

MONTE CARLO SIMULATION APPROACH TO STOCHASTIC PROGRAMMING

Alexander Shapiro

School of Industrial & Systems Engineering
 Georgia Institute of Technology
 Atlanta, Georgia 30332, U.S.A.

ABSTRACT

Various stochastic programming problems can be formulated as problems of optimization of an expected value function. Quite often the corresponding expectation function cannot be computed exactly and should be approximated, say by Monte Carlo sampling methods. In fact, in many practical applications, Monte Carlo simulation is the only reasonable way of estimating the expectation function. In this talk we discuss convergence properties of the sample average approximation (SAA) approach to stochastic programming. We argue that the SAA method is easily implementable and can be surprisingly efficient for some classes of stochastic programming problems.

1 INTRODUCTION

Consider the optimization problem

$$\min_{x \in X} \{g(x) := \mathbb{E}_P[G(x, \omega)]\}, \quad (1)$$

where $G : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$, the expectation is taken with respect to probability measure P defined on a sample space (Ω, \mathcal{F}) , and $X \subset \mathbb{R}^n$. We assume that for every $x \in X$ the expectation $g(x)$ is well defined.

The function $G(x, \omega)$ in itself can be defined by an optimization problem. For example, in two-stage stochastic programming with recourse, $G(x, \omega)$ is given by the optimal value of the corresponding second stage problem. In particular, in two-stage linear stochastic programming with recourse, $G(x, \omega)$ is the optimal value of

$$\begin{aligned} \text{Min}_y \quad & c^T x + q^T y \\ \text{s.t.} \quad & T x + W y = h, \quad y \geq 0, \end{aligned} \quad (2)$$

where $(q(\omega), h(\omega), T(\omega), W(\omega))$ is the random data of the problem.

If the space Ω is finite, say $\Omega := \{\omega_1, \dots, \omega_K\}$ with respective probabilities $p_k, k = 1, \dots, K$, then

$$g(x) = \sum_{k=1}^K p_k G(x, \omega_k). \quad (3)$$

Consequently problem (1) can be viewed as a deterministic optimization problem. Note, however, that the number K of possible realizations (scenarios) of the data can be very large. Suppose, for instance, that ω is a random vector with 100 stochastically independent components each having 3 realizations. Then the total number of scenarios is $K = 3^{100}$. No computer in a foreseeable future will be able to handle that number of scenarios.

One can try to solve the optimization problem (1) by a Monte Carlo simulation. That is, by generating a random (say iid) sample $\omega^1, \dots, \omega^N \sim P$ one can estimate the expectation $g(x)$ by the corresponding sample average

$$\hat{g}_N(x) := N^{-1} \sum_{j=1}^N G(x, \omega^j). \quad (4)$$

There are two basic approaches to solving the optimization problem (1) by using such Monte Carlo sampling approximations. In one approach the sampling is performed inside a particular algorithm with a new sample generated at each iteration of the corresponding numerical procedure. In the other approach $\hat{g}_N(x)$ is considered as a function of x , associated with a generated random sample, and the “true” (expected value) problem (1) is approximated by the optimization problem

$$\min_{x \in X} \hat{g}_N(x). \quad (5)$$

It is assumed that for a given $\omega \in \Omega$, the function $G(\cdot, \omega)$, and may be its derivatives, is computable at any point $x \in X$. Therefore, once the random sample is generated, problem (5) becomes a deterministic optimization problem.

Required values of the sample average function $\hat{g}_N(x)$ can be calculated either by keeping the generated random sample in the computer memory or by using the common random numbers generation. We refer to this approach as the sample average approximation (SAA) method. Note that the SAA method is not an algorithm since one still has to choose a particular numerical procedure in order to solve the SAA problem (5).

In this talk we discuss various properties of the SAA method. We argue that in many instances the SAA method can be very efficient and easily implementable. It is difficult to point at the origin of that approach. The idea is simple and natural, and variants of the SAA method were suggested by various authors over the years. In a context of simulation models a variant of the SAA method, based on Likelihood Ratio transformations, was suggested in Rubinstein and Shapiro (1993). Independently, and more or less at the same time, similar ideas were employed in Statistics for computing Maximum Likelihood estimates by Monte Carlo techniques based on Gibbs sampling, Geyer and Thompson (1992). Let us also remark that the terminology “sample average approximation” is not uniform in the literature. For example, the term “sample-path optimization” was used in Plambeck, Fu, Robinson and Suri (1996).

2 RATES OF CONVERGENCE

Let us denote by v_0 and S_0 the optimal value and the set of optimal solutions, respectively, of the “true” problem (1), and similarly by \hat{v}_N and \hat{S}_N the optimal value and the set of optimal solutions, respectively, of the SAA problem (5). By the Law of Large Numbers we have that for a given x , the sample average $\hat{g}_N(x)$ converges to the corresponding expectation $g(x)$ w.p.1. It is possible then to show that, under mild regularity conditions, \hat{v}_N converges to v_0 and $\text{dist}(\hat{x}_N, S_0)$ converges to zero w.p.1 as $N \rightarrow \infty$, for any $\hat{x}_N \in \hat{S}_N$. That is, the estimators \hat{v}_N and \hat{x}_N are consistent.

However, by the Central Limit Theorem, $\hat{g}_N(x)$ converges to $g(x)$ at a rate of $O_p(N^{-1/2})$, and therefore the convergence is slow. That is, in order to improve the accuracy of the estimator $\hat{g}_N(x)$ by one digit, one needs to increase the sample size 100 times. It appears that this slow convergence is inherited by the estimator \hat{v}_N of the optimal value. It is possible to show (Shapiro, 1991) that

$$\hat{v}_N = \min_{x \in S_0} \hat{g}_N(x) + o_p(N^{-1/2}). \tag{6}$$

In particular, if $S_0 = \{x_0\}$ is a singleton, then \hat{v}_N converges to v_0 at the same rate as $\hat{g}_N(x_0)$ converges to $g(x_0)$.

Let us consider an estimator $\hat{x}_N \in \hat{S}_N$. In the Statistics literature such estimators are called M -estimators, and it is well known that under certain regularity conditions, and in particular if $S_0 = \{x_0\}$ is a singleton and the expecta-

tion function $g(x)$ is smooth near x_0 , then $N^{1/2}(\hat{x}_N - x_0)$ converges in distribution to normal (Huber, 1967). That is, \hat{x}_N converges to x_0 also at a rate of the square root of the sample size. Moreover, in such cases \hat{x}_N is asymptotically equivalent to stochastic approximation estimators implemented with optimal step sizes (Shapiro, 1996) (see, e.g., Kushner and Clark (1978) for a discussion of the stochastic approximation method).

It appears from the above discussion that the SAA approach inherits the slow convergence properties of the Monte Carlo sampling method. However, the situation changes drastically in cases where the expectation function $g(x)$ is not smooth (differentiable). It happens quite often in such cases that the true problem (1) has a sharp optimal solution x_0 . That is, there exists a constant $\kappa > 0$ such that

$$g(x) \geq g(x_0) + \kappa \|x - x_0\| \tag{7}$$

for all $x \in X$. For example, if $G(x, \omega)$ is given by the optimal value of the linear program (2), then $G(\cdot, \omega)$ is a piecewise linear convex function. If, moreover, Ω is finite (i.e., there is a finite number of scenarios), then $g(\cdot)$ is also piecewise linear and convex. If, further, the set X is defined by linear constraints and the set S_0 of optimal solutions is nonempty, then the following inequality always holds

$$g(x) \geq v_0 + \kappa \text{dist}(x, S_0) \tag{8}$$

for all $x \in X$. Of course, property (7) is a particular case of (8) if $S_0 = \{x_0\}$ is a singleton.

The following results are derived in Shapiro and Homem-de-Mello (2001).

Theorem 2.1 *Suppose that Ω is finite, for every $\omega \in \Omega$ the function $G(\cdot, \omega)$ is convex piecewise linear, the set X is polyhedral and the set S_0 is nonempty and bounded. Then the following holds:*

- (i) *w.p.1 for N large enough, the set \hat{S}_N of optimal solutions of the SAA problem is nonempty and $\hat{S}_N \subset S_0$,*
- (ii) *there exists a constant $\gamma > 0$ such that*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(\hat{S}_N \not\subset S_0) = -\gamma. \tag{9}$$

Property (i) of the above theorem asserts that for N large enough, any optimal solution \hat{x}_N of the SAA problem solves the “true” problem (1) *exactly*. Moreover, (9) implies that the probability of that event approaches one exponentially fast with increase of the sample size N . If, in addition to the assumptions of theorem 2.1, the set $S_0 = \{x_0\}$ is a singleton, then for N large enough, the SAA problem has unique optimal solution \hat{x}_N and $\hat{x}_N = x_0$, and moreover probability of that event approaches one exponentially fast. Under the condition (7) the same holds for a general distribution,

provided that the corresponding moment generating function is finite valued near zero.

The exponential rate of convergence indicates that one may not need a large sample in order to solve polyhedral problems (as, for example, the linear two-stage problems with a finite number of scenarios) exactly by using the SAA method. The exponential constant γ depends on how well the set of optimal solutions of the true problem is conditioned. The conditioning of stochastic problems, from the point of view of the SAA approach, is discussed in Shapiro, Homem-de-Mello and Kim (2000).

3 COMPLEXITY OF THE SAA METHOD

Suppose now that the feasible set X of the problem (1) is *finite*, although may be very large. That is, problem (1) is a discrete (or integer) stochastic programming problem. One may still use the same approach by applying an appropriate numerical algorithm to the SAA problem (5). Suppose that the SAA problem can be solved with a given accuracy $\delta \geq 0$. Denote by S^ε the set of ε -optimal solutions of (1), i.e., $\bar{x} \in S^\varepsilon$ iff $\bar{x} \in X$ and $g(\bar{x}) \leq v_0 + \varepsilon$. Similarly, denote by \hat{S}_N^δ the set of δ -optimal solutions of the SAA problem (5).

Consider a mapping $u : X \setminus S^\varepsilon \rightarrow S_0$ and let

$$H(x, \omega) := G(u(x), \omega) - G(x, \omega).$$

Note that since X is finite, the set S_0 of optimal solutions of the true problem is nonempty, and therefore such mapping $u(x)$ always exists. The following result is due to Kleywegt, Shapiro and Homem-de-Mello (2000).

Theorem 3.1 *Suppose that for every $x \in X$ the moment generating function of $H(x, \omega)$ is finite valued in a neighborhood of zero, and let ε and δ be nonnegative numbers such that $\delta \leq \varepsilon$. Then there is a constant $\gamma(\delta, \varepsilon) > 0$ such that*

$$\mathbb{P}(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq |X|e^{-N\gamma(\delta, \varepsilon)}. \quad (10)$$

For small ε and δ the above constant $\gamma(\delta, \varepsilon)$ can be estimated as follows

$$\gamma(\delta, \varepsilon) \geq \frac{(\varepsilon^* - \delta)^2}{3\sigma^2} > \frac{(\varepsilon - \delta)^2}{3\sigma^2}, \quad (11)$$

where

$$\varepsilon^* := \min_{x \in X \setminus S^\varepsilon} g(x) - v_0 \quad (12)$$

and

$$\sigma^2 := \max_{x \in X \setminus S^\varepsilon} \text{Var}[H(x, \omega)]. \quad (13)$$

Note that since X is finite, ε^* is always greater than ε , although the difference $\varepsilon^* - \varepsilon$ can be small.

For $\varepsilon > \delta \geq 0$ and a given significance level $\alpha \in (0, 1)$, the above analysis gives the following estimate of the sample size N that guarantees that the probability $\mathbb{P}(\hat{S}_N^\delta \subset S^\varepsilon)$ is greater than or equal to $1 - \alpha$:

$$N \geq \frac{3\sigma^2}{d(\varepsilon, \delta)} \left(\log |X| - \log \alpha \right), \quad (14)$$

where

$$d(\varepsilon, \delta) := (\varepsilon - \delta)^2.$$

This estimates shows that the sample size N required to compute an ε -optimal solution of (1) with a probability at least $1 - \alpha$, by solving the corresponding SAA problem with accuracy δ , grows as a logarithm of $|X|$.

Suppose now that X is a bounded (not necessarily finite) subset of \mathbb{R}^n . For a given $\nu > 0$, consider a finite subset X_ν of X such that for any $x \in X$ there is $x' \in X$ satisfying $\|x - x'\| \leq \nu$. If D is the diameter of the set X , then such set X_ν can be constructed with $|X_\nu| \leq (D/\nu)^n$. By reducing the feasible set X to its subset X_ν , we obtain the following estimate of the sample size, required to solve the reduced problem:

$$N \geq \frac{3\sigma^2}{d(\varepsilon, \delta)} \left(n \log \frac{D}{\nu} - \log \alpha \right). \quad (15)$$

Suppose, further, that the expectation function $g(x)$ is Lipschitz continuous on X modulus L . Then an ε -optimal solution of the reduced problem is an ε' -optimal solution of problem (1) with $\varepsilon' = \varepsilon + Lv$. By taking

$$\nu := (\varepsilon - \delta)/(2L)$$

we obtain the following estimate of the sample size N required to solve the true problem (1):

$$N \geq \kappa (n \log \beta - \log \alpha), \quad (16)$$

where

$$\kappa := \frac{12\sigma^2}{(\varepsilon - \delta)^2}$$

and

$$\beta := \frac{2DL}{(\varepsilon - \delta)^2}.$$

Although the estimates like (14) and (16) typically are too conservative for practical calculations, they indicate that the required sample size grows slowly with dimension of the

data. In particular, estimate (16) shows that the required sample size grows linearly in dimension n of the problem.

One can also solve the SAA problem repeatedly M times using M independent samples each of size N . Let $\hat{x}_N^1, \dots, \hat{x}_N^M$ be optimal solutions of the corresponding SAA problems. Probability that at least one of \hat{x}_N^m , $m = 1, \dots, M$, is an optimal solution of the true problem is at least $1 - (p_N)^M$, where $p_N := \mathbb{P}(\hat{S}_N \not\subset S_0)$. By the above exponential bounds we obtain that if complexity of solving SAA problems grows faster than at a linear rate in the sample size of the problem, then it may be advantages to solve a number, say M , of SAA problems with sample size N each than to solve one SAA problem with the sample size MN . Similar remark applies to ε -optimal solutions.

REFERENCES

- Geyer, C.J., and Thompson, E. A., Constrained Monte Carlo maximum likelihood for dependent data, (with discussion), *J. Roy. Statist. Soc. Ser. B*, 54, 657-699, 1992.
- Huber, P.J., The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp.Math. Statist. Probab.*, Univ. California Press, 1, 221-233, 1967.
- Kleywegt, A., Shapiro, A., and Homem-de-Mello, T., The sample average approximation method for stochastic discrete optimization, *SIAM Journal on Optimization*, to appear.
- Kushner, H.J. and Clark, D.S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, New York, NY, 1978.
- Plambeck, E.L., Fu, B.R., Robinson, S.M., and Suri, R., Sample-path optimization of convex stochastic performance functions, *Mathematical Programming, Series B*, 75, 137-176, 1996.
- Rubinstein, R.Y., and Shapiro, A., *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley & Sons, New York, NY, 1993.
- Shapiro, A., Asymptotic analysis of stochastic programs, *Annals of Operations Research*, 30, 169-186, 1991.
- Shapiro, A., Simulation based optimization - convergence analysis and statistical inference, *Stochastic Models*, 12, 425-454, 1996.
- Shapiro, A., and Homem-de-Mello, T., On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs, *SIAM Journal on Optimization*, 11, 70-86, 2001.
- Shapiro, A., Homem-de-Mello, T., and Kim, J., Conditioning of convex piecewise linear stochastic programs, Georgia Institute of Technology, preprint, 2000.

AUTHOR BIOGRAPHY

ALEXANDER SHAPIRO is a Professor in the Department of Industrial and Systems Engineering at Georgia Institute of Technology. He received a Ph.D. degree in applied mathematics from the Ben-Gurion University in 1981. His research interests are focused on stochastic programming, sensitivity analysis of optimization problems and multivariate statistical analysis.