Simulation Optimization and Optimal Sampling for Stochastically Constrained Systems

Raghu Pasupathy {pasupath@vt.edu}
Virginia Tech

Acknowledgements:
Susan Hunter, Cornell University
Loo Hay Lee, National University of Singapore
Nugroho Pujowidianto, National University of Singapore
Chun-Hung Chen, George Mason University

October 17, 2012

Simulation-Optimization (SO) Problem Statement

"Solve an optimization problem where the objective functions/constraints have to be sampled."

$$\begin{aligned} \text{minimize} \quad & h(x) \\ \text{subject to} \quad & g(x) \le 0, x \in \mathcal{D}; \end{aligned}$$

where

- $h : \mathcal{D} \to \mathbb{R}$ can only be estimated using $H_m(x) = m^{-1} \sum_{i=1}^{m} H_j(x)$, where $H_j(x)$ are iid random variables with mean $h(x)$;
- $g : \mathcal{D} \to \mathbb{R}^c$ can only be estimated using $G_m = m^{-1} \sum_{i=1}^{m} G_j(x)$, where $G_j(x)$ are iid random vectors with mean $g(x)$; and
- $\mathcal{D} \subseteq \mathbb{R}^q$ is some region.

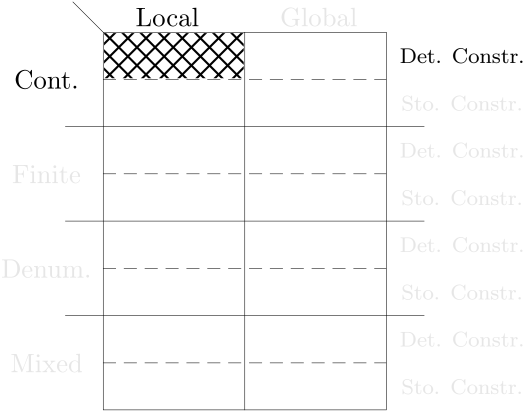
SO Examples

Visit the simulation optimization library at
`http://www.simopt.org`.

SO — Where do we stand?

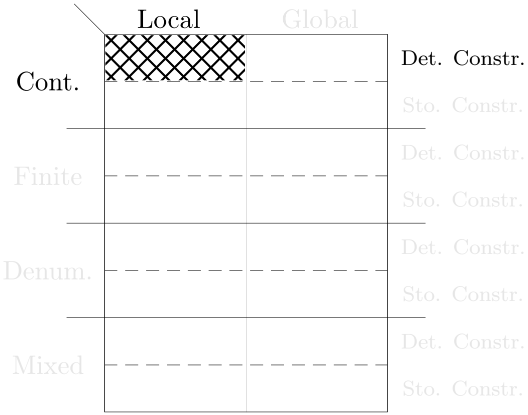
	Local	Global	
Cont.			Det. Constr.
			Sto. Constr.
Finite	NA		Det. Constr.
	NA		Sto. Constr.
Denum.			Det. Constr.
			Sto. Constr.
Mixed			Det. Constr.
			Sto. Constr.

SO — Where do we stand?



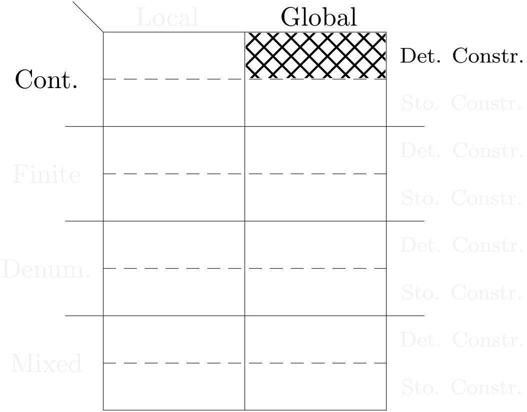
- Stochastic Approximation (SA) and Sample-Average Approximation (SAA) are the main algorithm classes.
- SA has an enormous amount of literature dating back to 1951 — Robbins and Monro's paper [31]. Excellent survey articles and books are widely available, e.g., [23, 7, ?].
- SA has had many resurgences, e.g., after 1997 paper by Polyak and Juditsky [30]. Most current work has been on

SO — Where do we stand?

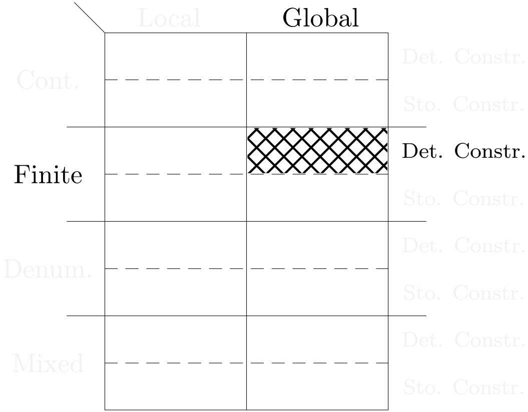


- SAA appeared around 1991 [15, 34] as a way to exploit advances in nlp and sample-path structure. A number of refinements are popular now [17, 28].
- Most current work is on dynamic sample-sizing, parameter choice, and solution quality estimation [33, 32, 28, 5, 6].

SO — Where do we stand?

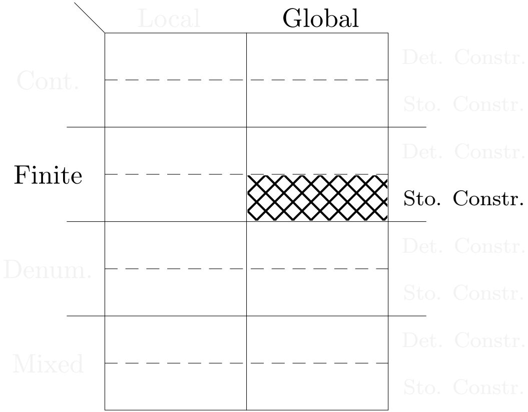


SO — Where do we stand?



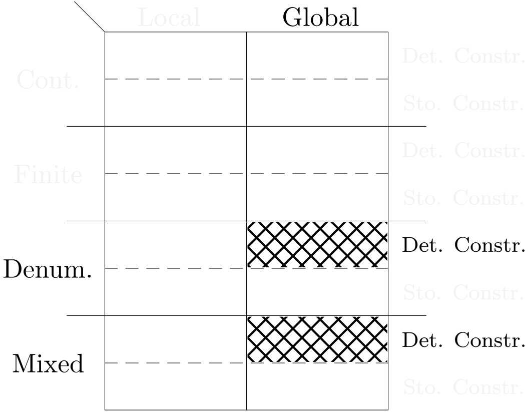
- Very mature existing theory and solution algorithms, see [16, 22]. Ready software is publicly available.
- Ongoing research is mostly on variations, e.g., incorporation of correlated sampling and crn [13, 38], incorporation of economics [10, 9, 11], other efficiencies [36, 14].

SO — Where do we stand?



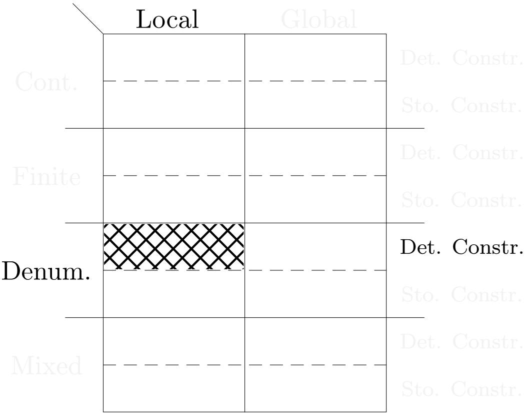
- This question is relatively new, with a surge in recent work [3, 2, 4, 20, 35]
- Generally, ongoing work is focused on appropriate treatment of stochastic constraints[27], optimal budget allocation[20, 21], and finite-time probabilistic guarantees [3, 2].

SO — Where do we stand?



- This question, like finite SO, has an enormous amount of existing literature, see [1] for an overview.
- Algorithms usually involve three steps: sampling candidate solution(s); estimating objective function; and update sampling strategy and relevant estimators.
- Ongoing research is predominantly about balancing exploration and exploitation (in various

SO — Where do we stand?



- Relatively specialized but important problem class as reflected by the fraction of submissions in the simulation library.
- The main (specialized) algorithms are COMPASS [18, 19, 39, 40],R-SPLINE [37], and discretized SA [24].
- A strong need for extensions to handle stochastic constraints.

SO Flavor of the Day

- The region $\mathcal{D}$ is finite but "large," and is categorical.
- Stochastic constraints are allowed.
- We seek a global minimizer.

SO Flavor of the Day (in more convenient notation)

We consider

$$\begin{aligned} \arg \min_{i=1,\ldots,k} \quad & h_i \\ \text{s.t.} \quad & g_{il} \leq \gamma_l, \;\; \text{for all } i = 1, \ldots, k \text{ and } l = 1, \ldots, s \end{aligned}$$

where k is a finite number of systems, s is a finite number of constraints, and

- design 1 is the optimal design,
- h_i and g_{il} are unknown expectations,
- estimates $\bar{H}_i$ of h_i and $\bar{G}_{il}$ of g_{il} may observed through simulation as iid sample means of random variables H_i and G_{il}, respectively,
- γ_l is a vector of known constants, and
- a unique solution exists.

Solution Context and Main Questions

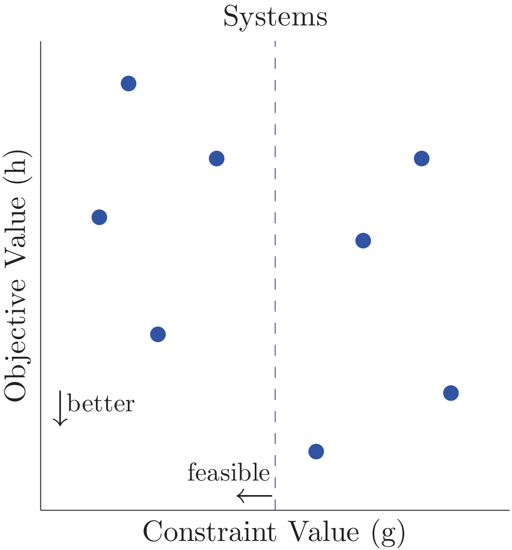
Solution Context:

1. System i is given fraction $\alpha_i \geq 0$ of the total budget t. Sample and construct estimators $(\overline{H}_i, \overline{G}_{il}), i \in \{1, 2, \ldots, k\}; l \in \{1, 2, \ldots, s\}$.
2. The estimated optimal system is $\hat{1} = \{i : i \in \hat{\Gamma}, \overline{H}_i \leq \overline{H}_j \text{ for all } j \in \hat{\Gamma}\}$ where $\hat{\Gamma} = \{i : \overline{G}_{il} \leq \gamma_l \text{ for } l \in \{1, 2, \ldots, s\}\}$.

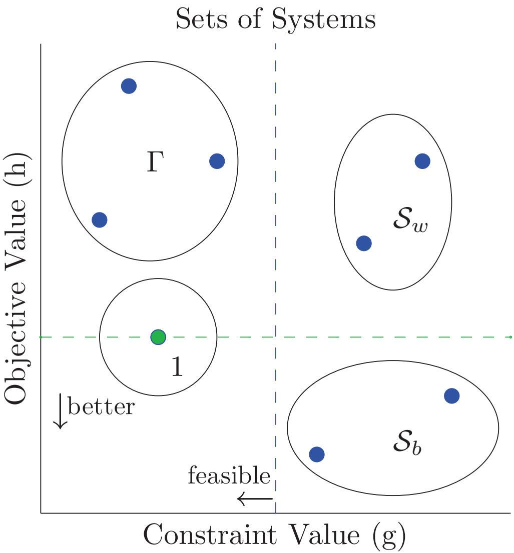
The Main Question:

What allocation vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ minimizes the probability of false selection $P(FS) = \Pr\{\hat{1} \neq 1\}$?

Understanding Probability of False Selection P(FS)



Understanding Probability of False Selection P(FS)



Understanding Probability of False Selection P(FS)

$$P(FS) = P\left(\overbrace{\left(\cup_{l=1}^{s} \overline{G}_{1l} > \gamma_l\right)}^{\text{best estimated infeasible}} \cup \underbrace{\left(\cup_{i \neq 1} \left(\cap_{l=1}^{s} \overline{G}_{il} \leq \gamma_l\right) \cap \left(\overline{H}_1 > \overline{H}_i\right)\right)}_{\text{best beaten by a system that is estimated to be feasible}} \right).$$

Main Question (Restatement)

(i) Answering the question of identifying $\alpha_i, i \in \{1, 2, \ldots, k\}$ such that P(FS) is minimized is in general very difficult.

(ii) Any allocation such that $\alpha_i > 0$ will ensure $P(FS) \to 0$ as $t \to \infty$.

Noting (i) and (ii), we ask:

What allocation vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ maximizes the rate of decay of P(FS) to zero?

10-minute Primer on Large Deviations

Let $\{X_i\}$ be iid random variables with $E[e^{tX_1}] < \infty$ for all t. Let $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any set $\mathcal{A}$, we know that

$$\lim_{n\to\infty} \Pr\{\overline{X}(n) \in \mathcal{A}\} \quad = \quad 0 \quad \text{if } E[X_1] \notin \mathcal{A}.$$

Cramér's Theorem [12] allows us to say more.

$$\Pr\{\overline{X}(n) \in \mathcal{A}\} \approx e^{-nI(x^*)}.$$

For (Borel measurable) sets $\mathcal{A} \subset \mathbb{R}$ with $E[X_1] \notin \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{\overline{X}(n) \in \mathcal{A}\} = \inf_{x\in\mathcal{A}} I(x) = I(x^*),$$

where $I(\cdot)$ is called the rate function of iid averages of X_i.

10-minute Primer on Large Deviations

Let $\{X_i\}$ be iid random variables with $E[e^{tX_1}] < \infty$ for all t. Let $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any set $\mathcal{A}$, we know that

$$\lim_{n\to\infty} \Pr\{\overline{X}(n) \in \mathcal{A}\} \quad = \quad 0 \quad \text{if } E[X_1] \notin \mathcal{A}.$$

Cramér's Theorem [12] allows us to say more.

$$\Pr\{\overline{X}(n) \in \mathcal{A}\} \approx e^{-nI(x^*)}.$$

For (Borel measurable) sets $\mathcal{A} \subset \mathbb{R}$ with $E[X_1] \notin \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{\overline{X}(n) \in \mathcal{A}\} = \inf_{x\in\mathcal{A}} I(x) = I(x^*),$$

where $I(\cdot)$ is called the rate function of iid averages of X_i.

10-minute Primer on Large Deviations

Let $\{X_i\}$ be iid random variables with $E[e^{tX_1}] < \infty$ for all t. Let $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any set $\mathcal{A}$, we know that

$$\lim_{n\to\infty} \Pr\{\overline{X}(n) \in \mathcal{A}\} \quad = \quad 0 \quad \text{if } E[X_1] \notin \mathcal{A}.$$

Cramér's Theorem [12] allows us to say more.

$$\Pr\{\overline{X}(n) \in \mathcal{A}\} \approx e^{-nI(x^*)}.$$

For (Borel measurable) sets $\mathcal{A} \subset \mathbb{R}$ with $E[X_1] \notin \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pr\{\overline{X}(n) \in \mathcal{A}\} = \inf_{x\in\mathcal{A}} I(x) = I(x^*),$$

where $I(\cdot)$ is called the rate function of iid averages of X_i.

10-minute Primer on Large Deviations

Let $\{X_i\}$ be iid random variables with $E[e^{tX_1}] < \infty$ for all t. Let $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any set $\mathcal{A}$, we know that

$$\lim_{n\to\infty} \Pr\{\overline{X}(n) \in \mathcal{A}\} \quad = \quad 0 \quad \text{if } E[X_1] \notin \mathcal{A}.$$

Cramér's Theorem [12] allows us to say more.

$$\Pr\{\overline{X}(n) \in \mathcal{A}\} \approx e^{-nI(x^*)}.$$

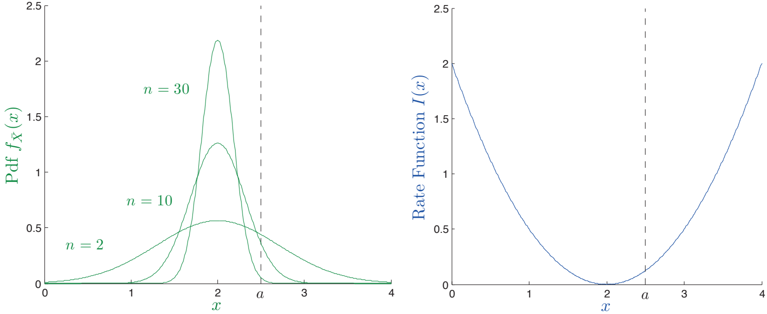
For (Borel measurable) sets $\mathcal{A} \subset \mathbb{R}$ with $E[X_1] \notin \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{\overline{X}(n) \in \mathcal{A}\} = \inf_{x\in\mathcal{A}} I(x) = I(x^*),$$

where $I(\cdot)$ is called the rate function of iid averages of X_i.

Example. For X_i iid normal$(\mu = 2, \sigma^2 = 1)$ and $\mu < a = 2.5$,

$$-\lim_{n\to\infty} \frac{1}{n} \log P\{\bar{X} \in [a, \infty)\} = I(a) = \frac{1}{2}\left(\frac{a-\mu}{\sigma}\right)^2 = 0.125$$



10-minute Primer on Large Deviations

Cramér's Theorem [12] holds in $\mathbb{R}^d$ as well. Suppose $(\overline{X}(n), \overline{Y}(n))$ is constructed as iid averages of X_i, Y_i with $E[e^{sX_i+tY_i}] < \infty$.

Then, for (Borel measurable) set $\mathcal{A} \subset \mathbb{R}^2$ with $(E[X_1], E[Y_1]) \notin \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{(\overline{X}(n), \overline{Y}(n)) \in \mathcal{A}\} = \inf_{(x,y)\in\mathcal{A}} I(x, y) = I(x^*, y^*),$$

where $I(\cdot, \cdot)$ is called the rate function of iid averages of (X_i, Y_i). (Interpret above as $\Pr\{\overline{X}(n), \overline{Y}(n) \in \mathcal{A}\} \approx e^{-nI(x^*,y^*)}$.)

10-minute Primer on Large Deviations

Suppose $E[X_i] < E[Y_i] < \gamma$, and we want to calculate the rate at which $\Pr\{\overline{X}(n) > \overline{Y}(n), \overline{Y}(n) > \gamma\}$.

Then the above probablity can be written as $\Pr\{\overline{X}(n) - \overline{Y}(n) > 0, \overline{Y}(n) > \gamma\}$ giving the rate

$$\lim_{n \to \infty} -\frac{1}{n} \log \Pr\{\overline{X}(n) - \overline{Y}(n) > 0, \overline{Y}(n) > \gamma\} = \inf_{z>0, y>\gamma} I(z, y),$$

where $I(\cdot, \cdot)$ is the rate function of iid averages of $(X_i - Y_i, Y_i)$.

Rate of Decay of P(FS)

The decay rate (to zero) of the probability of false selection is:

$$-\lim_{t\to\infty}\frac{1}{t}\log P(FS) = \min\left(\min_{l\in\{1,2,\ldots,s\}}\alpha_1 J_{1l}(\gamma_l), \min_{i\neq 1} R_i(\alpha_1,\alpha_i)\right)$$

where

- $J_{1,l}, l \in \{1,2,\ldots,s\}$ is the rate of decay of the best system being deemed infeasible;
- $R_i(\alpha_1,\alpha_i)$ is the rate of decay of the ith system being deemed feasible and beating the best system; and
- $$R_i(\alpha_1,\alpha_i) = \inf_{x_i\leq x_1, y_i\leq\gamma}\{\alpha_1 I_1(x_1) + \alpha_i I_i(x_i,y_i)\}.$$

Back to the Main Question

What should the α_i's be to maximize the rate of decay of the probability of false selection?

$$\max_{\alpha_1, \ldots, \alpha_k} \min \left(\min_{l \in \{1,\ldots,s\}} \alpha_1 J_{1l}(\gamma_l) \,, \; \min_{i \neq 1} R_i(\alpha_1, \alpha_i) \right)$$
$$\text{subject to } \sum_{i=1}^{k} \alpha_i = 1, \alpha \geq 0. \quad (1)$$

An Equivalent Reformulation

What should the α_i's be to maximize the rate of decay of the probability of false selection?

$$
\begin{aligned}
\max \quad z \quad \text{s.t.} \\
\alpha_1 J_{1j}(\gamma_j) \geq z,\ j = 1, 2, \ldots, l \\
R_i(\alpha_1, \alpha_i) \geq z,\ i = 2, 3, \ldots, k \\
\sum_{i=1}^{r} \alpha_i = 1,\ \alpha_i \geq 0.
\end{aligned} \tag{2}
$$

Characterization of the Exact Solution

After writing the KKT conditions, the optimal fractions $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*)$ are obtained as the unique solution to the following system.

$$\begin{aligned}
\alpha_1^* J_{1,l}(\gamma_l) &\geq z^*, \quad l \in \{1, 2, \ldots, s\}; \\
R_i(\alpha_1, \alpha_i) &= z^*, \quad i \neq 1; \\
\sum_{i \neq 1} \frac{\partial R_i(\alpha_1^*, \alpha_i^*)/\partial \alpha_1}{\partial R_i(\alpha_1^*, \alpha_i^*)/\partial \alpha_i} &= 1.
\end{aligned}$$

As the number of systems tend to ∞ ...

As $|\Gamma^*| + |\mathcal{S}_{\mathrm{w}}^*| \to \infty$, the following hold.

(i)

$$\frac{\alpha_i^*}{\alpha_1^*} \to 0 \quad \forall i \neq 1.$$

(ii)

$$\frac{R_i(\alpha_1^*, \alpha_i^*)}{\alpha_i^*} = \overbrace{\inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i)}^{\text{score } S_i}.$$

As the number of systems tend to ∞ ...

As $|\Gamma^*| + |\mathcal{S}_w^*| \to \infty$, the following hold.

(i)

$$\frac{\alpha_i^*}{\alpha_1^*} \to 0 \quad \forall i \neq 1.$$

(ii)

$$\frac{R_i(\alpha_1^*, \alpha_i^*)}{\alpha_i^*} = \overbrace{\inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i)}^{\text{score } S_i}.$$

The Proposed Solution

Recall that the KKT conditions dictate equating the rates $R_i(\alpha_1^*, \alpha_i^*)$ for $i \neq 1$. Using this and the previous result, we see that

$$\alpha_i^* \overbrace{\left(\inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i) \right)}^{\text{score } S_i} \approx \alpha_j^* \overbrace{\left(\inf_{x_j \leq h_1, y_j \leq \gamma} I_j(x_j, y_j) \right)}^{\text{score } S_j}, \quad i, j \neq 1.$$

Proposed Solution:

Choose allocations $\alpha_j, j = 2, 3, \ldots, k$ such that

$$\alpha_j \propto S_j^{-1},$$

where $S_j = \inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i)$.

The Proposed Solution

Recall that the KKT conditions dictate equating the rates $R_i(\alpha_1^*, \alpha_i^*)$ for $i \neq 1$. Using this and the previous result, we see that

$$\alpha_i^* \overbrace{\left(\inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i) \right)}^{\text{score } S_i} \approx \alpha_j^* \overbrace{\left(\inf_{x_j \leq h_1, y_j \leq \gamma} I_j(x_j, y_j) \right)}^{\text{score } S_j}, \quad i, j \neq 1.$$

<u>Proposed Solution</u>:

Choose allocations $\alpha_j, j = 2, 3, \ldots, k$ such that

$$\alpha_j \propto S_j^{-1},$$

where $S_j = \inf_{x_i \leq h_1, y_i \leq \gamma} I_i(x_i, y_i)$.

Two Examples.

1. If estimators are mutually independent normals,

$$S_j = \overbrace{\frac{1}{2}\frac{(h_j - h_1)^2}{\sigma^2}\mathbb{I}\{h_i > h_1\}}^{\text{suboptimality penalty}} + \sum_{l=1}^{s} \overbrace{\frac{1}{2}\frac{(g_{jl} - \gamma_l)^2}{\sigma_l^2}\mathbb{I}\{g_{il} > \gamma_l\}}^{\text{constraint violation penalty}}.$$

2. If estimators are mutually independent Bernoullis,

$$S_j = E(h_1, h_i)\mathbb{I}\{h_i > h_1\} + \sum_{l=1}^{s} E(g_{il}, \gamma_l)\mathbb{I}\{g_{il} > \gamma_l\},$$

where $E(a, b) = a \log \frac{a}{b} + (1 - a) \log \frac{1-a}{1-b}$.

Implementation

Outline of a sequential algorithm:

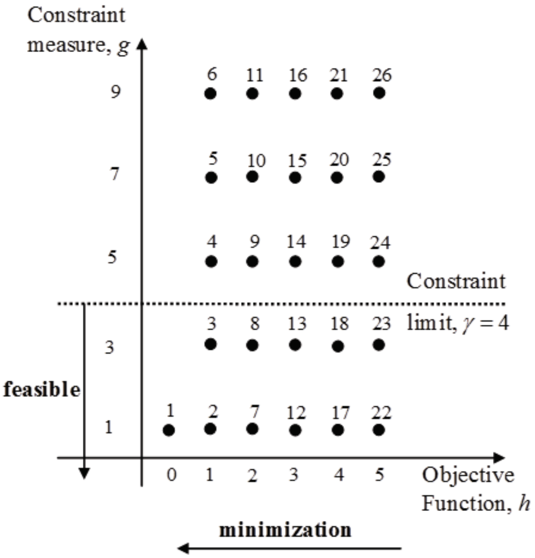
1. Collect δ_0 observations from each system $i \leq k$.
2. Set $n = r \times \delta_0$.
3. Update the estimators $\overline{H}_i, \overline{G}_{il}$ for $i \leq k, j \leq s$, the feasible set estimator $\hat{\Gamma}$, and the optimal solution estimator $\hat{1}$.
4. Update the score function estimators $\hat{S}_i, i \neq 1$ and the optimal allocations $\hat{\boldsymbol{\alpha}}^* = (\alpha_1, \alpha_2, \ldots, \alpha_k)$.
5. Use $\hat{\boldsymbol{\alpha}}^*$ as a sampling distribution from which to collect the next δ samples.
6. Set $n = n + \delta$ and go to step 3.

Numerical Example

Problem Design:

1. Objective and constraint function estimators are mutually independent and normal.
2. Number of constraints $s = 1$.
3. Number of systems $k = 401, 901, 1601, 2501, 3601$.
4. $h_1 = 0, g_{1,1} = 1$.
5. $|\Gamma| = 0.4(k-1)+1$, $\gamma = 2(|\Gamma|-1)/\sqrt{k-1}$;
6. Variance parameters $\sigma^2 = \sigma_1^2 = 9$.

Numerical Example



Numerical Example

Optimality gap and computation times for equal allocation (EA), proposed solution (CF), and exact solution (*).

k				
	$\Delta z(\alpha_{EA})$	$\Delta z(\alpha_{CF})$	Time(α_{CF})	Time(α^*)
26	5.661	0.94	0.01 s	0.978 s
101	3.926	0.488	0.011 s	1.526 s
401	2.453	0.227	0.014 s	9.785 s
901	1.807	0.140	0.019 s	54.809 s
1,601	1.439	0.099	0.027 s	227.746 s
2,501	1.195	0.069	0.037 s	615.115 s
3,601	N/A	N/A	0.048 s	N/A

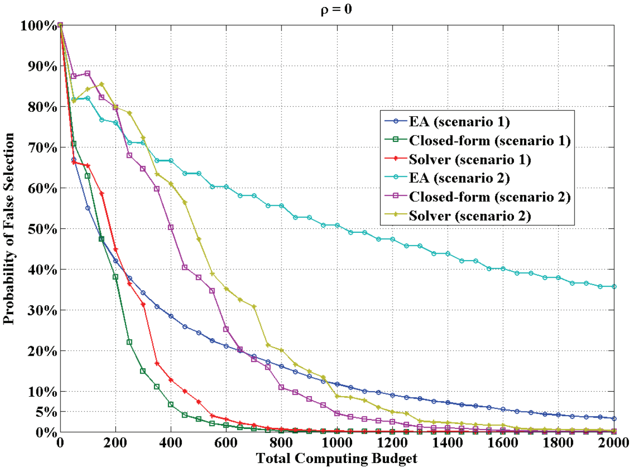
Numerical Example

The effect of constraints.

s	k = 901			
	$\Delta z(\alpha_{EA})$	$\Delta z(\alpha_{CF})$	$\text{Time}(\alpha_{CF})$	$\text{Time}(\alpha^*)$
1	1.807	0.140	0.019 s	54.809 s
5	1.907	0.134	0.031 s	1,691.612 s
10	1.933	0.131	0.047 s	1,696.179 s

Numerical Example

Probability of false selection as a function of the budget for k = 26 and k = 101.



Concluding Remarks

1. We propose a simple solution for solving constrained SO problems on large finite sets using score functions.
2. The score function is very easy to compute in many cases, particularly when the underlying distributions are known or assumed.
3. In general, this work should be seen as providing a theoretical basis for allocation using a model.
4. Very large constrained SO problems have recently been solved with surprising ease. (For example, a problem with $20,000$ systems and 100 constraints was solved recently within about 20 seconds.)
5. The proposed solution might have ramifications for continuous global simulation optimization, particularly when using many processors.

S. Andradóttir.
An overview of simulation optimization via random search.
In S. G. Henderson and B. L. Nelson, editors, Simulation, Handbooks in Operations Research and Management Science, pages 617–631. Elsevier, 2006.

S. Andradóttir, D. Goldsman, and S. -H. Kim.
Finding the best in the presence of a stochastic constraint.
In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, Proceedings of the 2005 Winter Simulation Conference, pages 732–738. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey.

S. Andradóttir and S. -H. Kim.
Fully sequential procedures for comparing constrained systems via simulation.
Naval Research Logistics, (57):403–421, 2010.

D. Batur and S. -H. Kim.

Procedures for feasibility detection in the presence of multiple constraints.
In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, Proceedings of the 2005 Winter Simulation Conference, pages 692–698. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey.

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

G. Bayraksan and D. P. Morton.
A sequential sampling procedure for stochastic programming.
Operations Research, 59(4):898–913, 2009.

V. S. Borkar.
Stochastic Approximation: A Dynamical Systems Viewpoint.
Cambridge University Press, Cambridge, UK, 2008.

M. Broadie, D. M. Cicek, and A. Zeevi.
General bounds and finite-time improvement for the kiefer-wolfowitz stochastic approximation algorithm.
Operations Research, 2010.
To appear.

S. E. Chick, J. Branke, and C. Schmidt.
Sequential sampling to myopically maximize the expected value of information.
INFORMS Journal on Computing, 22(1):71–80, 2010.

S. E. Chick and P. Frazier.
Sequential sampling with economics of selection procedures.
Management Science, 58(3):550–569, 2012.

S. E. Chick and N. Gans.
Economic analysis of simulation selection options,.
Management Science, 55(3):421–437, 2009.

A. Dembo and O. Zeitouni.

Large Deviations Techniques and Applications.
Springer-Verlag, New York, NY., 1998.

- M. C. Fu, J.-Q. Hu, C.-H. Chen, and X. Xiong.
Simulation allocation for determining the best design in the presence of correlated sampling.
INFORMS Journal on Computing, 19(1):101–111, 2007.

- C. Healy, D. Goldsman, and S. -H. Kim.
Ranking and selection techniques with overlapping variance estimators.
Sequential Analysis, 28(4):459–474, 2009.

- K. Healy and L. W. Schruben.
Retrospective simulation response optimization.
In B. L. Nelson, D. W. Kelton, and G. M. Clark, editors, Proceedings of the 1991 Winter Simulation Conference, pages 954–957. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 1991.

- S. G. Henderson and B. L. Nelson, editors.

volume 13 of Handbooks in Operations Research and Management Science: Simulation.
Elsevier, 2006.

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

J. Hong and B. L. Nelson.
Discrete optimization via simulation using compass.
Operations Research, 54(1):115–129, 2006.

L. J. Hong, B. L. Nelson, and Xu.
Speeding up compass for high-dimensional discrete optimization via simulation.
Operations Research Letters, (38):550–555, 2010.

S. R. Hunter and R. Pasupathy.
Optimal sampling laws for stochastically constrained simulation optimization.

INFORMS Journal on Computing, 2011.
To appear.

S. R. Hunter, N. A. Pujowidianto, L. H. Lee, C. H. Chen, and R. Pasupathy.
Optimal sampling laws for constrained simulation optimization on finite sets: The bivariate normal case.
In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, editors, Proceedings of the 2011 Winter Simulation Conference. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2011.

Seong-Hee Kim and B. L. Nelson.
Selecting the best system.
In S. G. Henderson and B. L. Nelson, editors, Simulation, Handbooks in Operations Research and Management Science, pages 501–534. Elsevier, 2006.

H. J. Kushner and G. G. Yin.

Stochastic Approximation and Recursive Algorithms and Applications.
Springer-Verlag, New York, NY., 2003.

E. Lim.
Stochastic approximation over multidimensional discrete sets with applications to inventory systems and admission control of queueing networks.
ACM TOMACS.
To appear.

A. Mokkadem and M. Pelletier.
A generalization of the averaging procedure: The use of two-time-scale algorithms.
SIAM Journal on Control and Optimization, 49:1523, 2011.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro.
Robust stochastic approximation approach to stochastic programming.
SIAM Journal on Optimization, 19(4):1574–1609, 2009.

C. Park and S. -H. Kim.
Handling stochastic constraints in discrete optimization via simulation.
In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, editors, Proceedings of the 2011 Winter Simulation Conference. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2011.

R. Pasupathy.
On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization.
Operations Research, 58:889–901, 2010.

R. Pasupathy and B. W. Schmeiser.
DARTS — dynamic adaptive random target shooting.
In B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, editors, Proceedings of the 2010 Winter

Simulation Conference. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, 2010.

B. T. Polyak and A. B. Juditsky.
Acceleration of stochastic approximation by averaging.
SIAM Journal on Control and Optimization, 30(4):838–855, 1992.

H. Robbins and S. Monro.
A stochastic approximation method.
Annals of Mathematical Statistics, 22:400–407, 1951.

J. Royset.
On sample size control in sample average approximations for solving smooth stochastic programs.
Journal of Computational and Applied Mathematics, 2011. Under Review.

J. Royset and R. Szechtman.
Optimal budget allocation for sample average approximation.

Operations Research, 2011.
Under Review.

R. Y. Rubinstein and A. Shapiro.
Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method.
John Wiley & Sons, New York, NY., 1993.

R. Szechtman and E. Yüesan.
A new perspective on feasibility determination.
In S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, editors, Proceedings of the 2008 Winter Simulation Conference, pages 273–280. Institute of Electrical and Electronics Engineers: Piscataway, New Jersey.

S. C. Tsai and B. L. Nelson.
Fully sequential selection procedures with control variates,.
IIE Transactions, 42:71–82, 2009.

H. Wang, R. Pasupathy, and B. W. Schmeiser.

Integer-ordered simulation optimization using r-spline: Retrospective search using piecewise-linear interpolation and neighborhood enumeration.
ACM TOMACS, 2012.
Under review.

X. Xiong, S. Juneja, and M. Fu.
Asymptotically optimal simulation allocation under dependent sampling.
Mathematics of Operations Research, 2012.
Under review.

J. Xu, L. J. Hong, and B. L. Nelson.
Industrial strength compass: A comprehensive algorithm and software for optimization via simulation.
ACM TOMACS, (20):1–29, 2010.

J. Xu, B. L. Nelson, and L. J. Hong.
An adaptive hyperbox algorithm for high-dimensional discrete optimization via simulation problems.

INFORMS Journal on Computing, 2011.
To appear.

- F. Yousefian, A. Nedić, and U.V. Shanbhag.
On stochastic gradient and subgradient methods with adaptive steplength sequences.
Automatica, 2011.