ON RATE OF CONVERGENCE OF OPTIMAL SOLUTIONS OF MONTE CARLO APPROXIMATIONS OF STOCHASTIC PROGRAMS

ALEXANDER SHAPIRO* AND TITO HOMEM-DE-MELLO†

Abstract. In this paper we discuss Monte Carlo simulation based approximations of a stochastic programming problem. We show that if the corresponding random functions are convex piecewise smooth and the distribution is discrete, then (under mild additional assumptions) an optimal solution of the approximating problem provides an *exact* optimal solution of the true problem with probability one for sufficiently large sample size. Moreover, by using theory of Large Deviations, we show that the probability of such an event approaches one exponentially fast with increase of the sample size. In particular, this happens in the case of two stage stochastic programming with recourse if the corresponding distributions are discrete. The obtained results suggest that, in such cases, Monte Carlo simulation based methods could be very efficient. We present some numerical examples to illustrate the involved ideas.

Key words. Two-stage stochastic programming with recourse, Monte Carlo simulation, Large Deviations theory, convex analysis

AMS subject classifications. 90C15, 90C25

1. Introduction.

We discuss in this paper Monte Carlo approximations of stochastic programming problems of the form

$$\operatorname*{Min}_{x\in\Theta}\{f(x) := I\!E_P h(x,\omega)\}, \tag{1.1}$$

where P is a probability measure on a sample space $(\Omega,\mathcal{F})$, Θ is a subset of $I\!R^m$ and $h : I\!R^m \times \Omega \to I\!R$ is a real valued function. We refer to the above problem as the "true" optimization problem. By generating an independent identically distributed (i.i.d.) random sample $\omega^1,...,\omega^N$ in $(\Omega,\mathcal{F})$, according to the distribution P, one can construct the corresponding approximating program

$$\operatorname*{Min}_{x\in\Theta}\left\{\hat{f}_N(x) := N^{-1}\sum_{j=1}^{N} h(x,\omega^j)\right\}. \tag{1.2}$$

An optimal solution $\hat{x}_N$ of (1.2) provides an approximation (an estimator) of an optimal solution of the true problem (1.1).

There are numerous publications where various aspects of convergence properties of $\hat{x}_N$ are discussed. Suppose that the true problem has a non empty set A of optimal solutions. It is possible to show that, under mild regularity conditions, the distance $\operatorname{dist}(\hat{x}_N, A)$, from $\hat{x}_N$ to the set A, converges with probability one (w.p.1) to zero as $N \to \infty$. There is a vast literature in Statistics dealing with such consistency properties of empirical estimators. In the context of stochastic programming we can mention recent works [9],[14],[17], where this problem is approached from the point of view of the epiconvergence theory.

* School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA. Email: ashapiro@isye.gatech.edu. This work was supported, in part, by grant DMI-9713878 from the National Science Foundation.

† Department of Industrial, Welding and Systems Engineering, The Ohio State University, Columbus, Ohio 43210-1271, USA. Email:homem-de-mello.1@osu.edu

It is also possible to give various estimates of the rate of convergence of $\hat{x}_N$ to A. Central Limit Theorem type results give such estimates of order $O_p(N^{-1/2})$ for the distance $\mathrm{dist}(\hat{x}_N, A)$ (e.g., [15], [20]), and the Large Deviations theory shows that one may expect that, for any given $\varepsilon > 0$, the probability of the event $\mathrm{dist}(\hat{x}_N, A) \geq \varepsilon$ approaches zero exponentially fast as $N \to \infty$ (see, e.g., [13],[16],[19]). These are general results and it seems that they describe the situation quite accurately in case the involved distributions are continuous. However, it appears that the asymptotics are completely different if the distributions are *discrete*. We show that in such cases, under rather natural assumptions, the approximating problem (1.2) provides an *exact* optimal solution of the true problem (1.1) for N large enough. That is, $\hat{x}_N \in A$ w.p.1 for sufficiently large N. Even more surprisingly we show that the probability of the event $\{\hat{x}_N \notin A\}$ tends to zero exponentially fast as $N \to \infty$. That is what happens in the case of two stage stochastic programming with recourse if the corresponding distributions are discrete. This indicates that, in such cases, Monte Carlo simulation based methods could be very efficient.

In order to motivate the discussion, let us consider the following simple example. Let $Y_1, ..., Y_m$ be independent identically distributed real valued random variables. Consider the following optimization problem

$$\min_{x \in \mathbb{R}^m} \left\{ f(x) := \mathbb{E}\left(\sum_{i=1}^{m} |Y_i - x_i| \right) \right\}. \tag{1.3}$$

This problem is a particular case of two stage stochastic programming with simple recourse. Clearly the objective function $f(x)$ can be written in the form $f(x) := \sum_{i=1}^{m} f_i(x_i)$, where $f_i(x_i) := \mathbb{E}\{|Y_i - x_i|\}$. Therefore the above optimization problem is separable. It is well known that a minimizer of $f_i(\cdot)$ is given by the median of the distribution of Y_i. Suppose that the distribution of the random variables Y_i is symmetrical around zero. Then $\bar{x} := (0, ..., 0)$ is an optimal solution of (1.3).

Now let $Y^1, ..., Y^N$ be an i.i.d. random sample of N realizations of the random vector $Y = (Y_1, ..., Y_m)$. Consider the following sample average approximation of (1.3)

$$\min_{x \in \mathbb{R}^m} \left\{ \hat{f}_N(x) := N^{-1} \sum_{j=1}^{N} h(x, Y^j) \right\}, \tag{1.4}$$

where $h(x, y) := \sum_{i=1}^{m} |y_i - x_i|$, with $x, y \in \mathbb{R}^m$. An optimal solution of the above approximating problem (1.4) is given by $\hat{x}_N := (\hat{x}_{1N}, ..., \hat{x}_{mN})$, where $\hat{x}_{iN}$ is the sample median of $Y_i^1, ..., Y_i^N$.

Suppose for the moment that $m = 1$, i.e. we are minimizing $\mathbb{E}\{|Y - x|\}$ over $x \in \mathbb{R}$. We assume that the distribution of Y is symmetrical around zero and hence $\bar{x} = 0$ is an optimal solution of the true problem. Suppose now that the distribution of Y is continuous with density function $g(y)$. Then it is well known (e.g., [6]) that the corresponding sample median $\hat{x}_N$ is asymptotically normal. That is, $N^{1/2}(\hat{x}_N - \bar{x})$ converges in distribution to normal with zero mean and variance $[2g(\bar{x})]^{-2}$. For example, if Y is uniformly distributed on the interval $[-1, 1]$, then $N^{1/2}(\hat{x}_N - \bar{x}) \Rightarrow N(0, 1)$. This means that for $N = 100$ we may expect $\hat{x}_N$ to be in the (so-called confidence) interval $[-0.2, 0.2]$ with probability of about 95%. Now for $m > 1$ we have that the events $\hat{x}_{iN} \in [-0.2, 0.2]$, $i = 1, ..., m$, are independent (this is because we assume that Y_i are independent). Therefore the probability that *each* sample

median $\hat{x}_{iN}$ will be inside the interval $[-0.2, 0.2]$ is about 0.95^m. For example, for $m = 50$, this probability becomes $0.95^{50} = 0.077$. If we want that probability to be about 0.95 we have to increase the interval to $[-0.3, 0.3]$, which constitutes 30% of the range of the random variable Y. In other words for that sample size and with $m = 50$ our sample estimate will be not accurate.

The situation becomes quite different if we assume that Y has a discrete distribution. Suppose now that Y can take values -1, 0 and 1 with equal probabilities 1/3. In that case the true problem has unique optimal solution $\bar{x} = 0$. The corresponding sample estimate $\hat{x}_N$ can be equal to -1, 0 or 1. We have that the event $\{\hat{x}_N = 1\}$ happens if more than half of the sample points are equal to one. Probability of that is given by $P(X > N/2)$, where X has a binomial distribution $B(N, 1/3)$. If exactly half of the sample points are equal to one, then the sample estimate can be any number in the interval $[0, 1]$. Similar conclusions hold for the event $\{\hat{x}_N = -1\}$. Therefore the probability that $\hat{x}_N = 0$ is at least $1 - 2P(X \geq N/2)$. For $N = 100$, this probability is 0.9992. Therefore the probability that the sample estimate $\hat{x}_N$, given by an optimal solution of the approximating problem (1.4) with the sample size $N = 100$ and the number of random variables $m = 50$, is at least $0.9992^{50} = 0.96$. With the sample size $N = 120$ and the number of random variables $m = 200$ this probability, of $\hat{x}_N = 0$, is about $0.9998^{200} = 0.95$. Note that the number of scenarios for that problem is 3^{200}, which is not small by any standard. And yet with sample size of only 120 the approximating problem produces an estimator which is exactly equal to the true optimal solution with probability of 95%.

The above problem, although simple, illustrates the phenomenon of exponential convergence referred to in the title of the paper. In the above example the corresponding probabilities can be calculated in a closed form, but in the general case of course we cannot expect to do so. The purpose of this paper is to extend this discussion to a class of stochastic programming problems satisfying some assumptions. Our goal is to exhibit some *qualitative* (rather than quantitative) results. We do not propose an algorithm, but rather show asymptotic properties of Monte Carlo simulation based methods.

The paper is organized as follows. In section 2 we show almost sure (w.p.1) occurrence of the event $\{\hat{x}_N \in A\}$ (recall that A is the set of optimal solutions of the "true" problem). In section 3 we take a step further and, using techniques from Large Deviations theory, we show that the probability of that event approaches one exponentially fast. In section 4 we discuss the median problem in more detail, and present some numerical results for a two-stage stochastic programming problem with complete recourse. Finally, section 5 presents some conclusions.

2. Almost sure convergence. Consider the "true" stochastic programming problem (1.1). For the sake of simplicity we assume that the corresponding expected value function $f(x) := \mathbb{E}_P h(x, \omega)$ exists (and in particular is finite valued) for all $x \in \mathbb{R}^m$. For example, if the probability measure P has a finite support (i.e. the distribution P is discrete and can take a finite number of different values), and hence the space Ω can be taken to be finite, say $\Omega := \{\omega_1, ..., \omega_K\}$, and P is given by the probabilities $P\{\omega = \omega_k\} = p_k$, $k = 1, ..., K$, we have

$$\mathbb{E}_P h(x, \omega) = \sum_{k=1}^{K} p_k h(x, \omega_k). \tag{2.1}$$

We assume that the feasible set Θ is closed and convex, and that for every $\omega \in \Omega$, the function $h(\cdot, \omega)$ is convex. This implies that the expected value function $f(\cdot)$ is also

convex, and hence the "true" problem (1.1) is convex. Also if P is discrete and the functions $h(\cdot,\omega_k)$, $k=1,...,K$, are piecewise linear and convex, then $f(\cdot)$ is piecewise linear and convex. That is what happens in two stage stochastic programming with a finite number of scenarios.

Let $\omega^1,...,\omega^N$ be an i.i.d. random sample in $(\Omega,\mathcal{F})$, generated according to the distribution P, and consider the corresponding approximating program (1.2). Note that, since the functions $h(\cdot,\omega^j)$ are convex, the approximating (sample average) function $\hat{f}_N(\cdot)$ is also convex, and hence the approximating program (1.2) is convex.

We show in this section that, under some natural assumptions which hold for instance in the case of two stage stochastic programming with a finite number of scenarios, with probability one (w.p.1) for N large enough any optimal solution of the approximating problem (1.2) belongs to the set of optimal solutions of the true problem (1.1). That is, problem (1.2) yields an *exact* optimal solution (w.p.1) when N is sufficiently large.

The statement: "w.p.1 for N large enough" should be understood in the sense that for P-almost every $\omega\in\Omega$ there exists $N^*=N^*(\omega)$, such that for any $N\geq N^*$ the corresponding statement holds. The number N^* is a function of ω, i.e. depends on the random sample, and therefore in itself is random. Note also that, since convergence w.p.1 implies convergence in probability, the above statement implies that the probability of the corresponding event to happen tends to one as the sample size N tends to infinity.

We denote by A the set of optimal solutions of the true problem (1.1), and by $f'(x,d)$ the directional derivative of f at x in the direction d. Note that the set A is convex and closed, and since f is a real valued convex function, the directional derivative $f'(x,d)$ exists, for all x and d, and is convex in d. We discuss initially the case when A is a singleton; later we will consider the general setting.

Assumption (A) The true problem (1.1) possesses unique optimal solution $\bar{x}$, i.e. $A=\{\bar{x}\}$, and there exists a positive constant c such that

$$f(x)\geq f(\bar{x})+c\|x-\bar{x}\|,\quad \forall\, x\in\Theta. \tag{2.2}$$

Of course condition (2.2), in itself, implies that $\bar{x}$ is the unique optimal solution of (1.1). In the approximation theory optimal solutions satisfying (2.2) are called sharp minima. It is not difficult to show, since problem (1.1) is convex, that assumption (A) holds iff

$$f'(\bar{x},d)>0,\quad \forall\, d\in T_\Theta(\bar{x})\setminus\{0\}, \tag{2.3}$$

where $T_\Theta(\bar{x})$ denotes the tangent cone to Θ at $\bar{x}$. In particular, if $f(x)$ is differentiable at $\bar{x}$, then assumption (A) (or equivalently (2.3)) holds iff $-\nabla f(\bar{x})$ belongs to the interior of the normal cone to Θ at $\bar{x}$. Note, that since $f'(\bar{x},\cdot)$ is a positively homogeneous convex real valued (and hence continuous) function, it follows from (2.3) that $f'(\bar{x},d)\geq\varepsilon\|d\|$ for some $\varepsilon>0$ and all $d\in T_\Theta(\bar{x})$. We refer to a recent paper [4], and references therein, for a discussion of that condition and some of its generalizations.

If the function $f(x)$ is piecewise linear and the set Θ is polyhedral, then problem (1.1) can be formulated as a linear programming problem, and the above assumption (A) always holds provided $\bar{x}$ is the unique optimal solution of (1.1). This happens, for example, in the case of a two stage linear stochastic programming problem with a finite number of scenarios provided it has a unique optimal solution. Note that assumption (A) is not restricted to such situations only. In fact, in some of our numerical experiments sharp minima (i.e. assumption (A)) happened quite often in

the case of continuous (normal) distributions. Furthermore, because the problem is assumed to be convex, sharp minima is equivalent to first order sufficient conditions. Under such conditions, first order (i.e. linear) growth (2.2) of $f(x)$ holds *globally*, i.e. for all $x \in \Theta$.

THEOREM 2.1. *Suppose that:* (i) *for every* $\omega \in \Omega$ *the function* $h(\cdot, \omega)$ *is convex,* (ii) *the expected value function* $f(\cdot)$ *is well defined and is finite valued,* (iii) *the set* Θ *is closed and convex,* (iv) *assumption (A) holds. Then w.p.1 for* N *large enough the approximating problem* (1.2) *has a unique optimal solution* $\hat{x}_N$ *and* $\hat{x}_N = \bar{x}$.

Proof of the above theorem is based on the following proposition. Results of that proposition (perhaps not exactly in that form) are basically known, but since its proof is simple we give it for the sake of completeness. Denote by $h'_\omega(x, d)$ the directional derivative of $h(\cdot, \omega)$ at the point x in the direction d, and by $\mathcal{H}(B, C)$ the Hausdorff distance between sets $B, C \subset \mathbb{R}^m$, that is

$$\mathcal{H}(B, C) := \max \left\{ \sup_{x \in C} \operatorname{dist}(x, B), \sup_{x \in B} \operatorname{dist}(x, C) \right\}. \tag{2.4}$$

PROPOSITION 2.2. *Suppose that the assumptions* (i) *and* (ii), *of Theorem* 2.1, *are satisfied. Then, for any* $x, d \in \mathbb{R}^m$, *the following holds:*

$$f'(x, d) = \mathbb{E}_P \{ h'_\omega(x, d) \}, \tag{2.5}$$

$$\lim_{N \to \infty} \sup_{\|d\| \leq 1} \left| f'(x, d) - \hat{f}'_N(x, d) \right| = 0, \quad w.p.1, \tag{2.6}$$

$$\lim_{N \to \infty} \mathcal{H} \left(\partial \hat{f}_N(x), \partial f(x) \right) = 0, \quad w.p.1. \tag{2.7}$$

Proof. Since $f(\cdot)$ is convex we have that

$$f'(x, d) = \inf_{t > 0} \frac{f(x + td) - f(x)}{t}, \tag{2.8}$$

and the ratio in the right hand side of (2.8) decreases monotonically as t decreases to zero, and similarly for the functions $h(\cdot, \omega)$. It follows then by the Monotone Convergence Theorem that

$$f'(x, d) = \mathbb{E}_P \left\{ \inf_{t > 0} \frac{h(x + td, \omega) - h(x, \omega)}{t} \right\}, \tag{2.9}$$

and hence the right hand side of (2.5) is well defined and the equation follows.

We have that

$$\hat{f}'_N(x, d) = N^{-1} \sum_{j=1}^{N} h'_{\omega^j}(x, d). \tag{2.10}$$

Therefore by the strong form of the Law of Large Numbers it follows from (2.5) that for any $d \in \mathbb{R}^m$, $\hat{f}'_N(x, d)$ converges to $f'(x, d)$ w.p.1 as $N \to \infty$. Consequently for any countable set $D \subset \mathbb{R}^m$ we have that the event: "$\lim_{N \to \infty} \hat{f}'_N(x, d) = f'(x, d)$ for all $d \in D$" happens w.p.1. Let us take a countable and dense subset D of $\mathbb{R}^m$. Recall that if a sequence of real valued convex functions converges pointwise on a

dense subset of $I\!R^m$, then it converges uniformly on any compact subset of $I\!R^m$ (e.g., [18, Theorem 10.8]). Therefore, since the functions $\hat{f}'_N(x,\cdot)$ are convex, it follows from the pointwise convergence of $\hat{f}'_N(x,\cdot)$ on D, that the convergence is uniform on the unit ball $\{d : \|d\| \leq 1\}$. This proves (2.6).

Recall that if g is a real valued convex function, then $g'(x,\cdot)$ coincides with the support function of its subdifferential $\partial g(x)$. Therefore the Hausdorff distance between the subdifferentials of f and $\hat{f}_N$, at x, is equal to the supremum on the left hand side of (2.6) (see, e.g., [12, Theorem V.3.3.8]). Consequently (2.7) follows from (2.6). ∎

Proof of Theorem 2.1 As we discussed earlier, assumption (A) is equivalent to condition (2.3) which, in turn, implies that $f'(\bar{x},d) \geq \varepsilon$ for some $\varepsilon > 0$ and all $d \in T_\Theta(\bar{x}) \cap S^{m-1}$, where

$$S^{m-1} := \{d \in I\!R^m : \|d\| = 1\}.$$

By (2.6) it follows that w.p.1 for N large enough

$$\hat{f}'_N(\bar{x},d) > 0, \quad \forall\, d \in T_\Theta(\bar{x}) \cap S^{m-1}. \tag{2.11}$$

Since the approximating problem is convex, this implies that $\bar{x}$ is a sharp (and hence unique) optimal solution of the approximating problem. This completes the proof. ∎

Let us consider now a situation where the true problem (1.1) may have multiple optimal solutions, i.e. the set A is not necessarily a singleton. In that case Theorem 2.1 can be generalized, under stronger assumptions, as follows.

THEOREM 2.3. *Suppose that:* (i) *the set Ω is finite,* (ii) *for every $\omega \in \Omega$ the function $h(\cdot,\omega)$ is piecewise linear and convex,* (iii) *the set Θ is closed, convex and polyhedral,* (iv) *the true problem* (1.1) *has a non empty bounded set A of optimal solutions. Then the set A is compact convex and polyhedral, and w.p.1 for N large enough the approximating problem* (1.2) *has a non empty set A_N of optimal solutions and A_N is a face of the set A.*

Proof of the above theorem is based on the following lemma which may have an independent interest.

LEMMA 2.4. *Suppose that the assumptions* (i) *and* (ii)*, of Theorem* 2.3*, are satisfied. Then the following holds.* (a) *There exists a finite number of points $z_1,...,z_r$ (independent of the sample) such that for every $x \in I\!R^m$, there is $k \in \{1,...,r\}$ such that $\partial f(x) = \partial f(z_k)$ and $\partial \hat{f}_N(x) = \partial \hat{f}_N(z_k)$ for any realization of the random sample.* (b) *With probability one the subdifferentials $\partial \hat{f}_N(x)$ converge to $\partial f(x)$ uniformly in $x \in I\!R^m$, i.e.*

$$\lim_{N\to\infty} \sup_{x\in I\!R^m} \mathcal{H}\left(\partial \hat{f}_N(x), \partial f(x)\right) = 0, \quad w.p.1. \tag{2.12}$$

(c) *If, in addition, the assumptions* (iii) *and* (iv) *are satisfied, then there exists a finite number of points $x_1,...,x_q$ (independent of the sample) such that the points $x_1,...,x_\ell$, $\ell < q$, form the set of extreme points of A and if the following condition holds*

$$\hat{f}_N(x_i) < \hat{f}_N(x_j) \text{ for any } i \in \{1,...,\ell\} \text{ and } j \in \{\ell+1,...,q\}, \tag{2.13}$$

then the set A_N is non empty and forms a face of the set A.

Proof. It follows from the assumptions (i) and (ii) that the expected value function $f(x)$ is piecewise linear and convex, and hence $f(x)$ can be represented as a

maximum of a finite number of affine functions $\ell_i(x)$, $i = 1, ..., n$. Consequently the space $\mathbb{R}^m$ can be partitioned into a union of convex polyhedral sets $C_1, ..., C_n$ such that $f(x)$, restricted to C_i, coincides with $\ell_i(x)$, $i = 1, ..., n$.

Let us make the following observations. Suppose that $f(x)$ is affine on a convex polyhedral set C. Then function $h(\cdot, \omega)$ is also affine on C for every $\omega \in \Omega$. Indeed, suppose for a moment that the set C has a non empty interior and that for some $\omega \in \Omega$ the corresponding function $h(\cdot, \omega)$ is not affine on C. Since $h(\cdot, \omega)$ is piecewise linear and convex, this can happen only if there is a point $\hat{x}$ in the interior of C such that $\partial h(\hat{x}, \omega)$ is not a singleton. By the Moreau-Rockafellar theorem (see [18, Theorem 23.8]) we have that $\partial f(\hat{x}) = \sum_{k=1}^{K} p_k \partial h(\hat{x}, \omega_k)$. Therefore if $\partial h(\hat{x}, \omega)$ is not a singleton, then $\partial f(\hat{x})$ is also not a singleton. This, however, cannot happen since $f(x)$ is affine on C. In case the interior of C is empty, we can restrict the problem to the linear space generated by C and to proceed as above. Now, since the sample average function $\hat{f}_N(x)$ is a linear combination of the functions $h(\cdot, \omega)$, $\omega \in \Omega$, with nonnegative coefficients, it follows that $\hat{f}_N(x)$ is also affine on C for *any* realization of the random sample.

Our second observation is the following. Let $g(x)$ be a convex function taking a constant value over a convex set S. Then $\partial g(x)$ is constant over the relative interior of S (e.g., [3, Lemma 1.115]). By adding to $g(x)$ an affine function, we obtain that the same property holds if $g(x)$ is affine over S.

By the above observations we can take points z_i in the relative interior of each face of the sets $C_1, ..., C_n$. Note that an extreme point of a set C_i is viewed as its face, of dimension zero, and its relative interior coincides with the considered extreme point. Since each set C_i is polyhedral, it has a finite number of faces, and hence the total number of such points will be finite. This completes the proof of the assertion (a). Assertion (b) follows immediately from Proposition 2.2 and assertion (a).

Let us prove (c). Since the function $f(x)$ is piecewise linear, the set A is a convex polyhedral set, and by assumption (iv), A is compact.

Let us observe that by adding a barrier function of the form $\psi(x) := \alpha \operatorname{dist}(x, \Theta)$ to the objective function $f(x)$, for sufficiently large $\alpha > 0$, we can reduce the true problem to the unconstrained problem

$$\operatorname*{Min}_{x \in \mathbb{R}^m} \mathbb{E}_P h^*(x, \omega), \tag{2.14}$$

where $h^*(x, \omega) := h(x, \omega) + \psi(x)$. It is well-known that, for α large enough, the optimal solutions of problems (1.1) and (2.14) coincide (see, e.g., [2, Proposition 5.4.1]). Since Θ is convex, the barrier function, and hence the functions $h^*(\cdot, \omega)$, are also convex. Moreover, since by the assumption (iii) the set Θ is polyhedral, the barrier function is also polyhedral if we take distance with respect to the ℓ_1 norm in $\mathbb{R}^m$. Therefore, without loss of generality, we can assume in the subsequent analysis that $\Theta = \mathbb{R}^m$, i.e. that the problem under consideration is *unconstrained*.

Let S be a sufficiently large convex compact polyhedral set (e.g. a cube) such that the set A is included in the interior of the set S. Such set exists since A is bounded. Consider the sets $C_i' := C_i \cap S$, $i = 1, ..., n$. These sets are polyhedral and compact. We can assume that all these sets are different from each other and that A coincides with the set C_1'. Now let $\{x_1, ..., x_q\}$ be the set of all extreme points (vertices) of the sets $C_1', ..., C_n'$ such that, for some $\ell < q$, points $x_1, ..., x_\ell$ form the set of extreme points of A. Since each set C_i' is polyhedral, there is a finite number of such points. Suppose that condition (2.13) holds, and let C_k', $k \geq 2$, be a set from the above collection such that the intersection of C_k' with A is non empty. Since $\hat{f}_N(x)$ is

linear on C'_k and C'_k is compact, it follows from condition (2.13) that the minimum of $\hat{f}_N(x)$ over C'_k is attained on a non empty subset of the set A. Consider a collection of such sets C'_k that their union forms a neighborhood of the set A. Then $\hat{f}_N(x)$ attains its minimum over that union on a non empty subset A^*_N of A. By convexity of $\hat{f}_N(x)$ it follows then that the set A_N coincides with A^*_N, and hence is non empty and is a subset of A. Finally, since $\hat{f}_N(x)$ is linear on A, it follows that A_N is a face of A. ∎

We give now two proofs of Theorem 2.3, which give a different insight into the problem.

Proof of Theorem 2.3 As it was shown in the proof of the above lemma, by adding a barrier function, we can reduce the problem to an unconstrained one. Therefore without loss of generality, we can assume that $\Theta = \mathbb{R}^m$, i.e. that the problem is *unconstrained*.

It follows from the assumptions (i) and (ii) that the expected value function $f(x)$ is piecewise linear and convex. Therefore the set A of optimal solutions of the true problem is a convex polyhedral, and by (iv), compact set. By the strong Law of Large Numbers we have that w.p.1 the approximating functions $\hat{f}_N(x)$ converge pointwise to $f(x)$. Moreover, by the same arguments as in the proof of Proposition 2.2 we have that this convergence is uniform on any compact subset of $\mathbb{R}^m$. Let V be a compact neighborhood of the set A. Then w.p.1 for N large enough $\hat{f}_N(x)$ has a minimizer over V which is arbitrarily close to A and hence lies in the interior of V. By convexity this minimizer will be a global minimizer of $\hat{f}_N(x)$. This shows that w.p.1 for N large enough the set A_N of optimal solutions of the approximating problem is non empty.

Since $f(x)$ is piecewise linear and convex, we have that subdifferentials of $f(x)$ are convex compact polyhedral sets and, by Lemma 2.4, it follows that the total number of the extreme points of all subdifferentials $\partial f(x)$ is finite. Moreover, since for any $x \notin A$ we have that $0 \notin \partial f(x)$, it follows that there exists $\varepsilon > 0$ such that the distance from the null vector $0 \in \mathbb{R}^m$ to $\partial f(x)$ is greater than ε for all $x \notin A$. Together with (2.12) this implies that w.p.1 for N large enough, $0 \notin \partial \hat{f}_N(x)$ for all $x \notin A$, and hence any $x \notin A$ cannot be an optimal solution of the approximating problem. This shows that w.p.1 for N large enough the inclusion $A_N \subset A$ holds. Let us finally observe that since $f(x)$, and hence $\hat{f}_N(x)$, are linear on A, and A_N is the set of minimizers of $\hat{f}_N(x)$ over A, it follows that A_N is a face of A.

Let us give now the second proof. Let $\{x_1, ..., x_q\}$ be the set of points constructed in the assertion (c) of Lemma 2.4. Since this set is finite and A is the set of minimizers of $f(x)$, we have that there exists $\varepsilon > 0$ such that $f(x_i) + \varepsilon < f(x_j)$ for any $i \in \{1, ..., \ell\}$ and $j \in \{\ell + 1, ..., q\}$. By the Law of Large Numbers we have that $\hat{f}_N(x_i)$ converges to $f(x_i)$, w.p.1 as $N \to \infty$, for every $i \in \{x_1, ..., x_q\}$. Therefore w.p.1 for N large enough we have that $\hat{f}_N(x_i) < f(x_i) + \varepsilon/2$ for $i \in \{1, ..., \ell\}$, and $\hat{f}_N(x_j) > f(x_j) - \varepsilon/2$ for $j \in \{\ell + 1, ..., q\}$, and hence condition (2.13) follows. Together with assertion (c) of Lemma 2.4 this proves that A_N is non empty and forms a face of A. ∎

Under the assumptions of the above theorem, the set A_N of optimal solutions of the approximating problem is convex and polyhedral. The above theorem shows that w.p.1 for N large enough, every optimal solution of the approximating problem is an optimal solution of the true problem and every vertex of the set of optimal solutions of the approximating problem is a vertex of the set of optimal solutions of the true problem.

In order to see what may happen consider the following example. Let $h(x, \omega) :=$

$|x_1 - \omega|$, where $x = (x_1, x_2) \in \mathbb{R}^2$ and $\omega \in \Omega$ with $\Omega := \{-2, -1, 1, 2\} \subset \mathbb{R}$. Suppose that the probability of ω being equal to any of the points of Ω is 0.25 and let $\Theta := \{x \in \mathbb{R}^2 : |x_2| \leq 1\}$. Then the set A of optimal solutions of the corresponding true problem is $A = \{x : |x_1| \leq 1,\ |x_2| \leq 1\}$. On the other hand, for large N, the set of optimal solutions of the approximating problem is given either by the face $\{x : x_1 = -1,\ |x_2| \leq 1\}$ or the face $\{x : x_1 = 1,\ |x_2| \leq 1\}$ of the set A.

3. Exponential rate of convergence. In the previous section we showed that, under appropriate assumptions, the approximating problem (1.2) yields an exact optimal solution of the true problem (1.1) w.p.1 for N large enough. Since convergence w.p.1 implies convergence in probability, it follows that the probability of this event tends to one as N tends to infinity. That result, however, does not say how large the sample size N should be in order for the approximating problem to provide such an exact solution.

Similarly to the example presented in the introduction, it turns out that, in the case under consideration (i.e. when Ω is finite and $h(\cdot, \omega)$ are piecewise linear), the convergence of the corresponding probability to one is *exponentially fast*. A consequence of this, somewhat surprising, fact is that one does not need a very large sample to find the optimal solution of (1.1), which shows that Monte Carlo approximations techniques can be an effective approach to solve such problems.

In this section we formalize and prove this result. We begin by considering again the case where the true problem (1.1) has a unique optimal solution $\bar{x}$. Suppose that the assumption (A) holds. Recall that S^{m-1} denotes the sphere in $\mathbb{R}^m$, and consider the Banach space $Z := C(S^{m-1})$ of real valued continuous functions defined on S^{m-1} and equipped with the sup-norm. By restricting a positively homogeneous function to S^{m-1}, we can identify Z with the space of continuous positively homogeneous functions on $\mathbb{R}^m$. Denote by Z^* the dual space of Z, i.e. the space of continuous linear functionals defined on Z.

Let $\mathcal{B}$ be the σ-algebra of Borel sets in Z. Consider the function

$$\eta(d, \omega) := h'_\omega(\bar{x}, d), \quad d \in \mathbb{R}^m,\ \omega \in \Omega. \tag{3.1}$$

The function $\eta(\cdot, \omega)$ is convex, and hence continuous, and is positively homogeneous. Therefore it can be considered as an element of Z. Moreover, the mapping $\omega \to \eta(\cdot, \omega)$, from $(\Omega, \mathcal{F})$ into $(Z, \mathcal{B})$, is measurable and hence $\eta(\cdot, \omega)$ can be considered as a random element of $(Z, \mathcal{B})$. Let $\mathbb{P}$ be the probability measure on Z induced by the measure P. Note that $\mathbb{E}_P \eta(d, \omega) = f'(\bar{x}, d)$, and that the measure $\mathbb{P}$ is concentrated on the subset of Z formed by convex positively homogeneous functions.

Assumption (B) There exists a constant $\kappa > 0$ such that

$$\|\eta(\cdot, \omega)\|_Z \leq \kappa, \quad \text{for } P\text{- almost every } \omega.$$

This assumption clearly holds if the set Ω is finite. Note that

$$\|\eta(\cdot, \omega)\|_Z = \sup_{d \in S^{m-1}} |h'_\omega(\bar{x}, d)|.$$

Therefore assumption (B) means that the subdifferentials $\partial h(\bar{x}, \omega)$ are uniformly bounded for P-almost every ω. Notice that this is what happens in two-stage stochastic programming problems with complete recourse if only the right-hand side is random, since in that case the dual feasibility set does not depend on ω. Complete recourse implies that the dual feasibility set is also bounded. Therefore, in such case the subdifferentials $\partial h(\bar{x}, \omega)$ are uniformly bounded for all ω.

Let us recall now a few facts about random variables on Banach spaces. Let $\eta_1, \eta_2, \ldots$, be an i.i.d. sequence of random elements of $(Z, \mathcal{B})$, with the common distribution $I\!P$, and define $\zeta_N := N^{-1} \sum_{j=1}^N \eta_j$. Note that assumption (B) implies that $\int_Z \|z\|_Z I\!P(dz) < \infty$. Then, by the strong Law of Large Numbers (for Banach spaces) we have that $\zeta_N \to \zeta := I\!E[\eta]$ w.p.1, where the convergence is in the norm of Z and the expectation operator corresponds to the so-called Bochner integral (see, e.g., Hiai [10]).

Let

$$M(z^*) := \int e^{z^*(z)} I\!P(dz), \quad z^* \in Z^*,$$

be the moment generating function of $I\!P$ (i.e. of $\eta(\cdot, \omega)$). A version of Cramér's Theorem for Banach spaces (see, e.g., Deuschel and Stroock [8]) can be stated as follows. If for any $\alpha \in [0, \infty)$ we have

$$\int_Z e^{\alpha \|z\|} I\!P(dz) < \infty, \tag{3.2}$$

then a Large Deviations Principle (LDP) holds for $\{\zeta_N\}$, i.e. for any $\mathcal{B}$-measurable set $\Gamma \subset Z$ we have that

$$\begin{array}{ll} -\inf_{z \in \mathrm{int}(\Gamma)} I(z) & \leq \liminf_{N\to\infty} N^{-1} \log[P(\zeta_N \in \Gamma)] \\ & \leq \limsup_{N\to\infty} N^{-1} \log[P(\zeta_N \in \Gamma)] \leq -\inf_{z \in \mathrm{cl}(\Gamma)} I(z). \end{array} \tag{3.3}$$

Here $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and the topological closure, respectively, of the set $\Gamma \subset Z$, and $I(z)$ is the large deviations rate function, which is given by

$$I(z) := \sup_{z^* \in Z^*} \{z^*(z) - \log M(z^*)\}. \tag{3.4}$$

Notice that (3.2) follows immediately from assumption (B).

For any $d \in S^{m-1}$ we can define a functional $z_d^* \in Z^*$ as $z_d^*(z) := z(d)$. Let $M_d(t) := M(tz_d^*)$. Note that we can also write

$$M_d(t) = I\!E_P \left\{ e^{t\eta(d,\omega)} \right\},$$

so we recognize $M_d(t)$ as the moment generating function of the (one dimensional) random variable $X := \eta(d, \omega)$. Note also that assumption (B) implies that $M_d(t) < \infty$ for all $t \in I\!R$. Consider the rate function of $\eta(d, \omega)$, that is,

$$I_d(\alpha) := \sup_{t \in I\!R} \left[t\alpha - \log M_d(t) \right]. \tag{3.5}$$

By taking z^* in the right hand side of (3.4) of the form $z^* := tz_d^*$, we obtain that, for any $z \in Z$,

$$I(z) \geq \sup_{d \in S^{m-1}} \sup_{t \in I\!R} \left[tz(d) - \log M_d(t) \right] = \sup_{d \in S^{m-1}} I_d(z(d)). \tag{3.6}$$

Let A_N be the set of optimal solutions of the approximating problem (1.2), and consider the following event

$$\mathcal{E}_N := \left\{ \text{ the set } A_N \text{ is non empty and } A_N = \{\bar{x}\} \right\}. \tag{3.7}$$

The above event $\mathcal{E}_N$ means that the approximating problem possesses unique optimal solution $\hat{x}_N$ and that $\hat{x}_N = \bar{x}$. Denote by $\mathcal{E}_N^c$ the complement of the event $\mathcal{E}_N$. Note that the probability $P(\mathcal{E}_N)$, of the event $\mathcal{E}_N$, is equal to $1 - P(\mathcal{E}_N^c)$. The following theorem shows that the probability of the event $\mathcal{E}_N^c$ approaches zero exponentially fast.

THEOREM 3.1. *Suppose that the assumptions of Theorem* 2.1 *are satisfied, and that assumption (B) holds. Then there exists a constant* $\beta > 0$ *such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log[P(\mathcal{E}_N^c)] \;\leq\; -\beta. \tag{3.8}$$

Proof. Consider $\zeta_N(\cdot) := N^{-1} \sum_{j=1}^N \eta(\cdot, \omega^j) = \hat{f}'_N(\bar{x}, \cdot)$, and the set

$$F := \left\{ z \in Z : \inf_{d \in T_\Theta(\bar{x}) \cap S^{m-1}} z(d) \leq 0 \right\}. \tag{3.9}$$

Since the topology on Z is that of uniform convergence, it follows that the min-function

$$\phi(z) := \inf_{d \in T_\Theta(\bar{x}) \cap S^{m-1}} z(d)$$

is continuous on the space Z, and hence the set F is closed in Z. By the definition of the set F, we have that if $\zeta_N \notin F$, then $\zeta_N(d) > 0$ for all $d \in T_\Theta(\bar{x}) \cap S^{m-1}$. Consequently, in that case, $\hat{x}_N = \bar{x}$ is the unique optimal solution of the approximating problem. Therefore we have that

$$P(\mathcal{E}_N^c) \leq P(\zeta_N \in F).$$

It follows then by the last inequality of (3.3) that we only need to show that the constant

$$\beta := \inf_{z \in F} I(z) \tag{3.10}$$

is positive.

Consider a fixed direction $d \in T_\Theta(\bar{x}) \cap S^{m-1}$, and let X denote the corresponding random variable $\eta(d, \omega)$. Let $\Lambda(t) := \log M_d(t) = \log I\!E[e^{tX}]$ be the logarithmic moment generating function of X. By the Dominated Convergence Theorem we have that $M_d(t)$ is differentiable for all $t \in I\!R$ and $M'_d(t) = I\!E[Xe^{tX}]$. It follows that $\Lambda'(t) = I\!E[Xe^{tX}]/I\!E[e^{tX}]$ and hence, since $|X| \leq \kappa$ by assumption (B),

$$|\Lambda'(t)| \;\leq\; \frac{I\!E[|X|e^{tX}]}{I\!E[e^{tX}]} \leq \kappa, \quad \forall t \in I\!R.$$

Similarly, we have

$$|\Lambda''(t)| \;= \left| \frac{I\!E[X^2 e^{tX}]}{I\!E[e^{tX}]} - (\Lambda'(t))^2 \right| \leq |\kappa^2 - (\Lambda'(t))^2| \leq\; \kappa^2, \quad \forall t \in I\!R. \tag{3.11}$$

By the Mean Value Theorem, (3.11) implies that, for all $t, s \in I\!R$,

$$|\Lambda'(t) - \Lambda'(s)| \;\leq\; \kappa^2 |t - s|. \tag{3.12}$$

Since the function $\Lambda(\cdot)$ is convex, it follows from a result in convex analysis (e.g., [12, Theorem X.4.2.2]) that the conjugate function $I_d = \Lambda^*$ is *strongly convex* modulus $1/\kappa^2$, that is,

$$I_d(\alpha_2) \ \geq \ I_d(\alpha_1) + I'_d(\alpha_1)(\alpha_2 - \alpha_1) + \frac{1}{2\kappa^2}|\alpha_2 - \alpha_1|^2$$

for all $\alpha_1, \alpha_2 \in I\!R$. Since at $\bar{\alpha}_d := I\!E[X] = f'(\bar{x}, d)$ we have that $I_d(\bar{\alpha}_d) = I'_d(\bar{\alpha}_d) = 0$, it follows that

$$I_d(\alpha) \ \geq \ \frac{1}{2\kappa^2}|\alpha - \bar{\alpha}_d|^2, \qquad \forall\, \alpha \in I\!R. \tag{3.13}$$

By the assumption (A) we have that $f'(\bar{x}, d) \geq c$ for all $d \in T_\Theta(\bar{x}) \cap S^{m-1}$, and hence we obtain that

$$I_d(0) \ \geq \ \frac{c^2}{2\kappa^2}, \qquad \forall\, d \in T_\Theta(\bar{x}) \cap S^{m-1}. \tag{3.14}$$

By the definition of the set F we have that if $z \in F$, then there exists $d \in T_\Theta(\bar{x}) \cap S^{m-1}$ such that $z(d) \leq 0$. It follows then by (3.6) and (3.14) that $I(z) \geq c^2/(2\kappa^2)$ for any $z \in F$. Consequently we obtain

$$\beta \geq \ \frac{c^2}{2\kappa^2}, \tag{3.15}$$

which completes the proof. ∎

The inequality (3.8) means that the probability that the approximating problem (1.2) has a unique optimal solution which coincides with the optimal solution of the true problem (1.1), approaches one exponentially fast. The inequality (3.15) also gives an estimate of the corresponding exponential constant.

Consider now a situation where the true problem (1.1) may have multiple solutions. As in the case of convergence w.p.1 presented in section 2, stronger assumptions are needed. Let A_N be the set of optimal solutions of the approximating problem (1.2), and consider the following event

$$\mathcal{M}_N := \{ \text{ the set } A_N \text{ is non empty and forms a face of the set } A \ \}. \tag{3.16}$$

THEOREM 3.2. *Suppose that the assumptions of Theorem* 2.3 *hold. Then there exists a constant* $\beta > 0$ *such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log[P(\mathcal{M}_N^c)] \ \leq \ -\beta. \tag{3.17}$$

Proof. It is possible to prove this theorem by using arguments of Theorem 3.1 combined with assertions (a) and (b) of Lemma 2.4. The proof becomes even simpler if we use assertion (c) of Lemma 2.4. Let $\{x_1, ..., x_q\}$ be the set of points constructed in the assertion (c) of Lemma 2.4. Recall that $\{x_1, ..., x_\ell\}$ forms the set of extreme points of A, and that $f(x_i) < f(x_j)$ for any $i \in \{1, ..., \ell\}$ and $j \in \{\ell + 1, ..., q\}$. Note that, by condition (2.13), we have that

$$\mathcal{M}_N^c \subset \left\{ \exists\, i \in \{1, ..., \ell\}, \exists\, j \in \{\ell + 1, ..., q\} \text{ such that } \hat{f}_N(x_i) \geq \hat{f}_N(x_j) \right\}. \tag{3.18}$$

Moreover, there is $\varepsilon > 0$ such that the event in the right hand side of (3.18) is included in the union of the events $\mathcal{A}_i := \left\{\hat{f}_N(x_i) \geq f(x_i) + \varepsilon\right\}$, $i = 1, ..., \ell$, and $\mathcal{A}_j := \left\{\hat{f}_N(x_j) \leq f(x_j) - \varepsilon\right\}$, $j = \ell + 1, ...q$. It follows that

$$P\left(\mathcal{M}_N^c\right) \leq \sum_{i=1}^{\ell} P\left(\hat{f}_N(x_i) \geq f(x_i) + \varepsilon\right) + \sum_{j=\ell+1}^{q} P\left(\hat{f}_N(x_j) \leq f(x_j) - \varepsilon\right).$$

Therefore, in order to prove (3.17) it suffices to show that, for any $i \in \{1, ..., \ell\}$, there exists $\beta_i > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log\left[P\left(\hat{f}_N(x_i) \geq f(x_i) + \varepsilon\right)\right] \leq -\beta_i$$

and, similarly, for any $j \in \{\ell + 1, ..., q\}$, there exists $\beta_j > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log\left[P\left(\hat{f}_N(x_i) \leq f(x_i) - \varepsilon\right)\right] \leq -\beta_j.$$

Both assertions follow immediately from the Large Deviations Principle (in a unidimensional setting), since $I\!E[\hat{f}_N(x_i)] = f(x_i)$, $i = 1, \ldots, q$. This completes the proof by taking $\beta := \min_{i\in\{1,...,q\}} \beta_i$. ∎

4. Examples . In this section we present some examples to illustrate the ideas discussed in sections 2 and 3.

4.1. The median problem, revisited. We begin by analyzing in more detail the median problem (1.3) discussed in the introduction. Let $Y_1, ..., Y_m$ be i.i.d. real valued random variables, each one taking values -1, 0 and 1 with equal probabilities $1/3$. Let $\hat{x}_N$ denote an optimal solution of the corresponding approximating problem (1.4). As it was shown in the introduction, $\hat{x}_N$ coincides with the true optimal solution $\bar{x} = 0$ with very high probability, even for small values of N compared to the size of the sample space.

We can approach this problem from the point of view of the Large Deviations theory. Let X be a binomial random variable $B(N, p)$, with $p = 1/3$. As it was discussed in the introduction, the probability of the event $\hat{x}_N = 0$ is at least $1 - 2P(X \geq N/2)$ (more precisely, when N is even this probability is exactly $1 - 2P(X \geq N/2) + \binom{N}{N/2} p^N$, the last term becoming negligible as N grows). By Cramér's Large Deviations theorem we have that (see, e.g., [7, Thm. 2.2.3])

$$\begin{aligned} -\inf_{z>1/2} I(z) &\leq \liminf_{N\to\infty} \tfrac{1}{N} \log\left[P\left(\tfrac{X}{N} \geq \tfrac{1}{2}\right)\right] \\ &\leq \limsup_{N\to\infty} \tfrac{1}{N} \log\left[P\left(\tfrac{X}{N} \geq \tfrac{1}{2}\right)\right] \leq -\inf_{z\geq 1/2} I(z). \end{aligned}$$

For a binomial distribution $B(N, p)$, the Large Deviations rate function $I(z)$ is given by

$$I(z) = z \log\left[\frac{(1-p)z}{p(1-z)}\right] - \log\left[1 - p + \frac{(1-p)z}{1-z}\right]. \tag{4.1}$$

Since $I(\cdot)$ is continuous, it follows that

$$\lim_{N\to\infty} \frac{1}{N} \log\left[P\left(\frac{X}{N} \geq \frac{1}{2}\right)\right] = -\inf_{z\geq 1/2} I(z) = -I(0.5),$$

the last equality arising from the fact that the function $I(\cdot)$ is increasing on the interval $[p, \infty)$. From (4.1) we obtain that

$$I(0.5) = \log\left[\frac{(p^{-1}-1)^{1/2}}{2(1-p)}\right]. \tag{4.2}$$

For $p = 1/3$ we have $I(0.5) = \log\left(\frac{3\sqrt{2}}{4}\right) = 0.0589$, and hence the probability $P(X/N \geq 1/2)$ converges to zero at the exponential rate $e^{-0.0589N}$. Note that in the considered (one dimensional) case the upper bound of Cramér's theorem holds for any N (and not just in the limiting sense). It follows that the probability that the sample estimate $\hat{x}_N$ is equal to the true optimal solution is greater than $(1-2e^{-0.0589N})^m$, which for large N is approximately equal to $1 - 2m\,e^{-0.0589N}$. Consequently the probability that the sample estimate $\hat{x}_N$ is not equal to the true optimal solution decreases exponentially fast with the sample size N and increases linearly with the number of variables m. For example, for $N = 100$ and $m = 50$ we have, by the above estimate, that the probability of the sample estimate $\hat{x}_N$ being equal to the true optimal solution is at least $(1-2e^{-5.89})^{50} = 0.76$. This can be compared with the exact probability of that event, which is about 0.96. This is quite typical for the large deviations estimates. For finite and not too "large" N, the large deviations estimates give poor approximations of the corresponding probabilities. What the Large Deviations theory provides, of course, is the exponential rate at which the corresponding probabilities converge to zero.

Suppose now that each variable Y_i has the following discrete distribution: it can take values -1, -0.5, 0.5 and 1 with equal probabilities 0.25. In that case the set of optimal solutions of the true problem (1.3) is not a singleton, and is given by the cube $\{x : -0.5 \leq x_i \leq 0.5\}$. We have that the probability that the sample estimate $\hat{x}_{iN}$ belongs to the interval $[-0.5, 0.5]$ is at least $1-2P(X \geq N/2)$, where $X \sim B(N, 0.25)$. Again we obtain that the probability that $\hat{x}_N$ is an exact optimal solution of the true problem is approaching one exponentially fast with increasing N.

Now let $m = 1$ and suppose that the distribution of Y is discrete with possible values given by an odd number $r = 2\ell + 1$ of points equally spaced on the interval $[-1, 1]$ with equal probabilities of $1/r$. For "large" r we can view this as a discretization of the uniform distribution on the interval $[-1, 1]$. Then by the same arguments as above we obtain that the probability that $\hat{x}_N = 0$ is at least $1-2P(X \geq N/2)$, where $X \sim B(N, p)$ with $p = \ell/r$.

An estimate of how fast N grows as a function of the number of variables m and the number of discretization points r can be obtained using again Large Deviations techniques. Suppose that $m \geq 1$ and that each random variable Y_i, $i = 1, ..., m$, has a discrete distribution as above. From (4.2) we have that in this case the constant $\beta := I(0.5)$ is given by

$$\beta = \frac{1}{2}\log\left[\frac{r^2}{r^2-1}\right], \tag{4.3}$$

and hence

$$P(\hat{x}_N = 0) \geq (1-2e^{-\beta N})^m \cong 1 - 2me^{-\beta N}.$$

Consequently, for a fixed $\varepsilon > 0$, a (conservative) estimate of the sample size N needed to obtain $P(\hat{x}_N = 0) \geq 1-\varepsilon$ is given by

$$N = \beta^{-1}\log(2m/\varepsilon) \cong (2r^2-1)\log(2m/\varepsilon),$$

so we see that N grows *quadratically* with the number of discretization points and *logarithmically* with the number of random variables.

4.2. A two-stage stochatic programming problem. We present now some numerical results obtained for the capacity expansion problem CEP1 described in [11], which can be modeled as a two-stage stochastic programming problem with complete recourse. The problem has 8 decision variables with 5 constraints (plus bound constraints) on the first stage, and 15 decision variables with 7 constraints (plus lower bound constraints) on the second stage. The random variables, which correspond to demand in the model, appear only on the right hand side of the second stage. There are three independent and identically distributed random variables, each taking six possible values with equal probability, so the sample space has size $6^3 = 216$.

For the sake of verification, we initially solved the problem exactly by solving the equivalent deterministic LP, and obtained the true minimizer $\bar{x}$. Notice that this optimal solution is unique. We then solved the corresponding Monte Carlo approximations, with sample sizes $N = 2, 5, 10, 15, 20, 35, 50$. For each sample size, we solved the approximating problem 400 times, and counted how many times the optimal solution $\hat{x}_N$, of the approximating problem, coincided with the true solution $\bar{x}$. The corresponding proportion $\hat{p}$ is then an estimate of the probability $P(\hat{x}_N = \bar{x})$. Since the generated replications are independent, it follows that an unbiased estimator of the variance of $\hat{p}$ is given by $\hat{p}(1-\hat{p})/399$. From this value we obtain a 95% confidence interval whose half-width is denoted by Δ. The results are displayed in Table 1.

N	$\hat{p}$	Δ
2	0.463	.049
5	0.715	.044
10	0.793	.040
15	0.835	.036
20	0.905	.029
35	0.958	.020
50	0.975	.015

TABLE 1
Estimated probabilities $P(\hat{x}_N = \bar{x})$

Notice again the exponential feature of the numbers on the table, i.e how fast $\hat{p}$ gets close to one. It is interesting to notice that convergence in the CEP1 model is even faster than in the median problem, even though the median problem is much more structured (in particular, the median problem is separable) with a smaller sample space (27 points for three random variables, as opposed to 216 points in the CEP1 model). For instance, in the median problem a sample size of 20 gives the true optimal solution with probability 0.544, whereas in the CEP1 problem that probability is approximately 0.9. These results corroborate the ideas presented in the previous sections, showing that convergence can be very fast if there is a sharp minimum such as in the case of the CEP1 model. The results also suggest that the separability inherent to the median problem was not a major factor to the speed of convergence, which encourages us to think that the numerical results reported here can be obtained in more complex problems. Of course, more research is needed to draw any definite conclusions.

5. Conclusions. We presented in this paper some results concerning convergence of Monte Carlo simulation-based approximations for a class of stochastic programming problems. As pointed out in the introduction, the usual approach to convergence analysis found in the literature consists in showing that optimal solutions of approximating problems converge, with probability one, to optimal solutions of the original problem, or in obtaining bounds for the rate of convergence via Central Limit Theorem or Large Deviations type asymptotics. We show, under some specific assumptions (in particular under the assumption that the sample space is finite) that the approximating problem provides an *exact* optimal solution w.p.1 for sample size N large enough and, moreover, that the probability of such an event approaches one at an *exponential* rate. This suggests that, in such cases, Monte Carlo simulation based algorithms could be efficient, since one may not need a large sample to find an exact optimal solution.

The median problem presented in section 4 illustrates that point. For a problem with 3^{200} scenarios, an approximating problem which employs only $N = 120$ samples, of a vector of dimension $m = 200$, yields the exact optimal solution approximately 95% of the time. Even more impressively, it is possible to show by the same type of calculations that $N = 150$ samples are enough to obtain the exact optimal solution with probability of about 95% for $m = 1000$ random variables, i.e. for 3^{1000} scenarios. Estimates of the sample size N, which were obtained in section 4 by the large deviations approximations, give slightly bigger values of N (for example, they give $N = 180$ instead of $N = 150$ for $m = 1000$). In either case the required sample size grows as a logarithm of the number m of random variables in that example. Of course, one must take into account the fact that this is a very structured problem, and in a more general case one may not get such drastically fast convergence; in fact, the flatter the objective function is around the optimal solution, the slower the convergence will be. Nevertheless, the CEP1 model studied in section 4 seems to indicate that fast convergence is obtained in more general problems, even in the absence of separability.

One should, however, be cautious about these results, especially with respect to the following aspect. The fact that the convergence is exponential does not necessarily imply that a small sample suffices. Indeed, the constant β in the corresponding exponential rate $e^{-\beta N}$ can be so small that one would need a large sample size N in order to achieve a reasonable precision. The lower bound (3.15) gives us an idea about the exponential constant β. In the median example, with r discretization points for each random variable Y_i, $i = 1, ..., m$, we have that we can take $c = 1/r$ and $\kappa = 1$, if we use ℓ_1 norm in the space $\mathbb{R}^m$. This gives us the lower bound $\beta \geq 1/(2r^2)$, which can be compared with the exact value of $\beta = \frac{1}{2}\log[r^2/(r^2-1)] \cong 1/(2r^2-1)$. Note that the estimate $\beta \geq 1/(2r^2)$ does not depend on the number m of random variables. This happens since any multiplicative constant before $e^{-\beta N}$ can be absorbed into the exponential rate as N tends to infinity.

Another remark concerns the assumption of Monte Carlo sampling in our analysis. By doing so, we were able to exploit properties of i.i.d. samples, which we used to derive our results. In practice, however, one might think of implementing *variance reduction techniques* in order to reduce even more the needed sample sizes. The incorporation of such techniques into stochastic optimization algorithms has been shown to be very effective in practice (see, e.g., [1, 5, 21]). Research on specific applications of variance reduction techniques to the type of problems discussed in this paper is underway.

REFERENCES

[1] T.G. Bailey, P.A. Jensen and D.P. Morton, "Response surface analysis of two-stage stochastic linear programming with recourse", to appear in *Naval Research Logistics*.

[2] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[3] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, to be published by Springer.

[4] J.V. Burke and M.C. Ferris, "Weak sharp minima in mathematical programming", *SIAM J. Control and Optimization*, 31, 1340-1359, 1993.

[5] G.B. Dantzig and P.W. Glynn, "Parallel processors for planning under uncertainty", *Annals of Operations Research* 22, 1-21, 1990.

[6] H.A. David, *Order Statistics*, 2nd. ed., Wiley, New York, 1981.

[7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd. ed., Springer-Verlag, New York, 1998.

[8] J.D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[9] 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), 1517-1549.

[10] F. Hiai, "Strong laws of large numbers for multivalued random variables", in *Multifunctions and Integrands*, eds. A. Dold and B. Eckmann, Springer-Verlag, Berlin, 1984.

[11] J.L. Higle and S. Sen, "Finite master programs in regularized stochastic decomposition", *Mathematical Programming*, 67 (1994), 143-168.

[12] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms I*, Springer-Verlag, Berlin, 1993.

[13] Y.M. Kaniovski, A.J. King and R.J.-B. Wets, "Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems", *Annals of Operations Research*, 56 (1995), 189–208.

[14] A.J. King and Roger J.-B. Wets, "Epi-consistency of convex stochastic programs", *Stochastics*, 34 (1991), 83-92.

[15] A.J. King and R.T. Rockafellar, "Asymptotic theory for solutions in statistical estimation and stochastic programming", *Mathematics of Operations Research*, 18 (1993), 148-162.

[16] G.Ch. Pflug, "Stochastic programs and statistical data", *Ann. Oper. Res.*, 85 (1999), 59-78.

[17] S.M. Robinson, "Analysis of sample-path optimization", *Mathematics of Operations Research*, 21 (1996), 513-528.

[18] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[19] W. Römisch and R. Schultz, "Lipschitz stability for stochastic programs with complete recourse", *SIAM J. Optimization*, 6 (1996), 531-547.

[20] A. Shapiro, "Asymptotic behavior of optimal solutions in stochastic programming", *Mathematics of Operations Research*, 18 (1993), 829-845.

[21] A. Shapiro and T. Homem-de-Mello, "A Simulation-Based Approach to Stochastic Programming with Recourse", *Mathematical Programming*, vol. 81 (1998), no. 3, 301-325.