Monte Carlo Sampling Methods in Stochastic Programming

David Love

Graduate Interdisciplinary Program in Applied Mathematics

Oral Comprehensive Exam
May 13, 2010

1. Introduction to Stochastic Programming
    - Notation and Terminology
    - Convenient Special Cases
2. Uncertain Convex Programs
3. Overlapping Batches
4. Conclusions

Introduction to Math Programming

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$
$$\text{subject to } \mathbf{x} \in \mathcal{X}$$

$\mathbf{x}$	Vector of decision variables.
f	Objective function.
$\mathcal{X}$	Feasible region.

- e.g., $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \ldots, m, \mathbf{x} \geq 0\}$.

Some Terminology

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \qquad \text{(GMP)}$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}$$

- $\hat{\mathbf{x}}$ is feasible if $\hat{\mathbf{x}}\in\mathcal{X}$.
- $\hat{\mathbf{x}}$ has objective function value or cost $f(\hat{\mathbf{x}})$.
- (GMP) is infeasible if $\mathcal{X}=\emptyset$.
- $\mathbf{x}^*$ is an optimal solution if $\mathbf{x}^*\in\operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$.

Theoretically Nice Assumptions

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) \qquad \text{(GMP)}$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}$$

Under some conditions

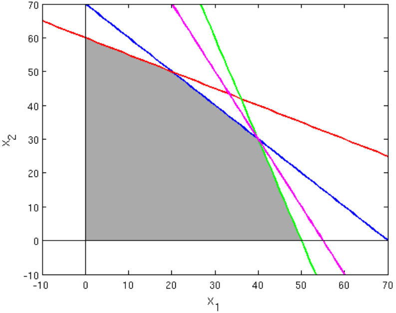
- $f(\mathbf{x})$ is convex in $\mathbf{x}$.
- $\mathcal{X}$ is a convex set.

Then locally optimal solutions are globally optimal!

- Then (GMP) is called a convex program.

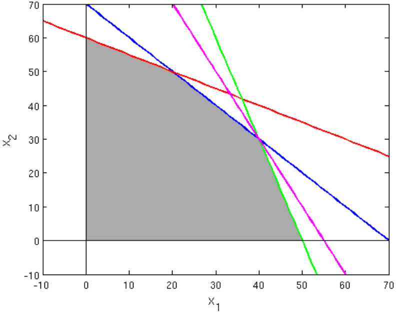
Example: Linear Programming

$$
\begin{array}{rrrl}
\min & -10x_1 & -8x_2 & \\
\text{subject to} & x_1 & +x_2 & \leq 70 \\
 & x_1 & +2x_2 & \leq 120 \\
 & 3x_1 & +x_2 & \leq 150 \\
 & 2x_1 & +x_2 & \leq 110 \\
 & (x_1, & x_2) & \in \mathbb{R}^2_+
\end{array}
$$



Example: Linear Programming

$$\begin{array}{lrrl} \min & -10x_1 & -8x_2 & \\ \text{subject to} & x_1 & +x_2 & \leq 70 \\ & x_1 & +2x_2 & \leq 120 \\ & 3x_1 & +x_2 & \leq 150 \\ & 2x_1 & +x_2 & \leq 110 \\ & (x_1, & x_2) & \in \mathbb{R}^2_+ \end{array}$$



What if the parameters aren't known?

Introduction to Stochastic Programming

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x},\xi)$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}(\xi)$$

$\mathbf{x}$	Vector of decision variables.
ξ	Vector of problem parameters.
f	Objective function, may depend on ξ.
$\mathcal{X}$	Feasible region, may depend on ξ.

- e.g., $\mathcal{X}(\xi)=\{\mathbf{x}\in\mathbb{R}^n : g_i(x,\xi)\le 0, i=1,\dots,m,\mathbf{x}\ge 0\}$.

Introduction to Stochastic Programming

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x},\xi)$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}(\xi)$$

$\mathbf{x}$ Vector of decision variables.

ξ Vector of problem parameters.

f Objective function, may depend on ξ.

$\mathcal{X}$ Feasible region, may depend on ξ.

- e.g.,
  $\mathcal{X}(\xi)=\{\mathbf{x}\in\mathbb{R}^n : g_i(x,\xi)\leq 0, i=1,\ldots,m,\mathbf{x}\geq 0\}$.

Uncertain Convex Program

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^n} \quad & c^T\mathbf{x} && \text{(UCP)}\\
\text{subject to} \quad & g(\mathbf{x},\xi) \le 0, && \xi \in \Xi\\
& \mathbf{x} \in \mathcal{X}.
\end{aligned}
$$

$c^T\mathbf{x}$	Linear objective function, independent of ξ.
$g(\mathbf{x},\xi)$	Convex scalar constraint, contains all ξ dependence.
$\mathcal{X}$	Feasible region, independent of ξ.
Ξ	Support of parameter vector ξ.

Robust Convex Program

- First idea: Find a solution that is feasible for every value of $\xi \in \Xi$.

$$\min_{\mathbf{x}\in\mathbb{R}^n} c^T\mathbf{x} \tag{RCP}$$
$$\text{subject to } \mathbf{x} \in \mathcal{X} \cap K.$$

$K \equiv \bigcap_{\xi\in\Xi}\{\mathbf{x} : g(\mathbf{x},\xi) \leq 0\}.$

Analysis of RCP

- One problem: $K = \bigcap_{\xi \in \Xi} \{\mathbf{x} : g(\mathbf{x}, \xi) \leq 0\}$ may be empty.
  - e.g., Ξ is unbounded.
  - Calafiori and Campi assume $K \neq \emptyset$.
- Another problem: (RCP) may be a semi-infinite program.
  - Finitely many variables, infinitely many constraints.
- (RCP) is still a convex program.
  - i.e., $\mathcal{X} \cap K$ is convex.

Chance Constrained Program

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \; & c^T\mathbf{x} \\ \text{subject to } & \mathbb{P}\left[g(\mathbf{x},\tilde{\xi}) > 0\right] \le \epsilon \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{CCP}$$

$\mathbb{P}$	Probability measure (distribution of $\tilde{\xi}$).
$\tilde{\xi}$	Parameters now take the form of a random variable.
ϵ	"Acceptable risk" of infeasible solution.

Analysis of CCP

- Does not suffer from infeasibility
    - i.e., there is (usually) a solution for sufficiently large tolerated risk.
- Feasible region is difficult to calculate.
    - Typically involves multi-dimension integral.
- Feasible region is, in general, **not** convex!

Example of Non-Convex Feasible Region

An example from Sen & Higle:

$$\min_{\mathbf{x}\in\mathbb{R}^2} x_1 + x_2$$
$$\text{subject to } \mathbb{P}\left[\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix}\right] \geq 0.5.$$

Random Variables $\tilde{\xi}_1, \tilde{\xi}_2$ standard Bernoulli distribution.

Dependence $\tilde{\xi}_1 = 1 - \tilde{\xi}_2$.

Feasible Region of Chance Constrained Program

$$\min_{\mathbf{x}\in\mathbb{R}^2} x_1 + x_2$$

$$\text{subject to } \mathbb{P}\left[2x_1 + x_2 \geq \tilde{\xi}_1, x_1 + 2x_2 \geq \tilde{\xi}_2\right] \geq 0.5.$$

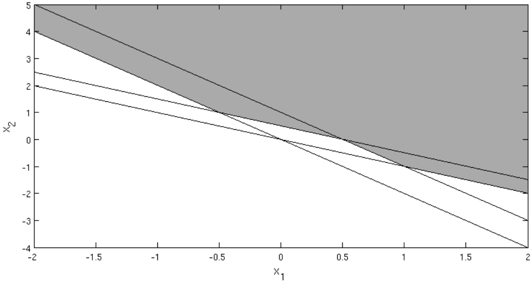


Figure: A graphical representation of feasible region. Feasible region is **not** convex.

Sampled Convex Program

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & c^T\mathbf{x} && && (\mathrm{SCP}_N) \\ \text{subject to } \quad & g(\mathbf{x},\tilde{\xi}^{(i)}) \leq 0, && i = 1,\ldots,N \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

$\tilde{\xi}^{(i)}$	Independent and identically distributed (iid) samples of $\tilde{\xi}$.
N	Number of samples taken.

Analysis of SCP_N

- Has some danger of infeasibility.
    - Is always feasible if (RCP) is feasible.
- Unlike (RCP), is a finite program!
- Is always a convex program.
- Feasibility of a point is easy to check.
- Solutions of (CCP) and (SCP_N) will outperform those of (RCP).

Benefits of Chance Constrained Problem

$$\min_{x \in \mathbb{R}} x$$

$$\text{subject to } g(x,\tilde{\xi}) \leq 0,$$

$$g(x,\tilde{\xi}) = \begin{cases} \frac{1}{\alpha} - x, & \tilde{\xi} \in [0,\alpha] \\ -x, & \tilde{\xi} \in (\alpha, 1] \end{cases}$$

$\tilde{\xi}$ $\sim U(0,1)$.

α A “bad” outcome with small probability.

(RCP) Feasible solutions are $x \geq \frac{1}{\alpha}$.

(CCP) If $\epsilon > \alpha$, feasible solutions are $x \geq 0$.

Optimal Cost $\frac{1}{\alpha}$ or 0.

Relating the Problems

We want to examine solutions to our sampling problem.

- How does the sampled problem compare to the chance constrained problem?
- How does the sampled problem compare to the robust problem?
- How many samples should we take?

Compared to CCP

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & c^T\mathbf{x} && && (\mathrm{SCP}_N) \\ \text{subject to} \quad & g(\mathbf{x},\tilde{\xi}^{(i)}) \le 0, && i = 1,\ldots,N \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

- Solve (SCP_N), get $\mathbf{x}_N^*$.
- What is the probability that $\mathbf{x}_N^*$ violates the constraint?

Principal Result

Theorem (Calafiori & Campi)

Let $\mathbf{x}_N^*$ be the optimal solution to (SCP_N). Then

$$\mathbb{E}_{\mathbb{P}^N}\left[\mathbb{P}_{\tilde{\xi}}\left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0\right]\right] \leq \frac{n}{N+1}.$$

n Dimension of problem ($\mathbf{x} \in \mathbb{R}^n$).

$\mathbb{P}^N$ $(= \mathbb{P} \times \cdots \times \mathbb{P}, N$ times)

- Measure on iid samples $\tilde{\xi}^{(1)}, \ldots, \tilde{\xi}^{(N)}$, i.e., on Ξ^N.

$\mathbb{P}_{\tilde{\xi}}$ $= \mathbb{P}$. Independent of $\mathbf{x}_N^*$.

Discussion of Theorem

$$\mathbb{E}_{\mathbb{P}^N}\left[\mathbb{P}_{\tilde{\xi}}\left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0\right]\right] \leq \frac{n}{N+1}.$$

- Linear increase with problem size (n) is good.
- Decreases with $1/N$, also good.
- Surprisingly, bound is independent of probability measure!

Moving Beyond Average Behavior

We want to know more than average behavior.

- Use sampling to find $\mathbf{x}_N^*$.
- Want an estimate on probability of constraint violation.
- What is the probability that our estimate is wrong?

ϵ Level Parameter. Probability of constraint violation is at most ϵ.

β Confidence Parameter. Probability that $\mathbf{x}_N^*$ fails to meet desired level parameter.

Principal Result

Corollary (Calafiori & Campi)

Fix two real numbers $\epsilon \in [0,1]$ *(level parameter) and* $\beta \in [0,1]$ *(confidence parameter) and let*

$$N \geq \frac{n}{\epsilon\beta} - 1. \tag{1}$$

Then

$$\mathbb{P}^N \left[\mathbb{P}_{\tilde{\xi}} \left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0 \right] \leq \epsilon \right] \geq 1 - \beta$$

Discussion of Results

$$N \geq \frac{n}{\epsilon\beta} - 1.$$

- Most surprising: both are independent of the distribution of $\tilde{\xi}$!
- Could use when $\mathbb{P}$ is not known explicitly.
    - Provided we can sample from it.
- Linear increase with problem size (n) is good.
- Hyperbolic in $\epsilon\beta$ is not too bad.

Least Squares

- We desire a polynomial of degree $n-1$ that interpolates the points $(a_i, y_i), i = 1, \ldots, m$.
- The problem is
$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - y\|^2$$
where,
$$A = \begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_m & \cdots & a_m^{n-1} \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \ y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$
- Solved by $\mathbf{x} = (A^T A)^{-1} A^T y$.

Robust Least Squares

- Suppose horizontal coordinates are known inexactly.
  - Belong to some interval
- Matrix A becomes $A(\xi)$.
- Want to solve the problem robustly

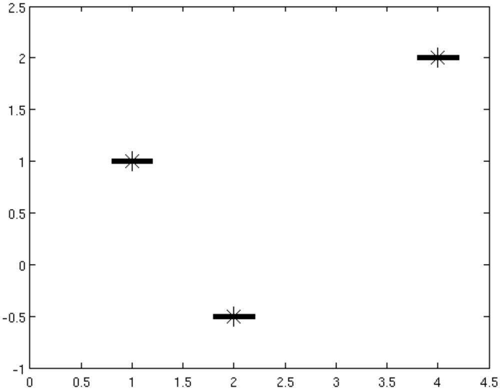
$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{\xi\in\Xi} \|A(\xi)\mathbf{x} - y\|^2$$

- One standard trick:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n,\gamma\in\mathbb{R}_+} \ \gamma \\ &\text{subject to } \ \|A(\xi)\mathbf{x} - y\|^2 \leq \gamma \qquad\qquad \forall\xi \in \Xi \end{aligned}$$

Robust Least Squares Example

- Fit a parabola.
- Three (unknown) points
    - $(1, 1)$
    - $(2, -0.5)$
    - $(4, 2)$
- x-coordinate may be ± 0.2.

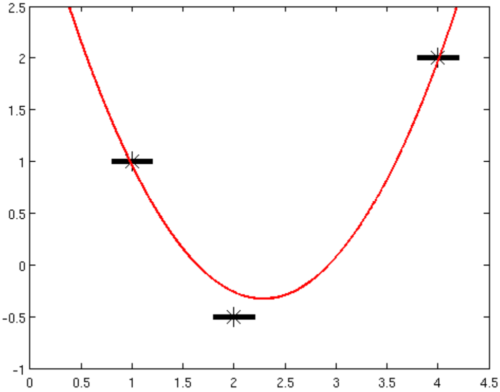


$\epsilon = 0.1$ $\beta = 0.1$

$N = 299$

Robust Least Squares Example

- Fit a parabola.
- Three (unknown) points
    - $(1, 1)$
    - $(2, -0.5)$
    - $(4, 2)$
- x-coordinate may be ± 0.2.
- Solution to Sampled Problem.

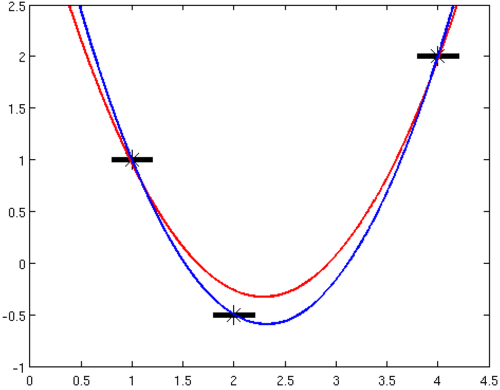


$\epsilon = 0.1$ $\beta = 0.1$

$N = 299$

Robust Least Squares Example

- Fit a parabola.
- Three (unknown) points
    - $(1, 1)$
    - $(2, -0.5)$
    - $(4, 2)$
- x-coordinate may be ± 0.2.
- Solution to Sampled Problem.
- Expected Value Solution.



$\epsilon = 0.1$ $\beta = 0.1$

$N = 299$

Estimating Solution Quality

Given a candidate solution, say $\mathbf{x}_N^*$, we can estimate its probability of infeasibility:

1. Generate more iid samples, $\tilde{\xi}^{(1')}, \ldots, \tilde{\xi}^{(M')}$.
2. Count the number of situations where the constraint is violated. Divide by total number of samples.

$$\hat{V}_{M'}(\mathbf{x}_N^*) = \frac{1}{M'} \sum_{i=1}^{M'} \mathbf{1}_{g(\mathbf{x}_N^*, \tilde{\xi}^{(i')})>0}$$

3. Then $\hat{V}_{M'}(\mathbf{x}_N^*)$ is an estimate of the constraint violation probability.

Statistics of Solution Quality

$$\mathbb{P}^N\left[\mathbb{P}_{\tilde{\xi}}\left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0 \leq \epsilon\right]\right] \geq 1 - \beta$$

- Hoeffding Inequality allows for statistical analysis of solution quality.
- Want estimate to be accurate to within ϵ'.

$$\mathbb{P}^{M'}\left[\left|\hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*)\right| \leq \epsilon'\right] \geq 1 - 2\exp\{-2(\epsilon')^2 M'\}$$

- If we desire confidence at least $1-\beta'$, draw $M' \geq \frac{\log(2/\beta')}{2(\epsilon')^2}$.
- Note: assessing solution quality does not require solving more optimization problems!

Statistics of Solution Quality

$$\mathbb{P}^N \left[\mathbb{P}_{\tilde{\xi}} \left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0 \leq \epsilon \right] \right] \geq 1 - \beta$$

- Hoeffding Inequality allows for statistical analysis of solution quality.
- Want estimate to be accurate to within ϵ'.

$$\mathbb{P}^{M'} \left[\left| \hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*) \right| \leq \epsilon' \right] \geq 1 - 2\exp\{-2(\epsilon')^2 M'\}$$

- If we desire confidence at least $1 - \beta'$, draw $M' \geq \frac{\log(2/\beta')}{2(\epsilon')^2}$.
- Note: assessing solution quality does not require solving more optimization problems!

Statistics of Solution Quality

$$\mathbb{P}^{M'}\left[\left|\hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*)\right| \leq \epsilon'\right] \geq 1 - \beta'$$

- Hoeffding Inequality allows for statistical analysis of solution quality.
- Want estimate to be accurate to within ϵ'.

$$\mathbb{P}^{M'}\left[\left|\hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*)\right| \leq \epsilon'\right] \geq 1 - 2\exp\{-2(\epsilon')^2 M'\}$$

- If we desire confidence at least $1 - \beta'$, draw $M' \geq \frac{\log(2/\beta')}{2(\epsilon')^2}$.
- Note: assessing solution quality does not require solving more optimization problems!

Statistics of Solution Quality

$$\mathbb{P}^{M'}\left[\left|\hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*)\right| \le \epsilon'\right] \ge 1 - \beta'$$

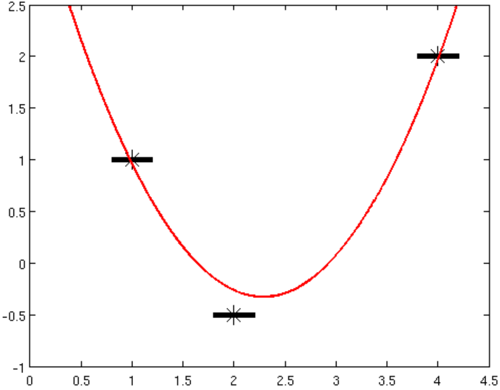
- Hoeffding Inequality allows for statistical analysis of solution quality.
- Want estimate to be accurate to within ϵ'.

$$\mathbb{P}^{M'}\left[\left|\hat{V}_{M'}(\mathbf{x}_N^*) - V(\mathbf{x}_N^*)\right| \le \epsilon'\right] \ge 1 - 2\exp\{-2(\epsilon')^2 M'\}$$

- If we desire confidence at least $1 - \beta'$, draw $M' \ge \frac{\log(2/\beta')}{2(\epsilon')^2}$.
- Note: assessing solution quality does not require solving more optimization problems!

Robust Least Squares Example

- Fit a parabola.
- Three (unknown) points
  - $(1, 1)$
  - $(2, -0.5)$
  - $(4, 2)$
- x-coordinate may be ±0.2.
- Solution to Sampled Problem.



$\epsilon = 0.1$ $\beta = 0.1$

$N = 299$

Robust Least Squares Result

- Parameters:

$$\begin{aligned} \epsilon &= 0.1 \\ \beta &= 0.1 \\ N &= 299 \end{aligned}$$

- Analysis shows solution has at most 0.6% risk of constraint violation (99.99% confidence).

Introduction to Stochastic Programming

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x},\xi)$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}(\xi)$$

$\mathbf{x}$ Vector of decision variables.

ξ Vector of problem parameters.

f Objective function, may depend on ξ.

$\mathcal{X}$ Feasible region, may depend on ξ.

- e.g.,
  $\mathcal{X}(\xi)=\{\mathbf{x}\in\mathbb{R}^n : g_i(x,\xi)\leq 0, i=1,\dots,m, \mathbf{x}\geq 0\}$.

Introduction to Stochastic Programming

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x},\xi)$$

$$\text{subject to } \mathbf{x}\in\mathcal{X}(\xi)$$

$\mathbf{x}$ Vector of decision variables.

ξ Vector of problem parameters.

f Objective function, may depend on ξ.

$\mathcal{X}$ Feasible region, may depend on ξ.

- e.g., $\mathcal{X}(\xi)=\{\mathbf{x}\in\mathbb{R}^n : g_i(x,\xi)\le 0, i=1,\dots,m,\mathbf{x}\ge 0\}$.

Stochastic Objective Function

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x},\xi)$$
$$\text{subject to } \mathbf{x}\in\mathcal{X}$$

- Now, uncertainty is in the objective function.
- What does it even mean to optimize?
    - Take the "average" cost or benefit?
    - Try to hedge against undesirable outcomes – minimize the probability of something bad?
    - Or maximize the probability of something good?
    - Hedge against undesirable outcomes – Make the "worst case scenario" as good as possible?
- Most general formulation: expected utility...

Stochastic Optimization with Uncertain Cost

$$z^* = \min \mathbb{E} f(\mathbf{x}, \tilde{\xi}) \qquad \text{(SP)}$$
$$\text{subject to } \mathbf{x} \in \mathcal{X}$$

We set up the approximating problem

$$z_N^* = \min \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}, \tilde{\xi}^i) \qquad (\text{SP}_N)$$
$$\text{subject to } \mathbf{x} \in \mathcal{X}$$

Project Goal

Given a “candidate solution,” $\hat{\mathbf{x}}$, to the stochastic problem, how can we estimate how “good” it is, compared to the optimal solution?

Optimality Gap Estimation Continued

The optimality gap $\mathbb{E} f(\hat{\mathbf{x}}, \tilde{\xi}) - \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E} f(\mathbf{x}, \tilde{\xi})$ is approximated by

$$\frac{1}{M} \sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^i) - \min_{\mathbf{x} \in X} \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{x}, \tilde{\xi}^i)$$

- Notice: need to optimize to assess quality!
- To do statistics, we repeat the above multiple times...

Nonoverlapping Batch Means

$N = 12$ Total sample size

$M = 4$ Batch size

$k = \frac{N}{M} = 3$ Number of batches

$\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4 \quad \tilde{\xi}^5 \quad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \quad \tilde{\xi}^9 \quad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$

Nonoverlapping Batch Means

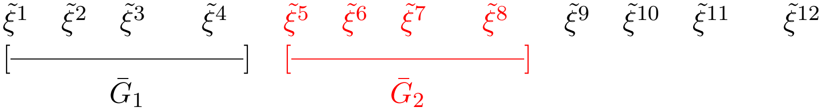
$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \tfrac{N}{M} &= 3 && \text{Number of batches} \end{aligned}$$

$$\underbrace{\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4}_{\bar{G}_1} \quad \tilde{\xi}^5 \quad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \quad \tilde{\xi}^9 \quad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$$

$$\bar{G}_1 = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(1-1)+i}) - \min_{\mathbf{x}\in\mathcal{X}} \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}, \tilde{\xi}^{M(1-1)+i})$$

Nonoverlapping Batch Means

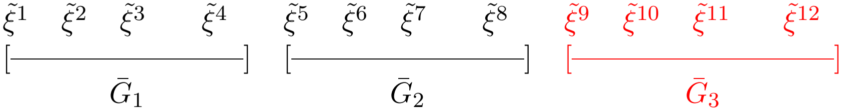
$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \tfrac{N}{M} &= 3 && \text{Number of batches} \end{aligned}$$



$$\bar{G}_2 = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(2-1)+i}) - \min_{\mathbf{x}\in\mathcal{X}} \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}, \tilde{\xi}^{M(2-1)+i})$$

Nonoverlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \frac{N}{M} &= 3 && \text{Number of batches} \end{aligned}$$



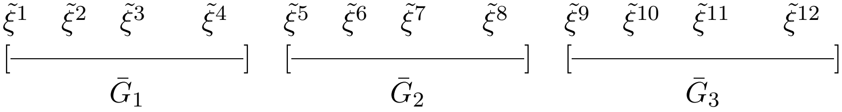
$\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4 \qquad \tilde{\xi}^5 \quad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \qquad \tilde{\xi}^9 \quad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$

$\bar{G}_1 \qquad \bar{G}_2 \qquad \bar{G}_3$

$$\bar{G}_3 = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(3-1)+i}) - \min_{\mathbf{x}\in\mathcal{X}} \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}, \tilde{\xi}^{M(3-1)+i})$$

Nonoverlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \tfrac{N}{M} &= 3 && \text{Number of batches} \end{aligned}$$



$$\bar{G}_j = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - \min_{\mathbf{x}\in\mathcal{X}} \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}, \tilde{\xi}^{M(j-1)+i})$$

$$VG = \frac{1}{(k-1)k}\sum_{j=1}^{k}(\bar{G}_j - \bar{\bar{G}})^2$$

Batch Means

Why not overlap the batches? (Meketon & Schmeiser, 1984)

Overlapping Batch Means

$N = 12$	Total sample size
$M = 4$	Batch size
$k = \left\lfloor \frac{N}{M} \right\rfloor = 3$	Number of nonoverlapping batches

$$\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4 \qquad \tilde{\xi}^5 \qquad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \qquad \tilde{\xi}^9 \qquad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$$

Overlapping Batch Means

$$\begin{aligned} N &= 12 & &\text{Total sample size} \\ M &= 4 & &\text{Batch size} \\ k = \left\lfloor \tfrac{N}{M} \right\rfloor &= 3 & &\text{Number of nonoverlapping batches} \end{aligned}$$

$$\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4 \quad \tilde{\xi}^5 \quad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \quad \tilde{\xi}^9 \quad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$$

$$[\text{———}\bar{G}_1\text{———}]$$

Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$

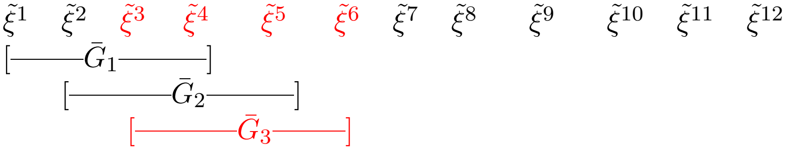
$$\tilde{\xi}^1 \quad \tilde{\xi}^2 \quad \tilde{\xi}^3 \quad \tilde{\xi}^4 \quad \tilde{\xi}^5 \quad \tilde{\xi}^6 \quad \tilde{\xi}^7 \quad \tilde{\xi}^8 \quad \tilde{\xi}^9 \quad \tilde{\xi}^{10} \quad \tilde{\xi}^{11} \quad \tilde{\xi}^{12}$$

$$[\text{———} \bar{G}_1 \text{———}]$$

$$[\text{———} \bar{G}_2 \text{———}]$$

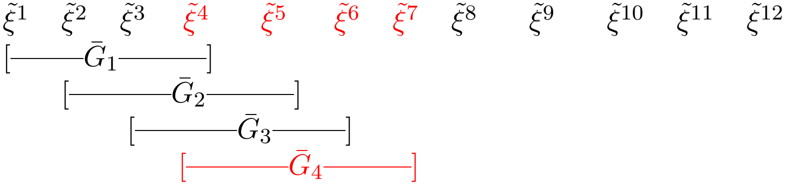
Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \tfrac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



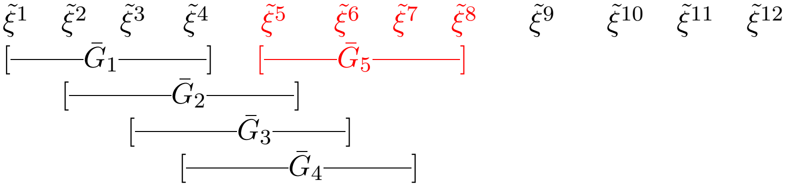
Overlapping Batches Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



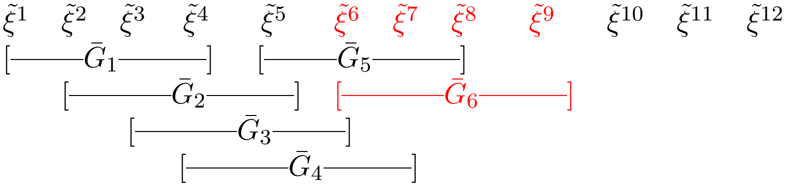
Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



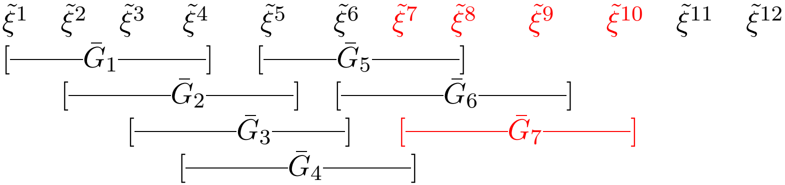
Overlapping Batch Means

$N = 12$ Total sample size
$M = 4$ Batch size
$k = \left\lfloor \frac{N}{M} \right\rfloor = 3$ Number of nonoverlapping batches



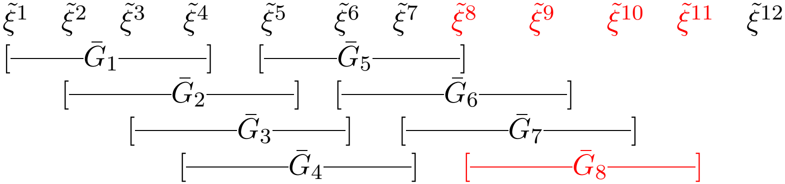
Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \frac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



$\tilde{\xi}^1$ $\tilde{\xi}^2$ $\tilde{\xi}^3$ $\tilde{\xi}^4$ $\tilde{\xi}^5$ $\tilde{\xi}^6$ $\tilde{\xi}^7$ $\tilde{\xi}^8$ $\tilde{\xi}^9$ $\tilde{\xi}^{10}$ $\tilde{\xi}^{11}$ $\tilde{\xi}^{12}$

$\bar{G}_1$ $\bar{G}_5$

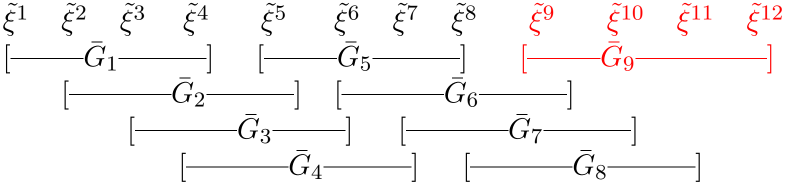
$\bar{G}_2$ $\bar{G}_6$

$\bar{G}_3$ $\bar{G}_7$

$\bar{G}_4$ $\bar{G}_8$

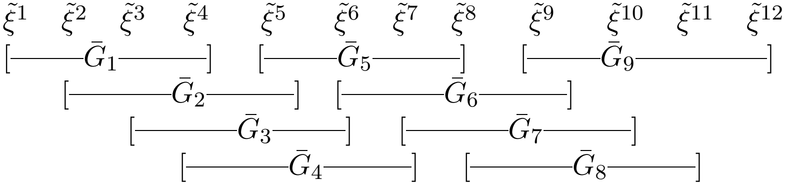
Overlapping Batch Means

$$\begin{aligned} N &= 12 && \text{Total sample size} \\ M &= 4 && \text{Batch size} \\ k = \left\lfloor \tfrac{N}{M} \right\rfloor &= 3 && \text{Number of nonoverlapping batches} \end{aligned}$$



Overlapping Batch Means

$N = 12$	Total sample size
$M = 4$	Batch size
$k = \left\lfloor \frac{N}{M} \right\rfloor = 3$	Number of nonoverlapping batches



$\tilde{\xi}^1$ $\tilde{\xi}^2$ $\tilde{\xi}^3$ $\tilde{\xi}^4$ $\tilde{\xi}^5$ $\tilde{\xi}^6$ $\tilde{\xi}^7$ $\tilde{\xi}^8$ $\tilde{\xi}^9$ $\tilde{\xi}^{10}$ $\tilde{\xi}^{11}$ $\tilde{\xi}^{12}$

$[\text{———}\bar{G}_1\text{———}]$ $[\text{———}\bar{G}_5\text{———}]$ $[\text{———}\bar{G}_9\text{———}]$

$[\text{———}\bar{G}_2\text{———}]$ $[\text{———}\bar{G}_6\text{———}]$

$[\text{———}\bar{G}_3\text{———}]$ $[\text{———}\bar{G}_7\text{———}]$

$[\text{———}\bar{G}_4\text{———}]$ $[\text{———}\bar{G}_8\text{———}]$

$$VG = \frac{1}{(k-1)(N-M+1)} \sum_{j=1}^{N-M+1} (\bar{G}_j - \bar{\bar{G}})^2$$

Partial Overlap

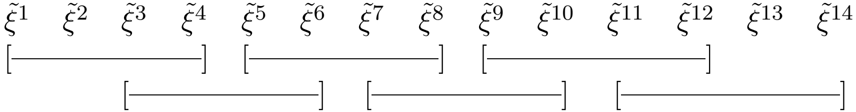


Figure: Graphical representation of $\gamma = 2$.

Let γ be the batch nonoverlap parameter.

- γ gives the number of new samples added to each new batch.
- $\gamma = 1$, maximally overlapping case discussed above.
- $\gamma = M$, classical nonoverlapping case.

Result

- The variance estimator obtained by use of overlapping batches has essentially the same bias as the nonoverlapping estimator.
- Both overlapping and nonoverlapping estimators are unbiased in the limit as $N, M, k \to \infty$.
- The variance of the sample variance is reduced when overlapping batches is used.

Variance reduction of Sample Variance

- $\gamma \equiv 1$, variance is reduced to $\dfrac{2}{3}$ of original.
- $\gamma = \dfrac{1}{3}M$, variance is reduced to $\dfrac{19}{27}$ of original.
- $\gamma = \dfrac{1}{2}M$, variance is reduced to $\dfrac{3}{4}$ of original.
- $\gamma = \dfrac{1}{L}M$, variance is reduced to $\dfrac{2L^2+1}{3L^2}$ of original.

Theoretical Conclusions

- Generate confidence intervals $\bar{\bar{G}}_l + t_{d,\alpha}\sqrt{VG_l}$.
- Confidence intervals are asymptotically valid.

Newsvendor Problem

- You purchase newspapers from a supplier at cost c, and sell them to the public at price r.
- Number of customers $\tilde{d} \geq 0$ is a random variable.
- Purchase x papers every day.
- Maximize expected profit $\mathbb{E}\left[-cx + r\min\left\{x, \tilde{d}\right\}\right]$.

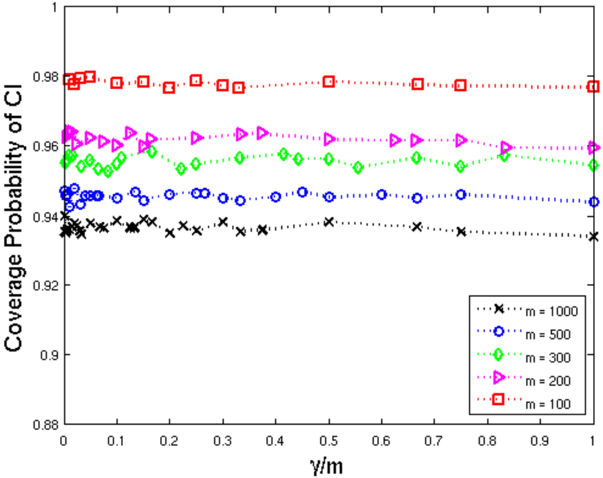
Optimal solution occurs when $\mathbb{P}\left[\tilde{d} \leq x\right] = \frac{r-c}{r}$.

Newsvendor Problem

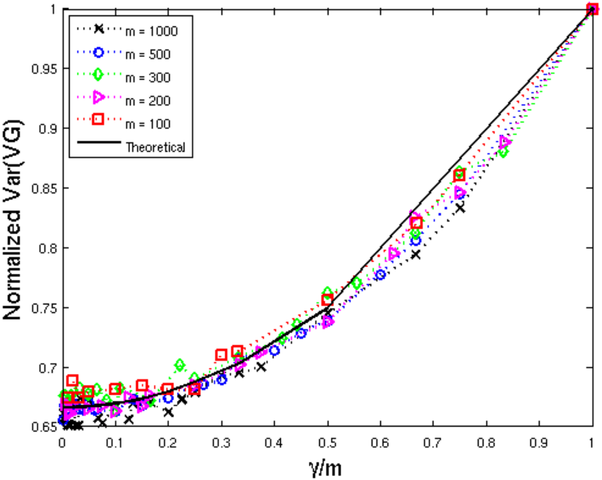
We tested the computational properties of our method on the newsvendor problem, with

- Selling price $r = 15$.
- Cost of paper $c = 5$.
- Optimal quantile $\frac{r-c}{r} = \frac{2}{3}$.
- Daily demand of newspapers is $\xi \sim U(0, 10)$.
- $\hat{\mathbf{x}} = 8.7749$.
- To solve:
    1. Take $\xi_1, \ldots, \xi_M$.
    2. Sort to get order statistics $\xi_{(1)} \leq \cdots \leq \xi_{(M)}$.
    3. Optimal solution is $\xi_{\left(\frac{2}{3}M\right)}$.

Coverage Probability



Variance Reduction of Variance Estimator



Conclusions

- My work successfully extended results from simulation into stochastic programming.
- By reusing data points, variance of the sample variance is decreased.
- Without sacrificing valid confidence intervals.
- Or increasing bias in sample variance.

Conclusions

- Calafiori and Campi give a new sampling method for solving uncertain convex programs.
- Method is likely to generate solutions that are feasible with high probability.
- Number of samples required is fairly low.
- Assumptions on the problem are fairly general.

Further Work

- Calafiori and Campi's paper inspired much further work.
- I will continue to review this.
- Focusing, in particular, on the "Ambiguous Chance Constrained Program."
- Learn about computational learning theory, probability metrics.

Questions?

Sketch of Proof

- Let $V(\mathbf{x}) = \mathbb{P}\left[g(\mathbf{x}, \tilde{\xi}) > 0\right]$.
- Proof proceeds in two steps:
    - Construct an unbiased estimator of $\mathbb{E}_{\mathbb{P}^N}\left[V(\mathbf{x}^*)\right]$.
    - Show that the estimator is bounded a.s.

Unbiased Estimator

- Let $\tilde{z}^{(1)}, \ldots, \tilde{z}^{(N+1)} \sim \mathbb{P}$ iid.
- Define

$$\min_{\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n} \left\{ c^T \mathbf{x} : g(\mathbf{x}, \tilde{z}^{(i)}) \leq 0, i = 1, \ldots, k-1, k+1, \ldots, N+1 \right\}, . \tag{SCP$_N^k$}$$

- Define indicator variables

$$v_k \equiv \mathbf{1}_{g((\mathbf{x}_N^*)^k, \tilde{z}^{(k)}) > 0}$$

- Then $\hat{V}_N = \frac{1}{N+1} \sum_{k=1}^{N+1} v_k$

Unbiased Estimator Cont.

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{N+1}}\left[v_k\right] &= \mathbb{E}_{\mathbb{P}^N}\left[\mathbb{E}_{\mathbb{P}}\left[v_k|\tilde{z}^{(1)},\ldots,\tilde{z}^{(k-1)},\tilde{z}^{(k+1)},\ldots,\tilde{z}^{N+1)}\right]\right] \\
&= \mathbb{E}_{\mathbb{P}^N}\left[\mathbb{P}\left[z\in\Xi : g((\mathbf{x}_N^*)^k,\tilde{z}^{(k)})>0\right]\right] \\
&= \mathbb{E}_{\mathbb{P}^N}\left[V((\mathbf{x}_N^*)^k)\right] \\
&= \mathbb{E}_{\mathbb{P}^N}\left[V(\mathbf{x}_N^*)\right],
\end{aligned}$$

$\hat{V}_N$ is unbiased.

Estimator is Bounded

Define

$$\min_{\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n} \left\{ c^T \mathbf{x} : g(\mathbf{x}, z^{(i)}) \leq 0, i = 1, \ldots, N+1 \right\}, \qquad (\mathrm{SCP}_{N+1})$$

- A constraint is a support constraint if removing it changes the optimal solutions.
- A convex program has, at most, n support constraints.
- Thus solutions to (SCP_N^k) differ from that of (SCP_{N+1}) at most n times.
- Then $\sum_{k=1}^{N+1} v_k \leq n \Rightarrow \hat{V}_N \leq \frac{n}{N+1}$.

◂ Return

Proof of Corollary

- By Markov's Inequality

$$\begin{aligned}\mathbb{P}^N\left[\mathbb{P}_{\tilde{\xi}}\left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0\right] > \epsilon\right] &\leq \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}^N}\left[\mathbb{P}_{\tilde{\xi}}\left[g(\mathbf{x}_N^*, \tilde{\xi}) > 0\right]\right] \\ &\leq \frac{1}{\epsilon}\frac{n}{N+1}\end{aligned}$$

- Set $\frac{1}{\epsilon}\frac{n}{N+1} = \beta$ and solve.

◂ Return

Hoeffding Inequality

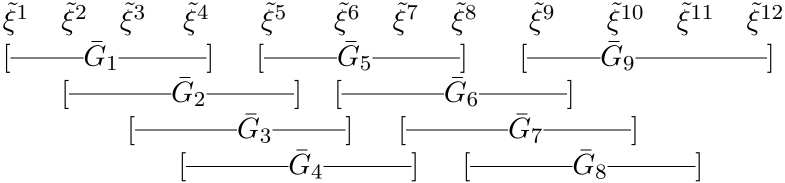
- Let $\tilde{\xi}^{(1)}, \ldots, \tilde{\xi}^{(M)}$ be independent random variables.
- Let $\mathbb{P}\left[a_i \leq \tilde{\xi}^{(i)} \leq b_i\right] = 1$.
- Let $S = \sum_{i=1}^{M} \tilde{\xi}^{(i)}$.
- Then the Hoeffding inequality is

$$\mathbb{P}\left[\left|S - \mathbb{E}\left[S\right]\right| \geq t\right] \leq 2\exp\left\{-\frac{2t^2}{\sum_{i=1}^{M}(b_i - a_i)^2}\right\}$$

- For us, $b_i \equiv 1, a_i \equiv 0, t = M\epsilon$.

◂ Return

Grand Average



- Define the grand mean estimator

$$\bar{\bar{G}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|N'(i)|}\sum_{j\in N'(i)}\left[f(\hat{\mathbf{x}},\tilde{\xi}^i) - f(\mathbf{x}_j^*,\tilde{\xi}^i)\right]$$

◂ Return # 1 ◂ Return # 2

Sketch of Proof

$$\bar{G}_j = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}_j^*, \tilde{\xi}^{M(j-1)+i})$$

$$\bar{\bar{G}} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{|N'(i)|} \sum_{j \in N'(i)} \left[f(\hat{\mathbf{x}}, \tilde{\xi}^{i}) - f(\mathbf{x}_j^*, \tilde{\xi}^{i}) \right]$$

$$\bar{D}_j = \frac{1}{M}\sum_{i=1}^{M} f(\hat{\mathbf{x}}, \tilde{\xi}^{M(j-1)+i}) - \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}^*, \tilde{\xi}^{M(j-1)+i})$$

$$\bar{\bar{D}} = \frac{1}{N}\sum_{i=1}^{N} f(\hat{\mathbf{x}}, \tilde{\xi}^{i}) - f(\mathbf{x}^*, \tilde{\xi}^{i})$$

Sketch Cont.

Show various convergences:

- $\bar{\bar{G}} - \bar{\bar{D}} \to 0$ as $N, M, k \to \infty$.
- Show $(G_j)^4$ is uniformly integrable.
- $VG - VD \xrightarrow{P} 0$.
- $VG - VD \xrightarrow{L^2} 0$.
- This gives us all the variance reduction results.
- Show that confidence interval estimators $\bar{\bar{G}} + t_{1-\alpha}\sqrt{VG} - (\bar{\bar{D}} + t_{1-\alpha}\sqrt{VD}) \xrightarrow{P} 0$.

◂ Return

Optimal Newsvendor Solution

- Suppose $\tilde{d}$ is continuous, with density ϕ, cdf Φ.
- Let $h(x, \tilde{d}) = -cx + r \max\{x, \tilde{d}\}$.

$$\begin{aligned}
\mathbb{E}\left[h(x, \tilde{d})\right] &= -cx + r\mathbb{E}\left[\max\{x, \tilde{d}\}\right] \\
&= -cx + r\int_0^x u\phi(u)du + rx\int_x^\infty \phi(u)du \\
&= -cx + r\int_0^x u\phi(u)du + rx(1-\Phi(x))
\end{aligned}$$

Optimal Newsvendor Solution Cont.

Find the optimal solution by means of the calculus:

$$\begin{aligned}\frac{d}{dx}\mathbb{E}\left[h(x,\tilde{d})\right] &= -c + rx\phi(x) + r(1-\Phi(x)) - rx\phi(x)\\ &= (r-c) - r\Phi(x)\\ &= 0\end{aligned}$$

which gives

$$\Phi(x) = \frac{r-c}{r}$$

◂ Return

Convex Objective Function

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \;& f(\mathbf{x}) \\ \text{subject to } & g(\mathbf{x},\xi) \le 0, && \xi \in \Xi \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

- New decision variable γ, $\gamma \ge f(\mathbf{x})$.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \;& \gamma \\ \text{subject to } & g(\mathbf{x},\xi) \le 0, && \xi \in \Xi \\ & f(\mathbf{x}) - \gamma \le 0 \\ & \mathbf{x} \in \mathcal{X}. \end{aligned}$$

- Combine with $g'(\mathbf{x},\xi) = \max\{g(\mathbf{x},\xi), f(\mathbf{x}) - \gamma\}$.

◂ Return