Proceedings of the 2009 INFORMS Simulation Society Research Workshop
L. H. Lee, M. E. Kuhl, J. W. Fowler, and S. Robinson, eds.

MONTE CARLO INFERENCE ON SAMPLE-PATH ROOTS

Raghu Pasupathy

Department of Industrial & Systems Engineering
250 Durham Hall
Blacksburg, VA 24061, USA

ABSTRACT

Our broad focus in this talk is Monte Carlo inference for sample average approximation (SAA) estimators within stochastic root finding problems (SRFPs). We derive insights on such issues as consistency, convergence rates, sample sizing, complexity, optimal budgeting, and correlated sampling within the context of SAA estimators. Some specific questions we tackle include: (i) when and at what rate do SAA solutions to SRFPs converge? (ii) what is the complexity of stochastic root finding, and what minimum sample size ensures that a δ-root of the sample-path problem is an ε-root of the true problem? (iii) what is the "optimal" sampling/searching tradeoff, and can sequential sampling schemes be devised to attain a solution of a specified quality? and (iv) what is the effect of correlated sampling? Our results parallel the enormous body of existing work in a similar SAA setting, but for the context of simulation optimization. In addition to detailing a selected subset of key results along with numerical illustration, our talk will clarify interesting connections between parallel results in SRFPs and simulation optimization.

1 INTRODUCTION AND MOTIVATION

Recall that the Stochastic Root-Finding Problem (SRFP) is that of finding $x \in \mathbf{R}^q$ such that $g(x) = 0$, given only a consistent "simulation estimator" G_n of the function $g : \mathbf{R}^q \to \mathbf{R}^q$. It is a "lesser-known cousin" of the simulation-optimization problem (SOP), where a real-valued function is to be minimized subject to constraints, given only consistent "simulation estimators" of the objective and constraint functions involved. Both of these problems have recently generated much attention amongst researchers and practitioners, owing primarily to their generality. Various flavors of SRFPs and SOPs have thus been incorporated into a wide range of application areas including vehicular transportation networks, quality control, telecommunication systems, and health care.

The broad setting of this work is *Sample Average Approximation* — a simple method that has recently found expediency amongst researchers and practitioners in solving SRFPs and SOPs. Our interest within SAA is Monte Carlo inference for SRFPs, by which we mean an investigation into such properties as consistency, convergence rates, sample sizing, complexity, and stability of SAA estimators for SRFPs. Specifically, we ask the following questions.

- Q.1 What conditions on the root-finding function g and its estimator G_n guarantee the consistency of the sample-path roots, i.e., what conditions guarantee that the distance between the SAA solution X_n and the set π^* of true roots of g converges to zero (as $n \to \infty$) in some precise sense?
- Q.2 What is the rate at which the sequence of SAA solutions converges to the set of true roots (when it does)? Specifically, how does this rate relate to the convergence rate of the given estimator G_n of the root finding function g?
- Q.3 A central limit theorem on the SAA solution has been shown previously under the assumption that g has a unique root. Can anything be said about the limiting distribution of the SAA solution X_n in more general cases? More importantly, what

can be said about the limiting distribution of the *quality-gap* $\|g(X_n)\|$ in such cases?

Q.4 What is the complexity of stochastic root finding?

Q.5 What minimum sample size should be chosen if a "δ-root" to a sample-path function G_n is to be an "ε-root" to the true function g with a minimum specified probability?

Q.6 Can sequential sampling schemes be devised with guarantees similar to that in Q.5?

Q.7 If a finite computational budget is specified, what is the "optimal" sampling/searching tradeoff?

Q.8 Are the answers to Q.3, Q.4, and Q.5 changed when the sampling is non-i.i.d?

Most of the above questions have been asked and answered in great detail in the context of SOPs. For a complete treatment of questions akin to Q.1 through Q.8 in the context of SOPs, see Shapiro and Homem-de-Mello (2000), Shapiro (2004), Bayraksan and Morton (2007), Homem-de-Mello (2003), Homem-de-Mello (2008), Bayraksan and Morton (2009), Mak, Morton, and Wood (1999), Pasupathy (2009), Yakowitz, L'Ecuyer, and Vazquez-Abad (2000) and the references listed therein. This motivates our exclusive focus on SRFPs. We will present individual results that address (partially or fully) each of the above questions, and whenever possible, clarify the connections with the corresponding known results for the SOP context.

In what follows, we provide a subset of the results that we have been able to derive in answering questions Q.1 through Q.6. Our intent is merely providing a flavor of what we will present at the simulation workshop. To ensure brevity, we have omitted proofs and numerical results, and limit ourselves to a listing of selected results along with intuition and connections to corresponding results in the SOP context.

We start with problem statement and notation in Section 2, followed by results on consistency (Section 3), speed of convergence (Section 4), and sample sizing (Section 5). We provide some concluding remarks in Section 6.

2 PROBLEM STATEMENT AND NOTATION

The SRFP version adopted in this paper is stated formally as follows.

Given: A simulation capable of generating, for any $x \in \mathfrak{D} \subset \mathbf{R}^q$, an estimator $G_n(x)$ of the function $g : \mathfrak{D} \to \mathbf{R}^q$ such that $G_n(x) \xrightarrow{d} g(x)$ as $n \to \infty$, for all $x \in \mathfrak{D}$.

Find: A zero $x^* \in \mathfrak{D}$ of g, i.e., find x^* such that $g(x^*) = 0$, assuming that one such exists.

As stated, the problem makes no assumptions about the nature of $G_n(x)$ except that $G_n(x) \xrightarrow{d} g(x)$ as $n \to \infty$. Also, the feasible set $\mathfrak{D}$ is assumed to be known in the sense that the functions involved in the specification of $\mathfrak{D}$ are observed without error.

The following is a list of key notation and definitions adopted in the paper: (i) π^* denotes the set of solutions to the SRFP; (ii) S_n^* denotes the (random) set of solutions to the sample-path problem obtained using sample size n, i.e., $S_n^* = \{x : G_n(x) = 0\}$; (iii) $X_n^* \in S_n^*$ denotes a solution to the sample-path problem obtained using sample size n; (iv) $X_n \to X$ wp1 means that the sequence of random variables $\{X_n\}$ converges to the random variable X with probability one; (v) $X_n \xrightarrow{d} X$ means that the sequence of random variables $\{X_n\}$ converges to the random variable X in distribution; (vi) $\text{dist}(x,B) = \inf\{\|x-y\| : y \in B\}$ denotes distance between a point $x \in \mathbf{R}^q$ and a set $B \subset \mathbf{R}^q$; (vii) $\text{dist}(A,B) = \sup\{\text{dist}(x,B) : x \in A\}$ denotes distance between two sets $A, B \subset \mathbf{R}^q$; (viii) $\mathbb{H}(A,B) = \max(\text{dist}(A,B), \text{dist}(B,A))$ refers to the Hausdorff distance between the sets A and B; (ix) $B(x,r)$ denotes a ball of radius r centered on x.

3 CONSISTENCY

Recall that X_n^* denotes a solution to the sample-path problem $G_n(x) = 0$, i.e., X_n^* satisfies $G_n(X_n^*) = 0$. Theorem 1 states that uniform convergence (wp1) of $\{G_n(x)\} \to g(x)$ is sufficient to ensure that $g(X_n^*) \to 0$ wp1. Ensuring that the sequence of solutions $\{X_n^*\}$ converges to the set of true solutions π^* requires additional stipulations, specifically continuity of g and the compactness of $\mathfrak{D}$.

Theorem 1. *Assume that the functional sequence $\{G_n(x)\} \to g(x)$ uniformly (in x) wp1. Also, let $\{X_n^*\}$ be any sequence of random variables satisfying $G_n(X_n^*) = 0$. Then, $g(X_n^*) \to 0$ wp1.*

Theorem 2. *Assume that (i) the set $\mathfrak{D} \subset \mathbf{R}^q$ is compact; (ii) the function $g : \mathfrak{D} \to \mathbf{R}^q$ is continuous; and (iii) the functional sequence $\{G_n(x)\}$ converges to $g(x)$ uniformly (in x) wp1. Then, the set of sample-path solutions $S_n^* = \{x : G_n(x) = 0\}$ converges to the set of true solutions $\pi^* = \{x : g(x) = 0\}$ in the sense that $\text{dist}(S_n^*, \pi^*) \to 0$ wp1.*

Theorems 1 and 2 parallel Propositions 5 and 6 in (Shapiro 2004) for SOPs. Theorem 1 is tight, i.e., examples can easily be constructed to show that pointwise convergence of $\{G_n(x)\}$ will not suffice, e.g., $g(x) = x - 1/2, G_n(x) = n^2 x I_{[0,1/n]} + x I_{(1/n,\infty)} - 1/2$ with $X_n^* = 1/2n^2$. Similarly, the continuity and compactness stipulations cannot be relaxed in Theorem 2.

4 SPEED OF CONVERGENCE

In this section, our interest is gaining insight on the speed of convergence of SAA solutions to SRFPs. Specifically, given the speed of convergence of the estimators $\{G_n(x)\}$

to $g(x)$ in some sense, we characterize the speed at which the set of sample-path solutions S_n^* converges (in Hausdorff distance) to the set of true solutions π^*. Also of interest is the rate at which $\{\|g(X_n^*)\|\}$ — the *quality-gap* sequence — converges to 0. Theorem 3 parallels results in (Pflug 2004).

Theorem 3. *Let $\beta(n)$ be a function that satisfies $\beta(n) \to \infty$ as $n \to \infty$. Also, let the random functions $G_n(x)$ converge to $g(x)$ such that the supremum-error sequence $\{\sup_x \|G_n(x) - g(x)\|\}$ is $O_p\left(\beta(n)^{-1}\right)$, i.e., for any given $\varepsilon > 0$, there exists K_ε such that $\sup_N Pr\{\beta(n)\sup_x \|G_n(x) - g(x)\| \geq K_\varepsilon\} \leq \varepsilon$. Then the sequences $\{\sup_{x \in S_n^*} \|g(x)\|\}$ and $\{\mathbb{H}(S_n^*, \pi^*)\}$ are also $O_p\left(\beta(n)^{-1}\right)$.*

A further analysis of the quality-gap sequence $\{\|g(X_n^*\|\}$ is of interest because, as will become evident soon, it plays a key role in answering questions surrounding the relation between sample size and solution quality. With this in mind, and using further assumptions on the nature of the estimator $\{G_n(x)\}$, we now present a simple central limit theorem on $g(X_n^*)$.

Theorem 4. *Assume $X_n^* \xrightarrow{d} X_\infty$. Let μ_n, ν_n, and μ_∞ be the probability measures of $\sqrt{n}(X_n^* - E[X_\infty])$, $\sqrt{n}g(X_n^*)$, and X_∞ respectively. Furthermore, assume (i) the functions $G_n(x)$ and $g(x)$ have nonzero gradients $\nabla G_n(x), \nabla g(x)$ in some neighborhood around x^* wp1, for each $x^* \in \pi^*$; (ii) the sequence $\{\nabla G_n(x)\}$ converges uniformly (elementwise) to $\nabla g(x)$ in some neighborhood around x^* wp1, for each $x^* \in \pi^*$; and (iii) a central limit theorem holds for $G_n(x)$, i.e., $\sqrt{n}(G_n(x) - g(x)) \xrightarrow{d} N(0, \Sigma_x)$, where $N(0, \Sigma_x)$ is the Gaussian random variable with mean zero and covariance Σ_x. Then, if $\phi(z_1, z_2)$ denotes the density function of a Gaussian random variable with mean z_1 and covariance z_2,*

(i) $\mu_n(A) \to \int_A \phi(x, \nabla G_n(x)^{-1}\Sigma_x(\nabla G_n(x)^{-1})^T)\mu_\infty(dx)$;
(ii) $\nu_n(A) \to \int_A \phi(0, \Sigma_x)\mu_\infty(dx)$.

Theorem 4 essentially states that, under certain conditions, the limiting distribution of X_n^* is an X_∞-mixture of normal random variables. Since X_∞ satisfies $g(X_\infty) = 0$ by definition, $g(X_n^*)$ turns out to be a mixture of multivariate normals, each of which has mean 0. In other words, Theorem 4 states that as n becomes larger, X_n^* looks like a random draw from a multivariate normal distribution that is centered on a point in π^*, chosen according to the random variable X_∞. Similarly, $g(X_n^*)$ looks like a random draw from a multivariate normal distribution that is centered on 0, but with covariance that is dependent on the structure of g at a point (in π^*) that is chosen according to the random variable X_∞.

Theorem 4 is the general central limit theorem for stochastic root finding. It parallels Theorem 10 in (Shapiro 2004), where the limiting optimal value of the SOP turns out to be the *infimum of a certain set of multivariate normals*, each of which is centered on the optimal value of the limiting problem.

5 SAMPLE SIZING AND COMPLEXITY

We know from Theorem 4 that the random variable $g(X_n^*)$ converges in distribution to an "X_∞-mixture of normal random variables." Let us denote this mixture random variable as $N(0, \Sigma_{X_\infty})$. In Theorem 5, with a view toward establishing the complexity of stochastic root finding, we derive the rate of convergence (to 0) of the δ-quantile of the quality-gap $\|g(X_n^*)\|$.

Theorem 5. *Denote $G_n(x) = Pr\{\|g(X_n^*)\| \leq x\}, G(x) = Pr\{\|N(0, \Sigma_{X_\infty})\| \leq x\}, \varepsilon(n;\delta) = \inf\{x : G_n(x) \geq \delta\}$, and $\varepsilon(\delta) = \inf\{x : G(x) \geq \delta\}$. Let the mixing random variable X_∞ be supported on some subset π_∞ of π^*, i.e., $Pr\{X_\infty \in \pi_\infty\} = 1$. In addition to $g(x)$, let the first $k \geq 0$ gradient functions $\nabla g(x), \nabla^{(2)} g(x), \ldots, \nabla^{(k)} g(x)$ have their zeros in π_∞, i.e., $g(x) = \nabla g(x) = \nabla^{(2)} g(x) = \cdots = \nabla^{(k)} g(x) = 0$. Then, $\lim_{n\to\infty} n^{(k+1)/2}\varepsilon(n;\delta) = \varepsilon(\delta)$.*

Theorem 5 says that the canonical rate of stochastic root finding is $O(n^{-(k+1)/2})$, where k is as defined. If $k = 0$, or if no information is available on the gradients of g, we see that a rate of $O(1/\sqrt{n})$ is all that can be guaranteed. The corresponding result for SOPs states that the δ-quantile of the optimality gap drops to zero at $O(n^{-1})$ when the objective function "behaves quadratic" at the optimal point (e.g., the optimal solution is at the interior of the feasible region), and $O(1/\sqrt{n})$ otherwise. During implementation, various available upper bounds on $\varepsilon(\delta)$ can be used.

We next address the question of choosing a sample size when implementing SAA algorithms for SRFPs. Specifically, suppose we want to obtain a solution X_n^* whose quality-gap $\|g(X_n^*)\|$ is at most $\varepsilon > 0$. How large should the sample size n be, and how fine should the sample-path problem be solved, so as to ensure that the resulting solution has a quality-gap of at most ε with a probability exceeding $1 - \alpha$? Theorem 6, paralleling Theorem 12 in (Shapiro 2004), answers this question for the case where $\mathfrak{D}$ is finite. An extension to continuous $\mathfrak{D}$ is obtained in the usual fashion, through the use of covering sets.

Theorem 6. *Let $\mathfrak{D} \subset \mathbb{R}^q$ be a finite set. Define the sets $\pi^*(\varepsilon) = \{x \in \mathfrak{D} : \|g(x)\| \leq \varepsilon\}, \Pi_n(\delta) = \{x \in \mathfrak{D} : \|G_n(x)\| \leq \delta\}$, and $c(\delta) = \{x \in \mathbb{R}^q : |x^1| \leq \delta_1, |x^2| \leq \delta_2, \ldots, |x^q| \leq \delta_q\}$. Then,*

(i) $\Pr\{\Pi_n(\delta) \not\subseteq \pi^*(\varepsilon)\} \leq |\mathfrak{D}|\exp(-n\tau(\delta,\varepsilon))$ *where* $\tau(\delta,\varepsilon) = \mathrm{Min}_{x\in\mathfrak{D}\setminus\pi^*(\varepsilon)} \inf_{z\in\partial c(\delta)} I(z)$, *$I(z)$ is the rate function governing $G_n(x)$;*
(ii) *Let Y_i, $i = 1, 2, \ldots, n$ be i.i.d. random variables such that $G_n(x) = n^{-1}\sum_{i=1}^n Y_i(x)$. Furthermore, assume*

that there exist a vector of positive constants $\sigma = [\sigma^1, \sigma^2, \ldots, \sigma^q]^T$*, and a* $q \times q$ *correlation matrix* ρ*, satisfying* $M(t) \leq (1/2)\sigma\rho\sigma^T \quad \forall t \in \boldsymbol{R}^q$*, where* $M(t)$ *is the moment-generating function of* $Y_i - \mathrm{E}[Y_i]$*. Then,*

$$n \geq \frac{2\left(\mathrm{Max}(\sigma^1, \sigma^2, \ldots, \sigma^q)\right)^2 \left(1 + \sum_{i \neq j} \rho_{ij}\right)}{(\varepsilon - \delta)^T (\varepsilon - \delta)} \log \frac{|\mathfrak{D}|}{\alpha}$$

implies $\Pr\{\Pi_n(\delta) \not\subseteq \pi^*(\varepsilon)\} \leq \alpha$.

6 REMARKS AND ONGOING WORK

Assertions in Theorems 3, 4, and 5 are illustrated quite beautifully by our ongoing numerical experiments. Not so surprisingly, however, the minimum sample size specified by Theorem 6 turns out to be conservative for a substantial set of problems. Accordingly, we are currently engaged in constructing sequential sampling designs, and a corresponding stopping rule, akin to that in (Bayraksan and Morton 2009) constructed for SOPs. The sequential stopping rule will guarantee that the SAA solution obtained upon stopping has a quality gap that is at most $\varepsilon > 0$, with a probability exceeding $1 - \alpha$.

In addition to the above, our ongoing work focuses on two other important questions (Q.7 and Q.8), both of which will be discussed in the workshop. In answering Q.7, we take a large-deviations viewpoint to optimal sampling. Specifically, predicated on the convergence rate of a specific algorithm in use, we derive the rate at which the probability of the quality-gap exceeding $\varepsilon > 0$ drops to zero, for a given sampling plan. This result raises the next obvious question of identifying a sampling plan that maximizes the said convergence rate, something we are currently investigating. In answering Q.8, again relying largely on the Gärtner-Ellis Theorem (Dembo and Zeitouni 1992), we derive the convergence rates under non-i.i.d sampling. Paralleling work in (Homem-de-Mello 2008) for SOPs, we will specifically discuss the effect of popular schemes such as Latin Hypercube Sampling (Mckay, Conover, and Beckman 1979) and Quasi-Monte-Carlo (Niederreiter 1992) on the rates at which the sequences of quality-gaps and SAA solutions converge to their respective limits.

REFERENCES

Bayraksan, G., and D. P. Morton. 2007. Assessing solution quality in stochastic programs. *Mathematical Programming Series B* 108:495–514.

Bayraksan, G., and D. P. Morton. 2009. A sequential sampling procedure for stochastic programming. *Operations Research*. To Appear.

Dembo, A., and O. Zeitouni. 1992. *Large deviations techniques and applications*. Boston, MA.: Jones and Bartlett.

Homem-de-Mello, T. 2003. Variable-sample methods for stochastic optimization. *ACM Transactions on Modeling and Computer Simulation* 13:108–133.

Homem-de-Mello, T. 2008. On rates of convergence for stochastic optimization problems under non-i.i.d. sampling. *SIAM Journal on Optimization* 19 (2): 524–551.

Mak, W. K., D. P. Morton, and R. K. Wood. 1999. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* 24:47–56.

Mckay, M. D., W. J. Conover, and R. J. Beckman. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245.

Niederreiter, H. 1992. *Random number generation and quasi-Monte-Carlo methods*. Philadelphia, PA.: SIAM.

Pasupathy, R. 2009. On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization. To Appear in *Operations Research*.

Pflug, G. C. 2004. Stochastic optimization and statistical inference. In *Stochastic Programming*, ed. A. Ruszczynski and Shapiro, Handbooks in Operations Research and Management Science, 427–482. Elsevier.

Shapiro, A. 2004. Monte Carlo sampling methods. In *Stochastic Programming*, ed. A. Ruszczynski and Shapiro, Handbooks in Operations Research and Management Science, 353–426. Elsevier.

Shapiro, A., and T. Homem-de-Mello. 2000. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM Journal on Optimization* 11 (1): 70–86.

Yakowitz, S., P. L'Ecuyer, and F. Vazquez-Abad. 2000. Global stochastic optimization with low-dispersion point sets. *Operations Research* 48:939–950.

AUTHOR BIOGRAPHY

RAGHU PASUPATHY is an assistant professor in the Industrial and Systems Engineering department at Virginia Tech. His research interests lie in general simulation methodology, with a current focus on simulation optimization, stochastic root finding, and random vector generation. He serves as an associate editor of ACM TOMACS and the newsletter/communications editor of I-Sim. His publications appear in various journals including Operations Research, INFORMS Journal on Computing, IIE Transactions, and ACM TOMACS. His email and web addresses are < `pasupath@vt.edu` > and < `filebox.vt.edu/users/pasupath/pasupath.htm` > respectively.