On the consistency of the risk evaluation in the scenario approach

Simone Garatti and Marco C. Campi

***Abstract*— The scenario approach is a well-established methodology that allows one to generate solutions from a sample of observations (data-driven decision making). In the recent wait-and-judge paradigm to the scenario approach, the *risk* (i.e., the probability with which a scenario solution does not satisfy new, out-of-sample, constraints) is estimated from an observable called the *complexity* and this result is used to compute intervals that contain with high confidence the value of the risk. In this paper, we establish a new analytical expression for these confidence intervals and we show that they are centered around the complexity divided by the sample size N while their width uniformly (in the complexity) shrinks to zero for increasing N at the rate $O(\ln(N)/\sqrt{N})$ (which is close to the convergence rate of the central limit theorem). This result bears profound implications: (i) it proves the asymptotic consistency of the evaluation of the risk; (ii) as a corollary, it shows that the *complexity* is an observable that carries the fundamental information on the risk (a quantity that is not directly accessible); (iii) it extends the result that the empirical mean tends to the true probability of an event to the case when the event is chosen based on observations via a scenario decision scheme.**

I. Introduction

The scenario approach, [1], is a well-established paradigm for data-driven decision making, which allows the user to design reliable solutions based on observations. One of the simplest design schemes within the scenario approach is the following, [2], [3]:

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

where x is the vector of decision variables belonging to an optimization domain $\mathcal{X}$,[1] $c(x)$ is a cost function, and $\mathcal{X}_{\delta_i}$ are instances of a constraint set $\mathcal{X}_\delta$, which depends on the uncertainty element δ. Parameter δ is modeled as a random outcome from a probability space $(\Delta, \mathcal{D}, \mathbb{P})$ and δ_i, $i = 1, \dots, N$, is a sample of N *independent* draws from this space. The idea is that $(\Delta, \mathcal{D}, \mathbb{P})$ represents the mechanism through which uncertainty is generated; however, this mechanism is unknown to the user and the sole available source of knowledge on uncertainty are the N observations $\delta_1, \delta_2, \dots, \delta_N$, which are called "scenarios". The solution to (1), denoted by x_N^* and called the scenario solution, empirically safeguards against the worst by minimizing the cost function $c(x)$ over the values of x that are feasible for all the scenarios at hand.

The design scheme in (1) is already quite general and instances of (1) have indeed found application to control system design, [4], [5], [6], [7], [8], [9], system identification, [10], [11], [12], [13], [14], and machine learning, [15], [16], [17], [18]. Moreover, design schemes alternative to (1) have been also introduced within the scenario framework, accommodating diverse design requirements, [19], [20], [21], [22], [23], [24], [25], [26], [27] – see also [28], [29], [30] for general paradigms encompassing most of the existing schemes as special cases. While in this paper we prefer to limit ourselves to (1) for the sake of simplicity, the presented results are generally applicable to other design schemes.

One fundamental issue when using the scenario approach is to ascertain how guaranteed x_N^* is in relation to the satisfaction of the constraint $x \in \mathcal{X}_\delta$ for out-of-sample instances of $\delta \in \Delta$. In this context, the following notion of risk is central.

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

As is clear, we are interested in the risk of the scenario solution $V(x_N^*)$, that is, $V(\cdot)$ evaluated at x_N^*. However, it is important to remark that this risk cannot be directly computed because it depends on the probability $\mathbb{P}$, which is unknown (or only partly known) to the user.

In [31] and [28] a new line of attack to the problem of evaluating $V(x_N^*)$, the so-called wait-&-judge paradigm, has been introduced, establishing that $V(x_N^*)$ can be assessed by means of an observable quantity called "complexity". The concept of complexity is formalized in the following definition.

Definition 2 (complexity): A constraint $x \in \mathcal{X}_{\delta_i}$ of the scenario optimization problem (1) is said to be a *support constraint* if its removal (while all other constraints are maintained) changes the solution x_N^*. The *complexity* s_N^* is the number of support constraints of (1). $\star$

Differently from $V(x_N^*)$, the complexity s_N^* can be computed once x_N^* has been calculated.[2] In a nutshell, the achievement of [31] and [28] is that $V(x_N^*)$ and s_N^* seen as functions of the scenarios $\delta_1, \delta_2, \dots, \delta_N$ are always,

S. Garatti is with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, piazza L. da Vinci 32, 20133 Milano, Italy. Email: `simone.garatti@polimi.it`.

M.C. Campi is with the Department of Information Engineering, University of Brescia, via Branze 38, 25123 Brescia, Italy. Email: `marco.campi@unibs.it`.

[1] $\mathcal{X}$ can be any generic space; hence, (1) can well be a program over $\mathbb{R}^d$, but also a calculus of variations problem, in which case $\mathcal{X}$ is an infinite dimensional space.

[2] To this purpose it is enough to apply the definition, which requires to solve N times a problem of the same type as (1) (each time removing one different constraint). Note that in many circumstances shortcuts exist to further reduce the computational burden.

for any Δ and $\mathbb{P}$, two highly dependent random variables, with s_N^* thus carrying fundamental information about $V(x_N^*)$. The results of [31] and [28] have been further refined in [29], where a more precise characterization of the dependence between $V(x_N^*)$ and s_N^* is provided. This has led to the computation of a complexity-dependent interval $[\underline{\epsilon}(s_N^*), \overline{\epsilon}(s_N^*)]$ that is guaranteed to contain the risk $V(x_N^*)$ with high confidence, yielding the remarkable property that the non-accessible quantity $V(x_N^*)$ can be accurately estimated from s_N^*.

As we also recall in the next Section II, the expressions for $\underline{\epsilon}(s_N^*)$ and $\overline{\epsilon}(s_N^*)$ are implicit and so far only numerical computations have revealed that the assessment of $V(x_N^*)$ through $[\underline{\epsilon}(s_N^*), \overline{\epsilon}(s_N^*)]$ is tight and improves as N increases. On the other hand, to date there has not been any attempt to theoretically study the tightness of this result.

The present paper aims at filling this gap. We provide explicit upper and lower bounds for $\underline{\epsilon}(s_N^*)$ and $\overline{\epsilon}(s_N^*)$ from which we show for the first time that the interval $[\underline{\epsilon}(s_N^*), \overline{\epsilon}(s_N^*)]$ shrinks to zero at a fast rate as N increases, uniformly with respect to the value taken by s_N^*. Besides its interest for applications, this result bears the important theoretical implication that the complexity s_N^* is indeed an observable from which the risk can be consistently estimated.

The structure of the paper is rather simple. In the next Section II, after briefly recalling the results of [29], the main result of this paper, Theorem 2, is stated, followed by a discussion about its implications. The rather long proof of Theorem 2 is then given in Section III.

II. Main result

We first recall the theory of [29], which serves as starting point for the findings of this paper. The result holds true under the following assumptions.

Assumption 1 (existence and uniqueness): For every N and for all values of $(\delta_1, \delta_2, \ldots, \delta_N)$, the optimization problem (1) admits at least one solution. If more than one solution exists, the solution x_N^* is singled out by the application of a tie-break rule, that is, by minimizing an additional function $t_1(x)$, and, possibly, other functions $t_2(x)$, $t_3(x)$, …if the tie still occurs. ⋆

The following is a technical non-degeneracy assumption (see Definition 3 of [29] and the discussion therein).

Assumption 2 (non-degeneracy): For every N and with probability one, the solution to (1) coincides with the solution that is obtained after eliminating all the constraints $x \in \mathcal{X}_{\delta_i}$ that are not of support. ⋆

We next recall the result from [29] that provides a quantitative evaluation of the risk $V(x_N^*)$ through s_N^*.

Theorem 1 (Theorem 2 in [29]): For a given value in $(0,1)$ of the confidence parameter β, consider for any $k = 0, 1, \ldots, N-1$ the polynomial equation in the t variable

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

and for $k = N$ the polynomial equation

$$1 - \frac{\beta}{6N} \sum_{i=N+1}^{4N} \binom{i}{N} t^{i-N} = 0. \tag{3}$$

For any $k = 0, 1, \ldots, N-1$ equation (2) 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, equation (3) 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 1 and 2, for any Δ and $\mathbb{P}$ it holds that

$$\mathbb{P}^N \{\underline{\epsilon}(s_N^*) \leq V(x_N^*) \leq \overline{\epsilon}(s_N^*)\} \geq 1 - \beta, \tag{4}$$

i.e. the interval $[\underline{\epsilon}(s_N^*), \overline{\epsilon}(s_N^*)]$ contains the risk $V(x_N^*)$ with confidence $1 - \beta$. ⋆

The main contribution of the present paper is given by the following Theorem 2, which provides explicit bounds for $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ showing thus their dependence on N, k and β.

Theorem 2: Functions $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ introduced in Theorem 1 are subject to the following bounds:

$$\overline{\epsilon}(k) \leq \frac{k}{N} + C \frac{\sqrt{k} \ln \frac{1}{\beta} + \sqrt{k} \ln k + 1}{N} \tag{5}$$

$$\underline{\epsilon}(k) \geq \frac{k}{N} - C \frac{\sqrt{k} \ln \frac{1}{\beta} + \sqrt{k} \ln k + 1}{N} \tag{6}$$

where C is a suitable constant (independent of k, N and β) and the bounds hold for $1 \leq k \leq N$ and $\beta \in (0,1)$, while, for $k = 0$, we have $\overline{\epsilon}(0) \leq (\ln(1/\beta) + 1) \cdot C/N$ and $\underline{\epsilon}(0) \geq 0$. ⋆

Proof: see Section III ⋆

In (5) and (6), the dependence in β is inversely logarithmic, which shows that "confidence is cheap", so much so that very small values of β can be enforced without significantly affecting the width of $[\underline{\epsilon}(k), \overline{\epsilon}(k)]$. For any fixed k, we see that $\overline{\epsilon}(k)$ and $\underline{\epsilon}(k)$ merge onto the same value k/N as fast as $O(1/N)$, while for k that grows at the same rate as N, say $k = \mu N$, convergence towards $k/N = \mu$ takes place at a rate $O(\ln(N)/\sqrt{N})$, which is almost the convergence rate of the central limit theorem. Hence, we see that we can construct a strip around k/N whose size goes to zero as $O(\ln(N)/\sqrt{N})$ and, thanks to (4), the bi-variate distribution of risk and complexity all lies in the strip but a slim tail that expands beyond the strip whose probability is no more than β.

Apart from showing that $V(x_N^*)$ can be tightly estimated form s_N^*, the result of Theorem 2 has the very important implication that the ratio s_N^*/N is always an asymptotically exact estimator of $V(x^*)$ irrespective of the problem at hand. As a matter of fact, combining Theorem 1 and 2, it is almost immediate to show that $|V(x_N^*) - s_N^*/N|$ converges to zero both in the mean square sense and almost surely. This result rigorously proves that s_N^* is an observable allowing one to consistently estimate $V(x^*)$ and it grounds the wait-&-judge paradigm of [31], [29] on a more solid basis.

Remark 1: Before the results of [29], it was shown in [28] that, for any Δ and $\mathbb{P}$, $\mathbb{P}^N\{V(x_N^*) \leq \tilde{\epsilon}(s_N^*)\} \geq 1-\beta$ where $\tilde{\epsilon}(k) = 1 - \sqrt[N-d]{\frac{\beta}{N\binom{N}{d}}}$. Let alone that in this result there is no lower bound to $V(x_N^*)$, it is worth noticing that the findings of Theorem 2 are not valid for $\tilde{\epsilon}(k)$. In fact, for $k = \mu N$ the upper bound $\tilde{\epsilon}(k)$ does not converge to k/N as $N \to \infty$, since, as it is shown in Appendix A, $\tilde{\epsilon}(k) \geq 1-(1-k/N)(k/N)^{\frac{k/N}{1-k/N}} = 1-(1-\mu)\mu^{\frac{\mu}{1-\mu}}$ asymptotically. This substantially different behavior between $\tilde{\epsilon}(k)$ and $\overline{\epsilon}(k)$

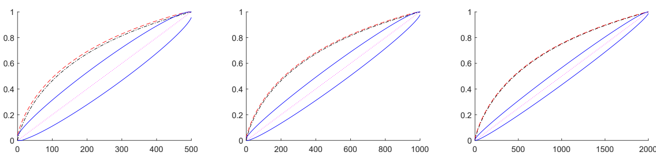

Fig. 1. $\tilde{\epsilon}(k)$ (red dashed line), $1-(1-k/N)(k/N)^{\frac{k/N}{1-k/N}}$ (black dash-dotted line), $\overline{\epsilon}(k)$ and $\underline{\epsilon}(k)$ (blue solid lines), and k/N (pink dotted line) as functions of $k = 0, 1 \ldots, N$ for $N = 500$, $N = 1000$, and $N = 2000$.

can be also appreciated in Figure 1, where $\tilde{\epsilon}(k)$ along with $1-(1-k/N)(k/N)^{\frac{k/N}{1-k/N}}$ and $\overline{\epsilon}(k)$ and $\underline{\epsilon}(k)$ along with k/N are plotted as functions of $k = 0, 1 \ldots, N$ for increasing values of N. ⋆

III. Proof of Theorem 2

Let $v := 1-t$. Equation (2) for $k = 0, \ldots, N-1$ becomes

$$\frac{\beta}{2N}\sum_{i=k}^{N-1}\binom{i}{k}(1-v)^{i-k} + \frac{\beta}{6N}\sum_{i=N+1}^{4N}\binom{i}{k}(1-v)^{i-k} = \binom{N}{k}(1-v)^{N-k}. \quad (7)$$

The fact that (2) has two solutions in $[0, +\infty)$, as stated in Theorem 1, translates into that equation (7) has two solutions in $(-\infty, 1]$, namely $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$. Observing that the left-hand side of (7) is equal to $\beta/2N > 0$ for $v = 1$, while the right-hand side is zero at the same point, we then conclude that, when running backward from 1 to $-\infty$, the left-hand side is first above, then below, and then above again of the right-hand side, as graphically illustrated in Figure 2.

Next consider the following two inequality conditions:

$$\frac{\beta}{2N}\sum_{i=k}^{N-1}\binom{i}{k}(1-v)^{i-k} \geq \binom{N}{k}(1-v)^{N-k}, \quad (8)$$

$$\frac{\beta}{6N}\sum_{i=N+1}^{4N}\binom{i}{k}(1-v)^{i-k} \geq \binom{N}{k}(1-v)^{N-k}. \quad (9)$$

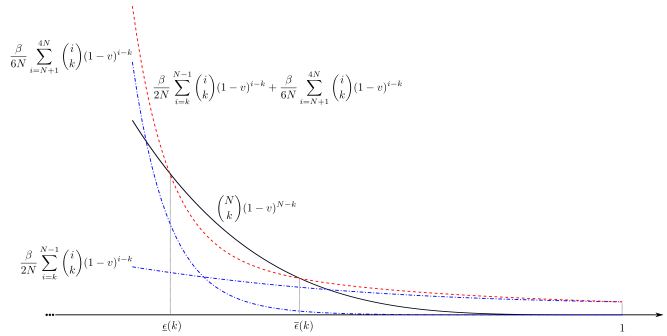

Fig. 2. Graph of functions in (7), (8), and (9) (right-hand sides of (7) = solid black line; left-hand side of (7) = dashed red line; left-hand sides of (8) and (9) = dashed-dotted blue lines).

These two inequalities can be used to effectively locate a suitable upper-bound for $\overline{\epsilon}(k)$ (inequality (8)) and lower-bound for $\underline{\epsilon}(k)$ (inequality (9)). This is explained as follows. Take the ratio of the left-hand side over the right-hand side of equation (8):

$$\frac{\beta}{2N}\sum_{i=k}^{N-1}\frac{\binom{i}{k}}{\binom{N}{k}}(1-v)^{i-N}.$$

Over $(-\infty, 1)$, this function is strictly increasing, moreover for $v = 0$ it is smaller than $\beta/2 < 1$ (note that $\frac{\binom{i}{k}}{\binom{N}{k}} < 1$) while it tends to $+\infty$ as $v \to 1$. Therefore, it picks the value 1 in one and only one point in $(0, 1)$, which shows that equality is attained in (8) for only one value of $v \in (0, 1)$. Hence, the two functions showing up in the left-hand and right-hand sides of (8) are mutually positioned as shown in Figure 2 (note that the right-hand side of (8) coincides with that of (7)).

Further, it is claimed that any v satisfying (8) is an upper-bound to $\overline{\epsilon}(k)$. Indeed, when moving from equation (7) to (8) we have removed from the left-hand side of (7) a positive term, so shifting to the right the point where equality is achieved in (8); then, owing to the mutual position of the two functions in (8) one immediately sees the correctness of the claim.

The inequality condition (9) can be studied in full analogy to (8) with the only advisory that the role of interval $(0, 1)$ is played by $(1, -\infty)$ when considering the second inequality (9).

⋄ *Preliminary calculations*

To study (8) and (9), we shall use a re-writing of the left-hand sides of these inequalities as given in the following.

Let

$$\varphi_{H,k}(v) = \sum_{i=k}^{H-1}\binom{i}{k}(1-v)^{i-k}.$$

Notice first that, for $k = 0$, we have $\varphi_{H,0}(v) = \sum_{i=0}^{H-1}(1-v)^i = \frac{1-(1-v)^H}{v}$. Next, for $k \leq H-1$, a direct verification proves the validity of the following updating rule

$$\varphi_{H,k}(v) = -\frac{1}{k}\frac{\mathrm{d}}{\mathrm{d}v}\varphi_{H,k-1}(v). \qquad (10)$$

A repeated use (a cumbersome but straightforward exercise) of (10) now gives

$$\varphi_{H,k}(v) = \frac{1-\sum_{i=0}^{k}\binom{H}{i}v^i(1-v)^{H-i}}{v^{k+1}} \qquad (11)$$

$$= \frac{\sum_{i=k+1}^{H}\binom{H}{i}v^i(1-v)^{H-i}}{v^{k+1}}. \qquad (12)$$

$\diamond$ *Upper bounding $\overline{\epsilon}(k)$*

Substituting (11) in (8), (8) becomes

$$\frac{\beta}{2}\left(1-\sum_{i=0}^{k}\binom{N}{i}v^i(1-v)^{N-i}\right) \geq N\binom{N}{k}v^{k+1}(1-v)^{N-k}. \qquad (13)$$

If we further decrease the left-hand side (and increase the right-hand side) we obtain an inequality the solutions of which are still upper-bounds to $\overline{\epsilon}(k)$. Starting with the left-hand side, we apply an argument first used in [8] and, for any $a > 1$, write:

$$\begin{aligned}
&\sum_{i=0}^{k}\binom{N}{i}v^i(1-v)^{N-i} \\
&\quad\leq a^k\sum_{i=0}^{k}\binom{N}{i}\left(\frac{v}{a}\right)^i(1-v)^{N-i} \\
&\quad\leq a^k\sum_{i=0}^{N}\binom{N}{i}\left(\frac{v}{a}\right)^i(1-v)^{N-i} \\
&\quad= a^k\left(1-v+\frac{v}{a}\right)^N \\
&\quad= (1-(1-a))^k\left(1-\frac{a-1}{a}v\right)^N \\
&\quad\leq \mathrm{e}^{-(1-a)k}\mathrm{e}^{-\frac{a-1}{a}vN}, \qquad (14)
\end{aligned}$$

where the last inequality follows from relation $1-z \leq \mathrm{e}^{-z}$. Similarly,

$$\begin{aligned}
&N\binom{N}{k}v^{k+1}(1-v)^{N-k} \\
&\quad\leq (k+1)\binom{N+1}{k+1}v^{k+1}(1-v)^{N+1-(k+1)} \\
&\quad\leq (k+1)\sum_{i=0}^{k+1}\binom{N+1}{i}v^i(1-v)^{N+1-i} \\
&\quad\leq (k+1)\mathrm{e}^{-(1-a)(k+1)}\mathrm{e}^{-\frac{a-1}{a}v(N+1)} \\
&\quad\leq (k+1)\mathrm{e}^{-(1-a)}\mathrm{e}^{-(1-a)k}\mathrm{e}^{-\frac{a-1}{a}vN}. \qquad (15)
\end{aligned}$$

Suppose now $k > 0$ (the case $k = 0$ will be considered separately) and take $a = 1 + 1/\sqrt{k}$. Using (14) and (15) in (13) yields that any v coming from the inequality

$$\frac{\beta}{2}\left(1-\mathrm{e}^{\sqrt{k}}\mathrm{e}^{-\frac{vN}{\sqrt{k+1}}}\right) \geq (k+1)\mathrm{e}^{\frac{1}{\sqrt{k}}}\mathrm{e}^{\sqrt{k}}\mathrm{e}^{-\frac{vN}{\sqrt{k+1}}}$$

is an upper bound to $\overline{\epsilon}(k)$. This inequality is equivalent to

$$\frac{\beta}{2(k+1)} \geq \mathrm{e}^{\sqrt{k}}\mathrm{e}^{-\frac{vN}{\sqrt{k+1}}}\left[\frac{\beta}{2(k+1)}+\mathrm{e}^{\frac{1}{\sqrt{k}}}\right]$$

and, solving for v, we obtain

$$v \geq \frac{k}{N}+\frac{\sqrt{k}+1}{N}\left(\lambda+\ln\frac{2}{\beta}+\ln(k+1)\right),$$

where $\lambda = \ln\left[\frac{\beta}{2(k+1)}+\mathrm{e}^{\frac{1}{\sqrt{k}}}\right]+\frac{\sqrt{k}}{\sqrt{k+1}}$. This shows that

$$\overline{\epsilon}(k) \leq \frac{k}{N}+\frac{\sqrt{k}+1}{N}\left(\lambda+\ln\frac{2}{\beta}+\ln(k+1)\right)$$

and the validity of (5) (for $k \neq 0, N$ – recall that we started from equation (2) that holds for $k < N$ and further left behind the case k=0) follows by noticing that $\lambda \leq 2$.
Turn now to the case $k = 0, N$.
Case $k = N$ is trivial because $\overline{\epsilon}(N) = 1$, which is clearly in agreement with (5).
As for $k = 0$, go back to (13) and use in it (14) and (15) with $a = 1 + 1/\sqrt{k+1}$, which, after substituting $k = 0$, gives $a = 2$ (adding 1 to k serves the purpose of avoiding division by zero). Operating the same manipulations as before we now obtain

$$v \geq \frac{2}{N}\left(\ln\left[\frac{\beta}{2}+\mathrm{e}\right]+\ln\frac{2}{\beta}\right),$$

which has the form of the upper bound for $\overline{\epsilon}(k)$ given in Theorem 2.

$\diamond$ *Lower bounding $\underline{\epsilon}(k)$*

First, we want to claim that for any k large enough there is a positive v satisfying equation (9). In fact, for $v = 0$ equation (9) reduces to $\frac{\beta}{6N}\sum_{i=N+1}^{4N}\binom{i}{k} \geq \binom{N}{k}$ and, using the hockey-stick identity (i.e., $\sum_{i=r}^{n}\binom{i}{r} = \binom{n+1}{r+1}$), we have

$$\begin{aligned}
&\frac{\beta}{6N}\frac{\sum_{i=N+1}^{4N}\binom{i}{k}}{\binom{N}{k}} \\
&\quad= \frac{\beta}{6N}\frac{\binom{4N+1}{k+1}-\binom{N+1}{k+1}}{\binom{N}{k}} \\
&\quad= \frac{\beta}{6N}\frac{(4N+1)\cdots(4N-k+1)-(N+1)\cdots(N-k+1)}{(N)\cdots(N-k+1)\cdot(k+1)} \\
&\quad\geq \frac{\beta}{6}\frac{2^{k+1}(N+1)\cdots(N-k+1)-(N+1)\cdots(N-k+1)}{(N+1)\cdots(N-k+1)\cdot(k+1)} \\
&\quad= \frac{\beta}{6}\frac{2^{k+1}-1}{k+1},
\end{aligned}$$

which is greater than 1 for any

$$k \geq c_1 + c_2\ln(1/\beta), \qquad (16)$$

where c_1 and c_2 are suitable constants. In what follows, we assume that this latter condition is satisfied and hence seek a positive solution of equation (9).
Using (12) to rewrite the left-hand side of equation (9) as

$\sum_{i=N+1}^{4N} \binom{i}{k}(1-v)^{i-k} = \varphi_{4N+1,k}(v) - \varphi_{N+1,k}(v)$, equation (9) becomes

$$\frac{\beta}{6}\left(\sum_{i=k+1}^{4N+1}\binom{4N+1}{i}v^i(1-v)^{4N+1-i} - \sum_{i=k+1}^{N+1}\binom{N+1}{i}v^i(1-v)^{N+1-i}\right) \geq N\binom{N}{k}v^{k+1}(1-v)^{N-k}, \tag{17}$$

where moving term v^{k+1} to the right-hand side does not change the inequality sign because v is positive. Similarly to what we did to find an upper bound for $\overline{\epsilon}(k)$, here we can decrease the left-hand side and increase the right-hand side of (17) to find a valid lower bound for $\underline{\epsilon}(k)$.

Notice first that $\sum_{i=k+1}^{4N+1}\binom{4N+1}{i}v^i(1-v)^{4N+1-i} \geq \frac{1}{2}$ for $v \geq \frac{k+1}{4N+2}$.[3] Thus, using also the fact $N\binom{N}{k} \leq (k+1)\binom{N+1}{k+1}$, we can take

$$\frac{\beta}{6}\left(\frac{1}{2} - \sum_{i=k+1}^{N+1}\binom{N+1}{i}v^i(1-v)^{N+1-i}\right) \geq (k+1)\binom{N+1}{k+1}v^{k+1}(1-v)^{N+1-(k+1)} \tag{18}$$

in place of (17) to obtain a lower bound to $\underline{\epsilon}(k)$ as long as we impose the additional condition that

$$v \geq \frac{k+1}{4N+2}. \tag{19}$$

For any $a > 1$, we now have

$$\begin{aligned}
&\binom{N+1}{k+1}v^{k+1}(1-v)^{N+1-(k+1)} \\
&\leq \sum_{i=k+1}^{N+1}\binom{N+1}{i}v^i(1-v)^{N+1-i} \\
&\leq \frac{1}{a^k}\sum_{i=k+1}^{N+1}\binom{N+1}{i}(av)^i(1-v)^{N+1-i} \\
&\leq \frac{1}{a^k}\sum_{i=0}^{N+1}\binom{N+1}{i}(av)^i(1-v)^{N+1-i} \\
&= \frac{1}{a^k}\left(1+(a-1)v\right)^{N+1} \\
&\leq \frac{\mathrm{e}^{(a-1)v(N+1)}}{a^k},
\end{aligned}$$

where the last inequality follows from relation $1 + z \leq \mathrm{e}^z$. Assume $k > 0$ and take $a = 1+1/\sqrt{k}$. Using the above chain of inequalities twice in (18) (for the term in the left-hand side of (18) we use the inequality obtained by comparing the second with the last term in the chain), we obtain the following condition that is more restrictive than (18)

$$\frac{\beta}{6}\left(\frac{1}{2} - \frac{\mathrm{e}^{\frac{v(N+1)}{\sqrt{k}}}}{(1+\frac{1}{\sqrt{k}})^k}\right) \geq (k+1)\frac{\mathrm{e}^{\frac{v(N+1)}{\sqrt{k}}}}{(1+\frac{1}{\sqrt{k}})^k}.$$

[3]This follows from the fact that $\sum_{i=k+1}^{4N+1}\binom{4N+1}{i}v^i(1-v)^{4N+1-i}$ is the cumulative distribution function of a Beta distribution and $\frac{k+1}{4N+2}$ is its mean, which is greater than the median, [32].

This inequality is equivalent to

$$\frac{\beta}{12(\frac{\beta}{6}+k+1)} \geq \frac{\mathrm{e}^{\frac{v(N+1)}{\sqrt{k}}}}{(1+\frac{1}{\sqrt{k}})^k},$$

which, solved for v, gives

$$\begin{aligned}
v \quad \leq \quad & \frac{k}{N+1}\ln\left[\left(1+\frac{1}{\sqrt{k}}\right)^{\sqrt{k}}\right] \\
& -\frac{\sqrt{k}}{N+1}\left(\ln\frac{12}{\beta} + \ln(\frac{\beta}{6}+k+1)\right).
\end{aligned}$$

Noticing now that $\ln(1+x) \geq x - x^2/2$ for all $x \geq 0$, we can finally replace the latter inequality with

$$\begin{aligned}
v \quad \leq \quad & \frac{k}{N+1}\left(1-\frac{1}{2\sqrt{k}}\right) \\
& -\frac{\sqrt{k}}{N+1}\left(\ln\frac{12}{\beta} + \ln(\frac{\beta}{6}+k+1)\right),
\end{aligned} \tag{20}$$

which, for a more handy use, we also rewrite as

$$v \leq \frac{k}{N} - g(k,N,\beta),$$

where function $g(k,N,\beta)$ is just the difference between k/N and the right-hand side of (20). Notice also that this equation is valid also for $k = N$ since (3) also leads to (9), which has been our starting point in the derivation.

To conclude the proof, we have to put together all inequalities that limit the choice of v, namely:

(i) $k \geq c_1 + c_2\ln(1/\beta)$ (equation (16));
(ii) $v \geq \frac{k+1}{4N+2}$ (equation (19));
(iii) $v \leq \frac{k}{N} - g(k,N,\beta)$.

Recall that (iii) makes sense only for $k \neq 0$ (the case $k = 0$ takes care of itself because Theorem 2 claims that $\underline{\epsilon}(0) \geq 0$ which is in agreement with the value of $\underline{\epsilon}(0)$ given in Theorem 1). For the time being, leave (i) behind. Now, one can take the value of v that achieves equality in (iii), i.e., $v = \frac{k}{N} - g(k,N,\beta)$, provided that this is compatible with (ii), that is, $\frac{k}{N} - g(k,N,\beta) \geq \frac{k+1}{4N+2}$. This can be re-written as $g(k,N,\beta) \leq \frac{k}{N} - \frac{k+1}{4N+2}$. Instead, for those values of k,N,β for which this latter inequality does not hold, we have $g(k,N,\beta) > \frac{k}{N} - \frac{k+1}{4N+2}$, from which an easy calculation shows that $2g(k,N,\beta) \geq \frac{k}{N}$, or, equivalently, $\frac{k}{N} - 2g(k,N,\beta) \leq 0$. Since $\underline{\epsilon}(k) \geq 0$, we conclude that in any case $\underline{\epsilon}(k) \geq \frac{k}{N} - 2g(k,N,\beta)$, no matter if $g(k,N,\beta) \leq \frac{k}{N} - \frac{k+1}{4N+2}$ is satisfied or not. Noticing now that $g(k,N,\beta)$ can be upper bounded by $C'\frac{\sqrt{k}\ln\frac{1}{\beta}+\sqrt{k}\ln k+1}{N}$ for a suitable value of the constant C', we conclude that

$$\underline{\epsilon}(k) \geq \frac{k}{N} - C\frac{\sqrt{k}\ln\frac{1}{\beta}+\sqrt{k}\ln k+1}{N}, \tag{21}$$

with $C = 2C'$. Consider now condition (i). When (i) is not satisfied we have that $\frac{k}{N} < (c_1 + c_2\ln(1/\beta))/N$. However, this latter inequality implies that the right-hand side of (21) is negative (possibly after enlarging the constant C in (21) to a value that, with a little abuse of notation, we still call C), so that (21) is always a valid lower bound because $\underline{\epsilon}(k)$ is always non-negative. This concludes the proof. $\star$

APPENDIX

A. Proof that $\tilde{\epsilon}(k) \geq 1-(1-k/N)(k/N)^{\frac{k/N}{1-k/N}}$ asymptotically

The proof mainly builds on the following lower and upper bounds to the factorial of an integer, which are due to [33]:

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n}}. \quad (22)$$

Start by noticing that

$$\tilde{\epsilon}(k) \geq 1-\sqrt[N-k]{\frac{\beta}{N\binom{N}{k}}} = 1-\left(\frac{\beta k!(N-k)!}{N!}\right)^{\frac{1}{N-k}}.$$

where the first inequality derives from $-\ln(x) \geq 1-x$. Using (22) to properly bound the factorials at the numerator and denominator in the last expression yields

$$\begin{aligned}\tilde{\epsilon}(k) \quad \geq \quad & 1-\left(\beta\sqrt{2\pi}\right)^{\frac{1}{N-k}} \times e^{\left(\frac{1}{12k}+\frac{1}{12(N-k)}-\frac{1}{12N+1}\right)\frac{1}{N-k}} \times \\ & \times \left(\frac{k(N-k)}{N}\right)^{\frac{1}{2(N-k)}} \times \left(1-\frac{k}{N}\right)\left(\frac{k}{N}\right)^{\frac{\frac{k}{N}}{1-\frac{k}{N}}}.\end{aligned}$$

Take now $k=\mu N$. The first three terms in the product in the last expression tend to 1 as $N \to \infty$. Whence,

$$\tilde{\epsilon}(k) \geq 1-\left(1-\frac{k}{N}\right)\left(\frac{k}{N}\right)^{\frac{\frac{k}{N}}{1-\frac{k}{N}}} = 1-(1-\mu)\,\mu^{\frac{\mu}{1-\mu}}$$

as $N \to \infty$. This concludes the proof. $\star$

REFERENCES

[1] M. Campi and S. Garatti, *Introduction to Scenario Oprimization*, ser. MOS-SIAM series on Optimization. SIAM, 2018.

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

[3] 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.

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

[5] M. 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.

[6] 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.

[7] S. Grammatico, X. Zhang, K. Margellos, P. Goulart, and J. Lygeros, "A scenario approach for non-convex control design," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 334–345, 2016.

[8] 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.

[9] M. Campi, S. Garatti, and M. Prandini, "Scenario optimization for mpc," in *Handbook of Model Predictive Control*, S. Raković and W. Levine, Eds. Cham, Switzerland: Birkhäuser, 2019, pp. 445–463.

[10] J. Welsh and C. Rojas, "A scenario based approach to robust experiment design," in *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.

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

[12] J. Welsh and H. Kong, "Robust experiment design through randomisation with chance constraints," in *Proceedings of the 18th IFAC World Congress*, Milan, Italy, 2011.

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

[14] S. Garatti, M. Campi, and A. Caré, "On a class of interval predictor models with universal reliability," *Automatica*, vol. 110, no. 108542, pp. 1–9, 2019.

[15] M. Campi, "Classification with guaranteed probability of error," *Machine Learning*, vol. 80, pp. 63–84, 2010.

[16] M. Campi and A. Carè, "Random convex programs with l_1-regularization: sparsity and generalization," *SIAM Journal on Control and Optimization*, vol. 51, no. 5, pp. 3532–3557, 2013.

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

[18] A. Caré, F. Ramponi, and M. Campi, "A new classification algorithm with guaranteed sensitivity and specificity for medical applications," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 393–398, 2018.

[19] M. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality," *Journal of Optimization Theory and Applications*, vol. 148, no. 2, pp. 257–280, 2011.

[20] S. Garatti and M. Campi, "Modulating robustness in control design: principles and algorithms," *IEEE Control Systems*, vol. 33, no. 2, pp. 36–51, 2013.

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

[22] A. Carè, S. Garatti, and M. Campi, "FAST - Fast Algorithm for the Scenario Technique," *Operations Research*, vol. 62, no. 3, pp. 662–671, 2014.

[23] ——, "Scenario min-max optimization and the risk of empirical costs," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2061–2080, 2015.

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

[25] D. Paccagnan and M. Campi, "The scenario approach meets uncertain game theory and variational inequalities," in *Proceeedings of the 58th IEEE Conference on Decision and Control*, 2019.

[26] F. Fele and K. Margellos, "Probably approximately correct nash equilibrium learning," *IEEE Transactions on Automatic Control*, 2020, early access.

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

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

[29] S. Garatti and M. Campi, "Risk and complexity in scenario optimization," *Mathematical Programming*, 2019, published on-line. DOI: https://doi.org/10.1007/s10107-019-01446-4.

[30] ——, "The risk of making decisions from data through the lens of the scenario approach," in *Proceeedings of the 19th IFAC Symposium on System Identification*, 2021.

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

[32] M. Payton, L. Young, and J. Young, "Bounds for the difference between median and mean of beta and negative binomial distributions," *Metrika*, vol. 36, p. 347?354, 1989.

[33] H. Robbins, "The american mathematical monthly," *Metrika*, vol. 62, p. 26?29, 1955.