A Tale of Two Topics:
(i) SAA Review and (ii) Testbed Update.

Shane Henderson
and
Raghu Pasupathy

24th May 2010

PART I

Problem Statement

$$\begin{aligned} \text{minimize} \quad & g(x) \\ \text{subject to} \quad & h(x) \geq 0, \\ & x \in \mathcal{D} \subset \mathbb{R}^q, \end{aligned}$$

where

- $g : \mathcal{D} \to \mathbb{R}$ can only be estimated using the "black box" estimator G_m, where $G_m(x) \Rightarrow g(x)$ for all $x \in \mathcal{D}$ and m is some measure of simulation effort;
- $h : \mathcal{D} \to \mathbb{R}^n$ can only be estimated using the "black box" estimator H_m, where $H_m(x) \Rightarrow g(x)$ for all $x \in \mathcal{D}$ and m is some measure of simulation effort;
- $\mathcal{D} \subseteq \mathbb{R}^q$ is a known set, e.g., the non-negative orthant.

Notes and Some Notation

- The case of known h has been studied far more.
- The feasible region resulting from the constraints h and the region $\mathcal{D}$ are usually assumed to be closed and convex.
- Denote (π^*, v^*) as the set of global minima and the global minimum value corresponding to the problem. Denote λ^* as the set of local minima (appropriately defined) of the problem.
- Usually an element of π^* or an element of λ^* is requested.

Agenda

What do we cover?

1. Broad overview of SAA and its refinements.
2. Some intuition on where these methods might be successfully applied.
3. Very basic but key theoretical results (relating to convergence, speed of convergence, solution quality, and choice of parameters) that apply in a simulation context.

What do we not cover?

1. SAA for problems where the constraint functions cannot be observed exactly. Specifically, we do not cover chance constrained problems.
2. "Sub-culture specific" results.
3. Results on complexity, epiconvergence, etc.

Sample Average Approximation (SAA)

Logic:

1. "Generate" a sample-path problem with sample size m.
2. Use a procedure to "solve" the sample-path problem

$$\begin{aligned} \text{minimize} \quad & \mathrm{G_m(x)} \\ \text{subject to} \quad & \mathrm{h(x)} \geq 0, \\ & \mathrm{x} \in \mathcal{D} \subset \mathbb{R}^{\mathrm{q}}. \end{aligned}$$

Algorithm Parameters:

(i) procedure for solving the sample-path problems;

(ii) sample size m;

(iii) if sample-path problem can only be solved numerically, the error-tolerance ϵ to within which the sample-path problem should be solved.

SAA Refinement — Retrospective Approximation (RA)

Logic:

1. "Generate" kth sample-path problem with sample-size m_k.
2. Use a procedure to solve the kth sample-path problem to within error-tolerance ϵ_k. Obtain a retrospective solution X_k.
3. $\overline{X}_k = \sum_{j=1}^{k} w_j X_j, w_j \geq 0, \sum_{j=1}^{k} w_j = 1$.
4. Update $k = k + 1$ and goto Step 1.

Algorithm Parameters:

(i) procedure for solving the sample-path problems;

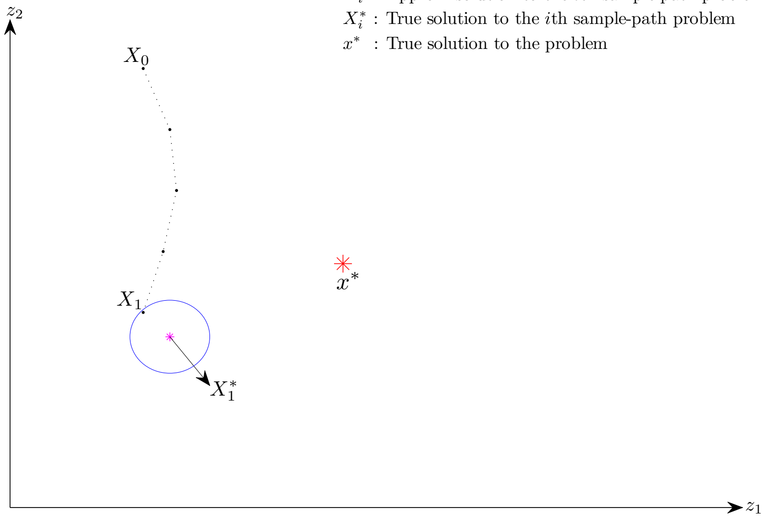
(ii) sample-size sequence $\{m_k\}$;

(iii) error-tolerance sequence $\{\epsilon_k\}$.

Notes: Framework and not an algorithm; "External" vs. "Internal" sampling.

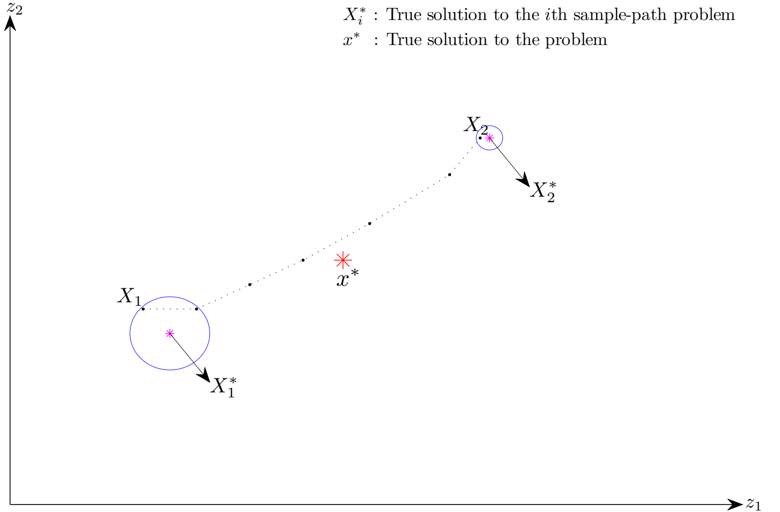
Retrospective Approximation

X_i : Approx. solution to the ith sample-path problem
X_i^* : True solution to the ith sample-path problem
x^* : True solution to the problem



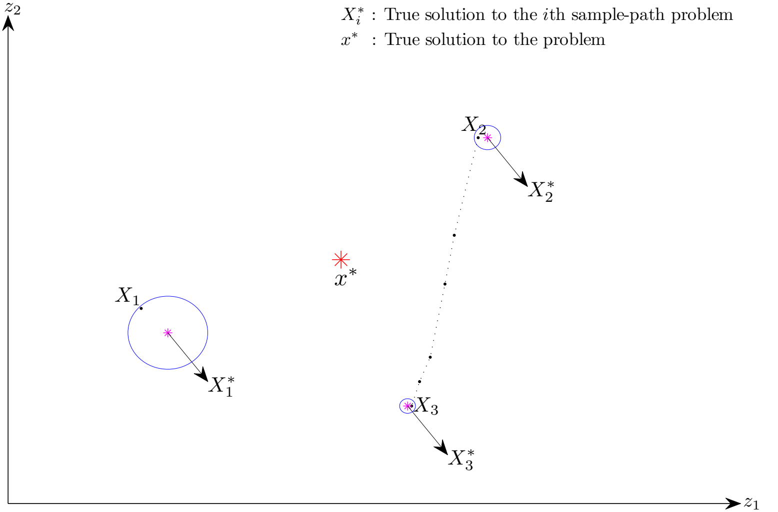
Retrospective Approximation

X_i : Approx. solution to the ith sample-path problem
X_i^* : True solution to the ith sample-path problem
x^* : True solution to the problem



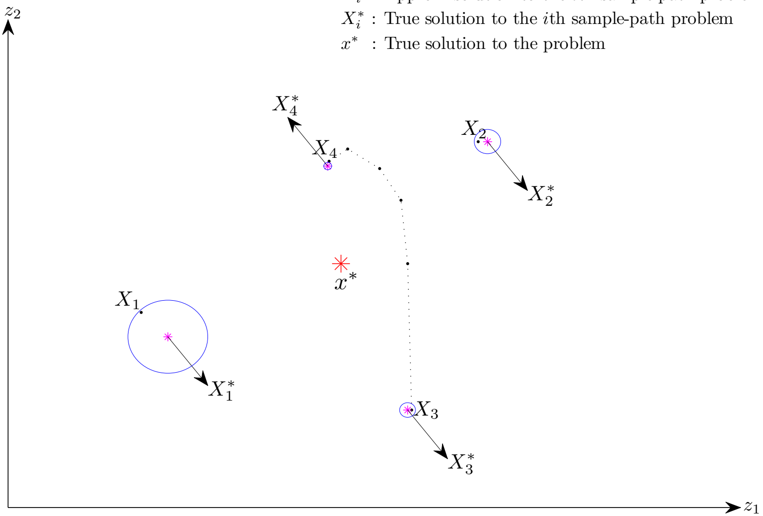
Retrospective Approximation

X_i : Approx. solution to the ith sample-path problem
X_i^* : True solution to the ith sample-path problem
x^* : True solution to the problem



Retrospective Approximation

X_i : Approx. solution to the ith sample-path problem
X_i^* : True solution to the ith sample-path problem
x^* : True solution to the problem



SAA and RA — When?

Advantages

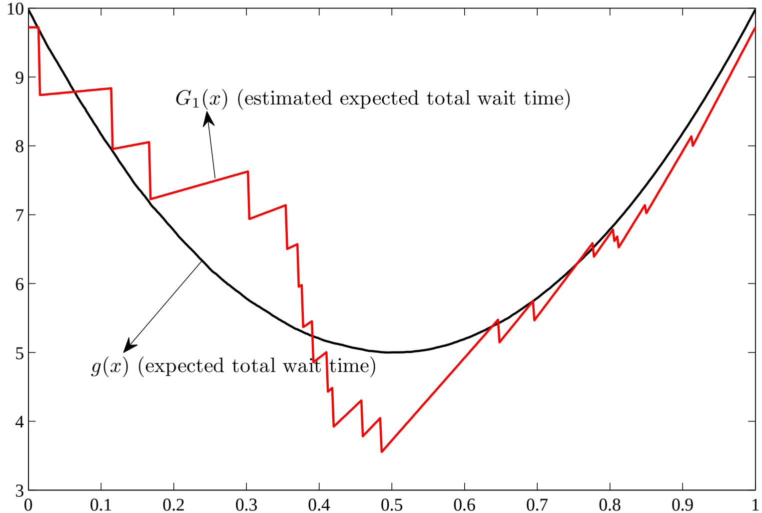
1. Advances in deterministic math. programming at our disposal, in principle.
2. When sample-path problems have structure so that generic search procedures are guaranteed to work well.
3. When sample-path problems have special structure that is known and can be utilized for efficiency. (Surprising counterexample provided by Nemirovski et al. [15].)

Disadvantages

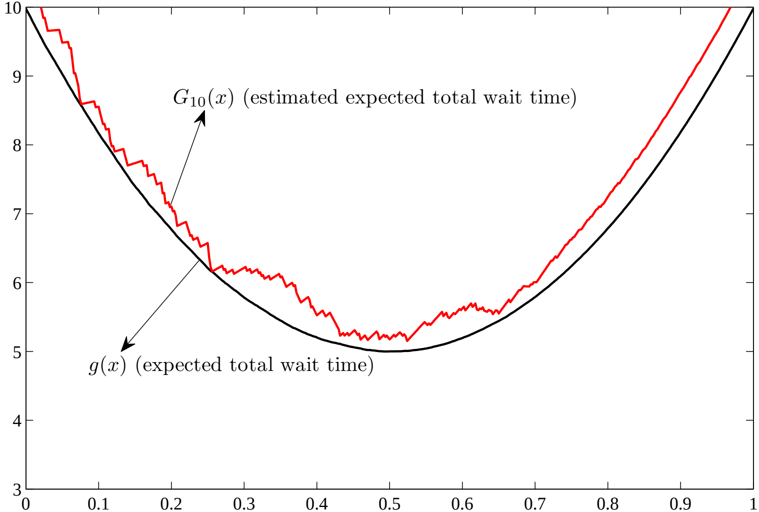
1. When the user cannot be expected to choose an appropriate procedure to solve sample-path problems.
2. Sample-paths are poorly behaved or have no known structure, and so choice of procedure is unclear.

Notes: See Kim and Henderson [10] for some nice relationships between sample-paths and their limit.

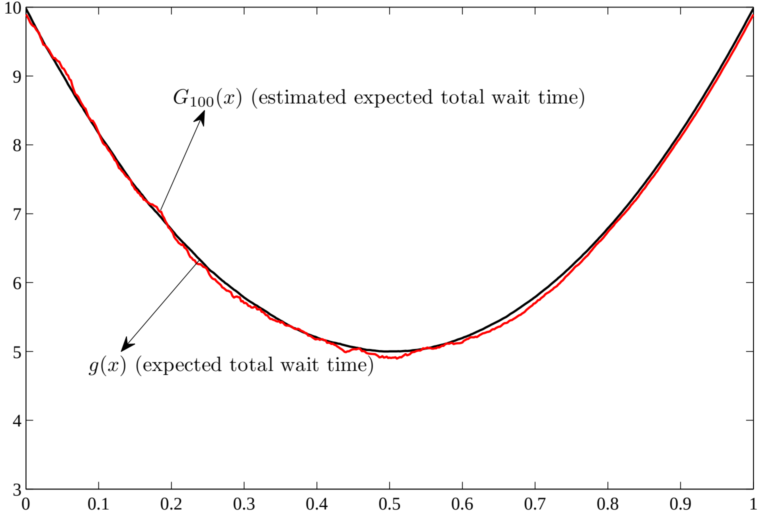
SAA and RA — Sample-Path Structure is Important



SAA and RA — Sample-Path Structure is Important



SAA and RA — Sample-Path Structure is Important



An Outline of Key Results

1. Consistency
   – Convergence of optimal value (SAA and RA).
   – Convergence of optimal solution (SAA and RA).
2. Speed of Convergence
   – CLT-type results for optimal value (SAA and RA).
   – CLT-type results for optimal solutions (SAA and RA).
   – Results under special conditions.
3. Algorithmic Results
   – Minimum sample size results (SAA).
   – Quality of solution/confidence interval type results (SAA).
   – Parameter choice results (RA).

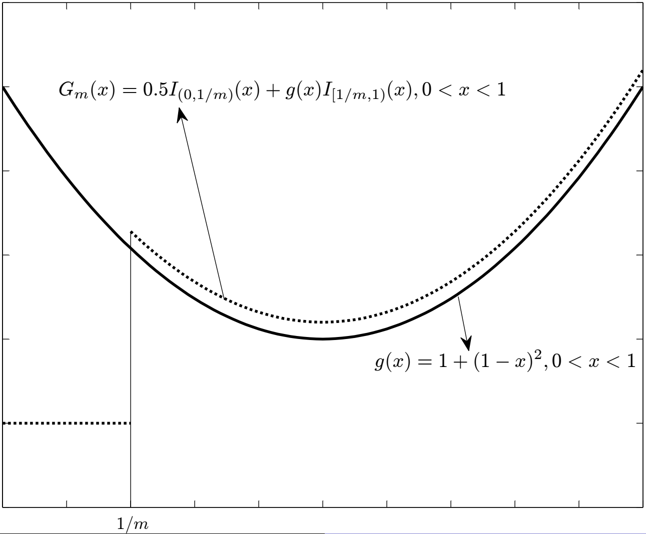
Consistency (SAA and RA)

Theorem (Shapiro [21])

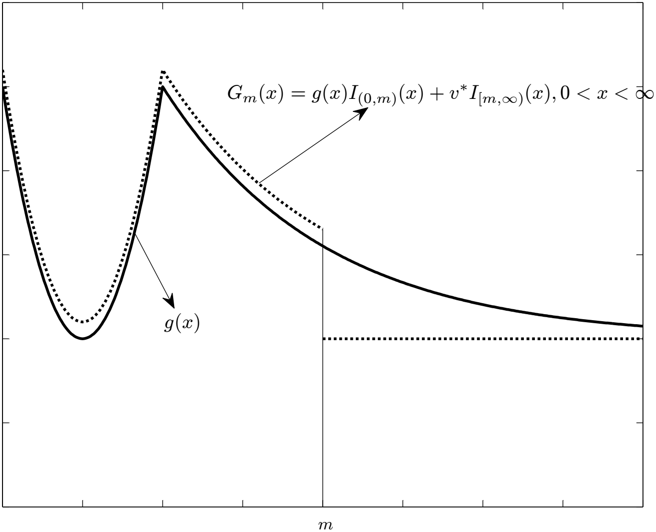
1. If $\lim_{m \to \infty} \sup_{x \in \mathcal{D}} |g(x) - G_m(x)| = 0$ wp1, then $V_m^* \to v^*$ wp1.
2. If $\lim_{m \to \infty} \sup_{x \in \mathcal{D}} |g(x) - G_m(x)| = 0$ wp1, $\mathcal{D}$ is compact, and g is continuous, then $\lim_{m \to \infty} \text{dist}(\Pi_m^*, \pi^*) \to 0$ wp1.

- Uniform convergence verification on a case by case basis. (e.g., if $\mathcal{D}$ is compact, G_m is an iid average that is continuous and dominated by an integrable function, uniform convergence is preserved.)
- Corresponding local results provided by Bastin et al. [1]
- Results on epiconvergence provided by Dupačová and Wets [5], Rockafellar and Wets [19], and Robinson [18].
- Results carry over to RA context easily.

Consistency — Optimal Value



Consistency — Optimal Solution



Speed of Convergence (SAA and RA) — Optimal Value

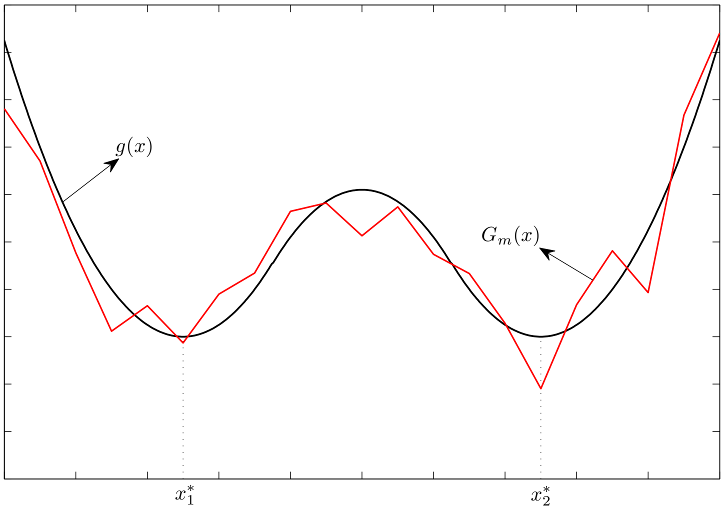
Theorem (Pflug [17], Shapiro [21])

Suppose $\beta(m)$ is a function satisfying $\lim_{m\to\infty} \beta(m) = \infty$ such that $\beta(m)(G_m(x) - g(x)) \Rightarrow Y(x) \in \mathcal{C}(\mathcal{D})$, where $\mathcal{C}(\mathcal{D})$ is the linear space of continuous functions on $\mathcal{D}$. Then,

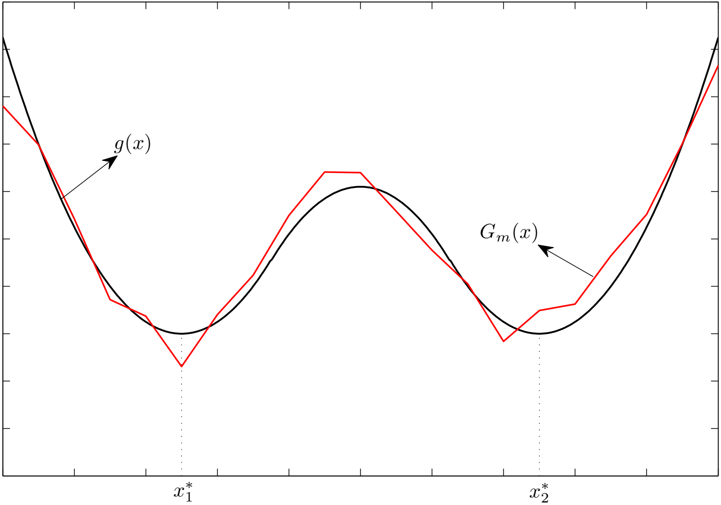
$$\beta(m)(V_m^* - v^*) \Rightarrow \min_{x\in\pi^*} Y(x).$$

- Rate of convergence of "black box" estimator transferred over to optimal value.
- When g is an expectation, the functional CLT condition is satisfied with a Lipschitz condition on G_m, where the Lipschitz constant has finite second moment.
- Corresponding CLT on optimal solution can be found in King and Rockefellar [11], and Shapiro [20].

Speed of Convergence (SAA and RA) — Optimal Value



Speed of Convergence (SAA and RA) — Optimal Value



Speed of Convergence — Important Special Cases

Theorem (Kleywegt et al. [12])

Let $\mathcal{D}$ be finite and $g(x) = E[G(x)]$, $G_m(x) = \sum_{i=1}^{m} Y_i(x)$ where $Y_1, Y_2, \ldots$ are iid copies of a random variable $Y(x)$. Then,

1. $\Pi_m^*(\epsilon) \subset \pi^*$ for large enough m wp1;
2. $\Pr\{\Pi_m^*(\delta) \not\subseteq \pi^*(\epsilon)\} \leq |\mathcal{D} \setminus \pi^*(\epsilon)| \exp\{-m\gamma(\delta,\epsilon)\}$ for $0 \leq \delta \leq \epsilon$, where $\gamma(\delta,\epsilon) = \min_{x \in \pi^* \setminus \pi^*(\epsilon)} I_x(-\delta)$, $I_x(\cdot)$ being the rate function associated with the sequence $\{G_m(x)\}$.

- For large-enough sample size, a true solution will be obtained wp1.
- The probability of not obtaining a true solution (at a specific sample size) drops exponentially in sample size.
- The result in 2. forms the essence of most minimum sample size results [13, 21].

Speed of Convergence — Important Special Cases

Theorem (Shapiro and Homem-de-Mello [22])

Let g be a finite-valued function having a sharp minimum, i.e., g satisfies $g(x) \geq g(x^*) + c\|x - x^*\|$ for $x \in \mathcal{D}$, where c is a positive constant and x^* is the unique minimum. Let $g(x) = E[G(x)]$ for $G_m(x) = \sum_{i=1}^{m} Y_i(x)$, where $Y_1(x), Y_2(x), \ldots$ are iid copies of a random variable $Y(x)$ having fixed finite support. Then, if G is convex, and the set $\mathcal{D}$ is closed and convex, $\Pi_m^* = \{x^*\}$ for large enough m wp1.

- The result is easily extended to contexts where π^* is not a singleton.
- The earlier result on exponential convergence holds in this special case as well.

Results for Solution Quality

Theorem (Mak et al. [14])

Let $g(x) = E[G_m(x)]$ for $G_m(x) = \sum_{i=1}^{m} Y_i(x)$, where $Y_1(x), Y_2(x), \ldots$ are iid copies of a random variable $Y(x)$. Then,

1. $E[V_m^*] \leq v^*$; $E[V_{m+1}^*] \geq E[V_m^*]$;
2. If $x \in \mathcal{D}$, $0 \leq g(x) - v^* \leq g(x) - E[V_m^*]$.

- These results are very general, e.g., $\mathcal{D}$ need not be convex. (See Birge [3], Broadie and Glasserman [4], Higle and Sen [7, 9, 6, 8] for similar results.)
- Mak et al. [14] use the above result to construct confidence intervals on the optimality gap of a candidate solution.
- Bayraksan and Morton [2] extend this to a sequential procedure for constructing confidence intervals.

Parameter Choice in RA

How to choose the sequence of sample sizes $\{m_k\}$, and the sequence of error tolerances $\{\epsilon_k\}$ in RA? Consider the following three conditions.

C.1. When the numerical procedure used to solve sample-path problems exhibits

(a) linear convergence: $\liminf_{k\to\infty} \epsilon_k \sqrt{m_{k-1}} > 0$;

(b) polynomial convergence: $\liminf_{k\to\infty} \frac{\log\left(1/\sqrt{m_{k-1}}\right)}{\log(\epsilon_k)} > 0$.

C.2. $\limsup_{k\to\infty} \left(\sum_{j=1}^{k} m_j\right) \epsilon_k^2 < \infty$.

C.3. $\limsup_{k\to\infty} \left(\sum_{j=1}^{k} m_j\right) m_k^{-1} < \infty$.

Parameter Choice in RA

Theorem (Pasupathy [16])

If the sequences $\{\epsilon_k\}$, $\{m_k\}$ satisfy conditions C.1, C.2, and C.3, and $\pi^* = \{x^*\}$, then $W_k \|X_k - x^*\|^2 = O_p(1)$.

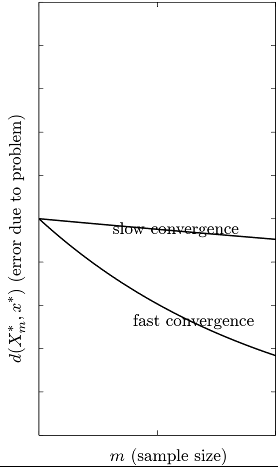
Theorem (Pasupathy [16])

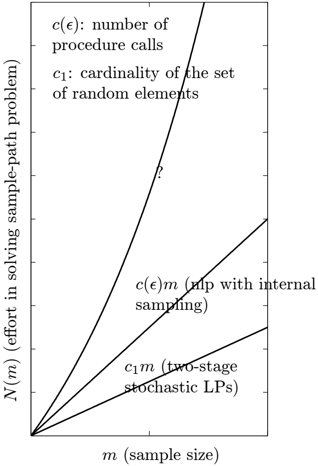
If even one of the conditions C.1, C.2, or C.3 is violated, and $\pi^* = \{x^*\}$, $W_k \|X_k - x^*\|^2 \xrightarrow{p} \infty$.

	Exp. Growth ($m_k = e^{1.1 m_{k-1}}$)	Pol. Growth ($m_k = m_{k-1}^{1.1}$)	Lin. Growth ($m_k = 1.1 m_{k-1}$)
Pol. Conv.	N	Y	Y
Lin. Conv.	N	N	Y
S-Lin. Conv.	N	N	NA

What Are Some "Burning" Questions in SAA/RA?

1. A general theory of optimal sample size increase?





What Are Some “Burning” Questions in SAA/RA? (contd.)

2. In Polyak-Juditsky type averaging, how should we trade-off variance and bias?
3. How to deduce solution quality on global SAA/RA contexts where sample-path problems cannot be solved to optimality easily, i.e., when the methods by Mak et al. [14] are not applicable?
4. Optimal sampling laws in contexts where both the objective function and constraints need to be sampled.

F. Bastin, C. Cirillo, and P. L. Toint.
Convergence theory for nonconvex stochastic programming with an application to mixed logit.
Mathematical Programming, 108:207–234, 2006.

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

J. R. Birge.
The value of the stochastic solution in stochastic linear programs with recourse.
Math. Programming, 24:314–325, 1982.

M. Broadie and P. Glasserman.
Pricing American-style options using simulation.
Journal of Economic Dynamics and Control, 21:1323–1352, 1997.

J. Dupačová and R. J. B. Wets.
Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems.
The Annals of Statistics, 16:1517–1549, 1988.

J. Higle and S. Sen.
Stististical verification of optimality conditions for stochastic programs with recourse.
Annals of Operations Research, 30:215–240, 1991.

J. Higle and S. Sen.
Stochastic decomposition: An algorithm for two-stage stochastic linear programs with recourse.
Mathematics of Operations Research, 16:650–669, 1991.

J. Higle and S. Sen.
Duality and statistical tests of optimality for two stage stochastic programs.
Mathematical Programming, 75:257–275, 1996.

J. Higle and S. Sen.
Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming.
Kluwer, The Netherlands, 1996.

S. Kim and S. G. Henderson.
The mathematics of continuous-variable simulation optimization.
In S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, and J. W. Fowler, editors, Proceedings of the 2008 Winter Simulation Conference, pages 122–132. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2008.

A. J. King and R. T. Rockefellar.
Asymptotic theory for solutions in statistical estimation and stochastic programming.
Mathematics of Operations Research, 18:148–162, 1993.

A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello.
The sample average approximation method for stochastic discrete optimization.
SIAM Journal on Optimization, 12:479–502, 2001.

J. Luedtke and S. Ahmed.
A sample approximation approach for optimization with probabilistic constraints.
SIAM Journal on Optimization, 2008.
To Appear.

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

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro.
Robust stochastic approximation approach to stochastic programming.
SIAM Journal on Optimization, (4), 2009.

R. Pasupathy.
On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization.
2010.
To Appear in Operations Research.

G. C. Pflug.
Optimization of Stochastic Models: The Interface Between Simulation and Optimization.
Kluwer, Boston, MA., 1996.

S. Robinson.
Analysis of sample-path optimization.
Mathematics of Operations Research, 21:513–528, 1996.

R. T. Rockafellar and R. J. B. Wets.
Variational Analysis.
Springer, Berlin, 1998.

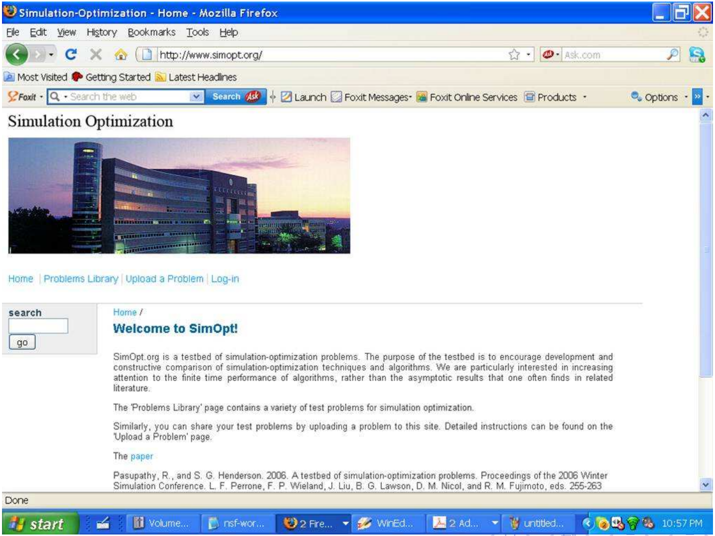
A. Shapiro.
Statistical inference of stochastic optimization problems.
In S. P. Uryasev, editor, Probabilistic Constrained Optimization: Methodology and Applications, pages 282–304. Kluwer Academic Publishers, 2000.

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

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

PART II

Introducing ... an SO Testbed

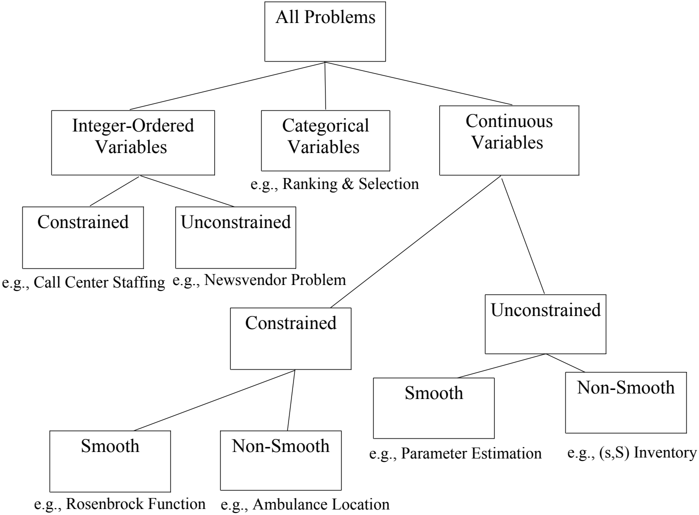


Recall Objectives

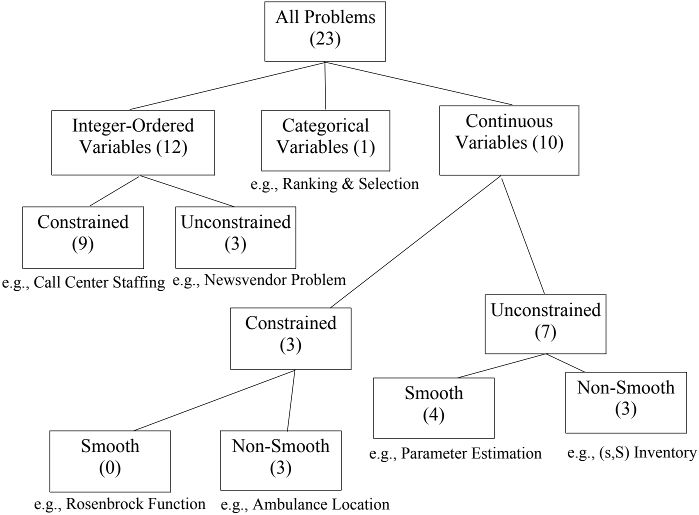
- Fill the stated and yet unfulfilled need for a carefully designed testbed of SO problems.

  (The stochastic programming community has a few of its own libraries, e.g., SIPLIB for stochastic integer programs, POSTS for linear recourse problems.)
- Actively draw attention to finite-time performance of algorithms, through the use of finite-time performance measures.
- Identify particular problem types that defy efficient solution.
- Increase visibility and usage of SO formulation and solution.

Problem Organization Within Testbed



Where do we stand?



Where do we stand and what's next?

1. Good: The testbed is close to having a critical mass of problems.
2. Bad: Not as much variety as we would like to see.
3. Bad: Small number of continuous-variable and categorical-variable problems.

1. One PhD and one undergrad comb past WSC proceedings.
2. Large source of potential continuous-variable problems: Approximate DP.

Group Discussion

1. What does the apparent dearth of problems tell us?
2. Should particular categories be coalesced?
3. Is it time yet to launch the testbed?
4. To submit or simply view, point your browser to www.simopt.org.