Noname manuscript No.
(will be inserted by the editor)

A note on sample complexity of multistage stochastic programs

M.M.C.R. Reaiche

Received: date / Accepted: date

Abstract We derive a *lower bound* for the *sample complexity* of the Sample Average Approximation method for a certain class of multistage stochastic optimization problems. In previous works, *upper bounds* for such problems were derived. We show that the dependence of the *lower bound* with respect to the complexity parameters and the problem's data are comparable to the upper bound's estimates. Like previous results, our *lower bound* presents an additional multiplicative factor showing that it is unavoidable for certain stochastic problems.

Keywords Stochastic programming · Monte Carlo sampling · Sample average method · Complexity

1 Introduction

Consider the following T-stage stochastic programming problem represented in the nested form

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

driven by the random data process $\xi_1,...,\xi_T$. Here, $x_t \in \mathbb{R}^{n_t}, t = 1,...,T$, are decisions variables, $F_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \to \mathbb{R}$ are continuous functions and $X_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}$, $t = 2,...,T$, are measurable multifunctions. The (continuous) function $F_1 : \mathbb{R}^{n_1} \to \mathbb{R}$, the (nonempty) closed set X_1 and the vector ξ_1 are deterministic. Moreover, $\xi_{[t]} := (\xi_1,\dots,\xi_t)$ denotes the history (information) available until stage t by the decision maker.

If the (conditional) distribution of ξ_t (given $\xi_{[t-1]}$) is continuous, problem (1) cannot be addressed directly, except for some trivial cases. In fact, the (conditional) expected value operators are multidimensional integrals on $\mathbb{R}^{d_t}$, that are typically impossible to evaluate with high accuracy even for moderate values of the dimension.

Hence, one usually makes a discretization of the random data of problem (1) building a scenario tree. A classical idea is to construct the tree via Monte Carlo conditional sampling techniques. Given the scenario tree, one solves the SAA problem, that is, problem (1) with the discrete random data. This is the basic idea of the SAA method.

In general, even if we solve the SAA problem exactly, its first-stage optimal decision will not be optimal for the true problem. So, there exists an *error* that comes from the fact that we are approximating the true stochastic process. Suppose that the true stochastic problem has an optimal solution. One can investigate sufficient conditions on the stage sample sizes $N_2,\dots,N_T$ in order to guarantee that the following conditions happen (jointly) with probability at least $1-\alpha$: (i) any first-stage δ-optimal solution of the SAA problem is a first-stage ϵ-optimal solution of the true problem, and (ii) the set of first-stage δ-optimal solutions of the SAA problem is nonempty; where $\epsilon > 0$, $\delta \in [0,\epsilon)$, and $\alpha \in (0,1)$ are specified parameters that we denote by complexity parameters. Let us point out that this notion of complexity (with condition (ii) being implicitly assumed) was proposed and studied in [3, 5, 7].

M.M.C.R. Reaiche
Brazilian Development Bank and Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina, 110, Jardim Botânico, Rio de Janeiro, RJ, 22460-320, Brazil
Skype: marcusreaiche
E-mail: mmcr@impa.br

In [4], it was given an explicit definition of the sample complexity of SAA method for *instances* and *classes* of T-stage stochastic optimization problems. In the same reference, it was argued that estimates of the sample sizes derived in [7] and [5] are *upper bounds* estimates for the *sample complexity* of *static* and *multistage* problems, respectively, satisfying some reasonable regularity conditions. In [4] it was obtained an explicit *upper bound* estimate for the complexity of T-stage problems under relaxed regularity conditions. We will see later that it was important to relax these conditions in order to make a fair comparison between the *upper* and *lower bounds* estimates of the sample complexity.

In section 2, we state the definition of sample complexity for T-stage stochastic problems and some extensions on the complexity's *upper bounds* obtained in [4]. In section 3, we present a family of T-stage convex stochastic optimization problems where it is possible to derive a *lower bound* for the sample complexity of each one of these problems. We apply this result to derive our *lower bound* for the sample complexity of a family of *convex* T-stage problems. In section 4, we compare our *lower bound* with the one derived for multistage financial optimization problems through no-arbitrage reasoning arguments. We also indicate one possible way to extend our results to the class of *linear* multistage optimization problems. In section 5, we make our final remarks. This section is followed by a technical appendix.

2 Definition of the sample complexity and its upper bound

We follow closely reference [4] where the respective definitions were stated. Consider a scenario tree with T-stages possessing the following node structure: every t^{th}-stage node has N_{t+1} successors nodes in stage $t+1$, for $t = 1, ..., T-1$. Under this assumption, the total number of scenarios in the tree is equal to

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

We denote the sets of (first-stage) ϵ-optimal solutions, respectively, of the true and the SAA problems as

$$S^{\epsilon} := \{x_1 \in X_1 : f(x_1) \leq v^* + \epsilon\} \tag{2}$$

and

$$\hat{S}^{\epsilon}_{N_2,\ldots,N_T} := \{x_1 \in X_1 : \hat{f}(x_1) \leq \hat{v}^* + \epsilon\}, \tag{3}$$

for $\epsilon \geq 0$. The quantities v^* and $\hat{v}^*$ are the optimal-values of the true and the SAA problems, respectively. Observe that $\hat{v}^*$ and $\hat{S}^{\epsilon}_{N_2,\ldots,N_T}$ depend on the sample realization.

Definition 1 (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 set of *viable samples sizes* $\mathcal{N}(\epsilon, \delta, \alpha; p)$ as

$$\left\{ (M_2, ..., M_T) : \begin{array}{l} \forall\, (N_2, ..., N_T) \geq (M_2, ..., M_T)\,, \\ \mathbb{P}\,(G \cap H) \geq 1 - \alpha \end{array} \right\},$$

where

$$G := \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \subseteq S^{\epsilon}\right] \text{ and } H := \left[\hat{S}^{\delta}_{N_2,\ldots,N_T} \neq \emptyset\right].$$

The sample complexity of (p) is defined as

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

Definition 2 (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 [4], *upper bounds* estimates of the sample complexity of T-stage problems were derived considering the *identical conditional sampling scheme* under the following regularity conditions:

(M0) The random data is stagewise independent.

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

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

(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 $\xi_t \in \text{supp}(\xi_t)$. Here, $\text{supp}(\xi_t)$ is the support of the random vector ξ_t.

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

$$\begin{aligned} M_{t,x}(s) &:= \mathbb{E}\left[\exp\left(s(\mathcal{Q}_{t+1}(x, \xi_{t+1}) - \mathcal{Q}_{t+1}(x))\right]\right. \\ &\leq \exp\left(\sigma_t^2 s^2/2\right), \ \forall s \in \mathbb{R}. \end{aligned} \tag{4}$$

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

$$\left|\mathcal{Q}_{t+1}(x_t', \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t, \xi_{t+1})\right| \leq \chi_t(\xi_{t+1})\, ||x_t' - x_t|| \tag{5}$$

holds, for all $x_t', x_t \in \mathcal{X}_t$. Moreover, its moment generating function $M_{\chi_t}(s)$ is finite-valued 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 the first-stage feasible set X_1 (e.g. $\mathcal{X}_1 = X_1$). The functions $\mathcal{Q}_{t+1} : \mathbb{R}^{n_t} \times \mathbb{R}^{d_{t+1}} \to \bar{\mathbb{R}}$, $t = 1, \ldots, T-1$, are the stage optimal-value functions and they satisfy the following dynamic programming equations

$$\mathcal{Q}_{t+1}(x_t, \xi_{t+1}) = \inf \left\{ \begin{array}{r} F_{t+1}(x_{t+1}, \xi_{t+1}) + \mathcal{Q}_{t+2}(x_{t+1}) \\ : x_{t+1} \in X_{t+1}(x_t, \xi_{t+1}) \end{array} \right\}, \tag{6}$$

where $\mathcal{Q}_{t+1}(x_t) := \mathbb{E}\left[\mathcal{Q}_{t+1}(x_t, \xi_{t+1})\right]$, for $t = 1, \ldots, T-1$, and $\mathcal{Q}_{T+1}(x_T) \equiv 0$. These functions do not depend on the history until stage t by the stagewise independence hypothesis (condition (M0)). For more details concerning this point, the reader should consult reference [6, Chapter 3].

Let us recall the *identical conditional sampling scheme*. Firstly, we generate independent observations of the stage random vectors, say

$$\mathfrak{S}_{N_2,\ldots,N_T} := \left\{ \xi_t^j : t = 2, \ldots, T,\ j = 1, \ldots, N_t \right\}, \tag{7}$$

where N_t is the number of copies (sample size) of the random vector ξ_t, for $t = 2, \ldots, T$. Given the sample realization, we consider the tree with the following set of scenarios or paths

$$\left\{ \left(\xi_1, \xi_2^{j_2}, \ldots, \xi_T^{j_T} \right) : 1 \le j_t \le N_t,\ t = 2, \ldots, T \right\}. \tag{8}$$

Moreover, we consider the empirical probability measure on the tree

$$\hat{\mathbb{P}}\left[\xi_2 = \xi_2^{j_2}, \ldots, \xi_T = \xi_T^{j_T}\right] = \prod_{t=2}^{T} \frac{\#\{1 \le i \le N_t : \xi_t^i = \xi_t^{j_t}\}}{N_t}, \tag{9}$$

for $1 \le j_t \le N_t$ and $t = 2, \ldots, T$. In this scheme, the empirical distribution is also stagewise independent.

The following result summarizes theorem 4 of [4] and the discussion that follows it. To the best of our knowledge, this kind of result, that concerns an estimation of an upper bound of the sample complexity of multistage stochastic optimization problems, was first derived on [5].

Proposition 1 *Consider a T-stage stochastic optimization problem that satisfies conditions (M0), (M1) and (Mt.1)-(Mt.3), for $t = 1, \ldots, T-1$. Let $N_2, \ldots, N_T$ be the sample sizes, $L_t := \mathbb{E}\left[\chi_t(\xi_{t+1})\right]$, for $t = 1, \ldots, T-1$ and let $\gamma > 1$ be arbitrary. Suppose that the scenario-tree is constructed following the identical conditional sampling scheme. Then, for $\epsilon > 0$, $\delta \in (0, \epsilon)$ and $\alpha \in (0,1)$, it follows that*

$$N(\epsilon, \delta, \alpha; p) \le \prod_{t=1}^{T-1} \max\{A_t, B_t\} =: \text{UPPER}(\epsilon, \delta, \alpha; p) \tag{10}$$

where, for $t = 1, \ldots, T-1$,

$$A_t := \left\lceil \frac{8\sigma_t^2 (T-1)^2}{(\epsilon-\delta)^2} \left[n_t \log\left(\frac{4\rho\gamma L_t D_t (T-1)}{\epsilon - \delta} \right) + \log\left(\frac{4(T-1)}{\alpha} \right) \right] \right\rceil, \tag{11}$$

$$B_t := \left\lceil \frac{1}{I_{\chi_t}(\gamma L_t)} \log\left(\frac{2(T-1)}{\alpha} \right) \right\rceil. \tag{12}$$

Moreover, if the problem also satisfies conditions (Mt.4), for $t = 1, \ldots, T-1$, then (10) also holds for $\delta = 0$ with the same values of A_t and B_t, $t = 1, \ldots, T-1$.

Here, the function $I_{\chi_t}(\cdot)$ is the convex conjugate (or Fenchel transform) of the function $\log(M_{\chi_t}(\cdot))$, for $t = 1, \ldots, T-1$; and $\rho > 0$ is an absolute constant that we know is at most 5 [4, Lemma 3]. For sufficiently small values of $\epsilon - \delta > 0$, we have that $A_t \ge B_t$ for each $t = 1, \ldots, T-1$. The dependence of A_t with respect to ϵ and δ is determined by the difference $\epsilon - \delta$. So, considering $\delta = 0$, $\alpha \in (0,1)$ fixed and $\epsilon > 0$ sufficiently small, we observe that the growth rate of UPPER$(\cdot)$ with respect to $\epsilon > 0$ is at most of order

$$\left(\frac{\sigma^2}{\epsilon^2} \left[n \log\left(\frac{LD(T-1)}{\epsilon} \right) \right] \right)^{T-1} (T-1)^{2(T-1)}, \tag{13}$$

where the problem's parameters n, σ, L and D are the maximum of the corresponding stagewise parameters. The estimate above was obtained for general multistage stochastic optimization problems. This class contains the class of multistage convex problems, and in particular, the classes of *linear* and *polyhedral* problems. To write the result for this class, we need to consider uniformly bounded conditions on the instances' parameters in order to prevent this quantity to be $+\infty$. In corollary 2 of [4], we have considered the following conditions

(UB) There exist positive (finite) constants σ, M, $n \in \mathbb{N}$, $\gamma > 1$ and β such that for every instance $(p) \in \mathcal{C}$ and $t = 1, \ldots, T-1$,

(i) $\sigma_t^2(p) \le \sigma^2$,
(ii) $D_t(p) \times L_t(p) \le M$,
(iii) $n_t(p) \le n$,

(iv) $(0 <)\, \beta \leq I_{\chi_t(p)}(\gamma L_t(p))$.

Then, it is immediate from the previous proposition the following upper bound for the class $\mathcal{C}$ of all T-stage stochastic optimization problems satisfying (M0), (M1), (Mt.1)-(Mt.4) and (UB):

$$N(\epsilon,\delta,\alpha;\mathcal{C}) = \sup_{P\in\mathcal{C}} N(\epsilon,\delta,\alpha;P) \leq \prod_{t=1}^{T-1} \max\{\bar{A}_t, \bar{B}_t\}, \tag{14}$$

where, for $t = 1,\ldots,T-1$,

$$\bar{A}_t := \left\lceil \frac{8\sigma^2(T-1)^2}{(\epsilon-\delta)^2}\left[n\log\left(\frac{4\rho\gamma M(T-1)}{\epsilon-\delta}\right) + \log\left(\frac{4(T-1)}{\alpha}\right)\right]\right\rceil, \tag{15}$$

$$\bar{B}_t := \left\lceil \frac{1}{\beta}\log\left(\frac{2(T-1)}{\alpha}\right)\right\rceil. \tag{16}$$

3 The main result

Here, we obtain a *lower bound* for the sample complexity of a class of T-stage stochastic problems that satisfies the previous regularity conditions and the uniformly bounded condition. So, when we compare the derived *lower bound* with the previous *upper bound*, we are obtaining estimates that hold for the same class of problems. Observe that in [5, 6] condition (Mt.3) was assumed in a more restrictive form. In fact, it was assumed that

$$\chi_t(\xi_{t+1}) = L_t, \tag{17}$$

for a.e. $\xi_{t+1} \in \operatorname{supp}(\xi_{t+1})$ and $t = 1,\ldots,T-1$. Here, we have assumed only that $\chi_t(\xi_{t+1})$ has a finite moment generating function in a neighborhood of zero. This is exactly the assumption that was considered in the derivation of the bound for static problems (see [7, Page 120, condition (A2)]). Moreover, to the best of our knowledge, for multistage stochastic problems, the case $\delta = 0$ have been firstly analyzed on [4]. It is worth noting that the examples below do not satisfy condition (17). This was one of the reasons that has motivated us to derive in [4] *upper bounds* under relaxed regularity assumptions when compared to the previous results in the literature. Moreover, we point out that each instance below is a multistage convex (continuous) stochastic optimization problem.

Let $T \geq 3$ be an integer. For each $k \in \mathbb{N}$, consider the instance (p_k) of a T-stage stochastic optimization problem (1) with the following data

- $\xi = (\xi_2,\ldots,\xi_T)$ is stagewise independent, $\xi_t \sim N(0, s^2 I_n)$, for every $t = 2,\ldots,T$, $s > 0$ and $n \in \mathbb{N}$,
- $F_t^k(x_t,\xi_t) := -2k\langle \xi_t, x_t\rangle$, for $t = 2,\ldots,T$,
- $X_t^k(x_{t-1},\xi_t) := \{x_{t-1}\}$, for $t = 2,\ldots,T$,

and $F_1^k(x_1) := ||x_1||^{2k}$ and $X_1^k := \frac{1}{k}\mathbb{B}_n$, where $\mathbb{B}_n$ is the closed unit euclidean ball of $\mathbb{R}^n$.

The scenario-tree is constructed following the identical conditional sampling scheme. Consider the independent random vectors

$$\mathfrak{S}_{N_2,\ldots,N_T} := \left\{ \xi_t^i \sim N(0, s^2 I_n) : \begin{array}{c} i = 1,\ldots,N_t, \\ t = 2,\ldots,T \end{array}\right\}, \tag{18}$$

and that each t^{th}-stage node has the same N_{t+1} successors nodes, $\xi_{t+1}^1,\ldots,\xi_{t+1}^{N_{t+1}}$, for $t = 1,\ldots,T-1$. Moreover, given the sample realization, the conditional probability of the states on the tree are equal to the unconditional ones and the discrete version of the random data is also stagewise independent.

Let us derive the objective functions $f^k(\cdot)$ and $\hat{f}^k(\cdot)$, respectively, of the true problem and the SAA problem given $\mathfrak{S}_{N_2,\ldots,N_T}$. To simplify the notation, we have dropped the subscript of the SAA objective function writing $\hat{f}(\cdot)$. We will do the same to the corresponding optimal-value functions, that we will derive on the way. We begin with the T^{th}-stage (optimal) value function obtained by the dynamic programming equation:

$$\begin{aligned}\mathcal{Q}_T(x_{T-1},\xi_T) &= \inf_{x_T\in X_T(x_{T-1},\xi_T)} \{-2k\langle \xi_T, x_T\rangle\} \\ &= -2k\langle \xi_T, x_{T-1}\rangle.\end{aligned} \tag{19}$$

The true problem and SAA problem T^{th}-stage cost-to-go functions are obtained, respectively, by taking the expected value of $\mathcal{Q}_T(x_{T-1},\xi_T)$ with respect to the true distribution of ξ_T and its discrete version:

$$\begin{aligned}\mathcal{Q}_T(x_{T-1}) &= \mathbb{E}\left[-2k\langle \xi_T, x_{T-1}\rangle\right] = 0, \\ \hat{\mathcal{Q}}_T(x_{T-1}) &= \hat{\mathbb{E}}\left[-2k\langle \xi_T, x_{T-1}\rangle\right] = -2k\left\langle \bar{\xi}_T, x_{T-1}\right\rangle,\end{aligned} \tag{20}$$

where $\bar{\xi}_T = \frac{1}{N_T}\sum_{i=1}^{N_T}\xi_T^i$. Continuing backward in stages, it is not difficult to verify that:

$$\begin{aligned}\mathcal{Q}_t(x_{t-1},\xi_t) &= -2k\langle \xi_t, x_{t-1}\rangle, \\ \hat{\mathcal{Q}}_t(x_{t-1},\xi_t) &= -2k\left\langle \bar{\xi}_t + \cdots + \bar{\xi}_T, x_{t-1}\right\rangle,\end{aligned} \tag{21}$$

where $\bar{\xi}_t := \frac{1}{N_t}\sum_{i=1}^{N_t}\xi_t^i$, for $t = 2,\ldots,T-1$. It follows from (21) that the true and SAA first-stage cost-to-go functions are

$$\begin{aligned}\mathcal{Q}_2(x_1) &= 0 \\ \hat{\mathcal{Q}}_2(x_1) &= -2k\left\langle \bar{\xi}_2 + \ldots + \bar{\xi}_T, x_1\right\rangle,\end{aligned} \tag{22}$$

Let us define $\eta := \bar{\xi}_2 + \ldots + \bar{\xi}_T$, so by (22), it follows that $f^k(x_1) = ||x_1||^{2k}$ and $\hat{f}^k(x_1) = ||x_1||^{2k} - 2k\langle \eta, x_1\rangle$, for $x_1 \in \frac{1}{k}\mathbb{B}_n$. The (unique) first-stage optimal solution of the true problem is $\bar{x}_1 = 0$, so its optimal-value is $v^* =$

0. Moreover, the (*exact*) first-stage optimal solution of the SAA problem is given by:

$$\hat{x}_1 = \begin{cases} 0 & \text{, if } ||\eta|| = 0 \\ \dfrac{1}{||\eta||^{\gamma_k}}\eta & \text{, if } 0 < ||\eta|| \le (\frac{1}{k})^{2k-1} \\ \dfrac{1}{||\eta||k}\eta & \text{, if } ||\eta|| > (\frac{1}{k})^{2k-1} \end{cases} \tag{23}$$

where $\gamma_k = \dfrac{2k-2}{2k-1}$.

Hence, given $\epsilon \in \left(0, \frac{1}{k^{2k}}\right)$, $\hat{x}_1$ is an ϵ-optimal solution of the true problem if, and only if, $||\eta||^{2k(1-\gamma)} \le \epsilon$. Define $v_k := 2k(1-\gamma_k) = 2k/(2k-1)$. By (18), $\eta \sim N\left(0, \sum_{t=2}^{T} \dfrac{s^2}{N_t} I_n\right)$. Considering the harmonic mean, say hm, of the numbers $N_2, ..., N_T$:

$$\frac{T-1}{\text{hm}} := \sum_{t=2}^{T} \frac{1}{N_t},$$

it follows that:

$$\eta \sim N\left(0, \frac{s^2(T-1)}{\text{hm}} I_n\right). \tag{24}$$

Here, we analyze the case $\delta = 0$, i.e. that we obtain an exact optimal-solution to the SAA problem. We write $\mathcal{N}(\epsilon, \alpha; p_k)$ and $N(\epsilon, \alpha; p_k)$, respectively, instead of $\mathcal{N}(\epsilon, 0, \alpha; p_k)$ and $\mathcal{N}(\epsilon, 0, \alpha; p_k)$.

Now, we will show that if $(N_2, \dots, N_T) \in \mathcal{N}(\epsilon, \alpha; p_k)$, for $\epsilon \in \left(0, \frac{1}{k^{2k}}\right)$ and $\alpha \in (0, \bar{\alpha})$, where $\bar{\alpha} := \mathbb{P}\left[\chi_1^2 > 1\right] \approx 0.3173$, then

$$N := \prod_{t=2}^{T} N_t \ge \left(\frac{s^2}{\epsilon^{2-\frac{1}{k}}}\right)^{T-1} \left[n(T-1)\right]^{T-1}. \tag{25}$$

Since $N(\epsilon, \alpha; p_k) = \inf\left\{\prod_{t=2}^{T} N_t : (N_2, \dots, N_T) \in \mathcal{N}(\epsilon, \alpha; p_k)\right\}$, the right-hand side of (25) will be a lower bound for the sample complexity of the instance (p_k).

Indeed, suppose that $(N_2, \dots, N_T) \in \mathcal{N}(\epsilon, \alpha; p_k)$ where ϵ and α are as specified before, then

$$\mathbb{P}\left[||\eta||^{v_k} \le \epsilon\right] \ge 1 - \alpha. \tag{26}$$

This is equivalent to $\alpha \ge 1 - \mathbb{P}\left[||\eta||^{v_k} \le \epsilon\right] = \mathbb{P}\left[||\eta||^{v_k} > \epsilon\right]$. It follows from (24) that $\dfrac{\text{hm}}{s^2(T-1)}||\eta||^2 \sim \chi_n^2$.

Observe also that

$$\mathbb{P}\left[\frac{\text{hm}}{s^2(T-1)}||\eta||^2 > \frac{\epsilon^{2/v_k}\,\text{hm}}{s^2(T-1)}\right] = \mathbb{P}\left[||\eta||^{v_k} > \epsilon\right].$$

Since the sequence $\mathbb{P}\left[\chi_n^2 > n\right]$ is monotone increasing and $\mathbb{P}\left[\chi_1^2 > 1\right] = \bar{\alpha}$, if $\alpha \in (0, \bar{\alpha})$ we must necessarily have $\frac{\epsilon^{2/v_k}\,\text{hm}}{s^2(T-1)} > n$, that is:

$$\text{hm} > \frac{s^2}{\epsilon^{2/v_k}} n(T-1). \tag{27}$$

It is a well known result that the harmonic mean of (positive) real numbers is always less than or equal to its *geometric mean*

$$\text{gm} := (N_2 ... N_T)^{1/(T-1)} = N^{1/(T-1)}. \tag{28}$$

So, we arrive at the following lower bound for N:

$$\begin{aligned} N &> \left(\frac{s^2 n}{\epsilon^{2-\frac{1}{k}}}\right)^{T-1} (T-1)^{T-1} \\ &= \left(\frac{\sigma_k^2 n}{4\epsilon^{2-\frac{1}{k}}}\right)^{T-1} (T-1)^{T-1}. \end{aligned} \tag{29}$$

We will show in a moment that $\sigma_k = 2s$, for all $k \in \mathbb{N}$. Observe the similarities between (29) and (13), without the logarithm term. The additional multiplicative factor of (29) is of order $(T-1)^{(T-1)}$, instead of $(T-1)^{2(T-1)}$. Even having a smaller growth order, this shows that such multiplicative factor is unavoidable, and that the number of scenarios for T-stage problems can grow much faster with respect to T than merely the number of scenarios for the *static* case to the power of $T-1$. This confirms the results firstly stated in [4]. Moreover, we conclude by (29) that

$$\lim_{\epsilon \to 0+} \frac{N(\epsilon, \alpha, \{p_k : k \in \mathbb{N}\})}{\epsilon^{2(T-1)-s}} = +\infty, \tag{30}$$

for all $s \in (0, 2(T-1))$, showing a growth order that is almost $1/\epsilon^{2(T-1)}$, when $\epsilon \to 0+$.

Now, let us verify that each instance (p_k) satisfies the regularity conditions (M0) and (Mt.1)-(Mt.4), for $t = 1, \dots, T-1$. Condition (M0) is trivially true. Defining $\mathcal{X}_t^k := \frac{1}{k}\mathbb{B}_n$, for $t = 1, \dots, T-1$, we see that $D_t^k := \text{diam}(\mathcal{X}_t^k) = 2/k$ and $X_t(x_{t-1}, \xi_t) \subseteq \mathcal{X}_t^k$, for every $x_{t-1} \in \mathcal{X}_{t-1}^k$ and $\xi_t \in \mathbb{R}^n$. So, conditions (Mt.1) and (Mt.4) hold, for every $t = 1, \dots, T-1$. Moreover,

$$\begin{aligned} \left|\mathcal{Q}_{t+1}(x_t', \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t, \xi_{t+1})\right| &= \\ 2k\left|\langle \xi_{t+1}, x_t' - x_t \rangle\right| &\le 2k||\xi_{t+1}||\,||x_t' - x_t||, \end{aligned} \tag{31}$$

for all $x_t', x_t \in \mathcal{X}_t^k$ and $\xi_{t+1} \in \mathbb{R}^n$, so condition (Mt.3) is satisfied with $\chi_t^k(\xi_{t+1}) = 2k||\xi_{t+1}||$. As we have pointed out, observe that does not exists $L \in \mathbb{R}$ such that the optimal-value function is L-Lipschitz-continuous, with probability one, in the first-variable. In fact, let $L > 0$ be arbitrary and take $x_t', x_t \in \mathcal{X}_t^k$ such that $z := x_t' - x_t \neq 0$. Consider the following (nonempty) open subset of $\mathbb{R}^n$

$$E := \{\xi \in \mathbb{R}^n : \cos(\theta_{\xi,z}) > 1/2,\ ||\xi|| > L/k\},$$

where $\theta_{\xi,z} \in [0, \pi]$ is the angle between (the nonzero) vectors ξ and z. Since $\text{int}(E) \neq \emptyset$ and $\xi_{t+1} \sim N(0, s^2 I_n)$

(with $s > 0$), we conclude that $\mathbb{P}[\xi_{t+1} \in E] > 0$. Moreover, for $\xi_{t+1} \in E$, we have

$$\left|\mathcal{Q}_{t+1}(x'_t, \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t, \xi_{t+1})\right| = 2k\left| \langle \xi_{t+1}, x'_t - x_t \rangle \right| = 2k\cos(\theta_{\xi_{t+1},z})||\xi_{t+1}||\, ||x'_t - x_t|| > L||x'_t - x_t||.$$

Finally, we show that (Mt.2) holds. In fact, for every $x_t \in \mathcal{X}^k_t$

$$\mathcal{Q}_{t+1}(x_t, \xi_{t+1}) - \mathcal{Q}_{t+1}(x_t) = -2k \langle \xi_{t+1}, x_t \rangle \sim N(0, 4k^2||x_t||^2 s^2), \tag{32}$$

so $\sigma_t = \sigma := 2s > 0$ is such that this family of random variables is σ-sub-Gaussian.

Now, we show that the (UB) condition is also satisfied. Conditions (i.) and (iii.) are trivially satisfied. Moreover,

$$L^k_t := \mathbb{E}\left[\chi^k_t(\xi_{t+1})\right] = 2ks\mathbb{E}\left[||Z||\right] = 2kc_n s, \tag{33}$$

where $Z \sim N(0, I_n)$ and the last equality follows from lemma 1 of the appendix. So, $D^k_t \times L^k_t = 4c_n s =: M$, for all $k \in \mathbb{N}$ and $t = 1, \dots, T-1$. Since $||\xi|| \leq ||\xi||_1$, we obtain the following

$$M_{\chi_k}(\theta) = \exp\left(2k||\xi||\theta\right) \leq 2^n \exp\left(2nk^2 s^2 \theta^2\right), \forall \theta \in \mathbb{R}. \tag{34}$$

Consequently, for $\gamma > 1$ (see also lemma 2)

$$I_{\chi^k_t}(\gamma L_k) \geq \frac{1}{4}\gamma^2 - n\log(2), \tag{35}$$

for all $t = 1, \dots, T-1$ and $k \in \mathbb{N}$. Taking $\gamma = 2\sqrt{n}$ we obtain that $I_{\chi^k_t}(\gamma L_k) \geq n(1 - \log(2)) \geq 1 - \log(2) =: \beta (> 0)$. So, we have verified all the items of (UB). We can summarize the discussion above in the following proposition.

Proposition 2 *Let $\mathcal{C}$ be the class of all T-stage stochastic convex problems satisfying the regularity conditions (M0), (Mt.1)-(Mt.4), for $t = 1, \dots, T-1$, and (UB) with arbitrary constants $\sigma > 0$, $M > 0$, $n \in \mathbb{N}$, $\gamma > 1$ and $\beta > 0$, where $\frac{1}{2}\gamma^2 \geq \beta + n\log(2)$. Then, for α (fixed) sufficiently small,*

$$\lim_{\epsilon \to 0+} \frac{N(\epsilon, \alpha; \mathcal{C})}{\epsilon^{2(T-1)-z}} = +\infty, \text{ for all } z \in (0, 2(T-1)). \tag{36}$$

The proof is immediate, since $\mathcal{C} \supseteq \{p_k : k \in \mathbb{N}\}$, for sufficiently small $s > 0$, which implies that $N(\epsilon, \alpha; \mathcal{C}) \geq N(\epsilon, \alpha; \{p_k : k \in \mathbb{N}\})$.

4 Further considerations

In this section, we discuss two issues that were pointed by an anonymous referee.

A stream of research on multistage financial stochastic optimization problems have derived some sample complexity's *lower bounds* through no-arbitrage reasoning arguments. Here, we give a very brief and incomplete review of this literature. In [2], it was addressed how the discretization of the random data for multistage financial stochastic programming models, whose state-variables are typically assumed continuous, can introduce arbitrage opportunities in the scenario tree. The author has also illustrated that even when arbitrage opportunities cannot be explored directly by the decision maker, because of trading restrictions (e.g. presence of no short-sale constraints) or markets with friction (e.g. presence of trading costs), they may introduce strong bias in the optimal investment strategy. In [1], the authors stated necessary conditions for the node structure of a scenario tree to rule out arbitrage opportunities. Indeed, they have asserted that the branching factor of each node of the tree must, at least, equal the number of non-redundant assets on the financial model. Then, they have studied how this condition affects some computational strategies applied on the scenario tree, like state aggregation and scenario reduction, in order to the reduce its number of scenarios.

In our framework, let us verify how this bound compares with ours. Consider a multistage financial optimization model that is possible to invest in n non-redundant securities at the beginning of each period. Following the reasoning that the branching factor of the scenario tree must be at least n to avoid arbitrage opportunities, the total number of scenarios in the tree must be at least

$$N = \prod_{t=2}^{T} N_t \geq n^{T-1}, \tag{37}$$

that is exponential with respect to the number of stages. For financial stochastic optimization problems, the *no-arbitrage condition* can be seen as a kind of *a priori* information that one wants to impose to the tree. In fact, this condition is fundamental to the *arbitrage pricing theory* in finance. Of course, it is interesting to obtain *lower bounds*, such as ours, that are applicable to any field of multistage stochastic problems, including *financial problems.*

Moreover, it is worth noting that our bound was derived without assuming this extra *a priori* condition. Observe also that, in our context, equation (37) could only be obtained assuming that the feasible sets of each stage are unbounded. In fact, the complete argument is that if the SAA problem admits arbitrage opportunities, then it would not have an optimal solution, since the optimal-value of the problem would be unbounded. However, for compact feasible sets, this is not the case (even for unbounded sets like the nonnegative orthant), so it is not clear how one could obtain this lower bound in this situation. Observe that, opposite to the situation when one derives *upper bounds*, the regularity conditions, such as compactness of the feasible sets (etc), make it difficult the derivation of *lower bounds* as they restrict the class of problems. So, it worth noting that our *lower bound* was derived under "nice" regularity conditions, that are, in fact, "nice" to derive *upper bounds*, and only turns the task harder when one tries to obtain *lower bounds*. Moreover, in our *lower bound* the base in which we observe the exponential dependence on the number of stages also depends on $(T-1)/\epsilon^{2-s}$, $0 < s < 2$, additionally to the dimension of the decisions variables n. Its growth order with respect to T is much higher than the one in (37).

So, to the best of our knowledge, estimates such as ours have never been derived before for multistage stochastic optimization problems and this work is a natural extension of the stream of research given by references [3, 5, 7], where *upper bounds* for the sample complexity of the SAA method were derived.

Another interesting consideration is that most multistage stochastic optimization problems that are solved in practice are linear. So, it would be interesting to obtain sample complexity's *lower bounds* for this class of problems. Here, we make only preliminary considerations concerning this point. First of all, observe that the *linear* and the *polyhedral* classes could be seen as essentially the same, since by adding one extra decision variable for each stage, we can cast any *polyhedral* problem as a *linear* one. Of course, the reciprocal is immediately true, since every *linear* problem is also *polyhedral*. Moreover, our examples can be approximated arbitrarily well by *polyhedral* problems. Indeed, one just need to approximate function $F_1^k(x_1)$, that is $||x_1||^{2k}$, by a polyhedral function (and also the stage feasible sets). Of course, the details must be carried out with care, but this indicates that it is possible to obtain a similar *lower bound* for the sample complexity of multistage linear stochastic optimization problems.

5 Discussion

In this paper, we have obtained a *lower bound* for a well-behaved family of multistage convex stochastic optimization problems. This was done by constructing a family of examples of T-stage problems that were possible to derive a *lower bound* for each one of its instances. The dependence of the *lower bound* with respect to some of the problem's parameters, like $\sigma > 0$ and n, were similar to the one observed in the *upper bound*. The dependence with respect to ϵ was almost of order $1/\epsilon^{2(T-1)}$. Moreover, additionally to the exponential behavior of the *upper* and *lower bounds* with respect to the number of stages, we have shown that a multiplicative factor at least of order $(T-1)^{(T-1)}$ is unavoidable for some problems. This represents an even more extreme growth behavior required for the total number of samples than just the static's behavior to the power of $T-1$. A topic for further investigation is the derivation of a *lower bound* for the class of multistage *linear* stochastic optimization problems. We believe that it is possible to obtain a similar *lower bound* for this class, but one must work out the details with care.

6 Appendix: some technical lemmas

Lemma 1 *Let ξ be a multivariate standard Gaussian random vector, i.e. $\xi \sim N(0, I_n)$, $n \in \mathbb{N}$. Then*

$$\frac{n}{\sqrt{n+1}} \leq \mathbb{E}\left[||\xi||\right] = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \leq \sqrt{n}, \tag{38}$$

where $\Gamma(s) := \int_0^{+\infty} u^{s-1}\exp\{-u\}du$, $s > 0$, is the gamma function.

Proof The probability density function of ξ is given by

$$f_\xi(x) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{||x||^2}{2}\right\}, \forall x \in \mathbb{R}^n. \tag{39}$$

Observe that $f_\xi(x) = g(||x||)$, where

$$g(r) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{r^2}{2}\right\}.$$

So, the expected-value of $||\xi||$ is

$$\mathbb{E}\left[||\xi||\right] = \int_{x\in\mathbb{R}^n} ||x||g(||x||)dx = \int_0^{+\infty} S_{n-1}(r)rg(r)dr, \tag{40}$$

where $S_{n-1}(r) = \dfrac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}r^{n-1}$ is the surface area of the sphere of $\mathbb{R}^n$ with radius r. So, we need to solve the following integral in one variable

$$\mathbb{E}\left[||\xi||\right] = \frac{n\pi^{n/2}}{(2\pi)^{n/2}\Gamma\left(\frac{n}{2}+1\right)}\int_0^{+\infty} r^n\exp\left\{-r^2/2\right\}dr.$$

(41)

Making the change of variables $u = r^2/2$, one can easily verify the equality in (38). The upper bound is an immediate consequence of Jensen's inequality, since $\mathbb{E}\left[||\xi||^2\right] = n$. Finally, using an induction argument on $k \in \mathbb{N}$, for $n = 2k-1$ and $n = 2k$ (separately), one can show the lower bound after some tedious calculations. It is not difficult to verify our claims and, for such, it is worth noting that $\Gamma(s) = (s-1)\Gamma(s-1)$, for $s > 1$, and $\Gamma(1/2) = \sqrt{\pi}$. □

Lemma 2 *Let $\chi_k(\xi) := 2k||\xi||$, where $\xi \sim N(0, s^2 I_n)$, $s > 0$, $c_n = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$ and $k \in \mathbb{N}$. The following conditions hold:*

i. $L_k := \mathbb{E}\left[\chi_k(\xi)\right] = 2kc_n s$, $\forall k \in \mathbb{N}$.
ii. $I_{\chi_k}(\gamma L_k) \geq \frac{1}{4}\gamma^2 - n\log(2)$, $\forall k \in \mathbb{N}$ *and* $\gamma > 1$.

Proof The first item is immediate from lemma 1. Let us show the second item. Taking the logarithm on (34), we obtain

$$m_{\chi_k}(t) := \log\left(M_{\chi_k}(t)\right) \leq n\log(2) + 2nk^2s^2t^2,\ \forall t \in \mathbb{R}. \tag{42}$$

Let $y > 0$ be arbitrary, then

$$\begin{aligned} I_{\chi_k}(y) &= \sup_{t\in\mathbb{R}} \left\{ty - m_{\chi_k}(t)\right\} \\ &\geq \sup_{t\in\mathbb{R}} \left\{ty - n\log(2) - 2nk^2s^2t^2\right\} \\ &= \frac{y^2}{8ns^2k^2} - n\log(2). \end{aligned} \tag{43}$$

Given $\gamma > 1$, take $y := \gamma L_k$ on (43) to obtain the following lower bound

$$I_{\chi_k}(\gamma L_k) \geq \frac{c_n^2\gamma^2}{2n} - n\log(2). \tag{44}$$

From (38), it follows that $\frac{c_n^2}{n} \geq \frac{n}{n+1} \geq 1/2$, for all $n \in \mathbb{N}$, and we obtain estimate (ii.). □

7 Acknowledgments

The author wishes to thank Prof. Alexander Shapiro for useful discussions, helpful comments and support in writing this paper. An anonymous referee made a number of helpful comments which also improved this document. The author also wishes to thank Prof. Alfredo Iusem for helping him preparing a first-version of this document. 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.

References

1. A. Geyer, M. Hanke, A. Weissensteiner, No-arbitrage conditions, scenario trees, and multi-asset financial optimization, European Journal of Operational Research 206 (2010) 609-613.
2. P. Klaassen, Discretized reality and spurious profits in stochastic programming models for asset/liability management, European Journal of Operational Research 101 (1997) 374-392.
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. M.M.C.R. Reaiche, Revisiting some results on the complexity of multistage stochastic programs and some extensions, E-print available at: `http://www.optimization-online.org/DB_FILE/2015/10/5183.pdf`, 2015.
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.