GENERAL FEASIBILITY BOUNDS FOR SAMPLE AVERAGE APPROXIMATION VIA VAPNIK-CHERVONENKIS DIMENSION*

HENRY LAM † AND FENGPEI LI ‡

Abstract. We investigate the feasibility of sample average approximation (SAA) for general stochastic optimization problems, including two-stage stochastic programs without relatively complete recourse. We utilize results from the *Vapnik–Chervonenkis* (VC) dimension and *probably approximately correct* learning to provide a general framework to construct feasibility bounds for SAA solutions under minimal structural or distributional assumption. We show that, as long as the hypothesis class formed by the feasible region has a finite VC dimension, the infeasibility of SAA solutions decreases exponentially with computable rates and explicitly identifiable accompanying constants. We demonstrate how our bounds apply more generally and competitively compared to existing results.

Key words. sample average approximation, feasibility, sample complexity, two-stage stochastic programming, Vapnik–Chervonenkisis dimension

AMS subject classifications. 90C15, 03E75

1. Introduction.

Consider the stochastic optimization problem

$$\inf_{x\in\mathcal{X}} F(x) \triangleq \mathbb{E}[f(\xi,x)], \tag{1.1}$$

where $\mathcal{X}$, the space for decision variables, is a nonempty closed subset of a Polish space, i.e., a complete, separable, metric space (e.g., $\mathcal{X}\subseteq\mathbb{R}^n$ or $\mathbb{R}^{n-p}\times\mathbb{Z}^p$ for mixed-integer decision sets, with Euclidean metric) equipped with the Borel σ-algebra $\mathscr{B}$ and $\xi:\Omega\to\Xi\subseteq\mathbb{R}^r$ is some random vector on a complete probability space $(\Omega,\mathcal{F},\mathbb{P})$. For each realization of $\xi(\omega)\in\Xi$ (henceforth, we suppress the dependence of ξ on $\omega\in\Omega$), we assume $f(\xi,\cdot):\mathcal{X}\to\mathbb{R}\cup\{+\infty\}$ is a lower semicontinuous function mapping to the extended real line and the function $f(\xi,x)$ is measurable with resepct to the completion of product σ-algebra $\mathcal{F}\otimes\mathscr{B}$. We also assume the set $\{x:x\in\mathcal{X} \text{ and } F(x)<+\infty\}$ is nonempty. Moreover, we assume all quantities of interest are measurable and defer the technical arguments on measurability to the appendix.

The class of problems under (1.1) is difficult to evaluate in general, especially for high–dimensional ξ. As a popular tractable approximation, the sample average approximation (SAA) method [46] solves the sampling-based counterpart of (1.1):

$$\inf_{x\in\mathcal{X}} \hat{F}_N(x) \triangleq \frac{1}{N}\sum_{i=1}^N f(\xi_i,x), \tag{1.2}$$

where $\xi^{[N]}\triangleq(\xi_1,\xi_2,...,\xi_N)$ are independent and identically distributed (IID) samples drawn from $\mathbb{P}$. An optimal solution of the SAA depends on the realization of $\xi^{[N]}$ and

*Submitted.

Funding: This work was partially supported by the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

†Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA (khl2114@columbia.edu).

‡Corresponding author. Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA. Current address: Morgan Stanley Machine Learning Research, New York, NY 10019, USA (fl2412@columbia.edu).

shall be denoted $x^\star(\xi^{[N]})$. Theoretical properties and numerical performances of SAA have been extensively studied in, e.g., [32, 46, 24], and its applications in stochastic optimization and chance-constrained programming can be found in, e.g., [27, 44, 54]. As an important class of (1.1), two-stage stochastic programming has applications in transportation planning [5, 33], disaster management [41], water recourse management [25], and inventory management [19]. Most of these studies rely on the assumption that $f(\xi, x) < +\infty$ with probability 1 for all $x \in \mathcal{X}$, which is called the relatively complete recourse condition. However, in many real-world applications, this condition is restrictive and there has been a growing literature studying two-stage stochastic programming without this assumption, i.e., $\mathbb{P}(\{\xi \in \Xi : f(\xi, x) = +\infty\}) > 0$ for some $x \in \mathcal{X}$ (see [15, 16, 37]). In such a setting, the SAA solution $x^\star(\xi^{[N]})$ is not necessarily feasible for the original problem (1.1) and it would be desirable to quantify the level of infeasibility for the SAA solution.

As nicely discussed in the recent works of [15, 37], the feasibility issue of SAA arises when $f(\xi, \cdot)$ maps to the extended real line. Following the notations in [37], we define dom $f_\xi \triangleq \{x \in \mathcal{X} : f(\xi, x) < +\infty\}$. Then by solving (1.2) we would obtain an optimal solution $x^\star(\xi^{[N]}) \in$ dom $\hat{F}_N \triangleq \bigcap_{i=1}^N$ dom f_{ξ_i} where dom $\hat{F}_N$ is the feasible region for the SAA. As we have assumed the set $\{x : x \in \mathcal{X} \text{ and } F(x) < +\infty\}$ is nonempty, the SAA feasible region is nonempty with probability 1 (since $F(x) < +\infty$ implies $\mathbb{P}(f(x, \xi) = +\infty) = 0$ and in turn $x \in$ dom $\hat{F}_N$ with probability 1). However, $x^\star(\xi^{[N]})$ may not be feasible for the original problem (1.1), i.e., $x^\star(\xi^{[N]}) \notin$ dom $F \triangleq \{x \in \mathcal{X} : F(x) < +\infty\}$. In other words, defining the violation probability $V(x)$ for $x \in \mathcal{X}$ as

$$V(x) \triangleq \mathbb{P}(\xi \in \Xi : x \notin \text{dom } f_\xi), \tag{1.3}$$

we could have $V(x^\star(\xi^{[N]})) > 0$ with positive probability.

In this paper, we utilize a framework based on the Vapnik–Chervonenkis (VC) dimension to analyze the feasibility of SAA solutions, including two-stage stochastic programming. Following [15, 37], we focus on showing the exponential decrease of $V(x^\star(\xi^{[N]}))$ as N grows. Specifically, letting $\mathbb{P}^N$ denote the sampling measure governing the generation of IID samples $\xi^{[N]}$ (notice the feasibility of $x^\star(\xi^{[N]})$ is random depending on $\xi^{[N]}$), we derive exponential bounds for $V(x^\star(\xi^{[N]}))$ under $\mathbb{P}^N$. As a key contribution, we show that, when the VC dimension of the feasible domain is finite, our framework produces feasibility bounds that are both general and explicit. In particular, the constants in the bounds are computable with respect to problem parameters. Moreover, aside from the finite VC dimension of the feasible domain, we impose no additional requirement on the geometric or distributional properties of (1.1) (i.e., whether the problem is convex or linear, whether the optimal solution is in the interior or on the boundary of dom F, whether $\mathcal{X}$ is finite or functions $\{f(\xi, \cdot)\}_{\xi \in \Xi}$ have a chain-constrained domain, as in [37]), or specific regularity conditions on $f(\xi, x)$ (i.e., Lipschitz continuity or the existence of certain moment generating functions, as in [15, 37]). Consequently, the analysis is widely applicable in both scenarios where some of the best-known results on SAA feasibility have been presented, and other scenarios where no related results have been established. Furthermore, the feasibility result under this framework is not restricted to the optimal solution of SAA, but any generic point within the SAA feasible region with probability 1. As a result, when the SAA problem is nonconvex and solvable only up to a local optimum, or when approximate algorithms are required, our feasibility guarantee would still hold. Finally, we show that the generality of this framework does not come at a cost of worse sample complexity as our bounds are comparable to, if not better than, existing ones.

There exists a large body of literature which leverages techniques from statistical/computational learning theory and empirical process theory to analyze the performance of SAA, including the (asymptotic) convergence rate of optimal solution, and (non asymptotic) concentration inequalities. The seminal work of [47] surveys the sample complexity of two-stage stochastic (linear) programming in obtaining so-called ϵ-optimal solutions using large deviations theory or the uniform law of large numbers. In [42, 43], Talagrand's "generic chaining" tail bounds and "localization" techniques for sub-Gaussian processes (see also [49]) are used to derive nonasymptotic risk bounds for the SAA optimal solution. Moreover, in [22], the Rademacher averages of functional class, a concept closely related to the VC dimension, is used to established convergence results for SAA in general compound stochastic optimization problems. As we can see, the concept of using statistical learning theory/complexity to analyze SAA is not entirely new. However, these results largely focus on the optimization accuracy of SAA as in [47], or the convergence rate of optimal SAA solution. In contrast, our paper provides a first application of the VC dimension and PAC learning in analyzing the feasibility of SAA solutions (optimal or otherwise). To this end, a related application is studied in [18, 21] under the context of chance-constrained programming, where the feasibility of constraint sampling is analyzed. It is worth noting that, the Rademacher complexity theory, as an alternative to the VC theory, typically obtains tighter (usually by a logarithmic factor) generalization error bounds which in fact subsume bounds provided by the VC theory [28]. However, the standard method for evaluating Rademacher complexity relies on a certain empirical risk minimization algorithm (ERM) [28] which may not be tractable, and one common way to bound the data-dependent Rademacher complexity is directly by the VC dimension. On the other hand, for the purpose of analyzing SAA feasibility, VC bounds are sufficient to establish results that are either new or comparable/superior to known ones. More importantly, the bounds we achieve do not depend on any data distribution, and carry explicit constants rather than a standalone big-O notation as in typical prior results on SAA feasibility or learning theory analysis. Thus, in this paper we focus on VC bounds which, although not the tightest possible (we do remark on where the bounds can be improved when appropriate), offer a great generality and a wide applicability, while maintaining a relatively simple, concise appeal.

Finally, we mention a growing body of literature exploiting low-dimensional structures (i.e., sparsity or low rankness) in high-dimensional problems to achieve a sample complexity that is logarithmic-in-dimension, instead of the usual linear-in-dimension sample complexity for the optimality of SAA solutions [10, 34, 35]. As we shall see, these improvements correspond to reduced VC dimensions for low-dimensional structures and also apply to our analysis of SAA feasibility.

The rest of the paper is organized as follows. In Section 2, we review the recent papers with closely related results, especially [15, 37]. In Section 3, we present the concepts of our framework and main result. Section 4 focuses on the applications of our main result. In the first part of this section, we present three novel results on the feasibility of two-stage stochastic programs: First, a general feasibility result (not limited to linear ones); Second, a feasibility bound for mixed-integer recourse, and third an exploitation of low-dimensional structures to obtain a logarithmic-in-dimension sample complexity in high-dimensional two-stage stochastic programs. In the second part of Section 4, we consider two special structures used to provide feasibility results in the existing literature with which we compare our results. The first is the chain-constrained domain considered in [37]. The second is a finite feasible domain considered in [15]. Finally, in the appendix, we provide a specially constructed

example that allows a direct verification of our feasibility results.

2. Review of Related Results. We discuss the existing results on SAA feasibility in [15, 37]. A considerable part of [15] discusses how to solve a so-called padded, modified version of SAA to obtain a *complete feasible* solution (i.e., $V(x) = 0$) with high confidence, which is somewhat different from the perspectives of our paper and [37]. In particular, we consider the feasibility for SAA in its original form and do not restrict our attention to completely feasible solutions. Both [15] and [37] discuss results of the form

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq \delta, \tag{2.1}$$

referred to as *high recourse likelihood solution* in [15]. In particular, [15] presents these bounds in two cases, one for finite $\mathcal{X}$ and another for two-stage stochastic linear programs. We shall discuss them in detail in Section 4 when we compare them with our results. On the other hand, the feasibility results in [37] are more general and can be summarized into three different scenarios.

- *Scenario 1*: In the presence of the so-called chain-constrained domain of order m (to be explained later) on dom f_ξ, [37] shows

$$\begin{aligned}\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) &\leq \sum_{k=0}^{m-1} \binom{N}{k} \epsilon^k (1-\epsilon)^{N-k} \\ &\leq \exp\left\{ - \frac{(N\epsilon - m + 1)^2}{2N\epsilon} \right\}.\end{aligned}$$

  The second inequality is given by the Chernoff bound, also shown in [13] and [37].
- *Scenario 2*: In the context of convexity, meaning $\mathcal{X}$ is closed and convex and the set of optimal solutions $\mathcal{X}^\star$ is nonempty, and $f(\xi, \cdot)$ is convex for all $\xi \in \Xi$, along with additional regularity conditions on $f(\xi, \cdot)$ and $\mathcal{X}$ (e.g., finite moment generating function for $f(\xi, x)$), [37] shows that for $\mathcal{X}^\star$ in the interior of dom F,

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > 0) \leq Ce^{-N\beta},$$

  where C and β are unknown constants.
- *Scenario 3*: In the context of convexity, if dom f_ξ is a chain-constrained domain as in Scenario 1, along with the additional regularity conditions, [37] shows that for $\mathcal{X}^\star$ which may have nonempty intersection with the boundary of dom F, one can still guarantee

$$\begin{aligned}\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) &\leq Ce^{-N\beta} + \sum_{k=0}^{|\mathcal{J}|-1} \binom{N}{k} \epsilon^k (1-\epsilon)^{N-k} \\ &\leq Ce^{-N\beta} + \exp\left\{ - \frac{(N\epsilon - |\mathcal{J}| + 1)^2}{2N\epsilon} \right\},\end{aligned}$$

  where C and β are again unknown constants as in Scenario 2 and $\mathcal{J}$ is the index set of active constraints at $\mathcal{X}^\star$ with the boundary of dom F. Notice it is shown in [12] that $|\mathcal{J}|$ is bounded by n, the dimension of the decision variable, which yields a useful upper bound regardless of the behavior of $\mathcal{J}$ (also note that in this case the order of the chain-constrained domain does not show up explicitly in the bound).

In all scenarios, a desirable exponential decrease of $V(\cdot)$ as N grows can be shown. However, there are several potential limitations. First, there exist hidden constants in the feasibility bound: In Scenarios 2 and 3, which are of great importance in stochastic convex programming, the rate of exponential decrease is established but governed by unknown constants C and, more importantly β which directly dictates the rate of exponential decrease. Second, the dependence of the bound on m, the order of the chain-constrained structure, can become potentially restrictive as m gets large (or even infinite) in many cases, a challenge that is also mentioned in [15]. Furthermore, even though it is motivated from practical examples in [37], the chain-constrained structure can be difficult to verify in general. Notably, the feasibility bound derived in Scenario 3 (the optimal solution of (1.1) intersects the boundary of dom F) is not explicitly dependent on the chain order m, but the chain-constrained structure is still required for the analysis in [37]. Finally, note that while an explicit bound (all constants known or verifiable) is presented in Scenario 1, it is a feasibility bound on the entire dom $\hat{F}_N$ instead of just $x^\star(\xi^{[N]})$, and is still under the chain-constrained domain assumption. Due to all these limitations, it is desirable to generalize the feasibility bound beyond the chain-constrained domain setting and with explicit constants.

3. Framework and Main Results. In this section we review the VC dimension framework and introduce our main results. In particular, we are interested in the VC dimension of a collection of subsets on Ξ. This approach gives bounds for any generic point in dom $\hat{F}_N$, the feasible region of the SAA, which in particular implies bounds for $x^\star(\xi^{[N]})$. Moreover, instead of looking at dom $f_\xi = \{x \in \mathcal{X} : f(\xi, x) < +\infty\}$, we investigate

$$H_x \triangleq \{\xi \in \Xi : f(\xi, x) < +\infty\} \quad \text{for } x \in \mathcal{X} \tag{3.1}$$

and define the class of subsets

$$\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}.$$

Consider the VC dimension of $\mathcal{H}$. The VC dimension is commonly used to describe the complexity of a collection of sets or functions [4, 29, 53], which is also known as the "hypothesis space" in computational learning theory. The concept applies to a class of subsets $\mathcal{H}$ (see [51]), and can be naturally generalized to binary functions. To define the VC dimension of a class of subsets $\mathcal{H}$, first we say a set of points $\{x_1, ..., x_d\}$ is "shattered" by $\mathcal{H}$ if any subset of $\{x_1, ..., x_d\}$ can be "picked out" by some element $C \in \mathcal{H}$, meaning that for any subset $D \subseteq \{x_1, ..., x_d\}$, there is a set $C \in \mathcal{H}$ such that $D \subseteq C$ and $(\{x_1, ..., x_d\} \setminus D) \cap C = \emptyset$. The VC dimension of $\mathcal{H}$ is defined to be the maximal cardinality of the sets it can shatter, denoted by $d_{VC}(\mathcal{H})$. For example, some classical results on the VC dimensions are the following:

- Positive intervals: If $\mathcal{H} = \big\{\{x \in \mathbb{R} : x \in [a, b]\} | 0 \le a \le b\big\}$, we have $d_{VC}(\mathcal{H}) = 2$.
- Affine hyperplanes (perceptrons): If
$$\mathcal{H} = \big\{\{x \in \mathbb{R}^d : a^T x + b \ge 0\} | a \in \mathbb{R}^d, b \in \mathbb{R}\big\},$$
we have $d_{VC}(\mathcal{H}) = d + 1$. If b is fixed to be 0, then $d_{VC}(\mathcal{H}) = d$.
- Convex sets: If $\mathcal{H} = \big\{C : C \subseteq \mathbb{R}^d \text{ and } C \text{ is convex}\big\}$ and $d \ge 2$, then $d_{VC}(\mathcal{H}) = +\infty$.

An important concept in computational learning theory tightly related to the VC dimension is *probably approximately correct* learning (PAC) (see, e.g., [29]). In this

context, the VC dimension of $\mathcal{H}$ can be used to derive bounds on the sample complexity needed to achieve a desired level of accuracy between "in-sample-error" and "generalization error" within class $\mathcal{H}$ (see, e.g. [4, 8, 29]). As it turns out, this type of result can transfer towards sample complexity results on the feasibility of SAA solutions.

Moreover, we note that the Ξ in (3.1) can be reparametrized and does not have to be viewed in $\mathbb{R}^r$ for fixed r defined in (1.1). For illustration, consider the following example.

Example 3.1. Suppose $x \in \mathcal{X} \subseteq \mathbb{R}$ and ξ is a random vector defined on $\mathbb{R}^r$ for some $r > 0$. Let $g(\cdot) : \mathbb{R}^r \to \mathbb{R}$ be a given function. Then, suppose $f(\xi, x) < +\infty$ if and only if $g(\xi) \cdot x \geq 1$. Then, H_x in (3.1) becomes

$$H_x = \{\xi : g(\xi) \cdot x \geq 1\} \subseteq \mathbb{R}^r \ \forall x \in \mathcal{X}.$$

On the other hand, if we define random variable $\xi' = g(\xi)$ on $\mathbb{R}$, then we can alternatively define

$$H'_x = \{\xi' : \xi' \cdot x \geq 1\} \subseteq \mathbb{R} \ \forall x \in \mathcal{X}.$$

Consequently, instead of fixing a canonical representation of ξ in (1.1), we sometimes utilize this flexibility to change representations for convenience.

Finally, we clarify the use of notation $[\cdot]$. For a positive integer q, $[q]$ denotes the set $\{1, ..., q\}$. For a vector $\boldsymbol{v} \in \mathbb{R}^q$, $[\boldsymbol{v}]_j$ denotes the jth component of $\boldsymbol{v}$ for $j \in [q]$, and $\|\boldsymbol{v}\|_0$ denotes the number of nonzero entries of $\boldsymbol{v}$.

3.1. Main Result. We now present our main theorem on SAA feasibility and its proof.

THEOREM 3.2. *Let $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ be the class of subsets defined in (3.1) and $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ be IID samples from $\mathbb{P}$ (consequently $\xi^{[N]} \sim \mathbb{P}^N$). Assume $\mathcal{H}$ has finite VC dimension d (i.e., $d_{VC}(\mathcal{H}) = d < +\infty$). Moreover, assume the feasible region of SAA (i.e., dom $\hat{F}_N = \cap_{i=1}^N \{x \in \mathcal{X} : f(\xi_i, x) < +\infty\}$) is nonempty with probability 1 and $x^\star(\xi^{[N]}) \in \mathcal{X}$ is the output of any algorithm that is guaranteed to be within the feasible region of SAA. Then, if*

$$N \geq \frac{4}{\epsilon}\left(d \log\left(\frac{12}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right), \tag{3.2}$$

we have, for any $0 < \delta, \epsilon < 1$,

$$\mathbb{P}^N\left(\sup_{x\in dom\ \hat{F}_N} V(x) > \epsilon\right) \leq \delta, \tag{3.3}$$

[1] *and consequently*

$$\mathbb{P}^N\left(V(x^\star(\xi^{[N]})) > \epsilon\right) \leq \delta. \tag{3.4}$$

[1] The measurability of (3.3) and (3.4) is discussed in the appendix.

Proof. We first adopt the following notations from [4]. For $x \in \mathcal{X}$, define $h_x : \Xi \to \{0,1\}$ as

$$h_x(\xi) \triangleq \mathbf{1}_{H_x}(\xi) = \begin{cases} 1 & \text{when } f(\xi,x) < +\infty, \\ 0 & \text{when } f(\xi,x) = +\infty, \end{cases}$$

and denote

$$\mathrm{H} \triangleq \{h_x\}_{x\in\mathcal{X}} \cup \{\mathbf{1}_\Xi(\cdot)\}$$

to be a hypothesis space as in the PAC learning context. Now, given a "target concept" $t \in \mathrm{H}$, the error of a hypothesis $h \in \mathrm{H}$ is defined in [4] as

$$\mathrm{er}_{\mathbb{P}}(h) \triangleq \mathbb{P}(\{\xi \in \Xi : h(\xi) \neq t(\xi)\}). \tag{3.5}$$

Building on (3.5), [4] further defines

$$\begin{aligned} \mathrm{H}[\xi^{[N]}] &= \{h \in \mathrm{H} \mid h(\xi_i) = t(\xi_i)\ \forall i \in [N]\}, \\ B_\epsilon &= \{h \in \mathrm{H} \mid \mathrm{er}_{\mathbb{P}}(h) \geq \epsilon\}. \end{aligned} \tag{3.6}$$

It is straightforward to see that H has the same VC dimension as $\mathcal{H}$ ($d_{VC}(\mathcal{H}) = d$). Thus, given H with VC dimension $d < +\infty$ and let the target function be

$$t(\cdot) = \mathbf{1}_\Xi(\cdot) \in \mathrm{H}, \tag{3.7}$$

Theorem 8.4.1 in [4] states that if

$$N \geq \frac{4}{\epsilon}\left(d\log\left(\frac{12}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right),$$

we have

$$\mathbb{P}^N(\{\xi^{[N]} \mid \mathrm{H}[\xi^{[N]}] \cap B_\epsilon \neq \emptyset\}) < \delta. \tag{3.8}$$

Given samples $\xi^{[N]}$, we know from (3.6) and (3.7) that the feasible region of SAA is

$$\begin{aligned} \mathrm{dom}\ \hat{F}_N &= \cap_{i=1}^N \{x \in \mathcal{X} \mid f(\xi_i,x) < +\infty\} \\ &= \{x \in \mathcal{X} \mid h_x(\xi_i) = 1,\ \forall i \in [N]\} \\ &= \{x \in \mathcal{X} \mid h_x \in \mathrm{H}[\xi^{[N]}]\} \end{aligned} \tag{3.9}$$

and we also know from (3.5) and (3.6) that,

$$V(x) \geq \epsilon \iff \mathrm{er}_{\mathbb{P}}(h_x) \geq \epsilon \iff h_x \in B_\epsilon \tag{3.10}$$

for any $x \in \mathcal{X}$. It then follows from (3.9) and (3.10) that, given samples $\xi^{[N]}$,

$$\{x \in \mathrm{dom}\ \hat{F}_N \mid V(x) \geq \epsilon\} \neq \emptyset \implies \mathrm{H}[\xi^{[N]}] \cap B_\epsilon \neq \emptyset. \tag{3.11}$$

Thus, if $x^\star(\xi^{[N]}) \in \mathrm{dom}\ \hat{F}_N$ with probability 1, it follows from (3.8) and (3.11) that

$$\begin{aligned} \mathbb{P}^N\left(\mathrm{H}[\xi^{[N]}] \cap B_\epsilon = \emptyset\right) > 1-\delta &\implies \mathbb{P}^N\left(\{x \in \mathrm{dom}\ \hat{F}_N \mid V(x) \geq \epsilon\} = \emptyset\right) > 1-\delta \\ &\implies \mathbb{P}^N\left(\sup_{x\in\mathrm{dom}\ \hat{F}_N} V(x) \leq \epsilon\right) > 1-\delta \\ &\implies \mathbb{P}^N\left(V(x^\star(\xi^{[N]})) \leq \epsilon\right) > 1-\delta \end{aligned}$$

which concludes the proof. ∎

Remark 3.3. First, in the proof of Theorem 3.2, the sample complexity in (3.2) comes from Theorem 8.4.1 of [4] and provides an $O(\frac{d}{\epsilon}\log(\frac{1}{\epsilon}) + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ bound. It is worth noting that a better sample complexity of $O(\frac{d}{\epsilon} + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ can be achieved by recent breakthroughs of [23, 48]. We choose to present the result from Theorem 8.4.1 in [4] because it is more concise and explicit. Nevertheless, under our framework, a better bound is indeed obtainable. Second, as shown in the proof, the feasibility result of the theorem holds not just for the optimal solution of SAA, but also for any generic point within the feasible region of SAA. In other words, Theorem 3.2 holds for any algorithm that can output a solution $x^\star(\xi^{[N]})$ (not necessarily the optimal solution of the SAA) in the feasible region of SAA with probability 1. This observation is particularly important when the considered SAA problem is nonconvex (i.e., mixed-integer programming), solvable up to a local optimum, or requires approximations algorithms.

There are several advantages when applying Theorem 3.2 to bound the feasibility of SAA solutions: (1) Other than a finite VC dimension, it does not rely on specific assumptions regarding the structures of (1.1) and (1.2). As we shall see later, even when the chain-constrained domain condition in [37] becomes restrictive, our analysis based on VC dimension would remain effective. (2) Our bound is explicit and computable with no hidden constants. (3) One might argue the generality of Theorem 3.2 would come at a cost of higher sample complexity. However, as we shall see, this is not the case even when we compare our bounds with some of the best-known ones. Moreover, thanks to (1) above, Theorem 3.2 yields some previously unattainable results.

While Theorem 3.2 is a result on sample complexity, it is convenient to convert it into an asymptotic rate of convergence with N. The infeasibility of the SAA solution still decreases exponentially as in [37], except the rate is now explicit. We summarize it in a corollary.

COROLLARY 3.4. *Under the same conditions as Theorem 3.2,*

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq 2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^d.$$

To compare with the existing results in [37], we note that direct comparisons on sample complexity are possible only when the rate of convergence is explicit, which only applies to Scenario 1 in [37] (summarized in Section 2). Moreover, it is shown in [14] that a relatively tight sufficient condition for N to satisfy

$$\sum_{k=0}^{m-1}\binom{N}{k}\epsilon^i(1-\epsilon)^{N-i} \leq \delta,$$

is

$$N \geq \frac{e}{e-1}\frac{1}{\epsilon}\left(m - 1 + \log\left(\frac{1}{\delta}\right)\right), \tag{3.12}$$

which provides a tight bound on sample complexity and we shall make use of it later.

4. Applications. In this section, we apply Theorem 3.2 in several problems of practical interests and compare with established results. For the examples formulated in this section, one can check that $f(\xi, x)$ satisfies the general conditions imposed in the introductory section (below (1.1)). Throughout the proofs, we make use of the following lemmas.

LEMMA 4.1 (Theorem 1.1 from [51]). *Given classes of subsets $\mathcal{C}_1$, $\mathcal{C}_2$,..., $\mathcal{C}_m$ with $d_{VC}(\mathcal{C}_j) = d_j < +\infty$, define*

$$\sqcap_{j=1}^m \mathcal{C}_j \triangleq \{\cap_{j=1}^m C_j : C_j \in \mathcal{C}_j, j = 1, ..., m\},$$
$$\sqcup_{j=1}^m \mathcal{C}_j \triangleq \{\cup_{j=1}^m C_j : C_j \in \mathcal{C}_j, j = 1, ..., m\},$$

and $d = \sum_{j=1}^m d_j$. Then,

$$\max(d_{VC}(\sqcap_{j=1}^m \mathcal{C}_j), d_{VC}(\sqcup_{j=1}^m \mathcal{C}_j)) \leq \frac{e}{(e-1)\log 2} d \log\Big(\frac{e}{\log 2} m\Big).$$

We also use a key result on the upper bound of the VC dimension for sets determined by finite-dimensional function spaces. Recall that $\mathcal{G}$ is a vector space of functions if $a_1 f_1 + a_2 f_2 \in \mathcal{G}$ for any $a_1, a_2 \in \mathbb{R}$ and $f_1, f_2 \in \mathcal{G}$. The lemma below comes directly from [20]. For a more concise proof, one is also referred to Lemma 2.6.15 in [52].

LEMMA 4.2. *Given arbitrary space $\mathcal{S}$, let $\mathcal{G}$ be a finite-dimensional vector space of functions $g(\cdot) : \mathcal{S} \to \mathbb{R}$ with dim $\mathcal{G} < +\infty$. Then, the class of sets*

$$\mathcal{H} = \{\{s \in \mathcal{S} : g(s) \geq 0\}\}_{g \in \mathcal{G}}$$

has VC dimension at most dim $\mathcal{G}$.

Note that since $\mathcal{G}$ is a vector space, $\mathcal{H}$ can then also be characterized as

$$\{\{s \in \mathcal{S} : g(s) \leq 0\}\}_{g \in \mathcal{G}}.$$

The following example will be used later.

Example 4.3. According to Lemma 4.2, $\mathcal{U} = \{\{(y, z) \in \mathbb{R}^d \times \mathbb{R} : y^T x \leq z\}\}_{x \in \mathbb{R}^d}$ has VC dimension at most d. In fact, the VC dimension for $\mathcal{U}$ is equal to d (see [18] or [20]).

Finally, related to the reparametrization discussion in Example 3.1, we have the following result.

LEMMA 4.4. *Suppose $\mathcal{H} = \{H_x\}_{x \in \mathcal{X}}$ is a collection of subsets on Ξ. Given $f(\cdot) : \Xi' \to \Xi$ and $H \subseteq \Xi$, define*

$$f^{-1}(H) \triangleq \{\xi' \in \Xi', f(\xi') \in H\}$$

with the convention $f^{-1}(\emptyset) = \emptyset$ and define

$$\mathcal{H}' \triangleq \{f^{-1}(H_x)\}_{x \in \mathcal{X}}.$$

Then, we have $d_{VC}(\mathcal{H}') \leq d_{VC}(\mathcal{H})$.

Proof. For any set of points $\{\xi_i'\}_{i \in [n]} \subseteq \Xi'$ shattered by $\mathcal{H}'$, let $\xi_i = f(\xi_i')$ and consider $\{\xi_i\}_{i \in [n]}$. If we can show $\{\xi_i\}_{i \in [n]}$ is shattered by $\mathcal{H}$, then we must have $n \leq d_{VC}(\mathcal{H})$. Since n is arbitrary, we have

$$\sup\{n \mid H' \subseteq \Xi', |H'| = n \text{ and } H' \text{ is shattered by } \mathcal{H}'\} \leq d_{VC}(\mathcal{H})$$

and consequently $d_{VC}(\mathcal{H}') \leq d_{VC}(\mathcal{H})$.

To show $\{\xi\}_{i\in[n]}$ is shattered by $\mathcal{H}$, pick any $I \subseteq [n]$, and we want to show there exists $x \in \mathcal{X}$ such that $\{\xi_i\}_{i\in I} \subseteq H_x$ and $\{\xi_i\}_{i\in[n]\cap I^c} \cap H_x = \emptyset$. Now, recall $\{\xi'\}_{i\in[n]}$ is shattered by $\mathcal{H}'$, so there exists $x \in \mathcal{X}$ such that $\{\xi'_i\}_{i\in I} \subseteq f^{-1}(H_x)$ and $\{\xi'_i\}_{i\in[n]\cap I^c} \cap f^{-1}(H_x) = \emptyset$. This concludes the proof, since for such $x \in \mathcal{X}$, we have ∎

$$\begin{cases} \xi'_i \in f^{-1}(H_x) \implies \xi_i = f(\xi'_i) \in H_x & \text{for } i \in I, \\ \xi'_i \notin f^{-1}(H_x) \implies \xi_i = f(\xi'_i) \notin H_x & \text{for } i \in [n] \cap I^c. \end{cases}$$

4.1. Two-Stage Stochastic Programming. One of the main motivating examples in studying SAA feasibility, mentioned in [15, 36, 37], is the two-stage stochastic program without relatively complete recourse ([15, 36, 37] focus on the linear recourse). This section is divided into three subsections to discuss feasibility in two-stage stochastic programming. The first subsection considers the general case with only continuous variables, the second subsection considers mixed-integer variables, and the third subsection considers low-dimensional structures. Finally, we note that, while we focus on two-stage stochastic programming, these feasibility results can be generalized to multistage stochastic programs using stagewise independence of the data process. For a detailed discussion, see Section 4 of [37].

4.1.1. General Case. We first consider the following general case. For clarity and generality, we rewrite (1.1) to define $f(\xi, x)$ as follows:

$$\inf_{x\in\mathcal{X}} F(x) \triangleq f_0(x) + \mathbb{E}[f(\xi, x)] \tag{4.1}$$

with $f_0(x)$ being a deterministic function and

$$\begin{aligned} f(\xi, x) \triangleq & \inf_y g(\xi, y) \\ \text{s.t.} \quad & W_\xi y + T_\xi x = h_\xi, \\ & y \geq 0, \end{aligned} \tag{4.2}$$

where we only assume $g(\xi, \cdot)$ is convex and finite everywhere, almost surely $\forall \xi \in \Xi$. Typical classes of problems that fall under this formulation include (see [36])

- two-stage stochastic linear programming: $f_0(x) = c^T x$ is linear, $\mathcal{X} = \{x \in \mathbb{R}^n : Ax \leq b\}$ is polyhedral, and $g(\xi, y) = q(\xi)^T y$ is linear in y;
- two-stage stochastic quadratic-linear programming: $f_0(x) = \frac{1}{2}x^T Q x + c^T x$ is quadratic in x, $\mathcal{X} = \{x \in \mathbb{R}^n : Ax \leq b\}$ is polyhedral, and $g(\xi, y) = q(\xi)^T y$ is linear in y;
- two-stage stochastic quadratic-quadratic programming: $f_0(x) = \frac{1}{2}x^T Q x + c^T x$ is quadratic in x, $\mathcal{X} = \{x \in \mathbb{R}^n : Ax \leq b\}$ is polyhedral, and $g(\xi, y) = \frac{1}{2}y^T P(\xi) y + q(\xi)^T y$ is quadratic in y.

Suppose $g(\xi, y)$ takes the general form above. To derive feasibility results on the SAA solution, [37] assumes there are only finitely many distinct values for W_ξ or T_ξ, i.e., $|\{W_\xi\}| = p$ and $|\{T_\xi\}| = q$, where $\{p, q\} \subseteq \mathbb{Z}^+$. By Farkas' lemma, we have $\{y \geq 0 : W_\xi y + T_\xi x = h_\xi\}$ is nonempty (i.e., $f(\xi, x) < +\infty$) if and only if $a^T(h_\xi - T_\xi x) \geq 0$ for all a such that $a^T W_\xi \geq 0$. Consequently, as shown in [37], suppose we let $W_i, i \in [p]$, be the ith distinct element in $\{W_\xi\}$, let $\{a_{ij}\}_{j\in J_i}$ be the set of nonequivalent extreme rays of polyhedral cone $\mathcal{C}_i = \{a : a^T W_i \geq 0\}$ with J_i being the index set for these extreme rays of $\mathcal{C}_i$, and let $I(\cdot) : \Xi \to [p]$ be the indexing function such that $I(\xi) = i$ when $W_\xi = W_i$. Then, we have

$$\text{dom } f_\xi = \{x \in \mathcal{X} \mid a_{I(\xi)j}^T T_\xi x \leq a_{I(\xi)j}^T h_\xi, \forall j \in J_{I(\xi)}\}, \tag{4.3}$$

where we assume the cones $\mathcal{C}_i$ are pointed, so they can be minimally generated by a unique set of nonequivalent extreme rays. Otherwise, a general polyhedral cone is still finitely generated according to Weyl–Minkowski's theorem, but the minimal set of generators may not be unique. Without of loss of generality (as in most linear programming problems), we assume the rows of W_i are linearly independent, which implies the cone $\mathcal{C}_i$ is pointed (see, e.g., [50]). Now, note that (4.3) allows [37] to use the so-called chain-constrained domain structure. Here, a chain-constrained domain is defined as follows.

DEFINITION 4.5. *A collection of functions $\{f(\xi,\cdot)\}_{\xi\in\Xi}$ has a chain-constrained domain of order m if there exist m chains $\{U_k^\xi\}_{\xi\in\Xi}$ and*

$$dom\ f_\xi = \bigcap_{k=1}^{m} U_k^\xi,$$

where a collection of sets $\{U^\omega\}_{\omega\in I}$ is a chain if for any $\omega_1,\omega_2\in I$, we have either $U_{\omega_1}\subseteq U_{\omega_2}$ or $U_{\omega_2}\subseteq U_{\omega_1}$.

It is shown in [37] that dom f_ξ in (4.3) is a chain-constrained domain of order $m = q\sum_{i=1}^{p}|J_i|$. Consequently, Scenario 1 in [37] can be applied to show that

$$\mathbb{P}^N(V(x^\star(\xi^{[N]}))>\epsilon)\le\sum_{k=0}^{m-1}\binom{N}{k}\epsilon^k(1-\epsilon)^{N-k},$$

which has a sample complexity

$$\frac{e}{e-1}\frac{1}{\epsilon}\left(m-1+\log\left(\frac{1}{\delta}\right)\right) \tag{4.4}$$

for achieving $\mathbb{P}^N(V(x^\star(\xi^{[N]}))>\epsilon)\le\delta$ according to (3.12).

Notice a necessary assumption made in [37] is that only finitely many distinct values for W_ξ or T_ξ are allowed. However, using Theorem 3.2, we can get a different sample complexity and concentration bounds, even when the cardinalities of $|\{W_\xi\}|$ and $|\{T_\xi\}|$ are infinite. In particular, we first show feasibility results on (4.2) when $|\{W_\xi\}|$ and $|\{T_\xi\}|$ are finite in Corollary 4.6. Then, we extend the result to the case when $|\{W_\xi\}|$ and $|\{T_\xi\}|$ are infinite in Corollary 4.7.

COROLLARY 4.6. *Consider (4.2) and assume there are only finitely many distinct values for W_ξ or T_ξ , i.e., $|\{W_\xi\}|=p$ and $|\{T_\xi\}|=q$, where $\{p,q\}\subseteq\mathbb{Z}^+$. Let $\xi^{[N]}=\{\xi_1,...,\xi_N\}$ be IID samples from $\mathbb{P}$ (consequently $\xi^{[N]}\sim\mathbb{P}^N$), and $x^\star(\xi^{[N]})$ be the output of any algorithm that is guaranteed to be within the feasible region of SAA. Then, if*

$$N\ge\frac{4}{\epsilon}\left(d_{VC}\log\left(\frac{12}{\epsilon}\right)+\log\left(\frac{2}{\delta}\right)\right), \tag{4.5}$$

where $|J|=\max_{i\in[p]}|J_i|$ and

$$d_{VC}=\left(\frac{e}{(e-1)\log 2}|J|(n+1)\log\left(\frac{e}{\log 2}\cdot|J|\right)\right), \tag{4.6}$$

we have

$$\mathbb{P}^N(V(x^\star(\xi^{[N]}))>\epsilon)\le\delta$$

for any $0 < \delta, \epsilon < 1$. *Equivalently, in terms of convergence rate, we have*

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq 2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^{d_{VC}}. \tag{4.7}$$

Proof. Define $I(\cdot) : \Xi \to [p]$ as the indexing function such that $I(\xi) = i$ when $W_\xi = W_i$. Then, it follows that $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ defined in (3.1) consists of

$$H_x = \{\xi : a_{I(\xi)j}^T T_\xi x \leq a_{I(\xi)j}^T h_\xi, \forall j \in J_{I(\xi)}\},$$

where $\{a_{ij}\}_{j\in J_i}$ is the set of nonequivalent extreme rays of the polyhedral cone

$$\{a : a^T W_i \geq 0\}.$$

Define $|J| = \max_{i\in[q]} |J_i|$ and, given $\xi \in \Xi$, define $\{(y_{\xi j}, z_{\xi j})\}_{j\in|J|} \subseteq \mathbb{R}^n \times \mathbb{R}$ to be

$$y_{\xi j}^T \triangleq \begin{cases} a_{I(\xi)j}^T T_\xi & \text{for } 1 \leq j \leq |J_{I(\xi)}|, \\ \mathbf{0} & \text{for } |J_{I(\xi)}| < j \leq |J|, \end{cases}$$

$$z_{\xi j} \triangleq \begin{cases} a_{I(\xi)j}^T h_\xi & \text{for } 1 \leq j \leq |J_{I(\xi)}|, \\ 0, & \text{for } |J_{I(\xi)}| < j \leq |J|. \end{cases}$$

Then, define $y_\xi^T = (y_{\xi 1}^T, y_{\xi 2}^T, ..., y_{\xi|J|}^T) \in \mathbb{R}^{|J|n}$ and $z_\xi = (z_{\xi 1}, z_{\xi 2}, ..., z_{\xi|J|}) \in \mathbb{R}^{|J|}$. Moreover, for $j \in [|J|]$, define $v_j(\cdot) : \mathbb{R}^n \to \mathbb{R}^{|J|n}$ to be

$$[v_j(x)]_i = \begin{cases} [x]_i & \text{for } (j-1)n + 1 \leq i \leq jn, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we can redefine

$$H_x = \bigcap_{j=1}^{|J|} \{(y_\xi, z_\xi) : y_\xi^T v_j(x) \leq [z_\xi]_j\}. \tag{4.8}$$

Fix $j \in [|J|]$, let $e_j \in \mathbb{R}^{|J|}$ be the vector with 1 in the jth component and 0 otherwise. Define a class of function $\mathcal{G}_j = \{g_{(x,c)}(\cdot)\}_{(x,c)\in\mathbb{R}^n\times\mathbb{R}}$ on $\mathbb{R}^{|J|(n+1)}$ such that, given $(y, z) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|}$,

$$g_{(x,c)}((y, z)) = [y, z]^T \begin{bmatrix} -v_j(x) \\ c \cdot e_j \end{bmatrix}.$$

It is straightforward to check $\mathcal{G}_j$ is a finite-dimensional vector space of functions with $\dim \mathcal{G}_j \leq n + 1$. Then, according to Lemma 4.2, the VC dimension of

$$\{\{(y, z) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} : g_{x,c}((y, z)) \geq 0\}\}_{(x,c)\in\mathbb{R}^n\times\mathbb{R}}$$

is at most $n + 1$. Moreover, by letting $(x, c) = \mathbf{0}$, it can be seen that the above collection of sets includes the set $\mathbb{R}^{|J|n} \times \mathbb{R}^{|J|}$. Consequently, as a smaller collection of sets, the VC dimension of

$$\{\{(y, z) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} : g_{x,1}((y, z)) \geq 0\}\}_{x\in\mathcal{X}} \cup \{\mathbb{R}^{|J|n} \times \mathbb{R}^{|J|}\}$$

is at most $n+1$. Thus, for each $j \in [|J|]$, it follows from Lemma 4.4 that the VC dimension of

$$\mathcal{U}_j = \{\{(y_\xi, z_\xi) : y_\xi^T v_j(x) \le [z_\xi]_j\}\}_{x\in\mathcal{X}} \cup \{\Xi\}$$

is at most $n+1$. Finally, it follows from Lemma 4.1 and (4.8) that

$$d_{VC}(\mathcal{H}) \le d_{VC}(\sqcap_{j=1}^{|J|}\mathcal{U}_j) \le \frac{e}{(e-1)\log 2}|J|(n+1)\log\left(\frac{e}{\log 2}\cdot|J|\right).$$

The analogous sample complexity and convergence rate follow from Theorem 3.2. ∎

Note that Corollary 4.6 does not require a convexity assumption on g_ξ or distributional assumptions on the random variables W_ξ and T_ξ. In fact, we can further extend our result to the case when $|\{W_\xi\}|$ and $|\{T_\xi\}|$ are infinite. In particular, the same proof can be applied as long as

$$|J| = \max_{\xi\in\Xi}\{ \text{ \# of extreme rays for the cone } \{a : a^T W_\xi \ge 0\}\}$$

is finite. However, it is known that the number of nonequivalent extreme rays of a polyhedral cone $\{a : a^T W \ge 0\}$ is finite and can be bounded by a term $\binom{n_1}{m_1-1}$, which only involves parameters m_1 and n_1 for $W \in \mathbb{R}^{m_1\times n_1}$ (similarly to the bound on the number of extreme points, $m_1 - 1$ linearly independent "constraints" are chosen; for details see [26, 45, 50]). Thus, as long as $\{W_\xi\} \subseteq \mathbb{R}^{m_1\times n_1}$ and m_1, n_1 are bounded almost surely, we have $|J| < +\infty$ regardless of the cardinalities of $\{W_\xi\}$. We summarize this in another corollary.

COROLLARY 4.7. *Consider (4.2) and assume $|J| < \infty$, where*

$$|J| = \max_{\xi\in\Xi}\{ \text{ \# of extreme rays for the cone } \{a : a^T W_\xi \ge 0\}\}.$$

Let $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ be IID samples from $\mathbb{P}$ (consequently $\xi^{[N]} \sim \mathbb{P}^N$), and $x^\star(\xi^{[N]})$ be the output of any algorithm that is guaranteed to be within the feasible region of SAA. Then the result of Corollary 4.6 still holds.

Proof. Let $\mathcal{A}_\xi$ be the set of nonequivalent extreme rays of the polyhedral cone

$$\{a : a^T W_\xi \ge 0\}.$$

Observe $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ defined in (3.1) consists of

$$H_x = \{\xi : a^T T_\xi x \le a^T h_\xi, \forall a \in \mathcal{A}_\xi\}. \tag{4.9}$$

For all $\xi \in \Xi$, since $|\mathcal{A}_\xi| \le |J| < +\infty$, we can label the elements in $\mathcal{A}_\xi$ by $\{a_{\xi j}\}_{j\in[|\mathcal{A}_\xi|]}$. Then, define $\{(y_{\xi j}, z_{\xi j})\}_{j\in|J|}$ as

$$y_{\xi j}^T = \begin{cases} a_{\xi j}^T T_\xi & \text{for } 1 \le j \le |\mathcal{A}_\xi|, \\ \mathbf{0} & \text{for } |\mathcal{A}_\xi| < j \le |J|, \end{cases}$$

$$z_{\xi j} = \begin{cases} a_{\xi j}^T h_\xi & \text{for } 1 \le j \le |\mathcal{A}_\xi|, \\ 0 & \text{for } |\mathcal{A}_\xi| < j \le |J|. \end{cases}$$

The rest of the proof follows exactly as in Corollary 4.6. ∎

Compared with our bound (4.5), the chain-constrained bound (4.4) relies on the order of the chain $m = q\sum_{i=1}^{p}|J_i|$. If the cardinality of $\{W_\xi\}$ or $\{T_\xi\}$ gets large (i.e., $qp \gg n$), or potentially infinite (for a continuous random variable), then the bound in (4.4) with a sample complexity of $O(\frac{qp|J|}{\epsilon} + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ becomes loose or even inapplicable. On the other hand, the VC bound (4.5) with a sample complexity $O(\frac{|J|n}{\epsilon}\log|J|\log(\frac{1}{\epsilon}) + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ maintains the same dependence on the dimension n regardless of the support of W_ξ or T_ξ. Moreover, if we use the PAC bound from [23, 48] as mentioned in Remark 3.3, the bound can be improved to $O(\frac{|J|n}{\epsilon}\log|J| + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$. Finally, in both bounds, the term $|J|$ appears. However, as mentioned previously, an explicit bound for $|J|$ in terms of m_1, n_1 can be obtained, where $\{W_\xi\} \subseteq \mathbb{R}^{m_1\times n_1}$. We omit it here as it is not essential for our comparison. Finally, the bound in Scenario 3 of [37] also applies to (4.2) and is not limited by the order of the chain structure. However, the bound there is not explicitly computable due to the hidden term β.

The dependence on the order of the chain m is also discussed in [15]. Using ideas similar to the scenario approximation of chance-constrained problems in [11, 40, 12], as well as specific properties of linear programming (e.g., existence of basic optimal solutions), [15] is able to provide a sample complexity for two-stage stochastic linear programming independent of the cardinalities of $\{W_\xi\}$ or the order of the chain. Nonetheless, the derivation of our bound in (4.5) does not depend on the linearity of the optimization problem and hence is not limited to two-stage stochastic programming with linear recourse. More specifically, in [15], the first stage $\mathcal{X}$ is defined by linear constraints $Ax = b$ for some $A \in \mathbb{R}^{m\times n}$ and the second-stage problem bears a linear objective $q(\xi)^T y$. In contrast, our bound is valid for general $\mathcal{X}$ in the first stage and $g(\xi, y)$ in the second-stage problem in (4.2). That being said, the bound derived in [15] has notable strengths in the linear case, in terms of the dependence on problem parameters, gained via a more efficient exploitation of the linear structure. Specifically, the sample complexity in [15] is (adapted to the notation in this paper)

$$O\left(\frac{1}{\epsilon}\left(nn_1\left(\log\left(\frac{m_1}{n_1+1}\right)+1\right)+n\left(\log\left(\frac{m}{n}+2\right)+\log\left(\frac{1}{\epsilon}\right)+1\right)+\log(\frac{1}{\delta})\right)\right), \tag{4.10}$$

which has better dependence on m_1, n_1, as the dependence on $|J|$ in (4.5) is $\binom{n_1}{m_1-1}$ in the worst case. Nonetheless, (4.5) has a similar dependence on n as the bound in (4.10), and does not depend on m in (4.10) at all. Omitting the dependence on these problem size parameters (e.g., constants based on n, m, m_1, n_1, and $|J|$), the bound derived in [15] is of order $O(\frac{1}{\epsilon}\log(\frac{1}{\delta}) + \frac{1}{\epsilon}\log(\frac{1}{\epsilon}))$, which of the same order as the bound (4.5). Moreover, (4.5) can be slightly improved to be of order $O(\frac{1}{\epsilon}\log(\frac{1}{\delta}) + \frac{1}{\epsilon})$ bound based on Remark 3.3.

4.1.2. Mixed-Integer Programming. The SAA method has also been applied in two-stage stochastic programming with mixed-integer recourse [2, 3, 31, 7]. However, due to the presence of an integer variable, currently known results based on (4.1) regarding the feasibility of SAA solutions do not apply. In this section, we provide an original feasibility bound in this case. We consider the following two-stage stochastic mixed-integer programming where $\mathcal{X} \subseteq \mathbb{R}^{n-p} \times \mathbb{Z}^p$ is allowed to contain integer components in the first stage,

$$\inf_{x\in\mathcal{X}} F(x) \triangleq f_0(x) + \mathbb{E}[f(\xi, x)],$$

and the second stage is a mixed-integer program (MIP),

$$
\begin{aligned}
f(\xi, x) \triangleq \inf_{y} \quad & g(\xi, y, y_0) \\
\text{s.t.} \quad & W_\xi y + W_\xi^0 y_0 + T_\xi x = h_\xi, \\
& y \in \mathbb{R}_+^{n'}, y_0 \in \mathcal{Z} \subseteq \mathbb{Z}_+^{p'},
\end{aligned} \tag{4.11}
$$

for given $n', p' \in \mathbb{Z}_+$. Here $g(\xi, y, y_0)$ can be a general function as in (4.2), although for much of theoretical and practical interest (also applicability), it is assumed to be in linear form $g(\xi, y, y_0) = q(\xi)^T y + q_0(\xi)^T y_0$. Moreover, most literature also assumes relatively complete recourse by fixing a deterministic recourse matrix (i.e., $W_\xi = W$ and $W_\xi^0 = W^0$ with probability 1) such that $\{\boldsymbol{v} \in \mathbb{R}_+^{n'} \times \mathbb{Z}_+^{p'} : [W|W^0]\boldsymbol{v} = \boldsymbol{w}\}$ is nonempty for all relevant $\boldsymbol{w}$. Consequently, the feasibility of the SAA solution for two-stage stochastic integer programming without relatively complete recourse has rarely been considered. In fact, due to the general nonconvex and discontinuous nature of MIPs, specialized approximation or iterative algorithms are usually required and the solutions are no longer guaranteed to be optimal. However, even without relatively complete recourse or optimality guarantee, as mentioned in Remark 3.3, as long as the solutions output from such algorithms are within the SAA feasible region with probability 1, the feasibility result from Theorem 3.2 still holds. Recall we have assumed the set $\{x : x \in \mathcal{X} \text{ and } F(x) < +\infty\}$ is nonempty and the SAA feasible region is nonempty with probability 1.

Under the setting of Theorem 3.2, it is possible to provide a feasibility bound for (4.11) when $|\mathcal{Z}| < +\infty$. This condition is satisfied when y_0 is restricted to be binary as in [7] (i.e., $y_0 \in \{0,1\}^{p'}$). On the other hand, if the solutions are polynomially bounded by the size of data (e.g., integer linear programming [9]), then it is also possible to consider solving (4.11) in a finite, although possibly large bounded set $\mathcal{Z} \subseteq \mathbb{Z}_+^{p'}$, thus satisfying $|\mathcal{Z}| < +\infty$.

COROLLARY 4.8. *Consider* (4.11). *Suppose* $|\mathcal{Z}| < +\infty$ *and* $|J| < +\infty$, *where*

$$
|J| = \max_{\xi \in \Xi} \{ \text{ \# of extreme rays for the cone } \{a : a^T W_\xi \geq 0\}\}.
$$

Then, let $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ *be IID samples from* $\mathbb{P}$ *(consequently* $\xi^{[N]} \sim \mathbb{P}^N$*), and* $x^\star(\xi^{[N]})$ *be the output of any algorithm that is guaranteed to be within the feasible region of SAA. Then, if*

$$
N \geq \frac{4}{\epsilon}\left(d_{VC} \log\left(\frac{12}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right), \tag{4.12}
$$

where

$$
d_{VC} = \left(\frac{e}{(e-1)\log 2}\right)^2 |\mathcal{Z}||J|(n+2)\log\left(\frac{e|J|}{\log 2}\right)\log\left(\frac{e|\mathcal{Z}|}{\log 2}\right),
$$

then we have $\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq \delta$ *for any* $0 < \delta, \epsilon < 1$. *Equivalently, in terms of convergence rate, we have*

$$
\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq 2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^{d_{VC}}. \tag{4.13}
$$

Proof. Let $\mathcal{A}_\xi$ be the set of nonequivalent extreme rays of the polyhedral cone

$$\{a : a^T W_\xi \geq 0\}.$$

Using Farkas' lemma as in (4.9), we construct elements of $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ defined in (3.1) as

$$H_x = \bigcup_{y_0 \in \mathcal{Z}} \{\xi : a^T(T_\xi x + W^0_\xi y_0) \leq a^T h_\xi \ \forall a \in \mathcal{A}_\xi\}$$

for $(x, y_0) \in \mathcal{X} \times \mathcal{Z}$. For all $\xi \in \Xi$, since $|\mathcal{A}_\xi| \leq |J| < +\infty$, we can label the elements in $\mathcal{A}_\xi$ by $\{a_{\xi j}\}_{j\in[|A_\xi|]}$. Then, given $\xi \in \Xi$, define $\{(y_{\xi j}, z_{\xi j}, w_{\xi j})\}_{j\in|J|} \subseteq \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^{p'}$ as

$$y^T_{\xi j} = \begin{cases} a^T_{\xi j} T_\xi & \text{for } 1 \leq j \leq |\mathcal{A}_\xi|, \\ \mathbf{0} & \text{for } |\mathcal{A}_\xi| < j \leq |J|, \end{cases}$$

$$z_{\xi j} = \begin{cases} a^T_{\xi j} h_\xi & \text{for } 1 \leq j \leq |\mathcal{A}_\xi|, \\ 0 & \text{for } |\mathcal{A}_\xi| < j \leq |J|, \end{cases}$$

$$w^T_{\xi j} = \begin{cases} a^T_{\xi j} W^0_\xi & \text{for } 1 \leq j \leq |\mathcal{A}_\xi|, \\ \mathbf{0}, & \text{for } |\mathcal{A}_\xi| < j \leq |J|. \end{cases}$$

Define $y^T_\xi = (y^T_{\xi 1}, y^T_{\xi 2}, ..., y^T_{\xi|J|}) \in \mathbb{R}^{|J|n}$, $z_\xi = (z_{\xi 1}, z_{\xi 2}, ..., z_{\xi|J|}) \in \mathbb{R}^{|J|}$ and $w^T_\xi = (w^T_{\xi 1}, w^T_{\xi 2}, ..., w^T_{\xi|J|}) \in \mathbb{R}^{|J|p'}$. Moreover, for $j \in [|J|]$, define $v_j(\cdot) : \mathbb{R}^n \to \mathbb{R}^{|J|n}$, $u_j : \mathbb{Z}^{p'} \to \mathbb{Z}^{|J|p'}$ so that

$$[v_j(x)]_i = \begin{cases} [x]_i & \text{for } (j-1)n + 1 \leq i \leq jn\ , \\ 0 & \text{otherwise,} \end{cases}$$

$$[u_j(x)]_i = \begin{cases} [y_0]_i & \text{for } (j-1)p' + 1 \leq i \leq jp', \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can redefine

$$H_x = \bigcup_{y_0\in\mathcal{Z}} \bigcap_{j=1}^{|J|} \{(y_\xi, z_\xi, w_\xi) : y^T_\xi v_j(x) + w^T_\xi u_j(y_0) \leq [z_\xi]_j\}. \tag{4.14}$$

Given $j \in [|J|]$ and $y_0 \in \mathcal{Z}$, let $e_j \in \mathbb{R}^{|J|}$ be the vector with 1 in the jth component and 0 otherwise. Define a class of functions $\mathcal{G} = \{g_{(x,c_1,c_2)}(\cdot)\}_{(x,c_1,c_2)\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}}$ on $\mathbb{R}^{|J|(n+1+p')}$ such that, given $(y, z, w) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} \times \mathbb{R}^{|J|p'}$,

$$g_{(x,c_1,c_2)}((y, z, w)) = [y, z, w]^T \begin{bmatrix} -v_j(x) \\ c_1 \cdot e_j \\ -c_2 \cdot u_j(y_0) \end{bmatrix}.$$

It is straightforward to check $\mathcal{G}$ is a finite-dimensional vector space of functions with $\dim \mathcal{G} \le n+2$. Then, according to Lemma 4.2, the VC dimension of

$$\{\{(y,z,w) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} \times \mathbb{R}^{|J|p'} : g_{(x,c_1,c_2)}((y,z,w)) \ge 0\}\}_{(x,c_1,c_2)\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}}$$

is at most $n+2$. Moreover, by letting $(x,c_1,c_2) = \mathbf{0}$, it can be seen that the above collection of sets includes the set $\mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} \times \mathbb{R}^{|J|p'}$. Consequently, as a smaller collection of sets, the VC dimension of

$$\{\{(y,z,w) \in \mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} \times \mathbb{R}^{|J|p'} : g_{(x,1,1)}((y,z,w)) \ge 0\}\}_{x\in\mathcal{X}} \cup \{\mathbb{R}^{|J|n} \times \mathbb{R}^{|J|} \times \mathbb{R}^{|J|p'}\}$$

is also at most $n+2$. Thus, for each $j \in [|J|]$, it follows from Lemma 4.4 that the VC dimension of

$$\mathcal{U}_j^{y_0} = \{\{(y_\xi, z_\xi, w_\xi) : y_\xi^T v_j(x) + w_\xi^T u_j(y_0) \le [z_\xi]_j\}\}_{x\in\mathcal{X}} \cup \{\Xi\}$$

is at most $n+2$. Consequently, given $y_0 \in \mathcal{Z}$, it follows from Lemma 4.1 that

$$d_{VC}(\sqcap_{j=1}^{|J|} \mathcal{U}_j^{y_0}) \le \frac{e}{(e-1)\log 2}|J|(n+2)\log\Big(\frac{e}{\log 2}|J|\Big)$$

and then

$$d_{VC}\Big(\sqcup_{y_0\in\mathcal{Z}}\Big(\sqcap_{j=1}^{|J|}\mathcal{U}_j^{y_0}\Big)\Big) \le \Big(\frac{e}{(e-1)\log 2}\Big)^2 |\mathcal{Z}||J|(n+2)\log\Big(\frac{e|J|}{\log 2}\Big)\log\Big(\frac{e|\mathcal{Z}|}{\log 2}\Big).$$

Thus, for $\mathcal{H}$ defined in (4.14), we have

$$d_{VC}(\mathcal{H}) \le \Big(\frac{e}{(e-1)\log 2}\Big)^2 |\mathcal{Z}||J|(n+2)\log\Big(\frac{e|J|}{\log 2}\Big)\log\Big(\frac{e|\mathcal{Z}|}{\log 2}\Big).$$

The rest of the proof follows as in Corollary 4.6. ∎

As we can see, the portion of infeasible SAA solutions (not necessarily optimal) still decreases exponentially as the sample size N increases, although it is worth noting that the rate now depends on $|\mathcal{Z}|$ as well.

4.1.3. Low-Dimensional Models. There has been a growing literature of applying and analyzing SAA in high-dimensional stochastic programming by leveraging low-dimensional structures, e.g., sparsity, low-rankness [10, 34, 35]. Most of these results focus on the optimization accuracy of SAA as in [47], which compute a sample complexity of

$$N = O\Big(\frac{n}{\epsilon^2}\log\frac{1}{\epsilon} + \frac{1}{\epsilon^2}\log\frac{1}{\delta}\Big) \tag{4.15}$$

to guarantee

$$\mathbb{P}^N\Big(F(x^\star(\xi^{[N]}) - \inf_{x\in\mathcal{X}} F(x) > \epsilon\Big) \le \delta. \tag{4.16}$$

Typical results exploring low-dimensional structure aim to reduce the dependence of (4.15) on n, especially for $n \gg \frac{1}{\epsilon}$. For example, the main result on sample complexity in [35] trades a worse dependence on ϵ for a better dependence on n in high-dimensional problems to obtain

$$N = O\Big(\frac{n_0}{\epsilon^3}\Big(\log\frac{n}{\epsilon}\Big)^{1.5} + \frac{1}{\epsilon^2}\log\frac{1}{\delta}\Big), \tag{4.17}$$

through certain sparse modifications of SAA, where n_0 is the number of nonzero entries for the optimal solution of (1.1) with $n_0 \ll n$. Compared to (4.15), the sample complexity (4.17) depends polynomially on n_0 and $\log n$, instead of n which could become prohibitively large in high-dimensional problems.

As we shall see, this type of trade-off in sample complexity is typical and it also applies to the feasibility guarantee. In this section, we illustrate how low-dimensional modeling assumptions can allow alternative feasibility bounds for two-stage stochastic programming, especially in high-dimensional settings similar to (4.17). In particular, we still consider (4.2), but focus on the setting where the optimal solution of (1.1) (i.e., argmin $F(x)$) and the solution output of SAA (i.e., $x^\star(\xi^{[N]})$) are both sparse,

$$\mathcal{X} \subseteq \{x \in \mathbb{R}^n : \|x\|_0 \leq n_0\}$$

with $n_0 \ll n$. As before, we do not require $x^\star(\xi^{[N]})$ to be optimal for SAA, but only to be within the feasible domain of SAA with probability 1.

The sparsity of the solution is usually achieved by a regularization penalty which includes convex penalty, i.e., ℓ_1-norm (lasso), ℓ_2-norm (ridge) or ℓ_p-norm ($1 \leq p \leq +\infty$), and nonconvex penalty, i.e., folded concave penalty (see, e.g., [35]) or ℓ_0-norm (subset selection with mixed-integer programming; see, e.g., [6]). While the level of sparsity of $x^\star(\xi^{[N]})$ can be controlled by hyperparameters, we do not generally know the true value of $\|\text{argmin } F(x)\|_0$. In practice, the choice of n_0 is chosen either by model selection techniques or fixed beforehand based on domain-specific knowledge. To simplify the problem, we assume the choice of n_0, the level of sparsity for the candidate solutions, is preset. Ideally, we have $\|\text{argmin } F(x)\|_0 \leq n_0 \ll n$ so that the space of candidate solutions is rich enough to include the true sparse solution, but still highly sparse compared to problem dimension n. To provide our alternative feasibility bound, we first present the following lemma on the VC dimensions of sparse linear classifiers.

LEMMA 4.9. *Let* $\mathcal{U} = \{\{(y,z) \in \mathbb{R}^d \times \mathbb{R} : y^T x \leq z\}\}_{x \in \mathbb{R}^d, \|x\|_0 \leq d_0}$. *Then*

$$d_{VC}(\mathcal{U} \cup \{\mathbb{R}^d \times \mathbb{R}\}) \leq 2(d_0+1)\log_2\Big(\frac{de+e}{d_0+1}\Big).$$

Proof. It follows from Lemma 1 in [1] that the VC dimension of

$$\mathcal{U}' \triangleq \{\{(y,z) \in \mathbb{R}^d \times \mathbb{R} : y^T x + z \cdot c \leq 0\}\}_{(x,c) \in \mathbb{R}^d \times \mathbb{R}, \|x\|_0 + \|c\|_0 \leq d_0+1}$$

satisfies $d_{VC}(\mathcal{U}') \leq 2(d_0+1)\log_2\left(\frac{de+e}{d_0+1}\right)$. Now notice $\mathcal{U} \subseteq \mathcal{U}'$, since we can rewrite

$$\mathcal{U} = \{\{(y,z) \in \mathbb{R}^d \times \mathbb{R} : y^T x + z \cdot c \leq 0\}\}_{x \in \mathbb{R}^d, c=-1, \|x\|_0 \leq d_0}.$$

On the other hand, $\mathbb{R}^d \times \mathbb{R} \in \mathcal{U}'$ by picking $(x,c) = \mathbf{0}$. Consequently,

$$d_{VC}(\mathcal{U} \cup \{\mathbb{R}^d \times \mathbb{R}\}) \leq d_{VC}(\mathcal{U}') \leq 2(d_0+1)\log_2\Big(\frac{de+e}{d_0+1}\Big).$$

∎

Now we present the alternative version of Corollary 4.7 under sparsity.

COROLLARY 4.10. *Consider* (4.2). *Suppose* $\mathcal{X} \subseteq \{x \in \mathbb{R}^n : \|x\|_0 \leq n_0\}$ *and* $|J| < +\infty$, *where*

$$|J| = \max_{\xi \in \Xi}\{\ \#\ \textit{of extreme rays for the cone}\ \{a : a^T W_\xi \geq 0\}\}.$$

Then, let $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ *be IID samples from* $\mathbb{P}$ *(consequently* $\xi^{[N]} \sim \mathbb{P}^N$*), and* $x^\star(\xi^{[N]})$ *be the output of any algorithm that is guaranteed to be within the feasible region of SAA. Then, if*

$$N \geq \frac{4}{\epsilon}\left(d_{VC}\log\left(\frac{12}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right), \tag{4.18}$$

where

$$d_{VC} = \left(\frac{e}{(e-1)\log 2}|J|\Big(2(n_0+1)\log_2\Big(\frac{ne+e}{n_0+1}\Big)\Big)\cdot\log\big(\frac{e}{\log 2}\cdot|J|\big)\right),$$

then we have $\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq \delta$ *for any* $0 < \delta, \epsilon < 1$*. Equivalently, in terms of convergence rate, we have*

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \leq 2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^{d_{VC}}. \tag{4.19}$$

Proof. As in Corollary 4.7, let $\mathcal{A}_\xi$ be the set of nonequivalent extreme rays of polyhedral cone $\{a : a^T W_\xi \geq 0\}$. Observe $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ defined in (3.1) consists of

$$H_x = \{\xi : a^T T_\xi x \leq a^T h_\xi, \forall a \in \mathcal{A}_\xi\}.$$

For all $\xi \in \Xi$, since $|\mathcal{A}_\xi| \leq |J| < +\infty$, we can label the elements in $\mathcal{A}_\xi$ by $\{a_{\xi j}\}_{j\in[|A_\xi|]}$. Then, define $\{(y_{\xi j}, z_{\xi j})\}_{j\in|J|}$ as

$$y_{\xi j}^T = \begin{cases} a_{\xi j}^T T_\xi & \text{for } 1 \leq j \leq |\mathcal{A}_\xi|, \\ \mathbf{0} & \text{for } |\mathcal{A}_\xi| < j \leq |J|, \end{cases}$$

$$z_{\xi j} = \begin{cases} a_{\xi j}^T h_\xi & \text{for } 1 \leq j \leq |\mathcal{A}_\xi|, \\ 0 & \text{for } |\mathcal{A}_\xi| < j \leq |J|. \end{cases}$$

Then, we can redefine

$$H_x = \bigcap_{j=1}^{|J|}\{\xi : y_{\xi j}^T x \leq z_{\xi j}\}. \tag{4.20}$$

Since $\mathcal{X} \subseteq \{x \in \mathbb{R}^n : \|x\|_0 \leq n_0\}$, it follows from Lemmas 4.1, 4.4, and 4.9 that

$$d_{VC}(\mathcal{H}) \leq \left(\frac{e}{(e-1)\log 2}|J|\Big(2(n_0+1)\log_2\Big(\frac{ne+e}{n_0+1}\Big)\Big)\cdot\log\Big(\frac{e}{\log 2}\cdot|J|\Big)\right).$$

The rest of the proof follows from Theorem 3.2. ∎

As we can see, by leveraging the sparsity of the solutions, the sample complexity in Corollary 4.10 is improved to $N = O\big(\frac{n_0 \log n}{\epsilon}\log\frac{1}{\epsilon} + \frac{1}{\epsilon}\log\frac{1}{\delta}\big)$ versus the $N = O\big(\frac{n}{\epsilon}\log\frac{1}{\epsilon} + \frac{1}{\epsilon}\log\frac{1}{\delta}\big)$ in Corollary 4.6, similar to the trade-off in high-dimensional SAA between (4.15) and (4.17).

4.2. Special Structures. In this section, we consider two special structures. The first one is the chain-constrained domain considered in [37]. The second one is a finite feasible domain considered in [15].

4.2.1. Chain-Constrained Domain. In previous sections, Theorem 3.2 is used to analyze example (4.2) without using the chain-constrained structure as in [37]. However, it is worth noting that Theorem 3.2 still offers an explicit bound on the feasibility of $x^\star(\xi^{[N]})$ based solely on the chain-constrained structure, although at a slightly worse sample complexity than [37]. To see this, notice the VC dimension of any chain-constrained domain can be directly bounded.

LEMMA 4.11. *If dom f_ξ has a chain-constrained domain of order m, then the VC dimension of $\mathcal{H} = \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ in (3.1) satisfies*

$$d_{VC}(\mathcal{H}) \leq \frac{e}{(e-1)\log 2} m \log\left(\frac{e}{\log 2}\cdot m\right) = O(m\log m).$$

Proof. Recall Definition 4.5. Since dom f_ξ is a chain-constrained domain of order m, we can write dom $f_\xi = \bigcap_{k=1}^m U_k^\xi$, where each $U_k^\xi \in \{U_k^{\xi'}\}_{\xi'\in\Xi}$ is a chain on $\mathcal{X}\subseteq\mathbb{R}^n$ indexed by $\xi\in\mathbb{R}^r$. Now, for $k\in[m]$, define $W_k^x \triangleq \{\xi : x\in U_k^\xi\}$, and we have from (3.1) that $H_x = \{\xi : x\in \text{ dom}f_\xi\} = \{\xi : x\in\bigcap_{k=1}^m U_k^\xi\} = \bigcap_{k=1}^m W_k^x$. We show, for each $k\in[m]$, $\{W_k^x\}_{x\in\mathcal{X}}$ is a chain as well. Suppose this is not the case, then there exists $x_1,x_2\in\mathcal{X}$ such that $W_k^{x_1}\not\subseteq W_k^{x_2}$ and $W_k^{x_2}\not\subseteq W_k^{x_1}$. This implies there exist $\xi_1\in W_k^{x_1}$ and $\xi_2\in W_k^{x_2}$ such that $\xi_1\notin W_k^{x_2}$ and $\xi_2\notin W_k^{x_1}$. This further implies $x_1\in U_k^{\xi_1}, x_2\notin U_k^{\xi_1}$ and $x_2\in U_k^{\xi_2}, x_1\notin U_k^{\xi_2}$. Consequently, neither $U_k^{\xi_1}\subseteq U_k^{\xi_2}$ nor $U_k^{\xi_2}\subseteq U_k^{\xi_1}$ is true, contradicting the assumption that $\{U_k^\xi\}_{\xi\in\Xi}$ is a chain. Thus, $\{W_k^x\}_{x\in\mathcal{X}}$ is a chain on Ξ for each $k\in[m]$. It then follows trivially $\{W_k^x\}_{x\in\mathcal{X}}\cup\{\Xi\}$ is also a chain on Ξ for each $k\in[m]$. Consequently, $\mathcal{H} = \{H_x\}_{x\in\mathcal{X}}\cup\{\Xi\}$ is a chain-constrained domain of order m.

On the other hand, the VC dimensions of chains $\{U^\omega\}_{\omega\in I}$ are at most 1 because they cannot shatter any two points. In particular, if $\{x_1,x_2\}$ are two points living on the same space as $\{U^\omega\}_{\omega\in I}$, the shattering of $\{x_1,x_2\}$ requires $x_1\in U^{\omega_1}, x_2\notin U^{\omega_1}$ and $x_2\in U^{\omega_2}, x_1\notin U^{\omega_2}$ for some $U^{\omega_1}, U^{\omega_2}\in\{U^\omega\}_{\omega\in I}$. If this were to happen, then neither $U^{\omega_1}\subseteq U^{\omega_2}$ nor $U^{\omega_2}\subseteq U^{\omega_1}$ would be true, contradicting the definition of a chain. Then, if $\{\mathcal{U}_k\}_{k\in[m]}$ are the m chains consisting of a chain-constrained domain $\mathcal{U}$ of order m where each $U\in\mathcal{U}$ is of the form $U=\bigcap_{k=1}^m U_k$ for some $U_k\in\mathcal{U}_k$, it again follows from Theorem 1.1 in [51] that

$$d_{VC}(\mathcal{U}) \leq d_{VC}(\sqcap_{k=1}^m \mathcal{U}_k) \leq \frac{e}{(e-1)\log 2} m\log\Big(\frac{e}{\log 2}\cdot m\Big),$$

where $\sqcap_{k=1}^m\mathcal{U}_k \triangleq \left\{\bigcap_{k=1}^m U_k : U_k\in\mathcal{U}_k, k\in[m]\right\}$. The result follows now from the fact that $\mathcal{H}$ is a chain of order m. ∎

Lemma 4.11 combined with Theorem 3.2 can provide an explicit sample complexity for feasibility.

COROLLARY 4.12. *If dom f_ξ has a chain-constrained domain of order m, then Theorem 3.2 guarantees that for*

$$N \geq \frac{4}{\epsilon}\left(\left(\frac{e}{(e-1)\log 2} m\log\left(\frac{e}{\log 2}\cdot m\right)\right)\log\left(\frac{12}{\epsilon}\right) + \log\left(\frac{2}{\delta}\right)\right), \tag{4.21}$$

we have

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \le \delta$$

for any $0 < \delta, \epsilon < 1$.

Corollary 4.12 provides a sample complexity $O(\frac{m}{\epsilon}\log m \log(\frac{1}{\epsilon}) + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ for chain-constrained domains, or $O(\frac{m}{\epsilon}\log m + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ using the PAC bounds from [23, 48], while Scenario 1 in [37] provides an $O(\frac{m}{\epsilon} + \frac{1}{\epsilon}\log(\frac{1}{\delta}))$ bound according to (3.12). As we can see, the more refined analysis on the chain-constrained structure in [37] leads to a better rate over Corollary 4.12 by log factors. However, the generality offered by Theorem 3.2 is still noteworthy, since its applicability in most situations does not hinge on the chain-constrained domain.

4.2.2. Finite Feasible Region. In this subsection, we apply Theorem 3.2 to the case where the decision set $\mathcal{X}$ is finite.

COROLLARY 4.13. *Suppose* $|\mathcal{X}| < +\infty$ *and let* $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ *be IID samples from* $\mathbb{P}$ *(consequently* $\xi^{[N]} \sim \mathbb{P}^N$*). Then, if*

$$N \ge \frac{4}{\epsilon}\Big(\log_2(|\mathcal{X}|+1)\cdot\log\Big(\frac{12}{\epsilon}\Big) + \log\Big(\frac{2}{\delta}\Big)\Big), \tag{4.22}$$

we have

$$\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) \le \delta$$

for any $0 < \delta, \epsilon < 1$.

Proof. Let $\mathcal{H} \triangleq \{H_x\}_{x\in\mathcal{X}} \cup \{\Xi\}$ be the class of subsets defined in (3.1). It follows that $|\mathcal{H}| \le |\mathcal{X}| + 1 < +\infty$. It is known that if $|\mathcal{H}| < +\infty$, then $d_{VC}(\mathcal{H}) \le \log_2|\mathcal{H}|$ (by definition of the VC dimension or see [4]). The result follows from Theorem 3.2. □

Note that since the VC dimension of a finite hypothesis class is bounded by the logarithm of its cardinality, we get the results in Corollary 4.13 for free. In Section 4 of [15], the case of finite feasible region $\mathcal{X}$ is also discussed, with a slightly different focus. In particular, with assumptions on the moment generating functions, [15] proves the exponential convergence of a δ-optimal set towards an ϵ-optimal set using large deviations (LD) theory. The rate of convergence also depends on constants from the LD analysis. However, [15] also offers a more direct analysis on the feasibility of SAA solution $x^\star(\xi^{[N]})$ when $|\mathcal{X}| < +\infty$ which does not rely on distributional assumptions of $f(\xi, x)$. To be specific, Lemma 9 of [15] states

$$\mathbb{P}^N(\hat{F}_N(x) < +\infty) \le (1-\eta)^N \text{ for } x \in \mathcal{X}^{Infea}, \tag{4.23}$$

where $\mathcal{X}^{Infea} = \{x : x \in \mathcal{X} \text{ and } V(x) > 0\}$ and $\eta = \min\{V(x) : x \in \mathcal{X}^{Infea}\}$. Building on (4.23), we can deduce the following direct bound regarding $x^\star(\xi^{[N]})$:

$$\begin{aligned}\mathbb{P}^N(V(x^\star(\xi^{[N]})) > \eta) &= \mathbb{P}^N(x^\star(\xi^{[N]}) \in \mathcal{X}^{Infea}) \\ &\le \mathbb{P}^N\Big(\bigcup_{x\in\mathcal{X}^{Infea}} \{\hat{F}_N(x) < +\infty\}\Big) \\ &\le \sum_{x\in\mathcal{X}^{Infea}} \mathbb{P}^N(\hat{F}_N(x) < +\infty) \le |\mathcal{X}^{Infea}|(1-\eta)^N, \end{aligned} \tag{4.24}$$

which leads to an $O(\frac{1}{\eta}\log(|\mathcal{X}|) + \frac{1}{\eta}\log(\frac{1}{\beta}))$ sample complexity, comparable to the $O(\frac{1}{\eta}\log(\frac{1}{\eta})\log_2(|\mathcal{X}|) + \frac{1}{\eta}\log(\frac{1}{\beta}))$ complexity in (4.22). Moreover, if we utilize the PAC bound from Remark 3.3, the bound in (4.22) could be improved to

$$O(\frac{1}{\eta}\log_2(|\mathcal{X}|) + \frac{1}{\eta}\log(\frac{1}{\beta}))$$

which is of the same order as (4.24).

Appendix A. A Verifiable Example. In this section, we validate the result of Corollary 4.7 on an artificial example that is specially constructed so that the optimal SAA solution $x^\star(\xi^{[N]})$ and its violation probability $V(x^\star(\xi^{[N]}))$ can be calculated analytically. This then allows us to directly check the validity of our bound in Corollary 4.7. Note that bounds obtained from computational learning theory and the VC dimension can be crude for small or midsize problems, and they become more interesting when N, n are large and ϵ is tiny. However, in the latter large-scale setting, it can be difficult to verify experimentally the quality of the bounds since an accurate estimation of the violation probability is costly or even infeasible. To overcome this difficulty, we build this example that exhibits a closed form so that we can directly check our bound's validity for any combination of N, n (arbitrarily large), and ϵ (arbitrarily small).

Our example is a modification of the two-stage resource planning problem in [15] (see also [38, 39]). The example has the form

$$\inf_{x\geq 0, x\in\mathbb{R}^n} F(x) \triangleq c^T x + \mathbb{E}[f(\xi, x)], \tag{A.1}$$

where $c > \mathbf{0} \in \mathbb{R}^n$ and

$$\begin{aligned} f(\xi, x) \triangleq \inf_{y\in\mathbb{R}^n}\ & q^T y \\ \text{s.t.}\quad & y \leq x - \xi, \\ & y \geq 0, \end{aligned} \tag{A.2}$$

where $q > \mathbf{0} \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^n$ with each entry $[\xi]_i$ independently and uniformly distributed on $[0, 1]$. We summarize our results in the following corollary.

Corollary A.1. *Let $\xi^{[N]} = \{\xi_1, ..., \xi_N\}$ be IID samples from $\mathbb{P}$ (consequently $\xi^{[N]} \sim \mathbb{P}^N$). Then the optimal SAA solution of the example in (A.1) satisfies*

$$\mathbb{P}^N(V(x^\star(\xi^{[N]}) > \epsilon) = \sum_{i=0}^{n-1} \frac{(N\log(\frac{1}{1-\epsilon}))^i}{i!} e^{-N\log(\frac{1}{1-\epsilon})}. \tag{A.3}$$

Moreover, compared with the bound in Corollary 4.7, we have

$$\sum_{k=0}^{n-1} \frac{(N\log(\frac{1}{1-\epsilon}))^k}{k!} e^{-N\log(\frac{1}{1-\epsilon})} \leq 2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^{d_{VC}} \tag{A.4}$$

for any integer $N, n \geq 1$ and $\epsilon \in (0, 1)$.

Proof. It is straightforward to check that the optimal solution of the SAA problem satisfies

$$[x^\star(\xi^{[N]})]_i = \max_{j\in[N]} [\xi_j]_i, \tag{A.5}$$

where $i \in [n]$. Moreover, given $\mathbf{0} \le x \le \mathbf{1} \in \mathbb{R}^n$, since each entry of ξ is independently and uniformly distributed on $[0,1]$, we have

$$\begin{aligned} V(x) &= \mathbb{P}(\cup_{i\in[n]}\ [\xi]_i > [x]_i) \\ &= 1 - \mathbb{P}(\cap_{i\in[n]}\ [\xi]_i \le [x]_i) \\ &= 1 - \prod_{i\in[n]} [x]_i. \end{aligned} \tag{A.6}$$

On the other hand, it is known that if U is a uniform random variable on $[0,1]$, then $-\log(U) \sim exp(1)$ follows an exponential distribution with parameter 1. Moreover, it is known that the minimum of a series of independent exponential distributions $\{V_j \sim exp(\lambda_j)\}_{j\in[N]}$ follows another exponential distribution $exp(\sum_{j\in[N]} \lambda_j)$. Thus, if $\{U_j\}_{j\in[N]}$ is a series of independent and uniform random variables on $[0,1]$, then

$$-\log(\max_{j\in[N]} U_j) = \min_{j\in[N]} -\log U_j \sim exp(N) \tag{A.7}$$

follows an exponential distribution with parameter N. Finally, it is also known that if we have n independent exponential random variables $\{W_i\}_{i\in[n]}$ with parameter N, then their sum,

$$\sum_{i\in[n]} W_i \sim gamma(n, N), \tag{A.8}$$

follows a gamma distribution with parameter (n, N) and cumulative distribution function,

$$F(x; n, N) = 1 - \sum_{i=0}^{n-1} \frac{(Nx)^i}{i!} e^{-Nx}. \tag{A.9}$$

Now, it follows from (A.5) and (A.6) that

$$\begin{aligned} \mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) &= \mathbb{P}^N(1 - \prod_{i\in[n]} \max_{j\in[N]} [\xi_j]_i > \epsilon) \\ &= \mathbb{P}^N\Big(\sum_{i\in[n]} -\log(\max_{j\in[N]}[\xi_j]) > \log\Big(\frac{1}{1-\epsilon}\Big)\Big) \\ &= \sum_{i=0}^{n-1} \frac{(N\log(\frac{1}{1-\epsilon}))^i}{i!} e^{-N\log(\frac{1}{1-\epsilon})}. \end{aligned}$$

This proves (A.3). To prove (A.4), note that since $\log(\frac{1}{1-\epsilon}) > \epsilon$ for $\epsilon \in (0,1)$, it follows from (A.7), (A.8), and (A.9) that

$$\begin{aligned} \mathbb{P}^N(V(x^\star(\xi^{[N]})) > \epsilon) &= \mathbb{P}^N\Big(\sum_{i\in[n]} -\log(\max_{j\in[N]}[\xi_j]) > \log\Big(\frac{1}{1-\epsilon}\Big)\Big) \\ &\le \mathbb{P}^N\Big(\sum_{i\in[n]} -\log(\max_{j\in[N]}[\xi_j]) > \epsilon\Big) \\ &= 1 - F(\epsilon; n, N) \\ &= \sum_{i=0}^{n-1} \frac{(N\epsilon)^i}{i!} e^{-N\epsilon}. \end{aligned}$$

Now, to prove (A.4), it suffices to prove

$$2\exp\left(-\frac{N\epsilon}{4}\right)\left(\frac{12}{\epsilon}\right)^{d_{VC}} \geq \sum_{i=0}^{n-1}\frac{(N\epsilon)^i}{i!}e^{-N\epsilon}.$$

First note from the formulations (A.2) and (4.6) that $d_{VC} \geq n$. Then, if we define $\theta \triangleq N\epsilon$, it suffices to show

$$2\left(\frac{12}{\epsilon}\right)^{n} \geq h(\theta;n) \triangleq \sum_{i=0}^{n-1}\frac{(\theta)^i}{i!}e^{-\frac{3}{4}\theta} \tag{A.10}$$

for any integer $n \geq 1$, $\theta \geq 0$, and $\epsilon \in (0,1)$. We prove (A.10) by contradiction. First note (A.10) is trivially true when $n = 1$. Suppose (A.10) is not true, then there exists $n \geq 2$ and $\epsilon \in (0,1)$ such that

$$\sup_{\theta\geq 0} h(\theta;n) > 2\left(\frac{12}{\epsilon}\right)^{n} \tag{A.11}$$

However, when $\theta \in [0,1]$, it is clear that

$$\sup_{\theta\geq 0} h(\theta;n) \leq \sum_{i=0}^{n-1}\frac{1}{i!}e^{0} \leq e < 2\left(\frac{12}{\epsilon}\right) \leq 2\left(\frac{12}{\epsilon}\right)^{n}.$$

On the other hand, when $\theta \to \infty$, the exponential decay dominates and we have $\lim_{\theta\to\infty} h(\theta;n) \to 0$. Consequently, there must exists some θ_u such that

$$\sup_{\theta\geq\theta_u} h(\theta;n) \leq 2\left(\frac{12}{\epsilon}\right)^{n}.$$

Combined with (A.11), the above inequalities imply $\sup_{1\leq\theta\leq\theta_u} h(\theta;n) > 2\left(\frac{12}{\epsilon}\right)^{n}$. By the extreme value theorem, there exists a local maximum $\theta^\star \in [1,\theta_u]$ such that

$$h(\theta^\star;n) > 2\left(\frac{12}{\epsilon}\right)^{n}, \tag{A.12}$$

and $h'(\theta^\star;n) = 0$. Solving for $h'(\theta^\star;n) = 0$, we obtain that

$$\sum_{i=0}^{n-2}\frac{(\theta^\star)^i}{i!} = \frac{3(\theta^\star)^{n-1}}{(n-1)!} \quad \text{and} \quad h(\theta^\star;n) = \frac{4(\theta^\star)^{n-1}}{(n-1)!}e^{-\frac{3}{4}\theta^\star}.$$

However, it can be checked that, for any integer $m \geq 1$,

$$\max_{\theta\geq 0}\frac{4\theta^m}{m!}e^{-\frac{3}{4}\theta} = 4\left(\frac{4}{3}\right)^{m}\frac{m^m}{m!}e^{-m}.$$

Then, using Stirling's formula (see [17]) we have $m! \geq m^m e^{-m}\sqrt{2\pi m}$ and

$$\max_{\theta\geq 0}\frac{4\theta^m}{m!}e^{-\frac{3}{4}\theta} \leq 2\left(\frac{4}{3}\right)^{m} < 2\left(\frac{12}{\epsilon}\right)^{m}. \tag{A.13}$$

Now, using the result in (A.13), we have

$$h(\theta^\star;n) \leq \max_{\theta\geq 0}\frac{4\theta^{n-1}}{(n-1)!}e^{-\frac{3}{4}\theta} < 2\left(\frac{12}{\epsilon}\right)^{n-1} < 2\left(\frac{12}{\epsilon}\right)^{n}$$

which contradicts (A.12). This concludes the proof. ∎

Appendix B. Technical Discussion on Measurability. As discussed in [30] or [46], an extended real valued function $f(\xi(\omega), x) : \Omega \times \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ defined on $\Omega \times \mathcal{X}$ equipped with a $\mathbb{P}$-complete measure and σ-algebra $\mathcal{F} \otimes \mathscr{B}$, where $\mathcal{X}$ is a Polish space (a complete, separable, metric space) is a random lower semicontinuous function if and only if

1. f is $\mathcal{F} \otimes \mathscr{B}$-measurable;
2. for every $\omega \in \Omega$, $f(\xi(\omega), \cdot)$ is lower semicontinuous (i.e., $\liminf_{x \to x_0} f(\xi, x) \geq f(\xi, x_0)$ for all $x_0 \in \mathcal{X}$),

which is the condition specified in the introduction following (1.1). Note that when $f : \mathcal{X} \to \mathbb{R}$, a notable class of random lower semicontinuous functions is given by the Carathéodory functions (also a standard setup in SAA literature [46]), which simply requires f's continuity in x and measurability in ξ. However, for $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, lower semicontinuity in x and measurability in ξ are not enough for f to be random lower semicontinuous. One must also require f to be $\mathcal{F} \otimes \mathscr{B}$-measurable [46, 30].

Given some $N \geq 1$ and the random lower semicontinuous function f, the sum of lower semicontinuous $\hat{F}_N$ is also random lower semicontinuous since the sum is always well-defined (there is no indeterminate form $-\infty + \infty$). Consequently, as shown in Theorems 7.36 and 7.37 of [46], the optimal solution $x^\star(\xi^{[N]})$ is measurable. On the other hand, fixing any $x_0 \in \mathcal{X}$, by Fatou's lemma

$$\begin{aligned}\liminf_{x \to x_0} V(x) = \liminf_{x \to x_0} \mathbb{E}\mathbf{1}\{f(\xi, x) = \infty\} &\geq \mathbb{E} \liminf_{x \to x_0} \mathbf{1}\{f(\xi, x) = \infty\} \\ &\geq \mathbb{E}\mathbf{1}\{f(\xi, x_0) = \infty\} = V(x_0),\end{aligned}$$

where the last inequality follows from the lower semicontinuity of $\mathbf{1}\{f(\xi, \cdot) = \infty\}$ due to the lower semicontinuity of $f(\xi, \cdot)$ and the nonnegativity of the indicator function. Thus, $V(\cdot)$ is a lower semicontinuous thus $\mathscr{B}$-measurable function on $\mathcal{X}$. It then follows $V(x^\star(\xi^{[N]}))$ is $\mathcal{F}$-measurable, suggesting (3.4) is well-defined.

For (3.3), we first note that for any $\epsilon > 0$, the set $\{\sup_{x \in \mathrm{dom}\ \hat{F}_N} V(x) > \epsilon\}$ is a.e. equivalent to $\big\{ \sup_{x \in \mathcal{X}} \big(V(x) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x) < \infty\}\big) > \epsilon \big\}$. It follows from the $\mathcal{F} \otimes \mathscr{B}$ measurability of f and the fact that the product of nonnegative lower semicontinuous functions (i.e., $\mathbf{1}\{f(\xi, \cdot) = \infty\}$) is also lower semicontinuous, that

$$V(x) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x) < \infty\} \tag{B.1}$$

is a random lower semicontinuous function. Thus, fixing any $x \in \mathcal{X}$, the set

$$\big\{ \big(V(x) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x) < \infty\}\big) > \epsilon \big\}$$

is $\mathbb{P}$-measurable. Now, letting $\{x_j\}_{j \in \mathbb{N}}$ be a countable dense subset of $\mathcal{X}$, the lower semicontinuity of (B.1) implies

$$\big\{ \sup_{x \in \mathcal{X}} \big(V(x) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x) < \infty\}\big) > \epsilon \big\}$$

is a.e. equivalent to

$$\big\{ \sup_{j \in \mathbb{N}} \big(V(x_j) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x_j) < \infty\}\big) > \epsilon \big\} = \bigcup_{j \in \mathbb{N}} \big\{ \big(V(x_j) \cdot \prod_{i=1}^N \mathbf{1}\{f(\xi_i, x_j) < \infty\}\big) > \epsilon \big\},$$

which is $\mathbb{P}$-measurable. The completeness of $\mathbb{P}$ now suggests the measurability of (3.3).

Acknowledgments. We greatly thank Alexander Shapiro for introducing to us the motivating problem and the helpful suggestions, and James Luedtke for the helpful feedback in the preparation of this manuscript.

REFERENCES

[1] F. Abramovich and V. Grinshtein, *High-dimensional classification by sparse logistic regression*, IEEE Transactions on Information Theory, 65 (2018), pp. 3068–3079.
[2] S. Ahmed, *Two-stage stochastic integer programming: A brief introduction*, Wiley encyclopedia of operations research and management science, (2010).
[3] S. Ahmed, A. Shapiro, and E. Shapiro, *The sample average approximation method for stochastic programs with integer recourse*, Submitted for publication, (2002), pp. 1–24.
[4] M. Anthony and N. Biggs, *Computational learning theory*, vol. 30, Cambridge University Press, 1997.
[5] G. Barbarosoğlu and Y. Arda, *A two-stage stochastic programming framework for transportation planning in disaster response*, Journal of the operational research society, 55 (2004), pp. 43–53.
[6] D. Bertsimas, A. King, and R. Mazumder, *Best subset selection via a modern optimization lens*, The annals of statistics, 44 (2016), pp. 813–852.
[7] H. M. Bidhandi and J. Patrick, *Accelerated sample average approximation method for two-stage stochastic programming with binary first-stage variables*, Applied Mathematical Modelling, 41 (2017), pp. 582–595.
[8] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, *Learnability and the vapnik-chervonenkis dimension*, Journal of the ACM (JACM), 36 (1989), pp. 929–965.
[9] I. Borosh and L. B. Treybig, *Bounds on positive integral solutions of linear diophantine equations*, Proceedings of the American Mathematical Society, 55 (1976), pp. 299–304.
[10] C. Bugg and A. Aswani, *Logarithmic sample bounds for sample average approximation with capacity-or budget-constraints*, Operations Research Letters, 49 (2021), pp. 231–238.
[11] G. Calafiore and M. C. Campi, *Uncertain convex programs: randomized solutions and confidence levels*, Mathematical Programming, 102 (2005), pp. 25–46.
[12] M. C. Campi and S. Garatti, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM Journal on Optimization, 19 (2008), pp. 1211–1230.
[13] M. C. Campi and S. Garatti, *A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality*, Journal of Optimization Theory and Applications, 148 (2011), pp. 257–280.
[14] A. Carè, S. Garatti, and M. C. Campi, *Fast—fast algorithm for the scenario technique*, Operations Research, 62 (2014), pp. 662–671.
[15] R. Chen and J. Luedtke, *On sample average approximation for two-stage stochastic programs without relatively complete recourse*, arXiv preprint arXiv:1912.13078, (2019).
[16] X. Chen, A. Shapiro, and H. Sun, *Convergence analysis of sample average approximation of two-stage stochastic generalized equations*, SIAM Journal on Optimization, 29 (2019), pp. 135–161.
[17] K. Conrad, *Stirling's formula*, Available in http://www. math. uconn. edu/kconrad/blurbs/analysis/stirling. pdf, (2016).
[18] D. P. De Farias and B. Van Roy, *On constraint sampling in the linear programming approach to approximate dynamic programming*, Mathematics of operations research, 29 (2004), pp. 462–478.
[19] M. Dillon, F. Oliveira, and B. Abbasi, *A two-stage stochastic programming model for inventory management in the blood supply chain*, International Journal of Production Economics, 187 (2017), pp. 27–41.
[20] R. M. Dudley, *Central limit theorems for empirical measures*, The Annals of Probability, (1978), pp. 899–929.
[21] E. Erdoğan and G. Iyengar, *Ambiguous chance constrained problems and robust optimization*, Mathematical Programming, 107 (2006), pp. 37–61.
[22] Y. M. Ermoliev and V. I. Norkin, *Sample average approximation method for compound stochastic optimization problems*, SIAM Journal on Optimization, 23 (2013), pp. 2231–2263.
[23] S. Hanneke, *The optimal sample complexity of pac learning*, Journal of Machine Learning

Research, 17 (2016), pp. 1–15, http://jmlr.org/papers/v17/15-389.html.
[24] J. L. Higle and S. Sen, *Stochastic decomposition: a statistical method for large scale stochastic linear programming*, vol. 8, Springer Science & Business Media, 2013.
[25] G. Huang and D. P. Loucks, *An inexact two-stage stochastic programming model for water resources management under uncertainty*, Civil Engineering Systems, 17 (2000), pp. 95–118.
[26] R.-J. Jing, M. Moreno-Maza, and D. Talaashrafi, *Complexity estimates for fourier-motzkin elimination*, in International Workshop on Computer Algebra in Scientific Computing, Springer, 2020, pp. 282–306.
[27] P. Jirutitijaroen and C. Singh, *Reliability constrained multi-area adequacy planning using stochastic programming with sample-average approximations*, IEEE Transactions on Power Systems, 23 (2008), pp. 504–513.
[28] M. Kääriäinen, *Relating the rademacher and vc bounds*, tech. report, Citeseer, 2004.
[29] M. J. Kearns, U. V. Vazirani, and U. Vazirani, *An introduction to computational learning theory*, MIT press, 1994.
[30] L. A. Korf and R. J.-B. Wets, *Random lsc functions: An ergodic theorem*, Mathematics of Operations Research, 26 (2001), pp. 421–445.
[31] S. Küçükyavuz and S. Sen, *An introduction to two-stage stochastic mixed-integer programming*, in Leading Developments from INFORMS Communities, INFORMS, 2017, pp. 1–27.
[32] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behavior of sampling methods for stochastic programming*, Annals of Operations Research, 142 (2006), pp. 215–241.
[33] C. Liu, Y. Fan, and F. Ordóñez, *A two-stage stochastic programming model for transportation network protection*, Computers & Operations Research, 36 (2009), pp. 1582–1590.
[34] H. Liu, C. Hernandez, and H. Y. Lee, *Regularized sample average approximation for high-dimensional stochastic optimization under low-rankness*, arXiv preprint arXiv:1904.03453, (2019).
[35] H. Liu, X. Wang, T. Yao, R. Li, and Y. Ye, *Sample average approximation with sparsity-inducing penalty for high-dimensional stochastic programming*, Mathematical programming, 178 (2019), pp. 69–108.
[36] J. Liu and S. Sen, *Asymptotic results of stochastic decomposition for two-stage stochastic quadratic programming*, SIAM Journal on Optimization, 30 (2020), pp. 823–852.
[37] R. P. Liu, *On feasibility of sample average approximation solutions*, SIAM Journal on Optimization, 30 (2020), pp. 2026–2052.
[38] X. Liu, S. Küçükyavuz, and J. Luedtke, *Decomposition algorithms for two-stage chance-constrained programs*, Mathematical Programming, 157 (2016), pp. 219–243.
[39] J. Luedtke, *A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support*, Mathematical Programming, 146 (2014), pp. 219–244.
[40] J. Luedtke and S. Ahmed, *A sample approximation approach for optimization with probabilistic constraints*, SIAM Journal on Optimization, 19 (2008), pp. 674–699.
[41] N. Noyan, *Risk-averse two-stage stochastic programming with an application to disaster management*, Computers & Operations Research, 39 (2012), pp. 541–559.
[42] R. I. Oliveira and P. Thompson, *Sample average approximation with heavier tails i: non-asymptotic bounds with weak assumptions and stochastic constraints*, arXiv preprint arXiv:1705.00822, (2017).
[43] R. I. Oliveira and P. Thompson, *Sample average approximation with heavier tails ii: localization in stochastic convex optimization and persistence results for the lasso*, arXiv preprint arXiv:1711.04734, (2017).
[44] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: theory and applications*, Journal of optimization theory and applications, 142 (2009), pp. 399–416.
[45] M. J. Panik, *Extreme Points and Directions for Convex Sets*, Springer Netherlands, Dordrecht, 1993, pp. 189–234, https://doi.org/10.1007/978-94-015-8124-0_8, https://doi.org/10.1007/978-94-015-8124-0_8.
[46] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, 2014.
[47] A. Shapiro and A. Nemirovski, *On complexity of stochastic programming problems*, in Continuous optimization, Springer, 2005, pp. 111–146.
[48] H. U. Simon, *An almost optimal pac algorithm*, in Proceedings of The 28th Conference on Learning Theory, P. Grünwald, E. Hazan, and S. Kale, eds., vol. 40 of Proceedings of Machine Learning Research, Paris, France, 03–06 Jul 2015, PMLR, pp. 1552–1563, http://proceedings.mlr.press/v40/Simon15a.html.
[49] M. Talagrand, *Upper and lower bounds for stochastic processes: modern methods and classical*

problems, vol. 60, Springer Science & Business Media, 2014.
[50] M. Terzer, *Large scale methods to enumerate extreme rays and elementary modes*, PhD thesis, ETH Zurich, 2009.
[51] A. Van Der Vaart and J. A. Wellner, *A note on bounds for vc dimensions*, Institute of Mathematical Statistics collections, 5 (2009), p. 103.
[52] A. W. Van Der Vaart and J. A. Wellner, *Weak convergence*, in Weak convergence and empirical processes, Springer, 1996, pp. 16–28.
[53] V. Vapnik, *The nature of statistical learning theory*, Springer science & business media, 2013.
[54] W. Wang and S. Ahmed, *Sample average approximation of expected value constrained stochastic programs*, Operations Research Letters, 36 (2008), pp. 515–519.