2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Average number of mistakes in sequential risk-averse scenario decision-making

S. Garatti, *Member, IEEE*, A. Carè, *Member, IEEE*, and Marco C. Campi, *Fellow, IEEE*

***Abstract*— Data-driven methods aim to design predictors and controllers that adapt to the environment by utilizing information sourced from data. Due to their reliance on a finite amount of data, these designs are inevitably subject to a degree of imprecision, which can result in mistakes when they are applied to new cases. In this contribution, we introduce a sequential decision scheme in which the user is provided at each step with both a design and an assessment of the associated risk of making mistakes. The user decides whether to apply the design based on a threshold on the acceptable risk level. Novel results are presented to evaluate the average number of mistakes in this sequential data-driven risk-averse decision making framework. This requires in-depth analyses because, as we will see, naive evaluations based on common sense may lead to misleading results. Many are the potential applications, including the optimization of control actions over shifting windows (as in MPC), investments with recourse, and sequential prediction approaches.**

I. Introduction and problem statement

The scenario approach is a data-driven paradigm for control, design and decision-making at large in the presence of uncertainty, [1]. Motivated by robust control techniques, the scenario approach was originally developed in a robust convex optimization framework, [2]–[4], but through the years has come to encompass much more general decision schemes, [5]–[10], with applications to machine learning, economics and game theory. See [11] for a recent survey. The robust convex optimization setup, however, remains relevant in many applications and the sake of concreteness suggests us to refer to this context in the present paper. In this framework, the decision process is abstracted as follows. Given a family of convex constraint sets $\mathcal{X}_\delta \subseteq \mathbb{R}^d$ parameterized by an uncertainty element δ, a decision vector of optimization variables $x \in \mathbb{R}^d$ has to be chosen within a convex decision set $\mathcal{X}$ so as to minimize a convex cost function $c(x)$, with the requirement that constraint $x \in \mathcal{X}_\delta$ is satisfied for a large enough portion of the possible values of δ (some values of δ can be disregarded to avoid over-conservatism in the design). δ is modeled as the outcome from a probability space $(\Delta, \mathcal{D}, \mathbb{P})$, while no knowledge on this probability space is assumed to be in the user's hands (unknown uncertainty generation mechanism). The user is given a sample of independent draws from $(\Delta, \mathcal{D}, \mathbb{P})$, denoted by δ_i, $i = 1, \ldots, N$. These draws are called the *scenarios*. They are used in a data-driven robust program where scenarios substitute the way vaster multitude of all possible δ's. This leads to the so-called *robust*[1] *scenario optimization* problem:

$$\begin{aligned} \min_{x \in \mathcal{X} \subseteq \mathbb{R}^d} \quad & c(x) \\ \text{subject to:} \quad & x \in \bigcap_{i=1,\ldots,N} \mathcal{X}_{\delta_i}, \end{aligned} \tag{1}$$

whose solution, assuming it exists and is unique, is denoted by x^*_N and is called the *scenario solution*.

While x^*_N is feasible for the observed scenarios $\delta_1, \ldots, \delta_N$, still it can be in violation of the constraints associated to δ's that have not be seen. Hence, it matters to quantify the probability for this to happen because this describes the robustness of the design against instances of δ that are encountered in the future. Such probability is called the risk of x^*_N. The main issue with its evaluation is that the risk is not directly accessible because of the lack of knowledge of $\mathbb{P}$.

In mathematical terms, this is captured by the following definition.

Definition 1 (risk of a generic x): The *risk* of a given $x \in \mathcal{X}$ is defined as $V(x) = \mathbb{P}\{\delta \in \Delta : x \notin \mathcal{X}_\delta\}$. $\star$

The risk of x^*_N is then given by $V(x^*_N)$. This is a random variable over $(\Delta^N, \mathcal{D}^N, \mathbb{P}^N)$ given the dependence of x^*_N on $\delta_1, \ldots, \delta_N$. The theory of the scenario approach then aims at characterizing $V(x^*_N)$ distribution-free, that is, the aim is to identify properties of $V(x^*_N)$ that are valid for every $\mathbb{P}$ and, therefore, can be applied even when $\mathbb{P}$ is unknown.

We quickly recall three milestones in the theory of the scenario approach that are relevant to the present work. In [12], $\mathbb{E}[V(x^*_N)]$, the expected value of $V(x^*_N)$, was proven to be bounded by $\frac{d}{N+1}$; in [13], the distribution of $V(x^*_N)$ was proven to be (first-order) stochastically dominated by a Beta distribution with parameters d and $N-d+1$; in [14], the risk $V(x^*_N)$ has been put in relation with a data-dependent quantity s^*_N, called *complexity*, which is the number of a

Paper supported by MUR under the PRIN 2022 project "The Scenario Approach for Control and Non-Convex Design" (project number D53D23001440006).

S. Garatti is with the Dipartimento di Elettronica, Informazione e Bioingegneria - Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italia (e-mail: simone.garatti@polimi.it).

A. Carè and M.C. Campi are with the Dipartimento di Ingegneria dell'Informazione - Università di Brescia, via Branze 38, 25123 Brescia, Italia (e-mail: [algo.care,marco.campi]@unibs.it).

[1]Problem (1) is robust in the sense that all the constraints indexed by $\delta_1, \ldots, \delta_N$ are rigidly enforced, so expressing an attitude to safeguard against the worst. We reiterate that the theory of the scenario approach has been extended to more flexible schemes such as relaxed optimization problems where a solution is allowed to violate some of the constraints (see [11, Section 3.4]). In the same vein, the results in this paper can also be extended to relaxed optimization problems; we do not delve into this extension due to space limitations.

minimal amount of scenarios that suffice to reconstruct x_N^* (see Definition 3 below). The complexity can be measured from data and has been proven to be a valid statistic to estimate $V(x_N^*)$.

In the next subsection I-A, we start with revising more deeply previous results on $\mathbb{E}[V(x_N^*)]$ and their relevance to the problem of evaluating the average number of mistakes in sequential data-driven decision making. This provides the starting point toward establishing our main contribution in this paper: the introduction of a new *risk-averse* data-driven decision making framework and the evaluation of its average number of mistakes.

A. Existing results on $\mathbb{E}[V(x_N^*)]$ and their implications

Two of the first investigations in the scenario approach, [12], [15], targeted explicitly the quantity $\mathbb{E}[V(x_N^*)]$. The key result in [12] is that, if s_N^* never exceeds a given value c, then

$$\mathbb{E}[V(x_N^*)] \leq \frac{c}{N+1}. \tag{2}$$

The aforementioned bound $d/(N+1)$ then follows because provably, see [12], $s_N^* \leq d$ holds for every problem within the framework of (1).

Under the assumption that scenarios are i.i.d. draws, it is easy to reinterpret $\mathbb{E}[V(x_N^*)]$ as the total probability that $N+1$ scenarios $\delta_1, \ldots, \delta_N, \delta_{N+1}$ are drawn such that the "test" constraint $x \in \mathcal{X}_{\delta_{N+1}}$ is violated by the solution x_N^* obtained by solving (1) with the first N scenarios, see [15]. In formulas, this is written as

$$\mathbb{E}[V(x_N^*)] = \mathbb{P}^{N+1}\{x_N^* \notin \mathcal{X}_{\delta_{N+1}}\}. \tag{3}$$

This interpretation makes the bound on $\mathbb{E}[V(x_N^*)]$ valuable in *sequential* decision schemes where the solution is built based on $\delta_1, \ldots, \delta_N$ and used against a new scenario δ_{N+1}, and then the process is repeated over and over, each time considering a new sample of scenarios $\delta_1, \ldots, \delta_N, \delta_{N+1}$. By the law of large numbers, the bound on $\mathbb{E}[V(x_N^*)] = \mathbb{P}^{N+1}\{x_N^* \notin \mathcal{X}_{\delta_{N+1}}\}$ provides an upper bound to the long-run frequency of failures of the solution against the $N+1$-th scenario. We mention here two applications that have been explored in the literature.

1) Investments based on a sliding window of data: In [16, Section 4.2], a sequence of time-indexed scenarios $\{\delta_t : t = 1, 2, 3, \ldots\}$ representing the rate of returns of a body of financial assets is considered. At a time $t \geq N$, the user decides how to invest a given budget on the assets by solving a problem like (1) based on the last N observed scenarios, over the time window $t, \ldots, t-N+1$. The process is repeated at t, $t+1$, $t+2$, $\ldots$ because every next day a new scenario is revealed and one wants to update the investment to better adhere to the market trends. In this context, the risk $V(x_N^*)$ is the probability of exceeding a pre-defined loss threshold (shortfall event), and, under the i.i.d. assumption, the law of large numbers ensures that the frequency of shortfalls in the long run can be bounded by the right-hand side of (2).[2] Note that the usage of a sliding window of fixed length N allows one to accommodate slow time-varying distribution $\mathbb{P}_t$ (as is the case in practice). The idea is that if $\mathbb{P}_t$ can be considered as constant along a time window, then the conclusions drawn under the ideal i.i.d. assumption remain approximately valid.

2) Receding Horizon Control of Dynamic Systems: In [17], the result in (2) is employed to bound the average number of constraint violations in Scenario-based Model Predictive Control, where at each time step a new control action is computed based on a sample of scenarios according to the receding horizon paradigm. In this dynamic setup, there exists a non vanishing correlation between problems at various time steps because of the coupling effect of the system state. While the classic law of large number is not directly applicable, martingale theorems are employed in [17] to relate the frequency of violations in the long run to $E[V(x_N^*)]$.

B. The contribution of this paper

A severe limitation of the result in (2) is that it relies on bounding s_N^* a priori by means of a constant c, which may be conservative. Indeed, for many problems $s_N^* \ll c$ with high probability. Results in [14] have shown that s_N^* is actually directly linked to the risk: with high confidence low values of s_N^* correspond to low risk, while a large s_N^* signals a high risk. Since s_N^* is accessible to the user, this suggests new sequential decision schemes, where solutions are still repeatedly computed on multiple samples of scenarios, but at each iteration the user decides *after computing the actual value of* s_N^* whether to employ the current solution x_N^*. For example:

1) in a sequence of investments (Section I-A.1), if s_N^* is "too high", suggesting a high risk of shortfall, the user may not invest or may invest in an alternative low risk asset;
2) in receding horizon control (Section I-A.2), if s_N^* is "too high", suggesting a high risk of violating the constraints, the user may opt for a backup control policy, cautious but safer.

Let us denote by $\bar{k}$, $\bar{k} \leq N$ the user-chosen complexity threshold, so that x_N^* is adopted when $s_N^* \leq \bar{k}$ (using a machine learning terminology, we can say that the solution "is rejected" when $s_N^* > \bar{k}$, [18]).[3] Then, the question arises: what is the probability $P_{\bar{k}}$ of the event that

> (i) a solution x_N^* is computed from (1) based on N scenarios $\delta_1, \ldots, \delta_N$,

[2] This is true in spite of the partial overlap between time windows. Indeed, the problems solved at the time instants $N, 2N, 3N, 4N, \ldots$ are the same problem instantiated with different i.i.d. data and the law of large numbers allows one to bound the average number of shortfalls incurred over the time instants $N, 2N, 3N, 4N, \ldots$ by means of $\mathbb{E}[V(x_N^*)]$. The same argument can be repeated for the average number of shortfalls along the time instants $k+N, k+2N, k+3N, \ldots$, for $k = 1, 2, \ldots, N-1$, yielding a bound for the average number of shortfalls over all t.

[3] In some contexts, rejecting to use a decision comes to a cost. Such a cost needs be considered when setting up a rejection policy. However, such an issue is not explored in this paper, which only aims at investigating how the number of mistakes computed in this scheme evolve.

(ii) the complexity of the solution is small enough ($s_N^* \leq \bar{k}$),
(iii) the solution violates the next constraint associated with δ_{N+1} (i.e., $x_N^* \notin \mathcal{X}_{\delta_{N+1}}$)?

As is clear, by the law of large numbers, $P_{\bar{k}}$ relates to the frequency of mistakes ($x_N^* \notin \mathcal{X}_{\delta_{N+1}}$) in this new sequential scheme, and the question above, which has thus far remained unanswered, is crucial to quantify the advantage offered by incorporating a rejection option in a sequential scenario scheme.

The main contribution of this paper lies in obtaining a valid upper bound for $P_{\bar{k}}$ for any $\bar{k}$. Specifically, in the next Section II formal definitions and assumptions are introduced. Section III then contains the main results of this paper: first, it is shown that (contrary to what one may be tempted to think by a hasty analogy with (2)) $\frac{\bar{k}}{N+1}$ is *not* a valid upper bound to $P_{\bar{k}}$; then, valid distribution-free bounds for $P_{\bar{k}}$ are provided. Finally, Section IV illustrates the usage of our bounds on a simple sequential prediction problem, while some conclusions are drawn in Section V.

II. Mathematical preliminaries

We start with some formal definition and standard assumptions borrowed from [14].

Assumption 1 (existence and uniqueness): For every N and for every sample $\delta_1, \delta_2, \ldots, \delta_N$, problem (1) admits a solution. If more than one solution exists, x_N^* 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. ⋆

Definition 2 (support constraint): 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) yields a new solution, different from x_N^*. ⋆

The following assumption rules out situations where the boundary of distinct constraints accumulate anomalously. The reader is referred to [14] for a thoroughgoing discussion on this assumption.

Assumption 2 (non-degeneracy): For every N, the solution x_N^* to problem (1) 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. ⋆

Definition 3 (complexity): The complexity s_N^* of the scenario optimization problem (1) is the number of its support constraints. ⋆

The complexity s_N^* depends on the random sample $\delta_1, \delta_2, \ldots, \delta_N$ and is therefore a random variable over $(\Delta^N, \mathcal{D}^N, \mathbb{P}^N)$ like $V(x_N^*)$. Importantly, given a set of scenarios $\delta_1, \ldots, \delta_N$, the user can compute the value of s_N^*, while computing $V(x_N^*)$ requires knowledge of the distribution $\mathbb{P}$, which is unknown.

Remark 1: The terminology "complexity" was first introduced in [19]. It hints at the fact that the solution can be reconstructed from s_N^* scenarios and, hence, s_N^* represents a "complexity of representation" of the solution. The recent literature on the scenario approach has determined that the joint distribution of $V(x_N^*)$ and s_N^* has notable problem-invariant properties. In particular, it is possible to construct valid and effective confidence intervals for $V(x_N^*)$ based on s_N^*, [19]. This is the theoretical ground justifying using s_N^* as a proxy for the inaccessible risk $V(x_N^*)$; ⋆

Going back to the problem stated in Section I-B, we recognize that

$$\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)]$$

($\mathbb{1}(\cdot)$ is the indicator function: $\mathbb{1}(\cdot) = 1$ if $\cdot$ is true, $\mathbb{1}(\cdot) = 0$ otherwise) is the crucial quantity to be bounded instead of $\mathbb{E}[V(x_N^*)]$. In fact, by the definition of $V(x_N^*)$,

$$\begin{aligned}
&\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)] \\
&= \int_{\Delta^N} \left[\int_{\Delta} \mathbb{1}\left(x_N^* \notin \mathcal{X}_\delta\right) \mathbb{P}(\mathrm{d}\delta) \right] \cdot \mathbb{1}\left(s_N^* \leq \bar{k}\right) \\
&\qquad\qquad \mathbb{P}^N(\mathrm{d}\delta_1, \ldots, \mathrm{d}\delta_N) \\
&= \int_{\Delta^{N+1}} \mathbb{1}\left(x_N^* \notin \mathcal{X}_{\delta_{N+1}} \wedge s_N^* \leq \bar{k}\right) \\
&\qquad\qquad \mathbb{P}^N(\mathrm{d}\delta_1, \ldots, \mathrm{d}\delta_N, \mathrm{d}\delta_{N+1}) \\
&= \mathbb{P}^{N+1}\{x_N^* \notin \mathcal{X}_{\delta_{N+1}} \wedge s_N^* \leq \bar{k}\}. \qquad (4)
\end{aligned}$$

That is, $\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)]$ is the quantity $P_{\bar{k}}$ defined in the Introduction.

III. Main results

The main achievement of this paper is offering a valid (and useful) upper-bound for $\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)]$. This is stated as Theorem 1 in Section III-B. Before doing this, we show that the naive bound $\frac{\bar{k}}{N+1}$ is not valid and, therefore, that it is *necessary* to take a margin from it.

A. An impossibility result

Motivated by (2), it is tempting to conjecture that

$$\text{(conjecture)} \qquad \mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)] \leq \frac{\bar{k}}{N+1}. \qquad (5)$$

Proposition 1: The conjecture (5) is false in general. ⋆

Proof: We quickly note that the conjecture (5) was refuted for exchangeable (but non-i.i.d.) sequences of scenarios in [20, Section 3]. Here, we construct a simple counterexample under the current assumptions, building on the same optimization problem as in [14, Section 8]. The problem is briefly recalled in Fig. 1. Let us consider $N = 2$ constraints, and set $\bar{k} = 1$. The probability of $s_N^* \leq 1$ is $(1-p)^2 + 2p(1-p)$, which is the probability of sampling two U-shaped constraints or one V-shaped constraint together with one U-shaped constraint. In both cases, the risk is no smaller than the probability of drawing a V-shaped constraint, which is p. Therefore, $\mathbb{E}\left[V(x_N^*)\mathbb{1}\left(s_N^* \leq 1\right)\right] \geq p \cdot [(1-p)^2 + 2p(1-p)]$. With $p = \frac{1}{\sqrt{3}}$, we obtain $\mathbb{E}\left[V(x_N^*)\mathbb{1}\left(s_N^* \leq 1\right)\right] \geq \frac{1}{\sqrt{3}} \cdot \frac{2}{3} > \frac{1}{3} = \frac{1}{2+1}$. ■

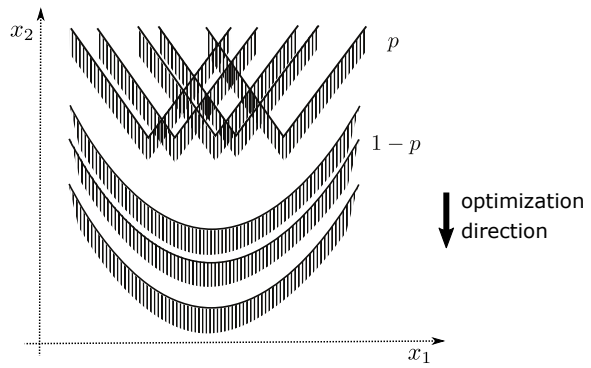


Fig. 1. Representation of the 2-dimensional optimization problem of [14, Section 8], where one minimizes x_2 subject to random constraints belonging to two classes: V-shaped constraints have probability p to be drawn, while U-shaped constraints have probability $1-p$.

B. Valid bound for $\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)]$

Theorem 1: Under Assumptions 1 and 2, it holds that

$$\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)] \leq \begin{cases} r_{N,\bar{k}} & \text{if } \bar{k} < \frac{N}{2} \\ \frac{\bar{k}}{N} & \text{if } \bar{k} \geq \frac{N}{2} \end{cases},$$

where

$$r_{N,\bar{k}} := \min_{\ell=0,1,\ldots,N} \frac{\binom{N}{\bar{k}}}{\binom{\ell}{\bar{k}}} \left(\frac{N-\ell}{N-\ell+1}\right)^{N-\ell} \frac{1}{N-\ell+1}.$$

⋆

Proof: The proof can be found in the internal report [21], which is available at the following URL: https://algocare.it/IntRep/CDC2024AverageProofTh1.pdf. ■

Remarkably, the bound is rather close to the *unattainable* value $\frac{\bar{k}}{N+1}$, even for small values of N, see Figure 2.

IV. Numerical example

We illustrate an application of Theorem 1 to sequential predictions, performed by an Interval Predictor Model (IPM), [22]–[25]. IPM is a rule that assigns to each value of an explanatory variable u (system input) an interval that is used to predict the corresponding system output y. The prediction scheme is described in the next section.

A. A sequential prediction scheme

In the present context the scenarios are input-output pairs $(u, y) \in \mathbb{R}^2$ that are sequentially generated by an unknown system, in such a way that $(u_1, y_1), (u_2, y_2), \ldots$ are i.i.d. draws from a probability measure $\mathbb{P}$. N and d below are two given positive integers with $d \geq N$.

At every time $t \geq N$, we have access to the last N data points (u_{t-N+1}, y_{t-N+1}),(u_{t-N+2}, y_{t-N+2}),. . .,(u_t, y_t); an interval predictor is trained on these N data points and possibly rejected based on its complexity; if it is not rejected, then, as soon as the next input u_{t+1} becomes available, the predictor is used to predict the corresponding, not yet available, output y_{t+1}. The exact details of the procedure are provided below. Note that the procedure depends on two user-chosen hyperparameters:

- $\bar{k}$ is the complexity threshold that, when exceeded, determines the rejection of the predictor;
- $w \in \mathbb{R}$ is the enforced width of the prediction interval.

Moreover, M is the number of runs after which we stop the iterative procedure. The procedure is initialized at time $t := N$, by setting a misprediction counter m_t to zero, i.e., $m_N := 0$.

ITERATIVE PROCEDURE

P.1 Solve the following polynomial fitting problem (with L2-norm Morozov regularization[4]), where θ_j denotes the j-th component of vector $\theta \in \mathbb{R}^d$:

$$\begin{aligned} \min_{\theta \in \mathbb{R}^d} \quad & \|\theta\|^2 \\ \text{subject to:} \quad & \left| y_i - \sum_{j=1}^{d} \theta_j \cdot (u_i)^{j-1} \right| \leq w, \\ & i = t-N+1, \ldots, t \end{aligned} \tag{6}$$

P.2 Denote by θ_N^* the solution to the problem above. Compute s_N^*, the number of support constraints of (6).

P.3 IF $s_N^* > \bar{k}$, THEN *reject* the predictor θ_N^*; wait until the next input u_{t+1} and the next output y_{t+1} are available, and record them; set $m_{t+1} := m_t$ and GO TO step P.6; ELSE proceed with the next step P.4.

P.4 Wait until the next input u_{t+1} becomes available, and record it. Construct the prediction interval $[\sum_{j=1}^{d} \theta_{N,j}^* \cdot (u_{t+1})^{j-1} - w, \sum_{j=1}^{d} \theta_{N,j}^* \cdot (u_{t+1})^{j-1} + w]$ for the output y_{t+1}.

P.5 Wait until the actual value of y_{t+1} becomes available and record it. IF $y_{t+1} \notin [\sum_{j=1}^{d} \theta_{N,j}^* \cdot (u_{t+1})^{j-1} - w, \sum_{j=1}^{d} \theta_{N,j}^* \cdot (u_{t+1})^{j-1} + w]$ (i.e., there is a misprediction), THEN set $m_{t+1} := m_t + 1$; ELSE set $m_{t+1} := m_t$.

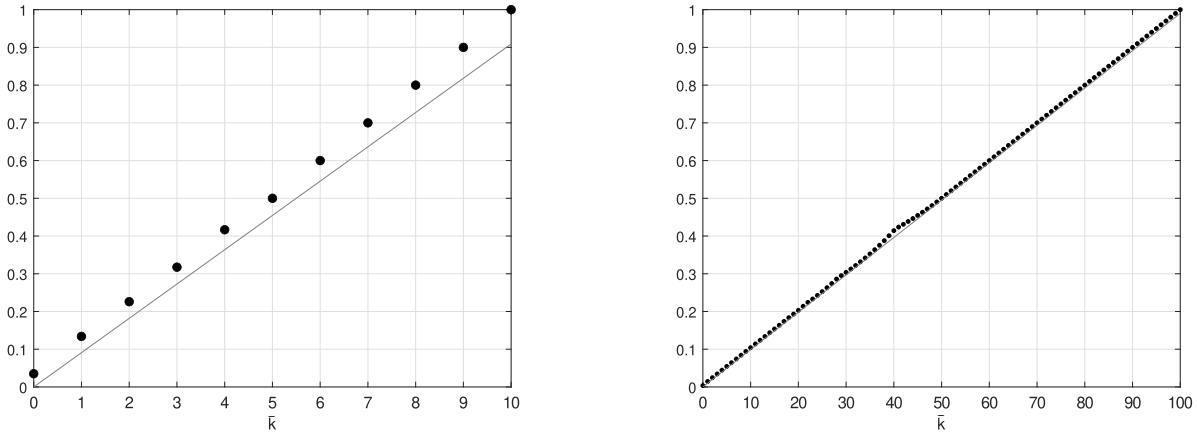
P.6 Let the time windows shift by setting $t := t+1$. IF $t < N + M$, THEN GO TO P.1; ELSE end the procedure.

When the procedure ends, the misprediction rate is given by $\frac{1}{M} m_{N+M}$. We show next that choosing $\bar{k}$ according to Theorem 1 allows one to keep control on the misprediction rate.

B. Usage of Theorem 1

For the sake of concreteness, let $N = 100$ and $d = 101$, and assume that delivering a prediction is desirable, *provided that* the misprediction rate be no larger than 7.5% in the long run. Theorem 1 can be applied to achieve this goal. To do this, note that (6) is a convex scenario program (θ_N^* is just a concrete instance of x_N^*), and the probability of misprediction $\mathbb{P}\{(u, y) : |y - \sum_{j=1}^{d} \theta_{N,j}^* \cdot (u)^{j-1}| > w\}$ associated with the predictor θ_N^* is the risk $V(\theta_N^*)$. Using Theorem 1, we can select the largest $\bar{k}$ such that $\mathbb{E}[V(x_N^*)\mathbb{1}\left(s_N^* \leq \bar{k}\right)] \leq 0.075$.

[4]In (6), the L2 norm of θ is a measure of the capacity of the fitting polynomial; thus, the program (6) can be interpreted as follows: we look for the "simplest" polynomial that fits the data with tolerance w. Though heuristic, this interpretation is rigorous for the limit cases: if $w = 0$, then the solution is a polynomial of degree N that interpolates all data and the complexity is at its maximum, i.e. $s_N^* = N$; if $w = \infty$, then $\theta = 0$ is a data-independent solution and $s_N^* = 0$.



(a) $N = 10$ (b) $N = 100$

Fig. 2. The black dots depicts the bound given by Theorem 1 for $\bar{k} = 0, 1, \ldots, N$, when (a) $N = 10$ and (b) $N = 100$. The gray line corresponds to the unattainable value $\frac{\bar{k}}{N+1}$.

This yields $\bar{k} = 7$. Finally, (4) and a suitable application of the strong law of large numbers (akin to what done in Section I-A.1) yields that $\lim_{M\to\infty} \frac{1}{M} m_{N+M} \leq 7.5\%$ with probability 1.

Remark 2 (a technical clarification): Theorem 1 applies to (6) under Assumptions 1 and 2. When carefully examined, the existence assumption (Assumption 1) looks limiting for (6) because it requires that the problem be feasible not only for $N = 100$, but for any other sample size N given the (fixed) optimization dimension $d = 101$. It is worth remarking, however, that this difficulty can be avoided by defining the solution through a suitable generalized decision map that takes a special value (with complexity N) when the optimization problem is unfeasible. Such a map can be seen to satisfy the conditions in [19, Section 5], and the proof of Theorem 1 in [21] carries over in this more general setup. Regarding Assumption 2, non-degeneracy holds if the regular conditional distribution of y given u admits a density with probability 1. This can be proven using the same arguments as in [9, Section 3.1].

C. Numerical results

We applied our prediction scheme in a concrete example with artificially generated data,[5] with $w = 0.3$ (and $N = 100$, $d = 101$, $\bar{k} = 7$).

We iterated the fitting and prediction procedure $M = 10000$ times. The predictor was rejected in 804 cases out of 10000, and there were overall 670 mispredictions, see Figure 3. As it appears, the misprediction rate keeps below 7.5% in the long run.

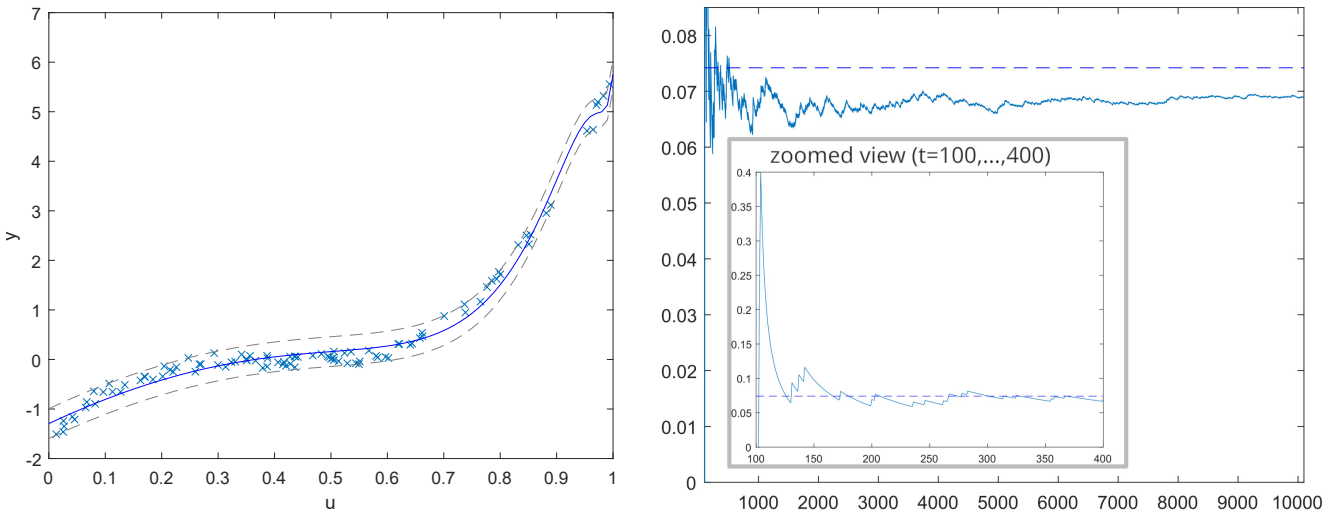
[5]In our simulation, u was uniformly distributed over $[0, 1]$, and $y = 3(u-0.4)^3 + 0.1v$, where v had standard normal distribution, independent of u. This information is shared with the reader to ensure reproducibility: clearly, in a real setup, the generating distribution is unknown to the user.

V. Conclusions and final remarks

We have provided valid bounds for the average number of mistakes in sequential applications of data-driven solutions, allowing the user the possibility to accept or reject each solution based on its complexity. The bounds are rather close to the bound $\frac{\bar{k}}{N+1}$, which we proved to be unattainable. By inspecting Figure 2(b), a curious, relatively small yet visible departure from the general trend is experienced when $\bar{k} \approx 0.4N$. This somehow entices us to seek for (slightly) improved bounds.

An interesting development of the approach of this paper would involve the extension to the case where N is not fixed in advance but is possibly chosen by the user in an iterative manner, by adding constraints until the complexity of the solution is small enough with respect to N. It is known that procedures of this kind can be used to keep the tail probability of $V(x_N^*)$ under control, see [10]; similar schemes aimed at regulating the expected number of mistakes would be highly valuable in applications such as receding horizon control.

Taking a step back from the technical results, we observe that the theory here presented sheds new light on the role of prior knowledge in risk assessments. In fact, it is often the case that a user has freedom in setting up problem (1). For example, in the prediction example of Section IV, the usage of polynomial regression and of a specific kind of regularization in (6) is typically a user's choice, which depends on educated guesses. A common theoretical approach to explaining how preliminary guesses relate to predictive performances is demonstrating that certain specific predictive models work well under restricted data generating mechanisms. Instead, the theory here presented follows a radically different approach: it shows that the data-dependent complexity s_N^* *reveals* the degree to which the (rather free) prior guesses align with the (unrestricted) data generating reality. Since the complexity can be measured in real-time,



(a) (b)

Fig. 3. Results on simulated data. (a) shows the data points at $t = N = 100$, with a representation of the interval predictor around the fitted polynomial. In this case, the predictor had 6 support constraints; thus, it was not rejected. After observing $u_{101} = 0.78$, it yielded the prediction interval [0.945,1.545]; as the actual value of y_{101} turned out to be 1.33 (not shown), the prediction was correct. (b) shows the misprediction rate $\frac{1}{t-100+1} m_{t+1}$ from $t = 100$ to $t = 10099$ (with a zoomed view until $t = 400$). The dotted line is the guaranteed bound 7.5% as provided by Theorem 1.

our findings enable the safe exploitation of prior, tentative knowledge, in an otherwise agnostic setup.

REFERENCES

[1] M. C. Campi and S. Garatti, *Introduction to the scenario approach*. SIAM, 2018.

[2] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.

[3] M. C. Campi, S. Garatti, and M. Prandini, "The scenario approach for systems and control design," *Annual Reviews in Control*, vol. 33, no. 2, pp. 149 – 157, 2009.

[4] T. Alamo, R. Tempo, A. Luque, and D. R. Ramirez, "Randomized methods for design of uncertain systems: sample complexity and sequential algorithms," *Automatica*, vol. 51, pp. 160–172, 2015.

[5] K. Margellos, M. Prandini, and J. Lygeros, "On the connection between compression learning and scenario based single-stage and cascading optimization problems," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2716–2721, 2015.

[6] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, "Performance bounds for the scenario approach and an extension to a class of non-convex programs," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 46–58, 2015.

[7] M. C. Campi, S. Garatti, and F. A. Ramponi, "A general scenario theory for nonconvex optimization and decision making," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4067–4078, 2018.

[8] M. Assif, D. Chatterjee, and R. Banavar, "Scenario approach for minmax optimization with emphasis on the nonconvex case: Positive results and caveats," *SIAM Journal on Optimization*, vol. 30, no. 2, pp. 1119–1143, 2020.

[9] M. C. Campi and S. Garatti, "A theory of the risk for optimization with relaxation and its application to support vector machines," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 13 160–13 197, 2021.

[10] S. Garatti, A. Carè, and M. C. Campi, "Complexity is an effective observable to tune early stopping in scenario optimization," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 928–942, 2022.

[11] M. Campi, A. Carè, and S. Garatti, "The scenario approach: A tool at the service of data-driven decision making," *Annual Reviews in Control*, vol. 52, pp. 1–17, 2021.

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

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

[14] M. C. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Mathematical Programming*, vol. 167, no. 1, pp. 155–189, 2018.

[15] M. C. Campi and G. C. Calafiore, "Notes on the scenario design approach," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 382–385, 2009.

[16] F. A. Ramponi and M. C. Campi, "Expected shortfall: Heuristics and certificates," *European Journal of Operational Research*, vol. 267, no. 3, pp. 1003–1013, 2018.

[17] G. Schildbach, L. Fagiano, C. Frei, and M. Morari, "The scenario approach for Stochastic Model Predictive Control with bounds on closed-loop constraint violations," *Automatica*, vol. 50, no. 12, pp. 3009–3018, 2014.

[18] K. Hendrickx, L. Perini, D. Van der Plas, W. Meert, and J. Davis, "Machine learning with a reject option: A survey," *arXiv preprint arXiv:2107.11277*, 2021.

[19] S. Garatti and M. C. Campi, "Risk and complexity in scenario optimization," *Mathematical Programming*, vol. 191, pp. 243–279, 2022.

[20] M. C. Campi, "Inductive knowledge under dominance," *Synthese*, vol. 201, no. 6, p. 184, 2023.

[21] S. Garatti, A. Carè, and M. C. Campi, "Proof of Theorem 1 in the paper "Average number of mistakes in sequential data-driven risk-averse decision making"," 2024, internal report, available at https://algocare.it/IntRep/CDC2024AverageProofTh1.pdf.

[22] M. Campi, G. Calafiore, and S. Garatti, "Interval predictor models: identification and reliability," *Automatica*, vol. 45, no. 2, pp. 382–392, 2009.

[23] L. G. Crespo, S. P. Kenny, and D. P. Giesy, "Random predictor models for rigorous uncertainty quantification," *International Journal for Uncertainty Quantification*, vol. 5, no. 5, pp. 469–489, 2015.

[24] S. Garatti, M. C. Campi, and A. Carè, "On a class of interval predictor models with universal reliability," *Automatica*, vol. 110, p. 108542, 2019.

[25] C. Wang, C. Shang, F. Yang, D. Huang, and B. Yu, "Robust interval prediction model identification with a posteriori reliability guarantee," in *Proceedings of the 21th IFAC World Congress*, Berlin, Germany, 2020.