

# ON FEASIBILITY OF SAMPLE AVERAGE APPROXIMATION SOLUTIONS\*

RUI PENG LIU†

**Abstract.** When there are infinitely many scenarios, the current studies of two-stage stochastic programming problems rely on the relatively complete recourse assumption. However, such an assumption can be unrealistic for many real-world problems. This motivates us to study general stochastic programming problems where the sample average approximation (SAA) solutions are not necessarily feasible. When the problems are convex and the true solutions lie in the interior of feasible solutions, we show that the portion of infeasible SAA solutions decays exponentially as the sample size increases. We also study functions with chain-constrained domain and show that the portion of SAA solutions having a low degree of feasibility decays exponentially as the sample size increases. This result is then extended to multistage stochastic programming.

**Key words.** (multistage) stochastic programming, sample average approximation method, feasibility, convergence, exponential rate

**AMS subject classifications.** 90C15, 03E75

**DOI.** 10.1137/19M1253447

**1. Introduction.** We consider the stochastic programming problem

$$(1.1) \quad \inf_{x \in \mathcal{X}} F(x) := \mathbb{E}[f_{\xi}(x)],$$

where  $\mathcal{X} \subseteq \mathbb{R}^n$  is a nonempty set; the random vector  $\xi : \Omega \rightarrow \mathbb{R}^{n_0}$  is defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and the support of  $\xi$  is  $\Xi := \xi(\Omega) \subseteq \mathbb{R}^{n_0}$  (we may also use  $\xi$  to denote the outcomes in  $\Xi$ ); and for each outcome  $\xi \in \Xi$ ,  $f_{\xi} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is an extended real-valued function. Throughout the paper, we assume  $F(x) < +\infty$  for some  $x \in \mathcal{X}$ .

An important class of (1.1) is the two-stage stochastic programming. In two-stage problems,  $f_{\xi}(x)$  is given by the optimal value of the second stage problem, i.e.,

$$f_{\xi}(x) := \inf_{y \in \mathcal{Y}(x, \xi)} g_{\xi}(y),$$

where  $\mathcal{Y}$  is a multivalued function that maps  $(x, \xi)$  to sets, and  $g_{\xi}$  is a real-valued function for each  $\xi \in \Xi$ . By definition,  $f_{\xi}(x) = +\infty$  if the set  $\mathcal{Y}(x, \xi)$  is empty. In two-stage stochastic programming, the first stage decisions  $x$  should be implemented before a realization of the random data becomes available and hence should be independent of the random data. The second stage decisions  $y$  are made after observing the random data and are functions of the data. The model has found wide applications, such as transportation planning [1], water resource management [9], power production [22], etc.

When  $\Xi$  contains infinitely many outcomes, the current studies of two-stage problems rely on the relatively complete recourse assumption; that is, for every  $x \in \mathcal{X}$  and almost every  $\xi \in \Xi$ , the set  $\mathcal{Y}(x, \xi)$  is nonempty. In terms of  $f_{\xi}$ , the assumption

\*Received by the editors April 1, 2019; accepted for publication (in revised form) May 4, 2020; published electronically August 3, 2020.  
<https://doi.org/10.1137/19M1253447>

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (rayliu@gatech.edu).

states that  $\mathbb{P}\{f_\xi(x) = +\infty\} = 0$  for every  $x \in \mathcal{X}$ . However, such an assumption can be unrealistic for many real-world applications. For example, when deciding the size of a reservoir for water supply during a potential drought, it could happen that some size is too small to store enough water. Although one can forge relatively complete recourse by passing to a penalized problem [18], the optimal solutions of the penalized problem may be inferior with respect to the original formulation.

In general, it is difficult to evaluate the expectation in (1.1) directly, because the underlying distribution is usually unknown. Even if the distribution is given, it is computationally hard to evaluate high-dimensional integrals to a high accuracy. The sample average approximation (SAA) method is a sampling-based method that aims to approximate the expectation using the Monte Carlo sampling technique; theoretical results [10], [17] and numerical experiments [11], [12], [21] indicate that the approximations could be reasonably accurate when the SAA method is applied to the two-stage stochastic programming problems with relatively complete recourse. In this paper, we study the SAA method applied to the problems (1.1) where the probability  $\mathbb{P}\{f_\xi(x) = +\infty\}$  could be positive for some  $x \in \mathcal{X}$ . To this end, we make the following assumption.

*Assumption 1.1.* It is possible to generate an independent and identically distributed (i.i.d.) sample  $\xi^1, \xi^2, \dots$  of realizations of the random vector  $\xi$ .

The SAA method generates a (random) sample  $\xi^{[N]} := (\xi^1, \dots, \xi^N)$  of size  $N$  and approximates the expectation function  $F$  by the SAA function

$$\hat{F}_N(x) := \frac{1}{N} \sum_{i=1}^N f_{\xi^i}(x),$$

thereby approximating the true problem by the SAA problem

$$(1.2) \quad \inf_{x \in \mathcal{X}} \hat{F}_N(x).$$

Note that the SAA problem (1.2) depends on the generated sample  $\xi^{[N]}$ . We assume there is a mapping that assigns an SAA solution  $x^*(\xi^{[N]})$  to each SAA problem, e.g., an optimization algorithm that outputs  $x^* = x^*(\xi^{[N]})$  by solving (1.2). We only assume that  $x^*$  is a feasible solution of (1.2) (i.e.,  $x^* \in \mathcal{X}$  and  $\hat{F}_N(x^*) < +\infty$ ), while  $x^*$  need not be an optimal solution; indeed, it is unrealistic to require  $x^*$  to be optimal when the SAA problem is nonconvex. In section 3, we explicitly require  $x^*$  to be optimal in stochastic convex programming.

The SAA solutions are not necessarily feasible for the true problem (1.1); it could happen that  $F(x^*(\xi^{[N]})) = +\infty$ . To better understand the quality of the SAA solutions, it is important to study how feasible the solution  $x^*(\xi^{[N]})$  is. The following definition of degree of feasibility involves the concept of the domain of  $f_\xi$ , i.e.,  $\text{dom } f_\xi = \{x : f_\xi(x) < +\infty\}$ .

**DEFINITION 1.2** (degree of feasibility). *The degree of feasibility of an SAA solution  $x^*(\xi^{[N]})$  is*

$$d(x^*(\xi^{[N]})) := \mathbb{P}\{\xi \in \Xi : x^*(\xi^{[N]}) \in \text{dom } f_\xi\}.$$

*One can also view  $\xi$  as a random vector independent of the random sample  $\xi^{[N]}$ ; then  $d(x^*(\xi^{[N]}))$  is the conditional probability*

$$d(x^*(\xi^{[N]})) = \Pr(x^*(\xi^{[N]}) \in \text{dom } f_\xi \mid \xi^{[N]}).$$

*We adopt the former view for other similar definitions.*

We intend to demonstrate in this paper that, for a broad class of problems, the portion of the SAA solutions  $x^*(\xi^{[N]})$  such that  $d(x^*(\xi^{[N]})) < 1 - \alpha$  (for some  $\alpha \in (0, 1)$ ) decreases exponentially in  $N$ ; here, the portion is measured with respect to the product measure  $\mathbb{P}^N$  associated with  $\xi^{[N]}$ . A similar problem regarding chance-constrained stochastic programming was studied in [2], [3] (and references therein), but the nature of their problem is somewhat different from the one considered in this paper. The asymptotic epiconvergence of the SAA function to the expectation function was studied in [6]. Their result establishes the convergence of the optimal SAA solutions to the optimal solutions of the true problem (1.1) when the sample size  $N$  tends to infinity. However, the degree of feasibility cannot be deduced from their result when  $N$  is finite.

The rest of this paper is organized as follows. In section 2, we investigate an exponential rate of convergence of degree of feasibility when  $\{f_\xi\}$  has chain-constrained domain (the proofs are delayed to Appendix A). In section 3, stochastic convex programming is considered. Under the uniform convergence property of the SAA method, we show that the portion of infeasible SAA solutions decays exponentially in  $N$  when the true solutions lie in the interior of the feasible region. Also, by combining uniform convergence and chain-constrained domain, we significantly improve the rate presented in section 2.1. In section 4, the result is extended to multistage stochastic programming.

We use the following notation and terminology throughout the paper. We denote  $F^*$  and  $\mathcal{X}^*$  to be the optimal value and the set of optimal solutions of the true problem (1.1), respectively. The Euclidean norm of a vector  $x \in \mathbb{R}^n$  is  $\|x\|$ . For a set  $U \subset \mathbb{R}^n$ , we denote  $\text{int } U$ ,  $\text{bd } U$ ,  $\text{cl } U$ ,  $U^c$ ,  $|U|$  to be its interior, boundary, closure, complement, and cardinality, respectively; also, we denote the distance of  $x$  to  $U$  by  $\text{dist}_U(x) = \text{dist}(x, U) = \inf_{y \in U} \|x - y\|$ . We denote  $\mathbb{R}$ ,  $\mathbb{Q}$ ,  $\mathbb{Q}^c$ ,  $\mathbb{N}$  to be real, rational, irrational, and natural numbers, respectively. Given two sets  $U$  and  $V$ , we denote  $U \subset V$  to be  $U \subseteq V$  and  $U \neq V$ . The effective domain of a function  $f$  is  $\text{dom } f = \{x : f(x) < +\infty\}$ . For a natural number  $m$ , we denote  $[m] := \{1, \dots, m\}$ . The preimage of a set  $T$  under the function  $f$  is  $f^{-1}T = \{x : f(x) \in T\}$ .

**2. Chain-constrained domain.** In this section, we investigate the degree of feasibility of SAA solutions when  $\{f_\xi\}$  has chain-constrained domain. Let us first define chain-constrained domain.

**DEFINITION 2.1.** A collection of sets  $\{U^\omega\}_{\omega \in I}$  is a chain if for any  $\omega_1, \omega_2 \in I$ , either  $U^{\omega_1} \subseteq U^{\omega_2}$  or  $U^{\omega_1} \supseteq U^{\omega_2}$ .

If the index set  $I$  is finite, then the sets can be put into order:

$$U^{\omega_1} \subseteq U^{\omega_2} \subseteq \dots \subseteq U^{\omega_n};$$

in particular,  $\bigcap_{\omega \in I} U^\omega = U^{\omega_1}$ , i.e.,  $U^{\omega_1}$  is the smallest element.

**DEFINITION 2.2** (chain-constrained domain). A collection of functions  $\{f_\xi\}_{\xi \in \Xi}$  has chain-constrained domain of order  $m \in \mathbb{N}$  if there exist  $m$  collections of sets  $\{U_k^\xi\}_{\xi \in \Xi}$ ,  $k \in [m]$ , such that each collection  $\{U_k^\xi\}$  is a chain, and for each  $\xi \in \Xi$ ,

$$(2.1) \quad \text{dom } f_\xi = \bigcap_{k=1}^m U_k^\xi.$$

**Remark 2.3.** Definition 2.2 is motivated by the functional constraints

$$(2.2) \quad \text{dom } f_\xi = \{x : c_k(x) \leq \ell_k(\xi), k \in [m]\} \quad \forall \xi \in \Xi,$$

where for each  $k$ ,  $c_k(x)$  is a function of  $x$  and  $\ell_k$  is a random variable supported on  $\Xi$ . Consider

$$U_k^\xi = \{x : c_k(x) \leq \ell_k(\xi)\};$$

observe that for any  $\xi_1$  and  $\xi_2 \in \Xi$ , either  $\ell_k(\xi_1) \leq \ell_k(\xi_2)$  or  $\ell_k(\xi_1) \geq \ell_k(\xi_2)$ , which implies  $U_k^{\xi_1} \subseteq U_k^{\xi_2}$  or  $U_k^{\xi_1} \supseteq U_k^{\xi_2}$ ; thus  $\{U_k^\xi\}_{\xi \in \Xi}$  is the chain induced by both the function  $c_k$  and the random variable  $\ell_k$ . One advantage of studying the form (2.1) instead of (2.2) is that (2.1) helps recognize sets that are not commonly represented as functional-constrained domain, e.g., when  $\{U^\xi\}$  is a chain of discrete sets.

Chain-constrained domain covers a broad range of two-stage stochastic programming problems.

*Example 2.4.* Suppose  $f_\xi$  is given by the second stage problem

$$\begin{aligned} f_\xi(x) &:= \inf_y g_\xi(y) \\ \text{s.t. } &W_\xi y + T_\xi x = h_\xi, \\ &y \geq 0, \end{aligned}$$

where the data  $(h_\xi, g_\xi, T_\xi, W_\xi)$  satisfy the following conditions:

1. the functions  $g_\xi$  are finite everywhere;
2. there are only finitely many distinct  $W_\xi$  and  $T_\xi$ , i.e.,  $|\{W_\xi\}| = p$  and  $|\{T_\xi\}| = q$  for some  $p, q \in \mathbb{N}$  (though there is no restriction on  $h_\xi$ ).

Note that  $f_\xi$  is convex when  $g_\xi$  is a convex function.

Denote  $\{W_1, \dots, W_p\}$  and  $\{T_1, \dots, T_q\}$  to be the set of distinct matrices. Observe that  $f_\xi(x) < +\infty$  if and only if the set  $\{y \geq 0 : W_\xi y + T_\xi x = h_\xi\}$  is nonempty, which by Farkas' lemma is if and only if  $a^\top(h_\xi - T_\xi x) \geq 0$  for all  $a$  such that  $a^\top W_\xi \geq 0$ . For each  $i \in [p]$ , let  $\{a_{ij}\}_{j \in J_i}$  denote the set of extreme rays of the polyhedral cone  $\{a : a^\top W_i \geq 0\}$ ; then

$$\text{dom } f_\xi = \{x : a_{ij}^\top T_k x \leq a_{ij}^\top h_\xi, W_\xi = W_i, j \in J_i, T_\xi = T_k\} \quad \forall \xi \in \Xi.$$

For each  $i \in [p]$ ,  $j \in J_i$ , and  $k \in [q]$ , consider the chain  $\{U_{ijk}^\xi\}$  such that

$$U_{ijk}^\xi := \begin{cases} \{x : a_{ij}^\top T_k x \leq a_{ij}^\top h_\xi\} & \text{if } W_\xi = W_i \text{ and } T_\xi = T_k, \\ \mathbb{R}^n & \text{otherwise;} \end{cases}$$

then

$$\text{dom } f_\xi = \cap_{i \in [p], j \in J_i, k \in [q]} U_{ijk}^\xi \quad \forall \xi \in \Xi,$$

and hence  $\{f_\xi\}$  has chain-constrained domain of order  $q \sum_{i=1}^p |J_i|$ . If  $W_i$  is never coupled with  $T_k$  in any outcome  $\xi$ , then  $U_{ijk}^\xi = \mathbb{R}^n$  for all  $\xi$ , and we can safely remove  $\{U_{ijk}^\xi\}$  to reduce the order.

As illustrated by Example 2.5, chain-constrained domain is more general than functional-constrained domain.

*Example 2.5.* Let  $\tau$  be an exponential random variable; then  $\tau$  is supported on  $(0, \infty)$ . Consider the chain  $\{U^\tau\}_{\tau > 0}$ ,

$$U^\tau := \begin{cases} \tau \bar{B} & \text{if } \tau \in \mathbb{Q} \\ \tau B & \text{if } \tau \in \mathbb{Q}^c \end{cases} \quad \forall \tau > 0,$$

where  $\overline{B}$  and  $B$  are the closed and open unit intervals in  $\mathbb{R}$ . We claim that  $\{U^\tau\}$  cannot be characterized by sublevel sets of any function  $c$ , i.e., there does not exist  $c, \ell$  such that  $U^\tau = c^{-1}(-\infty, \ell(\tau)]$  for all  $\tau > 0$ .

For contradiction, suppose there exist  $c$  and  $\ell$  such that  $U^\tau = c^{-1}(-\infty, \ell(\tau)]$  for all  $\tau$ . For  $0 < \tau_1 < \tau_2$ , we have  $U^{\tau_1} \subset U^{\tau_2}$ , which implies  $\ell(\tau_1) < \ell(\tau_2)$ . Let  $x$  be a positive irrational number; then  $x \notin U^x$  but  $x \in \cap_{\tau > x} U^\tau$ , which translates to  $\ell(x) < c(x) \leq \inf_{\tau > x} \ell(\tau)$ . Consider the gap  $\gamma_x := \inf_{\tau > x} \ell(\tau) - \ell(x)$ ; then  $\gamma_x > 0$  for each positive irrational number  $x$ . For any  $0 < \tau_1 < \tau_2$ ,

$$\ell(\tau_2) - \ell(\tau_1) \geq \sum_{x \in \mathbb{Q}^c \cap [\tau_1, \tau_2)} \gamma_x = +\infty,$$

since the sum of uncountably many positive numbers necessarily diverges to infinity, a contradiction.

**2.1. Exponential rate of convergence.** For a specified threshold  $\alpha \in (0, 1)$ , we are interested in the probability

$$(2.3) \quad \mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1 - \alpha\}.$$

When  $\{f_\xi\}$  has chain-constrained domain, (2.3) decreases exponentially in  $N$ . We bound (2.3) by bounding the degree of feasibility of the domain of SAA function  $\text{dom } \hat{F}_N = \cap_{i=1}^N \text{dom } f_{\xi^i}$ .

DEFINITION 2.6. *The degree of feasibility of  $\text{dom } \hat{F}_N$  is*

$$D(\xi^{[N]}) := \mathbb{P}\{\xi \in \Xi : \cap_{i=1}^N \text{dom } f_{\xi^i} \subseteq \text{dom } f_\xi\}.$$

For each sample  $\xi^{[N]}$ , the SAA solution  $x^*(\xi^{[N]})$  is assumed to lie in the domain of the SAA function, i.e.,  $\hat{F}_N(x^*(\xi^{[N]})) < +\infty$ . If  $\cap_{i=1}^N \text{dom } f_{\xi^i} \subseteq \text{dom } f_\xi$  for some  $\xi$ , then  $x^*(\xi^{[N]}) \in \text{dom } f_\xi$ . Hence,  $d(x^*(\xi^{[N]})) \geq D(\xi^{[N]})$  and

$$\mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1 - \alpha\} \leq \mathbb{P}^N \{D(\xi^{[N]}) < 1 - \alpha\}.$$

Suppose  $\{f_\xi\}$  has chain-constrained domain of order  $m$ . For fixed  $\xi^{[N]}$  and  $\xi \in \Xi$ , if  $\cap_{i=1}^N U_k^{\xi^i} \subseteq U_k^\xi$  for each  $k \in [m]$ , then

$$\bigcap_{i=1}^N \text{dom } f_{\xi^i} = \bigcap_{k=1}^m \left( \bigcap_{i=1}^N U_k^{\xi^i} \right) \subseteq \bigcap_{k=1}^m U_k^\xi = \text{dom } f_\xi.$$

It follows that

$$\mathfrak{D}(\xi^{[N]}) := \mathbb{P} \left\{ \xi \in \Xi : \cap_{i=1}^N U_k^{\xi^i} \subseteq U_k^\xi \ \forall k \in [m] \right\} \leq D(\xi^{[N]})$$

and

$$\mathbb{P}^N \{D(\xi^{[N]}) < 1 - \alpha\} \leq \mathbb{P}^N \{\mathfrak{D}(\xi^{[N]}) < 1 - \alpha\}.$$

Under the technical assumptions of Theorem A.10, Corollary A.14 gives

$$\mathbb{P}^N \left\{ \mathfrak{D}(\xi^{[N]}) < 1 - \alpha \right\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k};$$

roughly speaking, aside from the assumptions on the measurability of some sets, Theorem A.10 assumes that if  $Y \subseteq \Xi$  has a positive measure, then for each  $k \in [m]$ , there exist  $\omega_1, \omega_2 \in Y$  such that

$$\mathbb{P}\{\xi \in Y : U_k^{\omega_1} \subseteq U_k^\xi \subseteq U_k^{\omega_2}\} > 0;$$

this assumption resembles the inner regularity of measures on  $\mathbb{R}$ . Nevertheless, if the chains are induced by functional constraints (i.e.,  $U_k^\xi = \{x : c_k(x) \leq \ell_k(\xi)\}$  for random variables  $\ell_k$ ), then by Corollaries A.12 and A.15, the assumptions of Theorem A.10 are satisfied automatically.

**THEOREM 2.7.** *Let  $m \leq N$  and  $\alpha \in [0, 1]$ . Suppose  $\{f_\xi\}_{\xi \in \Xi}$  has chain-constrained domain of order  $m$ ; then under the assumptions of Theorem A.10,*

$$(2.4) \quad \mathbb{P}^N\{\mathfrak{D}(\xi^{[N]}) < 1 - \alpha\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k};$$

in particular,

$$(2.5) \quad \mathbb{P}^N\{d(x^*(\xi^{[N]})) < 1 - \alpha\} \leq \mathbb{P}^N\{D(\xi^{[N]}) < 1 - \alpha\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

If the chains are induced by functional constraints, then the assumptions of Theorem A.10 are satisfied automatically.

**Remark 2.8.** The bound  $\sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}$  does not necessarily depend on the dimension of the variable  $x$ , which makes it potentially useful in a high-dimensional setting. In section 3.2, we show that the dependence on  $m$  can be mitigated when the problem is convex.

**Remark 2.9.** The sum  $\sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}$  is the tail probability of the binomial distribution  $\text{Bin}(N, \alpha)$ . For  $N\alpha \geq m - 1$ , the Chernoff bound [20] gives the estimate

$$\sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k} \leq \exp \left\{ -\frac{(N\alpha - m + 1)^2}{2N\alpha} \right\}.$$

**3. The convex case.** Throughout this section, we assume  $\mathcal{X}$  is a closed convex set, the set of optimal solutions  $\mathcal{X}^*$  is nonempty and compact, and  $f_\xi$  is convex for all  $\xi \in \Xi$  (this implies  $\text{dom } f_\xi$  is a convex set). In addition, we assume  $F > -\infty$ , and hence  $F$  is a convex function. We also assume that each SAA solution  $x^*(\xi^{[N]})$  is an optimal solution of the SAA problem (1.2).

In the convex case, the feasibility of SAA solutions depends on the local geometry around  $\mathcal{X}^*$ ; the main idea is to combine convexity and the uniform convergence of  $\hat{F}_N$  to  $F$ . A result regarding uniform convergence is summarized in Theorem 3.1, and its proof can be found in [17, section 7.2.10].

For each  $x \in \text{dom } F$ , we define  $M_x(t) := \mathbb{E}[e^{t(f_\xi(x) - F(x))}]$  to be the moment generating function of the random variable  $f_\xi(x) - F(x)$ .

**THEOREM 3.1** (see [17, section 7.2.10]). *Let  $X \subseteq \text{dom } F$  be a compact set of diameter  $D$ . Suppose the following:*

- (C1) *For every  $x \in X$ , the moment generating function  $M_x(t)$  is finitely valued for all  $t$  in a neighborhood of zero.*

(C2) There exists a (measurable) function  $\kappa : \Xi \rightarrow \mathbb{R}_+$  such that  $\mathbb{E}[\kappa(\xi)] = L$  and  $f_\xi$  is  $\kappa(\xi)$ -Lipschitz on  $X$  for all  $\xi$ .

(C3) The moment generating function  $M_\kappa(t) := \mathbb{E}[e^{t\kappa(\xi)}]$  of  $\kappa(\xi)$  is finitely valued for all  $t$  in a neighborhood of zero.

Then for any  $\epsilon > 0$ , there exist positive constants  $C$  and  $\beta$ , independent of  $N$ , such that

$$\mathbb{P}^N \left\{ \sup_{x \in X} |\hat{F}_N(x) - F(x)| \geq \epsilon \text{ or } \frac{1}{N} \sum_{i=1}^N \kappa(\xi^i) > 2L \right\} \leq C e^{-N\beta}.$$

Moreover, if assumption (C1) is replaced by

(C4) there exists constant  $\sigma > 0$  such that for any  $x \in X$ , the inequality

$$M_x(t) \leq e^{\sigma^2 t^2 / 2} \quad \forall t \in \mathbb{R}$$

holds, then for some constants  $\ell$  and  $\rho$ ,

$$\mathbb{P}^N \left\{ \sup_{x \in X} |\hat{F}_N(x) - F(x)| \geq \epsilon \right\} \leq \exp(-N\ell) + 2 \left[ \frac{4\rho DL}{\epsilon} \right]^n \exp \left\{ -\frac{N\epsilon^2}{32\sigma^2} \right\}.$$

If  $\kappa(\xi) \equiv L$  for all  $\xi \in \Xi$ , then the term  $\exp(-N\ell)$  can be omitted.

A thorough discussion of assumptions (C1)–(C4) can be found in [17, section 5.3]. Basically, (C1), (C3), and (C4) assume the existence of moment generating functions in a neighborhood of 0 in order to invoke the large deviation theory. Assumptions (C1) and (C3) hold, for example, if the corresponding random variables are subexponential [20].

**3.1. Solutions in the interior.** When  $\mathcal{X}^*$  is contained in the interior of  $\text{dom } F$ , the uniform convergence alone can guarantee that the portion of infeasible SAA solutions (i.e.,  $\mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1\}$ ) decays exponentially in  $N$ . In particular, the result, which relies on Lemma 3.2, applies to general convex functions  $\{f_\xi\}$  without chain-constrained domain.

**LEMMA 3.2.** *Let  $U$  be a compact convex set and  $V$  be an open set such that  $U \subset V$ ; then there exists a compact convex set  $W$  such that  $U \subset W \subset V$  and  $U \cap \text{bd } W = \emptyset$ .*

*Proof.* Since  $U$  is convex, the function  $\text{dist}_U$  is convex, and so it is also continuous. Since  $U$  is compact and  $V^c$  is closed,

$$r := \inf_{y \in V^c} \text{dist}_U(y) > 0.$$

Consider

$$W := \{x : \text{dist}_U(x) \leq r/2\} = \text{dist}_U^{-1}[0, r/2].$$

Since  $U$  is compact and  $\text{dist}_U$  is convex and continuous,  $W$  is a compact convex set. Moreover,

$$U = \text{dist}_U^{-1}\{0\} \subset W \subset \text{dist}_U^{-1}[0, r] \subseteq V$$

and

$$U \cap \text{bd } W = \text{dist}_U^{-1}(\{0\} \cap \{r/2\}) = \emptyset. \quad \square$$

**THEOREM 3.3.** *Suppose  $\{f_\xi\}$  is a collection of convex functions that satisfy assumptions (C1), (C2), and (C3) in Theorem 3.1 on any compact set  $X \subseteq \text{dom } F$ . If  $\mathcal{X}^*$  is contained in the interior of  $\text{dom } F$ , then there exist positive constants  $C$  and  $\beta$ , independent of  $N$ , such that*

$$\mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1\} \leq Ce^{-N\beta}.$$

*Proof.* Consider

$$\mathcal{B} := \{B : B \text{ compact convex, } \mathcal{X}^* \subseteq B \cap \mathcal{X} \subset \text{int dom } F, \mathcal{X}^* \cap \text{bd } B = \emptyset\}.$$

By Lemma 3.2,  $\mathcal{B}$  is nonempty. We fix a  $B \in \mathcal{B}$  and let  $B_{\mathcal{X}} := B \cap \mathcal{X}$ ; then  $B_{\mathcal{X}}$  is compact and convex, and  $\text{bd } B_{\mathcal{X}} \subseteq \text{bd } B \cup \text{bd } \mathcal{X}$  (see Figure 1). By the extreme value theorem,  $F$  attains the minimum  $F_B^*$  on the compact set  $\text{bd } B_{\mathcal{X}} \cap \text{bd } B$ .

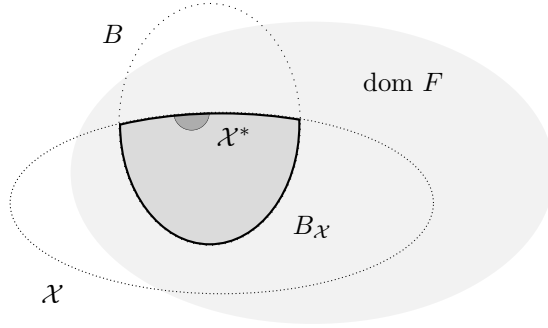


FIG. 1. Illustration of Theorem 3.3.

Fix a  $z \in \mathcal{X}^*$ ; then  $F_B^* > F^* = F(z)$  since  $\mathcal{X}^*$  and  $\text{bd } B$  are disjoint; moreover, the set  $X_B := \{z\} \cup (\text{bd } B_{\mathcal{X}} \cap \text{bd } B)$  is compact. By Theorem 3.1, there exist positive constants  $C_B$  and  $\beta_B$  such that

$$\mathbb{P}^N \left\{ \sup_{x \in X_B} |\hat{F}_N(x) - F(x)| \geq \frac{F_B^* - F^*}{2} \right\} \leq C_B e^{-N\beta_B}.$$

It remains to show

$$\mathbb{P}^N \left\{ \sup_{x \in X_B} |\hat{F}_N(x) - F(x)| \geq \frac{F_B^* - F^*}{2} \right\} \geq \mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1\}.$$

Note that the uniform convergence  $\sup_{x \in X_B} |\hat{F}_N(x) - F(x)| < \frac{F_B^* - F^*}{2}$  implies

$$\min_{x \in \text{bd } B_{\mathcal{X}} \cap \text{bd } B} \hat{F}_N(x) > \min_{x \in \text{bd } B_{\mathcal{X}} \cap \text{bd } B} F(x) - \frac{F_B^* - F^*}{2} = F(z) + \frac{F_B^* - F^*}{2} > \hat{F}_N(z).$$

Let  $x \in B_{\mathcal{X}}^c$ ; then there exists the smallest  $\lambda \in (0, 1]$  such that  $\tilde{x} := \lambda z + (1 - \lambda)x \in \text{bd } B_{\mathcal{X}}$ . If  $\tilde{x} \in \text{bd } B$ , then  $\hat{F}_N(\tilde{x}) > \hat{F}_N(z)$ ; by the convexity of  $\hat{F}_N$ ,

$$\hat{F}_N(\tilde{x}) \leq \lambda \hat{F}_N(z) + (1 - \lambda) \hat{F}_N(x) \implies \hat{F}_N(x) > \hat{F}_N(z),$$

and thus  $x$  cannot be the (optimal) SAA solution. If  $\tilde{x} \in (\text{bd } B_{\mathcal{X}} \setminus \text{bd } B) \subseteq (\text{bd } \mathcal{X} \setminus \text{bd } B)$ , then  $x \in \mathcal{X}^c$ ; i.e.,  $x$  is infeasible. Hence, under uniform convergence, the SAA solutions  $x^*(\xi^{[N]})$  are contained in  $B_{\mathcal{X}} \subseteq \text{dom } F$ , and

$$\mathbb{P}^N \left\{ \sup_{x \in X_B} |\hat{F}_N(x) - F(x)| < \frac{F_B^* - F^*}{2} \right\} \leq \mathbb{P}^N \{d(x^*(\xi^{[N]})) = 1\},$$



or equivalently,

$$\mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1\} \leq \mathbb{P}^N \left\{ \sup_{x \in X_B} |\hat{F}_N(x) - F(x)| \geq \frac{F_B^* - F^*}{2} \right\} \leq C_B e^{-N\beta_B}. \quad \square$$

*Remark 3.4.* We can modify the proof of Theorem 3.3 to obtain different properties of the SAA solutions. Recall

$$\mathcal{B} := \{B : B \text{ compact convex, } \mathcal{X}^* \subseteq B \cap \mathcal{X} \subset \text{int dom } F, \mathcal{X}^* \cap \text{bd } B = \emptyset\}.$$

Consider  $B \in \mathcal{B}$  such that the following hold:

1.  $B = \{x : F(x) \leq F^* + \epsilon\}$  is the  $\epsilon$ -optimal set of  $F$ ; then the result bounds the probability that SAA solutions are not the  $\epsilon$ -optimal solutions.
2.  $B = \{x : \text{dist}(x, \mathcal{X}^*) \leq r\}$  for some radius  $r > 0$ ; then the result bounds the probability that SAA solutions are more than  $r$  units away from  $\mathcal{X}^*$ .

**COROLLARY 3.5.** *Let  $B$  be an  $\epsilon$ -optimal set of  $F$  such that  $B \cap \mathcal{X}$  is compact and is contained in the interior of  $\text{dom } F$ , and let  $D$  denote the diameter of  $B \cap \mathcal{X}$ . Suppose  $\{f_\xi\}$  satisfies assumptions (C2), (C3), and (C4) in Theorem 3.1 on  $B \cap \mathcal{X}$ . Then*

$$\mathbb{P}^N \left\{ F(x^*(\xi^{[N]})) > F^* + \epsilon \right\} \leq \exp(-N\ell) + 2 \left[ \frac{8\rho DL}{\epsilon} \right]^n \exp \left\{ -\frac{N\epsilon^2}{128\sigma^2} \right\}$$

for constants  $\ell$  and  $\rho$  given in Theorem 3.1.

**3.2. Chain-constrained domain revisited.** In this section, we investigate the case that  $\mathcal{X}^*$  has a nonempty intersection with the boundary of  $\text{dom } F$ . We assume  $\{f_\xi\}$  has chain-constrained domain of order  $m$  and

$$(3.1) \quad \text{dom } f_\xi = \{x : c_k(x) \leq \ell_k(\xi), k \in [m]\} \quad \forall \xi \in \Xi,$$

where for each  $k$ ,  $c_k$  is a real-valued convex function and  $\ell_k$  is a random variable supported on  $\Xi$ . For  $\beta \in [0, 1)$ , we define

$$(\ell_k)_\beta := \inf\{t \in \mathbb{R} : \mathbb{P}\{\ell_k \leq t\} > \beta\}.$$

By the right continuity of cumulative distribution functions,  $\mathbb{P}\{\ell_k \leq (\ell_k)_\beta\} \geq \beta$ . Note that  $(\ell_k)_0$  is the essential infimum of  $\ell_k$  and that  $(\ell_k)_0 > -\infty$  for each  $k \in [m]$  since, by assumption,  $F(x) < +\infty$  for some  $x \in \mathcal{X}$ .

In Theorem 2.7, the bound  $\sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}$  was derived for general functions; in stochastic convex programming, the dependence on  $m$  can be mitigated. We show two different approaches, one in Example 3.6 and the other one in Theorem 3.7.

*Example 3.6.* Let us revisit Example 2.4. Suppose  $\{f_\xi\}$  is a collection of convex functions with chain-constrained domain

$$\text{dom } f_\xi = \cap_{i \in [p], j \in J_i, k \in [q]} U_{ijk}^\xi \quad \forall \xi \in \Xi;$$

for each  $i \in [p]$ ,  $j \in J_i$ ,  $k \in [q]$ ,

$$U_{ijk}^\xi := \begin{cases} \{x : a_{ijk}^\top x \leq b_{ij}(\xi)\}, & W_\xi = W_i, T_\xi = T_k, \\ \mathbb{R}^n & \text{otherwise,} \end{cases}$$

where  $a_{ijk}$  are deterministic vectors and  $b_{ij}$  are random variables.

Assume  $\text{dom } F = \{x : \mathbb{P}\{f_\xi(x) < +\infty\} = 1\}$ ; then

$$\text{dom } F = \{x : a_{ijk}^\top x \leq (b_{ij})_0, i \in [p], j \in J_i, k \in [q]\}.$$

For  $\epsilon > 0$ , we denote  $\mathcal{X}_\epsilon := F^{-1}\{F^* + \epsilon\}$ . In Figure 2,  $\text{dom } F$  is the polygon; for some  $\epsilon_1 > \epsilon_2 > 0$ , the sets  $\mathcal{X}_{\epsilon_1}$  and  $\mathcal{X}_{\epsilon_2}$  are the two dash-dotted curves, respectively; the solid line is part of the boundary of  $\text{dom } F$ , and  $\mathcal{X}^*$  is the dot on the solid line.

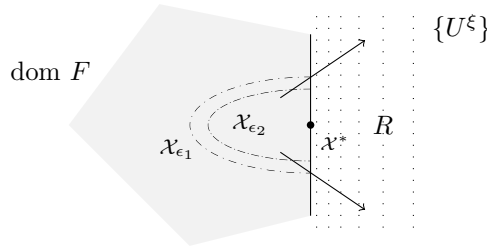


FIG. 2. Illustration of Example 3.6. The arrows pass through  $\mathcal{X}_{\epsilon_2}$  and  $\mathcal{X}_{\epsilon_1}$  consecutively. The dotted lines correspond to the half spaces in the chain  $\{U^\xi\}$ .

Let  $\xi^{[N]}$  be a sample such that  $\hat{F}_N$  approximates  $F$  uniformly on  $\mathcal{X}_{\epsilon_1} \cup \mathcal{X}_{\epsilon_2}$  by an error less than  $(\epsilon_1 - \epsilon_2)/2$ , i.e.,

$$\sup_{x \in \mathcal{X}_{\epsilon_1} \cup \mathcal{X}_{\epsilon_2}} |F(x) - \hat{F}_N(x)| < \frac{\epsilon_1 - \epsilon_2}{2};$$

then

$$\inf_{x \in \mathcal{X}_{\epsilon_1}} \hat{F}_N(x) > \inf_{x \in \mathcal{X}_{\epsilon_1}} F(x) - \frac{\epsilon_1 - \epsilon_2}{2} = \sup_{x \in \mathcal{X}_{\epsilon_2}} F(x) + \frac{\epsilon_1 - \epsilon_2}{2} > \sup_{x \in \mathcal{X}_{\epsilon_2}} \hat{F}_N(x).$$

Let  $x_1 \in \mathcal{X}_{\epsilon_1}$ ,  $x_2 \in \mathcal{X}_{\epsilon_2}$ , and  $\lambda \in (0, 1)$ , and consider  $x$  such that  $x_1 = \lambda x_2 + (1 - \lambda)x$ . Since  $\hat{F}_N(x_2) < \hat{F}_N(x_1)$ ,

$$\hat{F}_N(x_1) \leq \lambda \hat{F}_N(x_2) + (1 - \lambda) \hat{F}_N(x) \implies \hat{F}_N(x) > \hat{F}_N(x_2),$$

which implies  $x$  cannot be the (optimal) SAA solution; in other words, if one draws an arrow that passes through  $\mathcal{X}_{\epsilon_2}$  and  $\mathcal{X}_{\epsilon_1}$  consecutively, then it cannot point toward an SAA solution. In Figure 2, the possible location of SAA solutions is in either  $\text{dom } F$  or a region  $R$  between the arrows. Suppose the dotted lines correspond to the chain  $U^\xi = \{x : a^\top x \leq b(\xi)\}$  and the solid line is the hyperplane  $\{x : a^\top x = (b)_0\}$ . Let  $\alpha > 0$  such that

$$R \cap \{x : a^\top x \leq (b)_\alpha\} \subseteq \{x : a_{ijk}^\top x \leq (b_{ij})_0, a_{ijk} \neq a\};$$

i.e., the set  $R \cap \{x : a^\top x \leq (b)_\alpha\}$  does not intersect the hyperplanes in other chains. If  $\xi^{[N]}$  is a sample such that  $\min_{i \in [N]} b(\xi^i) \leq (b)_\alpha$ , then the SAA solution  $x^*(\xi^{[N]})$  lies in either  $R \cap \{x : a^\top x \leq (b)_\alpha\}$  or  $\text{dom } F$ , which implies  $d(x^*(\xi^{[N]})) \geq 1 - \alpha$ . Note that a random sample  $\xi^{[N]}$  satisfies  $\min_{i \in [N]} b(\xi^i) \leq (b)_\alpha$  with probability at least  $1 - (1 - \alpha)^N \geq 1 - e^{-\alpha N}$ ; if the aforementioned uniform approximation occurs with probability  $1 - O(e^{-cN})$  for some constant  $c$ , then by a union bound,

$$\mathbb{P}^N \{d(x^*(\xi^{[N]})) < 1 - \alpha\} = O(e^{-\min\{\alpha, c\}N}).$$

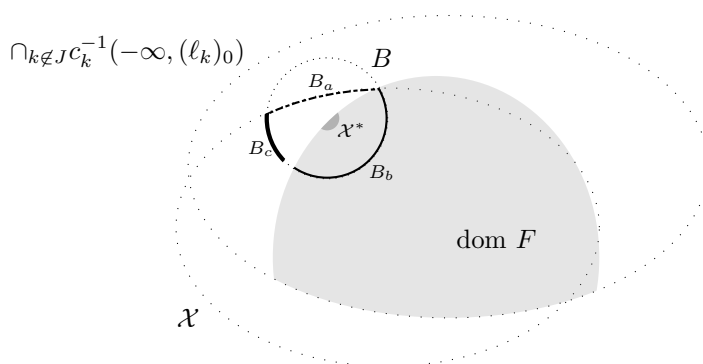


FIG. 3. Illustration of Theorem 3.7.

The following theorem adopts an approach similar to that of Theorem 3.3. It states that, under regularity conditions, the  $m$  in the bound  $\sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}$  can be replaced by  $|J|$ , where  $J$  is the index set of active constraints at  $\mathcal{X}^*$ , i.e.,

$$J := \{k \in [m] : \exists x \in \mathcal{X}^*, c_k(x) = (\ell_k)_0\}.$$

**THEOREM 3.7.** *Let  $\text{dom } f_\xi$  be in the form (3.1), and let  $J$  denote the index set of active constraints at  $\mathcal{X}^*$ . Suppose there exists a function  $\kappa : \Xi \rightarrow \mathbb{R}_+$  such that  $\mathbb{E}[\kappa(\xi)] = L$  and  $f_\xi$  is  $\kappa(\xi)$ -Lipschitz continuous on  $\text{dom } f_\xi$  for every  $\xi \in \Xi$ . Suppose in addition that  $\{f_\xi\}$  satisfies assumptions (C1) and (C3) in Theorem 3.1 on any compact set  $X \subset \text{dom } F$ . Then there exist  $\bar{\alpha} > 0$  and positive constants  $C$  and  $\beta$  (independent of  $N$ ) such that for each  $N \geq |J|$  and  $\alpha \in [0, \bar{\alpha})$ ,*

$$\mathbb{P}\{d(x^*(\xi^{[N]})) < 1 - \alpha\} \leq Ce^{-N\beta} + \sum_{k=0}^{|J|-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}.$$

*Proof.* Let  $x_1 \in \text{dom } F$ , and consider  $x_2$  such that  $\mathbb{P}\{f_\xi(x_2) < +\infty\} = 1$ ; then

$$|F(x_1) - F(x_2)| \leq \mathbb{E}|f_\xi(x_1) - f_\xi(x_2)| \leq \mathbb{E}[\kappa(\xi)\|x_1 - x_2\|] = L\|x_1 - x_2\|;$$

this implies  $F$  is  $L$ -Lipschitz continuous on  $\text{dom } F$ , and

$$\begin{aligned} \text{dom } F &= \{x : \mathbb{P}\{f_\xi(x) < +\infty\} = 1\} \\ &= \{x : c_k(x) \leq (\ell_k)_0 \ \forall k \in [m]\}, \end{aligned}$$

which is a closed convex set.

Since  $c_k$  is a continuous function,  $c_k^{-1}(-\infty, (\ell_k)_0)$  is an open set. For any  $k \notin J$ ,  $c_k^{-1}\{(\ell_k)_0\} \cap \mathcal{X}^* = \emptyset$ , and hence  $\mathcal{X}^* \subseteq \cap_{k \notin J} c_k^{-1}(-\infty, (\ell_k)_0)$ . Consider

$$\mathcal{B} := \{B : B \text{ compact convex}, \mathcal{X}^* \subseteq B \cap \mathcal{X} \subset \cap_{k \notin J} c_k^{-1}(-\infty, (\ell_k)_0), \mathcal{X}^* \cap \text{bd } B = \emptyset\}.$$

By Lemma 3.2,  $\mathcal{B}$  is nonempty. We fix a  $B \in \mathcal{B}$  and let  $B_\mathcal{X} := B \cap \mathcal{X}$ ; then  $B_\mathcal{X}$  is compact and convex, and  $\text{bd } B_\mathcal{X} \subseteq \text{bd } B \cup \text{bd } \mathcal{X}$ . Fixing a  $z \in \mathcal{X}^*$ , we proceed by considering the following three disjoint subsets of  $\text{bd } B_\mathcal{X}$  (see Figure 3):

1.  $B_a := \text{bd } B_\mathcal{X} \setminus \text{bd } B$ . Note that  $B_a \subseteq \text{bd } \mathcal{X} \setminus \text{bd } B$ .

2.  $B_b := \text{bd } B_{\mathcal{X}} \cap \text{bd } B \cap \text{dom } F$ .

Let

$$\epsilon := \min_{x \in B_b} F(x) - F^* > 0.$$

Consider the compact set  $\{z\} \cup B_b$ ; by Theorem 3.1, for some positive constants  $C_B$  and  $\beta_B$ ,

(3.2)

$$\mathbb{P}^N \left\{ \sup_{x \in \{z\} \cup B_b} |\hat{F}_N(x) - F(x)| \geq \frac{\epsilon}{3} \text{ or } \frac{1}{N} \sum_{i=1}^N \kappa(\xi^i) > 2L \right\} \leq C_B e^{-N\beta_B}.$$

Since the complement of the event in (3.2) implies

$$(3.3) \quad \min_{x: \text{dist}(x, B_b) \leq \epsilon/6L} \hat{F}_N(x) > \hat{F}_N(z),$$

we have

$$(3.4) \quad \mathbb{P}^N \left\{ \min_{x: \text{dist}(x, B_b) \leq \epsilon/6L} \hat{F}_N(x) \leq \hat{F}_N(z) \right\} \leq C_B e^{-N\beta_B}.$$

3.  $B_c := \text{bd } B_{\mathcal{X}} \cap \text{bd } B \cap \{x : \text{dist}(x, B_b) \geq \epsilon/6L\}$ .

We define

$$(3.5) \quad \gamma(x) := \mathbb{P}\{\ell_j < c_j(x) \text{ for some } j \in J\}.$$

Since  $B_c \cap B_b = \emptyset$ , we have  $B_c \cap \text{dom } F = \emptyset$  and  $\gamma > 0$  on  $B_c$ . If we can show that  $\gamma$  is lower semicontinuous, then  $\gamma$  attains its minimum  $\bar{\alpha} > 0$  on the compact set  $B_c$ . To see that  $\gamma$  is lower semicontinuous, let  $x \in \mathbb{R}^n$  and  $\eta > 0$ , and for every  $j \in J$ , we can find a  $t_j < c_j(x)$  such that  $\mathbb{P}\{t_j \leq \ell_j < c_j(x)\} \leq \frac{\eta}{|J|}$ . Since  $c_j$  is continuous, there exists  $\delta_j > 0$  such that  $c_j(y) \geq t_j$  whenever  $\|y - x\| < \delta_j$ . Let  $\delta = \min_{j \in J} \delta_j$ , and then for all  $y$  such that  $\|y - x\| < \delta$ , we have

$$\gamma(y) \geq \gamma(x) - \sum_{j \in J} \mathbb{P}\{t_j \leq \ell_j < c_j(x)\} \geq \gamma(x) - \eta;$$

since  $\eta > 0$  is arbitrary,  $\gamma$  is lower semicontinuous at every  $x \in \mathbb{R}^n$ .

We give conditions below that the (optimal) SAA solutions  $x^*(\xi^{[N]})$  are contained in  $B_{\mathcal{X}}$ . Let  $x \in B_{\mathcal{X}}^c$ ; then there exists the smallest  $\lambda \in (0, 1]$  such that  $\tilde{x} := \lambda z + (1 - \lambda)x \in \text{bd } B_{\mathcal{X}}$ . If  $\tilde{x} \in B_a$ , then  $x \in \mathcal{X}^c$ ; i.e.,  $x$  is infeasible. We now assume  $\tilde{x} \in \text{bd } B_{\mathcal{X}} \setminus B_a \subseteq \text{bd } B$ . Suppose  $\text{dist}(\tilde{x}, B_b) \leq \epsilon/6L$  and  $\min_{x: \text{dist}(x, B_b) \leq \epsilon/6L} \hat{F}_N(x) > \hat{F}_N(z)$ ; since  $\hat{F}_N(\tilde{x}) > \hat{F}_N(z)$ , the convexity of  $\hat{F}$  implies  $\hat{F}_N(x) > \hat{F}_N(z)$ , and thus  $x$  is not the SAA solution. It remains to address the case  $\tilde{x} \in B_c$ . We first claim that  $\gamma(x) \geq \gamma(\tilde{x})$ ; indeed, suppose  $c_j(\tilde{x}) > \ell_j(\xi) \geq (\ell_j)_0 \geq c_j(z)$  for some  $\xi \in \Xi$  and  $j \in J$  (note that  $\ell_j \geq (\ell_j)_0$  a.s.). Then by the convexity of  $c_j$ , we have  $c_j(x) > c_j(\tilde{x}) > \ell_j(\xi)$ ; it follows that  $\gamma(x) \geq \gamma(\tilde{x})$ . Let  $\alpha \in (0, \bar{\alpha})$ , and consider a sample  $\xi^{[N]}$  such that

$$\mathfrak{D}_r(\xi^{[N]}) := \mathbb{P} \left\{ \xi \in \Xi : \min_{i \in [N]} \ell_j(\xi^i) \leq \ell_j(\xi) \forall j \in J \right\} \geq 1 - \alpha;$$

then

$$\mathbb{P} \left\{ \ell_j < \min_{i \in [N]} \ell_j(\xi^i) \text{ for some } j \in J \right\} = 1 - \mathfrak{D}_r(\xi^{[N]}) < \bar{\alpha} \leq \gamma(\tilde{x}) \leq \gamma(x),$$

which by (3.5) implies  $c_j(x) > \min_{i \in [N]} \ell_j(\xi^i)$  for some  $j \in J$ ; thus  $x \notin \text{dom } \hat{F}_N$  cannot be the SAA solution. To summarize, we have  $x^*(\xi^{[N]}) \in B_{\mathcal{X}}$  under the following two conditions:

- $\min_{x: \text{dist}(x, B_b) \leq \epsilon/6L} \hat{F}_N(x) > \hat{F}_N(z)$ ;
- $\mathfrak{D}_r(\xi^{[N]}) \geq 1 - \alpha$ .

Moreover, if  $x^*(\xi^{[N]}) \in B_{\mathcal{X}} \subset \cap_{j \notin J} c_j^{-1}(-\infty, (\ell_j)_0)$ , then

$$(3.6) \quad d(x^*(\xi^{[N]})) = \mathbb{P}\{\xi \in \Xi : c_j(x^*(\xi^{[N]})) \leq \ell_j(\xi) \forall j \in J\} \geq \mathfrak{D}_r(\xi^{[N]}).$$

By Corollary A.15,

$$(3.7) \quad \mathbb{P}^N\{\mathfrak{D}_r(\xi^{[N]}) < 1 - \alpha\} \leq \sum_{k=0}^{|J|-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

Combining (3.4), (3.6), and (3.7),

$$\begin{aligned} & \mathbb{P}^N\{d(x^*(\xi^{[N]})) < 1 - \alpha\} \\ & \leq \mathbb{P}^N\left\{\min_{x: \text{dist}(x, B_b) \leq \epsilon/6L} \hat{F}_N(x) \leq \hat{F}_N(z) \text{ or } \mathfrak{D}_r(\xi^{[N]}) < 1 - \alpha\right\} \\ & \leq C_B e^{-N\beta_B} + \sum_{k=0}^{|J|-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}. \end{aligned} \quad \square$$

In the remaining part of this section, we discuss the value of  $|J|$ . Recall the following result from convex optimization.

LEMMA 3.8 (see [17, p. 234]). *Suppose  $f$  and  $f_j$ ,  $j \in [m]$ , are real-valued convex functions on  $\mathbb{R}^n$ ; then there exists an index set  $I \subseteq [m]$  of size at most  $n$  such that the optimal solutions of the convex optimization problem*

$$\min_x \{f(x) : f_j(x) \leq 0, j \in [m]\}$$

*are also the optimal solutions of*

$$\min_x \{f(x) : f_j(x) \leq 0, j \in I\}.$$

*Such an index set  $I$  of the minimum cardinality is called the support indices.*

Let us reinterpret stochastic convex programming problems so that Lemma 3.8 applies. For each  $\xi \in \Xi$ , let  $\hat{f}_\xi$  be a convex extension of  $f_\xi$  to  $\mathbb{R}^n$ . For example, if  $f_\xi$  is  $\kappa(\xi)$ -Lipschitz continuous on  $\text{dom } f_\xi$ , then

$$\hat{f}_\xi(x) := \inf_{y \in \text{dom } f_\xi} \{f_\xi(y) + \kappa(\xi)\|y - x\|\}$$

is a  $\kappa(\xi)$ -Lipschitz convex extension of  $f_\xi$  by the Kirszbraun theorem [13, section 3.3.9]. Let  $I_\xi$  denote the indicator function of  $\text{dom } f_\xi$ . If  $\mathbb{E}[\hat{f}_\xi(x)] > -\infty$  on  $\mathbb{R}^n$ , then

$$F(x) = \mathbb{E}[f_\xi(x)] = \mathbb{E}[\hat{f}_\xi(x) + I_\xi(x)] = \mathbb{E}[\hat{f}_\xi(x)] + \mathbb{E}[I_\xi(x)],$$

where

$$\mathbb{E}[I_\xi(x)] = \begin{cases} 0 & \text{if } c_k(x) \leq (\ell_k)_0 \quad \forall k \in [m], \\ +\infty & \text{otherwise.} \end{cases}$$

Hence,

$$\min_{x \in \mathcal{X}} F(x) \equiv \min_{x \in \mathcal{X}} \left\{ \mathbb{E}[\hat{f}_\xi(x)] : c_k(x) - (\ell_k)_0 \leq 0, k \in [m] \right\},$$

and the two optimization problems share the same set of optimal solutions  $\mathcal{X}^*$ . Since  $\mathbb{E}[\hat{f}_\xi(x)]$  is a convex function in  $x$ , if the active constraints at  $\mathcal{X}^*$  correspond to the support indices, then  $|J| \leq n$  by Lemma 3.8.

**4. Feasibility in multistage stochastic programming.** In this section, we extend the notion of degree of feasibility to multistage stochastic programming. Let  $T \geq 2$  be the number of stages, and let  $\xi_t, t = 2, \dots, T$ , be a stagewise independent data process; i.e.,  $\xi_t$ 's are mutually independent random vectors defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  (we may also use  $\xi_t$  to denote the outcomes in  $\Xi_t := \xi_t(\Omega)$ ). We denote  $\xi_{[t]} := (\xi_1, \dots, \xi_t)$  to be the history of the process up to time  $t$ . Consider the  $T$ -stage stochastic programming problem

$$(4.1) \quad \begin{aligned} \min_{x_1, \mathbf{x}_2, \dots, \mathbf{x}_T} \quad & \mathbb{E} [f_1(x_1) + f_2(\mathbf{x}_2(\xi_{[2]}), \xi_2) + \dots + f_T(\mathbf{x}_T(\xi_{[T]}), \xi_T)] \\ \text{s.t. } & x_1 \in \mathcal{X}_1, \mathbf{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\xi_{[t-1]}), \xi_t), t = 2, \dots, T. \end{aligned}$$

Here the decision vectors  $x_t = \mathbf{x}_t(\xi_{[t]}) \in \mathbb{R}^{n_t}$ ,  $t \in [T]$ , are functions of the data process  $\xi_{[t]}$  up to time  $t$ ,  $f_t$  are real-valued functions, and

$$\mathcal{X}_t(x_{t-1}, \xi_t) := \{x_t \geq 0 : A_t x_t + B_t x_{t-1} = b_{\xi_t}\},$$

where  $A_t$  and  $B_t$  are deterministic matrices, and  $b_{\xi_t}$  is a random vector supported on  $\Xi_t$ . The first stage data  $f_1$  and  $\mathcal{X}_1$  are deterministic. A thorough discussion of multistage stochastic programming can be found in [17].

The identical conditional sampling generates an i.i.d. sample  $\xi_t^{[N_t]} = (\xi_t^1, \dots, \xi_t^{N_t})$  of  $\xi_t$  for each  $t = 2, \dots, T$ . The (directed) scenario tree generated by the sample  $\xi^{[N, T]} := (\xi_2^{[N_2]}, \dots, \xi_T^{[N_T]})$  is constructed by first creating the root node  $\xi_1$  and then creating an arc from each node  $\xi_{t-1}^1, \dots, \xi_{t-1}^{N_{t-1}}$  to each node  $\xi_t^1, \dots, \xi_t^{N_t}$ ,  $t = 2, \dots, T$ . The multistage stochastic programming problem induced by the original problem (4.1) on the scenario tree is viewed as the SAA problem of (4.1); note that the data process of the induced SAA problem is still stagewise independent. We denote  $\mathcal{P}_t(\xi^{[N, T]})$  to be the collection of paths from the root to a node at stage  $t$ ; i.e., each  $p_t \in \mathcal{P}_t(\xi^{[N, T]})$  has the form  $p_t = (\xi_1, \xi_2^{i_2}, \dots, \xi_t^{i_t})$ ,  $i_s \in [N_s]$ .

For a generated sample  $\xi^{[N, T]}$ , we assume there is a mapping that assigns a decision  $x_t^*(p_t)$  to each  $p_t = (\xi_1, \xi_2^{i_2}, \dots, \xi_t^{i_t}) \in \mathcal{P}_t(\xi^{[N, T]})$  for all  $t \in [T]$ , such that  $x_1^* = x_1^*(\xi_1) \in \mathcal{X}_1$ , and for the subpath  $p_{s-1} = (\xi_1, \xi_2^{i_2}, \dots, \xi_{s-1}^{i_{s-1}})$ ,  $s = 2, \dots, t$ , we have  $x_s^*(p_s) \in \mathcal{X}_s(x_{s-1}^*(p_{s-1}), \xi_s^{i_s})$ ; for example, one such mapping is given by an optimization algorithm that solves the SAA problem and outputs feasible solutions. It is important to note that the solutions  $x_t^*(p_t)$  depend on both the sample  $\xi^{[N, T]}$  and the specific path  $p_t$  in the scenario tree generated by the sample  $\xi^{[N, T]}$ .

The topology of the scenario tree (e.g.,  $\mathcal{P}_t(\xi^{[N, T]})$ ) is determined by the parameters  $N_2, \dots, N_T$ . By fixing  $N_t$ 's and viewing  $\xi^{[N, T]}$  as random samples, we treat each  $x_t^*(p_t)$  as a random vector with respect to  $\xi^{[N, T]}$ . The relatively complete recourse for multistage stochastic programming states that for every  $t \leq T - 1$ , it holds that for almost every  $\xi_{[t]}$  and any sequence of feasible solutions  $x_i = \mathbf{x}_i(\xi_{[i]})$ ,  $i = 1, \dots, t$ , the set  $\mathcal{X}_{t+1}(x_t, \xi_{t+1})$  is nonempty for almost every  $\xi_{t+1}$ . Without relatively complete recourse, we are interested in the probability that  $\mathcal{X}_t(x_{t-1}^*(p_{t-1}), \xi_t)$  is nonempty.

DEFINITION 4.1. Let  $\xi^{[N,T]}$  be a generated sample. For each  $t = 2, \dots, T$  and  $p \in \mathcal{P}_{t-1}(\xi^{[N,T]})$ , the degree of feasibility of  $x_{t-1}^*(p)$  is

$$d_t(x_{t-1}^*(p)) := \mathbb{P}\{\xi_t \in \Xi_t : \mathcal{X}_t(x_{t-1}^*(p), \xi_t) \text{ is nonempty}\}.$$

Definition 4.1 coincides with Definition 1.2 when  $t = T = 2$ . We can apply Theorem 2.7 to obtain a probabilistic bound on  $\min_{p \in \mathcal{P}_{t-1}(\xi^{[N,T]})} d_t(x_{t-1}^*(p))$ , which is the minimum degree of feasibility over all solutions at stage  $t$ ; here the probability measure is the product measure  $\mathbb{P}^{N,T} = \mathbb{P}^{N_2} \times \dots \times \mathbb{P}^{N_T}$  associated with the random sample  $\xi^{[N,T]}$ .

COROLLARY 4.2. Consider a  $T$ -stage problem of the form (4.1); we denote  $m_t$  to be the number of extreme rays of the cone  $\{r : r^\top A_t \geq 0\}$  for each  $t = 2, \dots, T$ . Let  $\xi^{[N,T]} = (\xi_2^{[N_2]}, \dots, \xi_T^{[N_T]})$  be a random sample such that  $N_t \geq m_t$ . Then for  $I \subseteq \{2, \dots, T\}$  and  $\alpha_t \in [0, 1]$ ,

$$\mathbb{P}^{N,T} \left\{ \min_{p \in \mathcal{P}_{t-1}(\xi^{[N,T]})} d_t(x_{t-1}^*(p)) \geq 1 - \alpha_t, t \in I \right\} \geq \prod_{t \in I} \left( \sum_{k=m_t}^{N_t} \binom{N_t}{k} \alpha_t^k (1 - \alpha_t)^{N_t-k} \right);$$

in particular, for  $\alpha \in [0, 1]$  and  $t = 2, \dots, T$ ,

$$\mathbb{P}^{N,T} \left\{ \min_{p \in \mathcal{P}_{t-1}(\xi^{[N,T]})} d_t(x_{t-1}^*(p)) < 1 - \alpha \right\} \leq \sum_{k=0}^{m_t-1} \binom{N_t}{k} \alpha^k (1 - \alpha)^{N_t-k}.$$

*Proof.* For  $t = 2, \dots, T$ , consider

$$S_{\xi_t} := \{x_{t-1} : \mathcal{X}_t(x_{t-1}, \xi_t) \text{ is nonempty}\};$$

then  $d_t(x_{t-1}^*(p)) = \mathbb{P}\{\xi_t \in \Xi_t : x_{t-1}^*(p) \in S_{\xi_t}\}$ . We define

$$D(\xi_t^{[N_t]}) := \mathbb{P}\left\{\xi_t \in \Xi_t : \bigcap_{i=1}^{N_t} S_{\xi_t^i} \subseteq S_{\xi_t}\right\}.$$

Since  $x_{t-1}^*(p) \in S_{\xi_t^i}$  for each  $p \in \mathcal{P}_{t-1}(\xi^{[N,T]})$  and  $i = 1, \dots, N_t$ , we have

$$\min_{p \in \mathcal{P}_{t-1}(\xi^{[N,T]})} d_t(x_{t-1}^*(p)) \geq D(\xi_t^{[N_t]}).$$

By Farkas' lemma,  $S_{\xi_t} = \{x : r_i^\top B_t x \leq r_i^\top b_{\xi_t} \text{ for all } i\}$ , where  $\{r_i\}$  is the set of extreme rays of the polyhedral cone  $\{r : r^\top A_t \geq 0\}$ . By viewing  $\{S_{\xi_t^i}\}$  as chain-constrained domain of order  $m_t$ , we can apply Theorem 2.7 to obtain the bound

$$\mathbb{P}^{N,T} \left\{ D(\xi_t^{[N_t]}) < 1 - \alpha \right\} = \mathbb{P}^{N_t} \left\{ D(\xi_t^{[N_t]}) < 1 - \alpha \right\} \leq \sum_{k=0}^{m_t-1} \binom{N_t}{k} \alpha^k (1 - \alpha)^{N_t-k};$$

by stagewise independence,

$$\mathbb{P}^{N,T} \left\{ D(\xi_t^{[N_t]}) \geq 1 - \alpha_t, t \in I \right\} \geq \prod_{t \in I} \left( \sum_{k=m_t}^{N_t} \binom{N_t}{k} \alpha_t^k (1 - \alpha_t)^{N_t-k} \right).$$

Therefore,

$$\begin{aligned}
& \mathbb{P}^{N,T} \left\{ \min_{p \in \mathcal{P}_{t-1}(\xi^{[N,T]})} d_t(x_{t-1}^*(p)) \geq 1 - \alpha_t, t \in I \right\} \\
& \geq \mathbb{P}^{N,T} \left\{ D(\xi_t^{[N_t]}) \geq 1 - \alpha_t, t \in I \right\} \\
& \geq \prod_{t \in I} \left( \sum_{k=m_t}^{N_t} \binom{N_t}{k} \alpha_t^k (1 - \alpha_t)^{N_t-k} \right). \quad \square
\end{aligned}$$

**5. Conclusions.** In situations where the SAA solutions could be infeasible to the true problem (1.1), it is shown that for functions with chain-constrained domain, the portion of SAA solutions having a low degree of feasibility decays exponentially when the sample size increases. For convex problems, estimates of this rate can be improved; in particular, when the true solutions are contained in the interior of the domain of the expectation function, the portion of infeasible SAA solutions decays exponentially as the sample size increases, and this result holds even without chain-constrained domain. In the multistage case, the result is extended to bound the minimum degree of feasibility of SAA solutions over all sample paths.

**Appendix A. Probability on chains.** In this section, we address the mathematical derivation of the main concept introduced in section 2, chain-constrained domain. Let us first recall the definition of chains.

**DEFINITION A.1.** A binary relation “ $\leq_R$ ” on a set  $X$  is a partial ordering of  $X$  if it is transitive (i.e.,  $x \leq_R y$  and  $y \leq_R z$  implies  $x \leq_R z$ ), reflexive (i.e.,  $x \leq_R x$  for every  $x \in X$ ), and antisymmetric (i.e.,  $x \leq_R y$  and  $y \leq_R x$  implies  $x = y$ ). If  $\leq_R$  is a partial ordering on  $X$ , then  $(X, \leq_R)$  is called a partially ordered set.

We do not necessarily need antisymmetry; instead, if  $x \leq_R y$  and  $y \leq_R x$ , we write  $x =_R y$  and treat them as the equivalence class  $\{y : y =_R x\}$ . We use  $x \geq_R y$  to denote  $y \leq_R x$ , and we use  $x <_R y$  to denote  $x \leq_R y$  and  $x \neq_R y$ .

**DEFINITION A.2.** A nonempty set  $C \subseteq X$ , where  $(X, \leq_R)$  is a partially ordered set, is a chain in  $X$  if for any  $x, y \in C$ , we have either  $x \leq_R y$  or  $y \leq_R x$ .

Let  $(X, \mathcal{A}, \mathbb{P})$  be a probability space. Suppose there exist  $m$  binary relations  $\leq_k$ ,  $k \in [m]$ , such that for each  $k$ ,  $X$  is a chain (possibly without antisymmetry) with respect to the partially ordered set  $(X, \leq_k)$ . We assume the  $\sigma$ -algebra  $\mathcal{A}$  contains sets of the form  $\{y \in X : y =_k x\}$  and  $\{y \in X : y \leq_k x\}$  for each  $x \in X$  and  $k \in [m]$  (note that this is similar to the construction of the Borel  $\sigma$ -algebra of  $\mathbb{R}$ ). More assumptions on measurability will be stated when needed.

Let  $\mathbf{x} = (x_i)_i \in X^N$  be an  $N$ -tuple, where  $N \geq m$ . We define  $h_k(\mathbf{x})$  to be the minimum of  $x_i$ 's under the relation  $\leq_k$ ; i.e.,  $h_k$  is an entry  $y$  of  $\mathbf{x}$  such that  $y \leq_k x_i$  for all  $i \in [N]$ . Consider the function

$$\mathfrak{D}(\mathbf{x}) = \mathbb{P}\{x \in X : h_k \leq_k x \ \forall k \in [m]\};$$

in plain words, the function  $\mathfrak{D}$  maps the tuple  $\mathbf{x}$  to the measure of the set of  $x$  that is no less than the minimum entry of  $\mathbf{x}$  for each relation  $\leq_k$ . For example, suppose  $1 <_1 2 <_1 3$  and  $3 <_2 2 <_2 1$  and  $\mathbf{x} = (1, 3)$ ; then

$$\{x : h_k \leq_k x \text{ for } k = 1, 2\} = \{1, 2, 3\}.$$



Throughout this section, we are interested in the probability

$$(A.1) \quad \mathbb{P}^N\{\mathbf{x} \in X^N : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\},$$

where  $\alpha \in (0, 1)$  and  $\mathbb{P}^N$  is the completion of the (unique) product measure  $\prod_{i=1}^N \mathbb{P}$ .

*Assumption A.3.* The function  $\mathfrak{D}$  is a random variable (with respect to the product space).

**A.1. The “atomless” case.** We first bound (A.1) under the condition

$$(A.2) \quad \mathbb{P}^N\{\mathbf{x} \in X^N : x_i =_k x_{i'}\} = 0 \quad \forall k \in [m], i \neq i' \in [N].$$

The key step is to partition the event  $\{1 - \mathfrak{D} > \alpha\}$ . Observe that

$$1 - \mathfrak{D}(\mathbf{x}) = \mathbb{P}\{x \in X : x <_k h_k \text{ for some } k \in [m]\}.$$

For each  $\mathbf{x} \in X^N$ , let  $\tilde{\mathbf{x}} = (\tilde{x}_i)_i$  be a permutation of  $\mathbf{x}$  such that for each  $k \in [m]$ ,  $\tilde{x}_k$  is the minimum of  $\tilde{x}_i$ ,  $k \leq i \leq N$ , under the relation  $\leq_k$  (i.e.,  $\tilde{x}_k \leq_k \tilde{x}_{k'}$  for all  $k' \geq k$ ). For example, suppose  $3 =_1 2 <_1 1$  and  $1 <_2 2 <_2 3$  and  $\mathbf{x} = (1, 2, 3)$ ; then the possible choices of  $\tilde{\mathbf{x}}$  are  $(3, 1, 2)$  and  $(2, 1, 3)$ . Certainly, under condition (A.2), here is a unique choice of the first  $m$  indices of  $\tilde{\mathbf{x}}$  almost surely. Now we define  $\delta_0 = 0$  and for each  $k \in [m]$  define

$$\delta_k(\mathbf{x}) = \mathbb{P}\{x \in X : x <_j \tilde{x}_j \text{ for some } j \in [k]\}.$$

*Assumption A.4.* Each  $\delta_k$  is a random variable.

Consider the disjoint events

$$E_k = \{\delta_{k-1} \leq \alpha, \delta_k > \alpha\}, \quad k \in [m].$$

Under condition (A.2), the measure of  $E_k$  does not depend on the particular choice of  $\tilde{\mathbf{x}}$ . We claim that  $\{1 - \mathfrak{D} > \alpha\} \subseteq \cup_{k \in [m]} E_k$ . Indeed, let  $\mathbf{x} \in X^N$  such that  $1 - \mathfrak{D}(\mathbf{x}) > \alpha$ ; since  $h_k \leq_k \tilde{x}_k$  for each  $k \in [m]$ , we have  $\delta_m(\mathbf{x}) \geq 1 - \mathfrak{D}(\mathbf{x}) > \alpha$ . Since  $\delta_k(\mathbf{x})$  is nondecreasing in  $k$  and  $\delta_0 = 0$ , there exists  $k'$  such that  $\delta_{k'-1}(\mathbf{x}) \leq \alpha$  and  $\delta_{k'}(\mathbf{x}) > \alpha$ , i.e.,  $\mathbf{x} \in E_{k'}$ . Hence,

$$\mathbb{P}^N\{1 - \mathfrak{D} > \alpha\} \leq \mathbb{P}^N\left(\bigcup_{k=1}^m E_k\right) = \sum_{k=1}^m \mathbb{P}^N(E_k).$$

It remains to bound the probability  $\mathbb{P}^N(E_k)$ . For each  $k \in [m]$ , we define

$$S_k(x) = \mathbb{P}\{y \in X : y \geq_k x\}.$$

*Assumption A.5.* The sets  $\{x \in X : S_k(x) = \beta\}$ ,  $\beta \in [0, 1]$ , are measurable.

*Assumption A.6.* If  $Y \subseteq X$  is a measurable set with  $\mathbb{P}(Y) > 0$ , then

$$\sup_{x, y \in Y} \mathbb{P}\{z \in Y : x \leq_k z \leq_k y\} > 0.$$

Condition (A.2) implies  $\mathbb{P}\{y \in X : y =_1 x\} = 0$  for each  $x \in X$ ; combining this fact with Assumptions A.5 and A.6, one can verify

$$\mathbb{P}\{x \in X : S_1(x) = \beta\} = 0 \quad \forall \beta \in [0, 1].$$

Indeed, suppose for contradiction that  $\mathbb{P}\{x \in X : S_1(x) = \beta\} > 0$ ; by Assumption A.6, there exists  $x <_1 y$  such that  $S_1(x) = S_1(y) = \beta$  and  $\mathbb{P}\{z \in X : x \leq_1 z \leq_1 y\} > 0$ , but this also implies  $S_1(x) = S_1(y) + \mathbb{P}\{z \in X : x \leq_1 z <_1 y\} > \beta$ , a contradiction.

Observe that

$$\mathbb{P}^N(E_1) = \mathbb{P}^N\{1 - \delta_1 < 1 - \alpha\} = (\mathbb{P}\{x \in X : S_1(x) < 1 - \alpha\})^N.$$

Let  $\beta_0 = \sup\{S_1(x) : S_1(x) < 1 - \alpha, x \in X\}$ ; then  $\beta_0 \leq 1 - \alpha$ . If there exists  $x_0$  such that  $S_1(x_0) = \beta_0$ , then

$$\mathbb{P}\{x \in X : S_1(x) < 1 - \alpha\} \leq \mathbb{P}\{x \in X : S_1(x) = \beta_0\} + S_1(x_0) \leq 1 - \alpha;$$

otherwise, we can find a sequence  $\{y_n\}$  such that  $y_1 >_1, y_2 >_1 \cdots, S_1(y_n) \uparrow \beta_0$ , and

$$\mathbb{P}\{x \in X : S_1(x) < 1 - \alpha\} = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{y \in X : y \geq_1 y_n\}\right) = \lim_n S_1(y_n) \leq 1 - \alpha.$$

A similar argument gives  $\mathbb{P}\{x \in X : S_1(x) \geq 1 - \alpha\} \leq \alpha$ . Thus,

$$\mathbb{P}^N(E_1) = (\mathbb{P}\{x \in X : S_1(x) < 1 - \alpha\})^N = (1 - \alpha)^N.$$

In fact, we can replace  $\alpha$  by  $\beta \in [0, 1]$  in the argument above to derive

$$\mathbb{P}\{x \in X : S_1(x) < 1 - \beta\} = 1 - \beta \quad \forall \beta \in [0, 1].$$

*Assumption A.7.* The set  $\{\mathbf{x} \in X^N : x_j =_j \tilde{x}_j, j \in [k]\}$  is measurable.

We next bound  $\mathbb{P}^N(E_{k+1})$  for  $k \geq 1$ . Consider the event

$$\begin{aligned} E_{k+1}^* &= E_{k+1} \cap \{\mathbf{x} \in X^N : x_j =_j \tilde{x}_j, j \in [k]\} \\ &= \{\mathbf{x} \in X^N : \delta_k(\mathbf{x}) \leq \alpha, \delta_{k+1}(\mathbf{x}) > \alpha, x_j =_j \tilde{x}_j, j \in [k]\}. \end{aligned}$$

By permutation,

$$(A.3) \quad \mathbb{P}^N(E_{k+1}) = \frac{N!}{(N-k)!} \cdot \mathbb{P}^N(E_{k+1}^*).$$

For each  $\mathbf{x}_k \in X^k$ , consider the  $\mathbf{x}_k$ -section of  $E_{k+1}^*$ :

$$(E_{k+1}^*)_{\mathbf{x}_k} = \{\mathbf{u}_{N-k} \in X^{N-k} : (\mathbf{x}_k, \mathbf{u}_{N-k}) \in E_{k+1}^*\};$$

then Tonelli's theorem implies

$$(A.4) \quad \mathbb{P}^N(E_{k+1}^*) = \int \mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k} d\mathbb{P}^k(\mathbf{x}_k).$$

If we can show that

1. the probability  $\mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k}$  is either 0 or  $(1 - \alpha)^{N-k}$  and
  2. the set of  $\mathbf{x}_k$  with  $\mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k} = (1 - \alpha)^{N-k}$  has measure  $\frac{\alpha^k}{k!}$ ,
- then (A.3) and (A.4) together yield

$$\mathbb{P}^N(E_{k+1}) = \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

To show item 1, let us fix an  $\mathbf{x}_k = (x_j)_j \in X^k$  such that  $x_j \leq_j x_{j'}$  for  $j \leq j'$  and

$$\mathbb{P}\{x \in X : x <_j x_j \text{ for some } j \in [k]\} = \alpha_{\mathbf{x}_k} \leq \alpha.$$

Consider

$$\begin{aligned} X_{\mathbf{x}_k} &= X \setminus \{x \in X : x <_j x_j \text{ for some } j \in [k]\} \\ &= \{x \in X : x \geq_j x_j \ \forall j \in [k]\} \end{aligned}$$

and

$$S_{\mathbf{x}_k}(x) = \mathbb{P}\{y \in X_{\mathbf{x}_k} : y \geq_{k+1} x\}.$$

By condition (A.2) and Assumptions A.5 and A.6, the set  $\{x \in X_{\mathbf{x}_k} : S_{\mathbf{x}_k}(x) = \beta\}$  is measurable and has measure 0 for each  $\beta \in [0, 1 - \alpha_{\mathbf{x}_k}]$ . Using an argument similar to that in the  $E_1$  case, we can show that

$$(A.5) \quad \mathbb{P}\{x \in X_{\mathbf{x}_k} : S_{\mathbf{x}_k}(x) < 1 - \beta\} = 1 - \beta \quad \forall \beta \in [0, 1 - \alpha_{\mathbf{x}_k}].$$

It follows that

$$\mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k} = (\mathbb{P}\{x \in X_{\mathbf{x}_k} : S_{\mathbf{x}_k}(x) < (1 - \alpha_{\mathbf{x}_k}) - (\alpha - \alpha_{\mathbf{x}_k})\})^{N-k} = (1 - \alpha)^{N-k}.$$

Note that  $\mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k} = 0$  for other  $\mathbf{x}_k$ 's.

To show item 2, observe that

$$\begin{aligned} &\mathbb{P}^k\{\mathbf{x}_k \in X^k : \mathbb{P}^{N-k}(E_{k+1}^*)_{\mathbf{x}_k} = (1 - \alpha)^{N-k}\} \\ &= \mathbb{P}^k\{\mathbf{x}_k \in X^k : \mathbb{P}(X_{\mathbf{x}_k}^c) \leq \alpha, x_j \leq_j x_{j'} \ \forall j \leq j'\} \\ &= \int_{x_1 \in X : \mathbb{P}(X_{\mathbf{x}_1}^c) \leq \alpha} \int_{x_2 \in X_{\mathbf{x}_1} : \mathbb{P}(X_{\mathbf{x}_2}^c) \leq \alpha} \cdots \int_{x_k \in X_{\mathbf{x}_{k-1}} : \mathbb{P}(X_{\mathbf{x}_k}^c) \leq \alpha} \mathbf{1} \, d\mathbb{P}(x_k) \cdots d\mathbb{P}(x_1), \end{aligned}$$

where the second equality follows by applying Tonelli's theorem  $k$  times. In the multiple integral,  $\mathbf{x}_j \in X^j$  is the  $j$ -tuple with entries  $x_1, \dots, x_j$ ; in particular, if  $j \leq j'$ , then  $\mathbf{x}_j$  is the first  $j$  entries of  $\mathbf{x}_{j'}$ . For each  $j \in [k]$ , we denote  $t_j$  to be  $\mathbb{P}(X_{\mathbf{x}_j}^c) - \mathbb{P}(X_{\mathbf{x}_{j-1}}^c)$ . Let us fix an appropriate  $\mathbf{x}_{j-1}$  (that appears in the multiple integral); then  $\sum_{i=1}^{j-1} t_i \leq \alpha$  is fixed. Since

$$t_j = \mathbb{P}(X_{\mathbf{x}_j}^c \setminus X_{\mathbf{x}_{j-1}}^c) = \mathbb{P}\{x \in X_{\mathbf{x}_{j-1}} : x <_j x_j\} = 1 - S_{\mathbf{x}_{j-1}}(x_j),$$

by (A.5), for  $\beta \in [0, 1 - \sum_{i=1}^{j-1} t_i]$ , we have

$$\mathbb{P}\{x_j \in X_{\mathbf{x}_{j-1}} : t_j \leq \beta\} = \mathbb{P}\{x_j \in X_{\mathbf{x}_{j-1}} : S_{\mathbf{x}_{j-1}}(x_j) \geq 1 - \beta\} = \beta;$$

in particular, let  $d_j = \alpha - \sum_{i=1}^{j-1} t_i$  and  $n \in \mathbb{N}$ . Then

$$\mathbb{P}\left\{x_j \in X_{\mathbf{x}_{j-1}} : \frac{(a-1)d_j}{n} < t_j \leq \frac{ad_j}{n}\right\} = \frac{d_j}{n} \quad \forall a \in [n].$$

If we calculate the integral by simple functions on the sets

$$\left\{x_j \in X_{\mathbf{x}_{j-1}} : \frac{(a-1)d_j}{n} < t_j \leq \frac{ad_j}{n}\right\}, \quad j \in [k], n \in \mathbb{N}, a \in [n],$$

then

$$\begin{aligned}
& \int_{x_1 \in X: \mathbb{P}(X_{\mathbf{x}_1}^c) \leq \alpha} \int_{x_2 \in X_{\mathbf{x}_1}: \mathbb{P}(X_{\mathbf{x}_2}^c) \leq \alpha} \cdots \int_{x_k \in X_{\mathbf{x}_{k-1}}: \mathbb{P}(X_{\mathbf{x}_k}^c) \leq \alpha} \mathbf{1} \, d\mathbb{P}(x_k) \cdots d\mathbb{P}(x_1) \\
&= \int_{x_1 \in X: t_1 \leq \alpha} \int_{x_2 \in X_{\mathbf{x}_1}: \sum_{j=1}^2 t_j \leq \alpha} \cdots \int_0^{\alpha - \sum_{j=1}^{k-1} t_j} \mathbf{1} \, dt_k \cdots d\mathbb{P}(x_2) d\mathbb{P}(x_1) \\
&\vdots \\
&= \int_0^\alpha \int_0^{\alpha - t_1} \cdots \int_0^{\alpha - \sum_{j=1}^{k-1} t_j} \mathbf{1} \, dt_k \cdots dt_2 dt_1 \\
&= \frac{\alpha^k}{k!}.
\end{aligned}$$

Thus far, we have proved the following lemma.

LEMMA A.8. *Let  $N \geq m$  and  $\alpha \in [0, 1]$ . Suppose*

$$\mathbb{P}^N\{\mathbf{x} \in X^N : x_i =_k x_{i'}\} = 0 \quad \forall k \in [m], i \neq i' \in [N],$$

*and Assumptions A.3 to A.7 are satisfied; then*

$$\mathbb{P}^N\{\mathbf{x} \in X^N : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

*Remark A.9.* Assumption A.6 is the only assumption that is not related to measurability. Without this assumption, it may happen that

$$\mathbb{P}\{y \in X : y <_R x\} = 1$$

for every  $x \in X$ . The following construction relies on the axiom of choice:<sup>1</sup> by the well ordering principle, there exists the set of countable ordinals  $\Omega$  that is well ordered by  $\leq_w$  [7] (note that  $\Omega$  itself is uncountable); consider the probability space  $(\Omega, \mathcal{M}, \mu)$ , where

$$\mathcal{M} = \{U \subseteq \Omega : U \text{ countable or } U^c \text{ countable}\},$$

and  $\mu(U) = 0$  if  $U$  is countable and  $\mu(U) = 1$  if  $U^c$  is countable. Let  $\leq_R$  be the converse of  $\leq_w$  (i.e.,  $x \leq_R y \iff y \leq_w x$ ); it can be verified that  $\mathcal{M}$  contains sets of the form  $\{y =_R x\}$  and  $\{y \leq_R x\}$  for each  $x \in \Omega$ . Since the set

$$\{y \in \Omega : y \geq_R x\} = \{y \in \Omega : y \leq_w x\}$$

is countable for every  $x \in \Omega$ , we have

$$\mu\{y \in \Omega : y <_R x\} = 1 - \mu\{y \in \Omega : y \geq_R x\} = 1.$$

**A.2. The general case.** We deal with the general case by constructing a new probability space that satisfies condition (A.2). Let  $([0, 1], \mathcal{L}, \lambda)$  be the probability space such that  $\mathcal{L}$  is the Lebesgue  $\sigma$ -algebra on  $[0, 1]$ , and  $\lambda$  is the Lebesgue measure. Consider the product space  $(\mathfrak{X}, \mathcal{A} \otimes \mathcal{L}, \mathbb{P}_\lambda)$ , where  $\mathfrak{X} = X \times [0, 1]$ , and  $\mathbb{P}_\lambda = \mathbb{P} \times \lambda$  is the product measure. For each  $k \in [m]$ , we define a binary relation  $\leq_k$  on  $\mathfrak{X}$ : we say that

<sup>1</sup>The construction is based on the example given in <https://mathoverflow.net/q/343303>.

$(x, s) \leq_k (y, t)$  if either  $x <_k y$ , or  $x =_k y$  and  $s \leq t$ ; note that  $\mathfrak{X}$  is a chain (possibly without antisymmetry) with respect to the partially ordered set  $(\mathfrak{X}, \leq_k)$ . It can be verified that the  $\sigma$ -algebra  $\mathcal{A} \otimes \mathcal{L}$  contains sets of the form  $\{(y, t) : (y, t) =_k (x, s)\}$  and  $\{(y, t) : (y, t) \leq_k (x, s)\}$  for each  $(x, s) \in \mathfrak{X}$  and  $k \in [m]$ .

We denote  $\mathbb{P}_\lambda^N$  to be the completion of the product measure  $\prod_{i=1}^N \mathbb{P}_\lambda$ . Observe that

$$\{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : (x_i, s_i) =_k (x_{i'}, s_{i'})\} \subseteq \{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : s_i = s_{i'}\};$$

since the latter is a  $\mathbb{P}_\lambda^N$ -null set when  $i \neq i'$ , we have

$$\mathbb{P}_\lambda^N\{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : (x_i, s_i) =_k (x_{i'}, s_{i'})\} = 0 \quad \forall k \in [m], i \neq i' \in [N].$$

We define the function  $\mathfrak{D}_\lambda$  on  $\mathfrak{X}^N$  similarly; i.e., we let  $(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N$  and denote  $h_k(\mathbf{x}, \mathbf{s})$  to be the minimum of  $(x_i, s_i)'$ s under the relation  $\leq_k$ . Then

$$\mathfrak{D}_\lambda(\mathbf{x}, \mathbf{s}) = \mathbb{P}_\lambda\{(x, s) \in \mathfrak{X} : h_k \leq_k (x, s) \forall k \in [m]\}.$$

If  $(\mathfrak{X}, \mathcal{A} \otimes \mathcal{L}, \mathbb{P}_\lambda)$  satisfies Assumptions A.4 to A.7, then by the proof of Lemma A.8,

$$\mathbb{P}_\lambda^N\left(\bigcup_{k=1}^m E_k\right) = \sum_{k=1}^m \mathbb{P}_\lambda^N(E_k) = \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k},$$

where  $E_k$ 's are defined with respect to  $\mathfrak{X}^N$ . Since

$$\mathfrak{D}_\lambda(\mathbf{x}, \mathbf{s}) \leq \mathfrak{D}_\lambda(\mathbf{x}, \mathbf{0}) = \mathfrak{D}(\mathbf{x}),$$

we have

$$(\{1 - \mathfrak{D} > \alpha\} \times [0, 1]^N) \subseteq \{1 - \mathfrak{D}_\lambda > \alpha\} \subseteq \bigcup_{k=1}^m E_k.$$

Therefore,

$$\mathbb{P}^N\{\mathfrak{D} < 1 - \alpha\} \leq \mathbb{P}_\lambda^N(\cup_{k=1}^m E_k) = \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}.$$

**THEOREM A.10.** *Let  $N \geq m$  and  $\alpha \in [0, 1]$ . Suppose  $\mathfrak{D}$  satisfies Assumption A.3 and  $(\mathfrak{X}, \mathcal{A} \otimes \mathcal{L}, \mathbb{P}_\lambda)$  satisfies Assumptions A.4 to A.7; then*

$$(A.6) \quad \mathbb{P}^N\{\mathbf{x} \in X^N : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}.$$

*Example A.11.* One natural example of chains on probability space is the one where the orders are preserved by the inverse of random variables. Suppose there are  $m$  random variables  $\tau_k$ ,  $k \in [m]$ , on the probability space  $(X, \mathcal{A}, \mathbb{P})$ , such that for each  $k \in [m]$ ,

$$\tau_k(x) \leq \tau_k(y) \implies x \leq_k y, \quad x, y \in X;$$

certainly, it does not preclude the case where  $\tau_k(x) < \tau_k(y)$  but  $x =_k y$ . For each  $x \in X$ , the equivalence class  $\langle x \rangle_k = \{y \in X : y =_k x\}$  is the preimage of some intervals (open, closed, or half-open) under  $\tau_k$ ; indeed, let  $y =_k x$  and  $\tau_k(y) \leq \tau_k(x)$ . Then  $z =_k x$  if  $\tau_k(z) \in [\tau_k(y), \tau_k(x)]$ , and thus  $\langle x \rangle_k$  is the preimage of  $\text{conv } \tau_k(\langle x \rangle_k)$ , i.e., the convex hull of  $\tau_k(\langle x \rangle_k)$ ; this also implies there are only countably many  $\langle x \rangle_k$

such that  $\tau_k(\langle x \rangle_k)$  is not a singleton. Similarly, the set  $\{y \in X : y \leq_k x\}$  is the preimage of some half-infinite intervals under  $\tau_k$ . For  $s \in \mathbb{R}$ , we denote  $\langle s \rangle_k$  to be  $\{x \in X : s \in \text{conv } \tau_k(\langle x \rangle_k)\}$ ; when  $\langle s \rangle_k$  is nonempty, it is an equivalence class  $\langle x \rangle_k$  for some  $x \in X$ . We first show that Assumption A.3 is satisfied, i.e.,  $\mathfrak{D}$  is a random variable.

Assumption A.3: We need to show that  $\{\mathfrak{D} \leq \alpha\}$  is measurable for every  $\alpha \in [0, 1)$ . For each  $k \in [m]$  and  $p \in \mathbb{Q}$ , we define

$$B_{k,p} = \begin{cases} \{y \in X : y \geq_k x\} & \text{if } p \in \text{conv } \tau_k(\langle x \rangle_k), \\ \tau_k^{-1}(p, \infty) & \text{if no such } x \text{ exists} \end{cases}$$

and consider the collection of sets

$$\mathfrak{B}_k = \{B_{k,p} : p \in \mathbb{Q}\} \cup \{\{y \in X : y \geq_k x\} : \mathbb{P}(\langle x \rangle_k) > 0, x \in X\} \\ \cup \{\{y \in X : y >_k x\} : \mathbb{P}(\langle x \rangle_k) > 0, x \in X\};$$

each  $\mathfrak{B}_k$  is countable since there can be at most countably many  $\langle x \rangle_k$  such that  $\mathbb{P}(\langle x \rangle_k) > 0$ . Let

$$\mathfrak{E}_n = \bigcup_{B_k \in \mathfrak{B}_k : \mathbb{P}(\cap_{k=1}^m B_k) \leq \alpha + \frac{1}{n}} \left( \bigcap_{k=1}^m B_k \right)^N.$$

We claim that  $\{\mathfrak{D} \leq \alpha\} = \cap_{n=1}^{\infty} \mathfrak{E}_n$ . To see that  $\{\mathfrak{D} \leq \alpha\} \supseteq \cap_{n=1}^{\infty} \mathfrak{E}_n$ , let  $\mathbf{x} \in \mathfrak{E}_n$ ; then

$$\mathfrak{D}(\mathbf{x}) = \mathbb{P} \left( \bigcap_{k=1}^m \{x \in X : h_k \leq_k x\} \right) \leq \mathbb{P} \left( \bigcap_{k=1}^m B_k \right) \leq \alpha + \frac{1}{n},$$

and thus  $\mathfrak{D}(\mathbf{x}) \leq \alpha$  for each  $\mathbf{x} \in \cap_{n=1}^{\infty} \mathfrak{E}_n$ . On the other hand, let  $\mathbf{x} \in X^N$ ,  $\mathfrak{D}(\mathbf{x}) \leq \alpha$ ; for each  $k \in [m]$ , if  $\mathbb{P}(\langle h_k \rangle_k) > 0$  or there exists rational  $p_k \in \text{conv } \tau_k(\langle h_k \rangle_k)$ , then we set  $B_k = \{y \in X : y \geq_k h_k\}$ . Otherwise, we can find a rational sequence  $p_i \uparrow \tau_k(h_k)$ , and  $\mathbb{P}(\tau_k^{-1}[p_i, \tau_k(h_k)]) \downarrow \mathbb{P}(\tau_k^{-1}\{\tau_k(h_k)\}) = 0$ , and hence there exists rational  $p_0 < \tau_k(h_k)$  such that  $\mathbb{P}(\tau_k^{-1}[p_0, \tau_k(h_k)]) \leq \frac{1}{mn}$ ; if there exists  $y_k \in X$  such that  $\mathbb{P}(\langle y_k \rangle_k) > 0$  and  $p_0 \in \text{conv } \tau_k(\langle y_k \rangle_k)$ , then we set  $B_k = \{y \in X : y >_k y_k\}$ ; otherwise, we set  $B_k = B_{k,p_0}$ ; either way, we have  $\mathbf{x} \in (\cap_{k=1}^m B_k)^N$ , and

$$\mathfrak{D}(\mathbf{x}) = \mathbb{P} \left( \bigcap_{k=1}^m \{y \in X : y \geq_k h_k\} \right) \\ = \mathbb{P} \left( \bigcap_{k=1}^m B_k \setminus (B_k \setminus \{y \in X : y \geq_k h_k\}) \right) \\ \geq \mathbb{P} \left( \bigcap_{k=1}^m B_k \right) - \sum_{k=1}^m \frac{1}{mn}.$$

Hence,  $\mathbf{x} \in \mathfrak{E}_n$  for each  $n \in \mathbb{N}$ . Since each  $\mathfrak{E}_n$  is measurable,  $\{\mathfrak{D} \leq \alpha\}$  is measurable as desired.

Now consider the product space  $(\mathfrak{X}, \mathcal{A} \otimes \mathcal{L}, \mathbb{P}_\lambda)$ . Let  $\leq_\ell$  be the lexicographical order on  $\mathbb{R} \times [0, 1]$ ; note that  $(\tau_k(x), s) \leq_\ell (\tau_k(y), t)$  does *not* imply  $(x, s) \leq_k (y, t)$  in general, though it is still valid that

$$(\tau_k(x), s) \leq_\ell (\tau_k(y), t) \implies x \leq_k y, \quad (x, s), (y, t) \in \mathfrak{X}.$$

We show below that Assumptions A.4 to A.7 are satisfied automatically.

Assumption A.5: Suppose the set  $\{(x, s) \in \mathfrak{X} : S_k(x, s) = \beta\}$  is nonempty; let

$$(v_1, v_2) = \inf_{S_k(x, s) = \beta} (\tau_k(x), s), \quad (w_1, w_2) = \sup_{S_k(x, s) = \beta} (\tau_k(x), s);$$

here, the sup and the inf are taken with respect to the lexicographical order. In the case where the sup and the inf are attained,

$$\{(x, s) : S_k(x, s) = \beta\} = (\tau_k^{-1}[v_1, w_1] \times [0, 1]) \setminus ((\langle v_1 \rangle_k \times [0, v_2]) \cup (\langle w_1 \rangle_k \times (w_2, 1])),$$

and thus the set is measurable. The other cases are similar.

Assumption A.6: Suppose  $Y \subseteq \mathfrak{X}$  is a measurable set with  $\mathbb{P}_\lambda(Y) > 0$ . Consider

$$\hat{Y}_k = \{(\tau_k(x), s) : (x, s) \in Y\} \subseteq \mathbb{R} \times [0, 1];$$

then there exist two sequences  $\hat{\mathbf{v}}^1 \geq_\ell \hat{\mathbf{v}}^2 \geq_\ell \cdots$  and  $\hat{\mathbf{w}}^1 \leq_\ell \hat{\mathbf{w}}^2 \leq_\ell \cdots$  in  $\hat{Y}_k$ , such that

$$\hat{Y}_k = \bigcup_{n=1}^{\infty} \{\hat{\mathbf{y}} \in \hat{Y}_k : \hat{\mathbf{v}}^n \leq_\ell \hat{\mathbf{y}} \leq_\ell \hat{\mathbf{w}}^n\}.$$

For each  $n$ , we pick a  $\mathbf{v}^n \in Y$  such that  $\mathbf{v}^n \leq_k \mathbf{v}^{n-1}$ ,  $v_1^n = \tau_k^{-1}(\hat{v}_1^n)$ , and

$$v_2^n \leq \inf_{(x, s) \in Y : x =_k v_1^n} s + \frac{1}{n};$$

similarly, we pick a  $\mathbf{w}^n \in Y$  such that  $\mathbf{w}^n \geq_k \mathbf{w}^{n-1}$ ,  $w_1^n = \tau_k^{-1}(\hat{w}_1^n)$ , and

$$w_2^n \geq \sup_{(x, s) \in Y : x =_k w_1^n} s - \frac{1}{n};$$

then  $\mathbf{v}^1 \geq_k \mathbf{v}^2 \geq_k \cdots$  and  $\mathbf{w}^1 \leq_k \mathbf{w}^2 \leq_k \cdots$ , and

$$Y_o = Y \setminus \left( \bigcup_{n=1}^{\infty} \{(x, s) \in Y : \mathbf{v}^n \leq_k (x, s) \leq_k \mathbf{w}^n\} \right)$$

is a  $\mathbb{P}_\lambda$ -null set; indeed, let  $(x, s) \in Y$ . Then  $\hat{\mathbf{v}}^{n_0} \leq_\ell (\tau_k(x), s) \leq_\ell \hat{\mathbf{w}}^{n_0}$  for some  $n_0 \in \mathbb{N}$ , which implies  $v_1^n \leq_k x \leq_k w_1^n$  for any  $n \geq n_0$ ; note that  $(x, s)$  can only be in  $Y_o$  when either  $x =_k w_1^n$  for all  $n \geq n_0$  and  $w_2^n \uparrow s$ , or  $x =_k v_1^n$  for all  $n \geq n_0$  and  $v_2^n \downarrow s$ . Thus  $Y_o \subseteq X \times \{s_1, s_2\}$  for some  $s_1, s_2 \in [0, 1]$ , which has measure 0. Therefore, by continuity from below,

$$\sup_{\mathbf{v}, \mathbf{w} \in Y} \mathbb{P}_\lambda\{(x, s) \in Y : \mathbf{v} \leq_k (x, s) \leq_k \mathbf{w}\} = \mathbb{P}_\lambda(Y) > 0.$$

Assumption A.7: Since

$$\mathfrak{N}_k = \{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : (x_i, s_i) =_j (x_{i'}, s_{i'}) \text{ for some } j \in [k], i \neq i' \in [N]\}$$

is a  $\mathbb{P}_\lambda^N$ -null set, we only need to show that

$$\mathfrak{S}_k = \{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : (x_j, s_j) <_j (x_{j'}, s_{j'}) \ \forall j < j', j \in [k], j' \in [m]\}$$

is measurable. It suffices to show that  $\{(\mathbf{x}, \mathbf{s}) \in \mathfrak{X}^N : (x_j, s_j) <_j (x_{j'}, s_{j'})\}$  is measurable for each  $j < j', j \in [k], j' \in [m]$ . By definition,  $(x_j, s_j) <_j (x_{j'}, s_{j'})$  if and only if either  $x_j <_j x_{j'}$ , or  $x_j =_j x_{j'}$  and  $s_j < s_{j'}$ . Since  $\{x_j <_j x_{j'}\}$  is the permutation of the set

$$\bigcup_{q \in \mathbb{Q}} (\tau_j^{-1}(-\infty, q] \times \tau_j^{-1}(q, \infty) \setminus \langle q \rangle_j \times \langle q \rangle_j) \times X^{N-2} \times [0, 1]^N,$$

$\{x_j <_j x_{j'}\}$  is measurable. Also,  $\{x_j =_j x_{j'}\} = (\{x_j <_j x_{j'}\} \cup \{x_j >_j x_{j'}\})^c$  is measurable, and hence

$$\{(x_j, s_j) <_j (x_{j'}, s_{j'})\} = \{x_j <_j x_{j'}\} \cup (\{x_j =_j x_{j'}\} \cap \{s_j < s_{j'}\})$$

is measurable.

Assumption A.4: We need to show that  $\{\delta_k \geq \alpha\}$  is measurable for every  $\alpha \in (0, 1]$  and  $k \in [m]$ . If we can show that  $\{\delta_k \geq \alpha\} \cap \mathfrak{S}_k$  is measurable, then  $\{\delta_k \geq \alpha\}$  is measurable by permutation. For each  $j \in [k]$  and  $p, q \in \mathbb{Q}$ , we define

$$B_{j,p,q} = \begin{cases} \{(y, t) \in \mathfrak{X} : (y, t) \geq_j (x, q)\} & \text{if } p \in \text{conv } \tau_j(\langle x \rangle_j), \\ \tau_j^{-1}(p, \infty) \times [0, 1] & \text{if no such } x \text{ exists} \end{cases}$$

and consider the collection of sets

$$\mathfrak{B}_j = \{B_{j,p,q} : p, q \in \mathbb{Q}\} \cup \{(y, t) \in \mathfrak{X} : (y, t) \geq_j (x, q)\} : \mathbb{P}(\langle x \rangle_j) > 0, x \in X, q \in \mathbb{Q}\};$$

each  $\mathfrak{B}_j$  is countable since there can be at most countably many  $\langle x \rangle_j$  such that  $\mathbb{P}(\langle x \rangle_j) > 0$ . Let

$$\mathfrak{A}_n = \left[ \bigcup_{B_j \in \mathfrak{B}_j : \mathbb{P}_\lambda(\bigcup_{j=1}^k B_j^c) \geq \alpha - \frac{1}{n}} \left( \prod_{j=1}^k B_j \right) \times \mathfrak{X}^{N-k} \right] \cap \mathfrak{S}_k.$$

We claim that  $\{\delta_k \geq \alpha\} \cap \mathfrak{S}_k = \bigcap_{n=1}^\infty \mathfrak{A}_n$ . To see that  $\{\delta_k \geq \alpha\} \cap \mathfrak{S}_k \supseteq \bigcap_{n=1}^\infty \mathfrak{A}_n$ , let  $(\mathbf{x}, \mathbf{s}) \in \mathfrak{A}_n$ ; since  $(\mathbf{x}, \mathbf{s}) \in \mathfrak{S}_k$ , we have

$$\delta_k(\mathbf{x}, \mathbf{s}) = \mathbb{P}_\lambda \left( \bigcup_{j=1}^k \{(y, t) \in \mathfrak{X} : (y, t) <_j (x_j, s_j)\} \right) \geq \mathbb{P}_\lambda \left( \bigcup_{j=1}^k B_j^c \right) \geq \alpha - \frac{1}{n},$$

and thus  $\delta_k(\mathbf{x}, \mathbf{s}) \geq \alpha$  for each  $(\mathbf{x}, \mathbf{s}) \in \bigcap_{n=1}^\infty \mathfrak{A}_n$ . On the other hand, let  $(\mathbf{x}, \mathbf{s}) \in \mathfrak{S}_k$  such that  $\delta_k(\mathbf{x}, \mathbf{s}) \geq \alpha$ ; for each  $j \in [k]$ , if  $\mathbb{P}(\langle x_j \rangle_j) > 0$  or there exists rational  $p_j \in \text{conv } \tau_j(\langle x_j \rangle_j)$ , then we pick  $q_j \in \mathbb{Q}$ ,  $s_j \geq q_j \geq s_j - \frac{1}{kn}$ , and we set  $B_j = \{(y, t) \geq_j (x_j, q_j)\}$ ; otherwise, we can find a rational sequence  $p_i \uparrow \tau_j(x_j)$ , such that  $\mathbb{P}(\tau_j^{-1}[p_i, \tau_j(x_j)]) \downarrow \mathbb{P}(\tau_j^{-1}\{\tau_j(x_j)\}) = 0$ , and hence there exists rational  $p_0 < \tau_j(x_j)$  such that  $\mathbb{P}(\tau_j^{-1}[p_0, \tau_j(x_j)]) \leq \frac{1}{kn}$ , and we set  $B_j = B_{j,p_0,1}$ ; either way, we have  $(\mathbf{x}, \mathbf{s}) \in (\prod_{j=1}^k B_j) \times \mathfrak{X}^{N-k}$ , and

$$\begin{aligned} \delta_k(\mathbf{x}, \mathbf{s}) &= \mathbb{P}_\lambda \left( \bigcup_{j=1}^k \{(y, t) \in \mathfrak{X} : (y, t) <_j (x_j, s_j)\} \right) \\ &= \mathbb{P}_\lambda \left( \bigcup_{j=1}^k B_j^c \cup (B_j \setminus \{(y, t) : (y, t) \geq_j (x_j, s_j)\}) \right) \\ &\leq \mathbb{P}_\lambda \left( \bigcup_{j=1}^k B_j^c \right) + \sum_{j=1}^k \frac{1}{kn}. \end{aligned}$$



Hence,  $(\mathbf{x}, \mathbf{s}) \in \mathfrak{A}_n$  for each  $n \in \mathbb{N}$ . Now, since each  $\mathfrak{A}_n$  is measurable,  $\{\delta_k \geq \alpha\} \cap \mathfrak{S}_k$  is measurable as desired.

**COROLLARY A.12.** *Let  $N \geq m$  and  $\alpha \in [0, 1]$ . Suppose there exist random variables  $\tau_k$ ,  $k \in [m]$ , such that*

$$\tau_k(x) \leq \tau_k(y) \implies x \leq_k y, \quad x, y \in X;$$

then

$$\mathbb{P}^N\{\mathbf{x} \in X^N : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

We next give examples where the bound (A.6) is tight.

*Example A.13.* Consider the probability space  $([0, 1], \mathcal{L}, \lambda)$ . Letting  $m \in \mathbb{N}$ , we define  $m$  binary relations  $\leq_k$  on  $[0, 1]$ : for each  $k \in [m]$ , we say  $s <_k t$  if the following hold:

- $s < t$  and  $s, t \in [\frac{k-1}{m}, \frac{k}{m})$ ;
- $s < t$  and  $s, t \in [0, 1] \setminus [\frac{k-1}{m}, \frac{k}{m})$ ;
- $s \in [\frac{k-1}{m}, \frac{k}{m})$  and  $t \in [0, 1] \setminus [\frac{k-1}{m}, \frac{k}{m})$ .

In other words,  $\leq_k$  is derived from the standard ordering  $\leq$  by making the interval  $[\frac{k-1}{m}, \frac{k}{m})$  smaller. Letting  $N \geq m$  and  $\alpha \in [0, \frac{1}{m}]$ , we claim that

$$\lambda^N\{\mathbf{x} \in [0, 1]^N : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\} = \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

The idea is similar to the proof of Lemma A.8. We define  $\Delta_0 = 0$ , and for each  $k \in [m]$ , we define

$$\Delta_k(\mathbf{x}) = \lambda\{x \in [0, 1] : x <_j h_j(\mathbf{x}) \text{ for some } j \in [k]\},$$

where  $h_j(\mathbf{x})$  is the minimum of  $x_i$ 's under the relation  $\leq_j$ . Consider the disjoint events

$$E_k = \{\Delta_{k-1} \leq \alpha, \Delta_k > \alpha\}, \quad k \in [m];$$

then  $\{\mathfrak{D} < 1 - \alpha\} = \cup_{k=1}^m E_k$ . For almost every  $\mathbf{x}$  such that  $\Delta_{k-1}(\mathbf{x}) \leq \alpha \leq \frac{1}{m}$ , we have  $h_j(\mathbf{x}) \in [\frac{j-1}{m}, \frac{j}{m})$  for all  $j \in [k-1]$ . Hence,  $E_k$  has the same measure as the set

$$(A.7) \quad \left\{ \mathbf{x} \in [0, 1]^N : \begin{array}{l} h_j \in [\frac{j-1}{m}, \frac{j}{m}) \quad \forall j \in [k-1], \sum_{j=1}^{k-1} (h_j - \frac{j-1}{m}) \leq \alpha, \\ h_k \notin [\frac{k-1}{m}, \frac{k-1}{m} + \alpha - \sum_{j=1}^{k-1} (h_j - \frac{j-1}{m}) \end{array} \right\};$$

the measure of (A.7) is  $\binom{N}{k-1} \alpha^{k-1} (1-\alpha)^{N-k+1}$ , which can be calculated by considering the permutation  $x_j = h_j$ ,  $j \in [k-1]$ , and conditioning on the values of  $x_j$ .

So far, we have assumed  $N \geq m$ . As illustrated by the following example, there may be no formula that bounds the probability (A.6) uniformly when  $N < m$ . Let  $N = 1$  and  $m = 2$ , and consider again  $([0, 1], \mathcal{L}, \lambda)$ ; the relation  $\leq_1$  is  $\leq$ , and  $\leq_2$  is the converse of  $\leq$  (i.e.,  $s \leq_2 t \iff t \leq s$ ). For any  $\alpha \in [0, 1]$ , we have

$$\lambda\{\mathbf{x} \in X : \mathfrak{D}(\mathbf{x}) < 1 - \alpha\} = 1.$$

**A.3. Applications.** A collection of sets  $\{U^\omega\}_{\omega \in I}$  is a chain if for any  $\omega_1, \omega_2 \in I$ , either  $U^{\omega_1} \subseteq U^{\omega_2}$  or  $U^{\omega_1} \supseteq U^{\omega_2}$ . Let  $\xi$  be a random vector supported on the set  $\Xi$ . Suppose we are given  $m$  chains  $\{U_k^\xi\}_{\xi \in \Xi}$ ,  $k \in [m]$ ; then each chain  $\{U_k^\xi\}_{\xi \in \Xi}$  induces a binary relation  $\leq_k$  on  $\Xi$ , i.e.,

$$\omega_1 \leq_k \omega_2 \iff U_k^{\omega_1} \subseteq U_k^{\omega_2}, \quad \omega_1, \omega_2 \in \Xi.$$

Note that  $\Xi$  is a chain with respect to the partially ordered set  $(\Xi, \leq_k)$  for each  $k \in [m]$ . Let  $\xi^{[N]} = (\xi^1, \dots, \xi^N)$  be the i.i.d. (random) sample of  $\xi$  of size  $N$ ; then

$$\mathfrak{D}(\xi^{[N]}) = \mathbb{P} \left\{ \xi \in \Xi : \bigcap_{i=1}^N U_k^{\xi^i} \subseteq U_k^\xi \quad \forall k \in [m] \right\}.$$

COROLLARY A.14. *Under the assumptions of Theorem A.10,*

$$\mathbb{P}^N \left\{ \mathfrak{D}(\xi^{[N]}) < 1 - \alpha \right\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

Oftentimes,  $\{U_k^\xi\}$  are sublevel sets, i.e.,  $U_k^\xi = c_k^{-1}(-\infty, \ell_k(\xi)]$  for some function  $c_k$  and random variable  $\ell_k$  supported on  $\Xi$ . In this case,

$$\ell_k(\omega_1) \leq \ell_k(\omega_2) \implies U_k^{\omega_1} \subseteq U_k^{\omega_2} \implies \omega_1 \leq_k \omega_2, \quad \omega_1, \omega_2 \in \Xi.$$

We can also define the relation

$$\omega_1 \leq_k \omega_2 \iff \ell_k(\omega_1) \leq \ell_k(\omega_2), \quad \omega_1, \omega_2 \in \Xi,$$

and

$$\mathfrak{D}_r(\xi^{[N]}) = \mathbb{P} \left\{ \xi \in \Xi : \min_{i \in [N]} \ell_k(\xi^i) \leq \ell_k(\xi) \quad \forall k \in [m] \right\}.$$

Note that

$$\min_{i \in [N]} \ell_k(\xi^i) \leq \ell_k(\xi) \implies \bigcap_{i=1}^N U_k^{\xi^i} \subseteq U_k^\xi,$$

and hence  $\mathfrak{D}(\xi^{[N]}) \geq \mathfrak{D}_r(\xi^{[N]})$ . Since the orders are preserved by the inverse of the random variables  $\ell_k$ , we can apply Corollary A.12 instead of Theorem A.10.

COROLLARY A.15. *If  $U_k^\xi$ 's are sublevel sets, then*

$$\mathbb{P}^N \left\{ \mathfrak{D}(\xi^{[N]}) < 1 - \alpha \right\} \leq \mathbb{P}^N \left\{ \mathfrak{D}_r(\xi^{[N]}) < 1 - \alpha \right\} \leq \sum_{k=0}^{m-1} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k}.$$

**Acknowledgments.** The author is indebted to Alexander Shapiro and the anonymous referees for constructive comments which helped to improve the presentation of the manuscript.

#### REFERENCES

- [1] G. BARBAROSOĞLU AND Y. ARDA, *A two-stage stochastic programming framework for transportation planning in disaster response*, J. Oper. Res. Soc., 55 (2004), pp. 43–53.
- [2] G. CALAFIORE AND M. CAMPI, *Uncertain convex programs: Randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25–46.

- [3] M. C. CAMPI AND S. GARATTI, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM J. Optim., 19 (2008), pp. 1211–1230, <https://doi.org/10.1137/07069821X>.
- [4] M. C. CAMPI AND S. GARATTI, *Introduction to the Scenario Approach*, SIAM, Philadelphia, 2018, <https://doi.org/10.1137/1.9781611975444>.
- [5] X. CHEN, A. SHAPIRO, AND H. SUN, *Convergence analysis of sample average approximation of two-stage stochastic generalized equations*, SIAM J. Optim., 29 (2019), pp. 135–161, <https://doi.org/10.1137/17M1162822>.
- [6] J. DUPACOVA AND R. WETS, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Statist., 16 (1988), pp. 1517–1549.
- [7] G. FOLLAND, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., John Wiley & Sons, New York, 1999.
- [8] L. J. HALBEISEN, *Combinatorial Set Theory*, Springer-Verlag, New York, 2017.
- [9] G. H. HUANG AND D. P. LOUCKS, *An inexact two-stage stochastic programming model for water resources management under uncertainty*, Civ. Eng. Environ. Syst., 17 (2000), pp. 95–118.
- [10] A. J. KLEYWEGT, A. SHAPIRO, AND T. HOMEM-DE-MELLO, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2001), pp. 479–502, <https://doi.org/10.1137/S1052623499363220>.
- [11] J. LINDEROTH, A. SHAPIRO, AND S. WRIGHT, *The empirical behavior of sampling methods for stochastic programming*, Ann. Oper. Res., 142 (2006), pp. 215–241.
- [12] W. MAK, D. MORTON, AND K. WOOD, *Monte Carlo bounding techniques for determining solution quality in stochastic programs*, Oper. Res. Lett., 24 (1999), pp. 47–56.
- [13] C. NICULESCU AND L. PERSSON, *Convex Functions and Their Applications*, 2nd ed., CMS Books Math./Ouvrages Math. SMC, Springer, Cham, 2018.
- [14] G. C. PFLUG AND A. PICHLER, *Multistage Stochastic Optimization*, 2nd ed., Springer-Verlag, New York, 2014.
- [15] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Stochastic convex programming: Relatively complete recourse and induced feasibility*, SIAM J. Control Optim., 14 (1976), pp. 574–589, <https://doi.org/10.1137/0314038>.
- [16] A. SCHRIJVER, *Theory of Linear and Integer Programming*, John Wiley & Sons, Chichester, 1986.
- [17] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed., SIAM and MPS, Philadelphia, 2014, <https://doi.org/10.1137/1.9781611973433>.
- [18] A. SHAPIRO AND A. NEMIROVSKI, *On complexity of stochastic programming problems*, in Continuous Optimization, Springer, New York, 2005, pp. 111–146.
- [19] H. SUN AND H. XU, *A note on uniform exponential convergence of sample average approximation of random functions*, J. Math. Anal. Appl., 385 (2012), pp. 698–708.
- [20] R. VERSHYNIN, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, Cambridge, UK, 2018.
- [21] B. VERWEIJ, S. AHMED, A. KLEYWEGT, G. NEMHAUSER, AND A. SHAPIRO, *The sample average approximation method applied to stochastic routing problems: A computational study*, Comput. Optim. Appl., 24 (2003), pp. 289–333.
- [22] Q. WANG, Y. GUAN, AND J. WANG, *A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output*, IEEE Trans. Power Syst., 27 (2012), pp. 206–215.
- [23] H. XU, *Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming*, J. Math. Anal. Appl., 368 (2010), pp. 692–710.