# Bounds and Solution Quality Estimation

Güzin Bayraksan

Systems and Industrial Engineering
The University of Arizona

14 August 2010

The 12th International Conference on Stochastic Programming (SP XII)

# Outline

# Outline

# A General Stochastic Program

$$
\text{(SP)} \qquad \begin{aligned} \min_x \quad & \mathbb{E} f(x, \tilde{\xi}) \\ \text{s.t.} \quad & \mathbb{E} g_i(x, \tilde{\xi}) \le 0, \;\; i = 1, 2, \ldots, m, \\ & x \in X \end{aligned}
$$

$X \subseteq \mathbb{R}^{d_x}$, set of deterministic constraints (assume nonempty and compact)
$f : \mathbb{R}^{d_x} \times \mathbb{R}^{d_{\tilde{\xi}}} \to \mathbb{R}$, objective function
$g_i : \mathbb{R}^{d_x} \times \mathbb{R}^{d_{\tilde{\xi}}} \to \mathbb{R}$, other constraints

$\tilde{\xi}$ is a random vector on probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
Expectations are taken with respect to the distribution of $\tilde{\xi}$, which is assumed independent of $x$. Also assume all expectations are finite for all $x \in X$.

The type of problem depends on $f$, $m$, $g_i$, $i = 1, 2, \ldots m$ and $X$.

# A General Stochastic Program: Two Special Cases

- **Two-stage stochastic linear program with recourse:**

  $m = 0$, $X = \{Ax = b, x \geq 0\}$, and $f(x, \tilde{\xi}) = cx + h(x, \tilde{\xi})$, where

$$h(x, \tilde{\xi}) = \min_{y} \ \tilde{q}y$$
$$\text{s.t.} \quad \tilde{W}y = \tilde{r} - \tilde{T}x, \quad y \geq 0.$$

  $\tilde{\xi} = (\tilde{q},\ \tilde{W},\ \tilde{R},\ \tilde{T})$

- **Stochastic program with a probabilistic constraint:**

  $m = 1$, $f(x, \tilde{\xi}) = cx$ and

$$g_1(x, \tilde{\xi}) = \mathbb{I}(\tilde{A}'x \geq \tilde{b}') - \alpha,$$

  $\mathbb{I}(\cdot)$: indicator function, $\alpha \in (0, 1)$, $\tilde{\xi} = (\tilde{A}', \tilde{b}')$

# A General Stochastic Program: Difficulties

- **A major challenge:** Calculating the multidimensional expectations that appear in (SP)—even for a fixed $x$.

  Calculating Multidimensional Integrals:

  - Analytical integration,
  - Quadrature methods,
  - Monte Carlo simulation

- **Another major challenge:** Even if the expectations can be calculated for a fixed $x$, optimization can be challenging: nonconvex, discontinuous . . .

  Solution Methods:

  - Exact solution methods,
  - Bounding approximations,
  - Monte Carlo simulation (external to/internal to an optimization algorithm)

# Why and When Use Bounds and Solution Quality Estimation?

- Bounds play a key role in proving optimality or near-optimality of a candidate solution
  - In integer and nonlinear optimization, bounds are often obtained via integrality, Langrangian and semidefinite relaxations
- Indispensable part of "exact" solution algorithms *(but we won't cover this)*
- Given the difficulties in solving (SP), "approximate" solution methods are often used. In this case, what is the quality of the solution obtained? Can we bound this quality? Can we "estimate" this bound, if we cannot calculate exactly? *(we will discuss this in detail)*

# Why and When Use Bounds and Solution Quality Estimation?

- "Bounding approximations" replace the difficult expectations in (SP) by easier to calculate functions/probability distributions/etc. that provide upper/lower bounds on optimal value *(we'll not cover these)*

# Why and When Use Bounds and Solution Quality Estimation?

- "Bounding approximations" replace the difficult expectations in (SP) by easier to calculate functions/probability distributions/etc. that provide upper/lower bounds on optimal value *(we'll not cover these)*
  - Two/Multi-stage linear programs with recourse: Jensen's and Edmundson-Madansky inequalities (e.g., Birge, Edirisinghe, Fraundorfer, Kall, Wallace, Ziemba . . . )

# Why and When Use Bounds and Solution Quality Estimation?

- "Bounding approximations" replace the difficult expectations in (SP) by easier to calculate functions/probability distributions/etc. that provide upper/lower bounds on optimal value *(we'll not cover these)*
  - Two/Multi-stage linear programs with recourse: Jensen's and Edmundson-Madansky inequalities (e.g., Birge, Edirisinghe, Fraundorfer, Kall, Wallace, Ziemba . . . )
  - Probabilistic constraints: by replacing these constraints with expressions using probability inequalities, e.g., generalizations of Chebyshev's inequality (e.g., Pintér, 1989, Prékopa, 1995, Nemirovski and Shapiro, 2006)

# Why and When Use Bounds and Solution Quality Estimation?

- "Bounding approximations" replace the difficult expectations in (SP) by easier to calculate functions/probability distributions/etc. that provide upper/lower bounds on optimal value *(we'll not cover these)*
  - Two/Multi-stage linear programs with recourse: Jensen's and Edmundson-Madansky inequalities (e.g., Birge, Edirisinghe, Fraundorfer, Kall, Wallace, Ziemba . . . )
  - Probabilistic constraints: by replacing these constraints with expressions using probability inequalities, e.g., generalizations of Chebyshev's inequality (e.g., Pintér, 1989, Prékopa, 1995, Nemirovski and Shapiro, 2006)
  - Multi-stage models: by aggregating the time stages and appropriately changing the information structure and time of data revelation (e.g., Birge, 1985, Kuhn, 2008, Wright, 1994)

# Outline

# A Relaxation Bound for (SP)

(Let's assume $m = 0$ for a while)

- As mentioned above, the bounds obtained through relaxations such as integrality, Lagrangian and semidefinite relaxations are essential to many optimization algorithms

- In stochastic programming, another class of relaxation bounds arises:

Approximate

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

by

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

# A Relaxation Bound for (SP)

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

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

# A Relaxation Bound for (SP)

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

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

Then,

$$\begin{aligned}
\mathbb{E} z_n^* = \mathbb{E} \min_{x \in X} \left[\frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)\right] &\le \min_{x \in X} \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)\right] \\
&= \min_{x \in X} \mathbb{E} f(x, \tilde{\xi}) \\
&= z^*
\end{aligned}$$

# A Relaxation Bound for (SP)

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

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

Then,

$$\begin{aligned}
\mathbb{E} z_n^* = \mathbb{E} \min_{x \in X}\left[\frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)\right] &\leq \min_{x \in X} \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)\right] \\
&= \min_{x \in X} \mathbb{E} f(x, \tilde{\xi}) \\
&= z^*
\end{aligned}$$

So, $\mathbb{E} z_n^* \leq z^*$.

Statistical interpretation: $z_n^*$ is a biased estimator of $z^*$

# A Relaxation Bound for (SP)

Let $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ be independent and identically distributed (iid) from the distribution of $\tilde{\xi}$. Then, the bound improves as $n$ increases:

$$\mathbb{E}z_n^* \leq \mathbb{E}z_{n+1}^* \leq z^*$$

(Norkin, Pflug, Ruszczyński, 1998, Mak, Morton, Wood, 1999)

Statistical interpretation: The bias of the estimator $z_n^*$ shrinks as the sample size increases

# A Relaxation Bound for (SP)

Let $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ be independent and identically distributed (iid) from the distribution of $\tilde{\xi}$. Then, the bound improves as $n$ increases:

$$\mathbb{E} z_n^* \leq \mathbb{E} z_{n+1}^* \leq z^*$$

(Norkin, Pflug, Ruszczyński, 1998, Mak, Morton, Wood, 1999)

Statistical interpretation: The bias of the estimator $z_n^*$ shrinks as the sample size increases

*So, how can we use this bound in solution quality estimation?*

# Solution Quality

- Assume a candidate solution, $x \in X$, is given. *How to obtain such a solution?*

# Solution Quality

- Assume a candidate solution, $x \in X$, is given. *How to obtain such a solution?*

  1. Solve a sample average approximation problem:

$$x \in \arg\min_{y \in X} \left[ \frac{1}{n} \sum_{j=1}^{n} f(y, \tilde{\xi}^j) \right]$$

  2. Stochastic Decomposition (Higle and Sen, 1996)
  3. Stochastic Approximation (Robbins-Monro, 1951, Nemirovski et al., 2009)
  4. or any other method ...

# Solution Quality

- Assume a candidate solution, $x \in X$, is given. *How to obtain such a solution?*

  1. Solve a sample average approximation problem:

$$x \in \arg\min_{y \in X} \left[ \frac{1}{n} \sum_{j=1}^{n} f(y, \tilde{\xi}^j) \right]$$

  2. Stochastic Decomposition (Higle and Sen, 1996)
  3. Stochastic Approximation (Robbins-Monro, 1951, Nemirovski et al., 2009)
  4. or any other method ...

- We are interested in determining its quality

# Solution Quality Estimation

- Define the quality of $x$ by its optimality gap:

$$\mu_x = \mathbb{E} f(x, \tilde{\xi}) - z^*$$

# Solution Quality Estimation

- Define the quality of $x$ by its optimality gap:

$$\mu_x = \mathbb{E} f(x, \tilde{\xi}) - z^*$$

- We can bound the optimality gap by:

$$\mu_x = \mathbb{E} f(x, \tilde{\xi}) - z^* \leq \mathbb{E} f(x, \tilde{\xi}) - \mathbb{E} z_n^*$$

# Solution Quality Estimation

- Define the quality of $x$ by its optimality gap:

$$\mu_x = \mathbb{E} f(x, \tilde{\xi}) - z^*$$

- We can bound the optimality gap by:

$$\mu_x = \mathbb{E} f(x, \tilde{\xi}) - z^* \leq \mathbb{E} f(x, \tilde{\xi}) - \mathbb{E} z_n^*$$

- We can estimate an upper bound on optimality gap of $x \in X$ via:

$$G_n(x) = \frac{1}{n} \sum_{j=1}^{n} f(x, \tilde{\xi}^j) - \underbrace{\min_{x \in X} \frac{1}{n} \sum_{j=1}^{n} f(x, \tilde{\xi}^j)}_{z_n^*}.$$

- $G_n(x) \geq 0$ is not asymptotically normal due to constrained optimization of sample means *(how to deal with this?)*

# Multiple Replication Procedure (MRP)

*Input:* Confidence interval (CI) level $1-\alpha$, sample size $n$, replication size $n_g$, candidate solution $x \in X$

*Output:* Approximate $(1-\alpha)$-level CI on $\mu_x = \mathbb{E} f(x,\tilde{\xi}) - z^*$

1. For $k = 1, \ldots, n_g$
   1.1. Sample iid observations $\tilde{\xi}^{k1}, \ldots, \tilde{\xi}^{kn}$ from the distribution of $\tilde{\xi}$
   1.2. Solve sample approximation problem using $\tilde{\xi}^{k1}, \ldots, \tilde{\xi}^{kn}$ to obtain $x_n^{k*}$
   1.3. Calculate $G_n^k(x) = \frac{1}{n}\sum_{j=1}^n \left( f(x,\tilde{\xi}^{kj}) - f(x_n^{k*},\tilde{\xi}^{kj}) \right)$
2. Calculate gap estimate and sample variance:

$$\bar{G}_n(n_g) = \frac{1}{n_g}\sum_{k=1}^{n_g} G_n^k(x) \quad \text{and} \quad s_G^2(n_g) = \frac{1}{n_g-1}\sum_{k=1}^{n_g}\left( G_n^k(x) - \bar{G}_n(n_g) \right)^2$$

3. Let $\epsilon_g = t_{n_g-1,\alpha} s_G(n_g)/\sqrt{n_g}$, and output one-sided CI:

$$\left[0, \bar{G}_n(n_g) + \epsilon_g\right]$$

# Multiple Replication Procedure (MRP)

Advantages:

- Widely applicable: $X$ need not be convex, $\mathbb{E}f(x, \tilde{\xi})$ need not be convex, or smooth, etc.
- It is a Monte Carlo method, so, can adapt any efficient algorithm for the underlying problem

# Multiple Replication Procedure (MRP)

Advantages:

- Widely applicable: $X$ need not be convex, $\mathbb{E}f(x, \tilde{\xi})$ need not be convex, or smooth, etc.
- It is a Monte Carlo method, so, can adapt any efficient algorithm for the underlying problem

Disadvantages:

- The computational effort to solve $n_g$ problems can be prohibitive

# Multiple Replication Procedure (MRP)

Advantages:

- Widely applicable: $X$ need not be convex, $\mathbb{E}f(x, \tilde{\xi})$ need not be convex, or smooth, etc.
- It is a Monte Carlo method, so, can adapt any efficient algorithm for the underlying problem

Disadvantages:

- The computational effort to solve $n_g$ problems can be prohibitive

When the optimality gap estimator is large, it can be due to:

1. Bias is large
2. Sampling error is large
3. The candidate solution $x$ is far from optimal

# Outline

# Single Replication Procedure (SRP)

**Remedy 1:** *When the computational effort is large, instead of* $n_g \geq 30$ *replications, can do a single replication*

# Single Replication Procedure (SRP)

**Remedy 1:** *When the computational effort is large, instead of* $n_g \geq 30$ *replications, can do a single replication*

Recall that we estimate an upper bound on optimality gap of $x \in X$ via:

$$G_n(x) = \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j) - \underbrace{\min_{x \in X} \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)}_{z_n^*}.$$

# Single Replication Procedure (SRP)

**Remedy 1:** *When the computational effort is large, instead of* $n_g \geq 30$ *replications, can do a single replication*

Recall that we estimate an upper bound on optimality gap of $x \in X$ via:

$$G_n(x) = \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j) - \underbrace{\min_{x \in X} \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)}_{z_n^*}.$$

Let $\bar{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)$ and $x_n^* \in \arg\min_{x \in X} \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)$.

# Single Replication Procedure (SRP)

**Remedy 1:** *When the computational effort is large, instead of $n_g \geq 30$ replications, can do a single replication*

Recall that we estimate an upper bound on optimality gap of $x \in X$ via:

$$G_n(x) = \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j) - \underbrace{\min_{x \in X} \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)}_{z_n^*}.$$

Let $\bar{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)$ and $x_n^* \in \arg\min_{x \in X} \frac{1}{n}\sum_{j=1}^{n} f(x, \tilde{\xi}^j)$.

Then, we can equivalently write $G_n(x)$ as:

$$G_n(x) = \bar{f}_n(x) - \bar{f}_n(x_n^*)$$

# Single Replication Procedure (SRP)

Estimate variance by:

$$s_n^2(x) = \frac{1}{n-1}\sum_{j=1}^{n}\Big((f(x,\tilde{\xi}^j) - f(x_n^*,\tilde{\xi}^j)) - (\bar{f}_n(x) - \bar{f}_n(x_n^*))\Big)^2$$

and form a $(1-\alpha)$-level confidence interval on optimality gap of $x \in X$ by:

$$\left[0,\ G_n(x) + \frac{t_{n-1,\alpha} s_n(x)}{\sqrt{n}}\right],$$

where $t_{n,\alpha}$ is the $1-\alpha$ quantile of Student's $t$ distribution with $n$ degrees of freedom.

# Justification of SRP

Notation:

- $\sigma_{\hat{x}}^2(x) = \mathrm{var}[f(\hat{x}, \tilde{\xi}) - f(x, \tilde{\xi})]$
- $X^*$ is the set of optimal solutions
- $z_\alpha$ satisfies $\mathbb{P}(N(0,1) \le z_\alpha) = 1 - \alpha$

**Theorem:**

Assume $X \neq \emptyset$ and compact, $\mathbb{E} f(x, \tilde{\xi})$ is lower semicontinuous and $\mathbb{E} \sup_{x \in X} f^2(x, \tilde{\xi}) < \infty$. Let $\hat{x} \in X$, and $\tilde{\xi}^1, \ldots, \tilde{\xi}^n$ be iid as $\tilde{\xi}$. Assume

$$\inf_{x \in X^*} \sigma_{\hat{x}}^2(x) \le \liminf_{n \to \infty} s_n^2(x_n^*) \le \limsup_{n \to \infty} s_n^2(x_n^*) \le \sup_{x \in X^*} \sigma_{\hat{x}}^2(x),$$

Then, given $0 < \alpha < 1$ in the SRP,

$$\liminf_{n \to \infty} \mathbb{P}\left( \mathbb{E} f(\hat{x}, \tilde{\xi}) - z^* \le G_n(\hat{x}) + \frac{z_\alpha s_n(x_n^*)}{\sqrt{n}} \right) \ge 1 - \alpha.$$

# A2RP: Averaged 2 Replication Procedure

- To achieve better small-sample behavior, select $n$ even and randomly partition the $n$ iid observations $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ into two.
- Calculate gap and variance estimators on each random partition,

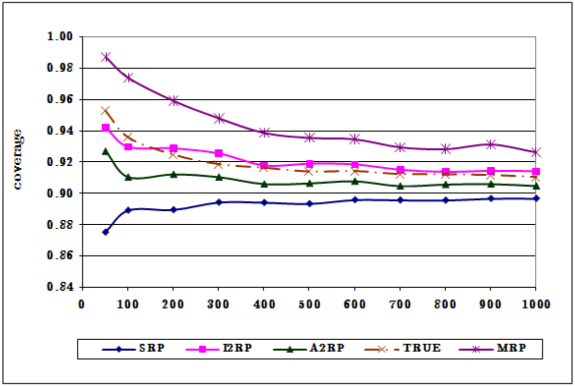
| | | |
|---|---|---|
| On Partition 1: | $G_{n/2,1}(x)$ | $s^2_{n/2,1}(x)$ |
| On Partition 2: | $G_{n/2,2}(x)$ | $s^2_{n/2,2}(x)$ |

- Average to obtain A2RP estimates:

$$
\begin{aligned}
G_n(x) &= \frac{1}{2} G_{n/2,1}(x) + \frac{1}{2} G_{n/2,2}(x) \\
s^2_n(x) &= \frac{1}{2} s^2_{n/2,1}(x) + \frac{1}{2} s^2_{n/2,2}(x)
\end{aligned}
$$

- Interval estimator formed same way.

# Computational Results




- I2RP is similar to A2RP but has uses gap estimator from one sample, and variance estimator from the other
- TRUE uses the true variance value instead of sample variance and has SRP gap estimator

# Bias and Variance Reduction

- **Remedy 2:** *When bias is large*, i.e., $\mathbb{E}z_n^* - z^* \ll 0$, the optimality gap estimator can be large even though the actual optimality gap is small.
  - Partani and Morton develop an **Adaptive Jacknife** procedure for MRP
  - It is also possible to use **stability results** to come up with heuristic bias reduction techniques

# Bias and Variance Reduction

- **Remedy 2:** *When bias is large*, i.e., $\mathbb{E}z_n^* - z^* \ll 0$, the optimality gap estimator can be large even though the actual optimality gap is small.
  - Partani and Morton develop an **Adaptive Jacknife** procedure for MRP
  - It is also possible to use **stability results** to come up with heuristic bias reduction techniques

- **Remedy 3:** *Variance reduction can be achieved via:*
  - **Randomized Quasi-Monte-Carlo**: empirical observations show that this not only reduces variance but also reduces bias.
  - Numerical Tests on other sampling schemes such as Latin Hypercube sampling can be found in e.g., Fremier et al., 2009, Linderoth et al., 2006, etc.

# Sequential Sampling Procedure

**Remedy 4:** *We can allow both candidate solution $x$ and sample size $n$ to change and stop at a "high-quality" solution*

**Step 1:** Generate a candidate solution,
**Step 2:** Assess the quality of the candidate solution,
**Step 3:** Check stopping criterion. If satisfied, stop.
Else, go to step 1.

Notation:

$k$ : iteration number
$x_k$ : candidate solution at iteration $k$
$G_k$ : optimality gap estimator of $x_k$ at iteration $k$
$s_k^2$ : associated variance estimator
$D_n(x) = \frac{1}{n}\sum_{j=1}^{n}\left[f(x,\tilde{\xi}) - f(x^*_{\text{min}})\right]$, where
$x^*_{\text{min}} \in \arg\min_{y \in X^*} \text{var}[f(x,\xi) - f(y,\xi)]$

# Assumptions

A1. The sequence of feasible candidate solutions $\{x_k\}$ has at least one limit point in $X^*$, w.p.1.

A2. Let $\{x_k\}$ be a feasible sequence (i.e., $x_k \in X$) with $x$ as one if its limit points. Let sample size $n_k$ satisfy $n_k \to \infty$ as $k \to \infty$. Then, $\liminf_{k\to\infty} P(|G_{n_k}(x_k) - \mu_x| > \delta) = 0$ for any $\delta > 0$.

A3. $G_n(x) \geq D_n(x)$, w.p.1, for all $x \in X$ and $n \geq 1$.

A4. $\liminf_{n\to\infty} s_n^2(x) \geq \sigma^2(x)$, w.p.1, for all $x \in X$.

A5. $\sqrt{n}(D_n(x) - \mu_x) \Rightarrow N(0, \sigma^2(x))$ as $n \to \infty$ for all $x \in X$, where $N(0, \sigma^2(x))$ is a normal random variable with mean zero and variance $\sigma^2(x)$. Here, "$\Rightarrow$" denotes convergence in distribution.

# Some Remarks on the Assumptions

- No need to know $x^*_{\min}$ or $D_n(x)$. Used for theoretical properties. But can view $D_n(x)$ as the "nominal" sample average estimate of $\mu_x$ with minimal variance. (Note that $\mathbb{E}D_n(x) = \mu_x$)
- We need a good way to generate the solutions (A1)
- We need good properties for point estimators (A2, A3, A4)
- Need sampling to be done via a method that satisfies the Central Limit Theorem (A5). For instance, i.i.d. sampling, anthitetic variates, bootstrapping (under certain conditions) work.

# Rules to Stop and to Increase the Sample Sizes

**Stopping Rule:** Terminate at iteration

$$T = \inf_{k \geq 1} \{k : G_k \leq h' s_k + \epsilon'\},$$

i.e., stop the first time $G_k$'s width relative to $s_k$ falls below $h' > 0$ plus a small positive number $\epsilon'$. Let $h > h'$.

# Rules to Stop and to Increase the Sample Sizes

**Stopping Rule:** Terminate at iteration

$$T = \inf_{k \geq 1} \{k : G_k \leq h' s_k + \epsilon'\},$$

i.e., stop the first time $G_k$'s width relative to $s_k$ falls below $h' > 0$ plus a small positive number $\epsilon'$. Let $h > h'$.

**Rules to Increase the Sample Size:** Select the sample size at iteration $k$ according to

$$n_k \geq \left(\frac{1}{h - h'}\right)^2 \left(c_q + 2q \ln^2 k\right),$$

where $c_q = \max\{2 \ln \left(\sum_{k=1}^{\infty} k^{-q \ln k} / \sqrt{2\pi}\alpha\right), 1\}$. Here, $q > 0$ is a parameter we can choose, which affects the number of samples we generate.

# Properties of the Sequential Procedure

**Quality Statement:** When stopped at iteration $T$, the sequential sampling procedure provides an approximate solution, $x_T$, and a confidence interval on its optimality gap, $\mu_T$, as

$$[0,\ hs_T + \epsilon],$$

# Properties of the Sequential Procedure

**Quality Statement:** When stopped at iteration $T$, the sequential sampling procedure provides an approximate solution, $x_T$, and a confidence interval on its optimality gap, $\mu_T$, as

$$[0,\ hs_T + \epsilon],$$

**Theorem:**

Let $\epsilon > \epsilon' > 0$ and $h > h' > 0$ and $0 < \alpha < 1$. Consider the above sequential sampling procedure.
(i) Assume A1 and A2 are satisfied. Then, $P(T < \infty) = 1$.
(ii) Assume A3-A5 and a moment generating function assumption are satisfied. Then,

$$\liminf_{h \downarrow h'} \mathbb{P}\left(\mu_T \leq hs_T + \epsilon\right) \geq 1 - \alpha.$$

# Outline

Now, let's consider problems with $m \geq 1$. For simplicity, consider $m = 1$:

$$\begin{aligned} z^*(\varepsilon) = \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \mathbb{P}(g(x, \tilde{\xi}) > 0) \leq \varepsilon \end{aligned}$$

# Stochastic Programs with Probabilistic Constraints

Now, let's consider problems with $m \geq 1$. For simplicity, consider $m = 1$:

$$z^*(\varepsilon) = \min_{x \in X} \quad f(x)$$
$$\text{s.t.} \quad \mathbb{P}(g(x, \tilde{\xi}) > 0) \leq \varepsilon$$

(Note: Equivalent constraint $\mathbb{P}(g(x, \tilde{\xi}) \leq 0) \geq 1 - \varepsilon$)

Now, let's consider problems with $m \geq 1$. For simplicity, consider $m = 1$:

$$\begin{aligned} z^*(\varepsilon) = \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \mathbb{P}(g(x, \tilde{\xi}) > 0) \leq \varepsilon \end{aligned}$$

(Note: Equivalent constraint $\mathbb{P}(g(x, \tilde{\xi}) \leq 0) \geq 1 - \varepsilon$)

Its sampling approximation given random sample $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ is:

$$\begin{aligned} z_n^*(\gamma) = \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(g(x, \tilde{\xi}^j) > 0) \leq \gamma \end{aligned}$$

# Stochastic Programs with Probabilistic Constraints

Now, let's consider problems with $m \geq 1$. For simplicity, consider $m = 1$:

$$
\begin{aligned}
z^*(\varepsilon) = \min_{x \in X} \quad & f(x) \\
\text{s.t.} \quad & \mathbb{P}(g(x, \tilde{\xi}) > 0) \leq \varepsilon
\end{aligned}
$$

(Note: Equivalent constraint $\mathbb{P}(g(x, \tilde{\xi}) \leq 0) \geq 1 - \varepsilon$)

Its sampling approximation given random sample $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^n$ is:

$$
\begin{aligned}
z_n^*(\gamma) = \min_{x \in X} \quad & f(x) \\
\text{s.t.} \quad & \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}(g(x, \tilde{\xi}^j) > 0) \leq \gamma
\end{aligned}
$$

Here, $\gamma \in [0, 1)$.

- $\gamma \leq \varepsilon$: a feasible solution to sampling problem is likely to be feasible to original problem
- $\gamma \geq \varepsilon$: optimal value of sampling problem can provide a lower bound on the optimal value.

# Stochastic Programs with Probabilistic Constraints

*Feasibility is an issue when sampling problem is solved!*

If $\gamma = 0$, $f(x)$ linear (w.l.o.g) and $g(\cdot, \tilde{\xi})$ is convex for all $\tilde{\xi}$ then the sampling problem is a convex program, which is computationally desirable. Let

$$B(k; p, n) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, k = 0, 1, \ldots, n.$$

be the cumulative distribution function (cdf) of a Binomial distribution with probability of success $p$ and number of trials $n$.

# Stochastic Programs with Probabilistic Constraints

*Feasibility is an issue when sampling problem is solved!*

If $\gamma = 0$, $f(x)$ linear (w.l.o.g) and $g(\cdot, \tilde{\xi})$ is convex for all $\tilde{\xi}$ then the sampling problem is a convex program, which is computationally desirable. Let

$$B(k; p, n) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, k = 0, 1, \ldots, n.$$

be the cumulative distribution function (cdf) of a Binomial distribution with probability of success $p$ and number of trials $n$.

**Campi and Garatti, 2008**

In the convex case, if sample size $n$ satisfies $B(d_x - 1, \varepsilon, n) \leq \beta$, then, with probability at least $1 - \beta$, the optimal solution to sampling problem with $\gamma = 0$ exists and is a **feasible** point of the stochastic program.

# Stochastic Programs with Probabilistic Constraints

When $m \geq 1$, solution quality estimation has two components:

1. Checking feasibility of candidate solution $x \in X$,
2. If feasible, then, determining how close its objective function value is to the optimal value.

Developed by Nemirovski & Shapiro (2006) and Luedtke & Ahmed (2008). Computations in Pagnoncelli, Ahmed & Shapiro (2009).

*We'll briefly review...*

# Checking Feasibility and Upper Bound

1. Sample $N_u$ iid observations $\tilde{\xi}^1, \tilde{\xi}^2, \ldots, \tilde{\xi}^{N_u}$ from the distribution of $\tilde{\xi}$.
2. Set $\hat{p}(x) = \frac{1}{N_u} \sum_{j=1}^{N_u} \mathbb{I}(g(x, \tilde{\xi}^j) > 0)$.
3. Let

$$\hat{\varepsilon} = \max_{\alpha \in [0,1]} \{\alpha : B(\hat{p}(x) N_u; \alpha, N_u) \geq \beta\}$$

(Note: can also use Normal approximation to Binomial)

4. If $\hat{\varepsilon} \leq \epsilon$, then, with $(1 - \beta)$-level confidence, $x$ is a feasible solution to original problem and $f(x)$ is an upper bound on $z^*(\varepsilon)$.

# Lower Bound

1. For $k = 1, \dots, M$

   1.1. Sample $N_l$ iid observations $\tilde{\xi}^{k1}, \dots, \tilde{\xi}^{kN_l}$ from the distribution of $\tilde{\xi}$

   1.2. Solve sample average approximation problem using $\tilde{\xi}^{k1}, \dots, \tilde{\xi}^{kN_l}$ to obtain $z^*_{N_l}(\gamma)$. If infeasible: $z^*_{N_l}(\gamma) = +\infty$; if unbounded: $z^*_{N_l}(\gamma) = -\infty$.

2. Sort the optimal values in nondecreasing order

$$z^*_{N_l}(\gamma)^{(1)} \leq z^*_{N_l}(\gamma)^{(2)} \leq \dots \leq z^*_{N_l}(\gamma)^{(M)}$$

3. Let $p_{N_l} = B(\lfloor \gamma N_l \rfloor; \varepsilon, N_l)$ and $L \leq M$ satisfy $B(L-1; p_{N_l}, M) \leq \beta$. Then,

$$z^*_{N_l}(\gamma)^{(L)}$$

   is a lower bound to $z^*(\varepsilon)$ with probability at least $1 - \beta$

**Possible Values:** $L = 1$ (use smallest value, $z^*_{N_l}(\gamma)^{(1)}$, as lower bound), $\gamma = \varepsilon$, $N\varepsilon \geq 10$, $\beta = 0.001$, $M \geq 10$

# Remarks: Cost vs. Risk

Consider:

$$z^*(\varepsilon) = \min_{x \in X} \; f(x) \quad \text{s.t.} \quad \mathbb{P}(g(x,\tilde{\xi}) > 0) \le \varepsilon$$

vs.

$$z_P^*(t) = \min_{x \in X} \; \mathbb{P}(g(x,\tilde{\xi}) > 0) \quad \text{s.t.} \quad f(x) \le t$$

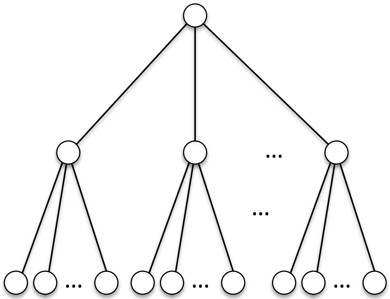
Let $\varepsilon \in [\underline{\varepsilon}, \bar{\varepsilon}]$ and $t \in [\underline{t}, \bar{t}]$ with $z_P^*(\underline{t}) = \bar{\varepsilon}$ and $z_P^*(\bar{t}) = \underline{\varepsilon}$.

Rengarajan and Morton (2009) examine the equivalence of these two problems and form confidence bounds on the efficient frontier using their sampling counterparts.

Note: the problem on the right is of the form we studied before with MRP, SRP, etc.

# Outline

# Some things we didn't cover: Multi-Stage SP



Multi-stage stochastic programs —extensions for estimating "policy" quality exists.

Some algorithms to generate policies and analysis: Pereira & Pinto (1991), Chen & Powell (1999), Birge & Donohue (1996, 2006), Philpott & Guan (2008), Linowsky & Philpott (2005)

Assessing policy quality: Chiralaksanakul (2003)

# Solution Quality Estimation in Sampling-Based Algorithms and Current Research

- Statistical tests based on Karush-Kuhn-Tucker optimality conditions (Higle & Sen, 1991; Shapiro & Homem-de-Mello, 1998)
- Within a particular algorithm, can use the knowledge gained by that algorithm and gain computational efficiencies:
  - Higle & Sen (1999) use bootstrapping to evaluate the candidate solutions generated within Stochastic Decomposition
  - Lan, Nemirovski & Shapiro (2010) form an easy to calculate lower bound with their Stochastic Approximation algorithm that is looser than our sampling lower bound.
- More are being developed, e.g., for:
  - stochastic programs with stochastic dominance constraints (Hu, Homem-de-Mello & Mehrohtra, 2010)
  - Nonlinear stochastic programs (Bastin et al., 2006, Royset, 2010)

# Summary

- Introduced bounds based on relaxations of the distributional information through Monte Carlo sampling for stochastic programs
- Reviewed methods to generate estimates of these lower bounds, hence, estimates of optimality gaps
- Reviewed a method to sequentially use these estimates in an algorithmic fashion
- Reviewed solution quality estimation in stochastic programs with probabilistic constraints

# Acknowledgments & References

Thanks to: David P. Morton.

Tutorial chapters:


- Bayraksan, G. and D.P. Morton, "Assessing Solution Quality in Stochastic Programs via Sampling," *Tutorials in Operations Research*, 5:102-122, 2009.
- Ahmed, S. and A. Shapiro, "Solving Chance-Constrained Stochastic Programs via Sampling and Integer Programming," *Tutorials in Operations Research*, 4:261-269, 2008.


Resources on the Web:

- COSP official website: `www.stoprog.org`

  Maintained by Maarten van der Vlerk and Stefan Vigerske. Contains a wealth of information.

- Searchable SP bibliography: `http://www.eco.rug.nl/mally/spbib.html`

  Maintained by Maarten van der Vlerk.

# General References and Textbooks on Stochastic Programming


- J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, 1997.
- A. Prékopa. *Stochastic Programming*. Springer Netherlands, 2009.
- A. Ruszczyński and A. Shapiro, editors. *Handbooks in Operations Research and Management Science, Volume 10: Stochastic Programming*. Elsevier, 2003.
- A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM Series on Optimization, Philadelphia, PA, 2009.

# Thank you... Questions?

(Please contact guzinb@sie.arizona.edu)