Monte Carlo Methods in Stochastic Programming

David Morton

Graduate Program in Operations Research

The University of Texas at Austin

August 2001

Monte Carlo Methods in Stochastic Programming
Overview

- Introduction
  - Problem statement
  - Some things we won't cover
  - Why Monte Carlo?

- The External Sampling Method
  - What properties should we want/expect?
  - Bias
  - Consistency
  - Rates of convergence and limiting distributions

- Internal Sampling Methods
  - Cutting-plane methods: two-stage
  - Cutting-plane methods: multi-stage
  - Stochastic quasi-gradient methods

- Testing Solution Quality

- An Example of Sequential Issues that Arise

Stochastic Programming Models

$$z^* = \min_{x \in X} E\, f(x, \tilde{\xi})$$

Such problems arise in statistics, simulation, and mathematical programming.

Our focus: mathematical programming with X deterministic.

We'll assume $X \neq \emptyset$ and closed, $z^* > -\infty$ and $E\, f^2(x, \tilde{\xi}) < \infty\ \forall x \in X$. Call these assumptions $(A1)$-$(A3)$.

$\tilde{\xi}$ is a random vector on $(\Xi, \mathcal{B}, P)$.

Choice of f determines problem class.

1. Simple recourse model (i.e., one-stage or static model)

   Examples:

   - $f(x, \tilde{\xi}) = \rho_1 \left(\sum_{i=1}^{K} \tilde{\xi}_i x_i - T \right)^+ + \rho_2 \left(\sum_{i=1}^{K} \tilde{\xi}_i x_i - T \right)^-$
   - $f(x, \tilde{\xi}) = I \left(\sum_{i=1}^{K} \tilde{\xi}_i x_i \leq T \right)$

2. Two-stage stochastic linear program with recourse:

$$
\begin{aligned}
f(x,\tilde{\xi}) \;=\; cx + \;\min_{y} \quad & \tilde{g}y \\
\text{s.t.} \quad & \tilde{D}y = \tilde{B}x + \tilde{d} \\
& y \geq 0
\end{aligned}
$$

where

$$\tilde{\xi} = \mathrm{vec}(\tilde{g}, \tilde{D}, \tilde{B}, \tilde{d})$$

and $X = \{x \,:\, Ax = b, x \geq 0\}$.

Classic example of a two-stage problem:

x decision variables determine system design

a realization of $\tilde{\xi}$ is then observed

y decision variables, adapted to ξ, are used to operate system

3. A T-stage stochastic program with recourse

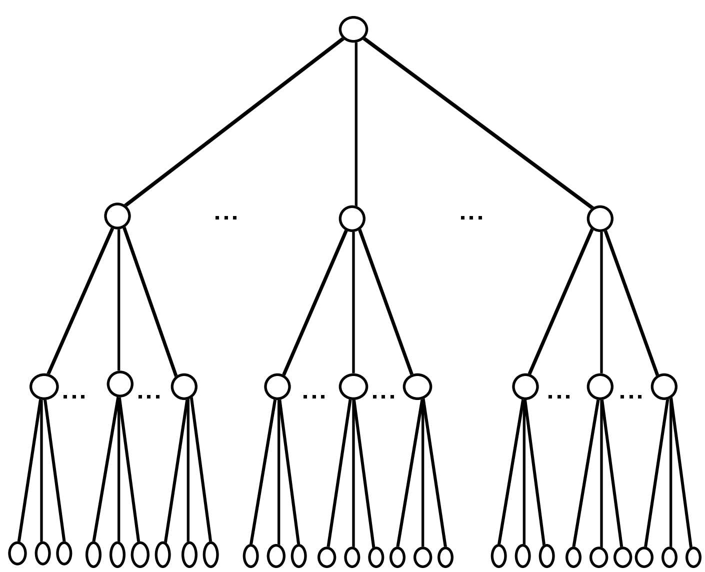
$$
\begin{aligned}
z^* \;=\; & \min_{x_1} \; Ef_1(x^0, x_1, \tilde{\xi}_2 | \tilde{\xi}^1) \\
& \text{s.t. } \; x_1 \in X_1(x^0, \tilde{\xi}^1)
\end{aligned}
$$

where for $t = 1, \ldots, T-2$,

$$
\begin{aligned}
f_t(x^t, \tilde{\xi}_{t+1} | \tilde{\xi}^t) \;=\; & \min_{x_{t+1}} \; Ef_{t+1}(x^t, x_{t+1}, \tilde{\xi}_{t+2} | \tilde{\xi}^{t+1}) \\
& \text{s.t. } \; x_{t+1} \in X_{t+1}(x^t, \tilde{\xi}^{t+1})
\end{aligned}
$$

and

$$
\begin{aligned}
f_{T-1}(x^{T-1}, \tilde{\xi}_T | \tilde{\xi}^{T-1}) \;=\; & \min_{x_T} \; f_T(x^{T-1}, x_T, \tilde{\xi}^T) \\
& \text{s.t. } \; x_T \in X_T(x^{T-1}, \tilde{\xi}^T).
\end{aligned}
$$



Some Things We Won't Talk About

- Models with Decision-Dependent Probability Distributions
- Optimizing (Discrete-Event) Simulation Models

  Gradient estimation: IPA, Likelihood-ratio method

  Grid search

  Meta-models
- Methods for Discrete X

  Comparing alternative system designs

  Bandit problems

  (actually much of what we will discuss holds for discrete X)
- Stochastic Dynamic Programming/Markov Decision Processes

Why Use Monte Carlo Methods?

For fixed x, $Ef(x,\tilde{\xi}) = \int_\Xi f(x,\xi)dP$

So, put optimization w.r.t. x aside for the moment

Let's consider the simpler problem of numerically computing the multi-dimensional integral

$$I = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} g(u_1,\ldots,u_d)du_d \ldots du_1$$

What methods are available?

Broadly speaking, the methods can be classified as:

1. Analytical Integration

   Quality: exact

   Viability: often impossible or impractical to perform

2. Numerical Quadrature (e.g., Simpson's rule in $\Re^d$)

   Quality: deterministic error bounds

   Viability: limited by dimension d, say, $d \leq 10\text{-}12$

3. Monte Carlo Integration

   Quality: statistical error bounds

   Viability: converges slowly ($\propto 1/\sqrt{n}$) but rate independent of d (although standard deviation tends to grow with d)

High-dimensional integral of a "difficult" function g

   Monte Carlo integration is the "method of choice"

Solving Stochastic Programs

- Exact Solution Procedures
  - Simplex method, interior point algorithms, decomposition methods

- Bounding and Approximation Schemes
  - Lower bounds: Jensen's inequality
  - Upper bounds: Edmundson-Madansky inequality
  - Apply bounds conditionally to partitions of Ξ

- Monte Carlo Sampling-Based Methods
  - External sampling
  - Internal sampling

Monte Carlo Integration

Suppressing the (fixed) decision x

Let $z = Ef(\tilde{\xi})$ and $\sigma^2 = \text{var} f(\tilde{\xi}) < \infty$

Let $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be i.i.d. as $\tilde{\xi}$

Let $z_n = \dfrac{1}{n} \sum_{i=1}^{n} f(\tilde{\xi}^i)$

z_n is the *sample mean* estimator of the *population mean* z

FACT 1. $Ez_n = z$

z_n is an *unbiased estimator* of z

FACT 2. $z_n \to z,$ wp1 (strong LLN)

z_n is a *strongly consistent* estimator of z

FACT 3. $\sqrt{n}(z_n - z) \Rightarrow N(0, \sigma^2)$ (CLT)

Rate of convergence is $1/\sqrt{n}$ and

scaled difference is normally distributed

FACTS 4,5,... law of iterated logarithm, large deviations results,...

Monte Carlo Methods in Stochastic Programming
External Sampling

- True/population problem:

$$(P) \qquad z^* = \min_{x \in X} E\, f(x, \tilde{\xi})$$

  Denote optimal solution x^*

- Solve sample problem:

$$(P_n) \qquad z_n^* = \min_{x \in X} \left[\frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) \right]$$

  Here, $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ i.i.d. as $\tilde{\xi}$

  Denote optimal solution x_n^*

- View z_n^* as an estimator of z^* and x_n^* as an estimator of x^*

- What can we say about z_n^* and x_n^* as $n \to \infty$?

- What *should* we want to say about z_n^* and x_n^* as $n \to \infty$?

Note: What I'm calling the *external sampling method* is also called *sample-path optimization*, *sample average approximation*, *stochastic counterpart*, and *non-recursive method*.

Monte Carlo Methods: External Sampling
Possible Goals

1. $x_n^* \to x^*$, wp1 and $\sqrt{n}(x_n^* - x^*) \Rightarrow N(0, \Sigma)$
2. $z_n^* \to z^*$, wp1 and $\sqrt{n}(z_n^* - z^*) \Rightarrow N(0, \sigma^2)$
3. $Ef(x_n^*, \tilde{\xi}) \to z^*, \text{wp1}$
4. $\lim_{n \to \infty} P\left\{Ef(x_n^*, \tilde{\xi}) \leq z^* + \tilde{\epsilon}_n\right\} = 1 - \alpha$ where $\tilde{\epsilon}_n \to 0$

Modeling Issues:

• If (P_n) is for maximum-likelihood estimation then goal 1. could be appropriate.

• If (P) is to price a financial option then goal 2. could be appropriate.

• When (P) is a decision-making model, 1. may be more then we need and 2. is of secondary interest. Goals 3. and 4. are arguably adequate.

Mathematical/Technical Issues:

• In general, we shouldn't expect $\{x_n^*\}_{n=1}^\infty$ to converge when (P) has multiple optimal solutions. In this case, we want: "limit points of $\{x_n^*\}_{n=1}^\infty$ solve (P)."

• If we achieve this "limit points" result, X is compact and $Ef(x, \tilde{\xi})$ is continuous then we obtain goal 3.

• The limiting distributions may not be normal.

Monte Carlo Methods: External Sampling
Example

$$(P) \qquad z^* = \min_{-1 \le x \le 1} \left[E\, f(x, \tilde{\xi}) = E\tilde{\xi} x \right]$$

where $\tilde{\xi} \sim N(0,1)$

$z^* = 0$ and every feasible solution is optimal

$$(P_n) \qquad z_n^* = \min_{-1 \le x \le 1} \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{\xi}^i \right) x$$

$$x_n^* = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} \tilde{\xi}^i \le 0 \\ -1 & \text{if } \sum_{i=1}^{n} \tilde{\xi}^i > 0 \end{cases} \quad \text{and } z_n^* = -\left| N(0, 1/n) \right|$$

Observations

1. $E z_n^* < z^* \, \forall n$ (Negative bias)
2. $z_n^* \to z^*$, wp1 (Strongly consistent)
3. $\sqrt{n}(z_n^* - z^*) = -\left| N(0,1) \right|$ (Non-normal error)

Another example:

Change the feasible region in (P) to $X = \Re$

Then, as before $z^* = 0$ but $z_n^* = -\infty,$ wp1 $\forall n$

So, *optimization* changes the nature of sample-mean estimators

Monte Carlo Methods: External Sampling
Bias

$$E\left[\frac{1}{n}\sum_{i=1}^{n} f(x,\tilde{\xi}^i)\right] = E\, f(x,\tilde{\xi}) \;\; \forall x \in X$$

Thus,

$$\min_{x\in X} E\left[\frac{1}{n}\sum_{i=1}^{n} f(x,\tilde{\xi}^i)\right] = \min_{x\in X} E\, f(x,\tilde{\xi}) = z^*$$

and so we obtain

$$Ez_n^* = E\left[\min_{x\in X}\frac{1}{n}\sum_{i=1}^{n} f(x,\tilde{\xi}^i)\right] \le \min_{x\in X} E\, f(x,\tilde{\xi}) = z^*$$

So, z_n^* has a negative bias, i.e., $Ez_n^* \le z^*$

Simple example when $n = 1$

$$E\min_{x\in X} f(x,\tilde{\xi}) \le \min_{x\in X} Ef(x,\tilde{\xi})$$

Interpretation: We'll do better if we get to "wait and see" $\tilde{\xi}$ before choosing x.

Also, can show bias decreases monotonically

$$Ez_n^* \le Ez_{n+1}^* \le z^*$$

Intuition?

Monte Carlo Methods: External Sampling
Consistency of z_n^*

As we've seen, the strong LLN

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) - Ef(x, \tilde{\xi}) \right| = 0, \text{ wp1}, \ \forall x \in X$$

is insufficient to ensure $z_n^* \to z^*$, wp1.

However, a *uniform* strong LLN hypothesis is sufficient

$$\lim_{n\to\infty} \sup_{x \in X} \left| \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) - Ef(x, \tilde{\xi}) \right| = 0, \text{ wp1}.$$

We have

$$|z_n^* - z^*| = \left| \frac{1}{n} \sum_{i=1}^{n} f(x_n^*, \tilde{\xi}^i) - Ef(x^*, \tilde{\xi}) \right| \leq \sup_{x \in X} \left| \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) - Ef(x, \tilde{\xi}) \right|$$

Why?

The right-hand side converges to 0 by the USLLN hypothesis so $z_n^* \to z^*$, wp1, i.e., z_n^* is a strongly consistent estimator of z^*.

Monte Carlo Methods: External Sampling
Consistency of x_n^*

Under USLLN and the additional hypothesis that $Ef(x, \tilde{\xi})$ is continuous we obtain: Every limit point of $\{x_n^*\}_{n=1}^{\infty}$ solves (P).

Idea of proof:

Assume $\hat{x}$ is a limit point of $\{x_n^*\}_{n=1}^{\infty}$ with $Ef(\hat{x}, \tilde{\xi}) > z^*$ and the subsequence is indexed by N. We already know

$$\lim_{\substack{n \to \infty \\ n \in N}} \left[\frac{1}{n} \sum_{i=1}^{n} f(x_n^*, \tilde{\xi}^i) \equiv z_n^* \right] = z^*, \text{ wp1.}$$

But,

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(x_n^*, \tilde{\xi}^i) - Ef(\hat{x}, \tilde{\xi}) \right|$$

$$\leq \left| \frac{1}{n} \sum_{i=1}^{n} f(x_n^*, \tilde{\xi}^i) - Ef(x_n^*, \tilde{\xi}) \right| + \left| Ef(x_n^*, \tilde{\xi}^i) - Ef(\hat{x}, \tilde{\xi}) \right|$$

First term $\to 0$ by USLLN and second by continuity of $Ef(x, \tilde{\xi})$.

So,

$$\lim_{\substack{n \to \infty \\ n \in N}} \frac{1}{n} \sum_{i=1}^{n} f(x_n^*, \tilde{\xi}^i) = Ef(\hat{x}, \tilde{\xi}) > z^* \qquad \square$$

- $\{x_n^*\}_{n=1}^{\infty}$ can have: no limit point, multiple limit points or a unique limit point
- To ensure existence of a limit point it's sufficient to assume X is compact
- If X is compact and (P) has a unique solution x^* then $\{x_n^*\}_{n=1}^{\infty}$ converges to x^*.

Monte Carlo Methods: External Sampling
Consistency of x_n^* and z_n^*
Epi-Convergence

Could now turn to studying conditions under which the USLLN holds...

Above discussion is simple enough to squeeze onto two slides but doesn't give a representative view of consistency analyses.

Epi-convergence is a more general (and hence powerful) tool for establishing consistency results.

DEFN: A sequence of functions $\{\phi_n\}_{n=1}^{\infty}$ is said to epi-converge to ϕ (written $\phi_n \overset{\text{epi}}{\to} \phi$) if the epi-graphs of ϕ_n, $\{(x, \beta) : \beta \geq \phi_n(x)\}$ converge to that of ϕ.

FACT: Suppose $\phi_n \overset{\text{epi}}{\to} \phi$. Then,
(i) $\limsup_{n\to\infty} \inf_x \phi_n(x) \leq \inf_x \phi(x)$
(ii) If $\hat{x}$ is an limit point of $\{x_n^*\}_{n=1}^{\infty}$ where $x_n^* \in \text{argmin}\, \phi_n(x)$, then $\hat{x} \in \text{argmin}\, \phi(x)$.

Epi-convergence is employed in a variety of settings when a true model (P) is approximated by a sequence of models (P_n).

The approximation might be deterministic, stochastic, arise algorithmically, etc.

Monte Carlo Methods: External Sampling
Rates of Convergence and Limiting Distributions: z_n^*

By the standard CLT for i.i.d.r.v.'s we have

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} f(x,\tilde{\xi}^i) - Ef(x,\tilde{\xi})\right] \Rightarrow N(0,\sigma^2(x)) \ \ \forall x \in X$$

where $\sigma^2(x) = \mathrm{var} f(x,\tilde{\xi})$.

Now, we've already seen that when we optimize the sample and population means with respect to x the limiting distribution need not be normal.

However, we can obtain the following. Assume $(A1)$-$(A3)$ and

• X is compact

• f satisfies the following Lipschitz condition

$$\left|f(x_1,\tilde{\xi}) - f(x_2,\tilde{\xi})\right| \le a(\tilde{\xi})\,\|x_1 - x_2\| \ \ \forall x_1, x_2 \in X,$$

where $Ea^2(\tilde{\xi}) < \infty$ for some real-valued function a

• (P) has a unique optimal solution x^*

Then,

$$\sqrt{n}\,(z_n^* - z^*) \Rightarrow N(0,\sigma^2(x^*)).$$

• The asymptotic analysis is done on a functional space.

• If solution to (P) isn't unique then the limiting distribution is non-normal and involves optimization of random variables, parameterized by x, over X^*, the set of optimal solutions.

Monte Carlo Methods: External Sampling
Rates of Convergence and Limiting Distributions: x_n^*

Assume $(A1)$-$(A3)$,

- $f(\cdot, \tilde{\xi})$ is convex and twice continuously differentiable
- $X = \{x : Ax \leq b\}$
- (P) has a unique optimal solution x^*
- $(x_1 - x_2)^T H^* (x_1 - x_2) > 0, \ \ \forall x_1, x_2 \in X, x_1 \neq x_2$
  where $H^* = E\nabla_x^2 f(x^*, \tilde{\xi})$.
- Assume $\nabla_x f(x, \tilde{\xi})$:
  — is measurable for all $x \in X$
  — satisfies the following Lipschitz condition

$$\left\|\nabla_x f(x_1, \tilde{\xi}) - \nabla_x f(x_2, \tilde{\xi})\right\| \leq a(\tilde{\xi}) \left\|x_1 - x_2\right\| \ \ \forall x_1, x_2 \in X,$$

where $Ea^2(\tilde{\xi}) < \infty$ for some real-valued function a

Then, $\sqrt{n}(x_n^* - x^*) \Rightarrow u$ where u solves the random QP:

$$\begin{array}{lll} \min\limits_u & \frac{1}{2}u^T H^* u + c^T u & \\ \text{s.t.} & A_{i\cdot} u \leq 0 & : i \in \{i : A_i x^* = b_i\} \\ & u^T E\nabla_x f(x^*, \tilde{\xi}) = 0 & \end{array}$$

and c is multivariate normal with mean 0 and covariance matrix Σ^* where $\Sigma^*_{ij} = \text{cov}\left(\frac{\partial f(x^*, \tilde{\xi})}{\partial x_i}, \frac{\partial f(x^*, \tilde{\xi})}{\partial x_j}\right)$.

Monte Carlo Methods for Stochastic Programming
External Sampling: Summary

- True/population problem:

$$(P) \qquad z^* = \min_{x \in X} E\, f(x, \tilde{\xi})$$

  Denote optimal solution x^*

- Solve sample problem:

$$(P_n) \qquad z_n^* = \min_{x \in X} \left[\frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) \right]$$

  Here, $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ i.i.d. as $\tilde{\xi}$

  Denote optimal solution x_n^*

Bias: z_n^* has an optimistic bias that decreases as n grows.

Consistency: Uniform SLLN yields consistency of z_n^*. USLLN plus continuity of $Ef(x, \tilde{\xi})$ ensures "limit pts solve (P)" notion of consistency for x_n^*. Epi-convergence is powerful analysis tool.

Rates of convergence and limiting distributions: Additional assumptions needed to obtain results. Convergence rate is $1/\sqrt{n}$. but in general, limiting distribution is non-normal.

Now, what if $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ are not i.i.d. but are drawn in another manner (e.g., in an attempt to reduce variance)?

Monte Carlo Methods for Stochastic Programming
Internal Sampling: Basic Idea

Consider the deterministic convex optimization model:

$$(P) \qquad \min_{x \in X} F(x)$$

where X is a convex set and F a convex function.

Algorithms to solve (P) (e.g., steepest-descent, cutting-plane) require evaluations of F and/or ∇F at iterates x_k.

Our stochastic program is (P) with $F(x) = Ef(x, \tilde{\xi})$. So, if we can evaluate $Ef(x, \tilde{\xi})$ and $\nabla Ef(x, \tilde{\xi})$ at iterates x_k we can employ the deterministic algorithm.

We consider Monte Carlo methods when exact evaluations of $Ef(x, \tilde{\xi})$ and $\nabla Ef(x, \tilde{\xi})$ are impossible.

Developing an "internal sampling" method consists of adapting a deterministic algorithm by replacing $Ef(x, \tilde{\xi})$ and/or $\nabla Ef(x, \tilde{\xi})$ with estimates obtained by Monte Carlo sampling.

Of course, we want the new algorithm to generate iterates $\{x_k\}$ with desirable properties.

New and typically independent observations of $\tilde{\xi}$ are generated on an as-needed basis at each iteration of the algorithm.

Caveat: $Ef(x, \tilde{\xi})$ is often nondifferentiable (so regard $\nabla Ef(x, \tilde{\xi})$ as a subgradient)

Monte Carlo Methods: Internal Sampling
A Cutting-Plane Algorithm

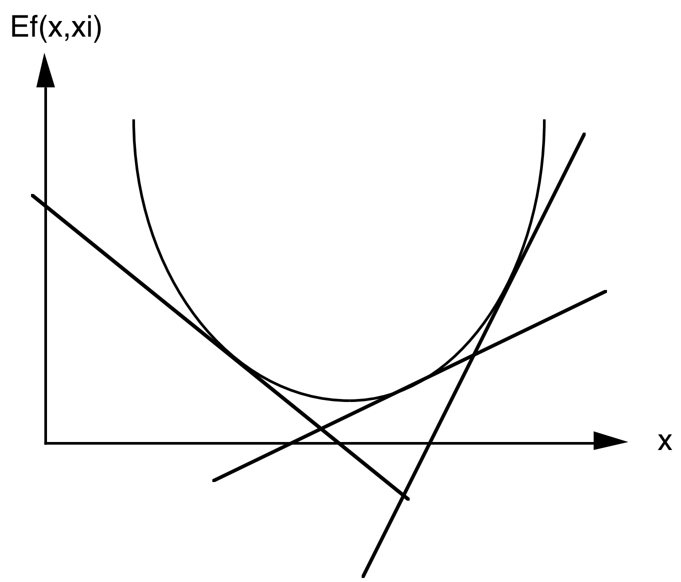
Solve a sequence of master programs of the form:

$$
\begin{aligned}
z_K = \ & \min_{x,\theta} \ \theta \\
& \text{s.t.} \quad \theta \geq Ef(x_k,\tilde{\xi}) + \nabla Ef(x_k,\tilde{\xi})(x - x_k), \ \ k \in \mathcal{K} \\
& \qquad \ \ x \in X
\end{aligned}
$$

where $\mathcal{K} = \{1, \ldots, K\}$.

Properties:

- $z_K \leq z^* \leq \min_{k \in \mathcal{K}} Ef(x_k,\tilde{\xi})$
- X compact, $\nabla Ef(x,\tilde{\xi})$ bounded yields optimal limit points
- Can't compute $Ef(x_k,\tilde{\xi})$ and $\nabla Ef(x_k,\tilde{\xi})$ (what to do?)



Monte Carlo Methods: Internal Sampling
A Cutting-Plane Algorithm

Suppose we generate $n = 500$ i.i.d. observations of $\tilde{\xi}$ and at each iteration use these $n = 500$ observations to replace

$$Ef(x_k, \tilde{\xi}) \text{ and } \nabla Ef(x_k, \tilde{\xi})$$

with

$$\frac{1}{n}\sum_{i=1}^{n} f(x_k, \tilde{\xi}^i) \text{ and } \frac{1}{n}\sum_{i=1}^{n} \nabla f(x_k, \tilde{\xi}^i)$$

This might be reasonable but it's not an internal sampling method. It's an external sampling method: We're simply solving an instance of $(P_{n=500})$ with a cutting-plane algorithm.

Now, if at each iteration we generate $n = 500$ *new* observations of $\tilde{\xi}$ that are independent of observations in previous iterations then we'd have an internal sampling algorithm.

Dantzig, Glynn and Infanger do this with a sampling-based version of the L-shaped method (called DECIS) for two-stage stochastic LPs. They also employ an importance sampling scheme to reduce variance.

Monte Carlo Methods: Internal Sampling
A Cutting-Plane Algorithm: Stochastic Decomposition

$$(P) \qquad \min_{x \in X} Ef(x, \tilde{\xi})$$

Assume $(A1)$-$(A3)$, $f(x, \tilde{\xi}) \geq 0$, wp1, X is a polytope, Ξ is compact,

$$\begin{aligned} f(x, \tilde{\xi}) \;=\; cx + \; & \min_{y \geq 0} \quad qy \\ & \text{s.t.} \quad Wy \;= \tilde{T}x + \tilde{h} \end{aligned}$$

and $\{\pi : \pi W \leq q\}$ is bounded.

Stochastic Decomposition Algorithm (Higle and Sen)

1. Find $x_1 \in X$ and let $K = 0$ and $\mathcal{K} = \{0\}$.
2. Let $K = K + 1$, $\mathcal{K} = \mathcal{K} \cup \{K\}$, and generate $\tilde{\xi}^K$ from distribution of $\tilde{\xi}$ and independent of other observations. Compute
$$g_K = \frac{1}{K} \sum_{k=1}^{K} f(x_K, \tilde{\xi}^k) - \frac{1}{K} \sum_{k=1}^{K} \nabla f(x_K, \tilde{\xi}^k) x_K$$
and $G_K = \dfrac{1}{K} \sum_{k=1}^{K} \nabla f(x_K, \tilde{\xi}^k)$
3. Let $g_k = \frac{K-1}{K} g_k$ and $G_k = \frac{K-1}{K} G_k$ for $k \in \mathcal{K} \setminus \{K\}$.
4. Let
$$\begin{aligned} x_{K+1} \in \; \text{argmin}_{x,\theta} \;\; & \theta \\ \text{s.t.} \quad & \theta \geq G_k x + g_k \;\; k \in \mathcal{K} \\ & x \in X \end{aligned}$$

   Go to step 2.

Monte Carlo Methods: Internal Sampling
A Cutting-Plane Algorithm: Stochastic Decomposition (cont.)

The entire history of observations $\tilde{\xi}^1, \ldots, \tilde{\xi}^K$ is used in computing the cut at iteration K.

"Old cuts never change" in deterministic cutting-plane methods. The danger with this in a sampling-based cutting-plane method is that a cut that's too high can "cut off" the optimal solution.

Step 3 is a "cure" to this danger because it (slowly) washes out old cuts.

The SD algorithm actually computes cuts only over the set of dual extreme points observed so far. Such cuts are cheaper to compute but are weaker than what we show.

Convergence result: Under hypotheses on previous slide, there is a subsequence of $\{x_k\}_{k=1}^{\infty}$ with limit points that solve (P), wp1.

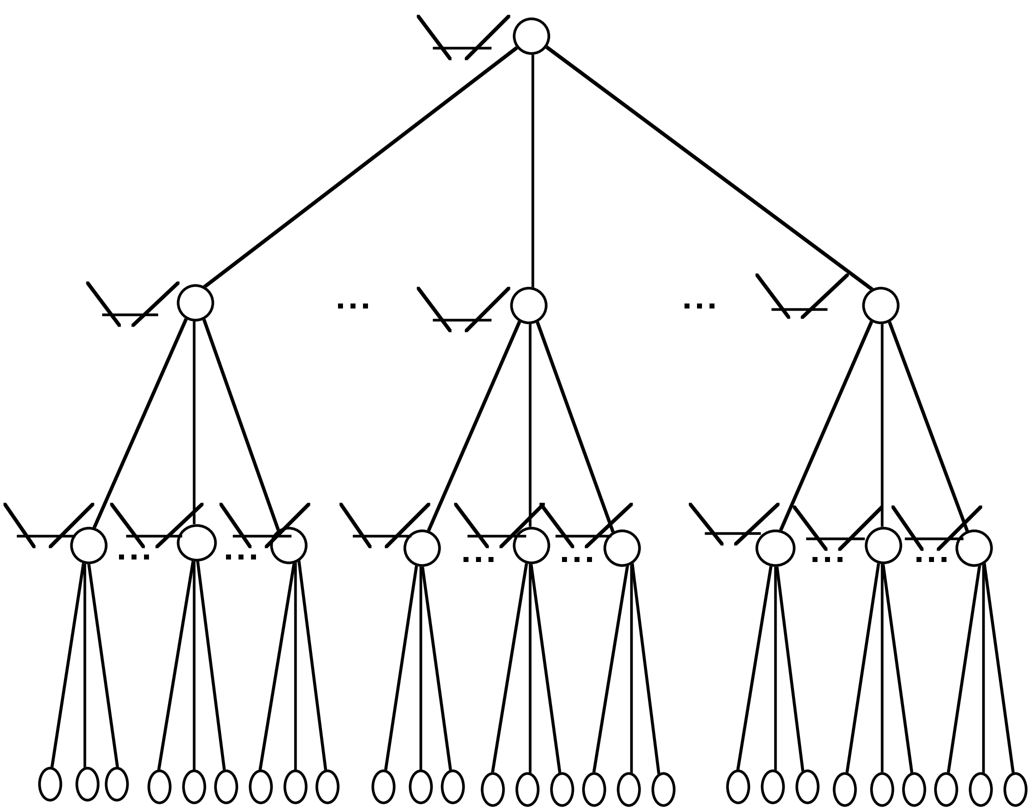
And, using a sequence of incumbent solutions with improving objective estimates it is possible to construct such a subsequence.

This establishes a consistency result for SD.

Monte Carlo Methods: Internal Sampling
A Multi-Stage Cutting-Plane Algorithm

Consider a multi-stage stochastic linear program.

A "general" implementation of a nested decomposition algorithm stores cuts separately at every node of the scenario tree.



What if the stochastic parameters are serially independent?

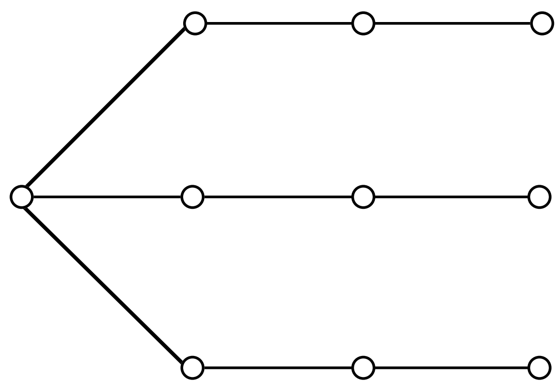
Then, the nodes on each stage t can "share" cuts, i.e., we only have to store one group of cuts per stage.

Monte Carlo Methods: Internal Sampling
A Multi-Stage Cutting-Plane Algorithm (cont.)

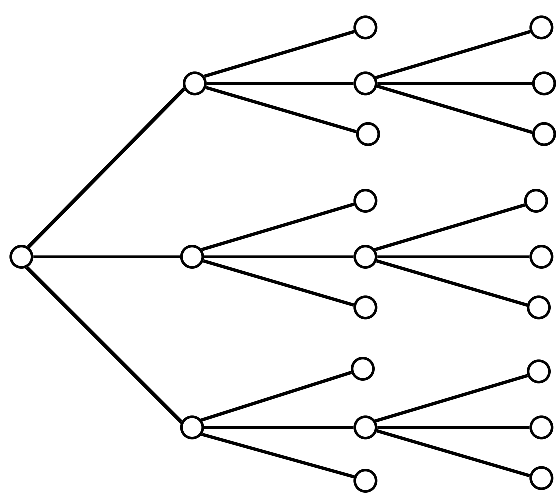
Assume inter-stage independence and assume the number of direct descendants of any one node in the tree is "manageable" (e.g., 50). For a T-stage problem this yields 50^{T-1} leaf nodes.

Algorithm (Pereira & Pinto and Birge & Donohue)

Upper Bound is a Sample Mean:



Lower Bound is Deterministic:



Can establish consistency.

Monte Carlo Methods: Internal Sampling
Variants of Steepest-Descent Methods

Steepest-Descent Method

$$\min_x f(x) \qquad\qquad x_{n+1} = x_n - t_n \nabla_x f(x_n)$$

$$\min_x Ef(x, \tilde{\xi}) \qquad\qquad x_{n+1} = x_n - t_n \nabla_x Ef(x_n, \tilde{\xi})$$

Stochastic variant of steepest descent replaces $\nabla_x Ef(x_n, \tilde{\xi})$ with an estimator.

- Robbins-Monro (1951): Gradient estimator is unbiased
- Kiefer-Wolfowitz (1952): Finite-difference estimator
- Stochastic Quasi-Gradient Methods (Ermoliev 1983): Allows non-differentiable $Ef(x, \tilde{\xi})$, constrained optimization (with a projector operator), biased gradient estimators (with bias shrinking to zero).

A focus in simulation: Constructing gradient estimators.

Finite-difference approximations lead to convergence rates slower than $n^{-1/2}$. For example, with forward differences can obtain $n^{-1/4}$, with central differences $n^{-1/3}$, with central differences and common random numbers $n^{-2/5}$.

Monte Carlo Methods
Testing Solution Quality

$$(P) \qquad\qquad z^* = \min_{x \in X} E\, f(x, \tilde{\xi})$$

Our goal: Given $\hat{x} \in X$ and α we want to find a (random) confidence interval width $\tilde{\epsilon}$ with:

$$P(E\, f(\hat{x}, \tilde{\xi}) \le z^* + \tilde{\epsilon}) \approx 1 - \alpha$$

Recall bias result $E z_n^* \le z^*$ where

$$(P_n) \qquad\qquad z_n^* = \min_{x \in X} \left[\frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) \right]$$

So,

$$E \underbrace{\left[\frac{1}{n} \sum_{i=1}^{n} f(\hat{x}, \tilde{\xi}^i) - \min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^i) \right]}_{G_n} \ge E f(\hat{x}, \tilde{\xi}) - z^*.$$

Remarks

- Anticipate $\operatorname{var} G_n \le \operatorname{var} U_n + \operatorname{var} L_n$
- $G_n \ge 0$
- G_n is not asymptotically normal (what to do?)

Monte Carlo Methods
Testing Solution Quality (cont.)

1. Find $\hat{x} \in X$, candidate solution
2. Generate n_g batches of size n:

$$\tilde{\xi}^{11}, \tilde{\xi}^{12}, \ldots, \tilde{\xi}^{1n}$$

$$\tilde{\xi}^{21}, \tilde{\xi}^{22}, \ldots, \tilde{\xi}^{2n}$$

$$\vdots$$

$$\tilde{\xi}^{n_g 1}, \tilde{\xi}^{n_g 2}, \ldots, \tilde{\xi}^{n_g n}$$

3. Calculate point estimate:

$$\bar{G}(n_g) = \frac{1}{n_g} \sum_{i=1}^{n_g} G_n^i = \frac{1}{n_g} \sum_{i=1}^{n_g} \left[\frac{1}{n} \sum_{j=1}^{n} f(\hat{x}, \tilde{\xi}^{ij}) - \min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} f(x, \tilde{\xi}^{ij}) \right]$$

Based on the central limit theorem for i.i.d.r.v.'s:

$$\sqrt{n_g} \left[\bar{G}(n_g) - EG_n \right] \Rightarrow N(0, \sigma_g^2) \text{ as } n_g \to \infty$$

Construct $(1-\alpha)$-level confidence interval for optimality gap:

$$\left[0, \bar{G}(n_g) \ + \ \frac{t_{n_g-1,\alpha} s_g(n_g)}{\sqrt{n_g}} \right]$$

Optimality gap $= Ef(\hat{x}, \tilde{\xi}) - z^*$

Monte Carlo Methods
Testing Solution Quality (cont.)
Methods for Finding $\hat{x}$

Option 1.

Using the external sampling method and solve

$$\hat{x} \in \underset{x \in X}{\operatorname{argmin}} \frac{1}{n'} \sum_{i=1}^{n'} f(x, \tilde{\xi}^i)$$

Option 2.

Use an internal sampling method (SD, SQG, . . .)

Option 3.

Employ a heuristic. . .

Monte Carlo Methods

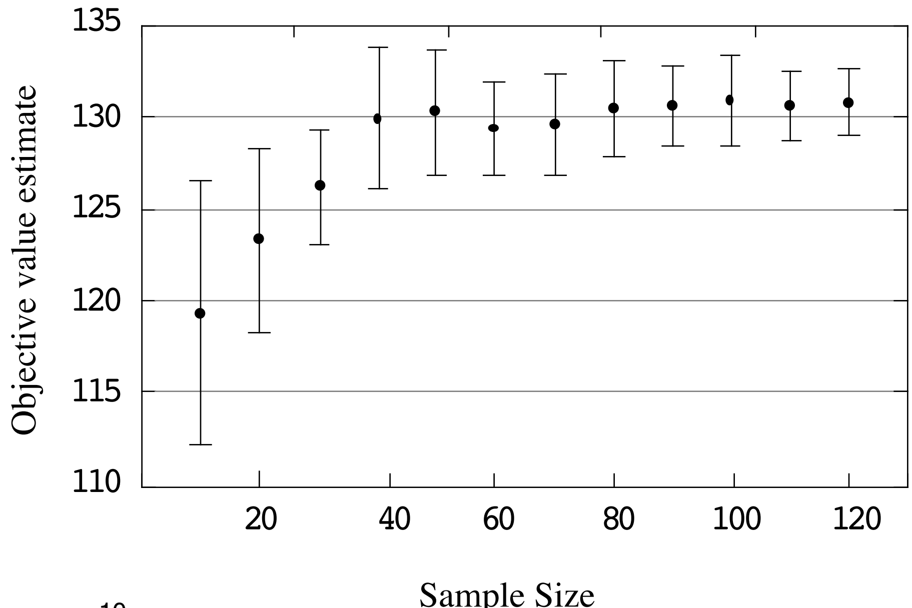
Testing Solution Quality (cont.)

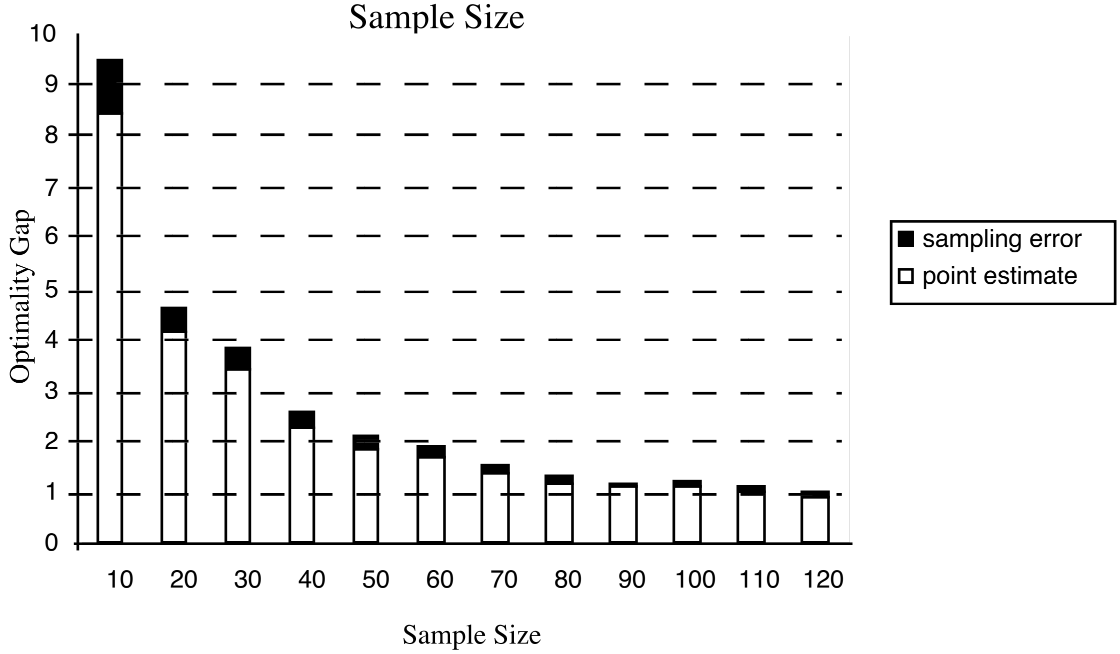
Computational Results: Two-Stage SLPs

Problem	"Core" Size	Dimension of ξ	Total scenarios
DB	$73 \times 107\,(357)$	46	4.5×10^{25}
WRPM	$345 \times 364\,(1226)$	11	5.6×10^{6}
20TERM	$128 \times 827\,(5336)$	40	1.1×10^{12}
SSN	$177 \times 795\,(2548)$	86	1.0×10^{70}

Problem	DB	WRPM	20TERM	SSN
n' in $SP_{n'}$ for $\hat{x}$	50	50	50	2000
		Optimality Gap		
n	25	25	25	1000
n_g	30	30	30	30
$\bar{G}(n_g)$	23	197	141	0.69
95% CI Width	0.2%	0.08%	0.5%	8%
$\times$ Var. Red.	4300	480	1300	17
		Upper Bound		
$\bar{U}(n_u)$	-17567	289186	254394	10.06

Monte Carlo Methods: Bias and Sampling Error
Capacity Expansion in a Manufacturing System





Value of Stochastic Solution $\approx 13\%$

$Ef(\hat{x}, \tilde{\xi}) \approx 131.7 \pm 2.3 \;\; (95\%)$ with 3000 samples

Monte Carlo Methods
An Example of Sequential Issues that Arise
Stopping Rules for Sampling-Based Algorithms

$$(P) \qquad z^* = \min_{x \in X} E\, f(x, \tilde{\xi})$$

Situation:

• Assume we have at hand an algorithm that at each iteration k requires us to select a sample size n_k

• Then, we observe $G_k(n_k)$ satisfying

$$\sqrt{n_k}(G_k(n_k) - \mu_k) \Rightarrow N(0, \sigma_k^2) \text{ as } n_k \to \infty$$

where $G_k = U_k - L_k$ with $EL_k \le z^*$ and $EU_k = E\, f(x_k, \tilde{\xi})$ for $x_k \in X$.

• We will stop the algorithm on iteration $T = \inf\{k : G_k(n_k) \le 0\}$

• Our goal: select sample sizes $n_k = n_k(\epsilon)$ to ensure

$$\lim_{\epsilon \downarrow 0} P\left\{\mu_{T(\epsilon)} \le \epsilon\right\} \ge 1 - \alpha$$

Note this result implies

$$\lim_{\epsilon \downarrow 0} P\left\{E\, f(x_{T(\epsilon)}, \tilde{\xi}) \le z^* + \epsilon\right\} \ge 1 - \alpha$$

Monte Carlo Methods
Stopping Rules (cont.)

Basic Idea: Increase the sample size as the algorithm proceeds

We want

$$
\begin{aligned}
P\{\mu_T > \epsilon\} &\leq P\{\mu_T > G_T + \epsilon\} \\
&= \sum_{k=1}^{\infty} P\{\text{Stop at } k, G_k(n_k) \leq \mu_k - \epsilon\} \\
&\leq \sum_{k=1}^{\infty} P\{G_k(n_k) \leq \mu_k - \epsilon\} \\
&\leq \alpha
\end{aligned}
$$

We'll bound this infinite series, term-by-term, by a series that converges to α.

Monte Carlo Methods
Stopping Rules (cont.)

Assume $\sup\limits_{k \geq 1} E \exp\left[\gamma \left(\frac{D_k - \mu_k}{\sigma_k / \sqrt{n_k}}\right)\right]$ is bounded for $|\gamma| < \gamma_o$.

If

$$n_k \geq \left(\frac{\sigma_k}{\epsilon}\right)^2 \left\{\beta' + 2p \ \ln^2 k\right\}$$

then

$$\lim_{\epsilon \downarrow 0} P\left\{\mu_{T(\epsilon)} \leq \epsilon\right\} \geq 1 - \alpha$$

where $0 < \alpha < 1$, $p > (2\gamma_o^2)^{-1}$, $\epsilon > 0$, $\beta' = \max\left\{2 \ln\left(\frac{\phi(p)}{\sqrt{2\pi}\ \alpha}\right),\ 1\right\}$

and $\phi(p) = \sum\limits_{k=1}^{\infty} k^{-p \ln k}$

Again, this yields

$$\lim_{\epsilon \downarrow 0} P\left\{E\, f(x_{T(\epsilon)}, \tilde{\xi}) \leq z^* + \epsilon\right\} \geq 1 - \alpha$$

Monte Carlo Methods in Stochastic Programming
Summary

- The External Sampling Method
  - What properties should we want/expect?
  - Bias
  - Consistency
  - Rates of convergence and limiting distributions
- Internal Sampling Methods
  - Cutting-plane methods: two-stage
  - Cutting-plane methods: multi-stage
  - Stochastic quasi-gradient methods
- Testing Solution Quality
- An Example of Sequential Issues that Arise
- With more time…Variance Reduction Techniques