm with discarding size  $r \in \mathbb{N}$  if there is (i) a consistent scenario decision algorithm  $\mathcal{A}$  and (ii) a selection algorithm  $\sigma$  with discarding size  $r$ , such that  $\mathcal{B} = \mathcal{A} \circ \sigma$ . The pair  $(\mathcal{A}, \sigma)$  is called a decomposition of  $\mathcal{B}$ .

**Example 3.** Continuing Examples 1–2, a sample-and-discard decision algorithm  $\mathcal{B}$  can be defined as follows: (i)  $\mathcal{A}$  is as in Example 2, and (ii)  $\sigma$  selects among  $(z_1, \dots, z_N) \in Z^*$ , with  $z_i = X_{a_i}$ , the  $N - r$  constraints for which  $\|a_i\|$  are the lowest (using a tie-breaking rule in case of ties).

##### A. Previous Bounds in the Literature

We review confidence bounds available in the literature for sample-and-discard decision algorithms:

**Theorem 3.** Consider a sample-and-discard decision algorithm  $\mathcal{B}$  with discarding size  $r \in \mathbb{N}$  and decomposition  $(\mathcal{A}, \sigma)$ , where  $\mathcal{A}$  satisfies Assumption 1. Let  $d \in \mathbb{N}$ , and assume that  $\rho(\mathcal{A}) \leq d$ . Then, confidence bounds for  $\mathcal{B}$  are given by

- 1)  $q(N, \epsilon) = \binom{N}{d} \sum_{i=0}^r \binom{N-d}{i} \epsilon^i (1 - \epsilon)^{N-d-i}$ ,  $N \in \mathbb{N}_{\geq r+d}$  and  $\epsilon \in (0, 1]$ ; see [17, Theorem 4];
- 2)  $q(N, \epsilon) = \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i}$ ,  $N \in \mathbb{N}_{\geq r+d-1}$  and  $\epsilon \in (0, 1]$ , under the “nondegeneracy”

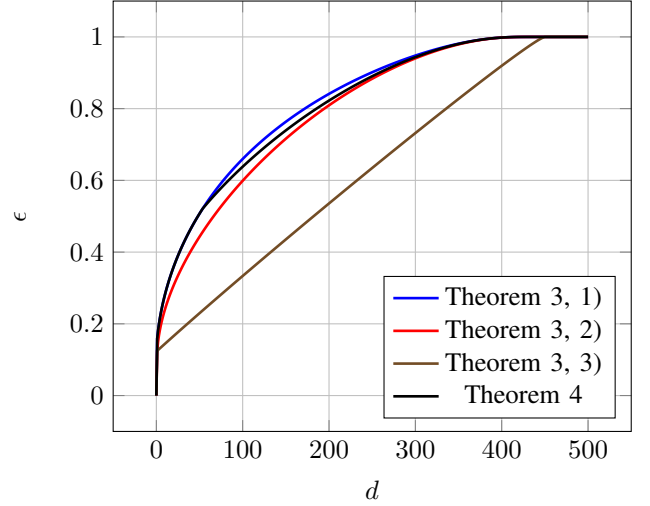


Fig. 3. The curves show the value of  $\epsilon$  so that  $q(N, \epsilon) = 0.05$  for  $N = 500$  and  $r = 50$  for different values of  $d$ . Lower values of  $\epsilon$  indicate better performance. We see that our bound (Theorem 4) is between the bounds in items 1) and 2) without requiring the nondegeneracy or conformity assumption. The bound in item 3) is significantly better but requires stronger assumptions.

- assumption [3, Assumption 2.1] and the “conformity” assumption [3, Assumption 2.2]; see [3, Theorem 2.1];
- 3)  $q(N, \epsilon) = \sum_{i=0}^{r+d-1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i}$ ,  $N \in \mathbb{N}_{\geq r+d-1}$  and  $\epsilon \in (0, 1]$ , under the “sequential nondegeneracy” assumption [16, Assumption 3] and the form [16, (21)] for  $\mathcal{A}$ ; see [16, Theorem 4].

**Remark 6.** Interestingly, [17, Theorem 4] assumes a “conformity” assumption [17, Assumption 4, item 3)], although it is not used in the proof. Therefore, we omit it in item 1).

**Remark 7.** The same comment as in Remark 3 applies for items 2) and 3).

Figure 3 compares the different bounds from Theorem 3. As we can see, the bound in item 3) is significantly better than those in items 1) and 2). However, it requires stronger assumptions. On the other hand, the bounds in items 1) and 2) are more or less identical, despite the fact that item 1) does not require the nondegeneracy assumption. In the next subsection, we provide a new compression-based confidence bound that is better than the one in items 1) in some regimes and does not require the nondegeneracy or conformity assumption. The proof is also elementary.

##### B. Our Bound

We will prove the following result, which is the second main contribution of this paper:

**Theorem 4.** Consider a sample-and-discard decision algorithm  $\mathcal{B}$  with discarding size  $r \in \mathbb{N}$  and decomposition  $(\mathcal{A}, \sigma)$ , where  $\mathcal{A}$  satisfies Assumption 1. Let  $d \in \mathbb{N}$ , and assume that  $\rho(\mathcal{A}) \leq d$ . A confidence bound for  $\mathcal{B}$  is given by

$$q(N, \epsilon) = \binom{N}{r} \binom{N-r}{d} \min_{m=d, \dots, N-r} \binom{m}{d}^{-1} (1 - \epsilon)^{N-r-m},$$

for all  $N \in \mathbb{N}_{\geq r+d}$  and  $\epsilon \in (0, 1]$ .

Figure 3 shows the bound provided in Theorem 4. As we can see, in some regimes (roughly,  $d > r$ ), the bound is better than the one in item 1) and requires the same assumptions.

### C. Proof of Theorem 4

The proof builds upon Theorem 2 and elementary combinatorics:

*Proof of Theorem 4.* Fix  $N \in \mathbb{N}_{\geq d+r}$ ,  $P$  a probability distribution on  $Z$  and  $\epsilon \in (0, 1]$ . Let  $T = \{z \in Z^N : V_P(\mathcal{A}(z)) > \epsilon\}$ . Define  $\Delta = \{(i_1, \dots, i_{N-r}) \in \mathbb{N}^{N-r} : 1 \leq i_1 < \dots < i_{N-r} \leq N\}$ , and for each  $I := (i_1, \dots, i_{N-r}) \in \Delta$ , define  $T_I = \{z \in T : \sigma(z) = z'\}$ , where  $z$  is a shorthand notation for  $(z_1, \dots, z_N)$ , and  $z'$  for  $(z_{i_1}, \dots, z_{i_{N-r}})$ . We will show that for all  $I \in \Delta$ ,

$$P^N(T_I) \leq \binom{N-r}{d} \min_{m=d, \dots, N-r} \binom{m}{d}^{-1} (1-\epsilon)^{N-r-m}. \quad (1)$$

Therefore, fix  $I := (i_1, \dots, i_{N-r}) \in \Delta$ . Note that  $T_I \subseteq \{z \in Z^N : V_P(\mathcal{A}(z')) > \epsilon\}$ , where  $z$  is a shorthand notation for  $(z_1, \dots, z_N)$ , and  $z'$  for  $(z_{i_1}, \dots, z_{i_{N-r}})$ . We deduce (1) from Theorem 2 and the fact that  $P^N(\{z \in Z^N : V_P(\mathcal{A}(z')) > \epsilon\}) = P^{N-r}(\{z' \in Z^{N-r} : V_P(\mathcal{A}(z')) > \epsilon\})$ . We obtain the final result by using the union bound:  $P^N(T) \leq \sum_{I \in \Delta} P^N(T_I)$ , and the observation that  $|\Delta| = \binom{N}{r}$ .  $\square$

## V. CONCLUSION

We reviewed and compared different confidence bounds available in the literature for scenario decision making based on the compression size of the problem. Some of these bounds are known to be optimal but require strong assumptions while others require weaker assumptions but are less good and are not known to be optimal. We improved the previous results by providing new bounds that do not require the strongest assumptions and are better than the available bounds with the same assumptions. For future work, we plan to improve these bounds even further or showing their optimality, we also plan to extend them to other approaches such as the “wait-and-judge” approach.

## REFERENCES

- [1] 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.
- [2] T. Alamo, R. Tempo, and E. F. Camacho, “Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems,” *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2545–2559, 2009.
- [3] M. C. Campi and S. Garatti, “A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality,” *Journal of Optimization Theory and Applications*, vol. 148, no. 2, pp. 257–280, 2011.
- [4] 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.
- [5] P. M. Esfahani, T. Sutter, and J. Lygeros, “Performance bounds for the scenario approach and an extension to a class of non-convex programs,” *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 46–58, 2015.

- [6] M. C. Campi and S. Garatti, “Wait-and-judge scenario optimization,” *Mathematical Programming*, vol. 67, no. 1, pp. 155–189, 2018.
- [7] M. C. Campi, S. Garatti, and F. A. Ramponi, “A general scenario theory for nonconvex optimization and decision making,” *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4067–4078, 2018.
- [8] S. Garatti and M. C. Campi, “Risk and complexity in scenario optimization,” *Mathematical Programming*, vol. 191, pp. 243–279, 2022.
- [9] M. C. Campi and S. Garatti, “Compression, generalization and learning,” *Journal of Machine Learning Research*, vol. 24, no. 339, pp. 1–74, 2023.
- [10] R. Rocchetto, A. Mey, and F. A. Oliehoek, “A survey on scenario theory, complexity, and compression-based learning and generalization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 16 985–16 999, 2024.
- [11] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: from theory to algorithms*. Cambridge, UK: Cambridge University Press, 2014.
- [12] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*, 2nd ed. Cambridge, MA: MIT, 2018.
- [13] N. Littlestone and M. Warmuth, “Relating data compression and learnability,” 1986, unpublished manuscript.
- [14] G. C. Calafiore, “Random convex programs,” *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3427–3464, 2010.
- [15] L. Romao, K. Margellos, and A. Papachristodoulou, “Tight sampling and discarding bounds for scenario programs with an arbitrary number of removed samples,” in *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, ser. Proceedings of Machine Learning Research, A. Jadbabaie, J. Lygeros, G. J. Pappas, P. Parrilo, B. Recht, C. J. Tomlin, and M. N. Zeilinger, Eds., vol. 144, 2021, pp. 312–323.
- [16] L. Romao, A. Papachristodoulou, and K. Margellos, “On the exact feasibility of convex scenario programs with discarded constraints,” *IEEE Transactions on Automatic Control*, vol. 68, no. 4, pp. 1986–2001, 2023.
- [17] K. Margellos, M. Prandini, and J. Lygeros, “On the connection between compression learning and scenario based single-stage and cascading optimization problems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2716–2721, 2015.
- [18] S. Garatti and M. C. Campi, “The risk of making decisions from data through the lens of the scenario approach,” *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 607–612, 2021.
- [19] L. G. Valiant, “A theory of the learnable,” *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.
- [20] S. Floyd and M. Warmuth, “Sample compression, learnability, and the Vapnik-Chervonenkis dimension,” *Machine learning*, vol. 21, pp. 269–304, 1995.