Scenario Generation and Sampling Methods

Güzin Bayraksan
Tito Homem-de-Mello

SVAN 2016 – IMPA
May 11th, 2016
Monte Carlo sampling-based methods for stochastic optimization

Tito Homem-de-Mello\textsuperscript{a,*}, Güzin Bayraksan\textsuperscript{b}

\textsuperscript{a} School of Business, Universidad Adolfo Ibañez, Santiago, Chile
\textsuperscript{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.
Recap: we were studying what happens when we approximate the problem

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

by

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

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

Let

\[ \hat{x}_N := \text{an optimal solution of } (SP_N) \]
\[ S_N := \text{the set of optimal solutions of } (SP_N) \]
\[ \nu_N := \text{the optimal value of } (SP_N) \]

and

\[ x^* := \text{an optimal solution of } (SP) \]
\[ S^* := \text{the set of optimal solutions of } (SP) \]
\[ \nu^* := \text{the optimal value of } (SP) \]

As the sample size \( N \) goes to infinity, does

- \( \hat{x}_N \) converge to some \( x^* \)?
- \( S_N \) converge to the set \( S^* \)?
- \( \nu_N \) converge to \( \nu^* \)?
So far our analysis has focused on the problem
\[
\min_{x \in X} \{ g(x) := \mathbb{E}[G(x, \xi)] \} \quad \text{(SP)}
\]
which has a \textit{deterministic} feasibility set \( X \), say, \( X = \{ x : h_i(x) \leq 0 \} \), \( i = 1, \ldots, m \).

**Issue:** What if we have \textit{stochastic} constraints? How to model the problem?

We will consider two classes of problems:

1. \( X \) has the form \( \mathbb{E}[H_i(x, \xi)] \leq 0, \ i = 1, \ldots, m \).

2. \( X \) has the form \( P(H_i(x, \xi) \leq 0) \geq 1 - \alpha, \ i = 1, \ldots, m \) (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 $\xi$ has distribution 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.
Now let $U_0^N$ denote the feasibility region if the SAA problem, i.e.,

$$U_0^N = \left\{ x \in X : \frac{1}{N} \sum_{j=1}^{N} H(x, \xi^j) \leq 0 \right\}.$$

**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*

$$P \left( U^{-\varepsilon} \subseteq U_0^N \subseteq U^{\varepsilon} \right) \geq 1 - Me^{-\beta \varepsilon^2 N}.$$
Application: Problems with CVaR constraints

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

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

where

$$\text{VaR}_\gamma[Z] := \min \{ t \mid P(Z \leq t) \geq \gamma \}.$$

It is well known that the CVaR can be written as

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

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

Also, the optimal solution $\eta^*$ of this problem is $\text{VaR}_{1-\alpha}[Z]$!
Note also that, when $Z$ has continuous distribution, we have

$$
\mathbb{E}[Z \mid Z > \text{VaR}_{1-\alpha}[Z]] = \mathbb{E}[\text{VaR}_{1-\alpha}[Z] + (Z - \text{VaR}_{1-\alpha}[Z]) \mid Z > \text{VaR}_{1-\alpha}[Z]]
$$

$$
= \text{VaR}_{1-\alpha}[Z] + \frac{\mathbb{E}[(Z - \text{VaR}_{1-\alpha}[Z])_+]}{P(Z > \text{VaR}_{1-\alpha}[Z])}
$$

$$
= \eta^* + \frac{1}{\alpha} \mathbb{E}[(Z - \eta^*)_+]
$$

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

$$
= \text{CVaR}_{1-\alpha}[Z].
$$

In particular, this implies that $\text{CVaR}_{1-\alpha}[Z] \geq \text{VaR}_{1-\alpha}[Z]$. 
Consider now the problem

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

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

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

which falls into the standard formulation by defining

\[H((x, \eta), \xi) := \eta + \frac{1}{\alpha} (F(x, \xi) - \eta)_+ - a.\]
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: $w_i$ 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
Each connection $c$ communicates with probability $\rho_c$ 
($\xi_c \sim \text{Bernoulli}(\rho_c)$)

We need to determine the minimum capacity $w_\ell$ for each link $\ell$ that will meet communication requirements with probability at least $1 - \alpha_\ell$.

The routing variables $x_\ell^c$ are equal to one if connection $c$ uses link $\ell$, zero otherwise.
A chance-constrained formulation

\[
\begin{align*}
\min_{x, w} & \quad \sum_{\ell \in L} w_{\ell} \\
\text{s.t.} & \quad \sum_{\ell \in \delta^+(s_c)} x_{\ell}^c - \sum_{\ell \in \delta^-(s_c)} x_{\ell}^c = -1 \quad \forall c \in C \\
& \quad \sum_{\ell \in \delta^+(t_c)} x_{\ell}^c - \sum_{\ell \in \delta^-(t_c)} x_{\ell}^c = 1 \quad \forall c \in C \\
& \quad \sum_{\ell \in \delta^+(n)} x_{\ell}^c - \sum_{\ell \in \delta^-(n)} x_{\ell}^c = 0 \quad \forall c \in C, \forall n \neq s_c, t_c \\
& \quad \mathbb{P} \left( \sum_{c \in C} \xi_c x_{\ell}^c \leq w_{\ell} \right) \geq 1 - \alpha_{\ell} \quad \forall \ell \in L.
\end{align*}
\]
Feasible regions: an example

\[
\begin{align*}
\min_{x \in \mathbb{R}^2} & \quad c_1 x_1 + c_2 x_2 \\
\text{s.t.} & \quad P(\xi x_1 + x_2 \geq 7) \geq 1 - \alpha
\end{align*}
\]

Assuming \( \xi \sim U[0, 1] \), draw the feasible region \( C(\alpha) \) for \( \alpha = 0.3 \) and \( \alpha = 0.7 \).
If \( \xi \sim U[0, 1] \) then the feasible set is
\[
C(\alpha) = C_+(\alpha) \cup C_0(\alpha) \cup C_-(\alpha), \alpha \in (0, 1),
\]
where
\[
C_+(\alpha) = \left\{ x \in \mathbb{R}^2 \mid x_1 > 0, \alpha x_1 + x_2 \geq 7 \right\},
\]
\[
C_0(\alpha) = \left\{ (0, x_2) \in \mathbb{R}^2 \mid x_2 \geq 7 \right\},
\]
\[
C_-(\alpha) = \left\{ x \in \mathbb{R}^2 \mid x_1 < 0, (1 - \alpha)x_1 + x_2 \geq 7 \right\}.
\]
Feasible regions

\[ \alpha = 0.3 \]

\[ \alpha = 0.7 \]
Connection with CVaR constraints

Consider the chance constraint

$$P(F(x, \xi) \leq 0) \geq 1 - \alpha.$$ 

Note that this is equivalent to

$$\text{VaR}_{1-\alpha}[F(x, \xi)] \leq 0.$$ 

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

- Therefore, if we replace the chance constraint $P(F(x, \xi) \leq 0) \geq 1 - \alpha$ with $\text{CVaR}_{1-\alpha}[F(x, \xi)] \leq 0$, we have a conservative approximation.

- The advantage of such an approximation is that the feasibility set is convex if $F(\cdot, \xi)$ is convex.
Sample approaches

- Non-convexity of chance-constraints does not occur when the distribution of $\xi$ 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.
Let $\xi^1, \ldots, \xi^N$ be a random sample from $\xi$.

Using that $P(\xi \in A) = \mathbb{E}[\mathbb{I}_A(\xi)]$, the SAA of a chance constrained problem is

$$\min_{x \in X} g(x)$$

$$\text{s.t. } p_N(x) := \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}_{(0,\infty)}(H(x, \xi^j)) \leq \gamma$$

(Compare with the original problem:)

$$\min_{x \in X} g(x)$$

$$\text{s.t. } p(x) := P(H(x, \xi) > 0) \leq \alpha$$
The scenario approach

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

$$\min_{x \in X} \ g(x)$$

s.t. $H(x, \xi^j) \leq 0, \ j = 1, \ldots, N$.

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 $\alpha$, that is,

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

regardless of the distribution of $\xi$. 
Consider now the case of $\gamma > 0$.

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

$$
\min_{x \in X} g(x)
$$

s.t. $H(x, \hat{\xi}^i) - Mz_i \leq 0 \quad i = 1, \ldots, N,$

$$
\frac{1}{N} \sum_{i=1}^{N} z_i \leq \gamma,
$$

$z_i \in \{0, 1\}^N$.

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 $\xi$. 

Bayraksan (OSU) & Homem-de-Mello (UAI)
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, $\nu_N \rightarrow \nu^*$ and $\text{dist}(\hat{S}_N, S^*) \rightarrow 0$ w.p.1 as $N \rightarrow \infty$. 
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 $k$th 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 $\alpha = 0$?

$\alpha = 0$ : Shortest Path Solution (optimal)

$\rho = 0.1$

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

Total cost: $18 \times 10 = 180$
Why not just use $\alpha = 0$?

$\alpha = 0$ : Shortest Path Solution (optimal)

$\rho = 0.1$

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

Total cost: $18 \times 10 = 180$

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

$\rho = 0.1$

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

Total cost: $18 \times 7 = 126$
Problems with stochastic constraints

Sampling approaches for chance constraints

Why not just use $\alpha = 0$?

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

$\rho = 0.1$

No. of connections routed on each **clockwise** link = 28
No. of connections routed on each **counterclockwise** link = 1

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

Total cost: $9 \times 12 + 9 \times 1 = 117$
Lessons from this example

- We cannot pretend that a very small $\alpha$ is equivalent to zero...
- On the other hand, when $\alpha$ 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!