Mathematical Programming
https://doi.org/10.1007/s10107-019-01446-4

FULL LENGTH PAPER

Series B

Risk and complexity in scenario optimization

S. Garatti[1] · M. C. Campi[2]

Received: 15 February 2017 / Accepted: 1 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2019

Abstract

Scenario optimization is a broad methodology to perform optimization based on empirical knowledge. One collects previous cases, called "scenarios", for the set-up in which optimization is being performed, and makes a decision that is optimal for the cases that have been collected. For convex optimization, a solid theory has been developed that provides guarantees of performance, and constraint satisfaction, of the scenario solution. In this paper, we open a new direction of investigation: the risk that a performance is not achieved, or that constraints are violated, is studied jointly with the complexity (as precisely defined in the paper) of the solution. It is shown that the joint probability distribution of risk and complexity is concentrated in such a way that the complexity carries fundamental information to tightly judge the risk. This result is obtained without requiring extra knowledge on the underlying optimization problem than that carried by the scenarios; in particular, no extra knowledge on the distribution by which scenarios are generated is assumed, so that the result is broadly applicable. This deep-seated result unveils a fundamental and general structure of data-driven optimization and suggests practical approaches for risk assessment.

Keywords Data-driven optimization · Scenario approach · Stochastic optimization · Probabilistic constraints

Mathematics Subject Classification 90C15 · 90C25 · 62C12

This research has been supported by the University of Brescia under the Project CLAFITE.

✉ M. C. Campi
marco.campi@unibs.it
http://bsing.ing.unibs.it/∼campi/

S. Garatti
simone.garatti@polimi.it
http://home.deib.polimi.it/sgaratti/

1 Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

2 Dipartimento di Ingegneria dell'Informazione, Università di Brescia, Via Branze 38, 25123 Brescia, Italy

Published online: 20 November 2019

1 Introduction

The scenario approach is a framework to perform optimization in uncertain environments where one has access to a record of past cases for the set-up where the present decision has to be made. The prototype convex scenario optimization problem is written as

$$\begin{aligned} &\min_{x \in \mathcal{X}} \quad c^T x \\ &\text{subject to}: \quad x \in \bigcap_{i=1,\dots,N} \mathcal{X}_{\delta_i}, \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^d$ is a vector of optimization variables and c is a constant vector of weights; $\mathcal{X} \subseteq \mathbb{R}^d$ is a convex set and $\mathcal{X}_{\delta_i}$ are instances of a family $\{\mathcal{X}_\delta\}$ of convex constraint sets parameterized by δ. Parameter δ is modeled as a random element from a probability space $(\Delta, \mathcal{F}, \mathbb{P})$, and δ_i, $i = 1, \dots, N$, is an independent random sample of δ values.[1] The interpretation of (1) is that the δ_i's are observations (or "scenarios") of an uncertain phenomenon (e.g., demand in the energy market or temperature in a given environment), and one makes a decision which is optimal according to the cost function $c^T x$ (e.g., minimize energy production or minimize the ranges for the forecast of monthly temperatures) while also satisfying the constraints that come from previous cases δ_i's (e.g., energy balance for a record of demands or correctly describing the temperatures that have been recorded in the past). See [8,28,60] for broader presentations of data-driven optimization.

The optimization problem (1) is convex, and this sets its fundamental structure: the optimization domain, $\mathcal{X}$, as well as the constraints, $x \in \mathcal{X}_{\delta_i}$, are convex and the cost function is linear. Note that linearity of the cost function is not a limiting assumption within a convex set-up because any problem with a convex, but nonlinear, cost function can be re-written as one with a linear cost function by an epigraphic reformulation, [9]. Convexity makes solving (1) computationally tractable even in the presence of many optimization variables.

Although clearly not all problems are convex, the set-up of (1) is truly vast and encompasses problems that come from a variety of fields that range from finance, [39, 51–53], to control, [20,34,36,55,68], from prediction, [15,23–25], to machine learning, [14,47]. A first common situation is the minimization of a loss function $\ell(v, \delta)$ that depends on one's choice v and on an uncertain variable δ.[2] Given a sample of scenarios, solving the worst-case problem

$$\min_v \max_{i=1,\dots,N} \ell(v, \delta_i)$$

[1] No limitations are imposed on Δ like e.g. that Δ is a subset of a Eucledian space or of a vector space, nor is Δ endowed with a metric or a topology. Δ is just a generic set that forms a probability space together with $\mathcal{F}$ and $\mathbb{P}$. Hence, ideas like "the sample δ_i, $i = 1, \dots, N$, covers, or fills up, Δ" are void of any meaning. This generality in the definition of Δ is important for the widespread applicability of the theory.

[2] We assume that function $\ell(v, \delta)$ is convex in v for any given value of δ, while its dependence on δ is arbitrary.

is equivalent to (1) after one introduces a new variable $h \in \mathbb{R}$, called the "performance variable", and set $x = (v, h)$, $c^T x = [0 \ \cdots \ 0 \ 1]x = h$, and $\mathcal{X}_{\delta_i} = \{(v, h) : h \geq \ell(v, \delta_i)\}$. In this context, enforcing the constraints $x \in \mathcal{X}_{\delta_i}$, $i = 1, \ldots, N$, leads to worst-case optimization and one makes a choice such that no other selection of the optimization variable v would lead to a better value simultaneously over all the scenario-based loss functions $\ell(v, \delta_i)$. In finance, this set-up has been described in [33]. More generally, constraints $x \in \mathcal{X}_{\delta_i}$ in (1) reflect needs of various type that go from saturation limits in control applications to obstacle avoidance in mobile robotics, from resource availability in management problems to bandwidth capacity in telecommunications.

1.1 A theory of generalization

In recent years, much effort has been spent in the stochastic optimization literature towards studying the properties that sample-based solutions exhibit when applied to new out-of-sample cases, [3,4,26,27,43,44,46,51,58]. By using a terminology imported from machine learning, this problem is also referred to as the "generalization" problem as it involves extending, or generalizing, properties to new and yet unseen situations. In this section, we specifically refer to the properties of the solution obtained by solving (1).

To describe the existing generalization results, we start by introducing the notation x_N^* for the solution to the optimization problem (1),[3] and the following definition of risk.

Definition 1 (*risk*) The *risk* of a given $x \in \mathcal{X}$ is defined as

$$V(x) = \mathbb{P}\{\delta \in \Delta : x \notin \mathcal{X}_\delta\}.$$

Hence, $V(x)$ is the probability with which constraints are not satisfied by x. Note that $V(\cdot)$ is a deterministic function from $\mathcal{X}$ to $[0, 1]$. The risk of x_N^* is the random variable $V(x_N^*)$ obtained by computing $V(\cdot)$ corresponding to the solution x_N^* of (1). Note that $V(x_N^*)$ is stochastic through the dependence of x_N^* on $\delta_1, \delta_2, \ldots, \delta_N$. □

When the constraints stem from an uncertain loss function $\ell(v, \delta)$ as described above, $V(x) = V(v, h)$ quantifies the probability that in a new case the loss associated with v exceeds h, so that the risk of (v_N^*, h_N^*) is the probability that applying the choice v_N^* results in a loss greater than h_N^*. More generally, $V(x_N^*)$ is a measure of the probability that some undesired event or condition occurs when the solution x_N^* is applied. If $V(x_N^*) \leq \epsilon$, then the risk for the solution to violate the random constraints $x \in \mathcal{X}_\delta$ is no more than ϵ. According to the stochastic programming terminology, this is expressed that x_N^* is a chance-constrained feasible point at level ϵ, [29,58].[4]

[3] Throughout, we assume that a solution exists. If more than one solution exists, a solution is singled out by means of a convex tie-break rule according to the approach of [10].

[4] We remark that $V(x_N^*)$ quantifies the risk, which refers to the chance-constrained feasibility, while the value is not at issue here.

Together with the cost value $c^T x_N^*$, $V(x_N^*)$ represents the fundamental quantity to evaluate the level of satisfaction one has in the solution x_N^*. Interestingly, the cost value $c^T x_N^*$ becomes available to the user after the solution x_N^* has been computed. In contrast, the value of $V(x_N^*)$ depends on the distribution $\mathbb{P}$ of δ, which in real applications is normally unknown or only imprecisely known: hence, $V(x_N^*)$ cannot be computed even after the optimization process has been completed. The problem of estimating $V(x_N^*)$ without resorting to extra observations or test sets (which may involve costly and limited resources) has attracted much attention over the past 10 years, and deep results have been established which (in a sense that we shall discuss in detail later) affirm that the distribution of $V(x_N^*)$ can be bounded even when no knowledge on the distribution of δ is available. This problem has been studied in [11,17] and then extended in various directions including constraint violation, [18], regularization, [16], non-convex optimization, [1,31,36], multi-stage problems, [62], and risks at various empirical levels, [22]. Moreover, papers [21,50] introduce algorithms to attain a solution that carries reduced risks. See also [5,44,45,48,49,69] for studies on the connection between scenario optimization and chance-constrained problems. All these results have put the scenario approach on solid quantitative grounds, a fact that has had a role in the widespread acceptance of this methodology in various application domains. In the next subsection, we specifically describe the mathematical results that are relevant to place the contribution of this paper in context, and then introduce the new perspective of this paper.

1.2 Previous results and the approach of this paper

In the paper [17], the fundamental relation

$$\mathbb{P}^N\{V(x_N^*) \leq \epsilon\} \geq 1 - \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \tag{2}$$

has been established, where $\mathbb{P}^N$ refers to the sample $(\delta_1, \delta_2, \ldots, \delta_N)$ by which x_N^* is determined ($\mathbb{P}^N$ is a product probability due to independence of $\delta_1, \delta_2, \ldots, \delta_N$). Equation (2) bounds the cumulative probability distribution of $V(x_N^*)$, the bound is universally valid for any scenario optimization problem in the form of (1) and, importantly, it is not improvable since it is exact (i.e., $\mathbb{P}^N\{V(x_N^*) \leq \epsilon\} = 1 - \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$) for a class of problems, the so-called "fully-supported" problems according to a definition introduced in [17]. The right-hand side of (2) is the cumulative distribution of a Beta variable with parameters $(d, N-d+1)$. Figure 1 displays in solid blue line the corresponding density when $d = 400$ and $N = 1000$ (respectively, number of optimization variables and number of scenarios). In the same figure one can also see in dashed yellow and dotted red lines the density of $V(x_N^*)$ for two scenario optimization problems that are not fully-supported (these two optimization problems are presented in detail in the simulation Sect. 4). For these problems, equation (2) holds with strict inequality.

In the later paper [19], it was observed that optimization problems encountered in applications are often not fully-supported, see also [56,63,64,67]. Moreover, by

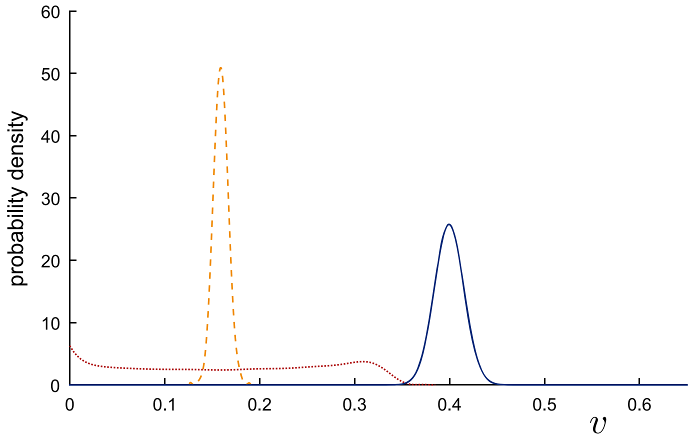

Fig. 1 Beta probability density for $d = 400$, $N = 1000$ (solid blue) and probability density of $V(x^*_N)$ for two non fully-supported scenario optimization problems (dashed yellow and dotted red) (color figure online)

counting the number of support constraints for the case at hand, one knows that the problem is not fully-supported, and, hence, can be unwilling to use the theory of [17] that is tight for fully-supported problems only. Based on this observation, a new approach was introduced in [19] where one waits before forming an evaluation on $V(x^*_N)$, and the evaluation is based on the number of support constraints that have been found in the instance of the scenario optimization problem at hand. This number of support constraints is interpreted as the complexity of the scenario optimization problem (1).

The present paper builds on the approach initiated in [19] and we herein fully develop a new theory for the study of the joint distribution of the risk and the complexity of scenario optimization problems. This theory reveals a fundamental correlation structure that links the risk to the complexity and has important implications in applications. We start by making formal in the next definition the concept of support constraint, taken from [10], and that of complexity.

Definition 2 (*support constraint and complexity*) A constraint $x \in \mathcal{X}_{\delta_i}$ of the scenario optimization problem (1) is called a *support constraint* if its removal (while all other constraints are maintained) changes the solution x^*_N.[5] The complexity of the scenario optimization problem (1) is the number of its support constraints. □

In paper [10] it is shown that the number of support constraints of (1) is always less than or equal to d, the number of optimization variables, and, in case of fully-supported problems, (1) has d support constraints with probability 1, whenever $N \geq d$.

A support constraint is necessarily an active constraint. The converse is not true in general, and an active constraint need not be a support constraint as it can be easily understood by considering a situation where, after finding the solution, one more

[5] Similarly to problem (1), it is assumed that the problems obtained after removing one constraint from (1) admit a unique solution, possibly after breaking the tie by means of a convex tie-break rule according to the approach of [10].

active constraint is added: removing this constraint does not change the solution and this constraint is not of support. In extreme cases, situations can occur where no support constraints exist. When all the active constraints are support constraints, which is the typical case, keeping the support constraints and removing all the other constraints leaves the solution unchanged. Following [19], we call this situation non-degenerate. Non-degeneracy rules out situations in which the constraints accumulate anomalously with nonzero probability.

Definition 3 (*non-degeneracy*) Optimization problem (1) is called non-degenerate if its solution coincides with probability 1 (with respect to the sample $(\delta_1, \delta_2, \ldots, \delta_N)$) with the solution that is obtained after eliminating all the constraints that are not of support. □

If (1) is non-degenerate, we can reconstruct the solution x_N^* by only using the support constraints. The number of support constraints is therefore a measure of the complexity of representation of x_N^* in a scenario optimization problem that has a reduced number of constraints. For short, we at times speak of "complexity of the solution" to mean the complexity of the optimization problem that has generated the solution.

Let s_N^* be the complexity of the optimization problem (1) and $(s_N^*, V(x_N^*))$ be the bivariate variable of complexity and risk. Since δ_i, $i = 1, \ldots, N$, are independent random elements from $(\Delta, \mathcal{F}, \mathbb{P})$, the N-dimensional sample $(\delta_1, \delta_2, \ldots, \delta_N)$ is a random element from $(\Delta^N, \mathcal{F}^N, \mathbb{P}^N)$ (we recall that the probability is a product probability due to independence of δ_i, $i = 1, \ldots, N$), and so $(s_N^*, V(x_N^*))$ is a bivariate random variable over $(\Delta^N, \mathcal{F}^N, \mathbb{P}^N)$ taking value in $\{0, 1, \ldots, d\} \times [0, 1]$. In this paper we study the distribution of $(s_N^*, V(x_N^*))$, that is, the joint distribution of complexity and risk, and Theorem 1 in Sect. 3 establishes a deep-seated result that this distribution is concentrated so that the risk $V(x_N^*)$ can be estimated from the complexity s_N^*. Figure 2 displays a 99% region obtained from Theorem 1 for the distribution of $(s_N^*, V(x_N^*))$

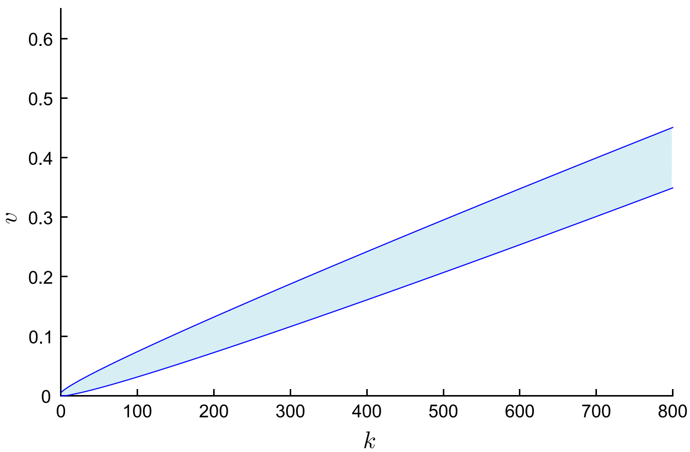

Fig. 2 99% region for the distribution of $(s_N^*, V(x_N^*))$ when $d = 800$ and $N = 2000$. Horizontal axis: value of s_N^* (k = value of s_N^* in the discrete set $\{0, 1, \ldots, 800\}$); vertical axis: value of $V(x_N^*)$ (v = value of $V(x_N^*)$ in the continuous interval [0, 1])

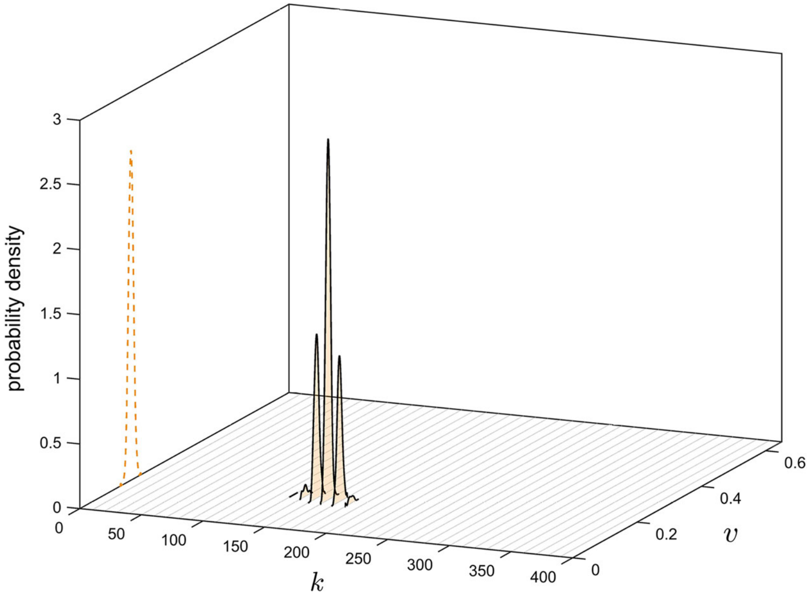

Fig. 3 Distribution of $(s_N^*, V(x_N^*))$ for an example discussed in Sect. 4

when $d = 800$ and $N = 2000$. Given the value of s_N^*, which can be easily computed,[6] $V(x_N^*)$ is estimated to be in the line segment that is obtained by cutting the 99% region with the vertical line that originates from the value of s_N^*. Importantly, the result in Theorem 1 holds independently of the probability $\mathbb{P}$ so that it can be used without requiring any additional information on the distribution of δ. In simple words, this means that the complexity carries universal information for the judgement of the risk, and this fact makes the theory of this paper broadly usable in applications where the underlying distribution that generates the δ's is not or only partly known.

Before closing this Sect. 1.2, we feel it advisable to make explicit two facts that are consequences of what has been discussed so far.

(i) First, we compare, and better contrast, the result from [17] with that of this paper. Figure 3 displays the distribution of $(s_N^*, V(x_N^*))$ for an example discussed in Sect. 4, along with the corresponding marginal distribution of $V(x_N^*)$ (dashed yellow line in the figure). This marginal has been already displayed in Fig. 1. Similarly, Fig. 4 shows the distribution of $(s_N^*, V(x_N^*))$ for another example also discussed in Sect. 4, and the marginal of $V(x_N^*)$ is that displayed in dotted red in Fig. 1.

The two bivariate distributions are quite different, but both are concentrated in the 99% region of Fig. 2. The two marginals for $V(x_N^*)$ have a dissimilar shape. As a result, if one studies the distribution of $V(x_N^*)$ alone, as is done in paper [17], then the problem arises that various behaviors are encountered depending on the scenario optimization problem at hand, so that tight results valid for all cases are not possible. On the other hand, adopting the broader point of view of studying jointly s_N^* and $V(x_N^*)$

[6] To this purpose, it is enough to eliminate one by one the constraints and recompute the solution, the support constraints are those whose elimination determines a change in the solution.

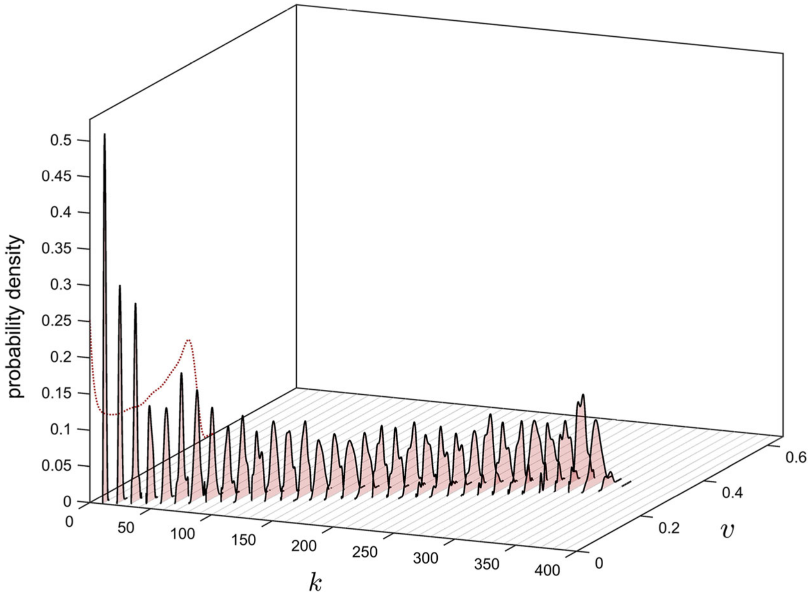

Fig. 4 Distribution of $(s_N^*, V(x_N^*))$ for another example discussed in Sect. 4

sheds light on the structure of dependence between these two variables and the value of the hidden variable $V(x_N^*)$ can be estimated from the value of the easy-to-measure variable s_N^* independently of all other elements that characterize each problem.

(ii) Strictly connected with point (i), a second fact is worth mentioning. For a given problem, one can investigate a number of specificities beyond complexity (which includes the shape of the constraints, the location of the constraints in the optimization domain, etc.) and one can also possibly use prior knowledge to refine the judgement of the risk. When doing so one has to keep in mind that the margin of improvement in the judgment of the risk is limited to only reducing the remaining spread in the value of $V(x_N^*)$ in the bivariate distribution of $(s_N^*, V(x_N^*))$, which shows that s_N^* already contains much of the information to judge the value of $V(x_N^*)$.

1.3 Extension to scenario-based decision problems

The scope of the present contribution goes beyond the formulation of (1) and Sect. 5 deals with a more abstract scenario-based decision framework that covers problems even beyond the domain of optimization. Leveraging this broader perspective, Sect. 5.1 presents a theory where the non-degeneracy condition is loosened. Further, Sect. 5.2 deals with "constraints relaxation": differently from (1), the solution is not required to satisfy all of the constraints introduced by the scenarios, instead the user is allowed to tune the level of empirical risk (proportion of violated scenario constraints) so as to achieve an improvement of the cost value. Interestingly, in the generalized context of Sect. 5 all the concepts introduced in the first part of the paper maintain their essential structure and mutual relation, while they are formulated more abstractly. In particular, the concept of complexity is still interpreted in terms of complexity of representation. The main thrust of Sect. 5 is that, under a non-degeneracy condition, the complexity is a tool that can be used to always tightly judge the level of risk (Theorem 2). This has important conceptual implications: two scenario decision problems with the same empirical risk can have quite different true risks depending on undisclosed mechanisms by which the satisfaction of some constraints implies the satisfaction of other constraints. Nonetheless, it is a universal fact that all these mechanisms are captured by the complexity, which, alone, allows one to derive tight evaluations.

1.4 Structure of the paper

After the next section where a comparison with other methods is provided, Sect. 3 presents the main result, Theorem 1, and a discussion of its use. An example is provided in Sect. 4 that aims to illustrate the theory. In Sect. 5, we broaden our point of view and show that the theory of this paper carries over to more general decision problems, including a scheme to trade risk for performance. All the proofs are given in Sect. 6 and conclusions are drawn in Sect. 7.

2 Comparison with other methods

The scenario approach is one method to deal with data-driven optimization and other approaches exist that address the problem from a different angle that focuses more on the domain of uncertainty. These alternative approaches may offer a more flexible tool of study than the scenario approach because working in the domain of uncertainty allows one to cast the problem more freely in relation to various risk measures. On the other hand, these alternative approaches may suffer from the fact that working directly with the domain of uncertainty is possible only when this domain is simple enough. In this section, we feel it advisable to present some comparison; we start by suggesting that, from the scenario approach's perspective, the domain of uncertainty is only a mathematical tool, which plays no role in the operation of the method. This is followed by a description of the advantages that may arise from working directly with the domain of uncertainty.

One aspect which is distinctive of the scenario approach is that it deals with observations only in terms of the impact they have on the decision problem. To better explain this assertion in mathematical terms, recall that the scenarios are realizations from a probability space $(\Delta, \mathcal{F}, \mathbb{P})$. One such realization, δ_i, introduces a constraint $x \in \mathcal{X}_{\delta_i}$ in the optimization problem (1), while Δ and $\mathbb{P}$ never enter the practice of scenario optimization, neither algorithmically, nor when certificates on the risk are generated. This is crucial for the operation of the method: in applications, the domain Δ may refer to a truly complex milieu, and indeed today's use of data-driven optimization in social sciences, engineering and medicine refers to ever more complex uncertainty domains. The reader may find an exemplification of this fact in the papers [2,13] where the scenario approach has been used with the goal of designing a model to predict the effect of a defibrillator shock given to patients in cardiac arrest. In this case, a realization of δ refers to the ECG trace of a patient and Δ is a complex infinite dimensional functional domain. While the domain Δ is often too complex to deal with, the domain for optimization is usually a simple Euclidean space. This is not surprising as Δ refers to the real world, while the optimization domain refers to the world of decisions as we, the users, construct it, often with an attention to simplicity of use and implementation.

Since the scenario approach operates at the optimization domain level, it lends itself very well to be applied independently of how complex Δ is or is meant to be (when adopting the scenario approach, the user is not even requested to describe Δ). On the other hand, as it has been noted before, in the optimization literature there has been a surge of interest for methods that make explicit use of the domain Δ. These methods offer valid alternatives when Δ has a simple enough structure, in which case they may suggest a more general perspective as reviewed in the following.

In these methods, data are used to construct an "ambiguity set" $\mathcal{P}$ of probability distributions over Δ and a decision which is robust with respect to $\mathcal{P}$ is made (data-driven Distributionally Robust Optimization—DRO). If $\mathcal{P}$ comes accompanied by a guarantee that $\mathbb{P} \in \mathcal{P}$ with a known probability $1-\beta$, then a given property of the decision which is guaranteed to hold over $\mathcal{P}$ translates into a property certified for $\mathbb{P}$ with probability $1 - \beta$. Three principal approaches have been adopted to construct ambiguity sets: (i) moment ambiguity sets, [12,28,35,38,65,69]; (ii) ambiguity sets defined as balls according to a metric, e.g., Prohorov [30] and Wasserstein [32,54,66], or to the Kullback-Leibler divergence [40,41,61]; (iii) ambiguity sets defined by statistical confidence regions, [6,7]. Moment ambiguity sets have better tractability properties, while they in general do not allow one to obtain asymptotic convergence, a fact already noticed in [7]. This is not surprising since along this approach the information in the data is compressed with loss into confidence intervals for the moments; a similar difficulty is not germane to the other two approaches (ii) and (iii). Regardless of the approach used to construct an ambiguity set, a difficulty with DRO methods is that it is not easy to obtain ambiguity sets that minimize the impact on the optimization cost while maintaining a desired level of confidence, which may reflect into conservatism of the solution. This aspect has been discussed in [37,42,61], which have also suggested remedies under suitable assumptions by using ambiguity sets that deviate from the requirement that the ambiguity set contains $\mathbb{P}$ with probability $1 - \beta$. One important advantage of the DRO approach is that an ambiguity set offers a tool that can be used independently of the adopted risk measure, so offering extra flexibility. Particularly, DRO methods lend themselves to studying the minimization of expected costs and to quantify the mismatch between the expected cost and its empirical counterpart, which is not within the scope of the the scenario approach.

Before closing this section, we offer one last remark in relation to the specific problem of evaluating the risk $V(x)$. The first part of this paper concentrates on the risk that is achieved when a value of x is obtained by solving problem (1). The data-driven problem (1) has an intuitive appeal, and the result of this paper aims at quantifying the risk of a methodology that pre-exists its analysis. In this context, the scenario theory offers results (the fundamental Theorem 1 in primis and also all the other theorems) that hold for any $\mathbb{P}$, with no limitations on the underlying data generative model. Importantly, despite their generality, the results are tight and informative, as described in detail in Sect. 3.1. While generality and tightness leave little room for improvement in the context of study of this paper, the DRO approach has a different goal of constructing methods to find an optimal solution while enforcing a desired level of risk. This approach, however, requires proper shaping of the ambiguity set which is prohibitive whenever Δ is a complex domain.

3 Main result and its practical use

Studying the optimization problem (1) that is based on N scenarios requires considering infinitely many other scenario optimization problems that have the same structure as (1) with, however, an arbitrary number of scenarios:

$$\begin{aligned} &\min_{x \in \mathcal{X}} \quad c^T x \\ &\text{subject to :} \quad x \in \bigcap_{i=1,\ldots,m} \mathcal{X}_{\delta_i}, \end{aligned} \tag{3}$$

where $m = 0, 1, 2, \ldots$ is any integer; it is meant that the "subject to" line is dropped when $m = 0$; and, δ_i, $i = 1, \ldots, m$, is an independent sample from $(\Delta, \mathcal{F}, \mathbb{P})$. Notice that (1) is the same as (3) when $m = N$.

Assumption 1 (*existence and uniqueness*) For every m and for every sample $(\delta_1, \delta_2, \ldots, \delta_m)$, optimization problem (3) admits a solution (i.e. the optimization problem is feasible and the infimum is achieved on the feasibility set). If more than one solution exists, one solution is singled out by the application of a convex tie-break rule, which breaks the tie by minimizing an additional convex function $t_1(x)$, and, possibly, other convex functions $t_2(x)$, $t_3(x)$, …if the tie still occurs. The so-obtained solution is indicated with x_m^* and it will be simply called the "solution" of (3). □

The approach for breaking the tie in Assumption 1 is the same as that in [10]. An example of a tie-break function is the norm of x, $t_1(x) = \|x\|$. Another example is the lexicographic rule, which consists in minimizing the components of x in succession, i.e., $t_1(x) = x_1$, $t_2(x) = x_2$, ….

The notion of support constraint in Definition 2 extends in an obvious manner to (3): a constraint $x \in \mathcal{X}_{\delta_i}$ of (3) is called a support constraint if its removal (while other constraints are maintained) changes the solution x_m^*. We introduce the following non-degeneracy assumption.

Assumption 2 (*non-degeneracy*) For every m, the solution x_m^* to the optimization problem (3) coincides with probability 1 (with respect to the sample δ_i, $i = 1, \ldots, m$) with the solution that is obtained after eliminating all the constraints that are not of support. □

It is perhaps worth noticing that, while the definition of support constraint concerns removing one constraint at time, in Assumption 2 the simultaneous removal of all the non-support constraints is considered. Assumption 2 is a mild condition that excludes that the constraints accumulate anomalously at the solution. More comments on this assumption are provided in Sect. 5.1 where it is shown that this assumption can be relaxed by using a more abstract theory.

The following theorem is our main result in this first part of the paper. It characterizes the distribution of the bivariate variable $(s_N^*, V(x_N^*))$, where s_N^* is the complexity of the solution (the number of support constraints) and $V(x_N^*)$ is its risk (the probability that a new constraint is not satisfied).

Theorem 1 *Consider the optimization problem (1) with $N > d$. Given a confidence parameter $\beta \in (0,1)$, for any $k = 0, 1, \ldots, d$ consider the polynomial equation in the t variable*

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0. \tag{4}$$

This equation has exactly two solutions in $[0, +\infty)$, which we denote with $\underline{t}(k)$ and $\overline{t}(k)$ ($\underline{t}(k) \leq \overline{t}(k)$). Let $\underline{\epsilon}(k) := \max\{0, 1 - \overline{t}(k)\}$ and $\overline{\epsilon}(k) := 1 - \underline{t}(k)$. Under Assumptions 1 and 2, for any Δ and $\mathbb{P}$ it holds that

$$\mathbb{P}^N\{\underline{\epsilon}(s_N^*) \leq V(x_N^*) \leq \overline{\epsilon}(s_N^*)\} \geq 1 - \beta. \tag{5}$$

Proof The proof is given in Sect. 6. □

The theorem assigns lower and upper bounds on $V(x_N^*)$ that hold with confidence $1-\beta$. The bounds depend on the random variable s_N^*, the number of support constraints, which can be assessed by the user after computing the solution x_N^*. A more explicit way of writing (5) is

$$\mathbb{P}^N\left(\bigcup_{k=0}^{d}\{s_N^* = k \text{ and } \underline{\epsilon}(k) \leq V(x_N^*) \leq \overline{\epsilon}(k)\}\right) \geq 1 - \beta. \tag{6}$$

Hence, a user who computes the number of support constraints and claims the risk to be between the limits $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ (where k is the assessed number of support constraints) incurs a probability $1 - \beta$ of making a wrong statement. This result holds true for all convex scenario optimization problems in the form of (1) under Assumptions 1 and 2. Figure 5 shows the confidence regions obtained from Theorem 1 for various values of N and d. For practical values of N and d, the two functions $\underline{\epsilon}(\cdot)$ and $\overline{\epsilon}(\cdot)$ are close enough to each other so that, after evaluating s_N^*, tight and useful information about $V(x_N^*)$ is obtained. As N grows, the two functions get progressively closer and eventually converge to one another.

3.1 Distribution-free has little cost

As we have seen, Theorem 1 applies independently of the distribution $\mathbb{P}$, that is, it is a distribution-free result. Hence, it can be used without knowledge of $\mathbb{P}$, a fact that plays an important role in applications. In this section we show that there are optimization problems corresponding to given probability measures $\mathbb{P}$ for which the distribution of the bivariate variable $(s_N^*, V(x_N^*))$ has support that is not too dissimilar to what is found by applying Theorem 1. Hence, the price paid for a distribution-free result is small relative to knowing that one of these optimization problems is being run. The interpretation is that the number of support constraints carries the fundamental information to judge the risk, and the residual uncertainty in the risk after the number

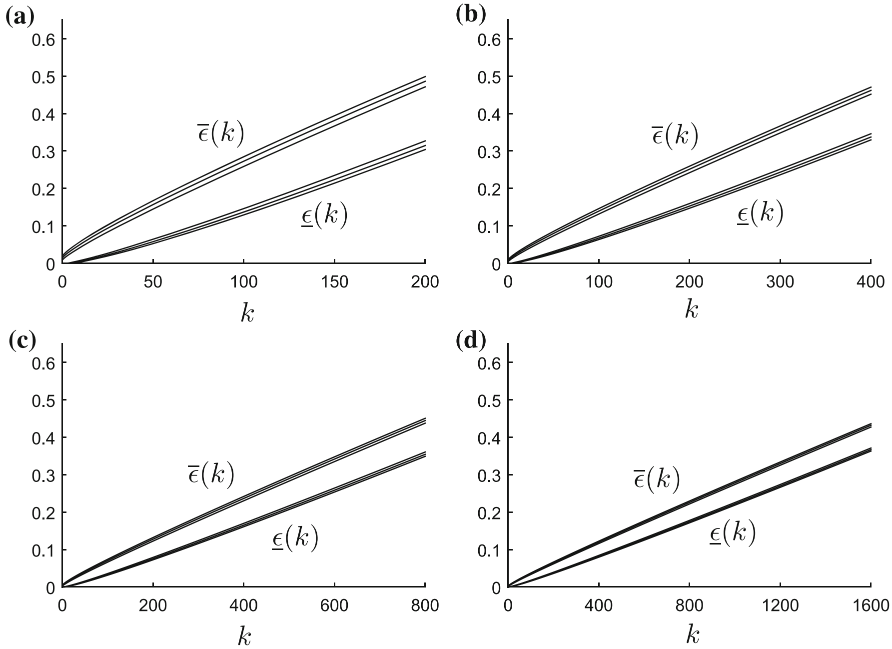

Fig. 5 Profile of curves $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$. **a** $N = 500$, $d = 200$; **b** $N = 1000$, $d = 400$; **c** $N = 2000$, $d = 800$; **d** $N = 4000$, $d = 1600$. The three curves are for $\beta = 0.1$, $\beta = 0.01$; $\beta = 0.001$

of support constraints has been seen (two samples of scenarios that lead to the same number of support constraints may carry a different risk) is only marginally increased by requiring that the result holds distribution-free.

To put the above discussion on solid grounds, consider a fully-supported problem in dimension k, see [17] for a definition of fully-supported problem. For such a problem, the number of support constraints is k with probability 1. It is not hard to embed this problem into another problem that has d optimization variables while it continues to have k support constraints with probability 1, so that $s_N^* = k$ with probability 1. If we now apply Theorem 1 in [17] to this problem we see that the distribution of $V(x_N^*)$ is a Beta$(k, N-k+1)$ distribution, i.e., $\mathbb{P}^N\{V(x_N^*) \leq \epsilon\} = 1 - \sum_{i=0}^{k-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$. Now let $\underline{\alpha}(k)$ be the value such that $1 - \sum_{i=0}^{k-1} \binom{N}{i} \underline{\alpha}(k)^i (1-\underline{\alpha}(k))^{N-i} = \beta$ (i.e., $\underline{\alpha}(k)$ is the threshold value that clips the left tail of the Beta distribution with probability β) and, similarly, let $\overline{\alpha}(k)$ be the threshold value that clips the right tail with probability β, i.e., $1 - \sum_{i=0}^{k-1} \binom{N}{i} \overline{\alpha}(k)^i (1-\overline{\alpha}(k))^{N-i} = 1 - \beta$. In order for Eq. (6) to be true for this problem it is necessary that

$$\underline{\epsilon}(k) \leq \underline{\alpha}(k) \;\text{ and }\; \overline{\alpha}(k) \leq \overline{\epsilon}(k), \tag{7}$$

for, otherwise, at least one side on the inequality $\underline{\epsilon}(k) \leq V(x_N^*) \leq \overline{\epsilon}(k)$ would be violated with a probability that exceeds β. Figure 6 profiles $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ against $\underline{\alpha}(k)$ and $\overline{\alpha}(k)$. The fact that the curves are close to each other shows that the price that is paid to make the result distribution-free is minor.

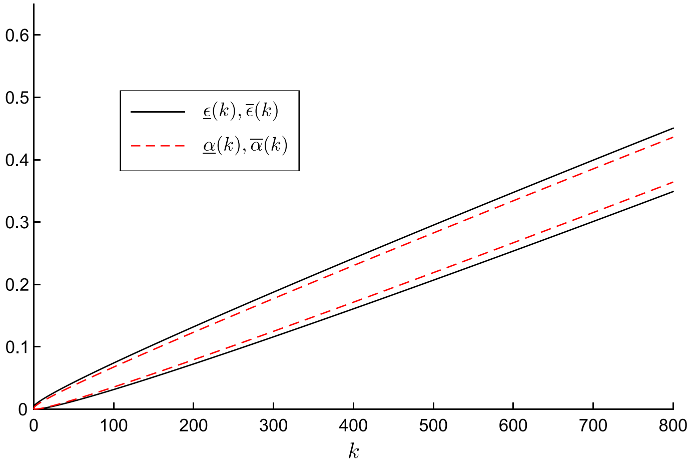

Fig. 6 Profile of curves $\underline{\epsilon}(k)$, $\overline{\epsilon}(k)$ and $\underline{\alpha}(k)$, $\overline{\alpha}(k)$, $N = 2000$, $d = 800$, $\beta = 0.001$

3.2 Computational aspects

The discussion in the previous section suggests an easy way to compute function $\underline{t}(\cdot)$ and $\overline{t}(\cdot)$ in Theorem 1. The two relations in (7) give respectively $\overline{t}(k) \geq 1 - \underline{\alpha}(k)$ and $\underline{t}(k) \leq 1 - \overline{\alpha}(k)$. Hence, the two solutions of Eq. (4) must lie in the two bold intervals in Fig. 7. To determine $\underline{t}(k)$, a bisection procedure can be run starting from the extreme points 0 and $1 - \overline{\alpha}(k)$, while, to determine $\overline{t}(k)$, one first checks if polynomial (4) has the same sign in $1 - \underline{\alpha}(k)$ and 1 (in which case one comes to know that $\overline{t}(k) > 1$ so that $\underline{\epsilon}(k)$ in Theorem 1 has value 0) and, when the signs in $1 - \underline{\alpha}(k)$ and 1 are different, a bisection procedure with initial extreme points $1 - \underline{\alpha}(k)$ and 1 is run to find $\overline{t}(k)$. A self-contained MATLAB code implementing these two bisection procedures is provided in "Appendix A".

4 An example

In this section, we describe the context in which Figs. 3 and 4 in the introduction have been generated and illustrate various aspects relating to the theory of the previous section.

1000 points p_i are independently sampled in $\mathbb{R}^{400}$ according to a probability density $\mathbb{P}$ and presented to us. We want to translate the negative orthant in $\mathbb{R}^{400}$ (i.e., the domain where all components are negative or zero) so that the translated orthant contains all the given points while the translation shift is minimized. This amounts to solving the scenario optimization problem

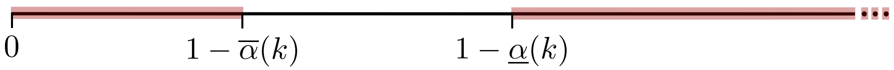

Fig. 7 Intervals to which $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ belong

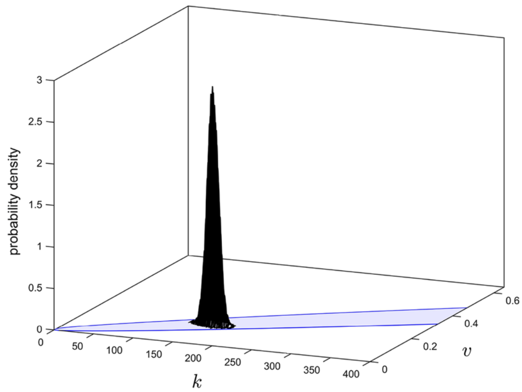

Fig. 8 Distribution of $(s^*_N, V(x^*_N))$ for the first probability density, $N = 1000$, $d = 400$. The shaded region is the 99.9% region given by Theorem 1

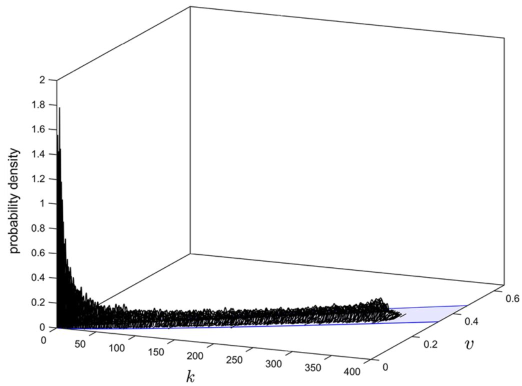

Fig. 9 Distribution of $(s^*_N, V(x^*_N))$ for the second probability density, $N = 1000$, $d = 400$. The shaded region is the 99.9% region given by Theorem 1

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^{400}} \quad \sum_{j=1}^{400} x_j \\
&\text{subject to :} \quad x_j \geq p_{i,j}, \quad i = 1, \ldots, N,
\end{aligned}
\tag{8}
$$

where j denotes component.

Optimization problem (8) was solved for two different probability densities $\mathbb{P}$. In the first case, the points were given by relation $p_i = q_i + c_i$ where the q_i's were independently drawn from a 400-dimensional Gaussian distribution with zero mean and unit variance, $G(0, I)$, and the c_i's were vectors with 400 equal components taken from $[0, 5]$ with uniform distribution. In the second case, the p_i's were again given by $p_i = q_i + c_i$ with the q_i's generated as in the previous case but the c_i's were this time vectors of equal components with value 0 with probability 99% and a value taken from a Gaussian distribution $G(0, 2^2)$ with probability 1%. The optimization problem (8) was solved 100000 times for both probability densities and each time the values of s^*_{1000} and $V(x^*_{1000})$ were recorded. This gave the empirical distributions reported in Figs. 8 and 9,[7] where the region given by Theorem 1 for $\beta = 0.001$ is also displayed.

[7] These are the same empirical distributions as in Figs. 3 and 4 (in Figs. 3 and 4 only some values of k were displayed).

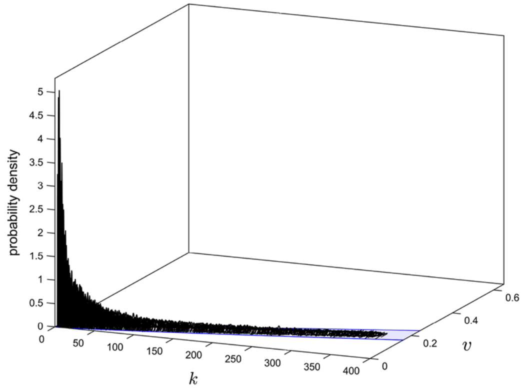

Fig. 10 Distribution of $(s_N^*, V(x_N^*))$ for the second probability density, $N = 2000$, $d = 400$. The shaded region is the 99.9% region given by Theorem 1

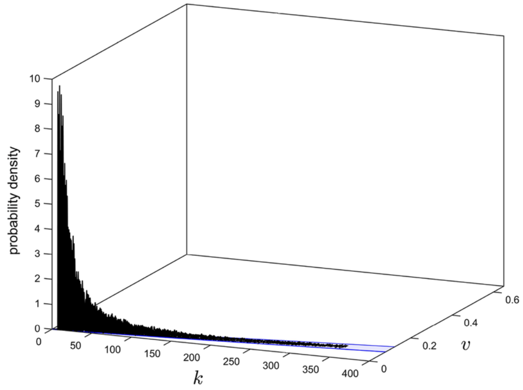

Fig. 11 Distribution of $(s_N^*, V(x_N^*))$ for the second probability density, $N = 4000$, $d = 400$. The shaded region is the 99.9% region given by Theorem 1

For the second probability density, the simulation set-up as above was then repeated for $N = 2000$ and $N = 4000$, while all other quantities were kept the same as for $N = 1000$. The results are on display in Figs. 10 and 11. One can notice the evolution of the empirical distribution, which, for any given k, tends to move towards lower values of the risk while the random spreads also progressively reduce.

5 A theory for scenario-based decision problems

The theory developed in previous sections can be carried over to a more abstract set-up which incorporates only the salient features that play a role in the derivation of the results. Such an abstract theory is presented here, followed by two application examples: in Sect. 5.1, the upper bound $V(x_N^*) \leq \overline{\epsilon}(s_N^*)$ is studied under a more general condition than the non-degeneracy Assumption 2, while Sect. 5.2 deals with the possibility of relaxing the constraints so as to improve the value of the cost function.

In analogy with Sect. 3, for any given $m = 0, 1, 2, \dots$ start by considering an independent sample $\delta_i, i = 1, \dots, m$, from $(\Delta, \mathcal{F}, \mathbb{P})$. The optimization domain $\mathcal{X}$ of Sect. 3 is here substituted by a generic set $\mathcal{Z}$, called the "decision set".[8] To any δ there is associated a set $\mathcal{Z}_\delta \subseteq \mathcal{Z}$. In Sect. 3, x_m^* was generated by optimization problem (3), that is, optimization problem (3) there defined maps (one for any m) from Δ^m to $\mathcal{X}$. Here, we consider generic maps

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

and write $z_m^* = M_m(\delta_1, \dots, \delta_m)$. The interpretation is that z_m^* is a decision made according to a rule M_m applied to a set of scenarios $\delta_1, \delta_2, \dots, \delta_m$. Throughout this Sect. 5 the word "decision" refers to z_m^*, and the symbol "z" will only be used in relation to a decision.

The following assumption applies to M_m.

Assumption 3 (*properties on M_m*)

(i) M_m is permutation-invariant: $M_m(\delta_1, \dots, \delta_m) = M_m(\delta_{i_1}, \dots, \delta_{i_m})$ if $\delta_{i_1}, \dots, \delta_{i_m}$ is a permutation of $\delta_1, \dots, \delta_m$;

(ii) given m values δ_i, $i = 1, \dots, m$, augment them with n more values $\delta_{m+1}, \dots, \delta_{m+n}$, where m and n are generic non-negative integers. If the decision z_m^* obtained from the first m values of δ is in the sets associated with the extra n values of δ, that is, $z_m^* \in \mathcal{Z}_{\delta_i}$ for $i = m+1, \dots, m+n$, then it holds that $M_{m+n}(\delta_1, \dots, \delta_{m+n}) = z_m^*$, that is, the decision obtained after adding the new δ's remains unchanged.

(iii) instead, if $\delta_i, i = 1, \dots, m$, is augmented with values $\delta_{m+1}, \dots, \delta_{m+n}$ such that one or more $\delta_i, i = m+1, \dots, m+n$, has associated a set which does not contain z_m^*, that is, $z_m^* \notin \mathcal{Z}_{\delta_i}$ for one or more $i = m+1, \dots, m+n$, then it holds that $M_{m+n}(\delta_1, \dots, \delta_{m+n}) \neq z_m^*$. □

Notice that Assumption 3 does not require that z_m^* is in $\mathcal{Z}_{\delta_i}, i = 1, \dots, m$.

The notion of risk of a $z \in \mathcal{Z}$ is an obvious extension from Sect. 3: $V(z) = \mathbb{P}\{\delta \in \Delta : z \notin \mathcal{Z}_\delta\}$. Also the notion of support constraint carries over from Sect. 3, but we here prefer to speak of "support element" since, as we have remarked above, z_m^* is not forced to be in $\mathcal{Z}_{\delta_i}, i = 1, \dots, m$, that is the $\mathcal{Z}_{\delta_i}$'s do not act here as constraints: $\mathcal{Z}_{\delta_i}$ is called a *support element* if $M_{m-1}(\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_m) \neq M_m(\delta_1, \dots, \delta_m)$. The number of support elements is denoted by $\tilde{s}_m^*$. Finally, the assumption of non-degeneracy remains essentially unchanged.

Assumption 4 (*non-degeneracy of M_m*)

For every m, the decision $z_m^* = M_m(\delta_1, \dots, \delta_m)$ coincides with probability 1 (with respect to the sample $\delta_i, i = 1, \dots, m$) with the decision that is obtained after eliminating all the elements that are not of support. □

[8] $\mathcal{Z}$ can be any set, without any Euclidean structure. We change notation from $\mathcal{X}$ to $\mathcal{Z}$ because in some applications $\mathcal{Z}$ is the same as $\mathcal{X}$ augmented with extra elements; concrete examples of decision sets are provided in Sects. 5.1 and 5.2.

The fact that the set-up of this section encompasses as a particular case that of Sect. 3 is shown in Sect. 6.2.

We are now ready to state the main theorem of this section, Theorem 2. The essence of this theorem is that the thesis of Theorem 1 carries over to the present more abstract set-up. One difference with Theorem 1 is that in Theorem 2 the number of support elements is not a-priori upper bounded (as it was the number of support constraints in Theorem 1, which could not exceed d), so that we here have to also account for the case $k = N$ (number of support elements equal to the number of scenarios), which leads to considering Eq. (10).

Theorem 2 *Given a confidence parameter $\beta \in (0, 1)$, for any $k = 0, 1, \ldots, N-1$ consider the polynomial equation in the t variable*

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0, \tag{9}$$

and, for $k = N$, consider the polynomial equation

$$1 - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-N} = 0. \tag{10}$$

For any $k = 0, 1, \ldots, N-1$, Eq. (9) has exactly two solutions in $[0, +\infty)$, which we denote with $\underline{t}(k)$ and $\overline{t}(k)$ ($\underline{t}(k) \leq \overline{t}(k)$). Instead, Eq. (10) has only one solution in $[0, +\infty)$, which we denote with $\overline{t}(N)$, while we define $\underline{t}(N) = 0$. Let $\underline{\epsilon}(k) := \max\{0, 1 - \overline{t}(k)\}$ and $\overline{\epsilon}(k) := 1 - \underline{t}(k)$, $k = 0, 1, \ldots, N$. Under Assumptions 3 and 4, for any Δ and $\mathbb{P}$ it holds that

$$\mathbb{P}^N \{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} \geq 1 - \beta. \tag{11}$$

□

Proof The proof is given in Sect. 6. □

Theorem 2 allows one to evaluate the risk for decision problems of various type. We next apply Theorem 2 to two specific setups.

5.1 Application no. 1: a theorem for problem (1) with a relaxed non-degeneracy condition

In Sect. 3, Theorem 1 was stated under the non-degeneracy Assumption 2. We here show that a statement on the generalization properties of (1) can be obtained under a condition that only refers to the instance at hand of (1), as opposed to Assumption 2, which pertains to all choices of scenario samples.

Suppose that in the instance at hand of (1) the set of active constraints coincides with the set of support constraints (which implies that the instance at hand of (1) is non-degenerate). Note that this condition is easily verifiable based on the available scenarios. The next theorem, which is a corollary of Theorem 2, provides an evaluation of the risk when this condition is satisfied.

Theorem 3 *Consider the optimization problem (1) with $N > d$. Given a confidence parameter $\beta \in (0, 1)$, for any $k = 0, 1, \dots, d$ consider the polynomial equation in the t variable*

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0. \tag{12}$$

This equation has exactly two solutions in $[0, +\infty)$, which we denote with $\underline{t}(k)$ and $\overline{t}(k)$ ($\underline{t}(k) \le \overline{t}(k)$). Let $\overline{\epsilon}(k) := 1 - \underline{t}(k)$. Under Assumptions 1, for any Δ and $\mathbb{P}$ it holds that

$$\mathbb{P}^N \big(\{ V(x_N^*) > \overline{\epsilon}(s_N^*) \} \cap C \big) \le \beta, \tag{13}$$

where C is the event where the set of active constraints coincides with the set of support constraints. □

Proof The proof is given in Sect. 6. □

5.2 Application no. 2: optimization with constraints relaxation

In this section, we consider scenario optimization problems where, unlike (1), one is allowed to violate constraints for the purpose of improving the cost value. We assume that constraint violation has itself a cost and in the limit when this cost goes to infinity the original problem (1) is recovered.

Matters of convenience suggest that constraints are written in this section as $f(x, \delta) \le 0$, where $f(x, \delta)$ is a convex function in x for any given δ (referring back to the notation in Sect. 1, we therefore have that $\mathcal{X}_\delta = \{x : f(x, \delta) \le 0\}$). The reason for this choice is that the value of the function f is used to express the “regret” for violating a constraint: for a given δ, the regret at x is $f(x, \delta)$ and the steepness of this function describes the marginal increase of regret when the solution is moved in a given direction. In this set-up, we consider the following scenario optimization problem, which is a penalty-based variant of model (1):

$$\begin{aligned} \min_{\substack{x \in \mathcal{X} \\ \xi_i \ge 0, i=1,\dots,N}} \quad & c^T x + \rho \sum_{i=1}^{N} \xi_i \\ \text{subject to :} \quad & f(x, \delta_i) \le \xi_i, \quad i = 1, \dots, N, \end{aligned} \tag{14}$$

where, as before, δ_i, $i = 1, \dots, N$, is an independent random sample from $(\Delta, \mathcal{F}, \mathbb{P})$. Note that (14) has $d + N$ optimization variables, namely, x and ξ_i, $i = 1, \dots, N$. If $\xi_i > 0$, the constraint $f(x, \delta_i) \le 0$ is relaxed to $f(x, \delta_i) \le \xi_i$ and this generates the regret ξ_i. Parameter ρ is used to set a suitable trade-off between the original cost function and the cost generated by the regret for violating constraints. When $\rho \to \infty$, one goes back to the original optimization problem (1) where no constraint violation is allowed. It turns out that the cost $c^T x_N^*$ achieved at the optimum of (14) (existence and uniqueness of the solution follows from Assumption 5 below) is an increasing function of ρ so that taking a lower value of ρ improves the cost. This fact, which is somehow intuitive as decreasing ρ reduces the penalty for constraint violation, is proven in the following (to highlight the dependence on ρ, the solution to (14) is written as $x_N^*(\rho)$, $\xi_{i,N}^*(\rho)$, $i = 1, \dots, N$, in this derivation). Let $\rho_1 > \rho_2 > 0$. Optimality of $x_N^*(\rho_1)$, $\xi_{i,N}^*(\rho_1)$, $i = 1, \dots, N$, for (14) with $\rho = \rho_1$ yields

$$c^T x_N^*(\rho_1) + \rho_1 \sum_{i=1}^N \xi_{i,N}^*(\rho_1) \le c^T x_N^*(\rho_2) + \rho_1 \sum_{i=1}^N \xi_{i,N}^*(\rho_2),$$

from which one has

$$\frac{1}{\rho_1}\big(c^T x_N^*(\rho_1) - c^T x_N^*(\rho_2)\big) \le \sum_{i=1}^N \xi_{i,N}^*(\rho_2) - \sum_{i=1}^N \xi_{i,N}^*(\rho_1). \tag{15}$$

Likewise,

$$c^T x_N^*(\rho_2) + \rho_2 \sum_{i=1}^N \xi_{i,N}^*(\rho_2) \le c^T x_N^*(\rho_1) + \rho_2 \sum_{i=1}^N \xi_{i,N}^*(\rho_1)$$

gives

$$\sum_{i=1}^N \xi_{i,N}^*(\rho_2) - \sum_{i=1}^N \xi_{i,N}^*(\rho_1) \le \frac{1}{\rho_2}\big(c^T x_N^*(\rho_1) - c^T x_N^*(\rho_2)\big). \tag{16}$$

From (15) and (16) one obtains

$$c^T x_N^*(\rho_1) - c^T x_N^*(\rho_2) \le \frac{\rho_1}{\rho_2}\big(c^T x_N^*(\rho_1) - c^T x_N^*(\rho_2)\big).$$

Since $\frac{\rho_1}{\rho_2} > 1$, one therefore has that $c^T x_N^*(\rho_1) - c^T x_N^*(\rho_2) \ge 0$, which gives the sought result.

The following assumption is the equivalent of the existence and uniqueness Assumption 1 for the generalized set-up of this section.

Assumption 5 (*existence and uniqueness*) Consider optimization problems as in (14) where N is substituted with an index $m = 0, 1, \dots$ and δ_i, $i = 1, \dots, m$, is an independent sample from $(\Delta, \mathcal{F}, \mathbb{P})$. For every m and for every sample $(\delta_1, \delta_2, \dots, \delta_m)$, these optimization problems admit a solution (i.e., the problems are feasible and the infimum is achieved on the feasibility set). If for one of these optimization problems more than one solution exists, one solution is singled out by the application of a convex tie-break rule, which breaks the tie by minimizing an additional convex function $t_1(x)$, and, possibly, other convex functions $t_2(x)$, $t_3(x)$, …if the tie still occurs.[9] □

Moreover, we make the following assumption.

Assumption 6 For every x, $\mathbb{P}\{\delta : f(x,\delta) = 0\} = 0$. □

This is a non-accumulation assumption on functions $f(x,\delta)$, and, when constraints $f(x,\delta_i) \leq 0$ are enforced in the scenario optimization problem as is done in (1), it implies the non-degeneracy Assumption 2.

In this context, we have the following theorem.

Theorem 4 *Consider the optimization problem (14). Given a confidence parameter $\beta \in (0,1)$, for any $k = 0, 1, \dots, N-1$ consider the polynomial equation in the t variable*

$$\binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-k} = 0, \tag{17}$$

and for $k = N$ consider the polynomial equation

$$1 - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{k} t^{i-N} = 0. \tag{18}$$

For any $k = 0, 1, \dots, N-1$, Eq. (17) has exactly two solutions in $[0, +\infty)$, which we denote with $\underline{t}(k)$ and $\overline{t}(k)$ ($\underline{t}(k) \leq \overline{t}(k)$). Instead, Eq. (18) has only one solution in $[0, +\infty)$, which we denote with $\overline{t}(N)$, while we define $\underline{t}(N) = 0$. Let $\underline{\epsilon}(k) := \max\{0, 1 - \overline{t}(k)\}$ and $\overline{\epsilon}(k) := 1 - \underline{t}(k)$, $k = 0, 1, \dots, N$. Under Assumptions 5 and 6, for any Δ and $\mathbb{P}$ it holds that

$$\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(x_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} \geq 1 - \beta, \tag{19}$$

where $V(x) = \mathbb{P}\{\delta \in \Delta : \; f(x,\delta) > 0\}$, x_N^* is the value of x at the optimum of (14), and $\tilde{s}_N^*$ is the number of δ_i's for which $f(x_N^*, \delta_i) \geq 0$. □

Proof The proof is given in Sect. 6. □

Theorem 4 provides a quantitative evaluation of the risk in the context of optimization with constraint relaxation. By a comparison with Theorem 1, we see that the number of support constraints s_N^* of the scenario optimization problem (1) is here substituted by $\tilde{s}_N^*$. $\tilde{s}_N^*$ accounts for the constraints of (1) that are violated, i.e., $f(x,\delta_i) > 0$, plus those that are active, i.e., $f(x,\delta_i) = 0$, at $x = x_N^*$. One can prove that $f(x_N^*, \delta_i) = 0$ in at most d cases with probability 1, showing that functions $\underline{\epsilon}(\cdot)$

[9] Note that only the tie with respect to x is broken by $t_1(x)$, $t_2(x)$, $t_3(x)$, …. On the other hand, for a given x_m^* the values of ξ_i, $i = 1, \dots, m$, remain unambiguously determined at optimum by relation $\xi_{i,m}^* = f(x_m^*, \delta_i)$, so that no tie on ξ_i, $i = 1, \dots, m$, can persist after the tie on x is broken.

and $\overline{\epsilon}(\cdot)$ have to be evaluated at an integer equal to the number of violated constraints plus at most an excess of d in order to compute the lower and upper bounds for the risk.[10]

As we have previously seen, decreasing ρ in (14) improves the cost $c^T x_N^*$, while, along the process, the number of violated constraints typically increases.[11] As ρ varies, the cost is computed, while Theorem 4 allows one to evaluate the risk, so providing the user with the information needed to achieve a sensible trade-off between cost and risk.[12] The next example illustrates this idea.

Example 1 A manufacturer produces goods in d different workplaces and x_j, $j = 1, \ldots, d$, is the quantity planned to be produced in workplace j. For the production, n different resources are employed. The quantity of resource k, $k = 1, \ldots, n$, used in workplace j to produce a unitary amount of goods is subject to random fluctuation and is denoted by $q_{j,k}(\delta)$. Each resource is available in a limited amount a_k. The goal of the manufacturer is to maximize the production while keeping low the probability of being in need of resources that exceed the available amount.

Assuming that a record $\{q_{j,k}(\delta_i),\ j = 1, \ldots, d,\ k = 1, \ldots, n\}$, $i = 1, \ldots, N$, of values for $\{q_{j,k}(\delta),\ j = 1, \ldots, d,\ k = 1, \ldots, n\}$ is available, the problem is modeled according to the scenario approach as follows

$$
\begin{aligned}
&\min_{x_j \geq 0,\ j=1,\ldots,d} \quad -\sum_{j=1}^{d} x_j \\
&\text{subject to :} \quad \begin{cases} \sum_{j=1}^{d} q_{j,1}(\delta_i) x_j \leq a_1, \\ \vdots \\ \sum_{j=1}^{d} q_{j,n}(\delta_i) x_j \leq a_n, \end{cases} \qquad i = 1, \ldots, N.
\end{aligned} \tag{20}
$$

A simulation was performed with $d = 50$, $n = 2$ and $N = 2000$, which gave the result $-\sum_{j=1}^{50} x_{j,2000}^* = -16.66$, $s_{2000}^* = 4$.[13] With the choice $\beta = 10^{-6}$, an application of

[10] Intuitively, the proportion of violated constraints (empirical risk) is not a valid indicator of the true risk $V(x_N^*)$ since optimization generates a bias towards larger risks by drifting the solution against the constraints. The excess with respect to the number of violated constraints that appears in the computation of $\tilde{s}_N^*$ captures this mechanism and offers one the possibility to obtain tight evaluations of the risk, as quantified by the lower and upper bounds in Eq. (19), independently of the problem under consideration.

[11] The increase is not always monotone.

[12] Other methods have been proposed in the literature to trade the risk for an improved cost. One method consists in allowing the solution to violate a preset proportion of the empirical constraints (chance-constrained problem over the empirical distribution). In the context of scenario optimization, this approach is described in [18], where practically useful, but untight, bounds on the risk are also derived. More generally, the problem of relating the empirical risk to chance-constrained feasibility is dealt with in many papers including [6,7,44,57]. The problem of finding a solution that violates a preset proportion of the empirical constraints is a non-convex problem that is difficult to solve in general. The formulation in (14) is convex and this eases the problem of finding a solution. Interestingly, as already noted, this formulation is amenable to tight evaluations of the risk.

[13] For reproducibility, we inform the reader about the mechanism by which $q_{j,k}(\delta)$ were generated. Let $\delta = (\alpha_1, \alpha_2, \gamma_{1,1}, \ldots, \gamma_{50,1}, \gamma_{1,2}, \ldots, \gamma_{50,2})$, where, for $k = 1, 2$ and $j = 1, \ldots, 50$, $\alpha_k \sim \mathcal{U}[10, 50]$ (i.e., α_k is uniformly distributed in [10, 50]), $\gamma_{j,k} \sim \mathcal{U}[-2.5, 2.5]$, and all these variables are independent

Theorem 1 provided the following interval for the risk that the available amounts of resources are exceeded: $0 \leq V(x_N^*) \leq 0.014$.

Further, the manufacturer decides to increase the production and towards this goal accepts some rise in the risk of running out of resources. To design the new production strategy, the manufacturer uses the constraints relaxation approach presented in this section and solves the optimization problem

$$\begin{aligned}
&\min_{\substack{x_j \geq 0, j=1,\ldots,50 \\ \xi_i \geq 0, i=1,\ldots,2000}} \quad -\sum_{j=1}^{50} x_j + \rho \sum_{i=1}^{N} \xi_i \\
&\text{subject to :} \quad \left(\max \left\{ \sum_{j=1}^{50} q_{j,1}(\delta_i) x_j - a_1, \sum_{j=1}^{50} q_{j,2}(\delta_i) x_j - a_2, 0 \right\} \right)^2 \leq \xi_i \\
&\qquad\qquad i = 1, \ldots, 2000.
\end{aligned} \tag{21}$$

Notice that this is the same as (14) with[14]

$$f(x, \delta) = \left(\max \left\{ \sum_{j=1}^{50} q_{j,1}(\delta) x_j - a_1, \sum_{j=1}^{50} q_{j,2}(\delta) x_j - a_2, 0 \right\} \right)^2. \tag{22}$$

As for the value of ρ, its selection can be tricky, because how ρ impacts on production/risk can be difficult to forecast, and we here refer to an approach that can be of general utility in other applications as well. The manufacturer sets out to solve (21) for an array of values of ρ. For each value, the production increase is calculated from the solution, while Theorem 4 gives an interval for the corresponding risk. Selecting $\beta = 10^{-6}$, the intervals that were found for 22 distinct values of ρ are displayed in Fig. 12 (in the figure, red crosses depict $\underline{\epsilon}(\tilde{s}_{2000}^*)$ and $\overline{\epsilon}(\tilde{s}_{2000}^*)$, where the numerical values are given on the right vertical axis).[15] In the same figure, the plot of cost $= -\sum_{j=1}^{50} x_{j,2000}^*$ is also profiled (blue dots, whose numerical values are given on the

one of the others. Then, $q_{j,k}(\delta) = (\alpha_k^{\frac{1}{4}} + \gamma_{j,k})^{-1}$, $j = 1, \ldots, 50$, $k = 1, 2$. Moreover, in the simulation, we took $a_1 = a_2 = 1$.

14 Notice that, strictly speaking, this choice of $f(x, \delta)$ does not satisfy Assumption 6. Reason is that setting to zero $f(x, \delta)$ when $\sum_{j=1}^{50} q_{j,1}(\delta) x_j - a_1$ and $\sum_{j=1}^{50} q_{j,2}(\delta) x_j - a_2$ are negative, as is done in (22), generates regions with positive volume in the domain in $\mathbb{R}^{50}$ for x where $f(x, \delta) = 0$. However, an easy inspection of the derivation of Theorem 4 shows that the requirement of Assumption 6 that, for every x, $\mathbb{P}\{\delta : f(x, \delta) = 0\} = 0$ can be relaxed to requiring that, for every x, $\mathbb{P}\{\delta : x \text{ is on the boundary of the constraint } \{f(x, \delta) \leq 0\}\} = 0$, and the theory goes through unaltered with the only modifications that, throughout, "$f(x, \delta) = 0$" becomes "x is on the boundary of the constraint $\{f(x, \delta) \leq 0\}$", "$f(x, \delta) < 0$" becomes "x is in the interior of the constraint $\{f(x, \delta) \leq 0\}$", and "$f(x, \delta) \geq 0$" becomes "x violates or is on the boundary of the constraint $\{f(x, \delta) \leq 0\}$". While we have preferred in the general presentation the simpler formulation of Assumption 6, this second formulation leads to zero volume regions in the domain in $\mathbb{R}^{50}$ for x in the present example.

15 Since the intervals in Fig. 12 are obtained by a repeated application of Theorem 4, the confidence that $\underline{\epsilon}(\tilde{s}_{2000}^*) \leq V(x_{2000}^*) \leq \overline{\epsilon}(\tilde{s}_{2000}^*)$ for all the 22 values of ρ simultaneously is $1 - 22 \cdot \beta$.

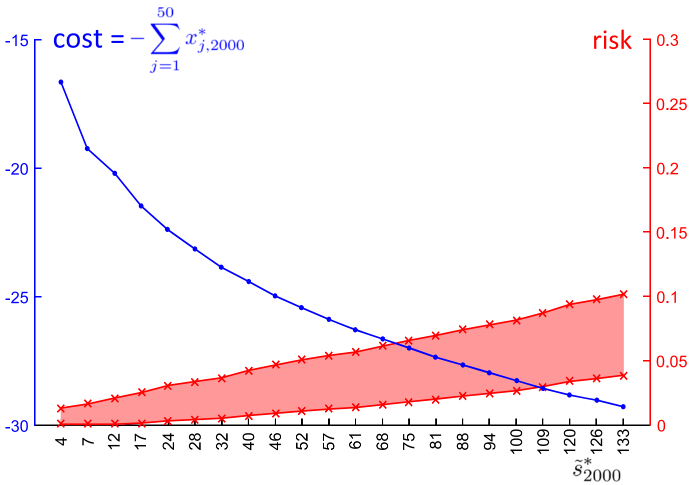

Fig. 12 Cost $= -\sum_{j=1}^{50} x^*_{j,2000}$ vs. risk. In abscissa is the number $\tilde{s}^*_{2000}$ of scenarios for which $f(x^*_{2000}, \delta_i) \geq 0$

left vertical axis). A suitable trade-off between production and risk can be obtained by a direct inspection of the figure. In this case, the value of ρ that gives $\tilde{s}^*_{2000} = 46$ was selected, resulting in a 50% increase in production, while the estimated interval for the risk moved from [0, 0.014] to [0.009, 0.047]. $\square$

6 Proofs

The proof of Theorem 2 is given first since Theorem 1 easily follows as a corollary of Theorem 2.

6.1 Proof of Theorem 2

We start by showing that Eq. (9) has two solutions in $[0, +\infty)$ and that Eq. (10) has one solution in $[0, +\infty)$.

Denote by $\varphi_k(t)$, $k = 0, 1, \ldots, N-1$, the polynomials in the left-hand side of (9) and by $\varphi_N(t)$ the polynomial in the left-hand side of (10), which we rewrite here by making explicit the number $H = 3N$ of terms in the rightmost summations:[16]

$$\varphi_k(t) = \begin{cases} \binom{N}{k} t^{N-k} - \frac{\beta}{2N} \sum_{i=k}^{N-1} \binom{i}{k} t^{i-k} - \frac{\beta}{2H} \sum_{i=N+1}^{N+H} \binom{i}{k} t^{i-k}, & 0 \leq k < N \\ 1 - \frac{\beta}{2H} \sum_{i=N+1}^{N+H} \binom{i}{N} t^{i-N}, & k = N. \end{cases} \tag{23}$$

[16] The reason for introducing H is that the theorem will be proven in a slightly more general form where H is any integer ≥ 1 and not just $3N$. The choice $H = 3N$ gives satisfactory evaluations in most cases, and this is why Theorem 2 was stated with $H = 3N$. However, the extra generality allowed by other values of H can turn out to be useful to tighten the bounds $\underline{\epsilon}(\cdot)$ and $\overline{\epsilon}(\cdot)$ in some cases when N is not too large. This issue is not further discussed in this paper.

Let us start with $\varphi_N(t)$. By construction, $\varphi_N(0) = 1$ and $\varphi_N(t)$ is strictly decreasing with $\varphi_N(t) \to -\infty$ as $t \to +\infty$. Hence, $\varphi_N(t)$ has a unique root $\bar{t}(N)$ in $[0, +\infty)$, and, moreover,

$$\varphi_N(t) > 0 \text{ for } t \in [0, \bar{t}(N)), \text{ while } \varphi_N(t) < 0 \text{ for } t \in (\bar{t}(N), +\infty). \tag{24}$$

Turn now to $\varphi_k(t)$, $k = 0, 1, \dots, N-1$. Notice first that the following recursive equation holds

$$\varphi_k(t) = -\frac{\beta}{2N} + (k+1)\int_0^t \varphi_{k+1}(\tau)\,\mathrm{d}\tau, \quad k = 0, 1, \dots, N-1, \tag{25}$$

as can be verified by a direct calculation. Using (25), we want to show that all the $\varphi_k(t)$, $k = 0, 1, \dots, N-1$, follow the same pattern: a. $\varphi_k(0) = -\frac{\beta}{2N} < 0$; b. with the sole exception of $\varphi_{N-1}(t)$ that is increasing in $t = 0_+$, $\varphi_k(t)$ is initially decreasing, then it is increasing with $\varphi_k(t) > 0$ at its maximum, and then decreasing again; c. $\varphi_k(t) \to -\infty$ as $t \to +\infty$. Facts a,b,c are obtained by using (25) for $k = N-1$ (and recalling the properties of $\varphi_N(t)$), and then proceeding backward, $k = N-2, N-3, \dots, 0$, where the only point that deserves an explanation is that $\varphi_k(t) > 0$ at its maximum. To show this, notice that from (23) we have

$$\varphi_0(1) = 1 - \frac{\beta}{2N}\sum_{i=0}^{N-1} 1 - \frac{\beta}{2H}\sum_{i=N+1}^{N+H} 1 = 1 - \beta > 0,$$

and, looking again at (25), $\varphi_0(1) > 0$ would not be possible if it were that $\varphi_k(t) \le 0$ $\forall t$ for some k. From a,b,c, it follows that each $\varphi_k(t)$, $k = 0, 1, \dots, N-1$, has exactly two roots, $\underline{t}(k)$ and $\bar{t}(k)$, in $[0, +\infty)$. Moreover,

$$\varphi_k(t) < 0 \text{ for } t \in [0, \underline{t}(k)) \cup (\bar{t}(k), +\infty), \text{ while } \varphi_k(t) > 0 \text{ for } t \in (\underline{t}(k), \bar{t}(k)). \tag{26}$$

We next prove relation (11). For all positive integers $k = 0, 1, \dots$ define

$$F_k(v) = \mathbb{P}^k\{V(z_k^*) \le v \ \wedge \ \tilde{s}_k^* = k\}, \tag{27}$$

where $z_k^* = M_k(\delta_1, \dots, \delta_k)$, $\tilde{s}_k^* =$ number of support elements, and $\wedge$ is the "and" operator. In words, $F_k(v)$ is the probability that, with a sample of k instances of δ, all of them are of support and the decision has risk no more than v. The F_k's are generalized distribution functions, [59]. Functions $F_0, F_1, \dots$ are different from one problem to another and, as we shall show, for a given problem the left-hand side of (11) can be computed from $F_0, F_1, \dots$.

Start by noting that the events $\{\tilde{s}_N^* = k_1\}$ and $\{\tilde{s}_N^* = k_2\}$ are not overlapping for $k_1 \neq k_2$, so that

$$\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} = \sum_{k=0}^{N} \mathbb{P}^N\{\underline{\epsilon}(k) \leq V(z_N^*) \leq \overline{\epsilon}(k) \ \wedge \ \tilde{s}_N^* = k\}. \tag{28}$$

Focus on one event $S_k := \{\delta_1, \ldots, \delta_N : \ \underline{\epsilon}(k) \leq V(z_N^*) \leq \overline{\epsilon}(k) \ \wedge \ \tilde{s}_N^* = k\} \subseteq \Delta^N$ and, for each sample $(\delta_1, \ldots, \delta_N) \in S_k$, evaluate the indexes of the δ_i's that correspond to the support elements. Group together all the samples with the same indexes. In this way, S_k is partitioned in $\binom{N}{k}$ subsets. All these subsets have the same probability because $\delta_1, \ldots, \delta_N$ are independent and identically distributed. Hence,

$$\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} = \binom{N}{k} \mathbb{P}^N\{A\}, \tag{29}$$

where A is one of these subsets, say the one where the indexes are $1, 2, \ldots, k$, that is,

$$A := \{\underline{\epsilon}(k) \leq V(z_N^*) \leq \overline{\epsilon}(k) \ \wedge \ \tilde{s}_N^* = k \ \wedge \ \delta_1, \ldots, \delta_k \text{ are of support}\}.$$

We show below that set A is equal to

$$B := \{\underline{\epsilon}(k) \leq V(z_k^*) \leq \overline{\epsilon}(k) \ \wedge \ \tilde{s}_k^* = k \ \wedge \ z_k^* \in \mathcal{Z}_{\delta_i}, \ i = k+1, \ldots, N\}$$

up to a zero probability set.

We first show that $A \subseteq B$ up to a zero probability set. Since in A the support elements are the first k, by the non-degeneracy Assumption 4, $z_N^* = z_k^*$ up to a zero probability set. Thus, $\underline{\epsilon}(k) \leq V(z_N^*) \leq \overline{\epsilon}(k)$ implies $\underline{\epsilon}(k) \leq V(z_k^*) \leq \overline{\epsilon}(k)$ up to a zero probability set. Moreover, $z_k^* \in \mathcal{Z}_{\delta_i}$, $i = k+1, \ldots, N$, because if $z_k^* \notin \mathcal{Z}_{\delta_{\bar{i}}}$ for some $\bar{i} \in \{k+1, \ldots, N\}$, then $z_N^* \neq z_k^*$ by property (iii) in Assumption 3. We finally show that all the $\delta_1, \ldots, \delta_k$ are of support for $M_k(\delta_1, \ldots, \delta_k)$, which gives $\tilde{s}_k^* = k$. Indeed, if one among $\delta_1, \ldots, \delta_k$, say δ_1, were not of support, then $M_{k-1}(\delta_2, \ldots, \delta_k) = z_k^*$. But then, by adding $\delta_{k+1}, \ldots, \delta_N$ – which correspond to $\mathcal{Z}_{\delta_i}$ such that $z_k^* \in \mathcal{Z}_{\delta_i}$ – one would obtain $M_{N-1}(\delta_2, \ldots, \delta_N) = z_k^* = z_N^*$ by property (ii) in Assumption 3, and, hence, δ_1 would not be of support for $M_N(\delta_1, \ldots, \delta_N)$, which is a contradiction.

Next we show that $B \subseteq A$ up to a zero probability set. Since in B it holds that $z_k^* \in \mathcal{Z}_{\delta_i}$, $i = k+1, \ldots, N$, by property (ii) in Assumption 3 we obtain that $z_N^* = z_k^*$ and, thus, relation $\underline{\epsilon}(k) \leq V(z_k^*) \leq \overline{\epsilon}(k)$ implies that $\underline{\epsilon}(k) \leq V(z_N^*) \leq \overline{\epsilon}(k)$. We next show that $\tilde{s}_N^* = k \wedge \delta_1, \ldots, \delta_k$ are of support, which is equivalent to say that $\delta_1, \ldots, \delta_k$ are the only support scenarios for $M_N(\delta_1, \ldots, \delta_N)$, up to a zero probability set. First, none of the $\delta_{k+1}, \ldots, \delta_N$ can be of support for $M_N(\delta_1, \ldots, \delta_N)$. Indeed, remove one of these scenarios, say δ_N, from the sample $\delta_1, \ldots, \delta_N$. Since $\delta_1, \ldots, \delta_{N-1}$ is $\delta_1, \ldots, \delta_k$ with the addition of δ_i, $i = k+1, \ldots, N-1$, for which $z_k^* \in \mathcal{Z}_{\delta_i}$, by property (ii) in Assumption 3 one obtains $M_{N-1}(\delta_1, \ldots, \delta_{N-1}) = z_k^* = z_N^*$. This shows that δ_N is not of support. To next show that $\delta_1, \ldots, \delta_k$ are of support for $M_N(\delta_1, \ldots, \delta_N)$ up to a zero probability set, proceed by contradiction, and assume that at least one of the first k scenarios is not of support with non-zero probability. Since $\delta_{k+1}, \ldots, \delta_N$ are not of support, then the support scenarios for $M_N(\delta_1, \ldots, \delta_N)$ would be with non-zero probability a strict subset of $\delta_1, \ldots, \delta_k$. However, since properties (ii)–(iii) in Assumption 3 imply that the decision obtained from a strict subset of $\delta_1, \ldots, \delta_k$ must be different from z_k^* (indeed, to be the same, by property (iii), the $\mathcal{Z}_{\delta_i}$ corresponding to the missing scenarios must include z_k^*; but, then, by property (ii), adding all the missing scenarios except one the solution would still be z_k^*, contradicting the assumption that all the $\delta_1, \ldots, \delta_k$ are of support for $M_k(\delta_1, \ldots, \delta_k)$), then we would have that the decision obtained from the support scenarios of $M_N(\delta_1, \ldots, \delta_N)$ would be different from z_k^*, and, hence, different from $z_N^* = M_N(\delta_1, \ldots, \delta_N)$, with non-zero probability. This, however, contradicts the non-degeneracy Assumption 4.

We next show that

$$\mathbb{P}^N\{B\} = \int_{[\underline{\epsilon}(k), \overline{\epsilon}(k)]} (1-v)^{N-k} \mathrm{d}F_k(v). \tag{30}$$

Indeed, owing to the independence of $\delta_1, \ldots, \delta_N$, $(1-v)^{N-k}$ is the conditional probability that $z_k^* \in \mathcal{Z}_{\delta_i}$, $i = k+1, \ldots, N$, given that $V(z_k^*) = v$ and $\tilde{s}_k^* = k$. Then, recalling the definition of F_k in (27), (30) follows from [59, Chapter II, Section 7, Equation (17)].

Since $\mathbb{P}^N\{A\} = \mathbb{P}^N\{B\}$ (which follows from the fact that $A = B$ up to a zero probability set), substituting (30) in (29) and further (29) in (28) yields

$$\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} = \sum_{k=0}^{N} \binom{N}{k} \int_{[\underline{\epsilon}(k), \overline{\epsilon}(k)]} (1-v)^{N-k} \mathrm{d}F_k(v). \tag{31}$$

Equation (31) provides the fundamental formula by which $\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\}$ can be computed from $F_0, F_1, \ldots$. To proceed in the evaluation of the right-hand side of (31), we have now to characterize the distributions $F_0, F_1, \ldots$, which is done in the following.

If the same argument used to derive (31) is repeated with z_m^*, $m = 0, 1, \ldots$, in place of z_N^* and with 0 in place of $\underline{\epsilon}(\tilde{s}_N^*)$ and and 1 in place of $\overline{\epsilon}(\tilde{s}_N^*)$, relation

$$\mathbb{P}^N\{0 \leq V(z_m^*) \leq 1\} = \sum_{k=0}^{m} \binom{m}{k} \int_{[0,1]} (1-v)^{N-k} \mathrm{d}F_k(v)$$

is found. Since $V(z_m^*)$ takes value in $[0, 1]$, the right-hand side of this equality is clearly equal to 1, which shows that $F_0, F_1, \ldots$ must satisfy the following generalized moment conditions:

$$\sum_{k=0}^{m} \binom{m}{k} \int_{[0,1]} (1-v)^{m-k} \mathrm{d}F_k(v) = 1, \quad m = 0, 1, \ldots. \tag{32}$$

Equation (32) provides a characterization of $F_0, F_1, \dots$ under which the right-hand side of (31) can be evaluated, and the optimization problem (34) below returns a lower bound to $\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\}$. That is,

$$\mathbb{P}^N\{\underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} \geq \gamma, \tag{33}$$

with ($\mathcal{C}$ denotes the positive cone of generalized distribution functions)

$$\begin{aligned}
\gamma = \inf_{F_k,\ k=0,1,\dots} \quad & \sum_{k=0}^{N} \binom{N}{k} \int_{[\underline{\epsilon}(k),\overline{\epsilon}(k)]} (1-v)^{N-k} \mathrm{d}F_k(v) \\
\text{subject to:} \quad & \sum_{k=0}^{m} \binom{m}{k} \int_{[0,1]} (1-v)^{m-k} \mathrm{d}F_k(v) = 1, \quad m = 0, 1, \dots \\
& F_k \in \mathcal{C}, k = 0, 1, \dots .
\end{aligned} \tag{34}$$

The last part of the proof consists in showing that $\gamma \geq 1-\beta$, from which (11) follows.

Consider the following truncated version of (34):

$$\begin{aligned}
\gamma_H = \inf_{F_0,F_1,\dots,F_{N+H}} \quad & \sum_{k=0}^{N} \binom{N}{k} \int_{[\underline{\epsilon}(k),\overline{\epsilon}(k)]} (1-v)^{N-k} \mathrm{d}F_k(v) \\
\text{subject to:} \quad & \sum_{k=0}^{m} \binom{m}{k} \int_{[0,1]} (1-v)^{m-k} \mathrm{d}F_k(v) = 1, \quad m = 0, 1, \dots, N+H, \\
& F_0, F_1, \dots, F_{N+H} \in \mathcal{C}.
\end{aligned} \tag{35}$$

Since in (34) and (35) the cost function only depends on $F_0, F_1, \dots, F_N$ and (35) is less constrained than (34), we have that

$$\gamma \geq \gamma_H. \tag{36}$$

The dual of (35) is

$$\begin{aligned}
\gamma_H^* = \sup_{\lambda_0,\lambda_1,\dots,\lambda_{N+H}} \quad & \sum_{m=0}^{N+H} \lambda_m \\
\text{subject to:} \quad & \sum_{m=k}^{N+H} \lambda_m \binom{m}{k} (1-v)^{m-k} \\
& \leq \begin{cases} \binom{N}{k}(1-v)^{N-k} \cdot \mathbf{1}_{[\underline{\epsilon}(k),\overline{\epsilon}(k)]}(v), & k = 0, 1, \dots, N \\ 0, & k = N+1, \dots, N+H, \end{cases} \\
& v \in [0, 1],
\end{aligned} \tag{37}$$

where $\mathbf{1}$ denotes the indicator function. The following derivation, provided for self-containedness, shows that $\gamma_H \geq \gamma_H^*$ (weak duality): for every $F_0, F_1, \dots, F_{N+H}$

feasible for (35) and $\lambda_0, \lambda_1, \ldots, \lambda_{N+H}$ feasible for (37) it holds that

$$
\begin{aligned}
&\sum_{k=0}^{N} \binom{N}{k} \int_{[\underline{\epsilon}(k),\overline{\epsilon}(k)]} (1-v)^{N-k} \mathrm{d}F_k(v) \\
&= \sum_{k=0}^{N} \int_{[0,1]} \binom{N}{k} (1-v)^{N-k} \cdot \mathbf{1}_{[\underline{\epsilon}(k),\overline{\epsilon}(k)]}(v) \mathrm{d}F_k(v) \\
&\geq \sum_{k=0}^{N+H} \int_{[0,1]} \sum_{m=k}^{N+H} \lambda_m \binom{m}{k} (1-v)^{m-k} \mathrm{d}F_k(v) \\
&= \sum_{m=0}^{N+H} \lambda_m \sum_{k=0}^{m} \binom{m}{k} \int_{[0,1]} (1-v)^{m-k} \mathrm{d}F_k(v) \\
&= \sum_{m=0}^{N+H} \lambda_m,
\end{aligned}
$$

so that taking the inf on the left-hand side and the sup on the right-hand side yields

$$
\gamma_H \geq \gamma_H^*. \tag{38}
$$

Inequalities (33), (36), and (38) give

$$
\mathbb{P}^N \left\{ \underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*) \right\} \geq \gamma \geq \gamma_H \geq \gamma_H^*, \tag{39}
$$

from which the theorem can be proven by showing that $\gamma_H^* \geq 1-\beta$.

To show that $\gamma_H^* \geq 1-\beta$, we perform the substitution $t := 1-v$ and rewrite (37) as

$$
\begin{aligned}
\gamma_H^* = \sup_{\lambda_0,\lambda_1,\ldots,\lambda_{N+H}} \quad & \sum_{m=0}^{N+H} \lambda_m \\
\text{subject to:} \quad & \sum_{m=k}^{N+H} \lambda_m \binom{m}{k} t^{m-k} \\
& \quad \leq \begin{cases} \binom{N}{k} t^{N-k} \cdot \mathbf{1}_{[1-\overline{\epsilon}(k),1-\underline{\epsilon}(k)]}(t), & k = 0, 1, \ldots, N \\ 0, & k = N+1, \ldots, N+H, \end{cases} \\
& \quad t \in [0, 1].
\end{aligned} \tag{40}
$$

Consider now

$$
\begin{aligned}
&\lambda_m = -\frac{\beta}{2N}, \quad m = 0, 1, \ldots, N-1, \\
&\lambda_m = 1, \quad m = N, \\
&\lambda_m = -\frac{\beta}{2H}, \quad m = N+1, N+2, \ldots, N+H.
\end{aligned} \tag{41}
$$

We want to show that the selection of λ_m's in (41) is feasible for (40), which gives the sought inequality because for these λ_m's we have that $\sum_{m=0}^{N+H} \lambda_m = 1 - \beta$, and $\gamma_H^* \geq 1 - \beta$ because γ_H^* is the sup over all the feasible choices of the λ_m's.

Consider first the constraints in (40) for $k > N$. These are trivially satisfied since for the λ_m's in (41) the left-hand side of the inequality is negative for $t \in [0, 1]$. When instead $k \leq N$, for the λ_m's in (41), the left-hand side of the inequality in (40) coincides with the polynomial $\varphi_k(t)$ in (23). Since $\varphi_k(t) \leq \binom{N}{k} t^{N-k}, k = 0, 1, \ldots, N$, and since $\varphi_k(t) \leq 0$ for t outside the interval $[1 - \overline{\epsilon}(k), 1 - \underline{\epsilon}(k)]$ (see (26), (24), and the definitions of $\overline{\epsilon}(k)$ and $\underline{\epsilon}(k)$ in Theorem 2), also the constraints in (40) for $k \leq N$ are satisfied.

Wrapping up, the selection of λ_m's in (41) is feasible for (40) and, hence, $\gamma_H^* \geq 1-\beta$, which used in (39) gives

$$\mathbb{P}^N \left\{ \underline{\epsilon}(\tilde{s}_N^*) \leq V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*) \right\} \geq 1 - \beta.$$

This concludes the proof. □

6.2 Proof of Theorem 1

The optimization problems (3) complemented with a convex tie-break rule as specified in Assumption 1 define a family of maps M_m from the sample $\delta_1, \ldots, \delta_m$ to the solution x_m^*. We show that this family of maps satisfies the assumptions of Theorem 2 so that Theorem 1 follows from Theorem 2 with the positions $z_m^* = x_m^*$, $\tilde{s}_m^* = s_m^*$ and $\mathcal{Z}_\delta = \mathcal{X}_\delta$, and by noting that Eq. (10) in Theorem 2 can be dropped in the context of Theorem 1 since $N > d$ implies $\tilde{s}_N^* < N$.

Consider Assumption 3. First of all note that the maps M_m are permutation invariant (point (i) in Assumption 3) because the solution to (3) clearly does not depend on the order in which constraints are sampled. Consider now (ii) in Assumption 3. In (3) the solution is selected as the feasible point that achieves the smallest cost $c^T x$ and, if a tie occurs, the tie is broken by minimizing the convex functions $t_1(x), t_2(x), \ldots$. By adding extra constraints that are satisfied at x_m^*, the feasibility domain shrinks while x_m^* remains feasible. Hence, x_m^* remains the optimal solution and (ii) follows. Referring to (iii), if some δ_i's are added such that at least one δ_i corresponds to a constraint that is not satisfied by x_m^*, then the solution x_m^* has to change and move to a feasible point, and this gives (iii). Finally, notice that the non-degeneracy Assumption 4 is guaranteed by Assumption 2. □

6.3 Proof of Theorem 3

Given a sample δ_i, $i = 1, \ldots, m$, the scenario optimization problem (3) defines a solution x_m^*, possibly after a tie-break rule is applied as indicated in Assumption 1. Here, we introduce a decision z_m^* generated by (3) which consists of x_m^* augmented with the values of the δ_i's that correspond to active constraints of (3), where each of these δ_i's is equipped with an integer number that indicates how many times the same value of δ_i has been seen (that is, if δ_1 is active, δ_3 takes on the same value as δ_1, and no other δ_i takes on the same value, then the value of δ_1 is included in the decision, followed by the number 2). Hence, formally,

$$z_m^* = \Big(x_m^*, \big\{(\delta; n) \text{ where } \delta = \delta_i \text{ for some } i \in \{1, \dots, m\} \text{ such that } \mathcal{X}_{\delta_i} \text{ is active for problem (3), and } n = \text{number of times with which the same value } \delta \text{ appears in (3)}\big\}\Big),$$

and $\mathcal{Z}$ is the set of all $z = (x, \{(\delta^{(j)}; n^{(j)}),\ \ j = 1, \dots, p\})$ for some integer $p \geq 0$, where $x \in \mathcal{X}$, $\delta^{(j)} \in \Delta$, $n^{(j)} \in \mathbb{N}$, $j = 1, \dots, p$. Given a δ, let $\mathcal{Z}_\delta = \{z \in \mathcal{Z} : x \in \mathcal{X}_\delta$ and $\mathcal{X}_\delta$ is non-active at $x\}$. Hence, $V(z)$ is defined as the probability of the set of δ's such that either $x \notin \mathcal{X}_\delta$ or $\mathcal{X}_\delta$ is active at x. Clearly, $V(x_m^*) \leq V(z_m^*)$ as the latter also includes active constraints. An easy inspection shows that Assumption 3 holds with these definitions. Moreover, the support elements are here those associated to the δ_i's such that $\mathcal{X}_{\delta_i}$ is active at x_m^* and it follows immediately that the non-degeneracy Assumption 4 also holds. Hence, Theorem 2 can be applied to this context to obtain a result on $V(z_N^*)$. From this result we next show that the statement of Theorem 3 is obtained. First, we drop in the event of Eq. (11) the left inequality and write $\mathbb{P}^N\{V(z_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} \geq 1 - \beta$; then, since $V(x_m^*) \leq V(z_m^*)$, we further obtain $\mathbb{P}^N\{V(x_N^*) \leq \overline{\epsilon}(\tilde{s}_N^*)\} \geq 1 - \beta$, or, equivalently, $\mathbb{P}^N\{V(x_N^*) > \overline{\epsilon}(\tilde{s}_N^*)\} \leq \beta$, from which

$$\mathbb{P}^N\big(\{V(x_N^*) > \overline{\epsilon}(\tilde{s}_N^*)\} \text{ and condition } C \text{ holds}\big) \leq \beta.$$

On the other hand, condition C implies that $\tilde{s}_N^*$ in Theorem 2 (number of active constraints) coincides with s_N^* in Theorem 1 (number of support constraints). This gives (13). Moreover, if $N > d$, then we certainly have $\tilde{s}_N^* = s_N^* \leq d$, so that the polynomial equation (10) in Theorem 2 can be dropped as it never happens that $\tilde{s}_N^* = N$. This concludes the proof. □

6.4 Proof of Theorem 4

Let $x_m^*, \xi_{i,m}^*, i = 1, \dots, m$, be the solution of (14) with m in place of N. The abstract theory of Sect. 5 does not apply directly to $x_m^*, \xi_{i,m}^*, i = 1, \dots, m$, and we have first to define the concept of decision z_m^*. This is x_m^* augmented with the values of the δ_i's that correspond to constraints $f(x, \delta_i) \leq 0$ that are violated at the solution (i.e., $f(x_m^*, \delta_i) > 0$), where each of these δ_i's is equipped with an integer number that indicates how many times the same value of δ_i has been seen. Formally,

$$z_m^* = \Big(x_m^*, \big\{(\delta; n) \text{ where } \delta = \delta_i \text{ for an } i \in \{1, \dots, m\} \text{ such that } \xi_{i,m}^* > 0, \text{ and } n = \text{number of times with which the same value } \delta \text{ appears in (14) with } m \text{ in place of } N\big\}\Big),$$

and $\mathcal{Z}$ is the set of all $z = (x, \{(\delta^{(j)}; n^{(j)}),\ \ j = 1, \dots, p\})$ for some integer $p \geq 0$, where $x \in \mathcal{X}$, $\delta^{(j)} \in \Delta$, $n^{(j)} \in \mathbb{N}$, $j = 1, \dots, p$. We take as $\mathcal{Z}_\delta$ the set of $z \in \mathcal{Z}$ for which $f(x, \delta) \leq 0$ and $(\delta^{(j)}; n^{(j)})$, $j = 1, \dots, p$, are arbitrary both in number (i.e., p is any positive integer) and value. Correspondingly, $V(z) = V(x, \{(\delta^{(j)}; n^{(j)}), j = 1, \dots, p\})$ is defined as the probability of the set of δ's such that $x \notin \mathcal{X}_\delta$. Assumptions 3 and 4 are satisfied in this context, as we next show. Condition (i) in Assumption 3 is clearly true. To show (ii), argue as follows. If $z_m^* \in \mathcal{Z}_{\delta_i}$ for $i = m+1, \dots, m+n$, then $f(x_m^*, \delta_i) \leq 0$ for $i = m+1, \dots, m+n$. Hence, augmenting the solution of (14) with m in place of N with $\xi_i = 0$ for $i = m+1, \dots, m+n$ gives a feasible point x_m^*, $\xi_i = \xi_{i,m}^*$, $i = 1, \dots, m$, $\xi_i = 0$, $i = m+1, \dots, m+n$ for (14) with $m+n$ in place of N that attains the same value as the optimal value of (14) with m in place of N. It is claimed that this is the optimal solution of (14) with $m+n$ in place of N. Indeed, if a better solution $\bar{x}, \bar{\xi}_i$, $i = 1, \dots, m+n$ existed for (14) with $m+n$ in place of N, then $\bar{x}, \bar{\xi}_i$, $i = 1, \dots, m$ would be superoptimal for (14) with m in place of N since the dropped $\bar{\xi}_i$, $i = m+1, \dots, m+n$ give a non-negative contribution. To the optimal solution $x_{m+n}^* = x_m^*, \xi_{i,m+n}^* = \xi_{i,m}^*, i = 1, \dots, m, \xi_{i,m+n}^* = 0, i = m+1, \dots, m+n$ of (14) with $m+n$ in place of N there corresponds $z_{m+n}^* = (x_{m+n}^*, \{(\delta; n)$, where $\delta = \delta_i$ for an $i \in \{1, \dots, m+n\}$ such that $\xi_{i,m+n}^* > 0$, and $n =$ number of times with which the same value δ appears in (14) with $m+n$ in place of $N\})$. This z_{m+n}^* equals $(x_m^*, \{(\delta; n)$, where $\delta = \delta_i$ for an $i \in \{1, \dots, m\}$ such that $\xi_{i,m}^* > 0$, and $n =$ number of times with which the same value δ appears in (14) with m in place of $N\})$, which is the same as z_m^*, showing the validity of (ii). Condition (iii) in Assumption 3 instead easily follows from the fact that if $z_m^* \notin \mathcal{Z}_{\delta_{\bar{i}}}$ for some $\bar{i} \in \{m+1, \dots, m+n\}$, then either x_{m+n}^* has to move to a new location where $f(x, \delta_{\bar{i}}) \leq 0$ (and therefore $z_{m+n}^* \neq z_m^*$) or $\delta_{\bar{i}}$ has to be added to the solution z_m^* to obtain z_{m+n}^* (and, again $z_{m+n}^* \neq z_m^*$). Turn now to assess the non-degeneracy Assumption 4. Consider optimization problem (14) with m in place of N and eliminate a δ_i such that $f(x_m^*, \delta_i) < 0$. The decision associated with the remaining $m-1$ δ_i's is the same decision as that associated with the original optimization problem (14) with m in place of N because the constraint corresponding to the eliminated δ_i is non-active. Hence, none of the δ_i such that $f(x_m^*, \delta_i) < 0$ is of support. We further claim that all δ_i's such that $f(x_m^*, \delta_i) \geq 0$ are of support with probability 1. Eliminate any one of them from the optimization problem (14) with m in place of N. If the eliminated one is such that $f(x_m^*, \delta_i) > 0$, then the decision clearly changes, so that the δ_i is of support. Suppose instead that the eliminated one is such that $f(x_m^*, \delta_i) = 0$ and, for the sake of contradiction, suppose also that the decision does not change. It follows that x_m^* is obtained by an optimization problem that does not contain δ_i and, due to the independence of $\delta_1, \delta_2, \dots, \delta_m$, it is easily seen that Assumption 6 implies that $f(x_m^*, \delta_i) = 0$ only happens with probability 0. Hence, with probability 1 the δ_i's such that $f(x_m^*, \delta_i) \geq 0$ are of support, and they give the original decision since the simultaneous elimination of the other δ_i's for which $f(x_m^*, \delta_i) < 0$ (non-active) does not change the decision. This means that the problem is non-degenerate.

Since all conditions of the abstract theory are satisfied, Theorem 2 applies. In the present context, $V(z_N^*) = V(x_N^*)$ and $\tilde{s}_N^* =$ number of the δ_i's for which $f(x_N^*, \delta_i) \geq 0$, which concludes the proof of the theorem. □

7 Conclusions

In recent years, the stochastic optimization community has witnessed a surge of interest for methods that are based on knowledge obtained from data. This is a timely trend in the current world where data analytics is gaining importance owing to the availability of increasing amounts of data and the growth of computational power for their treatment. Various approaches are making their way as emerging technologies in this context. One is data-driven Distributionally Robust Optimization (DRO) where the distribution of uncertainty is estimated from the data. This approach is quite flexible because availing of a distributional description of uncertainty allows one to apply various methods and risk measures, while it has been also shown that the computational complexity attendant to DRO methods can be maintained to levels of tractability under various conditions. The disadvantage of DRO is that it deals with uncertainty explicitly, which can be prohibitive in applications involving complex systems. In contrast, the scenario approach deals with data directly by considering the impact they have on the optimization problem, regardless of the underlying mechanisms that govern the generation of uncertainty. Scenario algorithms include optimizing over the set of feasibility of all observations, as well as relaxed schemes where an increase of risk is traded for an improvement of the performance. The core achievement of this paper is showing that there exists a profound, and quite general, link between two concepts: *risk*, which is important to the user to judge the danger of constraint violation, and *complexity*, a quantity that can be measured from observations. Exploiting this link furnishes fundamental tools to evaluate the risk of scenario solutions, so complementing the heuristic use of data in decision-making with a solid theory that enables one to certify the dependability of the decision.

A MATLAB code

The following MATLAB code returns $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ for user assigned k, N, and β.

```
function [epsL, epsU] = epsLU(k,N,bet)

 alphaL = betaincinv(bet,k,N-k+1);
 alphaU = 1-betaincinv(bet,N-k+1,k);
 m1 = [k:1:N];
 aux1 = sum(triu(log(ones(N-k+1,1)*m1),1),2);
 aux2 = sum(triu(log(ones(N-k+1,1)*(m1-k)),1),2);
 coeffs1 = aux2-aux1;
 m2 = [N+1:1:4*N];
 aux3 = sum(tril(log(ones(3*N,1)*m2)),2);
 aux4 = sum(tril(log(ones(3*N,1)*(m2-k))),2);
 coeffs2 = aux3-aux4;

 t1 = 1-alphaL;
 t2 = 1;
```

```
  poly1 = 1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1 - (N-m1')*log(t1)))...
    -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t1)));
  poly2 = 1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1 - (N-m1')*log(t2)))...
    -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t2)));
  if ((poly1*poly2) > 0)
    epsL = 0;
  else
    while t2-t1 > 1e-10
      t = (t1+t2)/2;
      polyt = 1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1 - (N-m1')*log(t)))...
        -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t)));
      if polyt > 0
        t1=t;
      else
        t2=t;
      end
    end
    epsL = 1-t2;
  end

  t1 = 0;
  t2 = 1-alphaU;
  poly1 = 1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1 - (N-m1')*log(t1)))...
    -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t1)));
  poly2 = 1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1 - (N-m1')*log(t2)))...
    -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t2)));
  if ((poly1*poly2) > 0)
    epsL = 0;
  else
    while t2-t1 > 1e-10
      t = (t1+t2)/2;
      polyt =1+bet/(2*N)-bet/(2*N)*sum(exp(coeffs1-(N-m1')*log(t)))...
        -bet/(6*N)*sum(exp(coeffs2 + (m2'-N)*log(t)));
      if polyt > 0
        t2=t;
      else
        t1=t;
      end
    end
    epsU = 1-t1;
  end
end
```

References

1. Alamo, T., Tempo, R., Camacho, E.: A randomized strategy for probabilistic solutions of uncertain feasibility and optimization problems. IEEE Trans. Autom. Control **54**(11), 2545–2559 (2009)
2. Baronio, F., Baronio, M., Campi, M., Caré, A., Garatti, S.: Ventricular defribillation: classification with GEM and a roadmap for future investigations. In: Proceedings of the 56th IEEE Conference on Decision and Control. Melbourne, Australia (2017)
3. Bayraksan, G., Morton, D.: Assessing solution quality in stochastic programs. Math. Program. **108**, 495–514 (2006)
4. Bayraksan, G., Morton, D.: Assessing solution quality in stochastic programs via sampling. In: Oskoorouchi, M. (ed.) Tutorials in Operations Research, pp. 102–122. Informs (2009)
5. Ben-Tal, A., Nemirovski, A.: On safe tractable approximations of chance-constrained linear matrix inequalities. Math. Oper. Res. **34**(1), 1–25 (2009)
6. Bertsimas, D., Gupta, V., Kallus, N.: Data-driven robust optimization. Math. Program. **167**, 235–292 (2018)
7. Bertsimas, D., Gupta, V., Kallus, N.: Robust sample average approximation. Math. Program. **171**, 217–282 (2018)
8. Bertsimas, D., Thiele, A.: Robust and data-driven optimization: modern decision-making under uncertainty. In: Tutorials on Operations Research. INFORMS (2006)
9. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
10. Calafiore, G., Campi, M.: Uncertain convex programs: randomized solutions and confidence levels. Math. Program. **102**(1), 25–46 (2005). https://doi.org/10.1007/s10107-003-0499-y
11. Calafiore, G., Campi, M.: The scenario approach to robust control design. IEEE Trans. Autom. Control **51**(5), 742–753 (2006)
12. Calafiore, G., El Ghaoui, L.: On distributionally robust chance-constrained linear programs. Math. Program. **130**(1), 1–22 (2006)
13. Carè, A., Ramponi, F.A., Campi, M.C.: A new classification algorithm with guaranteed sensitivity and specificity for medical applications. IEEE Control Syst. Lett. **2**, 393–398 (2018)
14. Campi, M.: Classification with guaranteed probability of error. Mach. Learn. **80**, 63–84 (2010)
15. Campi, M., Calafiore, G., Garatti, S.: Interval predictor models: identification and reliability. Automatica **45**(2), 382–392 (2009). https://doi.org/10.1016/j.automatica.2008.09.004
16. Campi, M., Carè, A.: Random convex programs with l_1-regularization: sparsity and generalization. SIAM J. Control Optim. **51**(5), 3532–3557 (2013)
17. Campi, M., Garatti, S.: The exact feasibility of randomized solutions of uncertain convex programs. SIAM J. Optim. **19**(3), 1211–1230 (2008). https://doi.org/10.1137/07069821X
18. Campi, M., Garatti, S.: A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. J. Optim. Theory Appl. **148**(2), 257–280 (2011)
19. Campi, M., Garatti, S.: Wait-and-judge scenario optimization. Math. Program. **167**(1), 155–189 (2018)
20. Campi, M., Garatti, S., Prandini, M.: The scenario approach for systems and control design. Annu. Rev. Control **33**(2), 149–157 (2009). https://doi.org/10.1016/j.arcontrol.2009.07.001
21. Carè, A., Garatti, S., Campi, M.: FAST—fast algorithm for the scenario technique. Oper. Res. **62**(3), 662–671 (2014)
22. Carè, A., Garatti, S., Campi, M.: Scenario min-max optimization and the risk of empirical costs. SIAM J. Optim. **25**(4), 2061–2080 (2015)
23. Crespo, L., Giesy, D., Kenny, S.: Interval predictor models with a formal characterization of uncertainty and reliability. In: Proceedings of the 53rd IEEE Conference on Decision and Control (CDC), pp. 5991–5996. Los Angeles, CA, USA (2014)
24. Crespo, L., Kenny, S., Giesy, D.: Random predictor models for rigorous uncertainty quantification. Int. J. Uncertain. Quantif. **5**(5), 469–489 (2015)
25. Crespo, L., Kenny, S., Giesy, D., Norman, R., Blattnig, S.: Application of interval predictor models to space radiation shielding. In: Proceedings of the 18th AIAA Non-Deterministic Approaches Conference. San Diego, CA, USA (2016)
26. de Mello, T.H.: Variable-sample methods for stochastic optimization. ACM Trans. Model. Comput. Simul. **13**, 108–133 (2003)
27. de Mello, T.H., Bayraksan, G.: Monte Carlo sampling-based methods for stochastic optimization. Surv. Oper. Res. Manag. Sci. **19**(1), 56–85 (2014)

28. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 596–612 (2010)
29. Dentcheva, D.: Optimization models with probabilistic constraints. In: Calafiore, G., Dabbene, F. (eds.) Probabilistic and Randomized Methods for Design Under Uncertainty. Springer, London (2006)
30. Erdogan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. Math. Program. **107**, 37–61 (2006)
31. Esfahani, P., Sutter, T., Lygeros, J.: Performance bounds for the scenario approach and an extension to a class of non-convex programs. IEEE Trans. Autom. Control **60**(1), 46–58 (2015)
32. Esfahani, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. Math. Program. (2017). https://doi.org/10.1007/s10107-017-1172-1
33. Fabozzi, F., Kolm, P., Pachamanova, D., Focardi, S.: Robust Portfolio Optimization and Management. Wiley, Hoboken (2010)
34. Garatti, S., Campi, M.: Modulating robustness in control design: principles and algorithms. IEEE Control Syst. Mag. **33**(2), 36–51 (2013). https://doi.org/10.1109/MCS.2012.2234964
35. Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Oper. Res. **58**, 902–917 (2010)
36. Grammatico, S., Zhang, X., Margellos, K., Goulart, P., Lygeros, J.: A scenario approach for non-convex control design. IEEE Trans. Autom. Control **61**(2), 334–345 (2016)
37. Gupta, V.: Near-optimal ambiguity sets for distributionally robust optimization. Manag. Sci. **65**(9), 4242–4260 (2019)
38. Hanasusanto, G., Roitch, V., Kuhn, D., Wiesemann, W.: A distributionally robust perspective on uncertainty quantification and chance constrained programming. Math. Program. **151**(1), 35–62 (2015)
39. Hong, L., Hu, Z., Liu, G.: Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. ACM Trans. Model. Comput. Simul. **24**(4), 22:1–22:37 (2014)
40. Hu, Z., Hong, L.: Kullback–Leiber divergence constrained distributionally robust optimization (2013). http://www.optimization-online.org/DB_HTML/2012/11/3677.html
41. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Math. Program. **158**(1–2), 291–327 (2016)
42. Lam, H.: Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. Oper. Res. **67**(4), 1090–1105 (2019)
43. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. **142**, 215–241 (2006)
44. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**, 674–699 (2008). https://doi.org/10.1137/070702928
45. Luedtke, J., Ahmed, S., Nemhauser, G.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **122**(2), 247–272 (2010)
46. Mak, W., Morton, D., Wood, R.: Monte Carlo bounding techniques for determing solution quality in stochastic programs. Oper. Res. Lett. **24**, 47–56 (1999)
47. Margellos, K., Prandini, M., Lygeros, J.: On the connection between compression learning and scenario based single-stage and cascading optimization problems. IEEE Trans. Autom. Control **60**(10), 2716–2721 (2015)
48. Nemirovski, A.: On safe tractable approximations of chance constraints. Eur. J. Oper. Res. **219**, 707–718 (2012)
49. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**(4), 969–996 (2006). https://doi.org/10.1137/050622328
50. Nemirovski, A., Shapiro, A.: Scenario approximations of chance constraints. In: Calafiore, G., Dabbene, F. (eds.) Probabilistic and Randomized Methods for Design Under Uncertainty. Springer, London (2006)
51. Pagnoncelli, B., Ahmed, S., Shapiro, A.: Sample average approximation method for chance constrained programming: theory and applications. J. Optim. Theory Appl. **142**(2), 399–416 (2009)
52. Pagnoncelli, B., Reich, D., Campi, M.: Risk-return trade-off with the scenario approach in practice: a case study in portfolio selection. J. Optim. Theory Appl. **155**(2), 707–722 (2012)
53. Pagnoncelli, B., Vanduffel, S.: A provisioning problem with stochastic payments. Eur. J. Oper. Res. **221**(2), 445–453 (2012)
54. Pflug, G., Wozabal, D.: Ambiguity in portfolio selection. Quant. Finance **7**, 435–442 (2007)

55. Schildbach, G., Fagiano, L., Frei, C., Morari, M.: The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations. Automatica **50**(12), 3009–3018 (2014)
56. Schildbach, G., Fagiano, L., Morari, M.: Randomized solutions to convex programs with multiple chance constraints. SIAM J. Optim. **23**(4), 2479–2501 (2013)
57. Shapiro, A.: Monte–Carlo sampling methods. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, volume 10 of Handbooks in Operations Research and Management Science. Elsevier, London (2003)
58. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. MPS-SIAM, Philadelphia (2009)
59. Shiryaev, A.: Probability. Springer, New York (1996)
60. Thiele, A.: Robust stochastic programming with uncertain probabilities. IMA J. Manag. Math. **19**(3), 289–321 (2008). https://doi.org/10.1093/imaman/dpm011
61. Van Parys, B., Esfahani, P., Kuhn, D.: From data to decisions: distributionally robust optimization is optimal. (2017). arxiv:1704.04118
62. Vayanos, P., Kuhn, D., Rustem, B.: A constraint sampling approach for multistage robust optimization. Automatica **48**(3), 459–471 (2012)
63. Welsh, J., Kong, H.: Robust experiment design through randomisation with chance constraints. In: Proceedings of the 18th IFAC World Congress, Milan, Italy (2011)
64. Welsh, J., Rojas, C.: A scenario based approach to robust experiment design. In: Proceedings of the 15th IFAC Symposium on System Identification. Saint-Malo, France (2009)
65. Wieseman, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**, 1358–1376 (2014)
66. Wozabal, D.: A framework for optimization under ambiguity. Ann. Oper. Res. **193**, 21–47 (2012)
67. Zhang, X., Grammatico, S., Schildbach, G., Goulart, P., Lygeros, J.: On the sample size of random convex programs with structured dependence on the uncertainty. Automatica **60**, 182–188 (2015)
68. Zhou, Z., Cogill, R.: Reliable approximations of probability-constrained stochastic linear-quadratic control. Automatica **49**(8), 2435–2439 (2013)
69. Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. Math. Program. **137**(1–2), 167–198 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.