

Revisiting some results on the sample complexity of multistage stochastic programs and some extensions

M.M.C.R. Reaiche*
IMPA, Rio de Janeiro, RJ, Brazil

March 17, 2016

Abstract

In this work we present explicit definitions for the *sample complexity* associated with the Sample Average Approximation (SAA) Method for *instances* and *classes* of multistage stochastic optimization problems. For such, we follow the same notion firstly considered in Kleywegt et al. (2001). We define the sample complexity for an arbitrary class of problems by considering its worst case behavior, as it is a common approach in the complexity theory literature. We review some sample complexity results for the SAA method obtained so far in the literature, for the static and multistage setting, under this umbrella. Indeed, we show that the derived sample sizes estimates in Shapiro(2006) are *upper bounds* for the *sample complexity* of the SAA method in the multistage setting. We extend one of this results, relaxing some regularity conditions, to address a more general class of multistage stochastic problems. In our derivation we consider the general (finite) multistage case $T \geq 3$ with details. Comparing the *upper bounds* obtained for the general finite multistage case, $T \geq 3$, with the *static* or *two-stage* case, $T = 2$, we observe that, additionally to the exponentially growth behavior with respect to the number of stages, a multiplicative factor of the order $(T - 1)^{2(T-1)}$ appears in the derived *upper bound* for the T -stage case. This shows that the *upper bound* for T -stage problems grows even faster, with respect to T , than the *upper bound* for the static case to the power of $T - 1$.

Keywords: Stochastic programming, Monte Carlo sampling, Sample average method, Complexity

1 Introduction

Consider the following general multistage stochastic optimization problem

$$\min_{x_1 \in X_1} \left\{ F_1(x_1) + \mathbb{E}_{|\xi_1} \left[\inf_{x_2 \in X_2(x_1, \xi_2)} F_2(x_2, \xi_2) \quad + \mathbb{E}_{|\xi_{[2]}} \left[\dots + \mathbb{E}_{|\xi_{[T-1]}} \left[\inf_{x_T \in X_T(x_{T-1}, \xi_T)} F_T(x_T, \xi_T) \right] \right] \right] \right] \right\}, \quad (1)$$

driven by the random data process (ξ_1, \dots, ξ_T) that is defined in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here, $x_t \in \mathbb{R}^{n_t}, t = 1, \dots, T$, are the decisions variables, $F_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \rightarrow \mathbb{R}$ are the stage (immediate) cost functions that we assume are continuous, and $X_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}, t = 2, \dots, T$, are the stage constraint (measurable) multifunctions. The (continuous) function $F_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$, the (nonempty) set X_1 and the vector ξ_1 are deterministic. In this work, when we refer to an instance of (1), all this features will be

*Ph.D. candidate at IMPA and economist of the Brazilian Development Bank. E-mail: mmcr@impa.br

automatically assumed. In (1) we have made explicitly the fact that the information available until each stage, $\xi_{[t]} := (\xi_1, \dots, \xi_t)$, is considered in order to calculate the expected values. This coincides with the regular expected values when the random data is stagewise independent.

An instance (p) of the T -stage problem (1) is completely characterized by its data, say $Data(p)$, that are

1. the stage cost functions F_t , $t = 1, \dots, T$,
2. the stage constraints multifunctions X_t , $t = 1, \dots, T$,
3. the probability distribution, say P , on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ of the discrete-time stochastic process (ξ_2, \dots, ξ_T)

$$P(B) := \mathbb{P}[(\xi_2, \dots, \xi_T) \in B], \forall B \in \mathcal{B}(\mathbb{R}^d).$$

One important aspect of problem's data is the probability distribution P on \mathbb{R}^d , induced by ξ and \mathbb{P} . In many applications, such as in finance, it is natural to assume that ξ has a density with respect to the Lebesgue measure on \mathbb{R}^d . Examples of such distributions are the multivariate normal distribution on \mathbb{R}^d and the uniform distribution on some convex body C of \mathbb{R}^d . In such cases, it is in general not possible to evaluate with high accuracy the expected-value operators on equation (1). Indeed, even for two-stage problems, the evaluation of the expectation consists in computing a multidimensional integral on \mathbb{R}^{d_2} of an optimal-value function that, typically, has to be evaluated by solving an optimization problem (the second-stage optimization problem).

When it is possible to sample from the probability distribution of the random data, one can resort to sampling techniques to circumvent this issue. In this work we will discuss only the SAA Method under the standard Monte Carlo sampling scheme. In this method, one discretize (the state-space) the random data constructing a scenario tree from the sampling realization. After that, one obtain the SAA problem that must be solved numerically by an appropriate optimization algorithm. It is interesting to note that the first two items of the SAA problem's data are equal to the true problem. However, in general, the third item is different, as for the SAA problem the probability distribution of the discrete-time stochastic process has finite state-space. This allows us to evaluate the integrals with respect to this empirical probability distribution. Now, the multidimensional integrals become just finite sums and the problem is typically much easier to address numerically than the original one.

It is important to remember that the problem we really want to solve is the original one, not the SAA problem. However, we will typically be only capable to obtain a (approximate) solution of the SAA problem. So, it is natural to ask how the solution sets of both problems relate. Moreover, how this relationship is affected by the number of samples in each stage. In general, one can make only probabilistic statements about these solutions sets.

In [7] it was derived, for *two-stage* stochastic optimization problems under some reasonable assumptions, a sufficient condition on the sample size to guarantee that, with high probability, every approximate solution of the SAA problem is also an approximate solution of the true problem. In section 2, we present this result with more details and define precisely what we mean by the *sample complexity* of a static stochastic optimization problem and of a *class* of such problems. We will see that the sufficient condition of the sample size derived in [7] is an upper bound for the sample complexity of a class of *static* optimization problems.

In [5] this result was extended to the multistage stochastic setting. In section 3 we define precisely what we mean by the *sample complexity* of a T -stage stochastic optimization problem and of a *class* of such problems. Moreover, we extend some results of [5] to address a more general class of multistage problems (relaxing

some regularity conditions), obtaining also sufficient conditions on the stage sample sizes to guarantee that, with high probability, every approximate solution of the SAA problem is also an approximate solution of the true problem. Again, we show that this estimates can be used to obtain an *upper bound* for the sample complexity of multistage stochastic problems. Then, we show that this *upper bound* exhibits an exponential behavior with respect to the number of stages.

Then, we make our concluding remarks and indicate some future work. This section is followed by three technical appendices, where we proof some results that are used along the main text.

2 Sample Complexity for Static and Two-Stage Problems

In this section we consider a *static* or *two-stage* stochastic optimization problem

$$\min_{x_1 \in X_1} \{f(x_1) := \mathbb{E}[G(x_1, \xi)]\} . \quad (2)$$

The *two-stage* formulation can be considered by writing

$$G(x_1, \xi) = F_1(x_1) + Q_2(x_1, \xi), \quad (3)$$

where

$$Q_2(x_1, \xi) := \inf_{x_2 \in X_2(x_1, \xi)} F_2(x_2, \xi) \quad (4)$$

is the (second-stage) optimal-value function. Let $\{\xi^1, \dots, \xi^N\}$ be a random sample of size N of i.i.d random vectors equally distributed as ξ . The SAA problem is

$$\min_{x_1 \in X_1} \left\{ \hat{f}_N(x_1) := \frac{1}{N} \sum_{i=1}^N G(x_1, \xi^i) = F_1(x_1) + \frac{1}{N} \sum_{i=1}^N Q_2(x_1, \xi^i) \right\}, \quad (5)$$

that is just the original one (2) with the empirical probability distribution. Observe that \hat{f}_N depends on the random sample $\{\xi^1, \dots, \xi^N\}$. Given a sample realization, problem (5) becomes a deterministic one that can be solved by an adequate optimization algorithm.

As pointed out in the last section, one wants to know the relationship between the solutions sets of problems (2) and (5) with respect to N . Indeed one is really aiming to obtain a solution of problem (2), although solving in practice problem (5). To study this issue with more details, let us establish some mathematical notation. We denote the optimal-value of problems (2) and (5) by

$$v^* := \inf_{x_1 \in X_1} f(x_1) \quad \text{and} \quad \hat{v}_N^* := \inf_{x_1 \in X_1} \hat{f}_N(x_1), \quad (6)$$

respectively. Given $\epsilon \geq 0$, we denote the ϵ -solution set of problems (2) and (5) by

$$S^\epsilon := \{x_1 \in X_1 : f(x_1) \leq v^* + \epsilon\} \quad \text{and} \quad \hat{S}_N^\epsilon := \left\{x_1 \in X_1 : \hat{f}_N(x_1) \leq \hat{v}_N^* + \epsilon\right\}, \quad (7)$$

respectively. When $\epsilon = 0$, we drop the superscripts in (7) and write S and \hat{S}_N . Observe that the set \hat{S}_N^ϵ depends on the sample's realization.

Let us assume that $S \neq \emptyset$, i.e. the true problem has an optimal solution. In order to define the *sample*

complexity for the SAA problem, we consider the following complexity parameters

1. $\epsilon > 0$: is the tolerance level associated with the solution set of problem (2);
2. $0 \leq \delta < \epsilon$: is the accuracy level of the solution obtained for problem (5);
3. $\alpha \in (0, 1)$: is the tolerance level for the probability that an unfavorable event happens (see below).

Let us explain the rationale of the parameters. We assume that the optimizer is satisfied to obtain an ϵ -solution of the true problem. For such, he will solve the SAA problem obtaining a δ -optimal solution. This strategy is guaranteed to work if every δ -optimal solution of the SAA problem is an ϵ -solution of the true problem and if there is any δ -optimal solution of the SAA problem¹. This defines the following favorable event

$$\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right], \quad (8)$$

that he wishes have a high likelihood of occurring, which is controlled by the parameter α ,

$$\mathbb{P} \left(\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right] \right) \geq 1 - \alpha. \quad (9)$$

Equivalently, the probability of the unfavorable event $\left[\hat{S}_N^\delta \not\subseteq S^\epsilon \right] \cup \left[\hat{S}_N^\delta = \emptyset \right]$ must be bound above by α . Given the problem's data and the parameter's values, the probability of event (8) depends also on the sample size N . The *sample complexity* of the SAA method says how large N should be in order that (9) holds. Of course, this quantity will also depend on the complexity parameters ϵ, δ and α .

Now we are ready to define the *sample complexity* of an instance of a *static* or *two-stage* stochastic problem. We consider an instance (p) and identify some of its data with a subscript to make clear that it is instance dependent².

Definition 1 (*The Sample Complexity of an instance of a Static or 2-Stage Stochastic Programming Problem*) *Let (p) be an instance of a static or 2-stage stochastic optimization problem. The sample complexity of this problem is, by definition, the following quantity depending on the parameters $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\alpha \in (0, 1)$*

$$N(\epsilon, \delta, \alpha; p) := \inf \left\{ M_2 \in \mathbb{N} : \mathbb{P}_p \left(\left[\hat{S}_{N_2}^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_{N_2}^\delta \neq \emptyset \right] \right) \geq 1 - \alpha, \forall N_2 \geq M_2 \right\}.$$

Here we adopt the usual notation that the infimum of the empty set is equal to $+\infty$. Observe that the infimum is taken on the set of sample sizes M_2 which the probability of the favorable event is at least $1 - \alpha$ for all sample sizes at least this large. Under reasonable regularity conditions, it is well known that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right] \right) = 1, \quad (10)$$

see [6, Theorem 5.18]. However, as we show in example 1, this sequence of numbers is not monotonically (non-decreasing), in general. This motivate us to define the sample complexity in such a way that if one takes a sample size at least this large, it is guaranteed that the favorable event occurs at least with the prescribed

¹If $\hat{S}_N^\delta = \emptyset$, then $\hat{S}_N^\delta \subseteq S^\epsilon$ immediately. However this situation is not favorable, since we do not obtain an ϵ -solution for the true problem.

²In fact, let us do this only with the probability measure of the problem \mathbb{P}_p in order to not make the notation too heavy. Of course the sets of ϵ -solutions and δ -solutions of the true and SAA problems, respectively, also depend on the instance (p).

probability. This would not be the case if we have defined this quantity as the infimum of the set

$$\left\{M_2 \in \mathbb{N} : \mathbb{P}_p \left(\left[\hat{S}_{M_2}^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_{M_2}^\delta \neq \emptyset \right] \right) \geq 1 - \alpha \right\}.$$

Before presenting the example, we extend this definition to contemplate an abstract *class*, say \mathcal{C} , of *static* or *2-stage* stochastic optimization problems.

Definition 2 (*The Sample Complexity of a Class of Static or 2-Stage Stochastic Programming Problem*)
Let \mathcal{C} be a nonempty class of static or 2-stage stochastic optimization problems. We define the sample complexity of \mathcal{C} as the following quantity depending on the parameters $\epsilon > 0$, $\delta \in [0, \epsilon]$ and $\alpha \in (0, 1)$

$$N(\epsilon, \delta, \alpha; \mathcal{C}) := \sup_{p \in \mathcal{C}} N(\epsilon, \delta, \alpha; p)$$

Observe that our definition has the flavor of others *complexity* notions, since it considers the *worst-case* behavior of a family of problems. Let us point out that another possibility would be to analyze the *typical* behavior of problems in a certain class, but we will not discuss this issue here. Now, let us consider the following example.

Example 1 Consider the following static stochastic optimization problem

$$\min_{x \in \mathbb{R}} \{f(x) := \mathbb{E}[|\xi - x|]\}, \quad (11)$$

where ξ is a random variable with finite expected-value. The set of all medians of ξ is the solution set of this problem. Let us denote the cumulative distribution function (c.d.f.) of ξ by $H(z) := \mathbb{P}[\xi \leq z]$. Recall that, by definition, $m \in \mathbb{R}$ is a median of ξ (or of $H(\cdot)$) if $H(m) = \mathbb{P}[\xi \leq m] \geq 1/2$ and $1 - H(m-) = \mathbb{P}[\xi \geq m] \geq 1/2$. Moreover, it is well known that, for every c.d.f. H , its set of medians is a nonempty closed bounded interval of \mathbb{R} .

Let $\{\xi^1, \dots, \xi^N\}$ be a random sample of ξ . The SAA problem is

$$\min_{x \in \mathbb{R}} \left\{ \hat{f}_N(x) := \hat{\mathbb{E}}[|\xi - x|] = \frac{1}{N} \sum_{i=1}^N |\xi^i - x| \right\}. \quad (12)$$

If $N = 2k - 1$ for some $k \in \mathbb{N}$, then the set of exact optimal solutions for the SAA problem is just $\hat{S}_N = \{\xi^{(k)}\}$, where $\xi^{(1)} \leq \dots \leq \xi^{(N)}$ are the orders statistics. If $N = 2k$ for some $k \in \mathbb{N}$, then $\hat{S}_N = [\xi^{(k)}, \xi^{(k+1)}]$. Now it is easy to show that, in general, the sequence of numbers $\{p_N : N \in \mathbb{N}\}$ is not monotonically, where

$$p_N := \mathbb{P} \left(\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right] \right), \forall N \in \mathbb{N}, \quad (13)$$

$\epsilon > 0$ and $\delta \in [0, \epsilon]$ are fixed.

Let us begin pointing that

$$f(x) = f(0) + \int_0^x (2H(s) - 1) ds, \quad \forall x \in \mathbb{R}. \quad (14)$$

In fact, f is (finite-valued) convex and its right-hand derivative on $x \in \mathbb{R}$ is given by $2H(x) - 1$.

Let us assume that ξ is a symmetric (integrable) random variable around the origin satisfying $\mathbb{P}[\xi \neq 0] > 0^3$. It follows that ξ and $-\xi$ are equally distributed and that $f(-x) = \mathbb{E}[|\xi + x|] = \mathbb{E}[|(-\xi) - x|] = \mathbb{E}[|\xi - x|] = f(x)$, for every $x \in \mathbb{R}$. So, f is an even function that assumes its minimum value at the origin. Moreover, by (14) we see that f is monotonically non-decreasing on \mathbb{R}_+ , since $H(s) \geq 1/2$, for all $s \geq 0$. So, given $\epsilon > 0$ (and $\delta = 0$), the set of ϵ -solutions for the true problem is $S^\epsilon = [-x^\epsilon, x^\epsilon]$, for some $x^\epsilon > 0$. For $N = 1$, $[\hat{S}_N \subseteq S^\epsilon]$ if, and only if, $|\xi^1| \leq x^\epsilon$. For $N = 2$, $[\hat{S}_N \subseteq S^\epsilon]$ if, and only if, $|\xi^1| \leq x^\epsilon$ and $|\xi^2| \leq x^\epsilon$. Observe also that the SAA problem always have an optimal solution, so $p_N = \mathbb{P}\left([\hat{S}_N^\delta \subseteq S^\epsilon]\right)$, for all $N \in \mathbb{N}$. We conclude that

$$p_2 = \mathbb{P}\left[|\xi^1| \leq x^\epsilon, |\xi^2| \leq x^\epsilon\right] = \mathbb{P}\left[|\xi^1| \leq x^\epsilon\right] \times \mathbb{P}\left[|\xi^2| \leq x^\epsilon\right] = p_1^2 < p_1,$$

as long as $p_1 < 1$. Of course, this will be the case for $\epsilon > 0$ sufficiently small. For a concrete example, just consider $\xi \sim U[-1, 1]$ and $\epsilon \in (0, 1/2)$. It is easy to verify that $x^\epsilon = \sqrt{2\epsilon} < 1$, so $p_1 < 1$ and $p_2 < p_1$. \square

The following proposition is immediate from the definitions.

Proposition 1 *Let \mathcal{C} be a nonempty class of static or 2-stage stochastic optimization problems. Let $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\alpha \in (0, 1)$ be such that $N(\epsilon, \delta, \alpha; \mathcal{C}) < \infty$. If the sample size N is at least $N(\epsilon, \delta, \alpha; \mathcal{C})$, then*

$$\mathbb{P}_p\left([\hat{S}_N^\delta \subseteq S^\epsilon] \cap [\hat{S}_N^\delta \neq \emptyset]\right) \geq 1 - \alpha, \forall p \in \mathcal{C}.$$

Moreover, if $N < N(\epsilon, \delta, \alpha; \mathcal{C}) \leq \infty$, then exists $p \in \mathcal{C}$ and $M_2 \geq N$ such that

$$\mathbb{P}_p\left([\hat{S}_{M_2}^\delta \subseteq S^\epsilon] \cap [\hat{S}_{M_2}^\delta \neq \emptyset]\right) < 1 - \alpha.$$

Proof: Let $N(\epsilon, \delta, \alpha; \mathcal{C}) \leq N < \infty$. Given $p \in \mathcal{C}$, we have $N \geq N(\epsilon, \delta, \alpha; p)$, then $\mathbb{P}_p\left([\hat{S}_N^\delta \subseteq S^\epsilon] \cap [\hat{S}_N^\delta \neq \emptyset]\right) \geq 1 - \alpha$ follows, which proof the first part. Now suppose that $N < N(\epsilon, \delta, \alpha; \mathcal{C}) \leq \infty$. Since $N(\epsilon, \delta, \alpha; \mathcal{C}) := \sup_{p \in \mathcal{C}} N(\epsilon, \delta, \alpha; p)$, it follows that exists $p \in \mathcal{C}$ such that $N < N(\epsilon, \delta, \alpha; p)$. Using the definition of $N(\epsilon, \delta, \alpha; p)$ we conclude that exists $M_2 \geq N$ such that $\mathbb{P}_p\left([\hat{S}_{M_2}^\delta \subseteq S^\epsilon] \cap [\hat{S}_{M_2}^\delta \neq \emptyset]\right) < 1 - \alpha$. \square

It is also standard to establish the following result, that we omit the proof.

Proposition 2 *Let (p) be an instance of a static or 2-stage stochastic optimization problem. Given $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\alpha \in (0, 1)$, we have that*

- a. $t \in (\delta, +\infty) \mapsto N(t, \delta, \alpha; p)$ is monotonically non-increasing;
- b. $t \in [0, \epsilon) \mapsto N(\epsilon, t, \alpha; p)$ is monotonically non-decreasing; and
- c. $t \in (0, 1) \mapsto N(\epsilon, \delta, t; p)$ is monotonically non-increasing.

In the sequel, we will see more explicitly how the *upper* bound behaves with respect to these parameters. In [3, 7] it were derived, under some regularity conditions, some estimates on the sample size N in order to equation (9) be satisfied. Here we follow closely the reference [6, Section 5.3.2]. We make only small modifications in our presentation with respect to the previous reference, so the correspondence between the assumptions of the static and multistage settings are more clearly seen. Let us consider the following assumptions:

³This is just to rule out the degenerate case $\xi = 0$.

(A0) For all $x \in X_1$, $f(x) = F(x) + \mathbb{E}\mathcal{Q}_2(x, \xi)$ is finite.

(A1) X_1 has finite diameter $D > 0$.

(A2) There exists a (finite) constant $\sigma > 0$ such that for any $x \in X_1$ the moment generating function $M_x(t)$ of the random variable $Y_x := G(x, \xi) - f(x) = \mathcal{Q}_2(x, \xi) - \mathbb{E}\mathcal{Q}_2(x, \xi)$ satisfies

$$M_x(t) \leq \exp(\sigma^2 t^2 / 2), \quad \forall t \in \mathbb{R}. \quad (15)$$

(A3) There exists a measurable function $\chi : \text{supp}(\xi) \mapsto \mathbb{R}_+$ such that its moment generating $M_\chi(t)$ is finite-valued for all t in a neighborhood of zero and for a.e. $\xi \in \text{supp}(\xi)$

$$|\mathcal{Q}_2(x', \xi) - \mathcal{Q}_2(x, \xi)| \leq \chi(\xi) \|x' - x\|, \quad (16)$$

for all $x', x \in X_1$.

By assumption (A3) the random variable χ has all k^{th} -moments finite, i.e. $\mathbb{E}[\chi^k] < +\infty$, for all $k \in \mathbb{N}$. Denoting $\mathcal{Q}_2(x) := \mathbb{E}\mathcal{Q}_2(x, \xi)$, it follows by (A0) and (A3) that $\mathcal{Q}_2(\cdot)$ is L -Lipschitz continuous on X_1 , where $L := \mathbb{E}[\chi(\xi)]^4$. Remembering that F_1 is continuous and X_1 is (nonempty) bounded and closed, we conclude that $S \neq \emptyset$. Moreover, it also follows from (16) that

$$|\hat{\mathcal{Q}}_2(x') - \hat{\mathcal{Q}}_2(x)| \leq \bar{\chi}_N \|x' - x\|, \quad \text{w.p.1 } \xi^1, \dots, \xi^N, \quad (17)$$

where $\hat{\mathcal{Q}}_2(x) = \frac{1}{N} \sum_{i=1}^N \mathcal{Q}_2(x, \xi^i)$ and $\bar{\chi}_N = \frac{1}{N} \sum_{i=1}^N \chi(\xi^i)$. Since $\chi(\xi)$ is integrable, it is finite with probability one, then $\hat{\mathcal{Q}}_2$ is Lipschitz continuous on X_1 (where the Lipschitz constant depends on the sample realization) with probability one. Since $\hat{f}_N(\cdot) = F_1(\cdot) + \hat{\mathcal{Q}}_2(\cdot)$, we conclude that $\mathbb{P}[\hat{S}_N \neq \emptyset] = 1$. We have shown that both events $[\hat{S}_N^\delta \subseteq S^\epsilon] \cap [\hat{S}_N^\delta \neq \emptyset]$ and $[\hat{S}_N^\delta \subseteq S^\epsilon]$ have the same probability and we don't need to bother with the event $[\hat{S}_N^\delta \neq \emptyset]$, for any $\delta \in [0, \epsilon]$. We also rule out from our analysis the trivial case $L := \mathbb{E}[\chi(\xi)] = 0$. In that case, $\mathcal{Q}(x) = \text{constant} = \hat{\mathcal{Q}}_N(x)$ for every $x \in X_1$, where the last equality holds with probability one. In this case both the true and the SAA problems coincide with probability one (so, just take $N = 1$).

Before continuing, let us introduce some mathematical notation from large deviation theory that will appear in the next result. Let Z be a random variable whose moment generating function is finite on a neighborhood of zero and let $\{Z_i : i = 1, \dots, N\}$ be i.i.d. copies of Z . In such situation, Cramer's Inequality provides the following upper bound for the event

$$\mathbb{P}[\bar{Z} \geq \mu'] \leq \exp(-NI_Z(\mu')), \quad (18)$$

where $\bar{Z} := \frac{1}{N} \sum_{i=1}^N Z_i$, $\mu' > \mathbb{E}[Z]$ and $I_Z(\mu') := (\log M_Z)^*(\mu')$. Here, we denote by h^* the Fenchel-Legendre transform (or convex conjugate) of h , i.e.

$$h^*(y) := \sup_{x \in \mathbb{R}} \{yx - h(x)\}. \quad (19)$$

⁴Observe that (A0) is necessary to conclude the Lipschitz-continuity of $\mathcal{Q}_2(\cdot)$, even if (A3) holds. For example, if $\mathcal{Q}(x, \xi) = \mathcal{Q}(x', \xi)$ for all $x, x' \in X_1$, then (A3) would be trivially satisfied: just take $\chi(\xi) := 0$, for all $\xi \in \text{supp}(\xi)$. However, if $\mathcal{Q}_2(x, \xi)$ has an infinity expected-value, $\mathcal{Q}_2(\cdot)$ would not be Lipschitz-continuous.

Moreover, it is possible to show that $I_Z(\mu') \in (0, +\infty]$ for $\mu' > \mathbb{E}[Z]$. Observe that we are not ruling out the possibility that this quantity assumes the value $+\infty$ and, in this case, the upper bound implies that the probability of the event is zero.

Here, we state **Theorem 7.75** of [6, Section 7.2.10] under straightforward modifications. This theorem gives a uniform exponential bound, under conditions (A0)-(A3), for the probability of the supremum on X_1 of the difference between the true objective function and the SAA objective function diverges more than $\epsilon > 0$. See the reference for a proof.

Theorem 1 *Consider a stochastic optimization problem which satisfies assumptions (A0), (A1), (A2) and (A3). Denote by $L := \mathbb{E}[\chi(\xi)] < \infty$ and let $\gamma > 1$ be given. Let $\{\xi^1, \dots, \xi^N\}$ be a random sample of ξ . Then*

$$\mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}_N(x_1) - f(x_1)| > \epsilon \right] \leq \exp\{-NI_\chi(\gamma L)\} + 2 \left[\frac{2\gamma\rho DL}{\epsilon} \right]^n \exp \left\{ -\frac{N\epsilon^2}{32\sigma^2} \right\}, \quad (20)$$

where ρ is an absolute constant⁵.

In **Theorem 7.75**, it was considered the case $L' := 2L$ and $l := I_\chi(L')$ to obtain the following bound

$$\mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}_N(x_1) - f(x_1)| > \epsilon \right] \leq \exp\{-Nl\} + 2 \left[\frac{4\rho DL}{\epsilon} \right]^n \exp \left\{ -\frac{N\epsilon^2}{32\sigma^2} \right\}. \quad (21)$$

This corresponds to take $\gamma = 2$ for theorem 1. It is not difficult to see that the same proof works for general $\gamma > 1$ and the only modification needed is to take the v -net with $v := \frac{\epsilon}{2\gamma L}$ (observe that it was taken $\frac{\epsilon}{4L}$ in **Theorem 7.75**). Here, we write the result depending explicitly on $\gamma > 1$. Observe that we can make the right-hand side of (20) arbitrarily small by increasing the sample size N ⁶. Moreover, one can verify that

$$\left[\hat{S}_N^\delta \not\subseteq S^\epsilon \right] \subseteq \left[\sup_{x_1 \in X_1} |\hat{f}_N(x_1) - f(x_1)| > \frac{\epsilon - \delta}{2} \right]. \quad (22)$$

This is equivalent to show that $\left[\sup_{x_1 \in X_1} |\hat{f}_N(x_1) - f(x_1)| \leq \frac{\epsilon - \delta}{2} \right] \subseteq \left[\hat{S}_N^\delta \subseteq S^\epsilon \right]$. To show this inclusion, suppose that the first event happened and let \hat{x}_1 be an arbitrary SAA's δ -optimal solution. We will show that it is also an ϵ -optimal solution for the true problem. In fact,

$$f(\hat{x}_1) - \frac{\epsilon - \delta}{2} \leq \hat{f}_N(\hat{x}_1) \leq \hat{f}_N(\bar{x}) + \delta \leq f(\bar{x}) + \frac{\epsilon - \delta}{2} + \delta, \quad (23)$$

where \bar{x} is an optimal solution of the true problem. By (23), we conclude that \hat{x}_1 is an ϵ -solution of the true problem. The following result is an immediate consequence of Theorem 1 and the discussion made above, see also [6, Theorem 5.18].

Theorem 2 *Consider a stochastic optimization problem which satisfies assumptions (A0), (A1), (A2) and (A3). Denote by $L := \mathbb{E}[\chi(\xi)] < \infty$ and let $\gamma > 1$ be arbitrary. Let $\{\xi^1, \dots, \xi^N\}$ be a random sample of ξ . If $\epsilon > 0$, $\delta \in [0, \epsilon)$, $\alpha \in (0, 1)$ and*

$$N \geq \frac{8\sigma^2}{(\epsilon - \delta)^2} \left[n \log \left(\frac{4\rho\gamma LD}{\epsilon - \delta} \right) + \log \left(\frac{4}{\alpha} \right) \right] \vee \left[\frac{1}{I_\chi(\gamma L)} \log \left(\frac{2}{\alpha} \right) \right], \quad (24)$$

⁵In fact, it is possible to show that $\rho \leq 5$, see lemma 4 of appendix A.

⁶Since $L > 0$ and $\gamma > 1$, $\gamma L > \mathbb{E}[\chi(\xi)]$ and $I_\chi(\gamma L) > 0$.

then $\mathbb{P} \left(\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right] \right) \geq 1 - \alpha$.

Inequality (24) was obtained by bounding from above each term of the right-hand side of (20) by $\alpha/2$. Moreover, considering the complementary event of $\left[\hat{S}_N^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_N^\delta \neq \emptyset \right]$, we obtain

$$\begin{aligned} \mathbb{P} \left(\left[\hat{S}_N^\delta \not\subseteq S^\epsilon \right] \cup \left[\hat{S}_N^\delta = \emptyset \right] \right) &\leq \mathbb{P} \left[\hat{S}_N^\delta \not\subseteq S^\epsilon \right] + \mathbb{P} \left[\hat{S}_N^\delta = \emptyset \right] \\ &= \mathbb{P} \left[\hat{S}_N^\delta \not\subseteq S^\epsilon \right] \leq \mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}_N(x_1) - f(x_1)| > \frac{\epsilon - \delta}{2} \right]. \end{aligned}$$

Considering our definition of the sample complexity, theorem 2 says that the right-hand side of (24) is an upper bound for $N(\epsilon, \delta, \alpha; p)$, for every instance (p) satisfying the regularity conditions (A0)-(A3). Of course, each instance (p) will have its own associated parameters, as $\sigma(p) > 0$, $n(p) \in \mathbb{N}$, $L(p) > 0$, $D(p) > 0$. Observe that the *upper bound* presents a modest growth rate in the parameter α , as α goes to zero. Since $I_\chi(\gamma L) \in (0, +\infty]$ is a constant, the maximum on (24) will be attained by the first-term for sufficiently small values of $\epsilon - \delta > 0$. The steepest behavior is with respect to $\epsilon - \delta$, since when it approaches zero, *ceteris paribus*, the rate of growth is of order $\frac{1}{(\epsilon - \delta)^2} \log \left(\frac{1}{\epsilon - \delta} \right)$. Moreover, it is worth noting that the estimate (24) is not dimension free, as it grows linearly with n . Other key parameters related to problem's data are σ^2 and LD . The product LD represents the cross effect of the diameter of the (first-stage) feasible set and the Lipschitz constant of the objective function f on this set and, as we see, it has a modest effect in the upper bound. Finally, σ^2 is a uniform upper bound in the variability of the random variables Y_x (see (A2)) and equation (15) says that this family of random variables is σ -sub-Gaussian, i.e. its tails decay as fast as a $N(0, \sigma^2)$. Finally, the growth behavior with respect to σ^2 is linear, although it is not unusual that σ^2 has a quadratic dependence on the diameter D .

Let us extend the previous result to a class of stochastic problems. In order to prevent the class' complexity upper bound to be $+\infty$, we need to consider finite upper bounds for the parameters instances $\sigma^2(p)$, $n(p)$, $L(p)D(p)$, and also a positive lower bound for $I_{\chi(p)}(\gamma L(p))$. Of course, if the upper bound is $+\infty$, it would not be clear how one could study its behavior with respect to the complexity parameters ϵ, δ and α . The following result is an immediate corollary from theorem 2.

Corollary 1 *Let \mathcal{C} be the class of all static or 2-stage stochastic optimization problems satisfying assumptions (A0)-(A3) and the uniformly bounded condition*

(UB): *there exist positive finite constants σ^2 , M , $n \in \mathbb{N}$, $\gamma > 1$ and β such that for every instance (p) $\in \mathcal{C}$ we have*

- i. $\sigma^2(p) \leq \sigma^2$,
- ii. $D(p) \times L(p) \leq M$,
- iii. $X_1(p) \subseteq \mathbb{R}^{n(p)}$ and $n(p) \leq n$,
- iv. $(0 <) \beta \leq I_{\chi(p)}(\gamma L_p)$,

where $D(p) := \text{diam}(X_1(p))$ and $L(p) := \mathbb{E}[\chi_p(\xi)]$. Then, it follows that

$$N(\epsilon, \delta, \alpha; \mathcal{C}) \leq \frac{O(1)\sigma^2}{(\epsilon - \delta)^2} \left[n \log \left(\frac{O(1)\gamma M}{\epsilon - \delta} \right) + \log \left(\frac{4}{\alpha} \right) \right] \vee \left[\frac{1}{\beta} \log \left(\frac{2}{\alpha} \right) \right], \quad (25)$$

where $O(1)$'s are absolute constants.

Proof: Let $(p) \in \mathcal{C}$ be arbitrary. By the previous theorem, it follows that

$$\begin{aligned} N(\epsilon, \delta, \alpha; p) &\leq \frac{8\sigma^2(p)}{(\epsilon - \delta)^2} \left[n(p) \log \left(\frac{O(1)\gamma D(p)L(p)}{\epsilon - \delta} \right) + \log \left(\frac{4}{\alpha} \right) \right] \vee \left[\frac{1}{I_{\chi(p)}(\gamma L(p))} \log \left(\frac{2}{\alpha} \right) \right] \\ &\leq \frac{O(1)\sigma^2}{(\epsilon - \delta)^2} \left[n \log \left(\frac{O(1)\gamma M}{\epsilon - \delta} \right) + \log \left(\frac{4}{\alpha} \right) \right] \vee \left[\frac{1}{\beta} \log \left(\frac{2}{\alpha} \right) \right], \end{aligned}$$

where the second inequality is an immediate consequence of conditions (i.) – (iv.). Taking the supremum on $p \in \mathcal{C}$ in the expression above, we obtain (25). \square

In the next section we consider the definitions of sample complexity for the multistage setting and some results about its upper bound.

3 Sample Complexity for Multistage Problems

Some of the results of this section were first obtained in [5] and were improved later on [6, Section 5.8.2], that we follow closely. We also present some new improvements with respect to [6, Section 5.8.2], see theorems 3 and 4.

To facilitate the analysis of the multistage setting, we suppose that the constructed scenario tree, via Monte Carlo sampling, possess a specific structure. Let assume that every t^{th} -stage node has N_{t+1} successors nodes in stage $t + 1$, for $t = 1, \dots, T - 1$. Under this assumption, the total number of scenarios in the tree is just

$$N = \prod_{t=2}^T N_t.$$

Later, we will specify how the Monte Carlo sampling is done under the additional hypothesis of stagewise independent random data. Now, consider the following definition.

Definition 3 (*The Sample Complexity of an instance of T-Stage Stochastic Optimization Problem*) Let (p) be a T -stage stochastic optimization problem. Given $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\alpha \in (0, 1)$, we define the subset of viable samples sizes as

$$\mathcal{N}(\epsilon, \delta, \alpha; p) := \left\{ (M_2, \dots, M_T) : \begin{array}{l} \forall (N_2, \dots, N_T) \geq (M_2, \dots, M_T), \\ \mathbb{P}_p([S_{N_2, \dots, N_T}^\delta \subseteq S^\epsilon] \cap [S_{N_2, \dots, N_T}^\delta \neq \emptyset]) \geq 1 - \alpha \end{array} \right\}.$$

The sample complexity of (p) is defined as

$$N(\epsilon, \delta, \alpha; p) := \inf \left\{ \prod_{t=2}^T M_t : (M_2, \dots, M_T) \in \mathcal{N}(\epsilon, \delta, \alpha; p) \right\}.$$

It is immediate to verify that the above definition recovers the previous one for $T = 2$. Now we define the sample complexity for a class of multistage problems.

Definition 4 (*The Sample Complexity of a class of T-Stage Stochastic Programming Problems*) Let \mathcal{C} be a nonempty class of T -stage stochastic optimization problems. We define the sample complexity of \mathcal{C} as the

following quantity depending on the parameters $\epsilon > 0$, $\delta \in [0, \epsilon)$ and $\alpha \in (0, 1)$

$$N(\epsilon, \delta, \alpha; \mathcal{C}) := \sup_{p \in \mathcal{C}} N(\epsilon, \delta, \alpha; p)$$

In [5, 6] it was assumed that the random data was stagewise independent in order to derive the sample size estimates. We will also make this assumption in this work. In this particular case, the number of samples required for each stage random vector, ξ_t , $t = 2, \dots, T$, can be dramatically reduced if we proceed in a particular way.

Firstly, let us consider a scheme that could be applied for any dependence structure of the random data, the *standard conditional Monte Carlo sampling scheme*. In this scheme, for each t -stage node, say $\xi_{[t]}^j$, we generate N_{t+1} scenarios for the next stage by sampling from the distribution of $\xi_{t+1} | \xi_{[t]} = \xi_{[t]}^j$. So, we see that, in general, the number of t -stage scenarios on the tree will be $\prod_{s=2}^t N_s$, for $t = 2, \dots, T$. Of course, we always have $N_1 = 1$, since ξ_1 is deterministic. Moreover, even when the random data is stagewise independent, if we construct the tree following this scheme, we obtain a discrete-state version that is not stagewise independent. So, additionally from memory constraints considerations in order to store the data in the computer, the optimization algorithms commonly used, like the SDDP method, to solve the multistage SAA problem could not take advantage of the *cut sharing* between the stage *cost-to-go* functions. Then, from a practical point of view, it is preferable to construct a scenario tree that is also stagewise independent, when the true random data is.

The following scheme is known as the *identical conditional sampling scheme*. Firstly, we generate independent observations of the stage random vectors, say

$$\mathfrak{S}_{N_2, \dots, N_T} := \left\{ \xi_t^j : t = 2, \dots, T, j = 1, \dots, N_t \right\}, \quad (26)$$

where N_t is the number of copies (sample size) of the random vector ξ_t , for $t = 2, \dots, T$. In this scheme, given the sample realization, the t -stage scenarios are just given by $\{\xi_t^j : j = 1, \dots, N_t\}$, for $t = 2, \dots, T$. Moreover, we consider a probability measure on the tree such that, independently of the history until stage $t - 1$, gives the same probability for visiting the scenarios on the stage t . In fact, we consider in the tree the empirical probability distribution given by

$$\hat{\mathbb{P}} \left[\xi_2 = \xi_2^{j_2}, \dots, \xi_T = \xi_T^{j_T} \right] = \prod_{t=2}^T \frac{\#\{1 \leq i \leq N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t}, \quad (27)$$

for $1 \leq j_t \leq N_t$ and $t = 2, \dots, T$. Observe that, differently from the *standard conditional scheme*, we generate only N_t observations of the t -stage random vector ξ_t , for $t = 2, \dots, T$. So, the data storage requirements are drastically reduced and the stage *cost-to-go* functions are simplified, since they are not path dependent as in the general case.

It is possible to obtain sample estimates considering the *standard conditional Monte Carlo sampling scheme*, as it was already done in [5]. However, in our presentation, similarly to [6], we will consider only the *identical conditional sampling scheme*. As pointed out above, this is particularly appealing for practical applications.

Consider the following regularity conditions for a T -stage stochastic optimization problem:

(M0) The random data is stagewise independent.

(M1) For all $x_1 \in X_1$, $f(x_1)$ is finite.

For each $t = 1, \dots, T-1$:

(Mt.1) There exist a compact set \mathcal{X}_t with diameter D_t such that $X_t(x_{t-1}, \xi_t) \subseteq \mathcal{X}_t$, for every $x_{t-1} \in \mathcal{X}_{t-1}$ and $\xi_t \in \text{supp}(\xi_t)$.

(Mt.2) There exists a (finite) constant $\sigma_t > 0$ such that for any $x \in \mathcal{X}_t$, the following inequality holds

$$M_{t,x}(s) := \mathbb{E}[\exp(s(\mathcal{Q}_{t+1}(x, \xi_{t+1}) - \mathcal{Q}_{t+1}(x)))] \leq \exp(\sigma_t^2 s^2 / 2), \quad \forall s \in \mathbb{R}. \quad (28)$$

(Mt.3) There exists a measurable function $\chi_t : \text{supp}(\xi_{t+1}) \rightarrow \mathbb{R}_+$ such that, for a.e. $\xi_{t+1} \in \text{supp}(\xi_{t+1})$, we have that

$$|\mathcal{Q}_{t+1}(x'_t, \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t, \xi_{t+1})| \leq \chi_t(\xi_{t+1}) \|x'_t - x_t\| \quad (29)$$

holds, for all $x'_t, x_t \in \mathcal{X}_t$. Moreover, its moment generating function $M_{\chi_t}(s)$ is finite in a neighborhood of zero.

(Mt.4) For almost every $\xi_{t+1} \in \text{supp}(\xi_{t+1})$, the constraint multifunction $X_{t+1}(\cdot, \xi_{t+1})$ restricted to the set \mathcal{X}_t is continuous.

Item (M1.1) just asserts that the compact set \mathcal{X}_1 contains X_1 (e.g. $\mathcal{X}_1 = X_1$). Moreover, the functions $\mathcal{Q}_{t+1} : \mathbb{R}^{n_t} \times \mathbb{R}^{d_{t+1}} \rightarrow \bar{\mathbb{R}}$, $t = 1, \dots, T-1$, are the stage optimal-value functions and $\mathcal{Q}_{t+1}(x_t) := \mathbb{E}[\mathcal{Q}_{t+1}(x_t, \xi_{t+1})]$. In Appendix C we review the concept of continuity of a multifunction. The following proposition summarizes some consequences from these regularity conditions.

Proposition 3 *Consider a general T -stage stochastic optimization problem as (1) satisfying the regularity conditions (M0), (M1) and (Mt.1), for $t = 1, \dots, T-1$. The following assertions hold*

- (a) *If the problem also satisfies conditions (Mt.3), for $t = 1, \dots, T-1$, then there exists $E_{t+1} \subseteq \text{supp}(\xi_{t+1})$ such that $\mathbb{P}[\xi_{t+1} \in E_{t+1}] = 1$ and $\mathcal{Q}_{t+1}(\cdot, \xi_{t+1})$ is a Lipschitz-continuous function on \mathcal{X}_t for all $\xi_{t+1} \in E_{t+1}$ and $t = 1, \dots, T-1$. In particular, $\mathcal{X}_t \subseteq \text{dom } X_{t+1}(\cdot, \xi_{t+1})$, for all $\xi_{t+1} \in E_{t+1}$ and $t = 1, \dots, T-1$.*
- (b) *Considering the same assumption of the previous item and denoting by $L_t := \mathbb{E}\chi_t(\xi_{t+1}) \in \mathbb{R}_+$, for $t = 1, \dots, T-1$, it follows that $\mathcal{Q}_{t+1}(\cdot)$ is L_t -Lipschitz-continuous on \mathcal{X}_t , for $t = 1, \dots, T-1$. Moreover, the restriction of the true objective function $f(\cdot)$ to X_1 is (finite-valued) continuous and the (first-stage) optimal solution set of the true problem is nonempty.*
- (c) *Suppose that, for a.e. $\xi_{t+1} \in \text{supp}(\xi_{t+1})$, $\mathcal{X}_t \subseteq \text{dom } X_{t+1}(\cdot, \xi_{t+1})$, for all $t = 1, \dots, T-1$. If condition (Mt.4) also holds, for $t = 1, \dots, T-1$, and we generate the scenario tree following the identical conditional sampling scheme, then the SAA objective function $\hat{f}_{N_2, \dots, N_T}$ restricted to the set X_1 is (finite-valued) continuous with probability one. In particular, $\mathbb{P}[\hat{S}_{N_2, \dots, N_T} \neq \emptyset] = 1$.*

Proof: Condition (Mt.3) states that there exists $E_{t+1} \subseteq \text{supp}(\xi_{t+1})$ such that $\mathbb{P}[\xi_{t+1} \in E_{t+1}] = 1$ and for all $\xi_{t+1} \in E_{t+1}$

$$|\mathcal{Q}_{t+1}(x'_t, \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t, \xi_{t+1})| \leq \chi_t(\xi_{t+1}) \|x'_t - x_t\|, \quad (30)$$

for all $x'_t, x_t \in \mathcal{X}_t$, $t = 1, \dots, T-1$. So, for fixed $\xi_{t+1} \in E_{t+1}$, the function $\mathcal{Q}_{t+1}(\cdot, \xi_{t+1})$ is Lipschitz-continuous on \mathcal{X}_t , since $\chi_t(\xi_{t+1})$ is finite. In particular, for every $\xi_{t+1} \in E_{t+1}$, $\mathcal{Q}_{t+1}(x_t, \xi_{t+1})$ is finite for $x_t \in \mathcal{X}_t$. Hence,

$$\text{dom } X_{t+1}(\cdot, \xi_{t+1}) = \{x_t \in \mathbb{R}^{n_t} : \text{dom } X_{t+1}(\cdot, \xi_{t+1}) \neq \emptyset\} \quad (31)$$

contains \mathcal{X}_t , for $\xi_{t+1} \in E_{t+1}$, which proves item (a).

To prove item (b), let us begin by showing that $\mathcal{Q}_{t+1}(x_t)$ is finite, for every $x_t \in \mathcal{X}_t$ and $t = 1, \dots, T-1$. The case $t = 1$ follows immediately from assumption (M1), since $f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1)$. Since $\mathcal{Q}_2(x_1) = \mathbb{E}\mathcal{Q}_2(x_1, \xi_2)$, we conclude that $\mathcal{Q}_2(x_1, \xi_2)$ is finite a.e. ξ_2 . So, take $\xi_2 \in \mathbb{R}^{d_2}$ such that

$$-\infty < \mathcal{Q}_2(x_1, \xi_2) = \inf_{x_2 \in X_2(x_1, \xi_2)} \{F_2(x_2, \xi_2) + \mathcal{Q}_3(x_2)\} < \infty. \quad (32)$$

It follows that for some $\tilde{x}_2 \in X_2(x_1, \xi_2) \subseteq \mathcal{X}_2$, $\mathcal{Q}_3(\tilde{x}_2) = \mathbb{E}\mathcal{Q}_3(\tilde{x}_2, \xi_3)$ is finite. Now, taking any $x_2 \in \mathcal{X}_2$, using condition (M2.3) with \tilde{x}_2 and x_2 , and applying lemma 1 of Appendix A, we conclude that $\mathcal{Q}_3(x_2) = \mathbb{E}\mathcal{Q}_3(x_2, \xi_3)$ is finite. This shows that $\mathcal{Q}_3(x_2)$ is finite for every $x_2 \in \mathcal{X}_2$. Applying a similar reasoning forward in stages, we conclude that $\mathcal{Q}_{t+1}(\cdot)$ is finite in \mathcal{X}_t , for all $t = 2, \dots, T-1$.

Now, the L_t -Lipschitz-continuity of $\mathcal{Q}_{t+1}(\cdot)$ on \mathcal{X}_t follows immediately from (30) and lemma 1 of Appendix A. Let us show the continuity of $f|_{X_1}$. Consider an arbitrary sequence $\{x_1^k : k \in \mathbb{N}\}$ in X_1 that converges to $x_1 (\in X_1)$. Since the (immediate) stage cost functions are continuous, as is the case of $F_1(\cdot)$, we conclude that

$$f(x_1^k) = F_1(x_1^k) + \mathcal{Q}_2(x_1^k) \longrightarrow F_1(x_1) + \mathcal{Q}_2(x_1), \text{ as } k \rightarrow \infty.$$

This proves the continuity of the $f|_{X_1}$ and, in particular, that $S \neq \emptyset$, which shows item (b). We show item (c) in appendix C. \square

We have seen that under the assumed regularity conditions the true stochastic optimization problem has an optimal solution and the same also holds for the SAA problem with probability one. In order to clarify some points, let us make the following remarks about the last proposition. Firstly, we are not claiming that f is continuous on X_1 , but that its restriction to this set is⁷. Observe also that condition (M0) was implicitly used when we have considered that the cost-to-go function $\mathcal{Q}_{t+1}(\cdot)$ does not depend on the history $\xi_{[t]}$ until stage t , for $t = 1, \dots, T-1$. The same observation holds for the functions $\mathcal{Q}_{t+1}(x_t, \xi_{t+1})$ that depends on ξ_{t+1} instead of $\xi_{[t+1]}$ under (M0). For more details concerning these points, the reader should consult reference [6, Chapter 3].

In appendix B, we prove the following generalization of Theorem 1.

Theorem 3 *Consider a T -stage stochastic optimization problem that satisfies conditions (M0), (M1) and (Mt.1)-(Mt.3), for $t = 1, \dots, T-1$. Denote the stage sample sizes by N_2, \dots, N_T and $L_t := \mathbb{E}[\chi_t(\xi_{t+1})]$, $t = 1, \dots, T-1$. Moreover, suppose that the scenario-tree is constructed using the identical conditional sampling scheme. Given $\gamma > 1$ and $\epsilon > 0$, the following estimate holds*

$$\mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}_{N_2, \dots, N_T}(x_1) - f(x_1)| > \epsilon \right] \leq \sum_{t=1}^{T-1} \left(\exp\{-N_{t+1}I_{\chi_t}(\gamma L_t)\} + 2 \left[\frac{2\rho D_t \gamma L_t}{\epsilon/(T-1)} \right]^{n_t} \exp \left\{ -\frac{N_{t+1}\epsilon^2}{32\sigma_t^2(T-1)^2} \right\} \right), \quad (33)$$

where $\hat{f}_{N_2, \dots, N_T}(\cdot)$ is the SAA objective function.

Similarly to the static case, we can apply the previous theorem to obtain an upper bound for the sample

⁷For example, the indicator function of the interval $[0, 1]$ is not continuous on 0 when we consider its whole domain of definition \mathbb{R} , but its restriction to $[0, 1]$ is.

complexity $N(\epsilon, \delta, \alpha; p)$ for an instance of T -stage stochastic problem satisfying the regularity conditions. In [5, 6] it was derived a bound considering the case $T = 3$, $\delta := \epsilon/2 > 0$ and under slightly stronger regularity conditions. The conditions (Mt.3) were considered with

$$\chi_t(\xi_{t+1}) = L_t, \text{ a.e. } \xi_{t+1}, t = 1, 2. \quad (34)$$

In these references, it was also pointed out that this result could be easily extended to address the general finite case $T \geq 3$. Here, we extend this result in both directions, i.e. considering the general case $T \geq 3$ and assuming only that the random variable $\chi_t(\xi_{t+1})$, $t = 1, \dots, T-1$, has finite moment generating function in a neighborhood of zero. Moreover, we also consider the general case $\delta \in [0, \epsilon)$. The case $\delta = 0$ can be tricky in the multistage setting. It is worth noting that when we deal with the general case $T \geq 3$, it appears in the bound an additional multiplicative factor.

Theorem 4 *Consider a T -stage stochastic optimization problem that satisfies conditions (M0), (M1) and (Mt.1)-(Mt.3), for $t = 1, \dots, T-1$. Denote the stage sample sizes by N_2, \dots, N_T , $L_t := \mathbb{E}[\chi_t(\xi_{t+1})]$, $t = 1, \dots, T-1$ and let $\gamma > 1$ be arbitrary. Moreover, suppose that the scenario-tree is constructed using the identical conditional sampling scheme. For $\epsilon > 0$, $\delta \in (0, \epsilon)$ and $\alpha \in (0, 1)$, $\mathcal{N}(\epsilon, \delta, \alpha; p) \supseteq \tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p)$, where*

$$\tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p) = \left\{ (N_2, \dots, N_T) : \sum_{t=1}^{T-1} \left(\exp\{-N_{t+1} I_{\chi_t}(\gamma L_t)\} + 2 \left[\frac{4\rho\gamma D_t L_t}{(\epsilon-\delta)/(T-1)} \right]^{n_t} \exp\left\{-\frac{N_{t+1}(\epsilon-\delta)^2}{8\sigma_t^2(T-1)^2}\right\} \right) \leq \alpha \right\} \quad (35)$$

As a consequence,

$$N(\epsilon, \delta, \alpha; p) \leq \inf \left\{ \prod_{t=2}^T N_t : (N_2, \dots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p) \right\} =: \tilde{N}(\epsilon, \delta, \alpha; p). \quad (36)$$

Moreover, if the problem also satisfies conditions (Mt.4), for $t = 1, \dots, T-1$, then (35) and (36) also hold for $\delta = 0$.

Proof: Supposing only that conditions (M0), (M1) and (Mt.1)-(Mt.3) hold, for $t = 1, \dots, T-1$, by theorem 3 we obtain a lower estimate for the probability of the event

$$\left[\sup_{x_1 \in X_1} |\hat{f}_{N_2, \dots, N_T}(x_1) - f(x_1)| \leq \frac{\epsilon - \delta}{2} \right], \quad (37)$$

for $\epsilon > 0$ and $\delta \in [0, \epsilon)$. Observe that if $\delta > 0$, then

$$\left[\sup_{x_1 \in X_1} |\hat{f}_{N_2, \dots, N_T}(x_1) - f(x_1)| \leq \frac{\epsilon - \delta}{2} \right] \subseteq \left[\hat{S}_{N_2, \dots, N_T}^\delta \subseteq S^\epsilon \right] \cap \left[\hat{S}_{N_2, \dots, N_T}^\delta \neq \emptyset \right]. \quad (38)$$

In fact, since f is bounded from below on X_1 , we conclude by lemma 2 of appendix A that

$$-\infty < \inf_{x_1 \in X_1} f(x_1) - \frac{\epsilon - \delta}{2} \leq \inf_{x_1 \in X_1} \hat{f}_{N_2, \dots, N_T}(x_1), \quad (39)$$

so $[\hat{S}_{N_2, \dots, N_T}^\delta \neq \emptyset]$, for $\delta > 0$. For $\delta = 0$, it is only possible to conclude that

$$\left[\sup_{x_1 \in X_1} |\hat{f}_{N_2, \dots, N_T}(x_1) - f(x_1)| \leq \frac{\epsilon - \delta}{2} \right] \subseteq [\hat{S}_{N_2, \dots, N_T}^\delta \subseteq S^\epsilon]. \quad (40)$$

However, supposing that conditions (Mt.4) also hold, for $t = 1, \dots, T-1$, we obtain that the event $[\hat{S}_{N_2, \dots, N_T}^\delta \neq \emptyset]$ has total probability, so

$$\mathbb{P} \left([\hat{S}_{N_2, \dots, N_T}^\delta \subseteq S^\epsilon] \cap [\hat{S}_{N_2, \dots, N_T}^\delta \neq \emptyset] \right) = \mathbb{P} [\hat{S}_{N_2, \dots, N_T}^\delta \subseteq S^\epsilon] \geq \mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}_{N_2, \dots, N_T}(x_1) - f(x_1)| \leq \frac{\epsilon - \delta}{2} \right], \quad (41)$$

and condition (35) is verified. Condition (36) is immediate from (35). \square

Let us analyze with more details the upper bound $\tilde{N}(\epsilon, \delta, \alpha; p)$. Maybe it is difficult to obtain an explicit solution of the variational formula that defines $\tilde{N}(\epsilon, \delta, \alpha; p)$, however it is easy to obtain upper and lower estimates for it. Of course, an upper estimate will also be an upper bound for $N(\epsilon, \delta, \alpha; p)$. The purpose in deriving a lower estimate is twofold. Firstly, it provides an interval for $\tilde{N}(\epsilon, \delta, \alpha; p)$ jointly with the upper estimate, giving also a measure of how crude are these estimates. Secondly, it provides the best we can expect for our upper bound, providing a kind of limitation of our analysis, in the sense that, the lower the *upper bound* is, the better (or tighter).

To obtain an upper estimate for $\tilde{N}(\epsilon, \delta, \alpha; p)$, observe that $\tilde{N}(\epsilon, \delta, \alpha; p)$ contains the following set

$$\left\{ (N_2, \dots, N_T) : \begin{array}{l} 2 \left[\frac{4\rho\gamma D_t L_t}{(\epsilon - \delta)/(T-1)} \right]^{n_t} \exp \left\{ -\frac{N_{t+1}(\epsilon - \delta)^2}{8\sigma_t^2(T-1)^2} \right\} \leq \frac{\alpha}{2(T-1)} \\ \text{and} \\ \exp\{-N_{t+1}I_{\chi_t}(\gamma L_t)\} \leq \frac{\alpha}{2(T-1)}, t = 1, \dots, T-1 \end{array} \right\}. \quad (42)$$

Now, similarly to the static case, we obtain separate estimates for each N_{t+1} of the form

$$N_{t+1} \geq \frac{8\sigma_t^2(T-1)^2}{(\epsilon - \delta)^2} \left[n_t \log \left(\frac{4\rho\gamma L_t D_t (T-1)}{\epsilon - \delta} \right) + \log \left(\frac{4(T-1)}{\alpha} \right) \right] \vee \left[\frac{1}{I_{\chi_t}(\gamma L_t)} \log \left(\frac{2(T-1)}{\alpha} \right) \right]. \quad (43)$$

The product of the right-hand side of (43)'s ceiling is an upper estimate for $\tilde{N}(\epsilon, \delta, \alpha; p)$, and consequently, also for $N(\epsilon, \delta, \alpha; p)$. Comparing with the static estimate to the power of $T-1$, we see that this one has an additional factor of the form

$$(T-1)^{2(T-1)}, \quad (44)$$

that accelerates its growth with respect to the number of stages. One may think that this may be a consequence that this upper estimate is too crude. However, as we show below, the lower estimate of $\tilde{N}(\epsilon, \delta, \alpha; p)$ presents the same factor $(T-1)^{2(T-1)}$, additionally to the simple exponential behavior of the static estimate to the power of $T-1$. This implies that $\tilde{N}(\epsilon, \delta, \alpha; p)$ is also subject to this additional factor.

Indeed, observe that each term of the sum is non-negative and we can take advantage of this fact to obtain a lower estimate of $\tilde{N}(\epsilon, \delta, \alpha; p)$, since there are no cancellations in the sum. In fact, for arbitrary

$(N_2, \dots, N_T) \in \tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p)$, we must necessarily have

$$2 \left[\frac{4\rho\gamma D_t L_t}{(\epsilon - \delta)/(T-1)} \right]^{n_t} \exp \left\{ -\frac{N_{t+1}(\epsilon - \delta)^2}{8\sigma_t^2(T-1)^2} \right\} \leq \alpha, \quad (45)$$

and

$$\exp\{-N_{t+1}I_{\chi_t}(\gamma L_t)\} \leq \alpha, \quad (46)$$

for $t = 1, \dots, T-1$. From (45) and (46) we obtain the following lower estimates

$$N_{t+1} \geq \frac{8\sigma_t^2(T-1)^2}{(\epsilon - \delta)^2} \left[n_t \log \left(\frac{4\rho\gamma L_t D_t (T-1)}{\epsilon - \delta} \right) + \log \left(\frac{2}{\alpha} \right) \right] \vee \left[\frac{1}{I_{\chi_t}(\gamma L_t)} \log \left(\frac{1}{\alpha} \right) \right], \quad (47)$$

for each $t = 1, \dots, T-1$. Considering the product of these estimates, we obtain a lower estimate for $\tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p)$ that has the same additional growth factor (44) observed for the upper estimate. Moreover, expression (47) is almost the same as the upper bound for the static case (the differences being the constants inside the logarithms).

Additionally, observe that now the growth rate of $\tilde{\mathcal{N}}(\epsilon, \delta, \alpha; p)$ with respect to the difference $\epsilon - \delta > 0$ is at least of order

$$\left(\frac{O(1)\sigma^2(T-1)^2}{(\epsilon - \delta)^2} \left[n \log \left(\frac{O(1)LD(T-1)}{\epsilon - \delta} \right) \right] \right)^{T-1}, \quad (48)$$

for sufficiently small values of $\epsilon - \delta$, where the problem's parameters n, σ, L and D are the minimum of the corresponding stage parameters. Here, the statement “for sufficiently small values of $\epsilon - \delta$ ” holds when the maximum in (47) is attained by the first term for each $t = 1, \dots, T-1$.

As was done for the static case, we will extend the previous sample complexity result to a class of T -stage stochastic problems. As before, we will consider an uniformly bounded condition for the instances parameters in the class to prevent it to be infinite. We will omit the proof, since it follows directly from the discussion above.

Corollary 2 *Let \mathcal{C} be the class of all T -stage stochastic optimization problems satisfying assumptions (M0), (M1) and (Mt.1)-(Mt.4), for $t = 1, \dots, T-1$ and the uniformly bounded condition*

(UB): *there exist positive finite constants $\sigma, M, n \in \mathbb{N}, \gamma > 1$ and β such that for every instance $(p) \in \mathcal{C}$ and all $t = 1, \dots, T-1$, we have*

- i. $\sigma_t^2(p) \leq \sigma^2$,
- ii. $D_t(p) \times L_t(p) \leq M$,
- iii. $n_t(p) \leq n$,
- iv. $(0 <) \beta \leq I_{\chi_t(p)}(\gamma L_t(p))$,

where $D_t(p), L_t(p) := \mathbb{E}[\chi_t(p)(\xi)]$, $\sigma_t(p) > 0$, $t = 1, \dots, T-1$ are the problem's parameters. Then, it follows that

$$N(\epsilon, \delta, \alpha; \mathcal{C}) \leq \left(\frac{O(1)\sigma^2(T-1)^2}{(\epsilon - \delta)^2} \left[n \log \left(\frac{O(1)\gamma M(T-1)}{\epsilon - \delta} \right) + \log \left(\frac{4(T-1)}{\alpha} \right) \right] \vee \left[\frac{1}{\beta} \log \left(\frac{2(T-1)}{\alpha} \right) \right] \right)^{T-1}, \quad (49)$$

where $O(1)$'s are absolute constants.

4 Concluding Remarks and Future Work

In this paper we have defined precisely the sample complexity of the SAA method for static and multistage stochastic optimization problems. We have shown that some sample sizes estimates obtained in the literature are *upper bounds* for the sample complexity of static and multistage stochastic optimization problems. We have extended an *upper bound* estimate for T -stage stochastic problems, for finite $T \geq 3$, under relaxed regularity conditions. In a future work, we take advantage of these modifications to obtain a *lower bound* for the sample complexity of the SAA method for an appropriate class of multistage problems. Moreover, we argue that these extensions are not arbitrary mathematical generalizations of the results obtained so far in the literature. Indeed, comparing the *upper bounds* obtained for the general finite multistage case, $T \geq 3$, with the *static* or *two-stage* case, $T = 2$, we observe that, additionally to the exponentially growth behavior with respect to the number of stages, a multiplicative factor of the order of $(T - 1)^{2(T-1)}$ appears in the derived *upper bound* for the T -stage case. To the best of our knowledge, this phenomenon have not been highlighted before and shows that the *upper bound* for T -stage problems grows even faster, with respect to T , than the *upper bound* for the static case to the power of $T - 1$. In a another work, we derive a *lower bound* estimate for the sample complexity of T -stage problems and analyze its dependence with respect to T and the remaining problem and complexity parameters.

5 Acknowledgments

This work was done while this author was visiting the School of Industrial and Systems Engineering of Georgia Institute of Technology, Atlanta, GA, 30332-0205. This work was supported by CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Appendices

Appendix A - Some Technical Lemmas

Appending $-\infty$ and $+\infty$ to the set of real numbers, we obtain the set of extended real numbers $\overline{\mathbb{R}} := [-\infty, +\infty] := \mathbb{R} \cup \{-\infty, +\infty\}$. It is possible to extend the order relation and the algebraic operations “+” and “ \times ” to this set in a natural way, although some conventions must be made to contemplate the following cases

- (a) $0 \times (+\infty) = 0 \times (-\infty) = (+\infty) \times 0 = (-\infty) \times 0 = 0$,
- (b) $(+\infty) + (-\infty) = (-\infty) + (+\infty) = +\infty$.

Lemma 1 *Let Y , W and Z be random variables defined on the same probability space. Suppose that Y and Z have finite expected-value and that*

$$|Y - W| \leq Z \text{ a.s.} \tag{50}$$

Then, W has finite-expected value and $|\mathbb{E}Y - \mathbb{E}W| \leq \mathbb{E}Z$.

Proof: Observe that

$$|Y| \leq |W| + |Y - W| \leq |W| + |Z| \text{ a.s.} \quad (51)$$

This shows that Y has finite expected-value. Moreover, it follows that

$$\mathbb{E}Z \geq \mathbb{E}|Y - W| \geq |\mathbb{E}[Y - W]| = |\mathbb{E}Y - \mathbb{E}W|, \quad (52)$$

where the second inequality is just Jensen's and the equality holds since both Y and W have finite expected-value. \square

Let us remark that it is not possible to conclude that $|\mathbb{E}Y - \mathbb{E}W| \leq \mathbb{E}Z$ without supposing that at least Y or W have finite expected-value. A trivial example would be to take $Y = W$ with $\mathbb{E}Y = +\infty$. Of course, $|Y - W| = 0$, so we can take $Z = 0$, but

$$|\mathbb{E}Y - \mathbb{E}W| = |\infty - \infty| = \infty > 0 = \mathbb{E}Z. \quad (53)$$

Lemma 2 *Let $f, g : X \rightarrow \overline{\mathbb{R}}$ be arbitrary functions, where X is a nonempty set, and suppose that at least one of these functions (maybe both) is bounded from below. Then,*

$$\left| \inf_{x \in X} f(x) - \inf_{x \in X} g(x) \right| \leq \sup_{x \in X} |f(x) - g(x)|. \quad (54)$$

Proof: Suppose, w.l.o.g., that $0 \leq \inf_{x \in X} f(x) - \inf_{x \in X} g(x)$. Of course, in this case, $a := \inf_{x \in X} f(x) > -\infty$, since at most one of the functions (maybe none) is unbounded from below. For every $y \in X$, we have that $\inf_{x \in X} f(x) \leq f(y)$, so $\inf_{x \in X} f(x) - g(y) \leq f(y) - g(y)$, $\forall y \in X$. Taking the supremum on $y \in X$ and using that $\sup_{y \in X} -g(y) = -\inf_{y \in X} g(y)$ and $a > -\infty$, we conclude that

$$\inf_{x \in X} f(x) - \inf_{y \in X} g(y) = a - \inf_{y \in X} g(y) = \sup_{y \in X} \{a - g(y)\} = \sup_{y \in X} (f(y) - g(y)) \leq \sup_{x \in X} |f(x) - g(x)|,$$

and the result is proved. \square

Observe that the second equality above is true, since $a > -\infty$. In fact the previous lemma is not true, in general, without supposing that at least one of the functions is bounded from below. Consider this counterexample: $f := g := Id : \mathbb{R} \rightarrow \mathbb{R}$. Then, $\inf_{x \in \mathbb{R}} f(x) = \inf_{x \in \mathbb{R}} g(x) = \inf_{x \in \mathbb{R}} x = -\infty$, and so $\inf_{x \in \mathbb{R}} f(x) - \inf_{x \in \mathbb{R}} g(x) = (-\infty) + (+\infty) = +\infty > 0 = \sup_{x \in \mathbb{R}} |f(x) - g(x)|$.

In the next two lemmas, we derive two crude estimates of the covering theory on \mathbb{R}^n . The first one is used in the derivation of the second, which provides an upper estimate for the absolute constant $\rho > 0$ that appears repeatedly in the text.

Lemma 3 *Let be given $D > d > 0$ finite constants and $n \in \mathbb{N}$. It is possible to cover the n -dimensional closed (euclidean) ball with radius D with K n -dimensional closed (euclidean) balls with radius d , where $K \leq (2D/d + 1)^n$.*

Proof: All balls considered below are closed euclidean balls of \mathbb{R}^n . In the following procedure, every small ball has radius $d/2$ and the big ball has radius D .

Place a small ball in the space, with center belonging to the big ball. Suppose that we have placed $k \geq 1$ disjoint small balls in \mathbb{R}^n , where the center of each one belongs to the big ball (we are not assuming that

each small ball is contained in the big one, only its center). If it is possible to place another small ball in the space satisfying both conditions above, proceed and make $k = k + 1$; otherwise, stop. Of course, by volume considerations (see below), the preceding algorithm stops after a finite number of iterations, say K . After the termination, we are in a configuration that for every point of the big ball, exists a small ball whose distance to this point is $\leq d/2$. Indeed, if there is a point in the big ball whose distance for each small ball is greater than $d/2$, then we can place an additional small ball with center at this point that is in the big ball. Moreover, this small ball is disjoint of all others, which contradicts the fact that the algorithm have stopped.

Now, duplicate the radius of each one of the K small balls, keeping the center the same. I claim that these K balls with radius d cover the big ball. In fact, consider an arbitrary point of the big ball. We know that exists one small ball whose distance to this point is $\leq d/2$. By the triangular inequality, we conclude that this point belongs to the enlarged ball with radius d .

Finally, observe that each one of the small balls (with radius $d/2$) is contained in a ball with radius $D + d/2$. Since the small balls are disjoint, we conclude that

$$K \cdot \text{Vol}(\mathbb{B}_n)(d/2)^n \leq (D + d/2)^n \text{Vol}(\mathbb{B}_n),$$

where $\text{Vol}(\mathbb{B}_n)$ is the volume of the unitary euclidean ball of \mathbb{R}^n and we have used that the volume of the ball with radius $r > 0$ is equal to $r^n \text{Vol}(\mathbb{B}_n)$. We obtain that $K \leq (2D/d + 1)^n$, and the lemma is proved. \square

Lemma 4 *Let $X \subseteq \mathbb{R}^n$ has finite diameter $D > 0$. For every $d \in (0, D)$, exists a d -net of X with $K \leq (5D/d)^n$ elements.*

Proof: Let B be a ball with radius D such that $B \supseteq X$. By the previous lemma, it is possible to cover B with $\{B_i : i = 1, \dots, K\}$, where each B_i has radius $d/2$ and $K \leq (4D/d + 1)^n \leq (5D/d)^n$. Consider the following procedure. For each $i = 1, \dots, K$, if $B_i \cap X \neq \emptyset$, select any of its elements, say x_i . Denote the set of selected elements by N . Of course, $N \subseteq X$ and $\#N \leq (5D/d)^n$. I claim that N is a d -net of X . In fact, let $x \in X$ be arbitrary. Since $X \subseteq B \subseteq \bigcup_{i=1}^K B_i$, exists i such that $x \in B_i$. So, there is $x_i \in N$ that belongs to B_i . Since B_i has radius $d/2$, the triangular inequality guarantees that $\|x_i - x\| \leq d$, and the lemma is proved. \square

Appendix B - Proof of Theorem 3

Let us simplify the notation by dropping the subscripts on N_2, \dots, N_T for the SAA objective function and stage cost-to-go functions. Since we adopt the identical conditional sampling scheme, the empirical probability distribution embedded in the scenario tree is also stagewise independent. Then, the SAA stage expected *optimal-value* functions (or *cost-to-go* functions) are path-independent and we can write $\hat{Q}_{t+1}(x_t)$ instead of $\hat{Q}_{t+1}(x_t, \xi_{[t]})$.

The idea of the proof is to obtain an upper bound for

$$\sup_{x_1 \in X_1} |\hat{f}(x_1) - f(x_1)|, \tag{55}$$

with probability one, in terms of a sum of random variables, whose terms we have a control of their tail decay with respect to the sample sizes N_{t+1} , for $t = 1, \dots, T - 1$. In fact, we will show that we can take each term

of the sum as

$$Z_t := \sup_{x_t \in X_t} \left| \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} \left[\mathcal{Q}_{t+1}(x_t, \xi_{t+1}^j) - \mathcal{Q}_{t+1}(x_t) \right] \right|, \quad t = 1, \dots, T-1. \quad (56)$$

In the final step, we will apply theorem 1 for each Z_t obtaining an upper bound for the probability of each Z_t be greater or equal than $\epsilon/(T-1)$ as a function depending on the specified constants such as: D_t , L_t and σ_t^2 , and the sample size N_{t+1} .

First of all, observe that by proposition (3.a), with probability one, $\mathcal{Q}_{t+1}(\cdot, \xi_{t+1}^j)$ is finite-valued on \mathcal{X}_t , for $j = 1, \dots, N_{t+1}$ and $t = 1, \dots, T-1$. So, let us assume as given a sample realization $\mathfrak{S}_{N_2, \dots, N_T}$ such that⁸

$$-\infty < \mathcal{Q}_t(x_{t-1}, \xi_t^j) = \inf_{x_t \in X_t(x_{t-1}, \xi_t^j)} \left\{ F_t(x_t, \xi_t^j) + \mathcal{Q}_{t+1}(x_t) \right\} < +\infty, \quad (57)$$

for all $x_{t-1} \in X_{t-1}$, $j = 1, \dots, N_t$ and $t = 2, \dots, T$.

Let $x_1 \in X_1$ be arbitrary, we have that

$$f(x_1) = F_1(x_1) + \mathcal{Q}_2(x_1) \quad \text{and} \quad \hat{f}(x_1) = F_1(x_1) + \hat{\mathcal{Q}}_2(x_1), \quad (58)$$

so $|\hat{f}(x_1) - f(x_1)| = |\hat{\mathcal{Q}}_2(x_1) - \mathcal{Q}_2(x_1)|$. Let us proof the following claim.

Claim: For each $t = 2, \dots, T-1$, with probability one,

$$\sup_{x_{t-1} \in X_{t-1}} \left| \hat{\mathcal{Q}}_t(x_{t-1}) - \mathcal{Q}_t(x_{t-1}) \right| \leq \sup_{x_{t-1} \in X_{t-1}} \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \left[\mathcal{Q}_t(x_{t-1}, \xi_t^j) - \mathcal{Q}_t(x_{t-1}) \right] \right| + \sup_{x_t \in X_t} \left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right| \quad (59)$$

Let $x_{t-1} \in X_{t-1}$ be arbitrary. Then,

$$\begin{aligned} \left| \hat{\mathcal{Q}}_t(x_{t-1}) - \mathcal{Q}_t(x_{t-1}) \right| &= \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \left[\hat{\mathcal{Q}}_t(x_{t-1}, \xi_t^j) - \mathcal{Q}_t(x_{t-1}) \right] \right| \leq \\ &\left| \frac{1}{N_t} \sum_{j=1}^{N_t} \left[\mathcal{Q}_t(x_{t-1}, \xi_t^j) - \mathcal{Q}_t(x_{t-1}) \right] \right| + \frac{1}{N_t} \sum_{j=1}^{N_t} \left| \hat{\mathcal{Q}}_t(x_{t-1}, \xi_t^j) - \mathcal{Q}_t(x_{t-1}, \xi_t^j) \right|, \end{aligned} \quad (60)$$

where we have applied just the triangle inequality above. Now, observe that for each $j = 1, \dots, N_t$,

$$\begin{aligned} \left| \hat{\mathcal{Q}}_t(x_{t-1}, \xi_t^j) - \mathcal{Q}_t(x_{t-1}, \xi_t^j) \right| &= \\ \left| \inf_{x_t \in X_t(x_{t-1}, \xi_t^j)} \left[F_t(x_t, \xi_t^j) + \hat{\mathcal{Q}}_{t+1}(x_t) \right] - \inf_{x_t \in X_t(x_{t-1}, \xi_t^j)} \left[F_t(x_t, \xi_t^j) + \mathcal{Q}_{t+1}(x_t) \right] \right| & \\ \leq \sup_{x_t \in X_t(x_{t-1}, \xi_t^j)} \left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right| &\leq \sup_{x_t \in X_t} \left| \hat{\mathcal{Q}}_{t+1}(x_t) - \mathcal{Q}_{t+1}(x_t) \right|, \end{aligned} \quad (61)$$

where the first inequality follows from lemma 2, since by (57) the infimum is greater than $-\infty$ and $X_t(x_{t-1}, \xi_t^j) \neq \emptyset$, and the last inequality holds since $X_t(x_{t-1}, \xi_t^j) \subseteq X_t$, for all $x_{t-1} \in X_{t-1}$ and $j = 1, \dots, N_t$. Observe that

⁸We must be careful in each step that we make, since we are potentially dealing with arithmetic operations involving extended real numbers.

the last term of (61) does not depend on $j = 1, \dots, N_t$. So, taking the supremum on $x_{t-1} \in X_{t-1}$ in equation (60), we obtain (59) and the claim is proved.

Observing that for the last stage we have $\hat{\mathcal{Q}}_T(x_{T-1}, \xi_T^j) = \mathcal{Q}_T(x_{T-1}, \xi_T^j)$, for all $x_{T-1} \in X_{T-1}$ and $j = 1, \dots, N_T$, we conclude that

$$\sup_{x_{T-1} \in X_{T-1}} \left| \hat{\mathcal{Q}}_T(x_{T-1}) - \mathcal{Q}_T(x_{T-1}) \right| = \sup_{x_{T-1} \in X_{T-1}} \left| \frac{1}{N_T} \sum_{j=1}^{N_T} \left[\mathcal{Q}_T(x_{T-1}, \xi_T^j) - \mathcal{Q}_T(x_{T-1}) \right] \right|. \quad (62)$$

Summing up, he have just shown that, with probability one,

$$\sup_{x_1 \in X_1} |\hat{f}(x_1) - f(x_1)| \leq \sum_{t=1}^{T-1} Z_t, \quad (63)$$

where $Z_t := \sup_{x_t \in X_t} \left| \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} \left(\mathcal{Q}_{t+1}(x_t, \xi_{t+1}^j) - \mathbb{E} \left[\mathcal{Q}_{t+1}(x_t, \xi_{t+1}^j) \right] \right) \right|$. Now, by hypothesis (Mt.1)-(Mt.3) and using also that $\mathcal{Q}_{t+1}(\cdot)$ is finite on \mathcal{X}_t , we can apply theorem 1 for each Z_t , obtaining the upper bound

$$\mathbb{P} \left[Z_t > \frac{\epsilon}{T-1} \right] \leq \left(\exp\{-N_{t+1} I_{\mathcal{X}_t}(\gamma L_t)\} + 2 \left[\frac{4\rho D_t \gamma L_t}{\epsilon/(T-1)} \right]^{n_t} \exp \left\{ -\frac{N_{t+1} \epsilon^2}{32\sigma_t^2 (T-1)^2} \right\} \right), \quad (64)$$

for each $t = 1, \dots, T-1$. Since

$$\left[\sup_{x_1 \in X_1} |\hat{f}(x_1) - f(x_1)| > \epsilon \right] \subseteq \bigcup_{t=1}^{T-1} \left[Z_t > \frac{\epsilon}{T-1} \right],$$

we conclude that

$$\mathbb{P} \left[\sup_{x_1 \in X_1} |\hat{f}(x_1) - f(x_1)| > \epsilon \right] \leq \sum_{t=1}^{T-1} \mathbb{P} \left[Z_t > \frac{\epsilon}{T-1} \right], \quad (65)$$

and theorem 3 is proved. \square

Appendix C - A glimpse on Set-Valued Analysis and a Proof of Proposition (3.c)

In this section we make a short revision of set-valued analysis, state a result of parametric optimization known as the Berge's Maximum Theorem and prove proposition (3.c). For a more detailed discussion about these topics, the reader should consult [1, 2, 4].

Let (X, d) be a metric space. We begin by reviewing two concepts of set convergence in metric spaces.

Definition 5 Let $\{A_k : k \in \mathbb{N}\}$ be a sequence of subsets of X , i.e. $A_k \subseteq X$ for all $k \in \mathbb{N}$. We consider the following definitions of the interior and exterior limits of a sequence of sets

$$\liminf_{k \in \mathbb{N}} A_k := \left\{ x \in X : \limsup_{k \in \mathbb{N}} d(x, A_k) = 0 \right\} \quad \text{and} \quad \limsup_{k \in \mathbb{N}} A_k := \left\{ x \in X : \liminf_{k \in \mathbb{N}} d(x, A_k) = 0 \right\}, \quad (66)$$

respectively.

It is easily seen that $\liminf_{k \in \mathbb{N}} A_k \subseteq \limsup_{k \in \mathbb{N}} A_k$ always. When both sets coincide, say $A := \liminf_{k \in \mathbb{N}} A_k = \limsup_{k \in \mathbb{N}} A_k$

$\limint_{k \in \mathbb{N}} A_k$, we say that the sequence $\{A_k : k \in \mathbb{N}\}$ converges to A . Now, we define the notions of inner-semicontinuity (I.S.C.) and outer-semicontinuity (O.S.C.) of a multifunction.

Definition 6 Let $C : X \rightrightarrows Y$ be a multifunction (also denoted a point-to-set mapping) between the metric spaces X and Y . Let $x \in X$ be arbitrary.

(a) We say that C is O.S.C. at x if whenever a sequence $\{x_k : k \in \mathbb{N}\} \subseteq X$ converges to x , then

$$\limext_{k \in \mathbb{N}} C(x_k) \subseteq C(x).$$

(b) We say that C is I.S.C. at x if whenever a sequence $\{x_k : k \in \mathbb{N}\} \subseteq X$ converges to x , then

$$C(x) \subseteq \limint_{k \in \mathbb{N}} C(x_k).$$

We say that C is continuous at $x \in X$ if it is I.S.C. and O.S.C. at x . We also consider the restriction of the multifunction C to an arbitrary subset Z of X in the same way that it is considered for functions. When we say that C restricted to Z is I.S.C. or O.S.C. at $x \in Z$, we take only sequences converging to x that are on Z . Finally, we define the domain of C by $\text{dom}(C) := \{x \in X : C(x) \neq \emptyset\}$ that is a subset of its domain of definition X . The following theorem establishes sufficient conditions for the continuity of a parametric optimization problem and it is known as the Berge's Maximum Theorem⁹.

Proposition 4 Let $g : X \times \Theta \rightarrow \mathbb{R}$ be a (jointly) continuous function on $X \times \Theta$ and $C : \Theta \rightrightarrows X$ be a multifunction. Consider the optimal-value function depending on the parameter $\theta \in \Theta$

$$h(\theta) := \inf_{x \in C(\theta)} g(x, \theta),$$

and the corresponding optimal-value set

$$S(\theta) := \operatorname{argmin}_{x \in C(\theta)} g(x, \theta).$$

Suppose that C is a compact-valued multifunction and that exists a neighborhood V of $\bar{\theta} \in \text{dom}(C)$ such that $C(V)$ is a compact metric space. If C is continuous at $\bar{\theta}$, then h is continuous at $\bar{\theta}$ and S is O.S.C. at $\bar{\theta}$.

Proof: Firstly, observe that $\bar{\theta}$ must belongs to $\text{int}(\text{dom}(C))$. Suppose the contrary by absurd, that is, $\bar{\theta} \notin \text{int}(\text{dom}(C))$, then exists $\{\theta_k : k \in \mathbb{N}\} \subseteq \Theta \setminus \text{dom}(C)$ such that $\lim_k \theta_k = \bar{\theta}$. Now, observe that $C(\theta_k) = \emptyset$, for all $k \in \mathbb{N}$, so $\limint_k C(\theta_k) = \emptyset \supseteq C(\bar{\theta})$, which contradicts that $\bar{\theta} \in \text{dom}(C)$.

Now, let $\{\theta_k : k \in \mathbb{N}\}$ be any sequence in Θ converging to $\bar{\theta}$. To prove that $h(\bar{\theta}) = \lim_k h(\theta_k)$, it is sufficient to show that for every subsequence $\{\theta_k : k \in N'\}$, where $N' \subseteq \mathbb{N}$, exists a subsubsequence $\{\theta_k : k \in N''\}$, $N'' \subseteq N'$, such that $h(\bar{\theta}) = \lim_{k \in N''} h(\theta_k)$. Observe that exists $K \in \mathbb{N}$ such that $\theta_k \in V \cap \text{dom}(C)$, for $k \geq K$. Consider any selection $x_k \in S(\theta_k) \subseteq C(\theta_k)$, for $k \geq K$ and $k \in N'$. Since $x_k \in C(V)$, for all $k \geq K$, and this set is compact, we conclude that exist a subsequence $\{x_k : k \in N''\}$, where $N'' \subseteq N'$, such that $x_k \rightarrow \bar{x} \in C(V) \subseteq X$, as $k \in N''$ goes to $+\infty$. Of course, $\bar{x} \in \limext_k C(\theta_k) \subseteq C(\bar{\theta})$, because C is O.S.C. at $\bar{\theta}$. Moreover, let $x \in C(\bar{\theta})$ be arbitrary. Since C is I.S.C. at $\bar{\theta}$, we conclude that for k sufficiently big, exists

⁹Here, we consider a minimization version, but the proof is identical for the maximization.

$y_k \in C(\theta_k)$ such that $y_k \rightarrow x$, as $k \rightarrow +\infty$. Summing up, we obtain that

$$g(x_k, \theta_k) \leq g(y_k, \theta_k), \text{ for all } k \in N'' \text{ and } k \geq K.$$

Letting $k \rightarrow +\infty$ above and using the (jointly) continuity of g , we obtain that $g(\bar{x}, \bar{\theta}) \leq g(x, \bar{\theta})$, for all $x \in C(\bar{\theta})$. This shows that $\bar{x} \in S(\bar{\theta})$ and that $h(\bar{\theta}) = \lim_{k \in N''} h(\theta_k)$, proving the continuity of h at $\bar{\theta}$.

Finally, given any sequence in Θ satisfying $\theta_k \rightarrow \bar{\theta}$, as $k \rightarrow +\infty$, let us show that $\limext_k S(\theta_k) \subseteq S(\bar{\theta})$. Take $\hat{x} \in \limext_k S(\theta_k)$ arbitrary. We know that exists a subsequence $\{x_{k_j} : j \in \mathbb{N}\}$ such that $x_{k_j} \rightarrow \hat{x}$, as $j \rightarrow +\infty$, where $x_k \in S(\theta_k)$, for all $k \in \mathbb{N}$. Following the previous reasoning, it is immediate to verify that $g(\hat{x}, \bar{\theta}) \leq g(x, \bar{\theta})$, for all $x \in C(\bar{\theta})$, i.e. $\hat{x} \in S(\bar{\theta})$. \square

The following lemma will be useful.

Lemma 5 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be (nonempty) compact sets and consider on these spaces the metric given by any norm on these euclidean spaces. If $C : X \rightrightarrows Y$ is a continuous multifunction, then C is compact-valued, $C(X) := \cup_{x \in X} C(x) \subseteq \mathbb{R}^m$ is compact and $\text{dom}(C)$ is open at X and compact.*

Proof: We consider the case $\text{dom}(C) \neq \emptyset$, since the result is trivially true in this other (uninteresting) case. It is well known that C is O.S.C. at X if, and only if, $\text{gph}(C) \subseteq X \times Y$ is closed on $X \times Y$. Since X and Y are compact, then $X \times Y$ is compact and also $\text{gph}(C)$. Denoting by $\pi_x(x, y) := x$ and $\pi_y(x, y) := y$ the projections in the x -variable and y -variable, respectively, it follows that $\text{dom}(C) = \pi_x(\text{gph}(C))$ and $C(X) = \pi_y(\text{gph}(C))$ are compact sets, since $\text{gph}(C)$ is compact. In particular, $C(x) \subseteq \mathbb{R}^m$ is compact, for all $x \in X$. To show that $\text{dom}(C)$ is open on X , we argue by contradiction. If exists $x \in \text{dom}(C) \setminus \text{int}_X(\text{dom}(C))$, then there is a sequence $\{x^k : k \in \mathbb{N}\}$ in $\text{dom}(C) \setminus \text{int}_X(\text{dom}(C))$ such that $x^k \rightarrow x$, as $k \rightarrow +\infty$. Since C is I.S.C. at x , we obtain that $C(x) \subseteq \limint_k C(x^k) = \emptyset$, which contradicts the fact that $x \in \text{dom}(C)$. \square

We are ready to prove proposition (3.c).

Proof of Proposition (3.c): In this proof, when we write $X_{t+1}(\cdot, \xi)$, for any $\xi_{t+1} \in \mathbb{R}^{d_{t+1}}$, we are considering the constraint multifunction restricted to the set \mathcal{X}_t , that is:

$$x_t \in \mathcal{X}_t \mapsto X_{t+1}(x_t, \xi_{t+1}), \tag{67}$$

where \mathcal{X}_t satisfies condition (Mt.1), $t = 1, \dots, T-1$. As a consequence, $\mathcal{X}_t \supseteq \text{dom} X_{t+1}(\cdot, \xi_{t+1}) := \{x_t \in \mathcal{X}_t : X_{t+1}(x_t, \xi_{t+1}) \neq \emptyset\}$. By the hypothesis considered in item (c), for $t = 1, \dots, T-1$, there exist $E_{t+1} \subseteq \text{supp}(\xi_{t+1}) \supseteq F_{t+1}$ such that $\mathbb{P}[\xi_{t+1} \in E_{t+1}] = 1 = \mathbb{P}[\xi_{t+1} \in F_{t+1}]$ and

- I. $\text{dom} X_{t+1}(\cdot, \xi_{t+1}) = \mathcal{X}_t$, for all $\xi_{t+1} \in E_{t+1}$,
- II. $X_{t+1}(\cdot, \xi_{t+1})$ is continuous, for all $\xi_{t+1} \in F_{t+1}$.

Following the *identical conditional sampling scheme*, the random sample $\mathfrak{S}_{N_2, \dots, N_T} := \{\xi_t^{j_t} : t = 2, \dots, T, j_t = 1, \dots, N_t\}$ is a family of independent random vectors where each $\xi_t^{j_t}$ is identically distributed as ξ_t , for $j_t = 1, \dots, N_t$ and $t = 2, \dots, T$. Observe that the event

$$\mathcal{E} := \bigcap_{t=2}^T \bigcap_{j_t=1}^{N_t} \left(\left[\xi_t^{j_t} \in E_t \right] \cap \left[\xi_t^{j_t} \in F_t \right] \right), \tag{68}$$

has probability one, since it is a finite (in particular, enumerable) intersection of events of total probability. Summing up, with probability one, the following conditions hold

- i. $\text{dom } X_{t+1}(\cdot, \xi_{t+1}^j) = \mathcal{X}_t$, for all $j = 1, \dots, N_{t+1}$ and $t = 1, \dots, T-1$,
- ii. $X_{t+1}(\cdot, \xi_{t+1}^j)$ is continuous, for all $j = 1, \dots, N_{t+1}$ and $t = 1, \dots, T-1$,

Now, we show that if a sample realization satisfies conditions (i.) and (ii.), then the SAA objective function $\hat{f}_{N_2, \dots, N_T} : X_1 \rightarrow \mathbb{R}$ is (finite) continuous. We begin arguing backwards in stages. For $t = T$ and $j = 1, \dots, N_T$, the T -stage optimal-value function of the SAA problem must satisfy the dynamic programming equations

$$\hat{Q}_T^j(x) = \inf_{x_T \in X_T(x, \xi_T^j)} F_T(x_T, \xi_T^j), \text{ for all } x \in \mathcal{X}_{T-1}. \quad (69)$$

We will show that $\hat{Q}_T^j : \mathcal{X}_{T-1} \rightarrow \mathbb{R}$ is (finite-valued) continuous. This follows from proposition 4 by considering $X := \mathcal{X}_T$, $\Theta := \mathcal{X}_{T-1}$, $C := X_T(\cdot, \xi_T^j)$ and $g(x_T, x_{T-1}) := F_T(x_T, \xi_T^j)$. We just need to verify that the proposition's hypotheses hold. In fact, since \mathcal{X}_{T-1} and \mathcal{X}_T are compact and $X_T(\cdot, \xi_T^j)$ is continuous, we conclude by lemma 5 that $X_T(\cdot, \xi_T^j)$ is compact-valued and $X_T(\mathcal{X}_{T-1}, \xi_T^j)$ is compact. This proves the continuity of \hat{Q}_T^j , for every $j = 1, \dots, N_T$, and so the continuity of the T^{th} -stage empirical *cost-to-go* function

$$\hat{Q}_T(x) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{Q}_T^j(x), \text{ for all } x \in \mathcal{X}_{T-1}, \quad (70)$$

Applying the same reasoning, it is easy to verify that each $(T-1)^{\text{th}}$ -stage optimal-value function $\hat{Q}_{T-1}^j : \mathcal{X}_{T-2} \rightarrow \mathbb{R}$, that satisfies the following dynamic programming equation

$$\hat{Q}_{T-1}^j(x) = \inf_{x_{T-1} \in X_{T-1}(x, \xi_{T-1}^j)} F_{T-1}(x_{T-1}, \xi_{T-1}^j) + \hat{Q}_T(x_{T-1}), \text{ for all } x \in \mathcal{X}_{T-2}, \quad (71)$$

is (finite-valued) continuous, for $j = 1, \dots, N_{T-1}$. So, we conclude that the empirical $(T-1)^{\text{th}}$ -stage *cost-to-go*

$$\hat{Q}_{T-1}(x) := \frac{1}{N_{T-1}} \sum_{j=1}^{N_{T-1}} \hat{Q}_{T-1}^j(x), \text{ for all } x \in \mathcal{X}_{T-2} \quad (72)$$

is also continuous. Repeating the argument, we conclude that $\hat{Q}_2 : \mathcal{X}_1 \rightarrow \mathbb{R}$ is (finite-valued) continuous, and so is $\hat{f}_{N_2, \dots, N_T}$. Since $X_1 \subseteq \mathcal{X}_1$ is (nonempty) compact, the (first-stage) solution set of SAA problem is nonempty. So, we have just proved that $\mathbb{P} \left[\hat{S}_{N_2, \dots, N_T} \neq \emptyset \right] = 1$. \square

References

- [1] C. Berge, Topological spaces: including a treatment of multi-valued functions, vector spaces and convexity. Oliver and Boyd Ltd, 1963.
- [2] R.S. Burachik and A.N. Iusem, Set-valued mappings and enlargements of monotone operators. Springer Optimization and Its Applications (Book 8). Springer, 2008.
- [3] A.J. Kleywegt, A. Shapiro, T. Homem-De-Mello, The sample average approximation method for stochastic discrete optimization, SIAM Journal on Optimization 12 (2001) 479-502.

- [4] R.T. Rockafellar and R.J.-B. Wets. Variational Analysis. Springer, Berlin, 3rd printing, 2009.
- [5] A. Shapiro, On complexity of multistage stochastic programs, Operations Research Letters 34 (2006) 1-8.
- [6] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia, 2014.
- [7] A. Shapiro and A. Nemirovski, On complexity of stochastic programming problems. In V. Jeyakumar and A.M. Rubinov, editors, Continuous Optimization: Current Trends and Applications, Springer-Verlag, Berlin, 2005, 111-144.