Scenario Generation and Sampling Methods

Güzin Bayraksan
Tito Homem-de-Mello

SVAN 2016 – IMPA
May 11th, 2016

Support material

Surveys in Operations Research and Management Science 19 (2014) 56–85

ELSEVIER

Contents lists available at ScienceDirect

Surveys in Operations Research and Management Science

journal homepage: www.elsevier.com/locate/sorms

Review

Monte Carlo sampling-based methods for stochastic optimization

CrossMark

Tito Homem-de-Mello [a,*], Güzin Bayraksan [b]

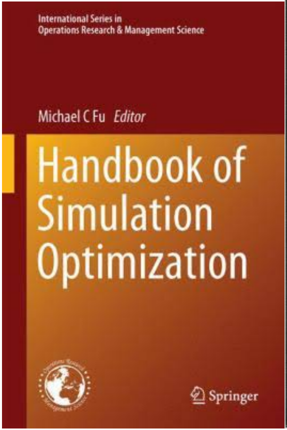
[a] *School of Business, Universidad Adolfo Ibañez, Santiago, Chile*

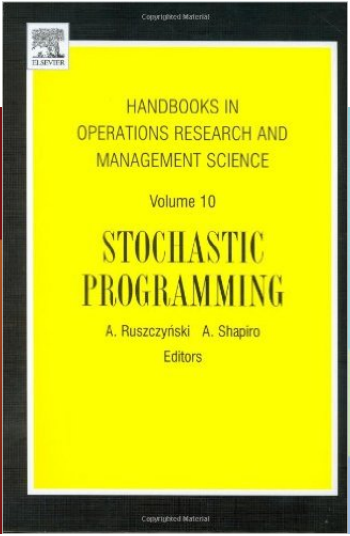
[b] *Integrated Systems Engineering, The Ohio State University, Columbus, OH, United States*

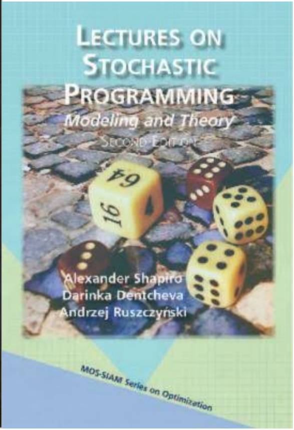
HIGHLIGHTS

- We survey the use of Monte Carlo sampling-based methods for stochastic optimization.
- We provide over 240 references from both the optimization and simulation literature.
- We discuss the convergence of optimal solutions/values for sampling approximations.
- Topics related to the implementation of sampling-based algorithms are discussed.
- An overview of alternative sampling techniques to reduce variance is presented.

Support material







The SAA approach

Recap: we were studying what happens when we approximate the problem

$$\min_{x \in X} \{g(x) := \mathbb{E}[G(x,\xi)]\} \tag{SP}$$

by

$$\min_{x \in X} \left\{ \hat{g}_N(x) \ := \ \frac{1}{N} \sum_{j=1}^{N} G(x,\xi^j) \right\}. \tag{SP$_N$}$$

This is called the Sample Average Approximation (SAA) approach.

Asymptotic properties of SAA

Let

$$
\begin{aligned}
\hat{x}_N &:= \text{an optimal solution of } (\mathrm{SP}_N)\\
S_N &:= \text{the set of optimal solutions of } (\mathrm{SP}_N)\\
\nu_N &:= \text{the optimal value of } (\mathrm{SP}_N)
\end{aligned}
$$

and

$$
\begin{aligned}
x^* &:= \text{an optimal solution of } (\mathrm{SP})\\
S^* &:= \text{the set of optimal solutions of } (\mathrm{SP})\\
\nu^* &:= \text{the optimal value of } (\mathrm{SP})
\end{aligned}
$$

As the sample size N goes to infinity, does

- $\hat{x}_N$ converge to some x^*?
- S_N converge to the set S^*?
- ν_N converge to ν^*?

Problems with stochastic constraints

So far our analysis has focused on the problem

$$\min_{x \in X} \{g(x) := \mathbb{E}[G(x,\xi)]\} \qquad \text{(SP)}$$

which has a *deterministic* feasibility set X, say, $X = \{x : h_i(x) \leq 0\}$, $i = 1, \dots, m$.

Issue: What if we have *stochastic* constraints? How to model the problem?

We will consider two classes of problems:

1. X has the form $\mathbb{E}[H_i(x,\xi)] \leq 0$, $i = 1, \dots, m$.
2. X has the form $P(H_i(x,\xi) \leq 0) \geq 1 - \alpha$, $i = 1, \dots, m$ (or, equivalently, $p(x) := P(H_i(x,\xi) > 0) \leq \alpha$).

Problems with expectation constraints

Let us consider the case where the stochastic constraint is $\mathbb{E}[H(x,\xi)] \leq 0$.

A natural approach is use SAA and replace the constraint with

$$\frac{1}{N}\sum_{j=1}^{N} H(x,\xi^j) \;\leq\; 0.$$

Does that work?

- Let us consider a simple example where the objective function is $g(x) = x$, and the constraint function is $H(x,\xi) = \xi - x$ where ξ has distribution $\text{Normal}(0,\sigma)$, i.e., the constraint is $x \geq 0 = \mathbb{E}[\xi]$.
- The SAA of this constraint is

$$x \;\geq\; \frac{1}{N}\sum_{j=1}^{N} \xi^j.$$

so the SAA solution is $\hat{x}_N = \frac{1}{N}\sum_{j=1}^{N} \xi^j$.

Problems with expectation constraints

Note that $\frac{1}{N}\sum_{j=1}^{N}\xi^j$ has distribution Normal$(0, \sigma/\sqrt{N})$.

So, there is a *50% chance* that the solution $\hat{x}_N$ will be infeasible for the original problem!

Idea: Perturb the feasibility set, writing it as
$U^\varepsilon := \{x \in X \,:\, \mathbb{E}[H(x,\xi)] \leq \varepsilon\}$.

- When $\varepsilon > 0$ we have a *relaxation* of the original problem.
- When $\varepsilon < 0$ we have a *tightening* of the original problem.

Problems with expectation constraints

Now let U_N^0 denote the feasibility region if the SAA problem, i.e.,

$$U_N^0 = \left\{ x \in X \;:\; \frac{1}{N} \sum_{j=1}^{N} H(x, \xi^j) \;\leq\; 0 \right\}.$$

Theorem

When X is compact, the function $H(\cdot, \xi)$ is Lipschitz and $H(x, \cdot)$ has finite moment generating function, there exist constants M and $\beta > 0$ such that

$$P\left(U^{-\varepsilon} \subseteq U_N^0 \subseteq U^{\varepsilon} \right) \;\geq\; 1 - Me^{-\beta\varepsilon^2 N}.$$

Application: Problems with CVaR constraints

Given a random variable Z, the conditional value-at-risk (CVaR) of Z is defined as

$$\mathrm{CVaR}_{1-\alpha}[Z] = \frac{1}{\alpha}\int_{1-\alpha}^{1} \mathrm{VaR}_{\gamma}[Z]\, d\gamma$$

where

$$\mathrm{VaR}_{\gamma}[Z] := \min\{t \mid P(Z \le t) \ge \gamma\}.$$

It is well known that the CVaR can be written as

$$\mathrm{CVaR}_{1-\alpha}[Z] = \min_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{\alpha}\mathbb{E}\left[(Z-\eta)_+\right] \right\},$$

where $(a)_+ := \max(a, 0)$.

Also, the optimal solution η^* of this problem is $\mathrm{VaR}_{1-\alpha}[Z]$!

Connection with CVaR constraints

Note also that, when Z has continuous distribution, we have

$$\begin{aligned} \mathbb{E}\left[Z \mid Z > \mathrm{VaR}_{1-\alpha}[Z]\right] &= \mathbb{E}\left[\mathrm{VaR}_{1-\alpha}[Z] + (Z - \mathrm{VaR}_{1-\alpha}[Z]) \mid Z > \mathrm{VaR}_{1-\alpha}[Z]\right] \\ &= \mathrm{VaR}_{1-\alpha}[Z] + \frac{\mathbb{E}\left[(Z - \mathrm{VaR}_{1-\alpha}[Z])_+\right]}{P(Z > \mathrm{VaR}_{1-\alpha}[Z])} \\ &= \eta^* + \frac{1}{\alpha}\mathbb{E}\left[(Z - \eta^*)_+\right] \\ &= \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha}\mathbb{E}\left[(Z - \eta)_+\right] \right\} \\ &= \mathrm{CVaR}_{1-\alpha}[Z]. \end{aligned}$$

In particular, this implies that $\mathrm{CVaR}_{1-\alpha}[Z] \geq \mathrm{VaR}_{1-\alpha}[Z]$.

Application: Problems with CVaR constraints

Consider now the problem

$$\min_{x \in X} \; g(x)$$
$$\text{s.t.} \quad \mathrm{CVaR}_{1-\alpha}[F(x,\xi)] \le a.$$

Then, by using the optimization representation of CVaR we can write the problem as

$$\min_{x \in X, \eta \in \mathbb{R}} \; g(x)$$
$$\text{s.t.} \quad \eta + \frac{1}{\alpha}\mathbb{E}\left[(F(x,\xi)-\eta)_+\right] \le a,$$

which falls into the standard formulation by defining
$H((x,\eta),\xi) := \eta + \frac{1}{\alpha}\left(F(x,\xi)-\eta\right)_+ - a.$

Chance-constrained problems

Chance constraints can be very helpful in modeling some situations.

This is true especially when what matters is *whether or not* a constraint was violated, not the *amount* of violation. For example,

- Reliability problems
- Problems with physical constraints

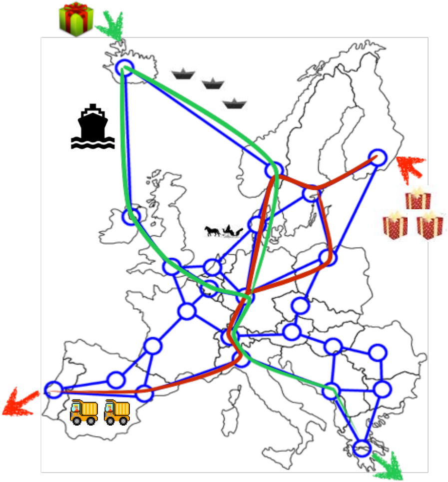
Also, it is often easier to choose the chance constraint level than to choose, say, penalties for violation.

Example: a telecommunication problem

Network: G=(V,A)

Commodities: $\mathcal{C}$, each one with possible demand d_c to be routed from s_c to t_c

Capacities: w_l for each link (need to be integral)

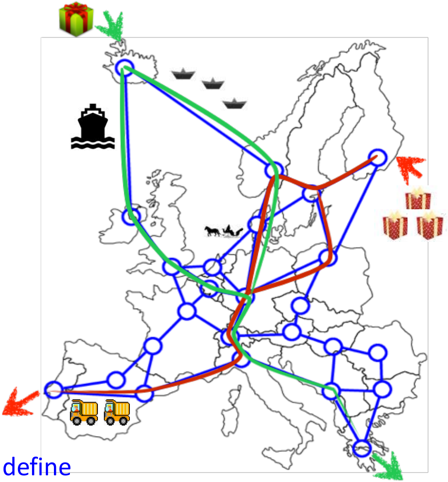


Example: a telecommunication problem

Network: G=(V,A)

Commodities: $\mathcal{C}$, each one with possible demand d_c to be routed from s_c to t_c

Capacities: w_l for each link (need to be integral)



Problem: To route each commodity and define capacities for each link that minimize the capacity installation cost, subject to a reliability constraint

Modeling the problem

- Each connection c communicates with probability ρ_c ($\xi_c \sim \text{Bernoulli}(\rho_c)$)
- We need to determine the minimum capacity w_ℓ for each link ℓ that will meet communication requirements with probability at least $1-\alpha_\ell$.
- The routing variables x_ℓ^c are equal to one if connection c uses link ℓ, zero otherwise

A chance-constrained formulation

$$\min_{x,w} \sum_{\ell \in L} w_\ell$$

$s.t.$

$$\begin{aligned}
\sum_{\ell \in \delta^+(s_c)} x_\ell^c - \sum_{\ell \in \delta^-(s_c)} x_\ell^c &= -1 \qquad \forall c \in \mathcal{C} \\
\sum_{\ell \in \delta^+(t_c)} x_\ell^c - \sum_{\ell \in \delta^-(t_c)} x_\ell^c &= 1 \qquad \forall c \in \mathcal{C} \\
\sum_{\ell \in \delta^+(n)} x_\ell^c - \sum_{\ell \in \delta^-(n)} x_\ell^c &= 0 \qquad \forall c \in \mathcal{C},\ \forall n \neq s_c, t_c \\
P\left(\sum_{c \in \mathcal{C}} \xi_c x_\ell^c \leq w_\ell \right) &\geq 1 - \alpha_\ell \qquad \forall \ell \in L.
\end{aligned}$$

Feasible regions: an example

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^2} & c_1x_1 + c_2x_2 \\ \text{s.t.} & P(\xi x_1 + x_2 \geq 7) \geq 1-\alpha \end{array}$$

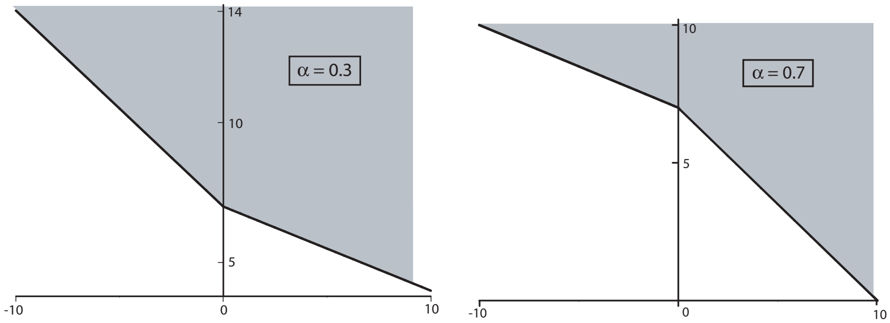
Assuming $\xi \sim U[0,1]$, draw the feasible region $C(\alpha)$ for $\alpha = 0.3$ and $\alpha = 0.7$.

Solution

If $\xi \sim U[0,1]$ then the feasible set is
$C(\alpha) = C_+(\alpha) \bigcup C_0(\alpha) \bigcup C_-(\alpha), \alpha \in (0,1)$, where

$$
\begin{aligned}
C_+(\alpha) &= \left\{x \in \mathbb{R}^2 \mid x_1 > 0, \alpha x_1 + x_2 \geq 7\right\}, \\
C_0(\alpha) &= \left\{(0, x_2) \in \mathbb{R}^2 \mid x_2 \geq 7\right\}, \\
C_-(\alpha) &= \left\{x \in \mathbb{R}^2 \mid x_1 < 0, (1-\alpha) x_1 + x_2 \geq 7\right\}
\end{aligned}
$$

Feasible regions



Connection with CVaR constraints

Consider the chance constraint

$$P(F(x,\xi) \le 0) \;\ge\; 1-\alpha.$$

Note that this is equivalent to

$$\mathrm{VaR}_{1-\alpha}[F(x,\xi)] \;\le\; 0.$$

Recall that we saw earlier that $\mathrm{CVaR}_{1-\alpha}[Z] \ge \mathrm{VaR}_{1-\alpha}[Z]$.

- Therefore, if we replace the chance constraint $P(F(x,\xi) \le 0) \ge 1-\alpha$ with $\mathrm{CVaR}_{1-\alpha}[F(x,\xi)] \le 0$, we have a *conservative approximation*.
- The advantage of such an approximation is that the feasibility set is <u>convex</u> if $F(\cdot,\xi)$ is convex.

Sampling approaches

- Non-convexity of chance-constraints does not occur when the distribution of ξ belongs to a certain class (called log-concave distributions).

- But what to do if the random parameters do not follow a tractable distribution?

- One alternative is to apply the SAA approach, which replaces the chance constraint by its sample average.

- The resulting problem is easier to solve, and provides useful information to the true problem.

SAA

- Let $\xi^1, \ldots, \xi^N$ be a random sample from ξ.
- Using that $P(\xi \in A) = \mathbb{E}[\mathbb{I}_A(\xi)]$, the SAA of a chance constrained problem is

$$\min_{x \in X} \; g(x)$$
$$\text{s.t.} \quad p_N(x) := \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}_{(0,\infty)}\left(H(x, \xi^j)\right) \leq \gamma$$

(Compare with the original problem:)

$$\min_{x \in X} \; g(x)$$
$$\text{s.t.} \quad p(x) := P(H(x, \xi) > 0) \; \leq \; \alpha$$

The scenario approach

Note that if we take $\gamma = 0$ in the above formulation we obtain

$$\begin{aligned} \min_{x \in X} \quad & g(x) \\ \text{s.t.} \quad & H(x, \xi^j) \leq 0, \quad j = 1, \ldots, N. \end{aligned}$$

If each function $H(\cdot, \xi)$ is *convex* and g is *convex*, then the above problem is convex.

- This is called the scenario approach.
- What is the relationship to the original problem?

The scenario approach

Theorem

Select a confidence parameter $\beta \in (0,1)$, and let d_x denote the dimension of x. Suppose that $H(\cdot,\xi)$ is convex. If

$$N \geq \frac{2}{\alpha}\left(\ln\frac{1}{\beta} + d_x\right),$$

then, with probability at least $1-\beta$ we have that $\hat{x}_N$ satisfy all constraints in the original problem but at most a fraction α, that is,

$$P\left(H(\hat{x}_N,\xi) > 0\right) \;\leq\; \alpha,$$

regardless of the distribution of ξ.

An equivalent IP formulation

Consider now the case of $\gamma > 0$.

- Given a sample of size N, we can write the problem as

$$\begin{aligned} \min_{x \in X} \quad & g(x) \\ \text{s.t.} \quad H(x, \hat{\xi}^i) - \mathcal{M} z_i & \le 0 \quad i = 1, \ldots, N, \\ \frac{1}{N} \sum_{i=1}^{N} z_i & \le \gamma, \\ z_i & \in \{0,1\}^N. \end{aligned} \tag{P}$$

- That is, we obtain an IP formulation, which is particularly helpful when H is linear in x.

Feasibility results

Similar results to the scenario approach theorem (i.e., feasibility of $\hat{x}_N$ guaranteed up to a confidence $1-\beta$) can be obtained, under various different settings:

- When X is finite;
- When $H(x,\xi)$ is of the form $H(x,\xi) = \xi - h(x)$;
- When $H(\cdot,\xi)$ is a Lipschitz function, with Lipschitz constant independent of ξ.

Asymptotic results

Condition (A): There is an optimal solution $\bar{x}$ of the true problem such that for any $\epsilon > 0$ there is $x \in X$ with $\|x - \bar{x}\| \leq \epsilon$ and $p(x) < \alpha$.

Consistency of SAA

Suppose that

(i) the significance levels of the true and SAA problems are the same, i.e., $\gamma = \alpha$,

(ii) the set X is compact,

(iii) the function $g(x)$ is continuous,

(iv) $H(x, \xi)$ is a Carathéodory function,

(v) condition (A) holds.

Then, $\nu_N \to \nu^*$ and $\mathrm{dist}(\hat{S}_N, S^*) \to 0$ w.p.1 as $N \to \infty$.

Dealing with small probabilities

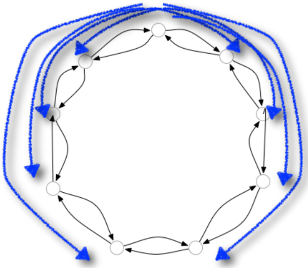
Let us consider again the reliability problem seen earlier, and suppose the reliability factor is very small, say, 10^{-6}.

What happens to the SAA approximation?

- As we saw earlier, the sample size estimates to achieve some desirable confidence are proportional to $1/\alpha$.
    - This is not surprising: the probability that the first violation occurs in the kth sample is $(1-\alpha)^{k-1}\alpha$.
    - Therefore, on average we need $(1-\alpha)/\alpha$ samples just to obtain one case for which violation occurs!
- So, we need a lot of samples.
- But each sample corresponds to a variable in the IP formulation!

Why not just use $\alpha = 0$?

α =0 : Shortest Path Solution (optimal)



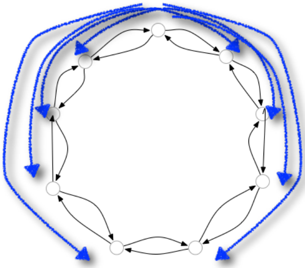
ρ=0.1

No. of connections routed on each link = 10
Capacity w_l on each link =10

Total cost: 18 × 10 =180

Why not just use $\alpha = 0$?

α =0 : Shortest Path Solution (optimal)

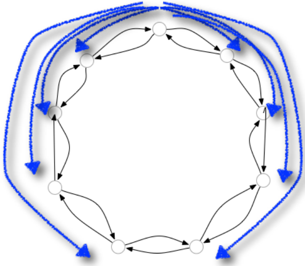


ρ=0.1

No. of connections routed on each link = 10
Capacity w_l on each link = 10

Total cost: 18 × 10 = 180

α = 10^{-6} : Shortest Path Solution



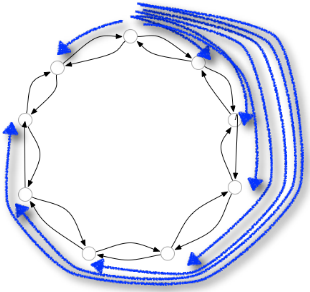
ρ=0.1

No. of connections routed on each link = 10
Capacity w_l on each link = 7

Total cost: 18 × 7 = 126

Why not just use $\alpha = 0$?

$\alpha = 10^{-6}$:Optimal Solution



ρ=0.1

No. of connections routed on each <u>clockwise</u> link = 28
No. of connections routed on each <u>counterclockwise</u> link = 1

Capacity w_l on each clockwise link = 12
Capacity w_l on each c/clockwise link = 1

Total cost: 9 × 12 + 9 × 1 = 117

Lessons from this example

- We cannot pretend that a very small α is equivalent to zero...
- On the other hand, when α is very small SAA will require a lot of samples!
- We need to do some "smarter sampling"
- One such strategy is importance sampling — see Guzin's talk!