

ON RATES OF CONVERGENCE FOR STOCHASTIC OPTIMIZATION PROBLEMS UNDER NON-I.I.D. SAMPLING

TITO HOMEM-DE-MELLO*

Abstract.

In this paper we discuss the issue of solving stochastic optimization problems by means of sample average approximations. Our focus is on rates of convergence of estimators of optimal solutions and optimal values with respect to the sample size. This is a well studied problem in case the samples are independent and identically distributed (i.e., when standard Monte Carlo is used); here, we study the case where that assumption is dropped. Broadly speaking, our results show that, under appropriate assumptions, the rates of convergence for *pointwise estimators* under a sampling scheme carry over to the optimization case, in the sense that convergence of approximating optimal solutions and optimal values to their true counterparts has the same rates as in pointwise estimation. We apply our results to two well-established sampling schemes, namely, Latin Hypercube Sampling (LHS) and randomized Quasi-Monte Carlo (QMC). The novelty of our work arises from the fact that, while there has been some work on the use of variance reduction techniques and QMC methods in stochastic optimization, none of the existing work — to the best of our knowledge — has provided a theoretical study on the effect of these techniques on rates of convergence for the optimization problem. We present numerical results for some two-stage stochastic programs from the literature to illustrate the discussed ideas.

Key words. Stochastic optimization, two-stage stochastic programming with recourse, sample average approximation, Monte Carlo simulation, Quasi-Monte Carlo methods, Latin Hypercube sampling, variance reduction techniques.

1. Introduction . In this paper we consider stochastic optimization problems of the form

$$\min_{x \in X} \{g(x) := \mathbb{E}[G(x, \boldsymbol{\xi})]\}, \quad (1.1)$$

where X is a subset of \mathbb{R}^n , $\boldsymbol{\xi}$ is a random vector in \mathbb{R}^s and $G : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}$ is a real valued measurable function. We refer to (1.1) as the “true” optimization problem. The class of problems falling into the framework of (1.1) is quite large, and includes two-stage stochastic programs as a particular subclass.

Oftentimes the expectation in (1.1) cannot be calculated exactly, particularly when G does not have a closed form. In those cases, approximations based on sampling are usually the alternative. One such approximation can be constructed as follows. Consider a family $\{\hat{g}_N(\cdot)\}$ of random approximations of the function $g(\cdot)$, each $\hat{g}_N(\cdot)$ being defined as

$$\hat{g}_N(x) := \frac{1}{N} \sum_{j=1}^N G(x, \boldsymbol{\xi}^j), \quad (1.2)$$

where $\{\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^N\}$ is a sample from the distribution of $\boldsymbol{\xi}$. When $\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^N$ — viewed as random variables — are independent and identically distributed (i.i.d.) the quantity $\hat{g}_N(x)$ is called a (*standard*) *Monte Carlo* estimator of $g(x)$.

Given the family of estimators $\{\hat{g}_N(\cdot)\}$ defined in (1.2), one can construct the corresponding approximating program

$$\min_{x \in X} \hat{g}_N(x). \quad (1.3)$$

* Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, U.S.A. (tito@northwestern.edu).

Let \hat{x}_N and $\hat{\nu}_N$ denote respectively an optimal solution and the optimal value of (1.3). Then, \hat{x}_N and $\hat{\nu}_N$ provide approximations respectively to an optimal solution x^* and the optimal value ν^* of the true problem (1.1). Note that the optimization in (1.3) is performed for a fixed sample; for that reason, this is called an *external sampling* approach. When $\hat{g}_N(\cdot)$ is a standard Monte Carlo estimator of $g(\cdot)$, such an approach is found in the literature under the names of sample average approximation method, stochastic counterpart, and sample-path optimization, among others.

The external sampling approach with standard Monte Carlo has been implemented in various settings, see for instance [14, 23, 43]. One advantage of that approach lies in its nice convergence properties; for example, it is possible to show that, when x^* is the unique optimal solution, $\hat{x}_N \rightarrow x^*$ and $\hat{\nu}_N \rightarrow \nu^*$ under fairly general assumptions (see, e.g., [10, 22, 44, 48, 49]). Two properties have proven particularly useful in terms of establishing *rates of convergence*: the first establishes that, under proper conditions, $P(|g(\hat{x}_N) - g(x^*)| \leq \varepsilon)$ and $P(\|\hat{x}_N - x^*\| \leq \varepsilon)$ converge to one *exponentially fast* in the sample size N for any fixed $\varepsilon > 0$ (see [6, 21]). Under some further conditions one can say more, namely, that $P(\hat{x}_N = x^*)$ converges to one exponentially fast in the sample size N [53]. Exponential rates of convergence have interesting consequences in terms of complexity of the underlying problems; see [51] for a discussion.

Another useful property establishes that the sequence of optimal values $\{\hat{\nu}_N\}$ satisfies a certain kind of *Central Limit Theorem* (CLT). More specifically, one has

$$N^{1/2}(\hat{\nu}_N - \nu^*) \xrightarrow{d} \text{Normal}(0, \sigma^*),$$

where “ \xrightarrow{d} ” denotes convergence in distribution and $\sigma^* := \text{Var}[G(x^*)]$ [48]. An immediate conclusion from the above result is that the rate of convergence of optimal values of (1.3) is of order $N^{-1/2}$. A compilation of these and other related results can be found in [50].

It is no surprise that the sequence of approximating optimal values converges at rate $N^{-1/2}$. Indeed, consider the estimator \hat{g}_N defined in (1.2), and fix $x \in X$. Under mild conditions, it follows from the Central Limit Theorem that $\sqrt{N}[\hat{g}_N(x) - g(x)]/\sigma(x)$ converges in distribution to the standard Normal, where $\sigma^2(x)$ is the variance of $G(x)$. This implies that the error $\hat{g}_N(x) - g(x)$ converges to zero at the rate $N^{-1/2}$. That is, even the pointwise estimators converge at rate $N^{-1/2}$. In many practical cases, the value of N necessary to obtain a reasonably small error under this scheme becomes prohibitively large, especially if evaluation of $G(x, \xi)$ for a given ξ is computationally expensive. This motivates the use of *variance reduction techniques* that can yield estimators with smaller variance than the ones obtained with standard sampling. Consequently, the same error can be obtained with less computational effort, which is a crucial step for the use of sampling-based methods in large-scale problems.

Several variance reduction techniques have been developed in the Simulation and Statistics literature, notably importance sampling, control variates, stratified sampling, and others (see, e.g., [11, 25]). However, incorporation of these techniques into a stochastic optimization algorithm is still at an early stage. Existing work [1, 7, 16, 19, 26, 52] already shows that significant benefits can be gained by implementing some of these methods, but these papers only provide *empirical* evidence of the gain.

Another approach to obtain better pointwise estimators is to choose the sample points in an appropriate manner. Such is the case of *Quasi-Monte Carlo methods*

(QMC); see [32] for a comprehensive discussion. This class of methods has been gaining popularity in the past few years, as it has been observed that these techniques can provide rates of convergence for pointwise estimators superior to the $N^{-1/2}$ obtained with standard Monte Carlo. However, because estimating the actual error of a QMC estimator relative to the quantity being estimated can be difficult, some procedures to randomize a QMC sequence have been proposed in the literature. We provide a brief review of the basic ideas of QMC methods in Section 3.2.

A few papers study the optimization problem $\min_{x \in X} \hat{g}_N(x)$ under QMC: In [20], empirical results are provided for the use of Hammersley sequences (one form of QMC) in stochastic optimization problems. In [39] and [40], the authors use the fact that the empirical measure defined by a QMC sequence converges weakly to the uniform distribution to show that, under mild assumptions, the estimator function \hat{g}_N constructed with QMC points *epiconverges* to the true function g , which guarantees convergence of optimal values and optimal solutions under appropriate further conditions; in [24] those results are applied to the case where the QMC sequence is randomized with the so-called Cranley-Patterson procedure. The numerical results in those papers also suggest considerable gains in terms of rates of convergence when using QMC methods. A different type of point sequences is studied in [41] whereby the sampling points are chosen in a way to minimize the Wasserstein distance between the original distribution and the empirical distribution generated by the points. A related approach is used in [15], which deals with the Fortet-Mourier metrics instead. Again, the numerical results presented in [42] suggest considerable advantage of these techniques over standard Monte Carlo.

The above discussion shows that, while there has been some important work on the use of variance reduction techniques and QMC methods in stochastic optimization, none of these papers has provided a theoretical study on the effect of these techniques on *rates of convergence*. The reason is that, without the i.i.d. assumption, many of the classical results in probability theory cannot be applied. One exception is the work in [6], which provides results on exponential rate of convergence of optimal solutions even without the i.i.d. assumption. However, that paper does not focus on any particular sampling technique; rather, they assume that certain conditions that allow for the application of the Gartner-Ellis Theorem in large deviations theory (see, e.g., [8]) are satisfied.

Another potential way to study convergence rates in general settings (i.e., without the i.i.d. assumption) is by means of *stability* theory. Broadly speaking, stability theory in the context of stochastic optimization quantifies how much the optimal value and the optimal solutions of the problem change when the underlying probability measures are perturbed. For example, by writing the optimal values of (1.1) and (1.3) respectively as $\nu(F)$ and $\nu(F_N)$ — where F is the distribution of ξ and F_N is the empirical distribution defined by a sample — it is possible to show that, under certain assumptions, $|\nu(F) - \nu(F_N)|$ is bounded by an appropriately defined distance between F and F_N . In the particular case of i.i.d. sampling one can write the latter distance in terms of the sample size N , which leads to a different way to view the $N^{-1/2}$ rate obtained via the Central Limit Theorem. We refer to [45] for a thorough exposition of stability results in stochastic programming.

In this paper we propose a study of rates of convergence for optimal solutions and optimal values of the approximating problem (1.3) *without imposing that the sample be independent or identically distributed*. More specifically, we show that, under certain conditions, if the proposed sampling scheme yields exponential rate

of convergence for pointwise estimators, then the convergence of optimal solutions will also have an exponential rate. Moreover, in case of discrete or piecewise linear problems, if the proposed sampling scheme yields a CLT for pointwise estimators then the convergence of optimal values will obey the CLT as well. Unless stated otherwise the setting is fairly general — i.e., the decision space can be continuous or discrete, and the distributions of the underlying random variables can be continuous or discrete, although some the results will not be valid in some of these cases.

We illustrate the ideas for the particular cases of Latin Hypercube Sampling (LHS) and a specific variation of randomized QMC called scrambled (t, m, s) -nets. We show that, for a particular class of functions, the exponential feature of the rate of convergence is preserved under LHS for pointwise estimators and therefore for estimators of optimal solutions. We also use CLT-type results available for LHS and randomized QMC to illustrate the convergence results for estimators of optimal values. In particular, we show that, under LHS, the estimators \hat{v}_N of optimal values converge at a rate of order $N^{-1/2}$, the same as standard Monte Carlo; for QMC, under appropriate assumptions the sequence $\{\hat{v}_N\}$ converges at a rate of order $[(\log N)^{s-1}/N^3]^{1/2}$, which asymptotically is better than $N^{-1/2}$.

We then apply our results to two-stage stochastic linear programs, and discuss the validity of our assumptions in that context. Numerical results are presented for two problems from the literature to illustrate the ideas.

The remainder of the paper is organized as follows: in Section 2 we describe our main results for rates of convergence of estimators of optimal solutions and optimal values. In Section 3 we apply these results to Latin Hypercube Sampling and randomized Quasi-Monte Carlo. We illustrate the ideas for two-stage stochastic programs in Section 4 and present numerical results in Section 5. Concluding remarks are presented in Section 6.

2. Rates of convergence . We discuss separately the results on rates of convergence for optimal solutions and optimal values. Throughout this paper, S^* and S_N denote the set of optimal solutions of respectively (1.1) and (1.3). Before we study the two cases, we shall make a general assumption.

Assumption A1: For each $x \in X$, $\hat{g}_N(x) \rightarrow g(x)$ with probability one (denoted w.p.1).

Assumption A1 is very natural, as it requires the estimators to be *consistent*. In the i.i.d. case, this is just the standard Strong Law of Large Numbers, which holds if $\mathbb{E}[|\hat{g}_N(x)|] < \infty$ for each $x \in X$.

2.1. Convergence of estimators of optimal solutions. We start by making the following probabilistic assumption on the estimators $\{\hat{g}_N(x)\}$:

Assumption B1: For each $x \in X$, there exist a number $C_x > 0$ and a function $\gamma_x(\cdot)$ such that $\gamma_x(0) = 0$, $\gamma_x(z) > 0$ if $z > 0$, and

$$P(|\hat{g}_N(x) - g(x)| \geq \delta) \leq C_x e^{-N\gamma_x(\delta)} \quad \text{for all } N \geq 1 \text{ and all } \delta > 0. \quad (2.1)$$

That is, the probability that the deviation between $\hat{g}_N(x)$ and $g(x)$ is bigger than δ goes to zero exponentially fast with N . Notice that (2.1) implies that $\hat{g}_N(x)$ converges in probability to $g(x)$, which is also ensured by Assumption A1.

Instead of (2.1), we can impose the following weaker condition.

Assumption B1': For each $x \in X$, there exists a function $\gamma_x(\cdot)$ such that $\gamma_x(0) = 0$, $\gamma_x(z) > 0$ if $z > 0$, and

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(|\hat{g}_N(x) - g(x)| \geq \delta) \leq -\gamma_x(\delta) \quad \text{for all } \delta > 0. \quad (2.2)$$

Some of our results will be stated assuming B1 holds; alternatively, B1' can be used, though in such cases the corresponding result will be stated in asymptotic form as well.

We study now a sufficient condition for Assumption B1 to hold. The main concept behind it arises from the theory of *large deviations*, a well-studied field. For a thorough exposition of the theory, we refer to any of the classical texts in the area, e.g., [8]. We present here a result from [9].

PROPOSITION 2.1. *Consider the sample ξ^1, \dots, ξ^N used in (1.2), and define the extended real-valued function*

$$\phi_N(x, t) := \frac{1}{N} \log \mathbb{E} \left[e^{tN\hat{g}_N(x)} \right]. \quad (2.3)$$

Suppose that for each $x \in X$ there exists an extended real-valued function ϕ_x^ such that $\phi_N(x, \cdot) \leq \phi_x^*(\cdot)$ for all N , and assume that ϕ_x^* satisfies the following conditions: (i) $\phi_x^*(0) = 0$; (ii) $\phi_x^*(\cdot)$ is continuously differentiable and strictly convex on a neighborhood of zero; and (iii) $(\phi_x^*)'(0) = g(x)$. Then, Assumption B1 holds, with the functions $\gamma_x(\cdot)$ given by $\gamma_x(\delta) := \min\{I_x(g(x) + \delta), I_x(g(x) - \delta)\}$, where $I_x(z) = \sup_{t \in \mathbb{R}} \{tz - \phi_x^*(t)\}$, and the constants C_x being all equal to 2.*

A simple setting where the conditions of Proposition 2.1 are satisfied is when the functions $\phi_N(x, \cdot)$, $N = 1, 2, \dots$, are bounded by the log-moment generating function of some random variable W_x (i.e., $\phi_x^*(t) = \log \mathbb{E}[e^{tW_x}]$) such that $\mathbb{E}[W_x] = g(x)$. Clearly, condition (i) holds in that case. Moreover, if there exists an open neighborhood \mathcal{N} of zero such that $\phi_x^*(\cdot)$ is finite on \mathcal{N} , then it is well known that ϕ_x^* is infinitely differentiable on \mathcal{N} (see, e.g., p. 35 of [8]) and (iii) holds. In that case, Proposition 1 in [54] ensures that ϕ_x^* is strictly convex on \mathcal{N} .

Note that when the samples $\{\xi^i\}$ are i.i.d. we have

$$\begin{aligned} \phi_N(x, t) &= \frac{1}{N} \log(\mathbb{E}[e^{tN\hat{g}_N(x)}]) = \frac{1}{N} \log(\{\mathbb{E}[e^{tG(x, \xi)}]\}^N) = \log(\mathbb{E}[e^{tG(x, \xi)}]) \\ &= \log M_x(t), \end{aligned}$$

where $M_x(t) := \mathbb{E}[e^{tG(x, \xi)}]$ is the moment generating function of $G(x, \xi)$ evaluated at t . Hence, in that case we have $\phi_N(x, t) = \phi_x^*(t) := \log M_x(t)$ for all N , so the resulting function I_x in Proposition 2.1 is the rate function associated with $G(x, \xi)$. Inequality (2.1) then yields the well-known Chernoff upper bounds on the deviation probabilities. It is also well known (Cramér's Theorem) that in that case $\gamma_x(\delta)$ in Proposition 2.1 is an *asymptotically exact* rate, in the sense that (2.2) holds with equality.

One important consequence of the above developments is the following: Suppose that the function ϕ_x^* in Proposition 2.1 is dominated by the log-moment generating function of the random variable $G(x, \xi)$, i.e., $\phi_x^*(t) \leq \phi_x^{\text{MC}}(t) := \log \mathbb{E}[e^{tG(x, \xi)}]$. This immediately implies that the rate function I_x dominates the rate function associated with the random variable $G(x, \xi)$, which as seen earlier is the asymptotically exact

rate function obtained with i.i.d. (i.e., Monte Carlo) sampling. In other words, if one uses a sampling technique that yields functions $\phi_N(x, \cdot)$ for which one can find ϕ_x^* in Proposition 2.1 such that $\phi_x^*(\cdot) \leq \phi_x^{\text{MC}}(\cdot)$, then *the pointwise convergence rate for this sampling technique — in the sense of (2.1) — is at least as good as the rate obtained with standard Monte Carlo*. We will use this basic argument repeatedly in the course of this paper.

In the subsections below we will study the convergence of optimal solutions in two different settings — one when the function $G(\cdot, \boldsymbol{\xi})$ is Lipschitz, and the other when either $G(\cdot, \boldsymbol{\xi})$ is piecewise linear, or the feasible set X is finite.

2.1.1. The Lipschitz case. We make now an assumption on the integrand G viewed as a function of its first argument:

Assumption A2: The feasibility set X is compact and there exists an integrable function $L : \mathbb{R}^s \rightarrow \mathbb{R}$ such that, for almost every $\boldsymbol{\xi}$ and all $x, y \in X$,

$$|G(x, \boldsymbol{\xi}) - G(y, \boldsymbol{\xi})| \leq L(\boldsymbol{\xi})\|x - y\|. \quad (2.4)$$

Clearly, Assumption A2 ensures that the function $G(\cdot, \boldsymbol{\xi})$ is continuous for almost every $\boldsymbol{\xi}$. Moreover, it implies that $\hat{g}_N(\cdot)$ and $g(\cdot)$ are also Lipschitz continuous with constants respectively equal to $\hat{L}_N := N^{-1} \sum_{j=1}^N L(\boldsymbol{\xi}^j)$ and $\mathbb{E}[L(\boldsymbol{\xi})]$. From classical results in convex analysis (e.g., [18, Theorem IV.3.1.2]), we see that if (i) the feasibility set X is compact and contained in the relative interior of the domain of $G(\cdot, \boldsymbol{\xi})$ for almost every $\boldsymbol{\xi}$, and (ii) $G(\cdot, \boldsymbol{\xi})$ is *convex* for almost every $\boldsymbol{\xi}$, then the existence of $L(\boldsymbol{\xi})$ in (2.4) is assured, so in that case only integrability of $L(\boldsymbol{\xi})$ needs to be checked.

Recall that \hat{x}_N is an optimal solution of (1.3) and S^* is the set of optimal solutions of (1.1). Below, $\text{dist}(z, A)$ denotes the usual Euclidean distance function between a point z and a set A , i.e., $\text{dist}(z, A) := \inf_{y \in A} \|z - y\|$. The following result is known (see, e.g., [46, p.67-70]), but we state it here for reference:

PROPOSITION 2.2. *Suppose Assumptions A1 and A2 hold. Then,*

- (i) $\hat{g}_N(x) \rightarrow g(x)$ uniformly on X w.p.1;
- (ii) $\hat{\nu}_N \rightarrow \nu^*$ w.p.1;
- (iii) $\text{dist}(\hat{x}_N, S^*) \rightarrow 0$ w.p.1.

Theorem 2.3 below shows a probabilistic rate of convergence of optimal solutions. In preparation for that result, we state the following assumption, which is similar to Assumption B1 but applied to the random variable $L(\boldsymbol{\xi})$ in Assumption A2. Conditions under which such an assumption holds are similar to those given in Proposition 2.1.

Assumption B1_L: Let \hat{L}_N be the estimator of $\mathbb{E}[L(\boldsymbol{\xi})]$ defined as $\hat{L}_N := N^{-1} \sum_{j=1}^N L(\boldsymbol{\xi}^j)$, where as before $\{\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^N\}$ is a sample from the distribution of $\boldsymbol{\xi}$. There exist a number $C_L > 0$ and a function $\gamma_L(\cdot)$ such that $\gamma_L(0) = 0$, $\gamma_L(z) > 0$ if $z > 0$, and

$$P\left(|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| \geq \delta\right) \leq C_L e^{-N\gamma_L(\delta)} \quad \text{for all } N \geq 1 \text{ and all } \delta > 0. \quad (2.5)$$

THEOREM 2.3.

Consider problem (1.3), and suppose that Assumptions A2, B1 and B1_L hold. Then, given $\varepsilon > 0$, there exist constants $K > 0$ and $\alpha > 0$ such that

$$P(\text{dist}(\hat{x}_N, S^*) \geq \varepsilon) \leq K e^{-\alpha N} \quad \text{for all } N \geq 1.$$

The constants K and α depend on the families of estimators $\{\hat{g}_N(\cdot)\}$ and $\{\hat{L}_N\}$ only through respectively the constants C_x and C_L and the exponent functions $\gamma_x(\cdot)$ and $\gamma_L(\cdot)$ in (2.1) and (2.5). More specifically,

$$\alpha = \min \left(\min_{k=1, \dots, r} \{\gamma_{x_k}(\delta/3)\}, \gamma_L(\delta/3) \right)$$

$$K = (r+1) \max \left(\max_{k=1, \dots, r} \{C_{x_k}\}, C_L \right),$$

where $\delta > 0$, r is a finite number and x_1, \dots, x_r are points in X .

The proof of Theorem 2.3 will be based on the following lemma:

LEMMA 2.4. *Suppose that Assumptions A2, B1 and B1_L hold. Then, for any $\delta > 0$ there exist positive constants $A = A(\delta)$ and $\alpha = \alpha(\delta)$ such that*

$$P(|\hat{g}_N(x) - g(x)| \geq \delta) \leq Ae^{-\alpha N}, \quad \text{for all } x \in X \text{ and all } N \geq 1. \quad (2.6)$$

Moreover, there exists a positive constant K (also dependent on δ) such that

$$P(|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X) \geq 1 - Ke^{-\alpha N} \quad \text{for all } N \geq 1. \quad (2.7)$$

Proof. Let $\eta := \delta/(3\mathbb{E}[L(\boldsymbol{\xi})] + \delta)$, and denote by $B(x, \eta)$ the open ball with center x and radius η . Let $\mathcal{X} = \{x_1, \dots, x_r\}$ be a collection of points in X such that $X \subset \cup_{k=1}^r B(x_k, \eta)$. Notice that the existence of \mathcal{X} is ensured by the compactness of X .

Consider now an arbitrary point $x \in X$. By construction, there exists some $x_k \in \mathcal{X}$ such that $\|x - x_k\| < \eta$. Thus, from (2.4) we have that

$$|\hat{g}_N(x) - \hat{g}_N(x_k)| \leq \frac{1}{N} \sum_{j=1}^N |G(x, \boldsymbol{\xi}^j) - G(x_k, \boldsymbol{\xi}^j)| < \hat{L}_N \eta = \frac{\delta}{3} \frac{\hat{L}_N}{\mathbb{E}[L(\boldsymbol{\xi})] + \delta/3} \quad (2.8)$$

$$|g(x) - g(x_k)| \leq \mathbb{E}[|G(x, \boldsymbol{\xi}) - G(x_k, \boldsymbol{\xi})|] < \mathbb{E}[L(\boldsymbol{\xi})]\eta < \delta/3. \quad (2.9)$$

Moreover, by Assumptions B1 and B1_L we have that

$$P(|\hat{g}_N(x_k) - g(x_k)| \geq \delta/3) \leq C_{x_k} e^{-N\gamma_{x_k}(\delta/3)} \quad (2.10)$$

$$P(|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| \geq \delta/3) \leq C_L e^{-N\gamma_L(\delta/3)}. \quad (2.11)$$

Finally, since

$$|\hat{g}_N(x) - g(x)| \leq |\hat{g}_N(x) - \hat{g}_N(x_k)| + |\hat{g}_N(x_k) - g(x_k)| + |g(x) - g(x_k)|,$$

it follows that

$$\begin{aligned} \{|\hat{g}_N(x) - g(x)| < \delta\} &\supseteq \{|\hat{g}_N(x) - \hat{g}_N(x_k)| < \delta/3\} \cap \{|\hat{g}_N(x_k) - g(x_k)| < \delta/3\} \\ &\quad \cap \{|g(x_k) - g(x)| < \delta/3\} \\ &\supseteq \{|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| < \delta/3\} \cap \{|\hat{g}_N(x_k) - g(x_k)| < \delta/3\} \end{aligned} \quad (2.12)$$

and then from (2.10)-(2.11) we have that

$$\begin{aligned} P(|\hat{g}_N(x) - g(x)| \geq \delta) &\leq P(|\hat{g}_N(x_k) - g(x_k)| \geq \delta/3) + P(|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| \geq \delta/3) \\ &\leq C_{x_k} e^{-N\gamma_{x_k}(\delta/3)} + C_L e^{-N\gamma_L(\delta/3)}. \end{aligned}$$

By taking

$$\alpha := \min \left(\min_{k=1, \dots, r} \{\gamma_{x_k}(\delta/3)\}, \gamma_L(\delta/3) \right) \quad (2.13)$$

$$A := 2 \max \left(\max_{k=1, \dots, r} \{C_{x_k}\}, C_L \right), \quad (2.14)$$

inequality (2.6) follows.

To show (2.7), notice that from (2.12) we have

$$\begin{aligned} & P(|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X) \\ & \geq P \left(\{|\hat{g}_N(x_k) - g(x_k)| < \delta/3, \quad k = 1, \dots, r\} \cap \{|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| < \delta/3\} \right) \\ & \geq 1 - \sum_{k=1}^r P(|\hat{g}_N(x_k) - g(x_k)| \geq \delta/3) - P(|\hat{L}_N - \mathbb{E}[L(\boldsymbol{\xi})]| \geq \delta/3), \end{aligned} \quad (2.15)$$

where the last inequality stems from a direct application of Bonferroni's inequality. It follows from (2.10), (2.11) and (2.15) that

$$P(|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X) \geq 1 - \frac{r+1}{2} A e^{-\alpha N},$$

so by taking

$$K := \frac{r+1}{2} A \quad (2.16)$$

we obtain (2.7). \square

We return now to the proof of Theorem 2.3.

Proof. Let $\varepsilon > 0$ be given. Assumption A2 implies the existence of some $\delta > 0$ such that $\text{dist}(\hat{x}_N, S^*) < \varepsilon$ whenever $|\hat{g}_N(x) - g(x)| < \delta$ for all $x \in X$; see, e.g., [46, p. 69] for a proof. By Lemma 2.4, the event $\{|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X\}$ occurs with probability at least $1 - K e^{-\alpha N}$ (where both K and α depend on δ). It follows that

$$P(\text{dist}(\hat{x}_N, S^*) \geq \varepsilon) \leq K e^{-\alpha N}$$

as asserted. Notice that δ does not depend on the particular approximation $\hat{g}_N(\cdot)$; therefore, from (2.13), (2.14) and (2.16) we see that the constants K and α depend on $\{\hat{g}_N(\cdot)\}$ and $\{\hat{L}_N\}$ only through respectively the constants C_x and C_L , and the exponent functions $\gamma_x(\cdot)$ and $\gamma_L(\cdot)$ in Assumptions B1 and B1_L. \square

In essence, Theorem 2.3 says that the existence of an exponential rate of convergence for *pointwise estimators* is enough to ensure an exponential rate of convergence for *optimal solutions* of the corresponding approximating problems, regardless of the sampling scheme adopted. Although reasonably intuitive, such a result has not — to the best of our knowledge — been stated or proved anywhere in the literature.

It is important to remark that the second part of Theorem 2.3 suggests that a better pointwise convergence rate leads to a better rate of convergence of optimal solutions. Indeed, suppose one has at hand two families of approximations, say, $\{\hat{g}_N(x)\}$ and $\{\tilde{g}_N(x)\}$, whose respective exponent functions $\tilde{\gamma}_x(\cdot)$ and $\tilde{\gamma}_x(\cdot)$ in (2.1) are such that $\tilde{\gamma}_x(\cdot) \geq \tilde{\gamma}_x(\cdot)$ for all $x \in X$. Then, the corresponding constants $\tilde{\alpha}$ and $\tilde{\alpha}$

will be such that $\bar{\alpha} \geq \tilde{\alpha}$, which suggests that the family $\{\bar{g}_N(\cdot)\}$ yields a better rate of convergence of \hat{x}_N to S^* . Of course, Theorem 2.3 only gives an *upper bound* on the deviation probability $P(\text{dist}(\hat{x}_N, S^*) \geq \varepsilon)$, so no definitive statements can be made.

Nevertheless, we shall see later specific situations where the pointwise rate of convergence yields an asymptotically exact rate of convergence for the optimization problem; in those cases, superiority of one sampling scheme over another can be established.

2.1.2. The finite/piecewise linear case. We derive now results that parallel the ones in Section 2.1.1, but with the following assumption in place of Assumption A2:

Assumption A3: Either (i) the feasibility set X is finite, or (ii) X is compact, convex and polyhedral, the function $G(\cdot, \xi)$ is piecewise linear for every value of ξ , and the distribution of ξ has finite support.

Assumption A3 is useful in the context of discrete stochastic optimization (case (i)) or stochastic linear programs (case (ii)). The proposition below shows consistency of the estimators. In the proposition (and elsewhere in this paper), the statement “w.p.1 for N large enough” means that, with probability one, there exists an N_0 such that, on each sample path of the underlying process, the condition holds for all $N > N_0$. The value of such N_0 depends on the particular sample path. The proof of the proposition follows a similar argument to that of Theorem 2.6 below and is therefore omitted.

PROPOSITION 2.5. *Suppose Assumptions A1 and A3 hold. Then,*

- (i) $\hat{g}_N(x) \rightarrow g(x)$ uniformly on X w.p.1;
- (ii) $\hat{\nu}_N \rightarrow \nu^*$ w.p.1;
- (iii) $\hat{x}_N \in S^*$ w.p.1 for N large enough.

THEOREM 2.6. *Consider problem (1.3), and suppose that Assumptions A3 and B1 hold. Then, there exist constants $K > 0$ and $\alpha > 0$ such that*

$$P(\hat{x}_N \notin S^*) \leq Ke^{-\alpha N} \quad \text{for all } N \geq 1.$$

Moreover, the constants K and α depend on the family of estimators $\{\hat{g}_N(\cdot)\}$ only through the constants C_x and the exponent functions $\gamma_x(\cdot)$ in (2.1).

The proof of Theorem 2.6 will be based on the following lemma:

LEMMA 2.7. *Suppose that Assumption B1 holds, and that the set X is finite. Then, for any $\delta > 0$ there exist positive constants $A = A(\delta)$ and $\alpha = \alpha(\delta)$ such that*

$$P(|\hat{g}_N(x) - g(x)| \geq \delta) \leq Ae^{-\alpha N}, \quad \text{for all } x \in X \text{ and all } N \geq 1. \quad (2.17)$$

Moreover, there exists a positive constant K (also dependent on δ) such that

$$P(|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X) \geq 1 - Ke^{-\alpha N} \quad \text{for all } N \geq 1. \quad (2.18)$$

Proof. By setting $\alpha := \min_{x \in X} \gamma_x(\delta)$ and $A := \max_{x \in X} C_x$ in (2.1), we immediately show (2.17). The proof of (2.18) follows a very similar argument to that in the proof of Lemma 2.4 and is therefore omitted. \square

We return now to the proof of Theorem 2.6.

Proof. Suppose initially that X is finite. Let δ be defined as $(1/2) \min_{x \in X \setminus S^*} g(x) - \nu^*$. It is clear that, if $|\hat{g}_N(x) - g(x)| < \delta$ for all $x \in X$, we have that $\hat{g}_N(x) < \hat{g}_N(y)$ for all $x \in S^*$ and all $y \in X \setminus S^*$, i.e., $\hat{x}_N \in S^*$. Now suppose that the conditions in

part (ii) of Assumption A3 hold. Then, from Lemma 2.4 in [53] we know that there exists a finite set of points $\{x_1, \dots, x_\ell\} \cup \{y_1, \dots, y_q\}$ such that $x_i \in S^*$, $y_j \in X \setminus S^*$ and, if $\hat{g}_N(x_i) < \hat{g}_N(y_j)$ for all $i \in \{1, \dots, \ell\}$ and all $j \in \{1, \dots, q\}$, then $\hat{x}_N \in S^*$ (in fact, the set S_N forms a face of S^*). Therefore, we can use the same argument as in the case where X is finite. We remark that similar results were derived in [23] and [53] in the i.i.d. context.

Next, by Lemma 2.7 the event $\{|\hat{g}_N(x) - g(x)| < \delta \text{ for all } x \in X\}$ occurs with probability at least $1 - Ke^{-\alpha N}$ (where both K and α depend on δ). It follows that

$$P(\hat{x}_N \notin S^*) \leq Ke^{-\alpha N},$$

as asserted. As argued in the proof of Theorem 2.3, δ does not depend on the particular approximation $\hat{g}_N(\cdot)$, so the constants K and α depend on $\{\hat{g}_N(\cdot)\}$ only through the constants C_x and the exponent functions $\gamma_x(\cdot)$ in Assumption B1. \square

We conclude this section by mentioning that an analogous form of Theorems 2.3 and 2.6 can be derived in case Assumption B1' holds instead of B1. We state the result below for completeness; the proof follows very similar steps to the proofs of those theorems and is therefore omitted.

THEOREM 2.8. *Consider problem (1.3), and suppose that Assumption B1' hold.*

1. *Suppose that Assumptions A2 and B1_L hold. Then, given $\varepsilon > 0$, there exists a constant $\alpha > 0$ such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\text{dist}(\hat{x}_N, S^*) \geq \varepsilon) \leq -\alpha. \quad (2.19)$$

2. *Suppose that Assumption A3 holds. Then, there exists a constant $\alpha > 0$ such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\hat{x}_N \notin S^*) \leq -\alpha. \quad (2.20)$$

2.2. Convergence of estimators of optimal values. We consider now the convergence of the optimal value of (1.3). In the previous section we showed that an exponential rate of convergence for pointwise estimators leads to an exponential rate of convergence for solutions of (1.3); here, we will show that, in the context of Assumption A3, a Central Limit Theorem-type result for pointwise estimators leads to a Central Limit Theorem-type result for the optimal value of (1.3). Outside the context of A3, however, one needs more than CLT for pointwise estimators.

We start by making the following probabilistic assumptions on the estimators $\{\hat{g}_N(x)\}$:

Assumption B2: For each $x \in S^*$, the random variable $W_N(x)$ defined as

$$W_N(x) := \frac{\hat{g}_N(x) - g(x)}{\sigma_N(x)}, \quad (2.21)$$

where $\sigma_N^2(x) := \text{Var}[\hat{g}_N(x)]$, is such that $W_N(x)$ converges in distribution to a standard Normal (denoted $W_N(x) \xrightarrow{d} \text{Normal}(0, 1)$).

Of course, Assumption B2 holds in case of i.i.d. sampling under very mild assumptions — in that case it corresponds to the classical Central Limit Theorem (with $\sigma_N(x) = \sqrt{\text{Var}[G(x, \boldsymbol{\xi})]/N}$). However, as we shall see later, B2 holds in other contexts

as well. Note that we impose Assumption B2 only on the set S^* of optimal solutions to (1.1).

The lemma below states a property that will be used in the sequel.

LEMMA 2.9. *Suppose Assumptions A1 and A3 hold. Then,*

$$\hat{g}_N(\hat{x}_N) - \min_{x^* \in S^*} \hat{g}_N(x^*) = 0 \quad \text{w.p.1 for } N \text{ large enough.}$$

Proof. We have already seen in Proposition 2.5 that, under Assumptions A1 and A3, we have $\hat{x}_N \in S^*$ w.p.1 for N large enough. Consider now an arbitrary sample path where such a condition holds. Then, there exists N_0 such that $\hat{x}_N \in S^*$ for all $N > N_0$. That is, for each $N > N_0$ there exists some point $x^*(N) \in S^*$ such that $\hat{x}_N = x^*(N)$. It follows that

$$\hat{g}_N(\hat{x}_N) - \hat{g}_N(x^*(N)) = 0 \quad \text{for all } N > N_0.$$

By definition, \hat{x}_N minimizes $\hat{g}_N(\cdot)$ over X . Together with above equality, this implies that

$$\hat{g}_N(\hat{x}_N) \leq \min_{x^* \in S^*} \hat{g}_N(x^*) \leq \hat{g}_N(x^*(N)) = \hat{g}_N(\hat{x}_N) \quad \text{for all } N > N_0$$

and hence

$$\hat{g}_N(\hat{x}_N) - \min_{x^* \in S^*} \hat{g}_N(x^*) = 0 \quad \text{for all } N > N_0.$$

□

We then have the following result for rates of convergence:

THEOREM 2.10. *Consider problem (1.3), and suppose that Assumptions A1 and A3 hold. Suppose also that the estimators $\hat{g}_N(x)$ have the same variance on the set S^* of optimal solutions to (1.1), i.e., the function $\sigma_N^2(\cdot)$ is constant on S^* , and let $(\sigma_N^*)^2$ denote that common value. Then,*

$$\frac{\hat{\nu}_N - \nu^*}{\sigma_N^*} - \min_{x^* \in S^*} W_N(x^*) \xrightarrow{d} 0. \quad (2.22)$$

If, in addition, Assumption B2 holds and problem (1.1) has a unique optimal solution (call it x^), then*

$$\frac{\hat{\nu}_N - \nu^*}{\sigma_N(x^*)} \xrightarrow{d} \text{Normal}(0, 1). \quad (2.23)$$

Proof. By Lemma 2.9 we have that

$$\frac{\hat{g}_N(\hat{x}_N) - \nu^*}{\sigma_N^*} - \frac{\min_{x^* \in S^*} \hat{g}_N(x^*) - \nu^*}{\sigma_N^*} = 0 \quad \text{w.p.1 for } N \text{ large enough.}$$

Since convergence w.p.1 implies convergence in distribution, it follows that

$$\frac{\hat{g}_N(\hat{x}_N) - \nu^*}{\sigma_N^*} - \frac{\min_{x^* \in S^*} \hat{g}_N(x^*) - \nu^*}{\sigma_N^*} \xrightarrow{d} 0,$$

and hence

$$\frac{\hat{g}_N(\hat{x}_N) - \nu^*}{\sigma_N^*} - \min_{x^* \in S^*} \frac{\hat{g}_N(x^*) - \nu^*}{\sigma_N^*} \xrightarrow{d} 0.$$

Note that the term inside the min operation is actually $W_N(x^*)$. Moreover, by definition $\hat{g}_N(\hat{x}_N) = \hat{\nu}_N$, which then shows (2.22).

Suppose now that B2 holds and that $S^* = \{x^*\}$. Then, since $W_N(x^*) \xrightarrow{d} \text{Normal}(0, 1)$, using a classical result in convergence of distributions (see, e.g., [3, Theorem 3.1]) we conclude that

$$\frac{\hat{\nu}_N - \nu^*}{\sigma_N(x^*)} \xrightarrow{d} \text{Normal}(0, 1).$$

□

The above result can be slightly strengthened in case the set S^* is finite (say, $S^* = \{x^1, \dots, x^\ell\}$) and a multivariate version of Assumption B2 holds — namely, that for some deterministic sequence $\{\tau_N\}$ such that $\tau_N \rightarrow \infty$ the multivariate process $\tau_N(\hat{g}_N(x^1) - g(x^1), \dots, \hat{g}_N(x^\ell) - g(x^\ell))$ converges in distribution to a random vector Y with Normal distribution with mean vector zero and covariance matrix Σ . In that case, using a very similar argument to that used in [23], one can show directly that $\tau_N(\hat{\nu}_N - \nu^*)$ converges in distribution to $\min_{x^* \in S^*} Y(x^*)$. We chose to present our result in the above form because it only requires a univariate CLT.

As mentioned earlier, outside the context of Assumption A3 stronger conditions are required. One possibility is to assume that Assumption A2 holds and that a version of Assumption B2 for functional spaces holds for the space $C(X)$ of continuous functions defined on X . As discussed in [48], Assumption A2 suffices to ensure that each $G(\cdot, \xi)$ is a random element of the space $C(X)$, and hence $\hat{g}_N(\cdot) := N^{-1} \sum_{j=1}^N G(\cdot, \xi_j)$ is also a random element of $C(X)$. The validity of a CLT in that functional space, in turn, implies that a convergence result such as (2.23) holds. This approach works well in the i.i.d. context, see [48] for a discussion. However, we are not aware of other contexts where a CLT in a functional space exists, so we do not elaborate further on this topic.

3. Applications. We discuss now the application of the results developed in Section 2 to two classes of non-i.i.d. sampling techniques — namely, Latin Hypercube Sampling and randomized Quasi-Monte Carlo methods. Note that these techniques are devised to sample s -dimensional random vectors U that are uniformly distributed on $[0, 1]^s$ and have independent components. Given an s -dimensional random vector ξ with arbitrary distribution and not necessarily independent components, it is always possible to write $\xi = \Psi(U)$ for some mapping $\Psi : [0, 1]^s \mapsto \mathbb{R}^s$, which is constructed by inverting the conditional distribution of ξ_j given ξ_1, \dots, ξ_{j-1} , $j = 1, \dots, s$; for details see, for example, [47]. In practice, it is difficult to generate Ψ for a general multivariate distribution, so such a method is typically used when either the distribution has a special form or the components of ξ are independent. In the latter case, $\Psi(u_1, \dots, u_s) = (F_1^{-1}(u_1), \dots, F_s^{-1}(u_s))$, where F_j^{-1} is the inverse of the cumulative distribution function F_j of ξ_j , defined as $F_j^{-1}(u) := \inf\{y \in \Xi_j : F_j(y) \geq u\}$, and Ξ_j denotes the support of the distribution F_j . For the remainder of this paper we assume that the components of ξ are independent. Moreover, since $G(x, \xi) = G(x, \Psi(U))$, we will restrict the domain of $G(x, \cdot)$ to $\Xi_1 \times \dots \times \Xi_s$. Sometimes we will refer to the function $G(x, \Psi(\cdot))$, which is defined on $[0, 1]^s$.

3.1. Latin Hypercube Sampling. Stratified sampling techniques have been used in Statistics and Simulation for years (see, for instance, [11] and references therein). Generally speaking, the idea is to partition the sample space and fix the number of samples on each component of the partition, which should be proportional

to the probability of that component. This way we ensure that the number of sampled points on each region will be approximately equal to the *expected* number of points to fall in that region. It is intuitive that such a procedure yields smaller variance than crude Monte Carlo; for proofs see [11]. Notice however that, though theoretically appealing, implementing such a procedure is far from trivial, since the difficulty is to determine the partition as well as to compute the corresponding probabilities.

There are many variants of this basic method, one of the most well-known being the so-called *Latin Hypercube Sampling* (LHS), introduced in [31]. The LHS method operates as follows: suppose we want to draw N samples from a random vector $\boldsymbol{\xi}$ with s independent components ξ_1, \dots, ξ_s , each of which has a Uniform(0,1) distribution. The algorithm consists repeating the two steps below for each dimension $j = 1, \dots, s$:

1. Generate

$$Y^1 \sim U\left(0, \frac{1}{N}\right), Y^2 \sim U\left(\frac{1}{N}, \frac{2}{N}\right), \dots, Y^N \sim U\left(\frac{N-1}{N}, 1\right);$$

2. Let $\xi_j^i := Y^{\pi(i)}$, where π is a random permutation of $1, \dots, N$.

In [31], it is shown that each sample ξ_j^i (viewed as a random variable) has *the same distribution* as ξ_j , which in turn implies the estimators generated by the LHS method are unbiased. In case of arbitrary distributions, the above procedure is easily modified by drawing the sample as before and applying the inversion method discussed at the beginning of Section 3 to generate the desired random variates.

It is also shown in [31] that, under some conditions, the LHS method does indeed reduce the variance compared to crude Monte Carlo. The papers [37, 55] show that, asymptotically (i.e., as the sample size N goes to infinity), LHS is never worse than crude Monte Carlo, even without the assumptions of [31]. More specifically, $V_{LHS} \leq N/(N-1)V_{MC}$, where V_{LHS} and V_{MC} are respectively the variances under LHS and crude Monte Carlo.

3.1.1. Exponential rate of convergence. Suppose the objective function $g(\cdot)$ in (1.1) is approximated by a sample average calculated using the LHS method, i.e., for each $i = 1, \dots, s$, ξ_i^1, \dots, ξ_i^N are samples of ξ_i (the i th component of $\boldsymbol{\xi}$) constructed using the LHS method. Call the resulting estimator in (1.2) $\hat{g}_N^{\text{LHS}}(x)$. To study convergence properties of the approximating problem in (1.3), we shall use the tools of Section 2. Our first goal is to show that the family $\{\hat{g}_N^{\text{LHS}}(\cdot)\}$ satisfies Assumption B1, so that we can apply Theorems 2.3 and 2.6 to ensure exponential rate of convergence.

We shall restrict our attention to functions satisfying the following assumption.

Assumption C1: For each $x \in X$, the function $G(x, \cdot)$ is monotone in each component. That is, for each $i = 1, \dots, s$ and each $\delta > 0$ we have

$$\text{either} \quad G(x, z + \delta e_i) \geq G(x, z) \quad \text{for all } z \in \mathbb{R}^s \quad (3.1)$$

$$\text{or} \quad G(x, z + \delta e_i) \leq G(x, z) \quad \text{for all } z \in \mathbb{R}^s, \quad (3.2)$$

where as customary e_i denotes the vector with 1 in the i th component and zeros otherwise.

An important case where such an assumption is satisfied is that of two-stage stochastic linear programs with fixed recourse. In Section 4 we discuss that case in detail.

An alternative assumption is the following:

Assumption C1': For each $x \in X$, the function $G(x, \cdot)$ is *additive*, i.e., there exist functions G_1, \dots, G_s (all of them mapping $\mathbb{R}^n \times \mathbb{R}$ to \mathbb{R}) such that $G(x, \boldsymbol{\xi}) = G_1(x, \xi_1) + \dots + G_s(x, \xi_s)$. Moreover, $|\mathbb{E}[G_j(x, \xi_j)]| < \infty$, the functions $G_j(x, F_j^{-1}(\cdot))$ have at most a finite number of singularities (i.e., points where the function approaches $\pm\infty$), and the set of points at which $G_j(x, F_j^{-1}(\cdot))$ is discontinuous has Lebesgue measure zero.

The importance of Assumptions C1 and C1' in the present context is given by the results below.

THEOREM 3.1. *Suppose that (i) Assumption C1 holds, and (ii) for each $x \in X$, the moment generating function of $G(x, \boldsymbol{\xi})$ (denoted $\phi_x^{\text{MC}}(t) := \mathbb{E}[e^{tG(x, \boldsymbol{\xi})}]$) is finite everywhere. Consider the LHS estimators $\hat{g}_N^{\text{LHS}}(\cdot)$ above defined and the corresponding problem $\min_{x \in X} \hat{g}_N^{\text{LHS}}(x)$. Let \hat{x}_N^{LHS} denote an optimal solution of that problem. Then,*

1. *If Assumption A2 holds with a uniform Lipschitz constant L (i.e., $L(\cdot) \equiv L$), then given $\varepsilon > 0$ there exists constants $\tilde{K} > 0$ and $\tilde{\alpha} > 0$ such that*

$$P(\text{dist}(\hat{x}_N^{\text{LHS}}, S^*) \geq \varepsilon) \leq \tilde{K}e^{-\tilde{\alpha}N} \quad \text{for all } N \geq 1. \quad (3.3)$$

2. *If Assumption A3 holds, then there exist a constant $\tilde{\alpha} > 0$ such that*

$$P(\hat{x}_N^{\text{LHS}} \notin S^*) \leq \tilde{K}e^{-\tilde{\alpha}N} \quad \text{for all } N \geq 1. \quad (3.4)$$

Moreover, in either case the exponent $\tilde{\alpha}$ is at least as large as the corresponding exponent obtained for standard Monte Carlo.

Proof. Let $\phi_N(x, t) := \frac{1}{N} \log \mathbb{E} \left[e^{tN\hat{g}_N^{\text{LHS}}(x)} \right]$. If conditions (i) and (ii) above hold, then by Proposition 6 in [9] we have that $\phi_N(x, t) \leq \phi_x^{\text{MC}}(t)$ for all x and all t and hence it follows from Proposition 2.1 that Assumption B1 holds for $\{\hat{g}_N^{\text{LHS}}(\cdot)\}$. Moreover, in case 1 of the theorem (i.e., when Assumption A2 holds with $L(\cdot) \equiv L$) Assumption B1_L is trivially satisfied. The two cases of the theorem then parallel Theorems 2.3 and 2.6, which shows (3.3) and (3.4).

The last assertion of the theorem is a consequence of the remark following the proof of Theorem 2.3. Indeed, the arguments in the previous paragraph show that the constants C_x and the exponent functions $\gamma_x(\cdot)$ used to show (2.1) are the same for both LHS and standard Monte Carlo. \square

Although Theorem 3.1 only guarantees the same bounds for both LHS and standard Monte Carlo, a closer look at the proof of the inequality $\phi_N(x, \cdot) \leq \phi_x^{\text{MC}}(\cdot)$ in [9] shows that such an inequality is essentially a consequence of Jensen's inequality, which often holds strictly; hence, generally speaking, LHS tends to behave better than Monte Carlo.

In case Assumption C1' holds instead of C1, we have the following stronger result:

THEOREM 3.2. *Suppose that the assumptions of Theorem 3.1 are satisfied, but Assumption C1' holds instead of C1. Then, the conclusions of Theorem 3.1 hold. In addition, we have*

1. *If Assumption A2 holds with a uniform Lipschitz constant L (i.e., $L(\cdot) \equiv L$), then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P(\text{dist}(\hat{x}_N^{\text{LHS}}, S^*) \geq \varepsilon) = -\infty. \quad (3.5)$$

2. *If Assumption A3 holds, then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P(\hat{x}_N^{\text{LHS}} \notin S^*) = -\infty. \quad (3.6)$$

Proof. The proof of the first part of the theorem follows the same steps as the proof of Theorem 3.1 (except that Proposition 4 in [9] is invoked instead of Proposition 6).

To show the second part, by writing each random variable ξ_j as $F_j^{-1}(U_j)$ (where $U_j \sim U(0,1)$), we have that conditions (i) and (ii) of Theorem 3.1 ensure that the assumptions of Theorem 3 in [9] are satisfied. The latter result, in turn, ensures that Assumption B1' holds with the function $\gamma_x = \infty$ everywhere except at zero, where it is equal to zero. Then, (3.5) and (3.6) follow from (2.19) and (2.20) in Theorem 2.8. \square

The strength of Theorem 3.2, of course, lies in the asymptotic results (3.5)-(3.6), which show that in the additive case the rate of convergence under LHS is *superexponential*.

3.1.2. Central Limit Theorem. We study now the convergence of optimal values of the approximating problem (1.3) under LHS. To do so we shall apply results of Section 2.2. Before that, however, we need to review some results related to the ANOVA decomposition of a function.

Let $U = (U_1, \dots, U_s)$ be an s -dimensional random vector of independent components with uniform distribution on $[0, 1]^s$, $f : [0, 1]^s \rightarrow \mathbb{R}$ an arbitrary measurable function and consider the problem of estimating $I := \mathbb{E}[f(U)]$. It is shown in [55] that, when $\mathbb{E}[f(U)^2] < \infty$, f can be decomposed as

$$f(u_1, \dots, u_s) = \mathbb{E}[f(U)] + \sum_{k=1}^s f_k(u_k) + r(u_1, \dots, u_s), \quad (3.7)$$

where $f_k(u_k) = \mathbb{E}[f(U) | U_k = u_k] - \mathbb{E}[f(U)]$ and $r(u)$ is the *residual* term, which satisfies $\mathbb{E}[r(U) | U_j = u_j] = 0$ for all j and all u_j . [55] also shows that the residual term can be viewed as a “residual from additivity”, in the following sense. We say that a function $g : [0, 1]^s \rightarrow \mathbb{R}$ is *additive* if there exist unidimensional functions g_1, \dots, g_s and a constant C such that g can be written as $g(u_1, \dots, u_s) = C + \sum_k g_k(u_k)$ for almost all $u \in [0, 1]^s$ (where “almost all” refers to the Lebesgue measure). Then, the additive function $f_a : [0, 1]^s \rightarrow \mathbb{R}$, defined as $f_a(u_1, \dots, u_s) = \mathbb{E}[f(U)] + \sum_{k=1}^s f_k(u_k)$ is the best additive fit to f in L^2 -norm, i.e., it minimizes $\mathbb{E}[(f(U) - g(U))^2]$ over all additive functions g . Note that if f is itself additive then the residual $r(u) = f(u) - f_a(u)$ will be equal to zero almost everywhere (a.e.); conversely, if $r(u) = 0$ a.e. then $f(u) = f_a(u)$ a.e., so in that case f is additive.

The variance of the estimator I_{LHS} (defined as $I_{\text{LHS}} := N^{-1} \sum_{i=1}^N f(U^i)$, where U^1, \dots, U^N are samples drawn with LHS) satisfies

$$\sigma_N^2 := \text{Var}[I_{\text{LHS}}] = N^{-1} \mathbb{E}[(r(U))^2] + o(N^{-1}); \quad (3.8)$$

see [55]. Using (3.7) and (3.8), it is shown in [33] that, when f is bounded, a Central Limit Theorem holds for the estimator I_{LHS} under Latin Hypercube Sampling. More specifically, it is shown that

$$N^{1/2}(I_{\text{LHS}} - I) \xrightarrow{d} \text{Normal}(0, \sigma^2), \quad \text{where } \sigma^2 := \mathbb{E}[(r(U))^2]. \quad (3.9)$$

Next, notice that from (3.8) we can write

$$\frac{I_{\text{LHS}} - I}{\sigma_N} = \frac{N^{1/2}(I_{\text{LHS}} - I)}{\left[\sigma^2 + \frac{o(N^{-1})}{N^{-1}}\right]^{1/2}}.$$

Since $N^{1/2}(I_{\text{LHS}} - I) \xrightarrow{d} \text{Normal}(0, \sigma^2)$ and the deterministic sequence $\{[\sigma^2 + \frac{o(N^{-1})}{N^{-1}}]^{1/2}\}$ converges to σ , it follows from a classical result in probability theory (see, e.g., [3, p. 29]) that, when $\sigma > 0$,

$$\frac{I_{\text{LHS}} - I}{\sigma_N} \xrightarrow{d} \frac{1}{\sigma} \text{Normal}(0, \sigma^2) = \text{Normal}(0, 1). \quad (3.10)$$

Notice that the condition $\mathbb{E}[f(U)^2] < \infty$ also implies that a Strong Law of Large Numbers holds for LHS, i.e.,

$$|I_{\text{LHS}} - I| \rightarrow 0 \quad \text{w.p.1}; \quad (3.11)$$

for a proof, see [27].

By applying (3.10) to our setting we see that Assumption B2 holds for LHS when, for every $x \in S^*$, the random variable $G(x, \boldsymbol{\xi})$ is bounded and the function $G(x, \Psi(\cdot))$ has a nonzero ANOVA residual. As seen above, the latter condition means that $G(x, \Psi(\cdot))$ is not additive, i.e., it cannot be written in the form $C + \sum_k g_k(u_k)$. It is easy to see that this is equivalent to saying that the function $G(x, \cdot)$ is not additive; note that here we extend the definition of additivity to the domain of $G(x, \cdot)$, which as discussed before is restricted to the support of $\boldsymbol{\xi}$.

On the other hand, if $\mathbb{E}[G(x, \boldsymbol{\xi})^2] < \infty$ for all $x \in X$ then (3.11) implies that Assumption A1 holds. Thus, under additional assumptions we can apply Theorem 2.10 and Propositions 2.2 and 2.5. We summarize the result in the theorem below.

THEOREM 3.3. *Consider the LHS estimators $\hat{g}_N^{\text{LHS}}(\cdot)$ above defined and the corresponding problem $\min_{x \in X} \hat{g}_N^{\text{LHS}}(x)$. Let \hat{x}_N^{LHS} and $\hat{\nu}_N^{\text{LHS}}$ denote respectively an optimal solution and the optimal value of that problem. Suppose that $\mathbb{E}[G(x, \boldsymbol{\xi})^2] < \infty$ for all $x \in X$.*

1. *If Assumption A2 holds, then $\text{dist}(\hat{x}_N^{\text{LHS}}, S^*) \rightarrow 0$ w.p.1 and $\hat{\nu}_N^{\text{LHS}} \rightarrow \nu^*$ w.p.1.*
2. *If Assumption A3 holds, then $\hat{x}_N^{\text{LHS}} \in S^*$ w.p.1 for N large enough and $\hat{\nu}_N^{\text{LHS}} \rightarrow \nu^*$ w.p.1. In addition, if problem (1.1) has a unique optimal solution (call it x^*), the random variable $G(x^*, \boldsymbol{\xi})$ is bounded, and the function $G(x^*, \cdot)$ is not additive, then*

$$\frac{\hat{\nu}_N^{\text{LHS}} - \nu^*}{\sigma_N(x^*)} \xrightarrow{d} \text{Normal}(0, 1),$$

where $\sigma_N^2(x^*)$ is the variance of $\hat{g}_N^{\text{LHS}}(x^*)$. Moreover, there exists a positive constant C such that

$$\sigma_N^2(x^*) = N^{-1}C + o(N^{-1}). \quad (3.12)$$

Theorem 3.3 shows that the rate of convergence of optimal values under LHS (under the conditions of case 2 of the theorem) is $N^{-1/2}$. Thus, compared to standard Monte Carlo we can see that, although LHS will likely reduce the variance of pointwise estimators, it cannot improve the *rate* of convergence unless $G(x^*, \cdot)$ is additive; in that case, we expect the convergence rate to be much faster. Indeed, recall from Theorem 3.2 that, under the assumptions of that theorem (which include additivity), convergence of optimal solutions is superexponential. Note also that, when S^* is finite (but not necessarily a singleton), and $G(x^*, \boldsymbol{\xi})$ is bounded and $G(x^*, \cdot)$ is not additive for all $x^* \in S^*$, the stronger result discussed in the paragraph following the proof of Theorem 2.10 applies with $\tau_N = N^{1/2}$, since the aforementioned CLT result proved in [33] is also valid in a multivariate context.

3.2. Randomized Quasi Monte Carlo. For completeness, we provide in this section a brief review of Quasi-Monte Carlo (QMC) techniques. We follow mostly [32], which we refer to for a comprehensive treatment of QMC concepts. Let U be an s -dimensional random vector with uniform distribution on $[0, 1]^s$, $f : [0, 1]^s \rightarrow \mathbb{R}$ an arbitrary function and consider the problem of estimating $I := \mathbb{E}[f(U)]$.

The basic idea of QMC is to calculate a sample average estimate as in the standard Monte Carlo but, instead of drawing a random sample from the uniform distribution on $[0, 1]^s$, a certain set of points u^1, \dots, u^N on space $[0, 1]^s$ is carefully chosen. The deterministic estimate

$$I_{\text{QMC}} := \frac{1}{N} \sum_{i=1}^N f(u^i) \quad (3.13)$$

is constructed. A key result is the so-called Koksma-Hlawka inequality which, roughly speaking, states that the quality of the approximation given by I_{QMC} depends on the quality of the chosen points (measured by the difference between the corresponding empirical measure and the uniform distribution, which is quantified by the so-called *star-discrepancy*) as well as on the nature of the function f (measured by its total variation). A great deal of the research on QMC methods aims at determining ways to construct *low-discrepancy sequences*, i.e., sequences of points u^1, u^2, \dots for which the star-discrepancy is small for all N . A particular type of sequence that has proven valuable is defined in terms of (t, m, s) -nets. We need some definitions before delving into more details, which we do next.

Let $b \geq 2$ be an arbitrary integer, called the base. An *elementary interval in base b* (in dimension s) is a subinterval E of $[0, 1]^s$ of the form

$$E = \prod_{j=1}^s \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right]$$

for nonnegative integers $\{a_j\}$ and $\{d_j\}$ such that $a_j < b^{d_j}$ for all j . The volume of E is $b^{-\sum_j d_j}$. Next, let t and m be nonnegative integers such that $t \leq m$. A finite sequence of b^m points is a (t, m, s) -net in base b if every elementary interval in base b of volume b^{t-m} contains exactly b^t points of the sequence. A sequence of points u^1, u^2, \dots is a (t, s) -sequence in base b if, for all integers $k \geq 0$ and $m > t$, the set of points consisting of the u^n such that $kb^m \leq n < (k+1)b^m$ is a (t, m, s) -net in base b .

The advantage of (t, m, s) -nets becomes clear from a result due to Niederreiter [32, Theorems 4.10 and 4.17], who shows that the error $|I_{\text{QMC}} - I|$ is: (i) of order $(\log N)^{s-1}/N$ when I_{QMC} is computed from a (t, m, s) -net in base b with $m > 0$; (ii) of order $(\log N)^s/N$ when I_{QMC} is computed from the first $N \geq 2$ terms of a (t, s) -sequence in base b . Note that in case (i) N must be equal to b^m , whereas in case (ii) N is arbitrary, which explains the weaker bound. In either case, it is clear that, asymptotically, the error is smaller than $N^{-1/2}$ given by standard Monte Carlo methods.

Despite the advantage of QMC with respect to error rates, the method has two major drawbacks:

- (a) The bounds provided by the Koksma-Hlawka inequality involve difficult-to-compute quantities such as the total variation of f , i.e., they yield qualitative (rather than quantitative) results; hence, obtaining an exact estimate of the error may be difficult.

- (b) A comparison of the functions $(\log N)^s/N$ and $N^{-1/2}$ shows that even though asymptotically the error from QMC is smaller than the error from standard Monte Carlo, such an advantage does not appear until N is very large, unless s is small.

These difficulties have long been realized by the QMC community, and various remedies have been proposed. A common way to overcome difficulty (a) above is to incorporate some randomness into the choice of QMC points. By doing so, errors can be estimated using standard methods, e.g., via multiple independent replications. This is the main idea of *randomized* QMC methods (RQMC), see [12] and [38] for detailed discussions.

One particular technique we are interested in using relies on “scrambling” the decimal digits of each point of a (t, s) -sequence in a proper way. This idea was proposed in [34], and has gained popularity due to the nice properties of the randomized sequence. We shall use these properties below.

3.2.1. Using QMC in optimization. Consider again the family of estimators defined in (1.2). Suppose that $\{\xi^i\}$ is generated by a (t, s) -sequence, and call the resulting family $\{\hat{g}_N^{\text{QMC}}(x)\}$.

Let us fix $x \in X$ for a moment. As argued in [40] — who in turn cite a result in [30] — the empirical measure defined by a QMC sequence converges weakly to the uniform distribution, provided that the star-discrepancy of that sequence goes to zero, which is the case of (t, s) -sequences. It follows from [3, Theorem 2.7] that, if $G(x, \xi)$ is bounded, then $\hat{g}_N^{\text{QMC}}(x) \rightarrow g(x)$ as $N \rightarrow \infty$. Now suppose that $\{\xi^i\}$ is generated by a scrambled (t, s) -sequence, and call the corresponding estimator \hat{g}_N^{RQMC} . In [35] it is shown that scrambled (t, s) -sequences are (t, s) -sequences with probability one, which then implies that

$$\hat{g}_N^{\text{RQMC}}(x) \rightarrow g(x) \text{ w.p.1.} \quad (3.14)$$

Moreover, $\hat{g}_N^{\text{RQMC}}(x)$ is an unbiased estimator of $g(x)$, i.e., $\mathbb{E}[\hat{g}_N^{\text{RQMC}}(x)] = g(x)$. Notice that the term “with probability one” above refers to the probability space where the random variables defining the permutations that are part of the scrambling algorithm lie. We assume that this probability space is the same as the one where the random vectors ξ are defined.

For some of the results that follow we will need the following assumption.

Assumption D1: The following conditions hold for each $x \in S^*$:

$$\left| \frac{\partial^s}{\partial u_1 \dots \partial u_s} G(x, \Psi(u_1, \dots, u_s)) - \frac{\partial^s}{\partial u_1 \dots \partial u_s} G(x, \Psi(v_1, \dots, v_s)) \right| \leq B \|u - v\|^\beta \quad (3.15)$$

(for some $B > 0$ and some $\beta \in (0, 1]$), and

$$\int_{[0,1]^s} \left[\frac{\partial^s}{\partial u_1 \dots \partial u_s} G(x, \Psi(u_1, \dots, u_s)) \right]^2 du > 0, \quad (3.16)$$

where $\Psi(u_1, \dots, u_s) = (F_1^{-1}(u_1), \dots, F_s^{-1}(u_s))$.

A few remarks about cases where Assumption D1 is satisfied are now in order. Suppose momentarily that G is infinitely differentiable in the second argument and

that each F_j^{-1} is differentiable as well. Then, we have that

$$\begin{aligned} \frac{\partial}{\partial u_1} G(x, F_1^{-1}(u_1), \dots, F_s^{-1}(u_s)) &= \\ &= \frac{\partial}{\partial \xi_1} G(x, \xi_1, F_2^{-1}(u_2), \dots, F_s^{-1}(u_s)) \Big|_{\xi_1 = F_1^{-1}(u_1)} \frac{\partial}{\partial u_1} F_1^{-1}(u_1) \end{aligned}$$

so by repeating the calculation for the higher-order mixed derivatives we obtain that

$$\begin{aligned} H(u_1, \dots, u_s) &:= \frac{\partial^s}{\partial u_1 \dots \partial u_s} G(x, F_1^{-1}(u_1), \dots, F_s^{-1}(u_s)) = \quad (3.17) \\ &= \frac{\partial^s}{\partial \xi_1 \dots \partial \xi_s} G(x, \xi_1, \dots, \xi_s) \Big|_{\xi_j = F_j^{-1}(u_j)} \frac{\partial}{\partial u_1} F_1^{-1}(u_1) \dots \frac{\partial}{\partial u_s} F_s^{-1}(u_s). \quad (3.18) \end{aligned}$$

It follows that, if the gradient of the function H defined in (3.17) is uniformly bounded for all $u \in [0, 1]^s$, then H is Lipschitz (see, e.g. [2, Corollary 40.6]), i.e., (3.15) holds. A sufficient condition for uniform boundedness of $\nabla H(u)$ on $[0, 1]^s$ is its continuity on $[0, 1]^s$. Equation (3.18) shows that continuous differentiability of G (up to order $s + 1$) and F_j^{-1} , $j = 1, \dots, s$ (up to second order) on the closed set $[0, 1]^s$ suffice for that. Of course, imposing a continuous differentiability assumption on F_j^{-1} restricts the type of distributions that can be used; we shall return to that issue shortly.

Condition (3.16) essentially says that interactions of order up to s are significant, at least on a set of positive probability. For example, (3.16) does not hold if $G(x, \cdot)$ is linear for $x \in S^*$, since the mixed derivatives of any order bigger than 1 are equal to zero. Situations like that suggest that the *effective dimension* (see [35]) of the problem is less than s — indeed, in the linear case the effective dimension is 1. In that case, one should apply Quasi-Monte Carlo only to the most significant variables, for which mutual interaction is significant.

Applying the above results on randomized QMC to the general context of Section 2.2 we obtain the following:

THEOREM 3.4. *Consider the RQMC estimators $\hat{g}_N^{RQMC}(\cdot)$ above defined and the corresponding problem $\min_{x \in X} \hat{g}_N^{RQMC}(x)$. Let \hat{x}_N^{RQMC} and $\hat{\nu}_N^{RQMC}$ denote respectively an optimal solution and the optimal value of that problem. Suppose that $G(x, \xi)$ is bounded for all $x \in X$.*

1. *If Assumption A2 holds, then $\text{dist}(\hat{x}_N^{RQMC}, S^*) \rightarrow 0$ w.p.1 and $\hat{\nu}_N^{RQMC} \rightarrow \nu^*$ w.p.1.*
2. *If Assumption A3 holds, then $\hat{x}_N^{RQMC} \in S^*$ w.p.1 for N large enough and $\hat{\nu}_N^{RQMC} \rightarrow \nu^*$ w.p.1. If, in addition, Assumption D1 holds, problem (1.1) has a unique optimal solution (call it x^*) and the samples $\{\xi^i\}$ are generated by a scrambled $(0, m, s)$ -net (i.e., $t = 0$), then*

$$\frac{\hat{\nu}_N^{RQMC} - \nu^*}{\sigma_N(x^*)} \xrightarrow{d} \text{Normal}(0, 1), \quad (3.19)$$

where $\sigma_N^2(x^*)$ is the variance of $\hat{g}_N^{RQMC}(x^*)$. Moreover, in the latter case there exist positive constants c and C such that

$$c \frac{(\log_b N)^{s-1}}{N^3} \leq \sigma_N^2(x^*) \leq C \frac{(\log_b N)^{s-1}}{N^3} \quad (3.20)$$

as $m \rightarrow \infty$.

Proof. Let us fix $x \in X$. The assertion in case 1 and the first assertion in case 2 follow directly from (3.14) (which implies that Assumption A1 holds) and Propositions 2.2 and 2.5.

Consider now the random variable $W(x)$ defined as

$$W(x) := \frac{\hat{g}_N^{\text{RQMC}}(x) - g(x)}{\sigma_N(x)}, \quad (3.21)$$

where $\sigma_N^2(x) := \text{Var}[\hat{g}_N^{\text{RQMC}}(x)]$. Here we resort to a key result on scrambled (t, m, s) -nets proved in [28] — building upon previous results in [35, 36] — that says that a Central Limit Theorem holds for pointwise estimators constructed with a scrambled $(0, m, s)$ -net. Assumption D1 translates the conditions in [28] into our notation. It follows that, under D1, $W(x)$ converges in distribution to the standard normal for each $x \in S^*$, i.e., Assumption B2 holds and hence the conclusion follows from Theorem 2.10. \square

Theorem 3.4 shows the benefits of using randomized Quasi-Monte Carlo methods in optimization. Essentially, it says that, in the setting of case 2 of the theorem, the convergence rate of optimal values is of order $[(\log N)^{s-1}/N^3]^{1/2}$, which asymptotically is better than the $N^{-1/2}$ obtained with standard Monte Carlo. This suggests that RQMC methods can be very efficacious for stochastic optimization. Note however that, strictly speaking, (3.19) applies only to the case where X is finite, since the assumption of finite support of ξ in the second case of Assumption A3 conflicts with the smoothness condition in Assumption D1. We will discuss the smoothness issue in more detail later. Note also that (3.19) is valid only for scrambled $(0, m, s)$ -nets, which restricts the choice of the base b — as shown in [32], when $m \geq 2$ a $(0, m, s)$ -net in base b can only exist if $b \geq s - 1$.

4. Two-Stage Stochastic Programs. In this section we discuss the application of the results outlined in the previous sections to two-stage stochastic linear programs (see, e.g., [4] for a comprehensive discussion of this class of problems). We consider problems of the form

$$\min_{x \in X} c^t x + \mathbb{E}[Q(x, \xi)], \quad (4.1)$$

where X is a convex polyhedral set,

$$Q(x, \xi) = \inf \{q^t y : W y \leq h - T x, y \geq 0\}, \quad (4.2)$$

and $\xi = (h, T)$. As before, ξ is an s -dimensional random vector with arbitrary distribution. Let $G(x, \xi)$ denote the function $c^t x + Q(x, \xi)$; then, we see that the above problem falls in the framework of (1.1).

The use of Monte Carlo sampling to solve two-stage problems has been extensively studied in the literature, both from algorithmic (e.g., [17, 19, 26, 52]) and theoretical perspectives (see, for instance, [50] for a compilation of results).

Note that the function $Q(x, \xi)$ can be written in the form $Q(x, \xi) = \tilde{Q}(h - T x)$, where

$$\tilde{Q}(z) = \inf \{q^t y : W y \leq z, y \geq 0\}. \quad (4.3)$$

By duality, we see that the function $\tilde{Q}(\cdot)$ can be represented in the form

$$\tilde{Q}(z) = \sup \{u^t z : W^t u \leq q, u \leq 0\}. \quad (4.4)$$

For the sake of simplicity we assume that: (i) for every vector z the system $Wy \leq z, y \geq 0$, has a solution (the recourse is complete), and (ii) the system $W^t u \leq q, u \leq 0$ has a solution (dual feasibility). Under these assumptions, $\bar{Q}(\cdot)$ is a finite valued, piecewise linear convex function. This in turn implies that the function $G(x, \xi)$ is also piecewise linear convex in each argument and can be written as

$$G(x, \xi) = \max_{k=1, \dots, r} c^t x + (v^k)^t (h - Tx), \quad (4.5)$$

where v^1, \dots, v^r are the vertices of the polyhedron $\{u : W^t u \leq q, u \leq 0\}$. Furthermore, by standard subdifferential calculus we have that the subdifferential set of $G(x, \xi)$ with respect to x is given by

$$\partial_x G(x, \xi) = \text{conv}\{c - T^t v^k : G(x, \xi) = c^t x + (v^k)^t (h - Tx), k = 1, \dots, r\}, \quad (4.6)$$

where “conv” denote the convex hull of the set.

In the discussion that follows we assume that the feasibility set X is compact.

4.1. LHS results. In order to apply the results for Latin Hypercube Sampling discussed in Section 3.1, we need to verify that the corresponding assumptions are satisfied.

Assume momentarily that the matrix T is deterministic, so that $\xi = h$. Consider Assumption A2. It follows from (4.6) that $\partial_x G(x, \xi)$ is uniformly bounded for all x and all ξ and thus, by a version of the mean-value theorem for subdifferentiable functions (see, e.g., [18, Theorem VI.2.3.3]), we conclude that A2 holds with a constant L such that $L(\cdot) \equiv L$. Next, notice that from (4.3) we have $G(x, \xi) = \min\{q^t y : Wy \leq \xi - Tx, y \geq 0\}$. Thus, for any $\delta > 0$ we have that $G(x, \xi + \delta e_i) \leq G(x, \xi)$, i.e., Assumption C1 holds.

It follows from the above discussion and from Theorem 3.1 that, if the moment generating function of $G(x, \xi)$ is finite everywhere for all x , then given $\varepsilon > 0$ there exist constants $\tilde{K} > 0$ and $\tilde{\alpha} > 0$ such that

$$P(\text{dist}(\hat{x}_N^{\text{LHS}}, S^*) \geq \varepsilon) \leq \tilde{K} e^{-\tilde{\alpha} N} \quad \text{for all } N \geq 1.$$

Moreover, the exponent $\tilde{\alpha}$ is at least as large as the corresponding exponent obtained for standard Monte Carlo. This suggests that convergence under LHS will indeed be faster than under standard Monte Carlo.

As mentioned earlier, $G(\cdot, \xi)$ is piecewise linear. Thus, if ξ has finite support then Assumption A3 holds, so from Theorem 3.1 we have that

$$P(\hat{x}_N^{\text{LHS}} \notin S^*) \leq \tilde{K} e^{-\tilde{\alpha} N} \quad \text{for all } N \geq 1.$$

It is fruitful to compare the above result with the i.i.d. case derived in [53]. Indeed, when problem (4.1) has a unique solution x^* , a slightly modified proof of Theorem 3.2 in [53] shows that there exists $\beta > 0$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\hat{x}_N \neq x^*) = -\beta, \quad (4.7)$$

where \hat{x}_N is the solution obtained with standard Monte Carlo. Moreover, the constant β is given by the minimum of a number of pointwise rates $-\gamma_x(\delta_0)$ in (2.1) (for a fixed $\delta_0 > 0$) over a *finite* number of x 's. Since finite support of ξ implies that the moment generating function of $G(x, \xi)$ is finite everywhere for all x , it follows from

Proposition 6 in [9] that the pointwise rates $-\gamma_x(\delta_0)$ under LHS are no worse than under Monte Carlo. It follows that, when LHS is applied, an equation similar to (4.7) holds and the resulting constant β is *no worse* than under Monte Carlo. Since the rate in (4.7) is exact, we conclude that LHS can only improve upon Monte Carlo in this setting.

Next, we apply Theorem 3.3 to the present context. As seen above, Assumption A2 holds if T is deterministic. Note that A2 holds even if we allow T to be random (i.e., $\boldsymbol{\xi} = (h, T)$), as long as the distribution of $\boldsymbol{\xi}$ has bounded support, since in that case $\partial_x G(x, \boldsymbol{\xi})$ in (4.6) is uniformly bounded for all x . Moreover, under such condition clearly (4.5) implies that $G(x, \boldsymbol{\xi})$ is bounded for each x . Theorem 3.3 then ensures that $\text{dist}(\hat{x}_N^{\text{LHS}}, S^*) \rightarrow 0$ w.p.1 and $\hat{\nu}_N^{\text{LHS}} \rightarrow \nu^*$ w.p.1. Now, suppose again that the distribution of $\boldsymbol{\xi}$ has finite support, so Assumption A3 holds. Then, $\hat{x}_N^{\text{LHS}} \in S^*$ w.p.1 for N large enough and $\hat{\nu}_N^{\text{LHS}} \rightarrow \nu^*$ w.p.1. It follows that, if problem (1.1) has a unique optimal solution x^* and $G(x^*, \cdot)$ is not additive, then

$$\frac{\hat{\nu}_N^{\text{LHS}} - \nu^*}{\sigma_N(x^*)} \xrightarrow{d} \text{Normal}(0, 1),$$

where $\sigma_N(x^*) := \text{Var}[\hat{g}_N^{\text{LHS}}(x^*)] = N^{-1}C + o(N^{-1})$ for some positive constant C . Note that the non-additivity assumption is reasonable in this setting, since at the optimal solution x^* typically it happens that the maximum in (4.5) is achieved by more than one k , so $G(x^*, \cdot)$ is not linear.

4.2. QMC results. We now apply the results from Section 3.2 to the two-stage stochastic programming model described above. As seen above, a sufficient condition for Assumption A2 to hold is that the distribution of $\boldsymbol{\xi}$ have bounded support. In that case, we have from Theorem 3.4 that $\text{dist}(\hat{x}_N^{\text{RQMC}}, S^*) \rightarrow 0$ w.p.1 and $\hat{\nu}_N^{\text{RQMC}} \rightarrow \nu^*$ w.p.1. When $\boldsymbol{\xi}$ has finite support (i.e., Assumption A3 holds), we obtain a stronger result, namely, that $\hat{x}_N^{\text{RQMC}} \in S^*$ w.p.1 for N large enough.

The second part of Theorem 3.4 — which deals with convergence rates — unfortunately is not applicable in this context. The reason, as pointed out in the discussion following Theorem 3.4, is that the smoothness condition stated in Assumption D1 cannot hold in this case, since $G(x, \cdot)$ is non-differentiable for each x . Moreover, the assumption that $\boldsymbol{\xi}$ has finite support causes the inverse cdfs F_j^{-1} to be discontinuous. Note however that, as recognized in [28], the smoothness condition in Assumption D1 is only *sufficient* for the proof of the Central Limit Theorem for scrambled $(0, m, s)$ -nets. Indeed, as the numerical experiments in Section 5 show, it appears the rates obtained with Theorem 3.4 are sometimes valid in the stochastic programming context considered above even though the smoothness condition does not apply.

5. Numerical experiments. To illustrate the ideas set forth in the previous sections, we discuss now some numerical experiments conducted with two small problems available in the literature. The first problem is **APL1P**, a model for electric power capacity expansion on a transportation network that was first described in [19]. The second problem is **LandS**, a modification of a simple problem in electrical investment planning originally presented in [29]. The modified version we study is the one discussed in [26].

APLIP. **APL1P** has two decision variables with 2 constraints (plus lower bound constraints) on the first stage, and 9 decision variables with 5 constraints (plus lower bound constraints) on the second stage. The random variables appear on both the right hand side and the technology matrix of the second stage. There are $s = 5$

independent random variables. The number of realizations per random variable yields a total of $4 \times 5 \times 4 \times 4 \times 4 = 1280$ scenarios. With current computing power, this problem can be easily solved exactly; nevertheless, we present the results with sampling because from that perspective the problem is ill-conditioned (cf. [54]), which means that the approximating solutions \hat{x}_N are likely to vary with replications. That, in turn, ensures that the objective value estimators \hat{v}_N do not correspond to the same solution — if they did, the analysis of rate of convergence would reduce to that of pointwise estimation. Thus, we view this case as a good test for the theoretical results presented in the paper.

We adopted the following methodology. We solved the approximating problem (1.3) using samples generated with standard Monte Carlo, Latin Hypercube Sampling and randomized (t, s) -sequences in base 5 — which, as discussed in Section 3.2, is a form of RQMC. For each sampling scheme, we solved the problem with sample sizes equal to successive powers of the base, ranging from 5^2 to 5^6 . The choice for such sample sizes was driven by two factors: (i) the restriction on the choice of the base in order for a $(0, m, s)$ -net to exist ($b \geq s - 1$, cf. discussion following Theorem 3.4), and (ii) the restriction on sample sizes to be powers of the base in order for a sequence to be a (t, m, s) -net. We feel that the two restrictions together would be rather limiting in practice; thus, we also present results with base 2, which in addition yields faster generators [13].

For each sample size, twenty-five replications were run and the standard deviation of the estimators \hat{v}_N over these replications was calculated. All simulations used independent random streams. By plotting the logarithm of the standard deviation against the logarithm of the sample size we can visualize the rate of convergence — for example, with standard Monte Carlo one expects to obtain a straight line with slope $-1/2$. We also calculated the *mean-squared error* of the estimators, but since the results were very similar (in this problem, bias was much smaller than standard deviation), we chose not to display them.

The sampling approximation problems were solved in two steps: first, we used the SUTIL library [5] to generate the linear programs corresponding to each sampled problem. SUTIL can construct MPS files for Monte Carlo sampling approximations of two-stage stochastic linear programs; we modified the library slightly to incorporate LHS and randomized (t, s) -sequences, using the publicly available routines developed in [13]. The resulting MPS files were fed into the software package Xpress-MPTM from Dash Optimization (available to us under the Academic Partnership Program).

Figure 5.1 shows the results. We can see that both Monte Carlo and LHS yield a convergence rate of $N^{-1/2}$, thus corroborating the results of [48] for Monte Carlo and of Theorem 3.3 for LHS. The rate for RQMC for both bases appears to be of order N^{-1} (although that is more evident with base 5), which is not as good as the rate in Theorem 3.4; a possible explanation is the absence of the smoothness assumed for that result. It is clear from the figure that the rate obtained with RQMC in either case is better than with both Monte Carlo and LHS. Note also that both LHS and RQMC yield estimators with smaller variance than Monte Carlo — even though the rate of convergence (in case of LHS) is the same as that of Monte Carlo — and that the variance with RQMC is smaller than with LHS except for very small sample sizes.

LandS. The LandS problem has 4 decision variables on the first stage, and 12 decision variables on the second stage. Randomness appears only on the right-hand side of the second stage, in the form of demand constraints. There are $s = 3$ independent random variables, each with 100 possible realizations. Thus, the total number

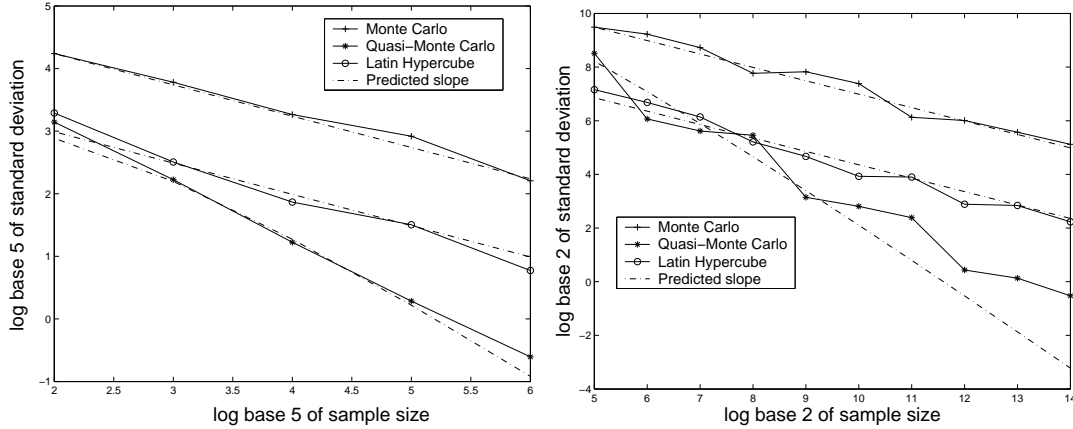


FIG. 5.1. Rates of convergence for APL1P problem using base 5 (left) and base 2 (right).

of scenarios is 10^6 .

The methodology we adopted was the same as in the APL1P case, except that used bases 3 and 2. Figure 5.2 shows the results. Again, we see that both Monte Carlo and LHS yield a convergence rate of $N^{-1/2}$, and the rate for RQMC appears to be of order N^{-1} . As in the previous example, the RQMC rate is better than the Monte Carlo and LHS rates, both LHS and RQMC yield estimators with smaller variance than Monte Carlo, and the variance with RQMC is smaller than with LHS.

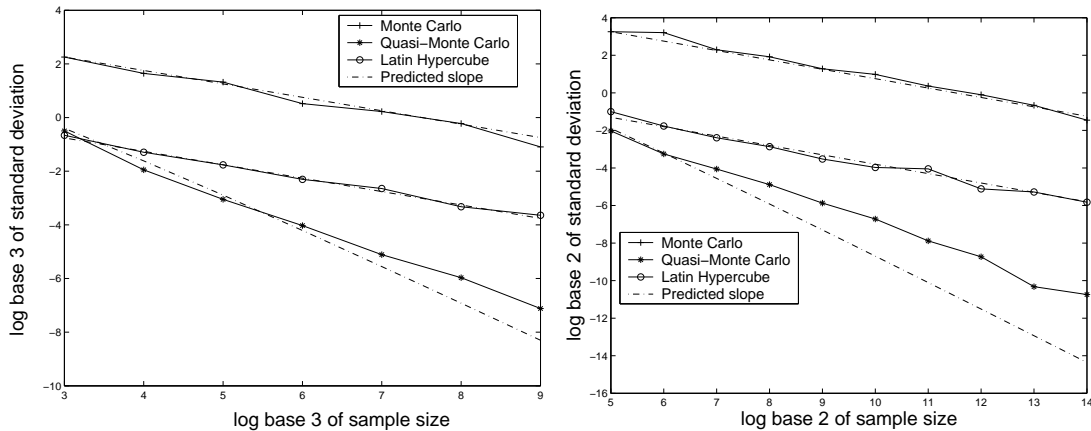


FIG. 5.2. Rates of convergence for LandS problem using base 3 (left) and base 2 (right).

A brief study of smoothness. To check the role of the lack of smoothness in the convergence rates, we considered the effect of smoothing the inverse cdfs F_i^{-1} . This was accomplished by replacing F_i^{-1} with a smooth function F_i^Δ such that F_i^{-1} and F_i^Δ coincide everywhere except on a interval of size 2Δ around each discontinuity point.

Figures 5.3 and 5.4 depict the results. In general, such a smoothing procedure may introduce bias; this is clearly seen in case of the APL1P problem, where smoothing does

not seem to help much, at least for base 2. With **LandS**, however, smoothing works perfectly — with $\Delta = 0.005$ we obtain the rate predicted by Theorem 3.4 without incurring virtually any bias, even though the theorem is not directly applicable in the absence of Assumption A3 (the smoothed distribution does not have finite support).

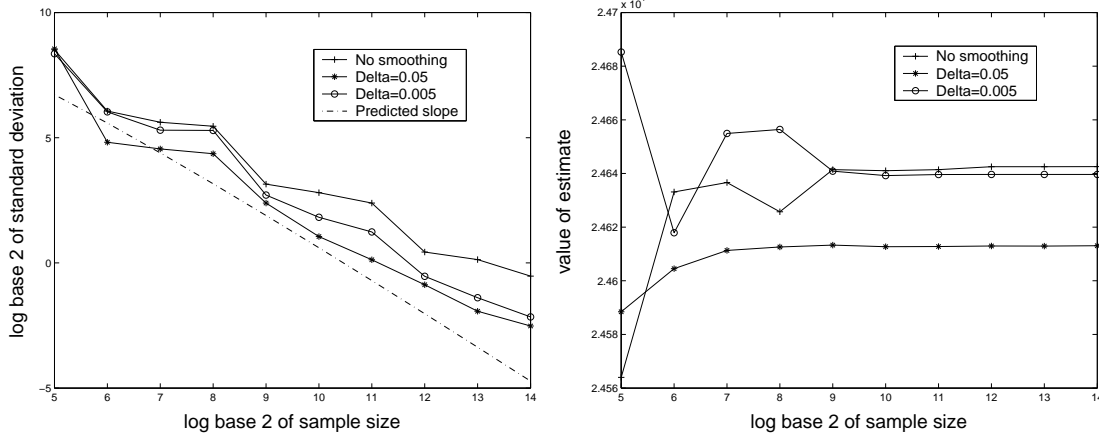


FIG. 5.3. Rates of convergence (left) and values of the estimate \hat{v}_N (right) for **APL1P** problem under RQMC with smoothing, using base 2.

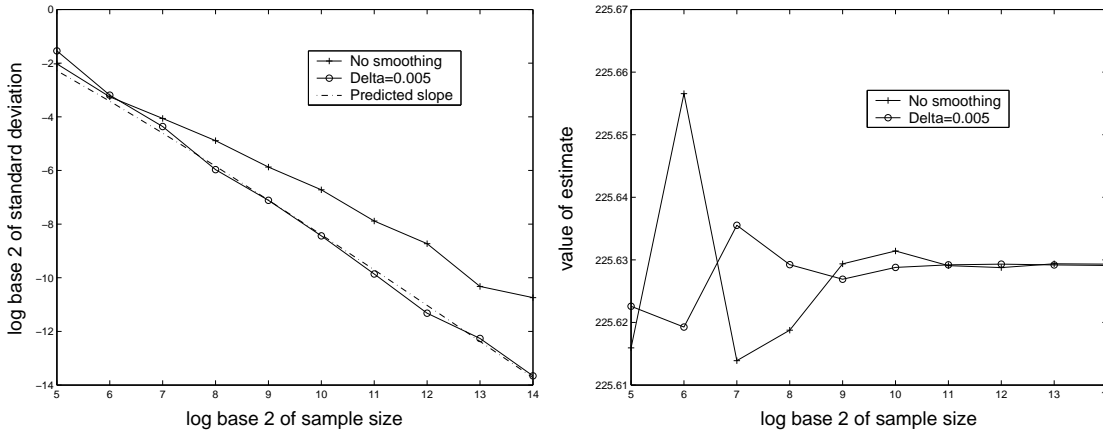


FIG. 5.4. Rates of convergence (left) and values of the estimate \hat{v}_N (right) for **LandS** problem under RQMC with smoothing, using base 2.

As another way to verify the effect of smoothness, we modified the **APL1P** problem by fitting continuous distributions to the discrete data. More precisely, we used Weibull distributions for each of the five random variables in the problem, in such a way that the mean and variance of the random variables were approximately the same as in the original data. Figure 5.5 depicts the results, again for bases 2 and 5. We see that the rates of convergence behave somewhat similarly to the discrete case, although the benefits of using RQMC seem slightly higher in the continuous case.

These results suggest that Theorem 3.4 may be valid under more general conditions than those we have used.

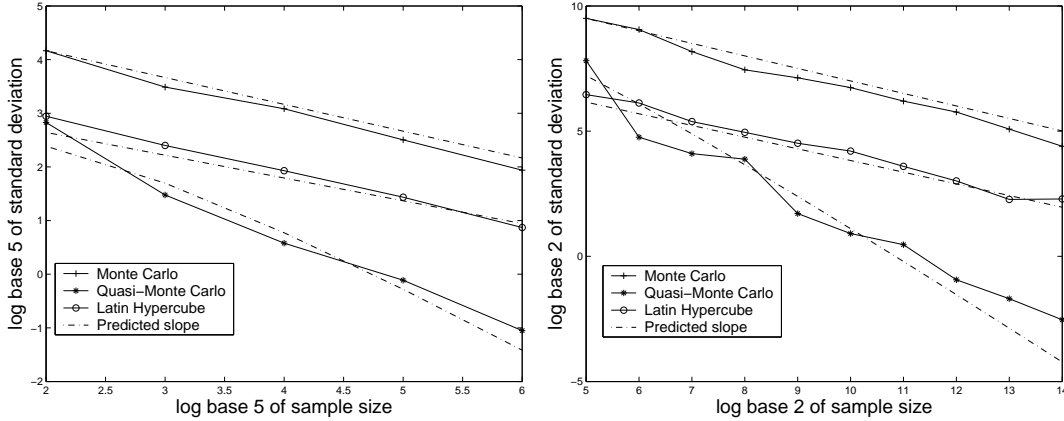


FIG. 5.5. Rates of convergence for modified APLIP problem with Weibull distributions, using base 5 (left) and base 2 (right).

6. Conclusions. The theoretical and numerical results in this paper suggest that alternative sampling methods such as Latin Hypercube Sampling and Quasi-Monte Carlo can be very effective when solving stochastic optimization problems via sample average approximations. The effectiveness is measured in terms of rates of convergence of estimators of optimal solutions and of optimal values as functions of the sample size. The main contribution of the paper is establishing that rates of convergence for pointwise estimators (i.e., estimation of integrals) carry over to estimators of optimal values/solutions, which allows for the use of results for pointwise estimation available in the literature. In particular, the results in the paper show that, under appropriate conditions, it is possible to obtain a rate of convergence of order $[(\log N)^{s-1}/N^3]^{1/2}$ for the approximating optimal values $\hat{\nu}_N$, which asymptotically is much better than the $N^{-1/2}$ obtained with standard Monte Carlo.

Such results are very encouraging, and at the same time raise some interesting issues for further investigation. One topic concerns the effect of smoothing on the rates of convergence when using RQMC — as discussed earlier, the “ideal” rate $[(\log N)^{s-1}/N^3]^{1/2}$ derived in Theorem 3.4 seems to require smoothness of the inverse cdf and of the objective function. However, our numerical results suggest such conditions may not be necessary. Moreover, it is important to mention that Theorem 2.10 is valid regardless of any smoothness conditions. That is, if one shows that a Central Limit Theorem holds for RQMC under nonsmooth (or potentially discontinuous) functions with a certain rate, then Theorem 2.10 will ensure that under appropriate conditions the optimal value estimators $\hat{\nu}_N^{\text{RQMC}}$ converge at the same rate. The result in [28] used in the proof of Theorem 3.4 is, however, the only CLT-type result available for RQMC, at least to the best of our knowledge.

On the other hand, our experiments also suggest that Theorem 3.4 is valid even when Assumption A3 does not hold. This is not surprising — as we mentioned earlier, it is possible that a functional version of Assumption B2 holds for the functional space $C(X)$ under RQMC, in which case the conditions of Assumption A3 would not be required; however, we are not aware of the existence of such a result.

It would also be interesting to study the precise effect of having multiple optimal solutions on the rates of convergence of optimal values — the main results we have obtained for that case under LHS and RQMC (Theorems 3.3 and 3.4) require uniqueness

of the optimal solution. Such a task, however, is likely to require again a functional or at least multivariate version of Assumption B2 (we note that multivariate CLTs have been proved for LHS, but not for RQMC).

Another important issue concerns the dimensionality of the problems. It is well known that the performance of RQMC methods worsens with the number of dimensions — indeed, it is easy to see that, when s is large, the term $[(\log_b N)^{s-1}/N^3]^{1/2}$ only becomes smaller than $N^{-1/2}$ for large N . For example, with $s = 30$ and $b = 2$ one needs $N \geq 2^{16}$ to get the benefits of the RQMC approach. This suggests that RQMC sampling should be used only with some of the random variables involved in the problem; however, determining which ones to select is a nontrivial issue. Research on this topic is underway.

Acknowledgments

We thank Shane Drew for help with the numerical experiments, Jeff Linderoth for assistance with the SUTIL library, and Mihai Anitescu and Alexander Shapiro for discussions on an early version of this paper. We also thank two anonymous referees and the associate editor for their comments.

REFERENCES

- [1] T. G. BAILEY, P. JENSEN, AND D. MORTON, *Response surface analysis of two-stage stochastic linear programming with recourse*, Naval Research Logistics, 46 (1999), pp. 753–778.
- [2] R. BARTLE, *The Elements of Real Analysis*, Wiley, New York, 2nd. ed., 1987.
- [3] P. BILLINGSLEY, *Convergence of Probability Measures*, Wiley, New York, 2nd. ed., 1999.
- [4] J. R. BIRGE AND F. LOUVEAUX, *Introduction to Stochastic Programming*, Springer Series in Operations Research, Springer-Verlag, New York, NY, 1997.
- [5] J. CZYZYK, J. LINDEROTH, AND J. SHEN, *SUTIL: A utility library for handling stochastic programs*, 2005. User’s Manual. Software available at <http://coral.ie.lehigh.edu/sutil>.
- [6] L. DAI, C. H. CHEN, AND J. R. BIRGE, *Convergence properties of two-stage stochastic programming*, J. Optim. Theory Appl., 106 (2000), pp. 489–509.
- [7] G. B. DANTZIG AND P. W. GLYNN, *Parallel processors for planning under uncertainty*, Annals of Operations Research, 22 (1990), pp. 1–21.
- [8] A. DEMBO AND O. ZEITOUNI, *Large Deviations Techniques and Applications*, Springer-Verlag, New York, NY, 2nd. ed., 1998.
- [9] S. S. DREW AND T. HOMEM-DE-MELLO, *Some large deviations results for Latin Hypercube Sampling*. Manuscript, Department of Industrial Engineering and Management Sciences, Northwestern University. Submitted to *Probability and Engineering in the Informational Sciences*, 2007.
- [10] J. DUPAČOVÁ AND R. J.-B. WETS, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, The Annals of Statistics, 16 (1988), pp. 1517–1549.
- [11] G. FISHMAN, *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag, New York, NY, 1997.
- [12] B. L. FOX, *Strategies for Quasi-Monte Carlo*, Kluwer Academic Publishers, Norwell, MA, 2000.
- [13] I. FRIEDEL AND A. KELLER, *Fast generation of randomized low-discrepancy point sets*, in Monte Carlo and quasi-Monte Carlo methods, 2000 (Hong Kong), Springer, Berlin, 2002, pp. 257–273. Software available at <http://www.multires.caltech.edu/software/libseq/>.
- [14] G. GÜRKAN, A. Y. ÖZGE, AND S. M. ROBINSON, *Sample-path solutions of stochastic variational inequalities*, Mathematical Programming, 84 (1999), pp. 313–334.
- [15] H. HEITSCH AND W. RÖMISCH, *Scenario reduction algorithms in stochastic programming*, Computational Optimization and Applications, 24 (2003), pp. 187–206.
- [16] J. L. HIGLE, *Variance reduction and objective function evaluation in stochastic linear programs*, INFORMS Journal on Computing, 10 (1998), pp. 236–247.
- [17] J. L. HIGLE AND S. SEN, *Stochastic decomposition: An algorithm for two stage linear programs with recourse*, Mathematics of Operations Research, 16 (1991), pp. 650–669.
- [18] J. B. HIRIART-URRUTY AND C. LEMARECHAL, *Convex Analysis and Minimization Algorithms*, vol. I, Springer-Verlag, Berlin, Germany, 1993.

- [19] G. INFANGER, *Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs*, Annals of Operations Research, 39 (1992), pp. 69–95.
- [20] J. KALAGNANAM AND U. DIWEKAR, *An efficient sampling technique for off-line quality control*, Technometrics, 39 (1997), pp. 308–319.
- [21] Y. M. KANIOVSKI, A. J. KING, AND R. J.-B. WETS, *Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems*, Ann. Oper. Res., 56 (1995), pp. 189–208.
- [22] A. J. KING AND R. T. ROCKAFELLAR, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 148–162.
- [23] A. KLEYWEGT, A. SHAPIRO, AND T. HOMEM-DE-MELLO, *The sample average approximation method for stochastic discrete optimization*, SIAM Journal on Optimization, 12 (2001), pp. 479–502.
- [24] M. KOIVU, *Variance reduction in sample approximations of stochastic programs*, Mathematical Programming, 103 (2005), pp. 463–485.
- [25] A. M. LAW AND W. D. KELTON, *Simulation Modeling and Analysis*, McGraw-Hill, New York, NY, 3rd. ed., 2000.
- [26] J. T. LINDEROTH, A. SHAPIRO, AND S. J. WRIGHT, *The empirical behavior of sampling methods for stochastic programming*, Annals of Operations Research, (2005). forthcoming.
- [27] W. LOH, *On Latin hypercube sampling*, The Annals of Statistics, 24 (1996), pp. 2058–2080.
- [28] W.-L. LOH, *On the asymptotic distribution of scrambled net quadrature*, Ann. Statist., 31 (2003), pp. 1282–1324.
- [29] F. LOUVEAUX AND Y. SMEERS, *Optimal investments for electricity generation: A stochastic model and a test problem*, in Numerical techniques for stochastic optimization problems, Y. Ermoliev and R. J.-B. Wets, eds., Springer-Verlag, Berlin, 1988, pp. 445–452.
- [30] R. LUCCHETTI, G. SALINETTI, AND R. WETS, *Uniform convergence of probability measures: topological criteria*, Journal of Multivariate Analysis, 51 (1994), pp. 252–264.
- [31] M. D. MCKAY, R. J. BECKMAN, AND W. J. CONOVER, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979), pp. 239–245.
- [32] H. NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, PA, 1992.
- [33] A. B. OWEN, *A central limit theorem for Latin hypercube sampling*, J. Roy. Statist. Soc. Ser. B, 54 (1992), pp. 541–551.
- [34] A. B. OWEN, *Randomly permuted (t, m, s) -nets and (t, s) -sequences*, in Monte Carlo and quasi-Monte Carlo methods in scientific computing (Las Vegas, NV, 1994), vol. 106 of Lecture Notes in Statist., Springer, New York, 1995, pp. 299–317.
- [35] ———, *Monte Carlo variance of scrambled net quadrature*, SIAM J. Numer. Anal., 34 (1997), pp. 1884–1910.
- [36] ———, *Scrambled net variance for integrals of smooth functions*, Ann. Statist., 25 (1997), pp. 1541–1562.
- [37] A. B. OWEN, *Latin supercube sampling for very high-dimensional simulations*, ACM Transactions on Modeling and Computer Simulation, 8 (1998), pp. 71–102.
- [38] A. B. OWEN, *Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo*, in Monte Carlo and quasi-Monte Carlo methods 1998 (Claremont, CA), Springer, Berlin, 2000, pp. 86–97.
- [39] T. PENNANEN, *Epi-convergent discretizations of multistage stochastic programs*, Mathematics of Operations Research, 30 (2005), pp. 245–256.
- [40] T. PENNANEN AND M. KOIVU, *Epi-convergent discretizations of stochastic programs via integration quadratures*, Numerische Mathematik, 100 (2005), pp. 141–163.
- [41] G. C. PFLUG, *Scenario tree generation for multiperiod financial optimization by optimal discretization*, Mathematical Programming, Series B, 89 (2001), pp. 251–271.
- [42] ———, *Scenario estimation and generation*. Tutorial presented at the X Stochastic Programming Conference, Tucson, AZ, 2004.
- [43] E. L. PLAMBECK, B. R. FU, S. M. ROBINSON, AND R. SURI, *Sample-path optimization of convex stochastic performance functions*, Mathematical Programming, Series B, 75 (1996), pp. 137–176.
- [44] S. M. ROBINSON, *Analysis of sample-path optimization*, Mathematics of Operations Research, 21 (1996), pp. 513–528.
- [45] W. RÖMISCH, *Stability of stochastic programming problems*, in Handbook of Stochastic Optimization, A. Ruszczyński and A. Shapiro, eds., Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2003.
- [46] R. Y. RUBINSTEIN AND A. SHAPIRO, *Discrete Event Systems: Sensitivity Analysis and Stochas-*

- tic Optimization by the Score Function Method*, John Wiley & Sons, Chichester, England, 1993.
- [47] L. RÜSCHENDORF AND V. DE VALK, *On regression representations of stochastic processes*, Stochastic Processes and Applications, 46 (1993), pp. 183–198.
 - [48] A. SHAPIRO, *Asymptotic analysis of stochastic programs*, Annals of Operations Research, 30 (1991), pp. 169–186.
 - [49] ———, *Asymptotic behavior of optimal solutions in stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 829–845.
 - [50] ———, *Monte Carlo sampling methods*, in Handbook of Stochastic Optimization, A. Ruszczyński and A. Shapiro, eds., Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2003.
 - [51] A. SHAPIRO, *On complexity of multistage stochastic programs*, Operations Research Letters, 34 (2006), pp. 1–8.
 - [52] A. SHAPIRO AND T. HOMEM-DE-MELLO, *A simulation-based approach to two-stage stochastic programming with recourse*, Mathematical Programming, 81 (1998), pp. 301–325.
 - [53] ———, *On rate of convergence of Monte Carlo approximations of stochastic programs*, SIAM Journal on Optimization, 11 (2000), pp. 70–86.
 - [54] A. SHAPIRO, T. HOMEM-DE-MELLO, AND J. C. KIM, *Conditioning of convex piecewise linear stochastic programs*, Mathematical Programming, 94 (2002), pp. 1–19.
 - [55] M. L. STEIN, *Large sample properties of simulations using Latin hypercube sampling*, Technometrics, 29 (1987), pp. 143–151.