IE 495 – Lecture 17

Sampling Methods for Stochastic Programming

Prof. Jeff Linderoth

March 19, 2003

Outline

- Review—Monte Carlo Methods
  - ◇ Estimating the optimal objective function value
  - ◇ Lower Bounds
  - ◇ Upper Bounds
  - ◇ Examples
- Variance Reduction
  - ◇ Latin Hypercube Sampling
- Convergence of Optimal Solutions
  - ◇ Examples

Monte Carlo Methods

$$\min_{x \in S} \{ f(x) \equiv \mathbb{E}_P g(x;\xi) \equiv \int_\Omega g(x;\xi) dP(\xi) \}$$

- Draw $\xi^1, \xi^2, \ldots \xi^N$ from P
- Sample Average Approximation:

$$\widehat{f}_N(x) \equiv N^{-1} \sum_{j=1}^{N} g(x, \xi^j)$$

- $\widehat{f}_N(x)$ is an unbiased estimator of $f(x)$ $(\mathbb{E}[\widehat{f}_N(x)] = f(x))$.
- We instead minimize the Sample Average Approximation:

$$\min_{x \in S} \{ \widehat{f}_N(x) \}$$

Lower Bound on the Optimal Objective Function Value

$$v^* = \min_{x \in S} \{f(x)\}$$

$$\hat{v}_N = \min_{x \in S} \{\widehat{f}_N(x)\}$$

Thm:

$$\mathbb{E}[\hat{v}_N] \leq v^*$$

- The expected optimal solution value for a sampled problem of size N is $\leq$ the optimal solution value.

Estimating $\mathbb{E}[\hat{v}_N]$

- Generate M independent SAA problems of size N.
- Solve each to get $\widehat{v}_N^j$

$$L_{N,M} \equiv \frac{1}{M} \sum_{j=1}^{M} \widehat{v}_N^j$$

- The estimate $L_{N,M}$ is an unbiased estimate of $\mathbb{E}[\widehat{v}_N]$.

$$\sqrt{M}\,[L_{N,M} - \mathbb{E}(\widehat{v}_N)] \rightarrow \mathcal{N}(0, \sigma_L^2)$$

- $\sigma_L^2 \equiv \mathrm{Var}(\widehat{v}_N)$

⋆ This variance depends on the sample!

Confidence Interval

$$s_L^2(M) \equiv \frac{1}{M-1} \sum_{j=1}^{M} \left(\widehat{v}_N^j - L_{N,M} \right)^2$$

$$\left[L_{N,M} - \frac{z_\alpha s_L(M)}{\sqrt{M}}, L_{N,M} + \frac{z_\alpha s_L(M)}{\sqrt{M}} \right]$$

- These only apply if the $\widehat{v}_N^j$ are i.i.d. random variables.
- But somehow, if I could choose the samples such that they were i.i.d, and the variance among the $\widehat{v}_N^j$ was reduced, I would get a tighter confidence interval.

Upper Bounds

$$f(\hat{x}) \geq v^* \quad \forall \hat{x} \in S$$

- Generate T independent batches of samples of size $\bar{N}$

$$\mathbb{E}\left[\widehat{f}_{\bar{N}}^{j}(x) := \bar{N}^{-1} \sum_{i=1}^{\bar{N}} g(x, \xi^{i,j})\right] = f(x), \quad \text{for all } x \in X.$$

$$U_{\bar{N},T}(\hat{x}) := T^{-1} \sum_{j=1}^{T} \widehat{f}_{\bar{N}}^{j}(\hat{x})$$

More Confidence Intervals

$$\sqrt{T}\left[U_{\bar{N},T}(\hat{x}) - f(\hat{x})\right] \Rightarrow N(0, \sigma_U^2(\hat{x})), \quad \text{as } T \to \infty,$$

- $\sigma_U^2(\hat{x}) \equiv \text{Var}\left[\widehat{f}_{\bar{N}}(\hat{x})\right]$
- Estimate $\sigma_U^2(\hat{x})$ by the sample variance estimator $s_U^2(\hat{x}, T)$

$$s_U^2(\hat{x}, T) \equiv \frac{1}{T-1} \sum_{j=1}^{T} \left[\widehat{f}_{\bar{N}}^{\,j}(\hat{x}) - U_{\bar{N},T}(\hat{x})\right]^2 .$$

$$\left[U_{\bar{N},T}(\hat{x}) - \frac{z_\alpha s_U(\hat{x};T)}{\sqrt{T}}, U_{\bar{N},T}(\hat{x}) + \frac{z_\alpha s_U(\hat{x};T)}{\sqrt{T}}\right]$$

Better Living Through Sampling

- Again, if I could reduce Var $\left[\widehat{f}_{\bar{N}}(\hat{x})\right]$, while keeping $\widehat{f}^{j}_{\bar{N}}$ i.i.d, I would get a tighter confidence interval.
- That is what *Variance Reduction Techniques* are all about.
- We can reduce the variance in our estimator by better sampling

A Neopyhte's Guide to Sampling Techniques

- The main goal is to reduce Var($\widehat{f}_N(x)$) or (Var($\widehat{v}_N$)).
- Uniform (Monte Carlo) Sampling
  - Sampling with replacement
- Latin Hypercube Sampling
  - Sampling without replacement
- Importance Sampling (Dantzig and Infanger)
- Control Variates
- Common Random Number Generation

Latin Hypercube Sampling — An example

- Suppose $\Omega = \{\omega \times \omega \times \omega\}$, where ω has the following distribution:
  - $P(\omega = A) = 0.5,\ P(\omega = B) = 0.25, P(\omega = C) = 0.25$
- $|\Omega| = 27$. We would like to draw a sample of size $N = 4$.

L.H. Sample

ω_1	B	C	A	A
ω_2	A	A	B	C
ω_3	A	C	A	B

M.C. Sample

ω_1	A	A	A	B
ω_2	A	A	B	C
ω_3	A	C	A	B

- The variance of $\widehat{f}_N(x)$ for the L.H. sample will likely be less

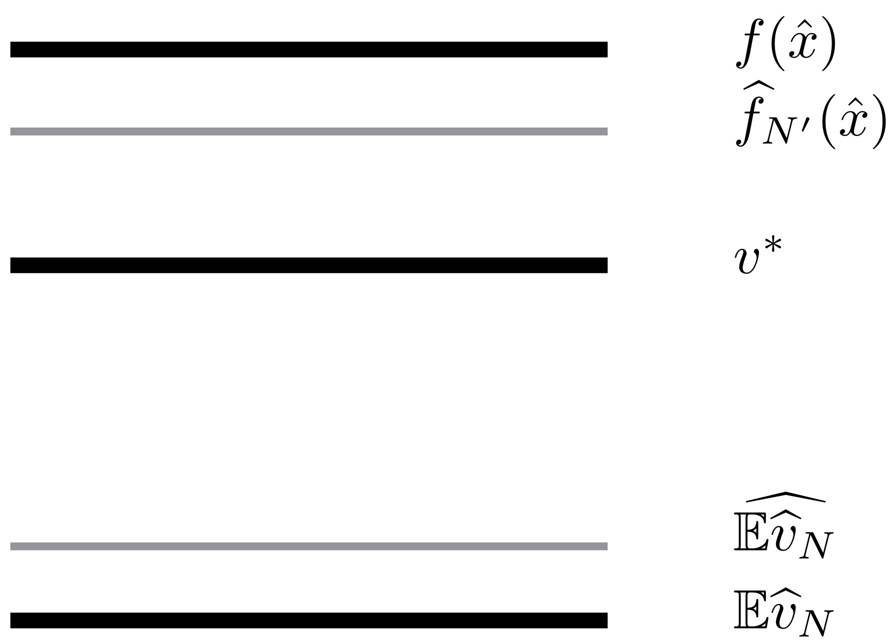
Fancy AMPL Demonstration

	Monte Carlo		Latin Hypercube	
N	$\widehat{\mathcal{Q}}(2,2)$	$\frac{s_U}{\sqrt{T}}$	$\widehat{\mathcal{Q}}(2,2)$	$\frac{s_U}{\sqrt{T}}$

Putting it all together

- $\widehat{f}_N(x)$ is the sample average function
  - ◇ Draw $\omega^1, \ldots \omega^N$ from P
  - ◇ $\widehat{f}_N(x) \equiv N^{-1} \sum_{j=1}^{N} g(x, \omega^j)$
  - ◇ For Stochastic LP w/recourse $\Rightarrow$ solve N LP's.
- $\widehat{v}_N$ is the optimal solution value for the sample average function:
  - ◇ $\widehat{v}_N \equiv \min_{x \in S} \left\{ \widehat{f}_N(x) := N^{-1} \sum_{j=1}^{N} g(x, \omega^j) \right\}$
- Estimate $\mathbb{E}(\widehat{v}_N)$ as $\widehat{\mathbb{E}(\widehat{v}_N)} = L_{N,M} = M^{-1} \sum_{j=1}^{M} \widehat{v}_N^j$
  - ◇ Solve M stochastic LP's, each of sampled size N.

The Gap



- Of most concern is the "bias" $v^* - \mathbb{E}\widehat{v}_N$.
- How fast can we make this go down in N?

A Biased Discussion

- Some problems are "ill-conditioned"
  - ◇ It takes a large sample to get an accurate estimate of the solution
- Variance reduction can help reduce the bias
  - ◇ You get the "right" small sample

An experiment

- M times – Solve a stochastic sampled approximation of size N. (Thus obtaining an estimate of $\mathbb{E}(\widehat{v}_N)$).
- For each of the M solutions $x^1, \ldots x^M$, estimate $f(\hat{x})$ by solving N' LP's.
- Test Instances

Name	Application	$\lvert\Omega\rvert$	(m_1, n_1)	(m_2, n_2)
LandS	HydroPower Planning	10^6	(2,4)	(7,12)
gbd	?	6.46×10^5	(?,?)	(?,?)
storm	Cargo Flight Scheduling	6×10^{81}	(185, 121)	(?,1291)
20term	Vehicle Assignment	1.1×10^{12}	(1,5)	(71,102)
ssn	Telecom. Network Design	10^{70}	(1,89)	(175,706)

Convergence of Optimal Solution Value

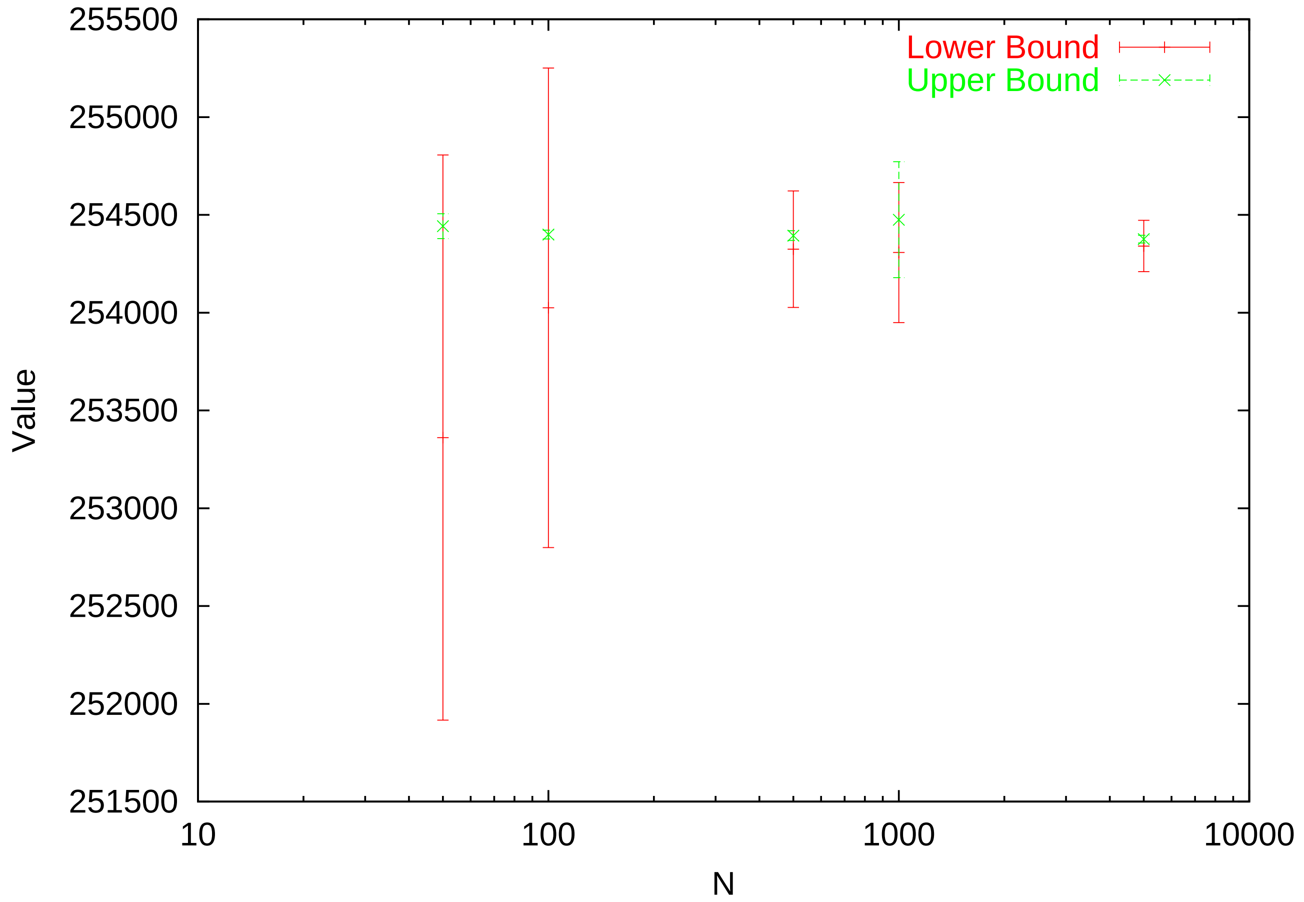
- $9 \leq M \leq 12$, $N' = 10^6$
- Monte Carlo Sampling

Instance	$N = 50$		$N = 100$		$N = 500$		$N = 1000$		$N = 5000$	
20term	253361	254442	254025	254399	254324	254394	254307	254475	254341	254376
gbd	1678.6	1660.0	1595.2	1659.1	1649.7	1655.7	1653.5	1655.5	1653.1	1655.4
LandS	227.19	226.18	226.39	226.13	226.02	226.08	225.96	226.04	225.72	226.11
storm	1550627	1550321	1548255	1550255	1549814	1550228	1550087	1550236	1549812	1550239
ssn	4.108	14.704	7.657	12.570	8.543	10.705	9.311	10.285	9.982	10.079

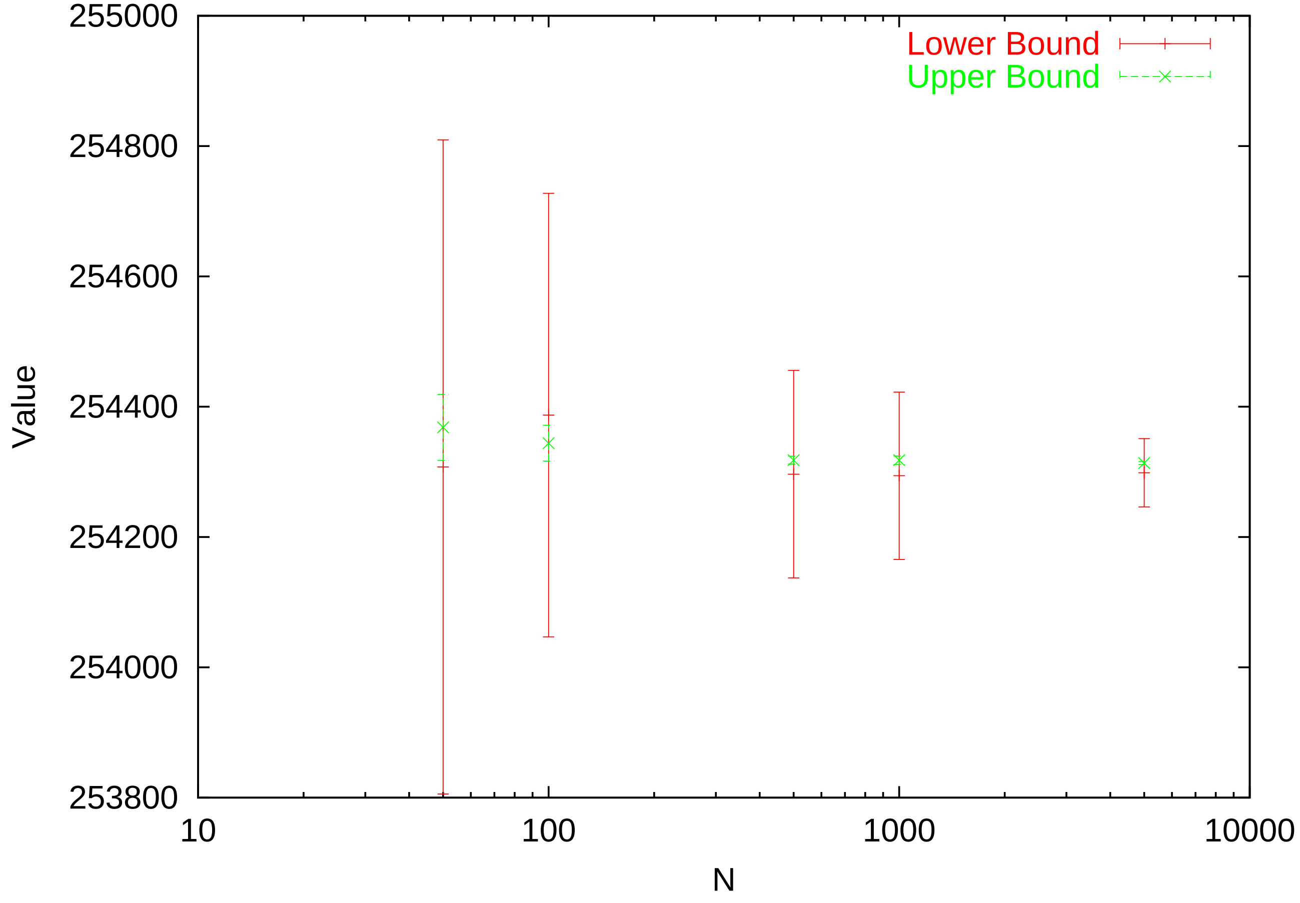
- Latin Hypercube Sampling

Instance	$N = 50$		$N = 100$		$N = 500$		$N = 1000$		$N = 5000$	
20term	254308	254368	254387	254344	254296	254318	254294	254318	254299	254313
gbd	1644.2	1655.9	1655.6	1655.6	1655.6	1655.6	1655.6	1655.6	1655.6	1655.6
LandS	222.59	222.68	225.57	225.64	225.65	225.63	225.64	225.63	225.62	225.63
storm	1549768	1549879	1549925	1549875	1549866	1549873	1549859	1549874	1549865	1549873
ssn	10.100	12.046	8.904	11.126	9.866	10.175	9.834	10.030	9.842	9.925

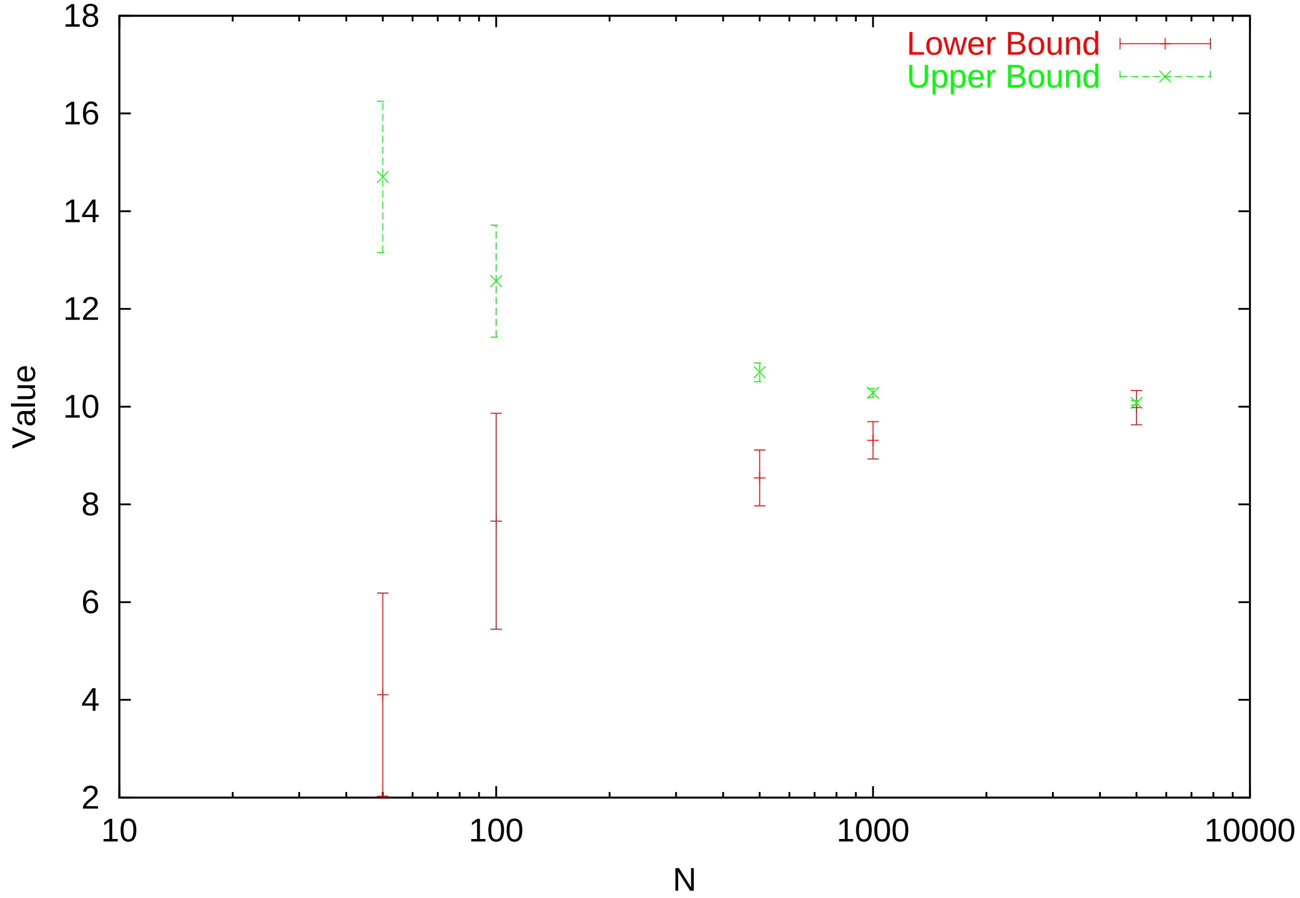
20term Convergence. Monte Carlo Sampling



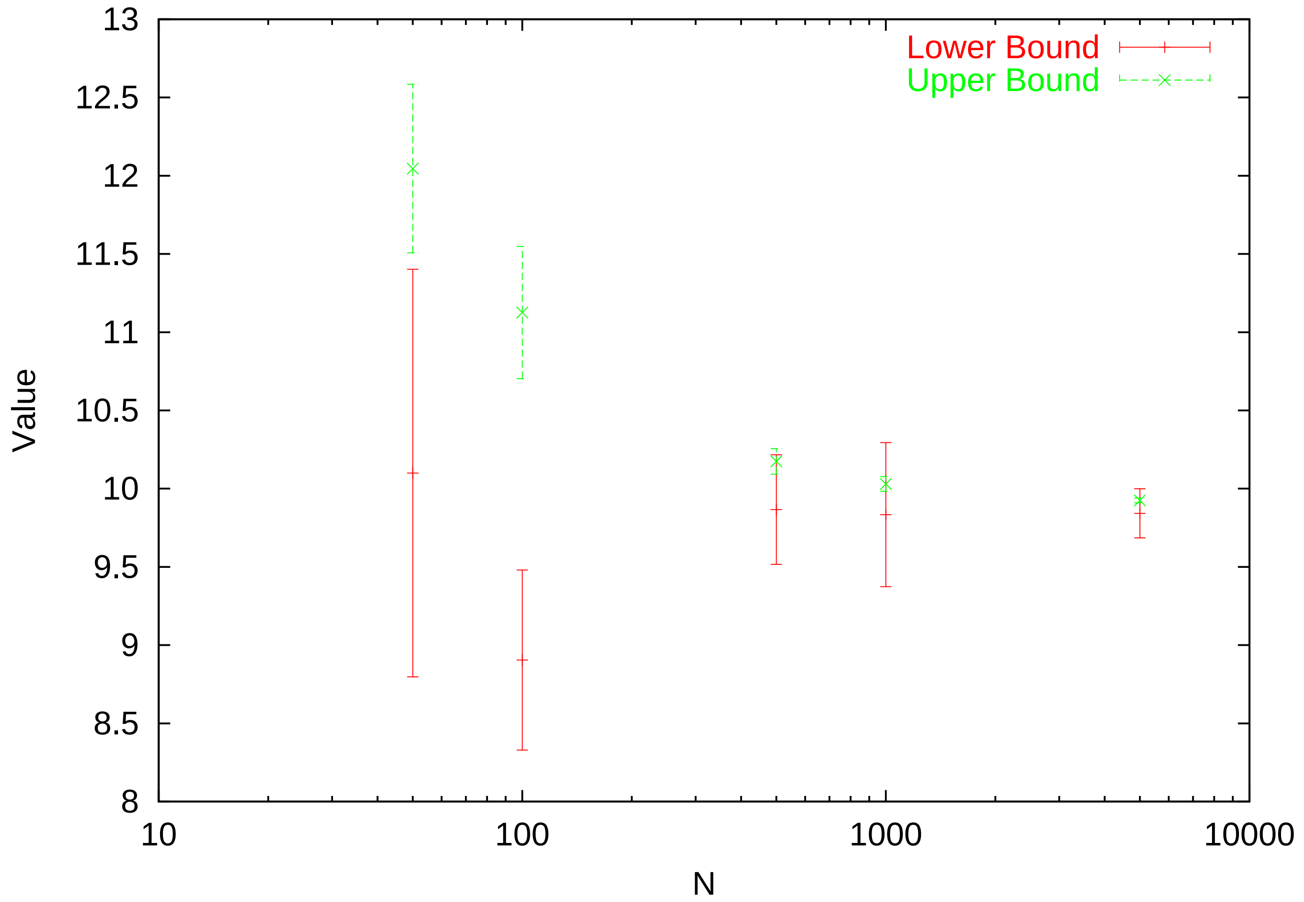
20term Convergence. Latin Hypercube Sampling



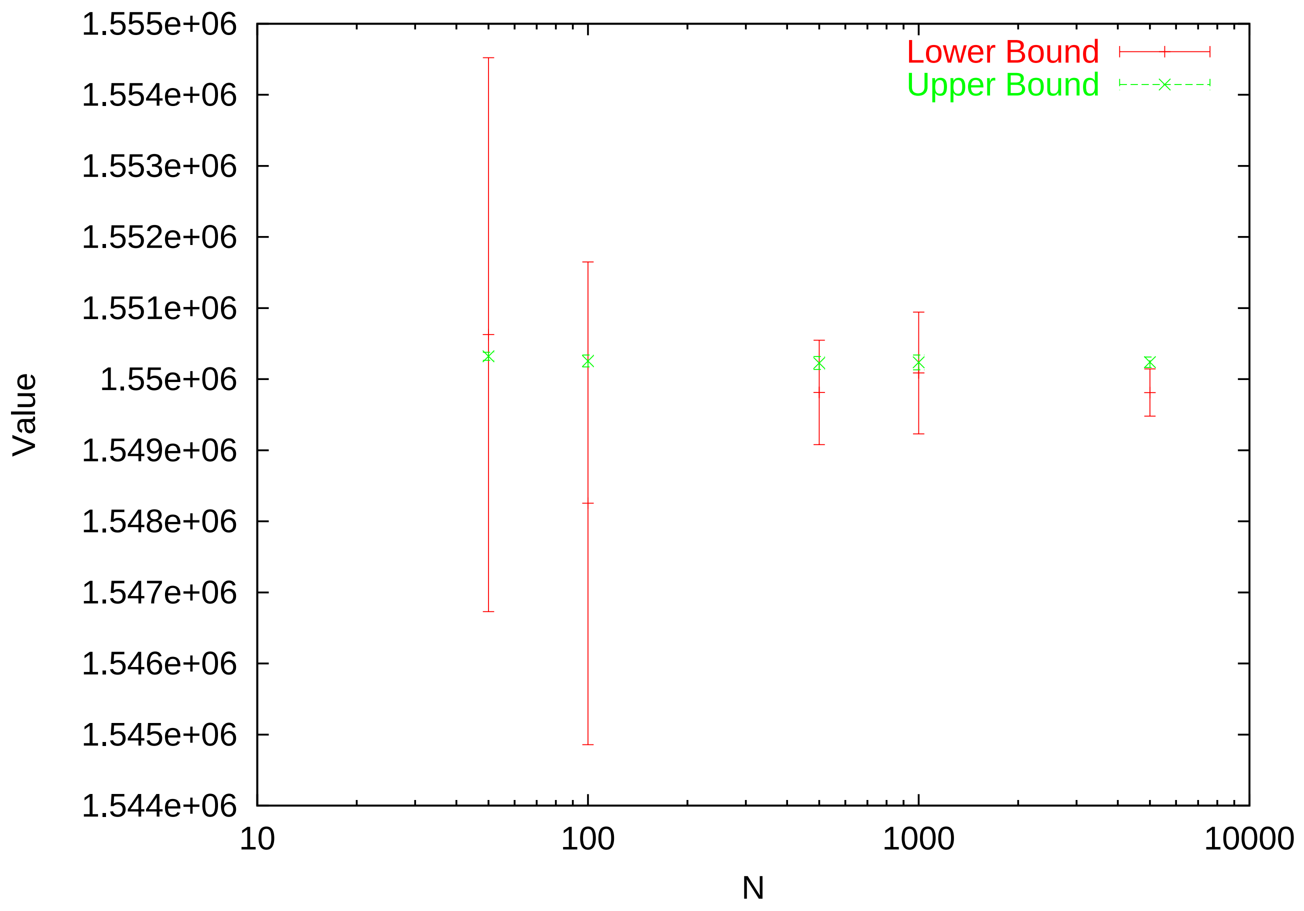
ssn Convergence. Monte Carlo Sampling



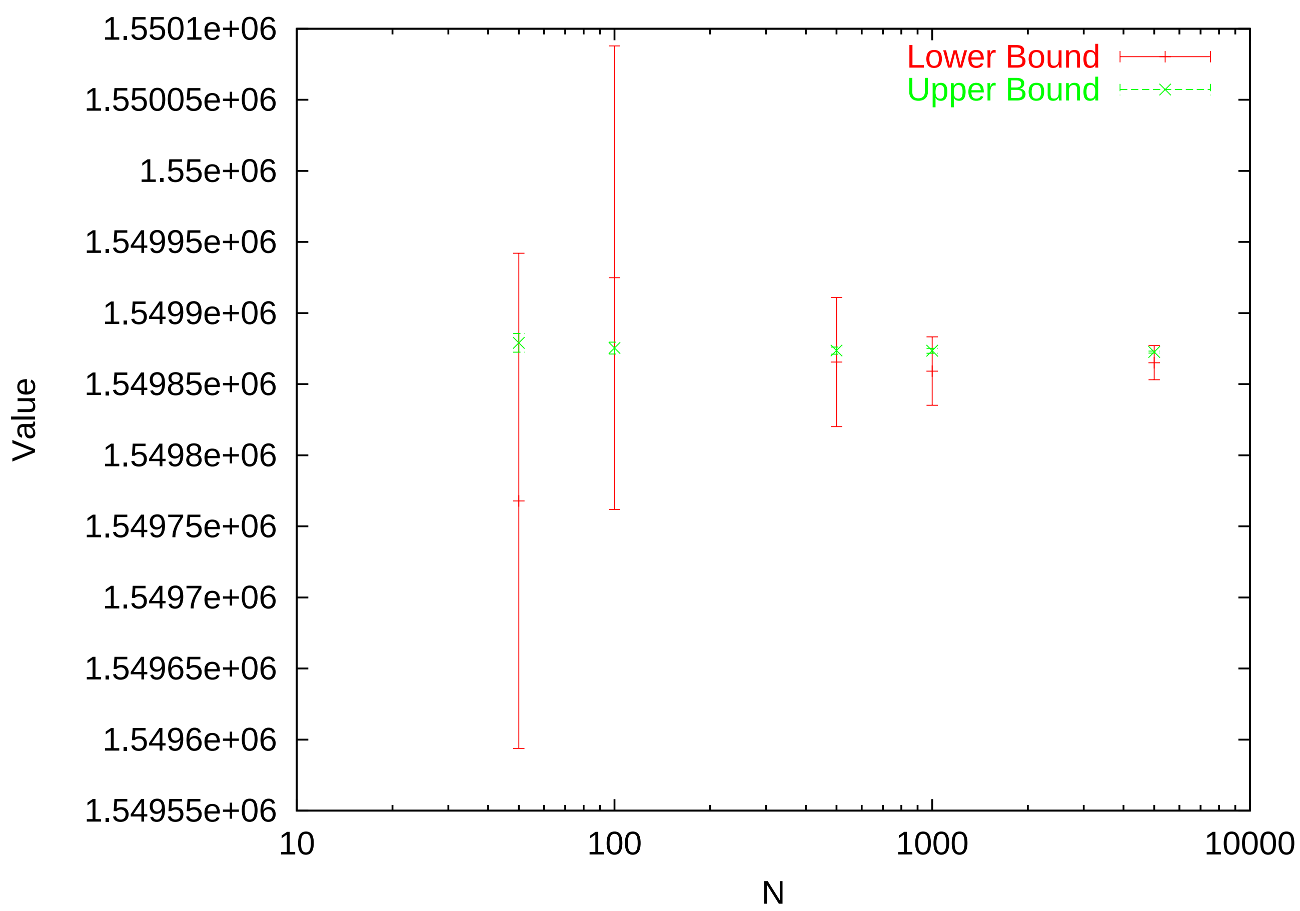
ssn Convergence. Latin Hypercube Sampling



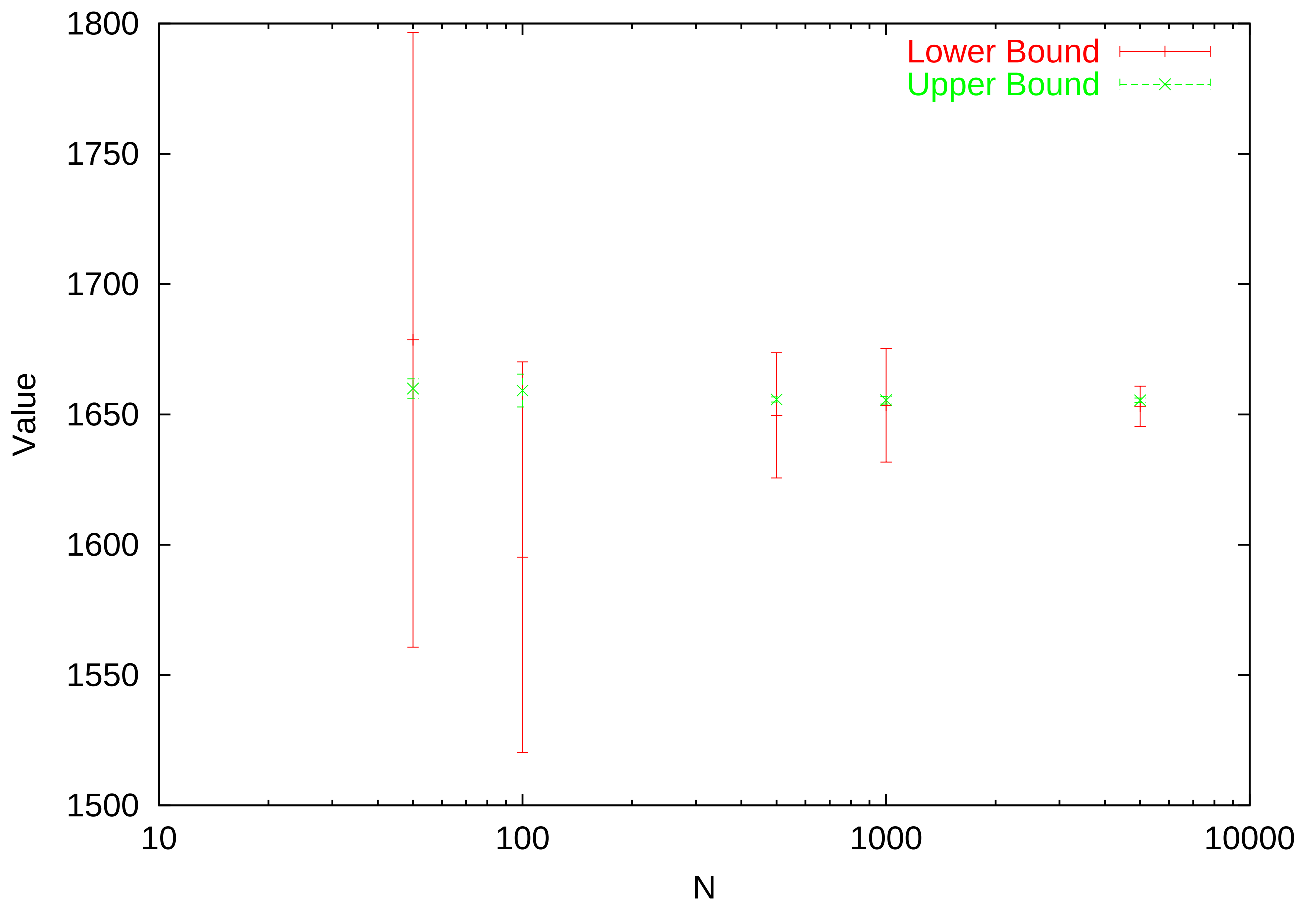
storm Convergence. Monte Carlo Sampling



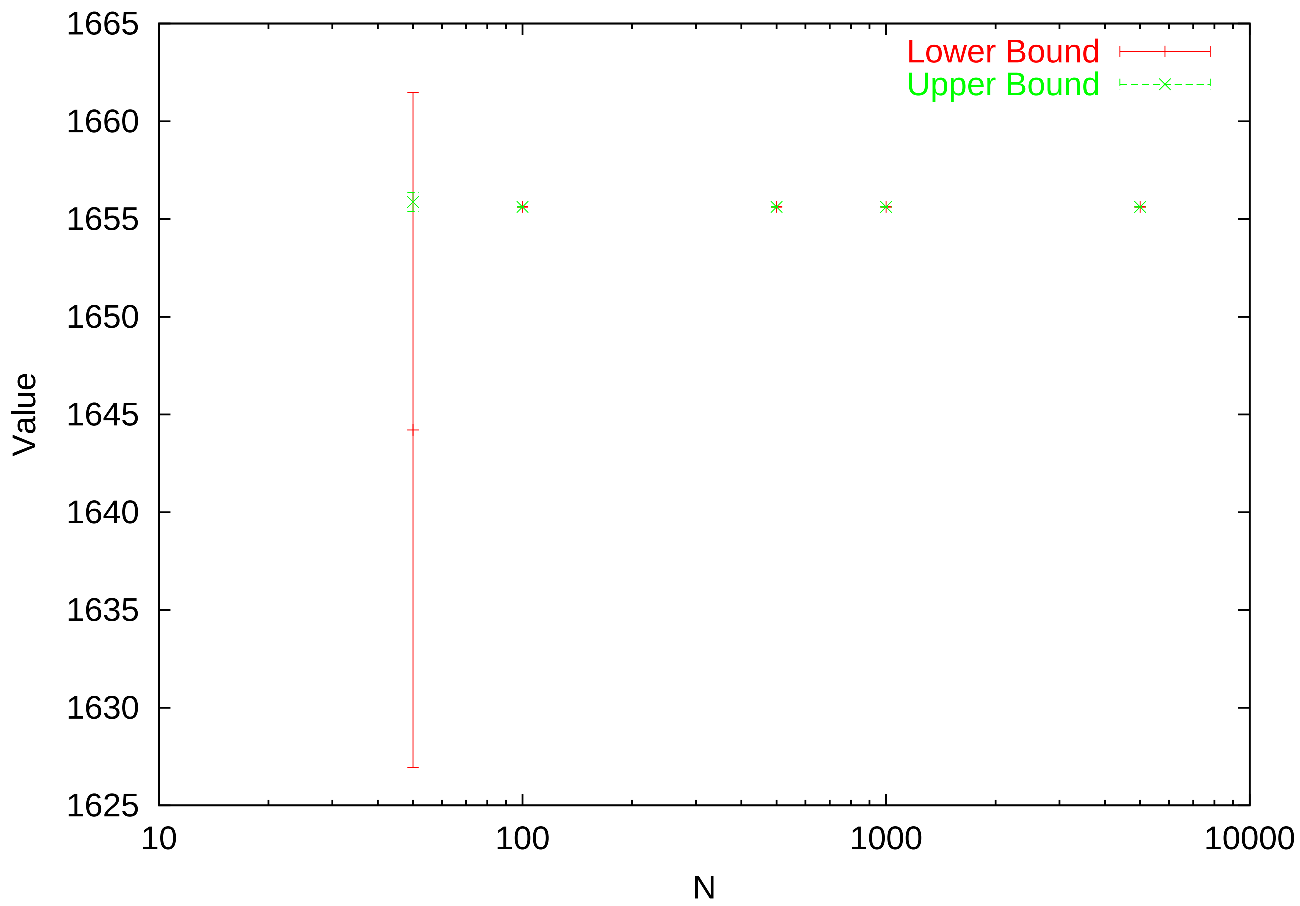
storm Convergence. Latin Hypercube Sampling



gbd Convergence. Monte Carlo Sampling



gbd Convergence. Latin Hypercube Sampling



Convergence of Optimal Solutions

- A *very interesting* recent result of Shapiro and Homem-de-Mello says the following:
- Suppose that $x^\star$ is the unique optimal solution to the "true" problem
- Let $\hat{x}_N$ be the solution to the sampled approximating problem
- Under certain conditions (like 2-stage stochastic LP with recourse with finite support), the event $(\hat{x}_N = x^\star)$ happens with probability 1 for N large enough.

⋆ The probability of this event approaches 1 exponentially fast as $N \to \infty$!!

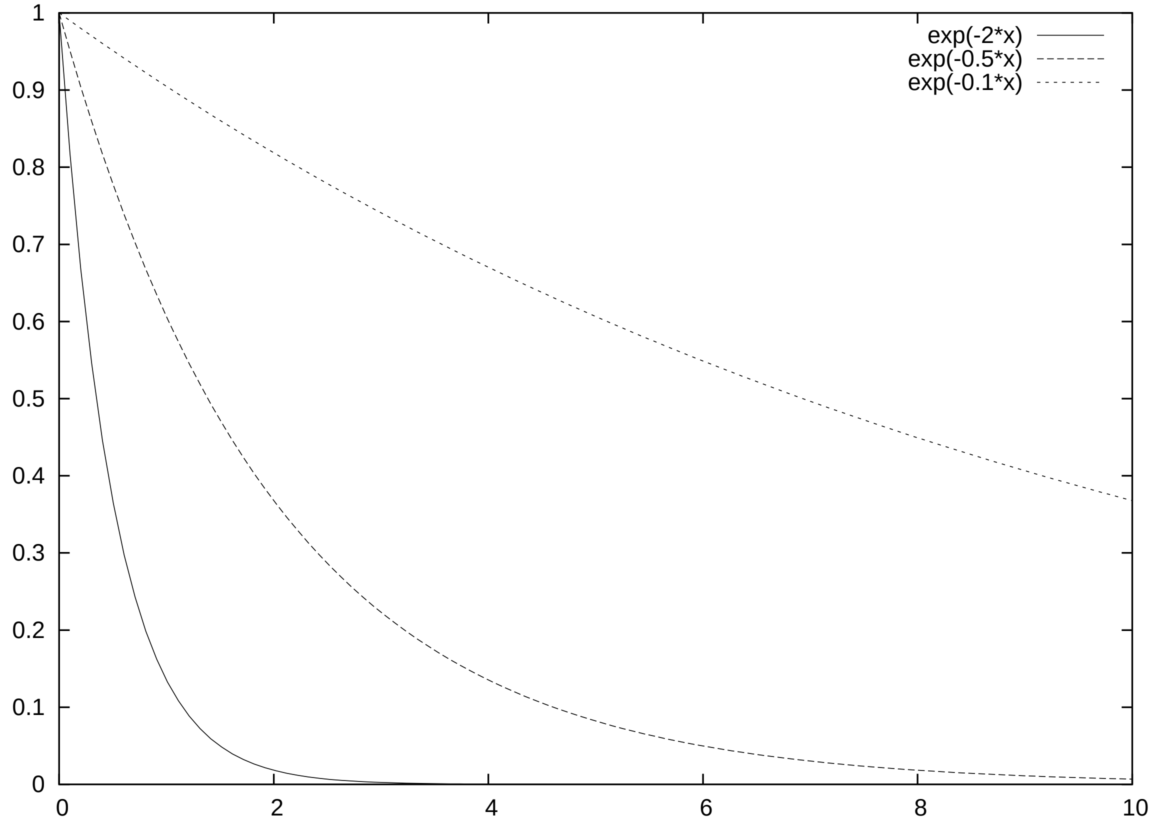
Convergence of Optimal Solutions

- There exists a constant β such that

$$\lim_{N \to \infty} N^{-1} \log[1 - P(\hat{x} = x^*)] \leq -\beta.$$

- This is a qualitative result indicating that it might not be necessary to have a large sample size in order to solve the true problem *exactly*.
- Determining a proper size N is of course difficult and problem dependent
  - ◇ Some problems are *well conditioned* – a small sample suffices
  - ◇ Others are *ill conditioned*

Function Shape



Problem Conditioning

- With the help of some heavy-duty analysis, Shapiro, Homem-de-Mello, and Kim go on to give a quantitative estimate of a stochastic program's condition.
- $g'_\omega(x^*, d)$ is the directional derivative of $g(\cdot, \omega)$ at $x^\star$ in the direction d
- $f'(x^\star, d)$ is the directional derivative of $f(\cdot)$ at $x^\star$ in the direction d
- The condition number κ of the true problem is

$$\kappa \equiv \sup_{d \in T_S(x^\star) \setminus \{0\}} \frac{\text{Var}\left[g'_\omega(x^\star, d)\right]}{\left[f'(x^\star, d)\right]^2}$$

Properties of κ

$$\kappa \equiv \sup_{d \in T_S(x^\star)\setminus\{0\}} \frac{\text{Var}\left[g'_\omega(x^\star, d)\right]}{[f'(x^\star, d)]^2}$$

- κ is related to the exponential convergence rate $\beta \approx 1/(2\kappa)$.

⋆ The sample size N required to achieve a given probability of the event $(\hat{x}_N = x^\star)$ is roughly proportional to κ

- If $f'(x^\star, d)$ is 0 (the optimal solution is not unique), then the condition number is essentially infinite.
  - ◇ (This is not really true).
- If $f'(x^\star, d)$ is small (f is "flat" in the neighborhood of the optimal solution), then κ is large
- You can also make similar statements about ϵ optimal solutions.

Examples

- Shapiro and Homem-de-Mello give a simple example of a well conditioned problem where the condition number can be computed exactly.

⋆ For a problem with 5^{1000} scenarios a sample of size $N \approx 400$ is required in order to find the true optimal solution with probability 95%!!!

⋆ Some "real" problems...

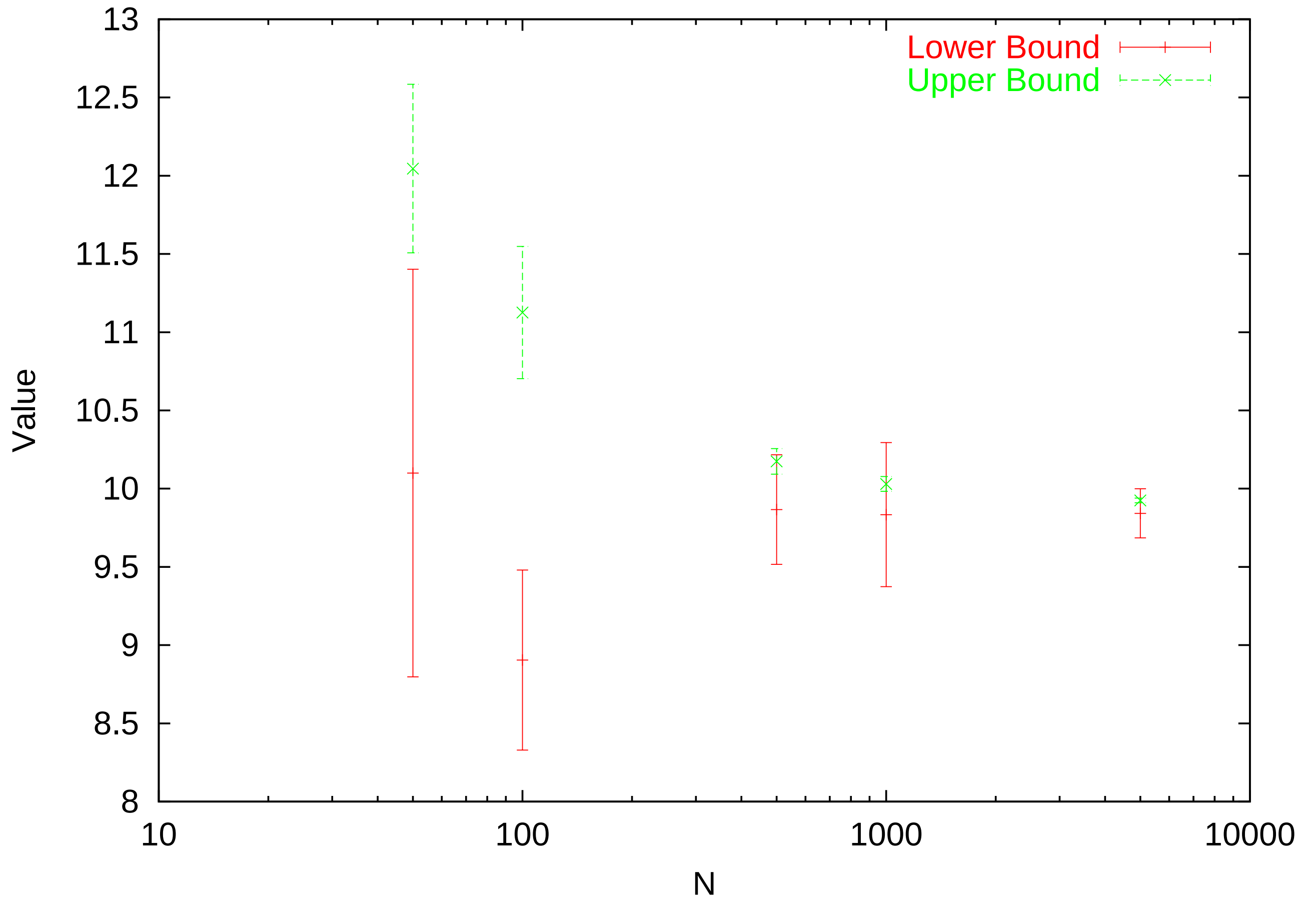
Instance	$\|\Omega\|$	$\hat{\kappa}$	$N_{\geq}$ (95%)
CEP1	216	17.45	54
APL1P	1280	1105.6	3363

Distance of ssn solutions

0.00	396.72	176.90	481.92	286.11	477.05
396.72	0.00	465.13	743.21	528.69	326.39
176.90	465.13	0.00	501.36	381.06	495.92
481.92	743.21	501.36	0.00	698.67	934.41
286.11	528.69	381.06	698.67	0.00	712.62
477.05	326.39	495.92	934.41	712.62	0.00

- For "large" sample size $(N = 5000)$, L.H. Sampling, the solutions $\hat{x}_N$ are very far apart, even though the objective functions are close to being the same.
- $f(x^\star)$ is "flat" $\Rightarrow$ This instance is ill-conditioned.
- We will require a large sample size to get an ϵ-optimal solution

ssn Convergence. Latin Hypercube Sampling



Distance of 20term solutions

0.00	259.36	700.39	87504.49	77043.47	68975.66
259.36	0.00	413.07	88080.16	77726.57	69723.88
700.39	413.07	0.00	84761.87	74631.07	67052.76
87504.49	88080.16	84761.87	0.00	2419.97	4485.82
77043.47	77726.57	74631.07	2419.97	0.00	1017.77
68975.66	69723.88	67052.76	4485.82	1017.77	0.00

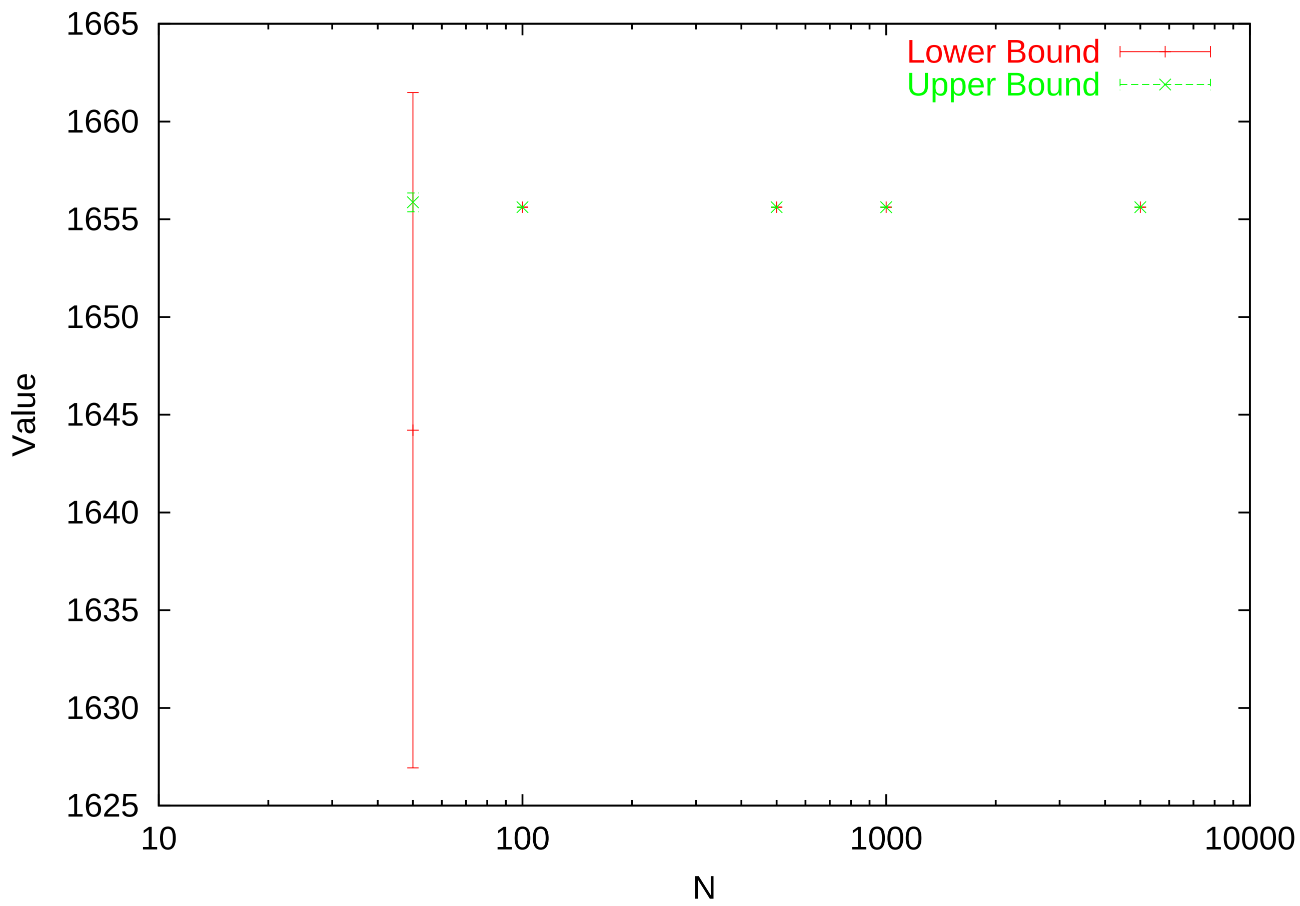
- This instance is also ill-conditioned

gbd solution distances

0.00e+00	2.49e-26	1.42e-27	6.68e-27	4.10e-28	1.47e-27
2.49e-26	0.00e+00	1.61e-26	6.72e-27	2.64e-26	3.22e-26
1.42e-27	1.61e-26	0.00e+00	3.17e-27	1.73e-27	2.96e-27
6.68e-27	6.72e-27	3.17e-27	0.00e+00	8.30e-27	1.19e-26
4.10e-28	2.64e-26	1.73e-27	8.30e-27	0.00e+00	8.17e-28
1.47e-27	3.22e-26	2.96e-27	1.19e-26	8.17e-28	0.00e+00

- This instance is extremely well conditioned

gbd Convergence. Latin Hypercube Sampling

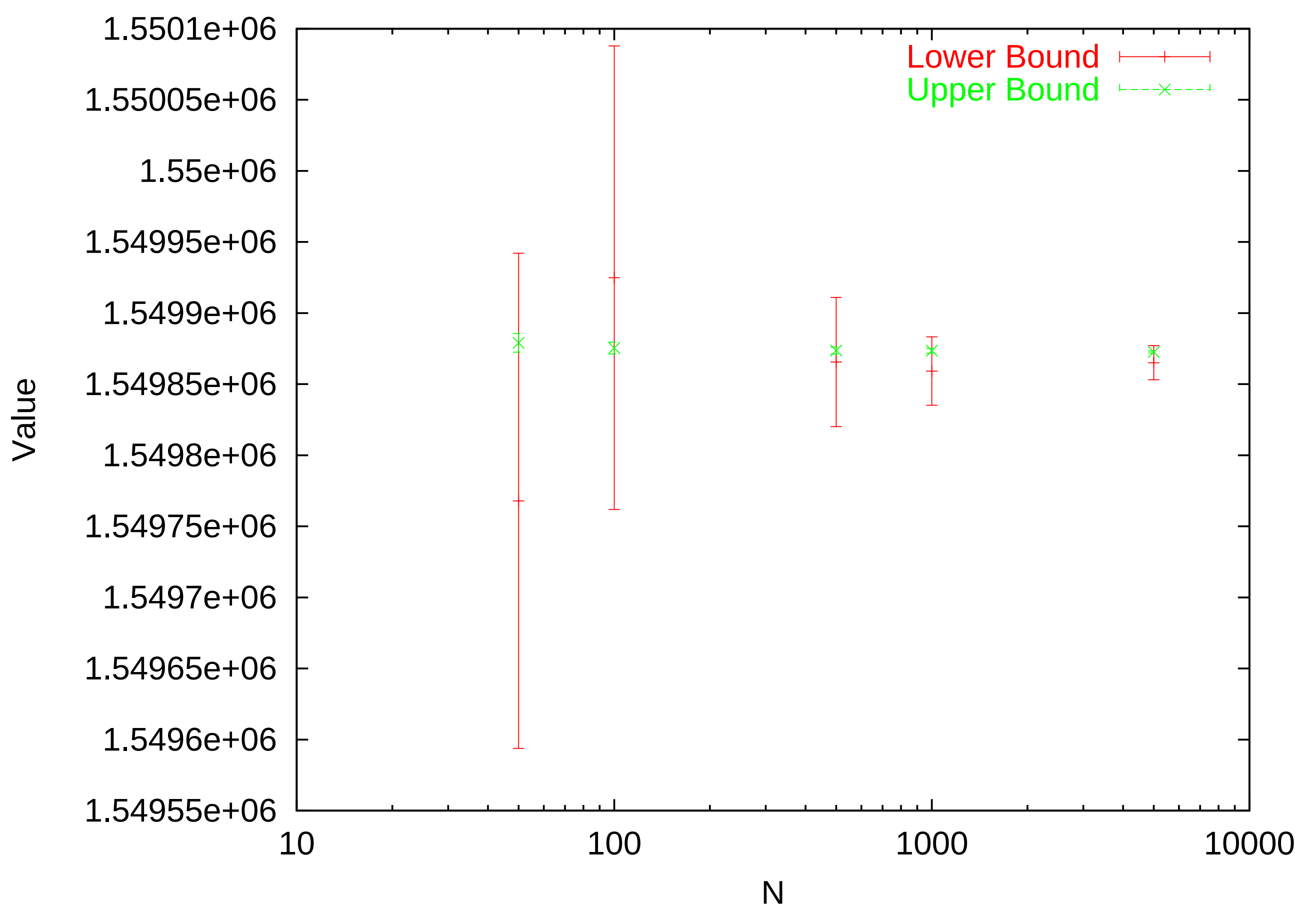


Distance of storm solutions

0.00e+00	3.51e-04	1.34e-05	1.11e-04	4.79e-05	5.27e-05
3.51e-04	0.00e+00	2.35e-04	8.94e-05	1.54e-04	1.47e-04
1.34e-05	2.35e-04	0.00e+00	4.73e-05	1.07e-05	1.30e-05
1.11e-04	8.94e-05	4.73e-05	0.00e+00	1.31e-05	1.08e-05
4.79e-05	1.54e-04	1.07e-05	1.31e-05	0.00e+00	1.12e-07
5.27e-05	1.47e-04	1.30e-05	1.08e-05	1.12e-07	0.00e+00

- This instance is also well conditioned

storm Convergence. Latin Hypercube Sampling



Conclusions

- Sometimes theory and practice do actually coincide
- You don't *need* to solve the whole problem – or consider all scenarios!
  - Using sampled approximations, you can quickly get good solutions (and bounds) to difficult stochastic programs
  - Variance reduction techniques will be very helpful
  - For "rare event" scenarios, likely importance sampling is the way to go

Next Time

- Interior Sampling Methods
  - Stochastic Decomposition