Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 54-7 (2021) 607–612

The risk of making decisions from data through the lens of the scenario approach

Simone Garatti * **Marco C. Campi** **

* *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy (e-mail: simone.garatti@polimi.it).*
** *Department of Information Engineering, University of Brescia, Brescia, Italy (e-mail: marco.campi@unibs.it).*

Abstract: In previous contributions, it has been shown that the "complexity" is a key indicator to quantify the "risk" associated to data-driven scenario-based solutions. Depending on the context of application, risk is interpreted as probability of misprediction, or probability of underperforming or meeting shortfalls in various control endeavors, and the acquired ability to tightly evaluate the risk is a vital element in a world where data-driven methods are being increasingly used not only for decision support but also for automated decision making. The present contribution is meant to significantly expand the area of applicability of these results: all achievements so far have been based on an assumption, called "non-degeneracy", that hardly applies e.g. to optimization problems that are not convex. Here, we show that these results maintain their integrity in a non-convex optimization setup, and beyond into a broad domain of decision making that contains non-convex optimization as a particular case.

Keywords: Scenario Approach, Data-Driven Decision Making, Data-Driven Control, Uncertainty Quantitifcation, Machine Learning.

1. INTRODUCTION AND PROBLEM STATEMENT

The scenario approach, Campi et al. (2009b); Campi and Garatti (2018a), is a relatively recent – and yet well established – approach to make data-driven designs in the presence of uncertainty. The driving idea of the method is that the effect of uncertainty can be controlled by knowledge that draws on past experience: one collects a sample of instances of the uncertain elements in the problem (the so-called "scenarios") and makes a decision that is somehow robust against these instances. The goal of the scenario approach is to build decision schemes endowed with precise guarantees against future, and yet unseen, realizations of the uncertainty. Originally, the scenario approach was developed in the context of empirical worst-case optimization, Calafiore and Campi (2005, 2006); Campi and Garatti (2008); Campi and Carè (2013); Schildbach et al. (2013); Carè et al. (2014); Grammatico et al. (2016); Mohajerin Esfahani et al. (2015); Zhang et al. (2015); Campi and Garatti (2018b); Assif et al. (2020); Shang and You (2020), but ever since then various alternative paradigms have emerged and the gallery of approaches has been enriched with methods that include constraints removal and relaxation, Campi and Garatti (2011); Garatti and Campi (2013, 2019); Picallo and Dörfler (2019); Romao et al. (2020), expected shortfall optimization, Ramponi and Campi (2017), variational inequalities and games, Paccagnan and Campi (2019); Fele and Margellos (2020), multi-agent budget constrained optimization, Falsone et al. (2017).

This paper specifically refers to a general and unitary framework – called "scenario decision-making" – that has been introduced in (Garatti and Campi, 2019, Section 5), which covers most of the existing paradigms. Let us briefly review the salient elements of this theoretical framework. Denote by $\mathcal{Z}$ the domain from which a decision z has to be drawn and let δ be a quantity cumulatively containing all the uncertain elements in the problem. It is assumed that δ lives in a probability space $(\Delta, \mathcal{D}, \mathbb{P})$, where $\mathbb{P}$ is a descriptor of the mechanisms through which uncertainty manifests, but the method does not assume any knowledge on $\mathbb{P}$; this accommodates conditions of partial knowledge – or even absence of knowledge – about the mechanisms governing uncertainty that is germane to many decision-making endeavors, especially those relating to complex systems. In the theory, no specific structure, e.g. that of vector space, is required for both $\mathcal{Z}$ and Δ, which are completely generic sets. To each δ, there is associated a set $\mathcal{Z}_\delta \subseteq \mathcal{Z}$ describing the solutions that are "suitable" for that δ, where what suitable means is up to the designer of the method who can capture by an appropriate selection of $\mathcal{Z}_\delta$ various needs that range from performance requirements to satisfaction of constraints of various nature, and the reader is referred to Garatti and Campi (2019) for more discussion on this point. Scenarios δ_i, $i = 1, \ldots, m$, are i.i.d. (independent and identically distributed) draws from $(\Delta, \mathcal{D}, \mathbb{P})$ and one considers decision maps

$$M_m : \Delta^m \to \mathcal{Z}, \quad m = 0, 1, 2, \ldots$$

from the m-dimensional uncertainty domain to the domain of decisions. M_m are required to satisfy the following Assumption 1 in which $z_m^* = M_m(\delta_1, \ldots, \delta_m)$.

Assumption 1. (**consistency**). For every non-negative integers m and n, and for every choice of $\delta_1, \ldots, \delta_m$, and $\delta_{m+1}, \ldots, \delta_{m+n}$, the following three properties hold:

10.1016/j.ifacol.2021.08.427

(i) if $\delta_{i_1}, \dots, \delta_{i_m}$ is a permutation of $\delta_1, \dots, \delta_m$, then it holds that $M_m(\delta_1, \dots, \delta_m) = M_m(\delta_{i_1}, \dots, \delta_{i_m})$;
(ii) if $z_m^* \in \mathcal{Z}_{\delta_{m+i}}$ for all $i = 1, \dots, n$, then it holds that $z_{m+n}^* = M_{m+n}(\delta_1, \dots, \delta_{m+n}) = M_m(\delta_1, \dots, \delta_m) = z_m^*$;
(iii) if $z_m^* \notin \mathcal{Z}_{\delta_{m+i}}$ for one or more $i = 1, \dots, n$, then it holds that $z_{m+n}^* = M_{m+n}(\delta_1, \dots, \delta_{m+n}) \neq M_m(\delta_1, \dots, \delta_m) = z_m^*$. □

The interpretation is that z_m^* is a decision made according to a rule M_m based on a sample of scenarios $\delta_1, \delta_2, \dots, \delta_m$. $\mathcal{Z}_\delta$ is the set of suitable decisions when uncertainty takes value δ and Assumption 1 enforces restrictions on the mechanism by which decisions are made from scenarios. Specifically, (i) requires that M_m is permutation-invariant; (ii) requires that, given m scenarios $\delta_1, \dots, \delta_m$, leading to a decision z_m^*, if more scenarios $\delta_{m+1}, \dots, \delta_{m+n}$ are introduced for which z_m^* is "suitable", then the decision does not change. Instead, if z_m^* is not suitable for at least one of the δ_{m+i}, then according to (iii) the decision must change.

As anticipated, the framework here described bears great generality and accommodates various algorithms that pursue diverse objectives. For the sake of concreteness we introduce here a couple of examples that are relevant to data-driven modeling.

Example 1. (**scenario worst-case optimization**). Let $\theta \in \mathbb{R}^d$ be a vector of optimization variables and let $f(\theta, \delta)$ be a cost function that also depends on the uncertain element δ. Given $\delta_1, \cdots, \delta_m$, let

$$\begin{aligned} \theta_m^* &= \arg\min_{\theta \in \mathbb{R}^d} \max_{i=1,\dots,m} f(\theta, \delta_i), \\ h_m^* &= \max_{i=1,\dots,m} f(\theta_m^*, \delta_i), \end{aligned} \tag{1}$$

i.e., the cost is minimized worst-case with respect to the scenarios, which gives θ_m^* (for simplicity assume that θ_m^* is unique) and the value h_m^*. In this context, we let $z = (\theta, h)$ and we take the definition $\mathcal{Z}_\delta := \{(\theta, h) : \ f(\theta, \delta) \leq h\}$ that expresses the wish that h is a valid upper bound to the cost corresponding to θ for the given δ. Formula (1) defines a map M_m that associates $z_m^* = (\theta_m^*, h_m^*)$ to $\delta_1, \cdots, \delta_m$. This map is clearly permutation invariant. Moreover, when n scenarios $\delta_{m+1}, \dots, \delta_{m+n}$ are added to the original pool, if $z_m^* \in \mathcal{Z}_{\delta_{m+i}}$ for $i = 1, \dots, n$, then $\max_{i=1,\dots,m+n} f(\theta_m^*, \delta_i) = \max_{i=1,\dots,m} f(\theta_m^*, \delta_i)$ and, since for all other θ it clearly holds that $\max_{i=1,\dots,m+n} f(\theta, \delta_i) \geq \max_{i=1,\dots,m} f(\theta, \delta_i)$, θ_m^* remains the optimal solution and h_m^* does not change as well; if instead $z_m^* \notin \mathcal{Z}_{\delta_{m+i}}$, then $h_m^* \neq \max_{i=1,\dots,m+n} f(\theta_m^*, \delta_i)$ and the decision has to change. Hence, M_m as defined by (1) satisfies the consistency Assumption 1. □

Example 2. (**scenario expected shortfall optimization**). Consider again the setup of Example 1 and, for a given θ, denote by $1_m(\theta)$ the index i among $\{1, \dots, m\}$ attaining the largest value of $f(\theta, \delta_i)$, by $2_m(\theta)$ that attaining the second largest value, and so on. Choose a $k \leq m$ and define

$$\begin{aligned} \theta_{k,m}^* &= \arg\min_{\theta \in \mathbb{R}^d} \ \frac{1}{k} \sum_{j=1}^{k} f(\theta, \delta_{j_m(\theta)}) \\ h_{k,m}^* &= f(\theta_{k,m}^*, \delta_{k_m(\theta_{k,m}^*)}) \end{aligned} \tag{2}$$

(again, for simplicity, assume that $\theta_{k,m}^*$ is unique). In (2), $\theta_{k,m}^*$ is chosen by minimizing the average of the k largest costs (this quantity is called expected-shortfall in the literature as the k scenarios that return the largest costs for a given θ are interpreted as "shortfalls"); $h_{k,m}^*$ is the k-th largest cost value corresponding to $\theta_{k,m}^*$ and is meant to be a guarantee on the cost value besides shortfalls. Note that $\theta_{1,m}^* = \theta_m^*$ and $h_{1,m}^* = h_m^*$, while for $k > 1$ the expect-shortfall decision can reduce the intrinsic conservatism of the worst-case decision. If we take z and $\mathcal{Z}_\delta$ as in Example 1, formula (2) defines a new map $M_{k,m}$ that associates $z_{k,m}^* = (\theta_{k,m}^*, h_{k,m}^*)$ to $\delta_1, \cdots, \delta_m$, which again is easily seen to satisfy the consistency Assumption 1. Indeed, $M_{k,m}$ is permutation invariant and, similarly to the worst-case case, when n new scenarios $\delta_{m+1}, \dots, \delta_{m+n}$ are added, if $z_{k,m}^* \in \mathcal{Z}_{\delta_{m+i}}$ for $i = 1, \dots, n$, then $\frac{1}{k}\sum_{j=1}^{k} f(\theta_{k,m}^*, \delta_{j_{m+n}(\theta_{k,m}^*)}) = \frac{1}{k}\sum_{j=1}^{k} f(\theta_{k,m}^*, \delta_{j_m(\theta_{k,m}^*)})$ and, since for all other θ it holds that $\frac{1}{k}\sum_{j=1}^{k} f(\theta, \delta_{j_{m+n}(\theta)}) \geq \frac{1}{k}\sum_{j=1}^{k} f(\theta, \delta_{j_m(\theta)})$, $\theta_{k,m}^*$ remains the optimal solution and also $h_{k,m}^*$ does not change; if instead $z_m^* \notin \mathcal{Z}_{\delta_{m+i}}$ for some i, then $h_{k,m}^* \neq f(\theta_{k,m}^*, \delta_{k_{m+n}(\theta_{k,m}^*)})$ and the decision has to change. □

Example 3. (**Examples 1 and 2 cont'd: application to modeling**). We here apply the setup of Examples 1 and 2 to what we name "coverage models", Campi et al. (2009a); Crespo et al. (2015); Garatti et al. (2019). In this context, δ is an input-output couple: $\delta = (u, y)$, where $u \in \mathbb{R}^q$ is an input and $y \in \mathbb{R}$ is the corresponding output and scenarios $\delta_1, \cdots, \delta_m$ are i.i.d. input/output pairs $(u_1, y_1), \dots, (u_m, y_m)$. We define the cost function as $f(\theta, \delta) = |y - \ell_\theta(u)|$, where $\ell_\theta(u)$ is any parametric map from the input space to the output space (obtained e.g. by means of polynomial or spline expansions or by a neural network). Using (1) one constructs a "layer" in the input/output domain delimited by the two functions $\ell_{\theta_m^*}(u) - h_m^*$ and $\ell_{\theta_m^*}(u) + h_m^*$ (notice that this layer contains all the observations (u_i, y_i)), which can be used to provide prediction intervals for the output y corresponding to a new value of the input u. As we shall see below, the theoretical achievements of this paper rigorously quantify the probability that this prediction is incorrect.

The worst-case layer described above may be adversely affected by the presence of outliers, resulting in wide and hence poorly informative intervals. Resorting to (2), one mitigates this difficulty and the layer $[\ell_{\theta_{k,m}^*}(u) - h_{k,m}^*, \ell_{\theta_{k,m}^*}(u) + h_{k,m}^*]$ has higher accuracy (i.e., smaller width), but, as is intuitive, reduced guarantees of correctness. Again, the theory of the present paper can be used as a tool to rigorously certify the prediction correctness of the layer, a result that can also be profitably used to tune the value of k. □

In what follows, we indicate with N the actual number of scenarios we have available to make a decision (our using a generic m before was because, to derive the result for a given N, we need to refer to any generic size m of the sample – see the proof of the main theorem). The goal of this study is to provide rigorous and generally-applicable tools to evaluate the probability with which the solution z_N^* is "unsuitable" for a new δ, that is $z_N^* \notin \mathcal{Z}_\delta$. We start with the notion of risk for a generic z.

Definition 1. Given any $z \in \mathcal{Z}$, the risk associated to z is $V(z) = \mathbb{P}\{\delta \in \Delta : z \notin \mathcal{Z}_\delta\}$. □

In the context of data-driven decision making, one would like to evaluate $V(z_N^*)$, which, however, is not directly computable because it depends on the unknown $\mathbb{P}$ and indeed most of the results within the scenario theory aim at providing assessments of the unaccessible quantity $V(z_N^*)$. One of the latest directions of investigation pursued in Campi and Garatti (2018b) and Garatti and Campi (2019) is the so-called wait-&-judge approach. The breakthrough result consists in recognizing that there exists an observable, the so called "complexity", from which one can construct an universal estimator of the risk $V(z_N^*)$. To describe the result, we first give the definition of complexity, which also requires to introduce the notion of support set.

Definition 2. (**support set and complexity**). Given a sample $\delta_1, \ldots, \delta_m$, a support set is a tuple of elements extracted from $\delta_1, \ldots, \delta_m$, i.e., $\delta_{i_1}, \ldots, \delta_{i_k}$ with $i_1 < i_2 < \cdots < i_k$, that:

i. gives the same solution as the original sample, that is, $M_m(\delta_1, \ldots, \delta_m) = M_k(\delta_{i_1}, \ldots, \delta_{i_k})$;
ii. is irreducible, that is, no element can be further removed from $\delta_{i_1}, \ldots, \delta_{i_k}$ leaving the solution unchanged.

The smallest cardinality of a support set of $\delta_1, \ldots, \delta_m$ is denoted by s_m^* and is called complexity. [1] Note that the support set can be void (while odd-looking, this situation is possible when the solution with no scenarios is suitable for the drawn scenarios) and in this case $s_m^* = 0$. □

For the given decision problem at hand, once $\delta_1, \ldots, \delta_N$ have been collected, s_N^* can be computed from the definition [2] and hence is an observable quantity. The main achievement of Garatti and Campi (2019) is provided in Theorem 2, which claims that, irrespective of M_m and $\mathbb{P}$, there exists a high correlation between the hidden quantity of interest $V(z_N^*)$ (the risk) and the observable quantity s_N^* (the complexity), so that $V(z_N^*)$ can be tightly estimated from s_N^*. However, this result of Garatti and Campi (2019) requires a non-degeneracy assumption, which considerably limits its applicability.

Assumption 2. (**non-degeneracy**). For any m, with probability 1 there is a unique support set for $\delta_1, \ldots, \delta_m$. □

Unfortunately, degeneracy occurs in many scenario schemes of interest. For example, in the context of Example 1 degeneracy is almost the rule for non-convex cost functions (inspect Figure 2 for an example where there are two support sets). Likewise, non-convexity hampers non-degeneracy in scenario expected shortfall optimization. While the list of examples might be made longer, we limit to these cases in the belief that they suffice to illustrate the stiffness of the non-degeneracy assumption.

This paper moves a fundamental step forward: Theorem 1 – presented and demonstrated for the first time in this

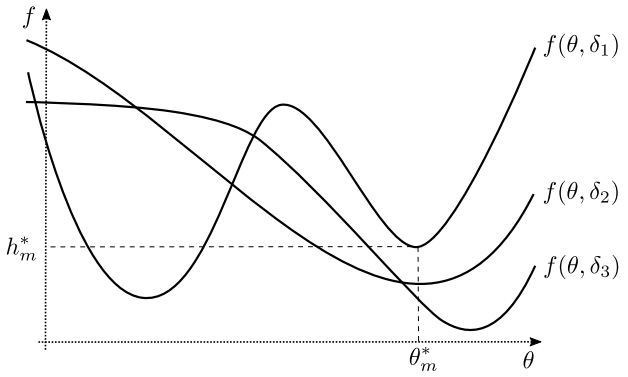

Fig. 1. An instance of scenario worst-case optimization with non-convex $f(x, \delta)$. Both δ_1, δ_2 and δ_1, δ_3 are (minimal) support sets, since both δ_1, δ_2 and δ_1, δ_3 alone suffice to obtain the solution (θ_m^*, h_m^*), while the solution with one single δ_i in place changes.

contribution – proves that $V(z_N^*)$ can be tightly bounded based on s_N^* without any non-degeneracy assumption. We believe that this new finding will foster the use of the scenario approach well beyond its present boundaries.

Theorem 1. Consider decision maps M_m, $m = 0, 1, \ldots$ satisfying Assumptions 1. Given a confidence parameter $\beta \in (0, 1)$, for any $k = 0, 1, \ldots, N-1$ consider the polynomial equation in the v variable

$$\binom{N}{k}(1-v)^{N-k} - \frac{\beta}{N}\sum_{m=k}^{N-1}\binom{m}{k}(1-v)^{m-k} = 0, \qquad (3)$$

and let $\epsilon(k)$ be the unique solution over the interval $(0, 1)$. [3] Also define $\epsilon(N) = 1$. For any $\mathbb{P}$ it holds that [4]

$$\mathbb{P}^N\Big\{V(z_N^*) > \epsilon(s_N^*)\Big\} \leq \beta, \qquad (4)$$

where $z_N^* = M_N(\delta_1, \ldots, \delta_N)$ and s_N^* is the complexity as defined in Definition 2. □

In words, the theorem says that $V(z_N^*) \leq \epsilon(s_N^*)$ holds true with high confidence $1 - \beta$ irrespective of M_m and $\mathbb{P}$ (distribution-free result). Notice that the definition of $\epsilon(k)$ has already appeared in Garatti and Campi (2019); the novelty of the present result is that (4) holds true without any non-degeneracy assumption. By setting β to very small values, like 10^{-6} or 10^{-7}, that are negligible for practical purposes, the result provides usable estimations of the risk $V(z_N^*)$. It is perhaps worth noticing that, though universal, the provided bound $\epsilon(s_N^*)$ is very tight and the interested reader is referred to the discussion in Garatti and Campi (2019) for this aspect.

The proof of Theorem 1 takes a major departure from the proof of the analogous result in Garatti and Campi (2019) and is highly technical. The rest of the paper is devoted to provide this derivation.

[1] This definition of complexity is stated differently from the definition of complexity in Garatti and Campi (2019) (Definition 2 for the case of optimization, then extended to generic decision maps in Section 5). However, under the non-degeneracy Assumption 2 stated below, which can be proven to be equivalent to Assumption 4 in Garatti and Campi (2019), the two definitions of complexity coincide.

[2] A direct application of the definition requires to consider all the combinations of scenarios, which can be a hard combinatorial problem; however, in many cases shortcuts exist to evaluate the complexity, see e.g. Campi et al. (2018).

[3] The fact that the solution is unique is easily seen since, over $(0, 1)$, the equation is equivalent to

$$1 = \frac{\beta}{N}\sum_{m=k}^{N-1}\frac{\binom{m}{k}}{\binom{N}{k}}\frac{1}{(1-v)^{N-m}}$$

whose right-hand side is a continuous strictly increasing function that takes value no bigger than β for $v = 0$ and goes to $+\infty$ as $v \to 1$.

[4] $\mathbb{P}^N$ is the probability distribution for $(\delta_1, \ldots, \delta_N)$ and it is a product probability because of the assumption that scenarios are independently drawn.

2. PROOF OF THEOREM 1

The proof becomes more straightforward in case of no concentrated masses, an assumption which is not satisfied in general by the distribution $\mathbb{P}$ for δ. Hence, to conform to the non-concentrated mass setup, we are well-advised to first augment δ with a second variable η drawn independently of δ in such a way that (δ, η) has no concentrated masses, and then we shall see that the results obtained for (δ, η) can be traced back to δ only. Let $\mathbb{U}$ be the uniform distribution over $[0,1]$ and define $\tilde{\Delta} = \Delta \times [0,1]$, $\tilde{\mathcal{D}} = \mathcal{D} \otimes \mathcal{B}_{[0,1]}$ ($\mathcal{B}_{[0,1]}$ are the Borel sets in $[0,1]$), and $\tilde{\mathbb{P}} = \mathbb{P} \times \mathbb{U}$. For any m, let $\tilde{\delta}_i = (\delta_i, \eta_i)$, $i = 1, \dots, m$, be i.i.d. draws from $(\tilde{\Delta}, \tilde{\mathcal{D}}, \tilde{\mathbb{P}})$. While having no concentrated masses plays a crucial role in some of the derivations to follows, introducing η also allows us to single out a unique support set of minimal cardinality associated to $\delta_1, \dots, \delta_m$. Precisely, let $\mathcal{S}_m : \tilde{\Delta}^m \to \bigcup_{k=0}^m \tilde{\Delta}^k$ be the map that select a subsample from $\tilde{\delta}_1, \dots, \tilde{\delta}_m$, say $\tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k}$, $1 \le i_1 < \cdots < i_k \le m$, such that the first components $\delta_{i_1}, \dots, \delta_{i_k}$ form a support set for $\delta_1, \dots, \delta_m$ of minimal cardinality and, in case this set is not unique, the second components $\eta_{i_1}, \dots, \eta_{i_k}$ minimize the following criterion: for any other support set $\delta_{j_1}, \dots, \delta_{j_k}$ of minimal cardinality it holds that $\sum_{\ell=1}^k \eta_{i_\ell} < \sum_{\ell=1}^k \eta_{j_\ell}$. Since a choice of minimal sum exists with probability one, the previous condition defines $\mathcal{S}_m(\tilde{\delta}_1, \dots, \tilde{\delta}_m)$ except for a zero-probability set. This zero-probability set plays no role in the following derivations and hence $\mathcal{S}_m(\tilde{\delta}_1, \dots, \tilde{\delta}_m)$ can be arbitrarily specified over it.

Turning now back to the problem of evaluating $\mathbb{P}^N\{V(z_N^*) > \epsilon(s_N^*)\}$, clearly $|\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_m)| = s_N^*$ and thus we have that

$$
\begin{aligned}
&\mathbb{P}^N\Big\{V(z_N^*) > \epsilon(s_N^*)\Big\} \\
&= \tilde{\mathbb{P}}^N\Big\{V(z_N^*) > \epsilon(s_N^*)\Big\} \\
&= \tilde{\mathbb{P}}^N\Big\{V(z_N^*) > \epsilon(|\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N)|)\Big\} \\
&= \sum_{k=0}^N \tilde{\mathbb{P}}^N\Big\{|\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N)| = k \text{ and } V(z_N^*) > \epsilon(k)\Big\} \\
&= \sum_{k=0}^N \tilde{\mathbb{P}}^N\Big(\bigcup_{\substack{i_1<i_2<\cdots<i_k: \\ \{i_1,\dots,i_k\}\subseteq\{1,\dots,N\}}} \Big\{\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k} \\
&\qquad\qquad \text{and } V(z_N^*) > \epsilon(k)\Big\}\Big) \\
&= \sum_{k=0}^N \sum_{\substack{i_1<i_2<\cdots<i_k: \\ \{i_1,\dots,i_k\}\subseteq\{1,\dots,N\}}} \tilde{\mathbb{P}}^N\Big\{\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k} \\
&\qquad\qquad \text{and } V(z_N^*) > \epsilon(k)\Big\}, \qquad (5)
\end{aligned}
$$

where the last equality is true because $\eta_1 \neq \eta_2 \neq \cdots \neq \eta_m$ holds with probability one, which implies that subsamples $\tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k}$ are all different from each other with probability one, and, hence, the equality $\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k}$ holds for one and only one choice of the indexes with probability one.

Now, for any fixed k, all the probabilities in the inner summation of (5) are equal because the $\tilde{\delta}_i$'s are i.i.d. draws [5] and so we can write

$$
\begin{aligned}
&\sum_{k=0}^N \sum_{\substack{i_1<i_2<\cdots<i_k: \\ \{i_1,\dots,i_k\}\subseteq\{1,\dots,N\}}} \tilde{\mathbb{P}}^N\Big\{\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_{i_1}, \dots, \tilde{\delta}_{i_k} \\
&\qquad\qquad \text{and } V(z_N^*) > \epsilon(k)\Big\} \\
&= \sum_{k=0}^N \binom{N}{k} \tilde{\mathbb{P}}^N\Big\{\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_1, \dots, \tilde{\delta}_k \\
&\qquad\qquad \text{and } V(z_N^*) > \epsilon(k)\Big\} \\
&= \sum_{k=0}^N \binom{N}{k} \tilde{\mathbb{P}}^N\Big\{\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_1, \dots, \tilde{\delta}_k \\
&\qquad\qquad \text{and } V(z_k^*) > \epsilon(k)\Big\} \\
&\quad \big(\text{this is because, by definition of support set,} \\
&\qquad z_N^* = z_k^* \text{ if } \mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N) = \tilde{\delta}_1, \dots, \tilde{\delta}_k \big) \\
&= \sum_{k=0}^N \binom{N}{k} \int_{(\epsilon(k),1]} \mathrm{d}\mathfrak{m}_{k,N}^+, \qquad (6)
\end{aligned}
$$

where $\mathfrak{m}_{k,N}^+$ is a (positive) measure on $[0,1]$ defined as follows (for future use we introduce a definition that holds for a generic m, and not just for $m = N$): for all $m = 0, 1, \dots$ and $k = 0, \dots, m$, let

$$
\begin{aligned}
&\mathfrak{m}_{k,m}^+(B) \\
&= \tilde{\mathbb{P}}^m\Big\{\mathcal{S}_m(\tilde{\delta}_1, \dots, \tilde{\delta}_m) = \tilde{\delta}_1, \dots, \tilde{\delta}_k \text{ and } V(z_k^*) \in B\Big\},
\end{aligned}
$$

with B any Borel set in $[0,1]$.

Next we show that Assumption 1 implies that measures $\mathfrak{m}_{k,m}^+$ satisfies conditions (i) and (ii) below. Later, these conditions will be enforced when maximizing the right-hand side of (6) with the goal of finding an upper bound to $\mathbb{P}^N\{V(z_N^*) > \epsilon(s_N^*)\}$.

(i) For $m = 0, 1, \dots$, it holds that

$$
\sum_{k=0}^m \binom{m}{k} \int_{[0,1]} \mathrm{d}\mathfrak{m}_{k,m}^+ = 1; \qquad (7)
$$

(ii) For $m = 0, 1, \dots$ and $k = 0, \dots, m$, it holds that

$$
\int_B \mathrm{d}\mathfrak{m}_{k,m+1}^+ - \int_B (1 - v)\, \mathrm{d}\mathfrak{m}_{k,m}^+ \le 0, \qquad (8)
$$

for any Borel set $B \subseteq [0,1]$.

For any given B, the left-hand side of (8) returns a numerical value and, when B ranges over the Borel sets in $[0,1]$, the left-hand side of (8) defines a signed measure. Condition (8) means that this measure is in fact negative. In the following, this measure will be denoted by $\mathfrak{m}_{k,m+1}^+ - (1-v)\mathfrak{m}_{k,m}^+$, [6] and condition (ii) can also be written as

[5] Since M_N is permutation invariant and, in case of ties in the support sets of minimal cardinality, the criterion to break the tie (minimizing $\sum_{\ell=1}^k \eta_{i_\ell}$) does not alter the permutation invariance, $\mathcal{S}_N$ applied to a permutation of $\tilde{\delta}_1, \dots, \tilde{\delta}_N$ returns the same elements as $\mathcal{S}_N(\tilde{\delta}_1, \dots, \tilde{\delta}_N)$ re-ordered according to the permutation. On the other hand, permutation preserves probability because of the i.i.d. property.

[6] Note that $(1-v)\mathfrak{m}_{k,m}^+$ cannot be interpreted as a product since $(1-v)$ is not a number as it depends on v; hence, "$\mathfrak{m}_{k,m+1}^+ - (1 -$

$$\mathfrak{m}^+_{k,m+1} - (1-v)\mathfrak{m}^+_{k,m} \in \mathcal{M}^-,$$

where $\mathcal{M}^-$ is the cone of negative finite measures on $[0,1]$.

Proof of (i): follow the same derivation in (5) and (6) replacing throughout "N" with "m" and both "$V(z^*_N) > \epsilon(s^*_N)$" and "$V(z^*_m) > \epsilon(k)$" with "$0 \le V(z^*_m) \le 1$"; then, notice that $\mathbb{P}^m\{0 \le V(z^*_m) \le 1\} = 1$. □

Proof of (ii): for any given Borel set B in $[0,1]$, we have that

$$\int_B \mathrm{d}\mathfrak{m}^+_{k,m+1} = \tilde{\mathbb{P}}^{m+1}\Big\{\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k \text{ and } V(z^*_k)\in B\Big\}. \tag{9}$$

Over the set where $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k$ (which is part of the condition defining the set on the right-hand side of (9)) it must hold that $z^*_k \in \mathcal{Z}_{\delta_{m+1}}$. As a matter of fact, if $z^*_k \notin \mathcal{Z}_{\delta_{m+1}}$, then, by (iii) in Assumption 1, $z^*_k := M_k(\delta_1,\dots,\delta_k) \neq M_{m+1}(\delta_1,\dots,\delta_k,\delta_{k+1},\dots,\delta_{m+1}) =: z^*_{m+1}$. This implies that $\delta_1,\dots,\delta_k$ is not a support set for $\delta_1,\dots,\delta_{m+1}$ and, therefore, that $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) \neq \tilde\delta_1,\dots,\tilde\delta_k$, which is a contradiction.
Over the set where $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k$ it must also hold that $\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_1,\dots,\tilde\delta_k$. The proof is by contradiction again. Note first that it is not possible that $z^*_m \neq z^*_k$. Indeed, by (ii) in Assumption 1, $M_m(\delta_1,\dots,\delta_k,\delta_{k+1},\dots,\delta_m) =: z^*_m \neq z^*_k := M_k(\delta_1,\dots,\delta_k)$ implies that $z^*_k \notin \mathcal{Z}_{\delta_j}$ for some $j \in \{k+1,\dots,m\}$ and, by (iii) in Assumption 1, this gives $z^*_{m+1} := M_{m+1}(\delta_1,\dots,\delta_k,\delta_{k+1},\dots,\delta_m,\delta_{m+1}) \neq M_k(\delta_1,\dots,\delta_k) =: z^*_k$, which is not possible given that $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k$. Hence, it must be that $z^*_m = z^*_k$ and this implies that $\delta_1,\dots,\delta_k$ is a support set for $\delta_1,\dots,\delta_m$ (note that the irreducibility of $\delta_1,\dots,\delta_k$ – which is in the definition of support set – follows from the fact that $\delta_1,\dots,\delta_k$ is a support set for $\delta_1,\dots,\delta_{m+1}$). To close the proof that $\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_1,\dots,\tilde\delta_k$, suppose for the sake of contradiction that $\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_{i_1},\dots,\tilde\delta_{i_\ell} \neq \tilde\delta_1,\dots,\tilde\delta_k$. This means that $\delta_{i_1},\dots,\delta_{i_\ell}$ is another support set for $\delta_1,\dots,\delta_m$ and that $\tilde\delta_{i_1},\dots,\tilde\delta_{i_\ell}$ is "preferred" by $\mathcal{S}_m$ either because $\tilde\delta_{i_1},\dots,\tilde\delta_{i_\ell}$ has smaller cardinality than $\tilde\delta_1,\dots,\tilde\delta_k$ or because $\tilde\delta_{i_1},\dots,\tilde\delta_{i_\ell}$ ranks better according to the η_i's. If so, however, we would have $M_\ell(\delta_{i_1},\dots,\delta_{i_\ell}) = z^*_m = z^*_k = z^*_{m+1}$, which means that $\delta_{i_1},\dots,\delta_{i_\ell}$ would be a support set for $\delta_1,\dots,\delta_{m+1}$ too. This gives a contradiction because $\tilde\delta_{i_1},\dots,\tilde\delta_{i_\ell}$ would be preferred to $\tilde\delta_1,\dots,\tilde\delta_k$ while, instead, $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k$.
Summarizing, we have proven that $\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k$ implies that $z^*_k \in \mathcal{Z}_{\delta_{m+1}}$ and that $\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_1,\dots,\tilde\delta_k$, yielding

$$\begin{aligned}&\tilde{\mathbb{P}}^{m+1}\Big\{\mathcal{S}_{m+1}(\tilde\delta_1,\dots,\tilde\delta_{m+1}) = \tilde\delta_1,\dots,\tilde\delta_k \text{ and } V(z^*_k)\in B\Big\}\\ &\le \tilde{\mathbb{P}}^{m+1}\Big\{z^*_k \in \mathcal{Z}_{\delta_{m+1}} \text{ and } \mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_1,\dots,\tilde\delta_k \\ &\qquad \text{and } V(z^*_k)\in B\Big\},\end{aligned} \tag{10}$$

because the set on the left-hand side is included in the set on the right-hand side. Using (10) in (9) now gives ($\mathbf{1}_{(\cdot)}$ is the indicator function)

$$\begin{aligned}&\int_B \mathrm{d}\mathfrak{m}^+_{k,m+1}\\ &\le \tilde{\mathbb{P}}^{m+1}\Big\{z^*_k \in \mathcal{Z}_{\delta_{m+1}} \text{ and } \mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m) = \tilde\delta_1,\dots,\tilde\delta_k \\ &\qquad \text{and } V(z^*_k)\in B\Big\}\\ &= \int_{\tilde\Delta^{m+1}} \mathbf{1}_{z^*_k\in\mathcal{Z}_{\delta_{m+1}}} \mathbf{1}_{\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m)=\tilde\delta_1,\dots,\tilde\delta_k \text{ and } V(z^*_k)\in B} \\ &\qquad \mathrm{d}\tilde{\mathbb{P}}^{m+1}(\tilde\delta_1,\dots,\tilde\delta_m,\tilde\delta_{m+1})\\ &= \int_{\tilde\Delta^m}\left(\int_{\tilde\Delta} \mathbf{1}_{z^*_k\in\mathcal{Z}_{\delta_{m+1}}}\ \mathrm{d}\tilde{\mathbb{P}}(\tilde\delta_{m+1})\right)\cdot \\ &\qquad \mathbf{1}_{\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m)=\tilde\delta_1,\dots,\tilde\delta_k \text{ and } V(z^*_k)\in B}\ \mathrm{d}\tilde{\mathbb{P}}^m(\tilde\delta_1,\dots,\tilde\delta_m)\\ &= \int_{\tilde\Delta^m} \big(1 - V(z^*_k)\big)\cdot \mathbf{1}_{\mathcal{S}_m(\tilde\delta_1,\dots,\tilde\delta_m)=\tilde\delta_1,\dots,\tilde\delta_k \text{ and } V(z^*_k)\in B} \\ &\qquad \mathrm{d}\tilde{\mathbb{P}}^m(\tilde\delta_1,\dots,\tilde\delta_m)\\ &= \int_B (1-v)\ \mathrm{d}\mathfrak{m}^+_{k,m},\end{aligned}$$

where the last equality is justified in view of (Rudin, 1970, Theorem 1.29). This concludes the proof of (ii). □

We are now ready to upper-bound $\mathbb{P}\{V(z^*_N) > \epsilon(s^*_N)\}$ by taking the sup of the right-hand side of (6) under conditions (i) and (ii) (in addition to the fact that measures $\mathfrak{m}^+_{k,m}$ belong to the cone $\mathcal{M}^+$ of positive finite measures on $[0,1]$). This gives

$$\mathbb{P}^N\{V(z^*_N) > \epsilon(s^*_N)\} \le \gamma,$$

where γ is defined as the value of the optimization problem

$$\gamma = \sup_{\substack{\mathfrak{m}^+_{k,m}\in\mathcal{M}^+\\ m=0,1,\dots,\\ k=0,\dots,m}} \sum_{k=0}^N \binom{N}{k} \int_{(\epsilon(k),1]} \mathrm{d}\mathfrak{m}^+_{k,N} \tag{11a}$$

$$\text{s.t.} \quad \sum_{k=0}^m \binom{m}{k} \int_{[0,1]} \mathrm{d}\mathfrak{m}^+_{k,m} = 1, \quad m = 0,1,\dots \tag{11b}$$

$$\mathfrak{m}^+_{k,m+1} - (1-v)\mathfrak{m}^+_{k,m} \in \mathcal{M}^-, \tag{11c}$$
$$m = 0,1,\dots;\ k = 0,\dots,m.$$

Problem (11) involves infinitely many constraints. On the other hand, as shown below, it is a fact that all constraints (11b) with $m > N$ and all constraints (11c) with $m > N-1$ are superfluous and can be removed without changing the optimal value of the problem. In formulas,

$$\gamma = \sup_{\substack{\mathfrak{m}^+_{k,m}\in\mathcal{M}^+\\ m=0,\dots,N,\\ k=0,\dots,m}} \sum_{k=0}^N \binom{N}{k} \int_{(\epsilon(k),1]} \mathrm{d}\mathfrak{m}^+_{k,N} \tag{12a}$$

$$\text{s.t.} \quad \sum_{k=0}^m \binom{m}{k} \int_{[0,1]} \mathrm{d}\mathfrak{m}^+_{k,m} = 1,\ m = 0,\dots,N \tag{12b}$$

$$\mathfrak{m}^+_{k,m+1} - (1-v)\mathfrak{m}^+_{k,m} \in \mathcal{M}^-, \tag{12c}$$
$$m = 0,\dots,N-1;\ k = 0,\dots,m.$$

To see this, first notice that the optimal value of (11) cannot be bigger than the optimal value of (12) because

$v)\mathfrak{m}^+_{k,m}$" has to be interpreted just as a symbol that indicates the measure defined via the left-hand side of equation (8).

(11) has more constraints than (12). On the other hand, for any feasible point of (12), say $\bar{\mathfrak{m}}^+_{k,m}$ for $m = 0, \dots, N$ and $k = 0, \dots, m$, by letting: $\mathfrak{m}^+_{k,m} = \bar{\mathfrak{m}}^+_{k,m}$ for $m = 0, \dots, N$ and $k = 0, \dots, m$; $\mathfrak{m}^+_{k,m} = 0$ for $m = N+1, N+2, \dots$ and $k = 0, \dots, m-1$; and $\mathfrak{m}^+_{m,m}$ be a unitary concentrated mass in $v = 1$ for $m = N+1, N+2, \dots$, we obtain a feasible point of (11) that clearly achieves the same cost value as that of $\bar{\mathfrak{m}}^+_{k,m}$ in (12). Hence, it is also true that the optimal value of (11) cannot be smaller than that of (12), and therefore the two optimal values must coincide.

To evaluate γ, we proceed by dualizing (12). A derivation, here omitted due to space limitations and that the reader can find in Garatti and Campi (2020), shows that there is no duality gap and γ can be computed by the dual problem

$$\gamma = \inf_{\lambda_m,\ m=0,\dots,N} \sum_{m=0}^{N} \lambda_m \tag{13a}$$

$$\text{s.t.} \quad \binom{N}{k}(1-v)^{N-k}\mathbf{1}_{v\in(\epsilon(k),1]} \le \sum_{m=k}^{N} \lambda_m \binom{m}{k}(1-v)^{m-k}, \quad \forall v \in [0,1], k = 0, \dots, N. \tag{13b}$$

Summarizing the results so far achieved, we have

$$\mathbb{P}^N \left\{ V(z_N^*) > \epsilon(s_N^*) \right\} \le \gamma,$$

where γ is given by (13). The proof of the theorem is concluded by showing that $\gamma \le \beta$.

To this purpose, take $\lambda_m = \frac{\beta}{N}$ for $m = 0, \dots, N-1$ and $\lambda_N = 0$. These λ_m's are feasible for (13) because (13b) for $k = N$ becomes $0 \le 0$ (recall that $\epsilon(N) = 1$ so that the indicator function is 1 over an empty set), which is true, while (13b) for $k = 0, \dots, N-1$ are satisfied in view of the definition of $\epsilon(k)$ through (3): for $v = \epsilon(k)$, equation (3) implies that (13b) holds with equality, while the monotonicity property shown in Footnote 3 suggests the validity of (13b) for all values of $v \in [0,1]$. Given the feasibility of these λ_m's, we then have $\gamma \le \sum_{m=0}^{N} \lambda_m = \beta$ and this concludes the proof. □

REFERENCES

Assif, M., Chatterjee, D., and Banavar, R. (2020). Scenario approach for minmax optimization with emphasis on the nonconvex case: Positive results and caveats. *SIAM Journal on Optimization*, 30(2), 1119–1143.

Calafiore, G. and Campi, M. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1), 25–46. doi:http://dx.doi.org/10.1007/s10107-003-0499-y.

Calafiore, G. and Campi, M. (2006). The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5), 742–753.

Campi, M., Calafiore, G., and Garatti, S. (2009a). Interval predictor models: identification and reliability. *Automatica*, 45(2), 382–392. doi:http://dx.doi.org/10.1016/j.automatica.2008.09.004.

Campi, M. and Carè, A. (2013). Random convex programs with l_1-regularization: sparsity and generalization. *SIAM Journal on Control and Optimization*, 51(5), 3532–3557.

Campi, M. and Garatti, S. (2008). The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3), 1211–1230. doi: http://dx.doi.org/10.1137/07069821X.

Campi, M. and Garatti, S. (2011). A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2), 257–280.

Campi, M. and Garatti, S. (2018a). *Introduction to Scenario Oprimization*. MOS-SIAM series on Optimization. SIAM.

Campi, M. and Garatti, S. (2018b). Wait-and-judge scenario optimization. *Mathematical Programming*, 167(1), 155–189. doi: https://doi.org/10.1007/s10107-016-1056-9.

Campi, M., Garatti, S., and Prandini, M. (2009b). The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2), 149 – 157.

Campi, M., Garatti, S., and Ramponi, F. (2018). A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12), 4067–4078.

Carè, A., Garatti, S., and Campi, M. (2014). FAST - Fast Algorithm for the Scenario Technique. *Operations Research*, 62(3), 662–671.

Crespo, L., Kenny, S., and Giesy, D. (2015). Random predictor models for rigorous uncertainty quantification. *International Journal for Uncertainty Quantification*, 5(5), 469–489.

Falsone, A., Margellos, K., Garatti, S., and Prandini, M. (2017). Linear programs for resource sharing among heterogeneous agents: the effect of random agent arrivals. In *Proceeedings of the 56th IEEE Conference on Decision and Control*.

Fele, F. and Margellos, K. (2020). Probably approximately correct nash equilibrium learning. *IEEE Transactions on Automatic Control*. doi:10.1109/TAC.2020.3030754. Early access.

Garatti, S. and Campi, M. (2013). Modulating robustness in control design: principles and algorithms. *IEEE Control Systems*, 33(2), 36–51. doi:http://dx.doi.org/10.1109/MCS.2012.2234964.

Garatti, S. and Campi, M. (2019). Risk and complexity in scenario optimization. *Mathematical Programming*. doi: https://doi.org/10.1007/s10107-019-01446-4. Published on-line.

Garatti, S. and Campi, M. (2020). Risk assessment in data-driven decision-making. Internal report.

Garatti, S., Campi, M., and Carè, A. (2019). On a class of interval predictor models with universal reliability. *Automatica*, 110, 108542.

Grammatico, S., Zhang, X., Margellos, K., Goulart, P., and Lygeros, J. (2016). A scenario approach for non-convex control design. *IEEE Transactions on Automatic Control*, 61(2), 334–345.

Mohajerin Esfahani, P., Sutter, T., and Lygeros, J. (2015). Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Transactions on Automatic Control*, 60(1), 46–58.

Paccagnan, D. and Campi, M. (2019). The scenario approach meets uncertain game theory and variational inequalities. In *Proceeedings of the 58th IEEE Conference on Decision and Control*.

Picallo, M. and Dörfler, F. (2019). Sieving out unnecessary constraints in scenario optimization with an application to power systems. In *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*, 6100–6105.

Ramponi, F. and Campi, M. (2017). Expected shortfall: Heuristics and certificates. *European Journal of Operational Research*, 267(3), 1003–1013.

Romao, L., Margellos, K., and Papachristodoulou, A. (2020). Tight generalization guarantees for the sampling and discarding approach to scenario optimization. In *Proceedings of the 589th IEEE Conference on Decision and Control (CDC)*.

Rudin, W. (1970). *Real and Complex Analysis*. McGraw-Hill.

Schildbach, G., Fagiano, L., and Morari, M. (2013). Randomized solutions to convex programs with multiple chance constraints. *SIAM Journal on Optimization*, 23(4), 2479–2501.

Shang, C. and You, F. (2020). A posteriori probabilistic bounds of convex scenario programs with validation tests. *IEEE Transactions on Automatic Control*. doi:10.1109/TAC.2020.3024273. Early access.

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