IE 495 – Lecture 16

Monte Carlo Methods for Stochastic Programming

Prof. Jeff Linderoth

March 17, 2003

Outline

- Review
  - ◇ Jensen's Inequality
  - ◇ Edmundson-Madansky Inequality
  - ◇ Partititioning
  - ◇ Using Bounds in Algorithms
- Monte Carlo Methods
  - ◇ Estimating the optimal objective function value
  - ◇ Lower Bounds
  - ◇ Upper Bounds
  - ◇ Examples

Review

- *Why* do we care about bounds in stochastic programming?
- What is Jensen's inequality?
- How is Jensen's Inequality used in stochastic programming?
- What is the Edmundson-Madansky Inequality?
- How is the Edmundson-Madansky Inequality used in stochastic programming?

Jensen's Inequality in Stochastic Programming

- If ϕ is a convex function ω of a random variable over its support Ω, then

$$\mathbb{E}_\omega \phi(\omega) \geq \phi(\mathbb{E}_\omega(\omega))$$

$$\mathbb{E}_\omega[Q(\hat{x},\omega)] \geq Q(\hat{x},\mathbb{E}_\omega[\omega])$$

- Let $\mathcal{S} = \{\Omega^l, l = 1, 2, \ldots v\}$ be some partition of Ω:

$$\mathbb{E}_\omega[Q(\hat{x},\omega)] \geq \sum_{l=1}^{v} P(\omega \in \Omega^l) Q(\hat{x}, \mathbb{E}_\omega(\omega | \omega \in \Omega^l))$$

Edmundson-Madansky Inequality

- If Ω has a (finite) set of extreme points $\text{ext}(\Omega)$, then there is some probability measure p_i (convex multipliers) so that

$$\sum_{e \in \text{ext}(\Omega)} p_{e_i} Q(\hat{x}, e) \geq Q(\hat{x}, \omega).$$

- If $\Omega = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d]$ and if the random variables in the d dimensions are independent...

$$U(\hat{x}, \omega) = \sum_{e \in \text{ext}(\Omega)} \prod_{i=1}^{d} \left(\frac{|\bar{\omega}_i - e_i|}{b_i - a_i} \right) Q(\hat{x}, e) \geq \mathbb{E}_\omega Q(w, \omega) = \mathcal{Q}(x)$$

Bounds in LShaped Method

- To use bounds, use the θ to underestimate a lower bounding function $\mathcal{Q}_L(x)$. (We called this $\mathbb{E}_\omega L(x, \omega)$ on Monday).
- Use the LShaped method to optimize the problem using $\mathcal{Q}_L(x)$.
  - ◇ Only include $\bar{\omega}$ "scenarios".
- When optimized with respect to $\mathcal{Q}_L(x)$, compare to $\mathcal{Q}_U(x)$ (what we called $\mathbb{E}_\omega U(x, \omega)$).
- If $\mathcal{Q}_U(x) - \mathcal{Q}_L(x)$ is "sufficiently small". Stop.
- Otherwise, refine the partition (which improves the bounds), and repeat.

Monte Carlo Methods

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

- Most of the theory presented holds for (*)—A very general SP problem
- Naturally it holds for our favorite SP problem:
  - ◇ $S \equiv \{x \,|Ax = b,\; x \geq 0\}$
  - ◇ $f(x) \equiv c^T x + \mathcal{Q}(x)$
  - ◇ $\mathcal{Q}(x) \equiv \mathbb{E}\{Q(x,\omega)\}$
  - ◇ $Q(x,\omega) \equiv \min_{y \geq 0} \{q(\omega)^T y | Wy = h(\omega) - T(\omega)x\}$

Sampling

- Instead of solving (*), we solve an approximating problem.
- Let $\xi^1, \ldots, \xi^N$ be N realizations of the random variable ξ:

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

- $\widehat{f}_N(x)$ is just the *sample average* function
- Since ξ^j drawn from P, $\widehat{f}_N(x)$ is an unbiased estimator of $f(x)$
  - ◇ $\mathbb{E}[\widehat{f}_N(x)] = f(x)$

Sample Variance

- Since ξ^j are independent, we can estimate $\text{Var}(\widehat{f}_N(x))$:
- This is known as the *sample variance*:

$$\hat{\sigma}^2(x) = \frac{1}{N(N-1)} \sum_{j=1}^{N} [(g(x, \xi^j) - \widehat{f}_N(x)]^2$$

Statistics Break

- Let $\chi_1, \chi_2, \ldots, \chi_n$ be independent, identically distributed (iid) random variables.
- Let $S_n = \sum_{i=1}^{n} \chi_i$
- Assume $\mu \equiv \mathbb{E}|\chi_i| < \infty$.

Weak Law of Large Numbers

$$\lim_{n \to \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) = 0 \quad \forall \delta > 0$$

Strong Law of Large Numbers

$$\lim_{n\to\infty} \frac{S_n}{n} \to \mu \quad \text{Almost surely}$$

- *Almost surely*.
  - ◇ Equivalent to "with probability 1", or..

$$P(\lim_{n\to\infty} \frac{S_n}{n} \neq \mu) = 0$$

Central Limit Theorem

- Further assume that $\chi_1, \chi_2, \ldots, \chi_n$ have finite nonzero variance σ^2:

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \mathcal{N}(0,1)$$

- $\mathcal{N}(\mu, \sigma^2)$: Normally distributed random variable with mean μ, variance σ^2.

⋆ This is an *amazing* theorem.

A More Convenient Form of the CLT

$$\begin{aligned} \frac{S_n - n\mu}{\sigma\sqrt{n}} &\approx \mathcal{N}(0,1) \\ \sqrt{n}\left(\frac{\bar{\chi} - \mu}{\sigma}\right) &\approx \mathcal{N}(0,1) \\ \sqrt{n}(\bar{\chi} - \mu) &\approx \mathcal{N}(0,\sigma^2) \end{aligned}$$

Sampling Methods

- "Interior" sampling methods.
- Sample during the course of the algorithm
  - ◇ LShaped Method (Dantzig and Infanger)
  - ◇ Stochastic Decomposition (Higle and Sen)
  - ◇ Stochastic Quasi-gradient methods (Ermoliev)
- "Exterior" sampling methods
  - ◇ Sample. Then solve problem approximating problem.
  - ◇ Can we get (statistical) bounds on key solution quantities?

Lower Bound on the Optimal Objective Function Value

$$v^* = \min_{x \in S} \{ f(x) \equiv \mathbb{E}_P g(x, \xi) \equiv \int_{\Omega} g(x; \xi) p(\xi) d\xi \}$$

For some sample $\xi^1, \ldots, \xi^N$, let

$$\hat{v}_N = \min_{x \in S} \{ \widehat{f}_N(x) \equiv N^{-1} \sum_{j=1}^{N} g(x, \xi^j) \}.$$

Thm:

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

Proof

$$v^* = \min_{x \in X} \mathbb{E}_P g(x;\xi) = \min_{x \in X} \mathbb{E}\left[N^{-1}\sum_{j=1}^{N} g(x,\xi_j)\right]$$

$$\begin{aligned}
\min_{x \in X} N^{-1}\sum_{j=1}^{N} g(x,\xi_j) &\leq N^{-1}\sum_{j=1}^{N} g(x,\xi_j) \quad \forall x \in X \quad \Leftrightarrow \\
\mathbb{E}\left[\min_{x \in X} N^{-1}\sum_{j=1}^{N} g(x,\xi_j)\right] &\leq \mathbb{E}\left[N^{-1}\sum_{j=1}^{N} g(x,\xi_j)\right] \quad \forall x \in X \Leftrightarrow \\
\mathbb{E}\left[\hat{v}_N\right] &\leq \mathbb{E}\left[N^{-1}\sum_{j=1}^{N} g(x,\xi_j)\right] \quad \forall x \in X \\
&\leq \min_{x \in X} \mathbb{E}\left[N^{-1}\sum_{j=1}^{N} g(x,\xi_j)\right] = v^*
\end{aligned}$$

Next?

- Now we need to somehow estimate $\mathbb{E}[\hat{v}_n]$
- The expected value $\mathbb{E}[\widehat{v}_N]$ can be estimated as follows.
- Generate M independent samples, $\xi^{1,j}, \dots, \xi^{N,j}$, $j = 1, \dots, M$, each of size N, and solve the corresponding SAA problems

$$\min_{x \in X} \left\{ \widehat{f}_N^j(x) := N^{-1} \sum_{i=1}^{N} g(x, \xi^{i,j}) \right\}, \tag{1}$$

- for each $j = 1, \dots, M$. Let $\widehat{v}_N^j$ be the optimal value of problem (1), and compute

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

Lower Bounds

- The estimate $L_{N,M}$ is an unbiased estimate of $\mathbb{E}[\widehat{v}_N]$.
- By our last theorem, it provides a statistical lower bound for the true optimal value v^*.
- When the M batches $\xi^{1,j}, \xi^{2,j}, \ldots, \xi^{N,j}$, $j = 1, \ldots, M$, are i.i.d. (although the elements *within* each batch do not need to be i.i.d.) have by the Central Limit Theorem that

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

Confidence Intervals

- The sample variance estimator of σ_L^2 is

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

Defining z_α to satisfy $P\{N(0,1) \le z_\alpha\} = 1 - \alpha$, and replacing σ_L by $s_L(M)$, we can obtain an approximate $(1-\alpha)$-confidence interval for $\mathbb{E}[\widehat{v}_N]$ to be

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

Let's do an example — Time Permitting

minimize

$$\mathcal{Q}(x_1, x_2) = x_1 + x_2 + 5 \int_{\omega_1=1}^{4} \int_{\omega_2=1/3}^{2/3} y_1(\omega_1, \omega_2) + y_2(\omega_1, \omega_2) d\omega_1 d\omega_2$$

subject to

$$\begin{aligned}
\omega_1 x_1 + x_2 + y_1(\omega_1, \omega_2) &\geq 7 \\
\omega_2 x_1 + x_2 + y_2(\omega_1, \omega_2) &\geq 4 \\
x_1 &\geq 0 \\
x_2 &\geq 0 \\
y_1(\omega_1, \omega_2) &\geq 0 \\
y_2(\omega_1, \omega_2) &\geq 0
\end{aligned}$$

Upper Bounds

$$v^* = \min_{x \in S} \{ f(x) \equiv \mathbb{E}_P g(x; \xi) \equiv \int_\Omega g(x; \xi) p(\xi) d\xi \}$$

- Quick, Someone prove...

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

- How can we estimate $f(\hat{x})$?

Estimating $f(\hat{x})$

- Generate T independent batches of samples of size $\bar{N}$, denoted by $\xi^{1,j}, \xi^{2,j}, \ldots, \xi^{\bar{N},j}$, $j = 1, 2, \ldots, T$, where each batch has the unbiased property, namely

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

We can then use the average value defined by

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

as an estimate of $f(\hat{x})$.

More Confidence Intervals

By applying the Central Limit Theorem again, we have that

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

where $\sigma_U^2(\hat{x}) := \text{Var}\left[\widehat{f}_{\bar{N}}(\hat{x})\right]$. We can estimate $\sigma_U^2(\hat{x})$ by the sample variance estimator $s_U^2(\hat{x};T)$ defined by

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

By replacing $\sigma_U^2(\hat{x})$ with $s_U^2(\hat{x};T)$, we can proceed as above to obtain a $(1-\alpha)$-confidence interval for $f(\hat{x})$:

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

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.

Recapping Theorems

Thm. $\mathbb{E}(\widehat{v}_N) \leq v^* \leq f(x) \forall x$

Thm. $\widehat{f}_{N'}(\hat{x}) - \mathbb{E}(\widehat{v}_N) \to f(\hat{x}) - v^*, \text{ as } N, M, N' \to \infty$

- We are mostly interested in estimating the quality of a given solution $\hat{x}$. This is $f(\hat{x}) - v^*$.
- $\widehat{f}_{N'}(\hat{x})$ computed by solving N' independent LPs.
- $\widehat{\mathbb{E}(\widehat{v}_N)}$ computed by solving M independent stochastic LPs.
- Independent $\Rightarrow$ no synchronization $\Rightarrow$ good for the Grid
- Independent $\Rightarrow$ can construct confidence intervals around the estimates

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	$\|\Omega\|$	(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

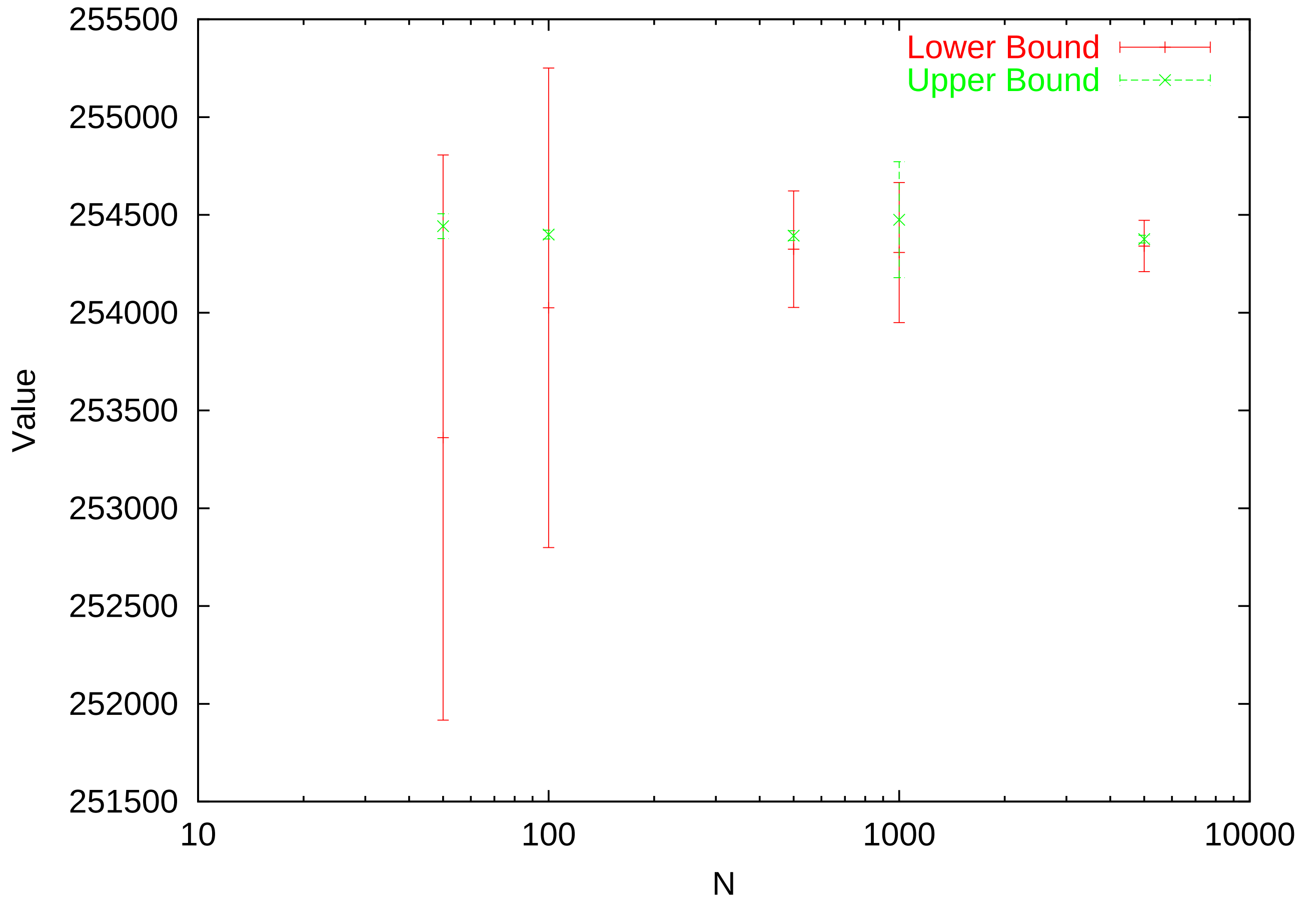
- $9 \leq M \leq 12$, $N' = 10^6$
- Monte Carlo Sampling

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

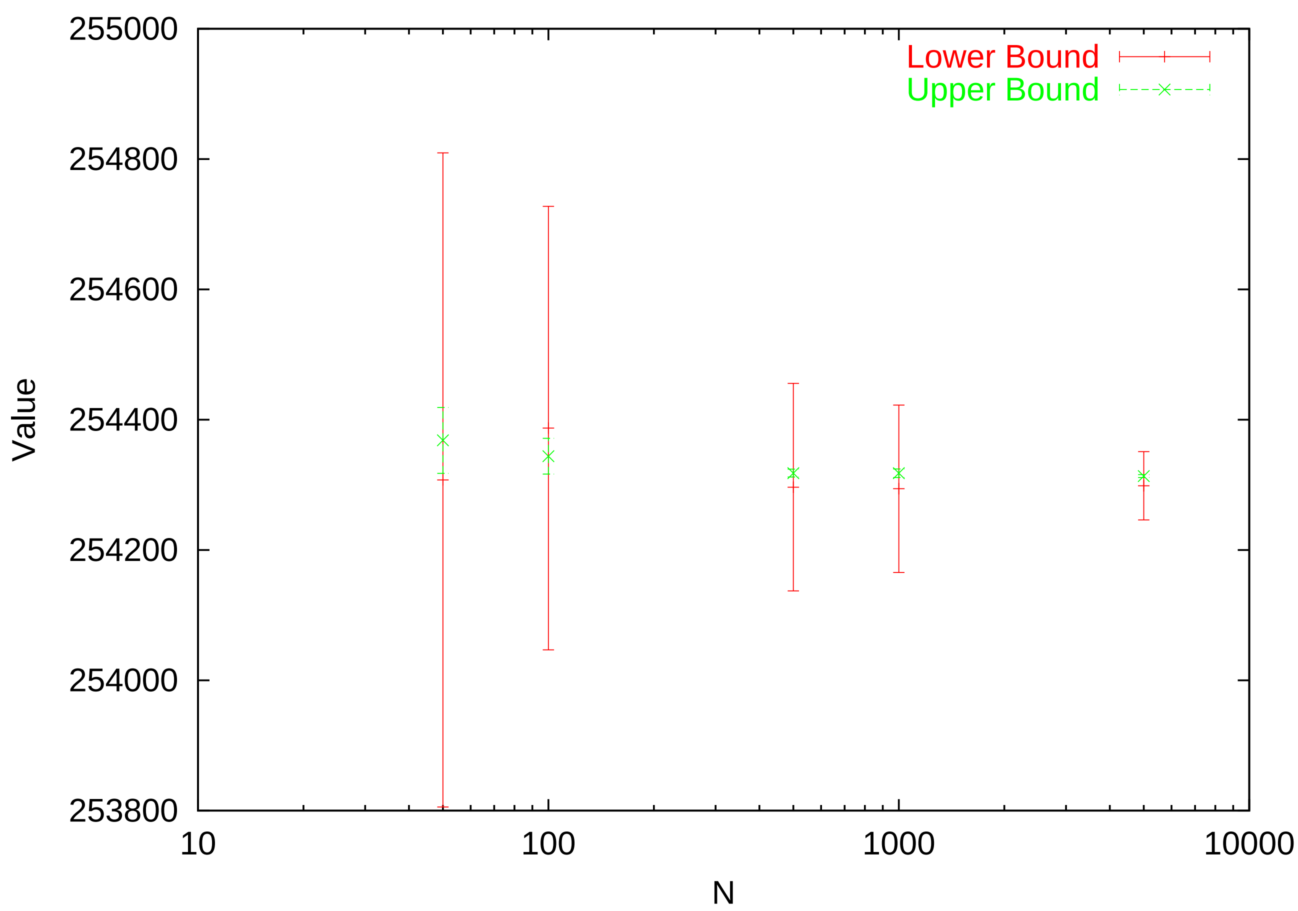
- Latin Hypercube Sampling

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

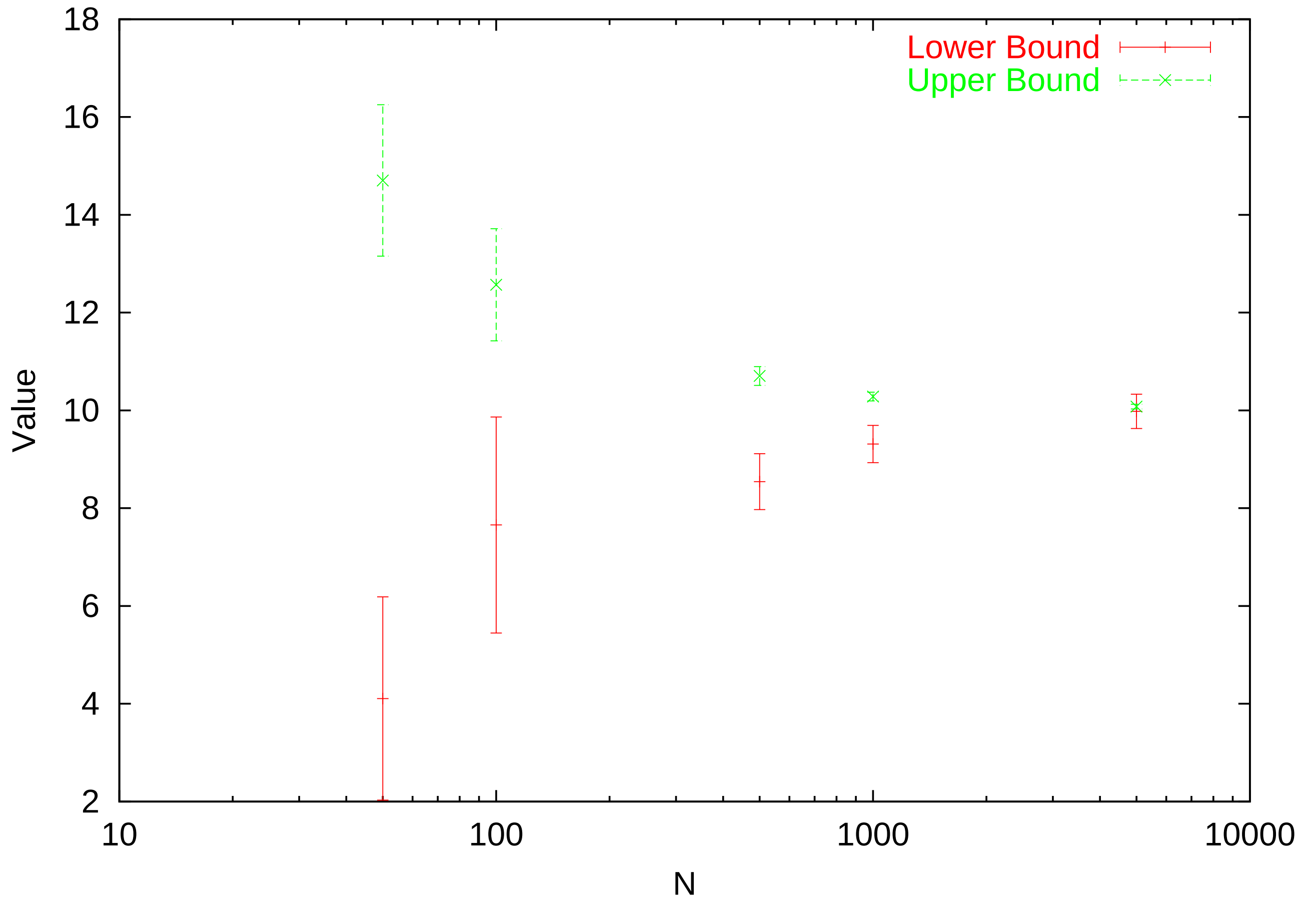
20term Convergence. Monte Carlo Sampling



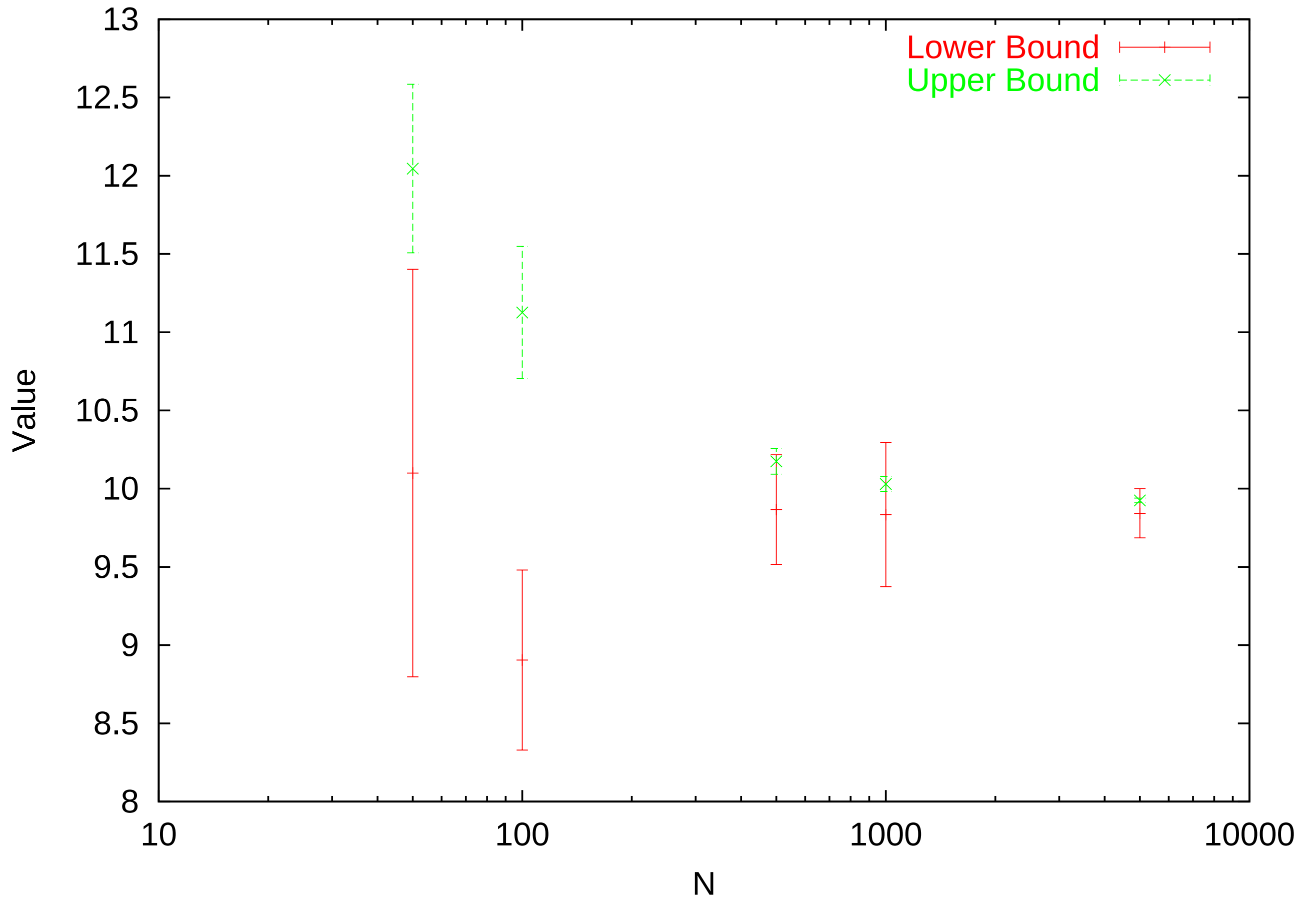
20term Convergence. Latin Hypercube Sampling



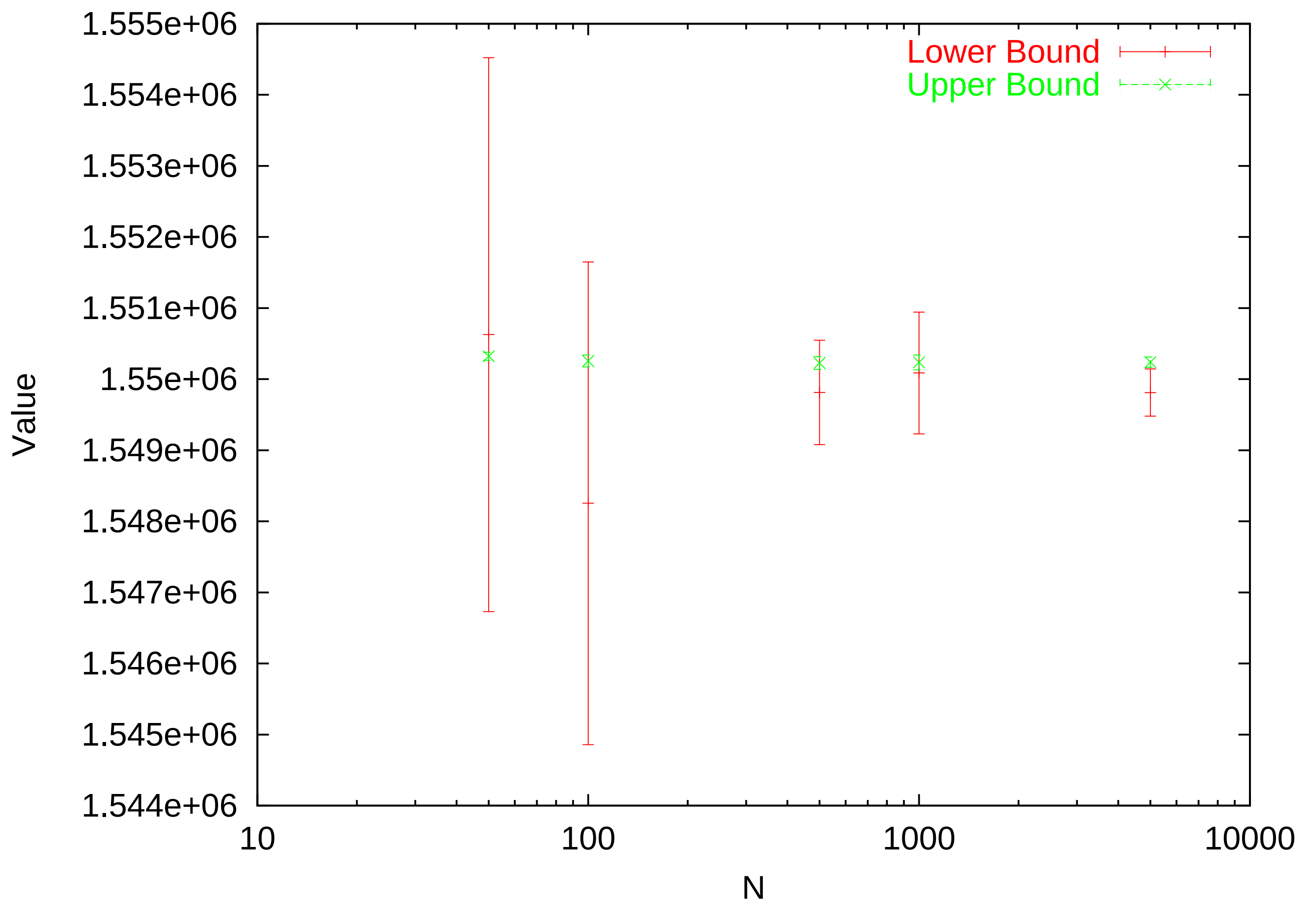
ssn Convergence. Monte Carlo Sampling



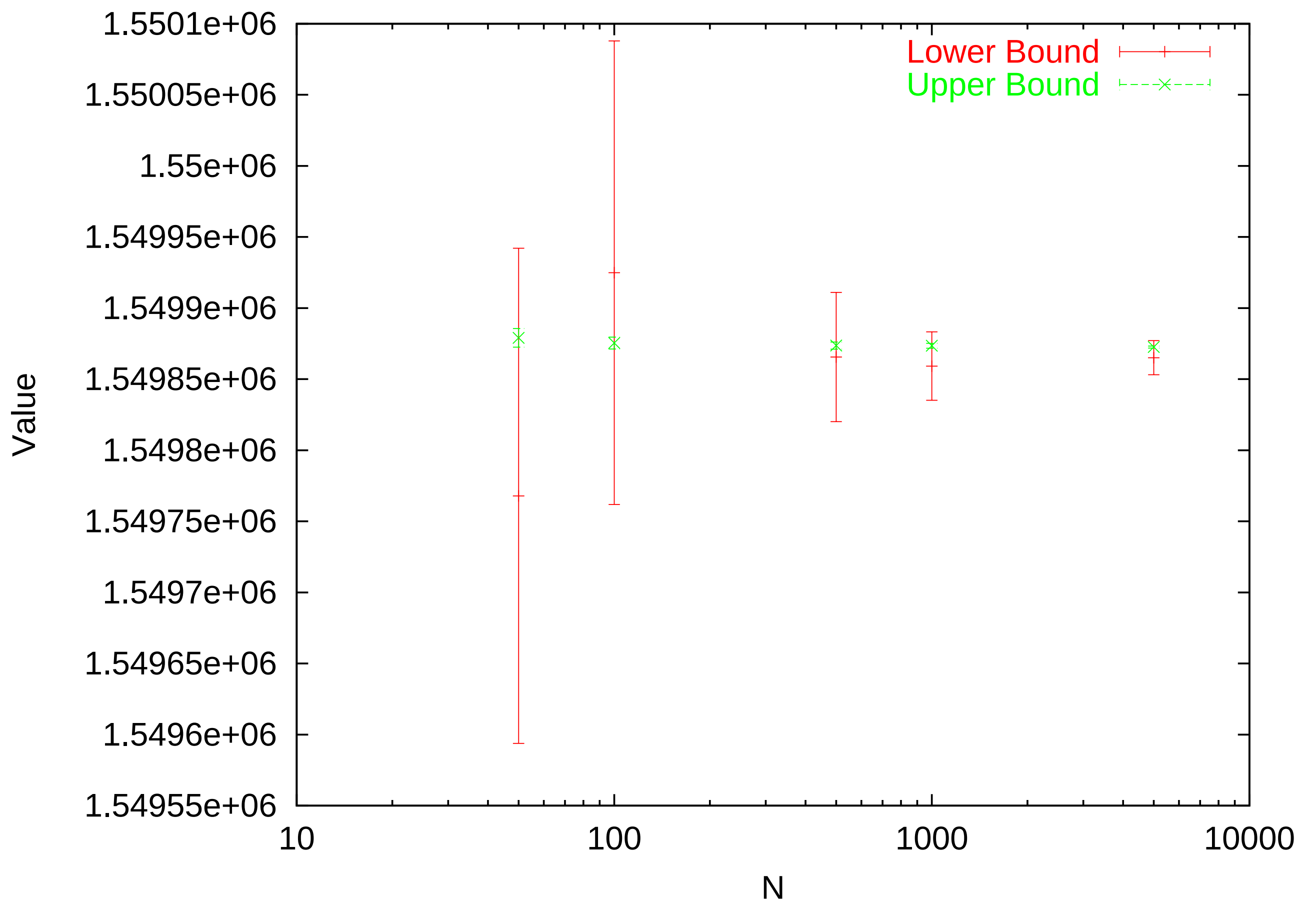
ssn Convergence. Latin Hypercube Sampling



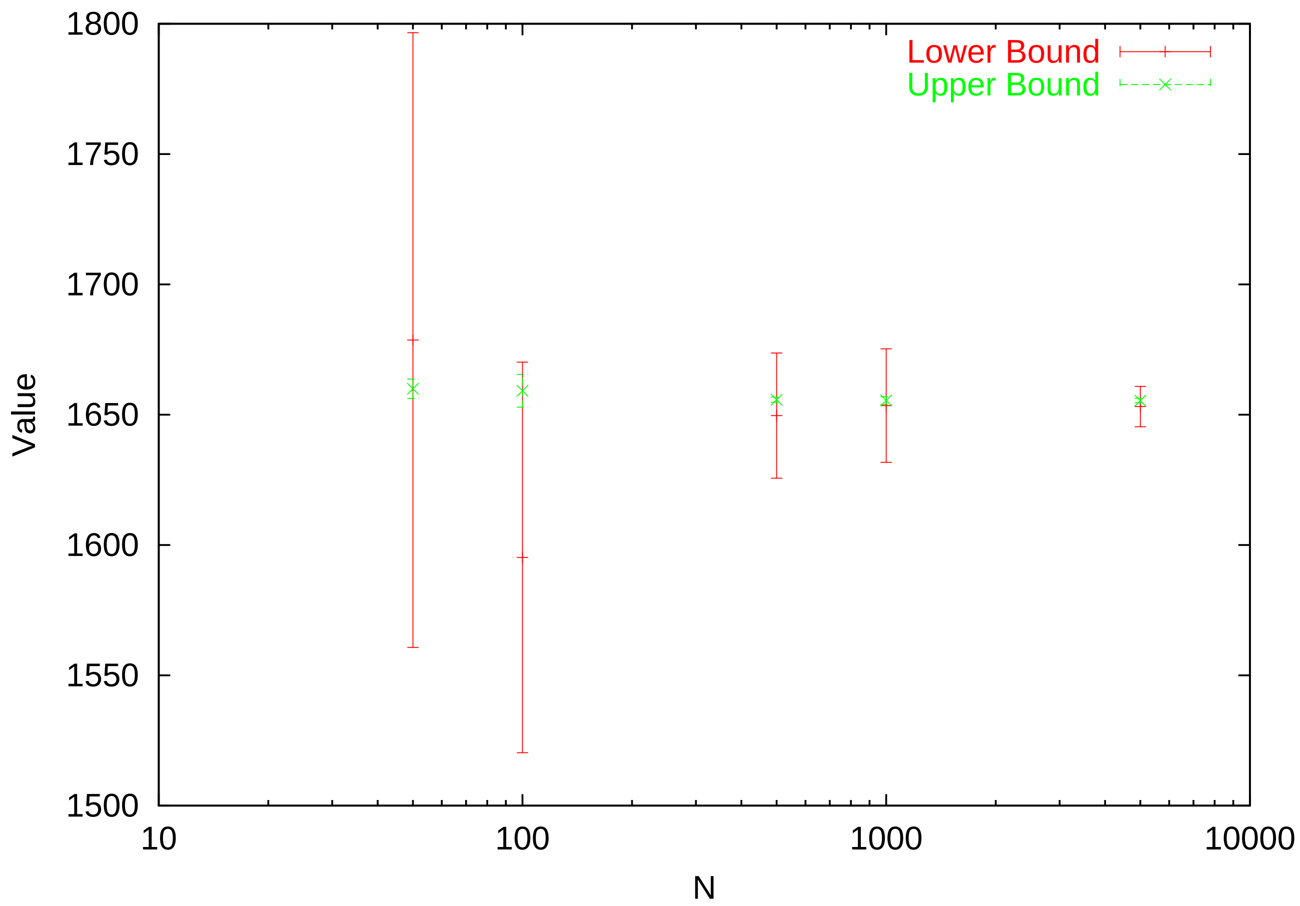
storm Convergence. Monte Carlo Sampling



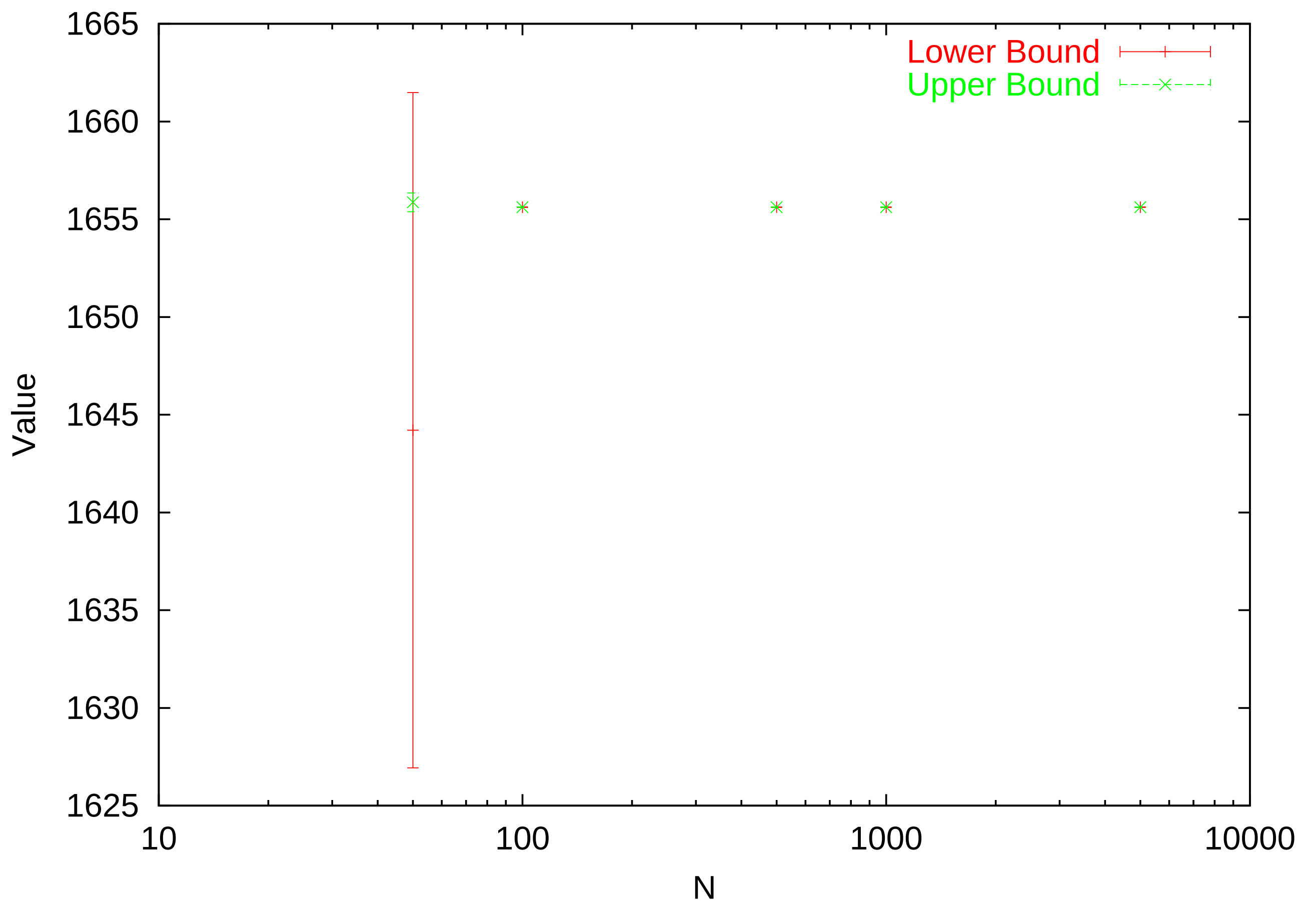
storm Convergence. Latin Hypercube Sampling



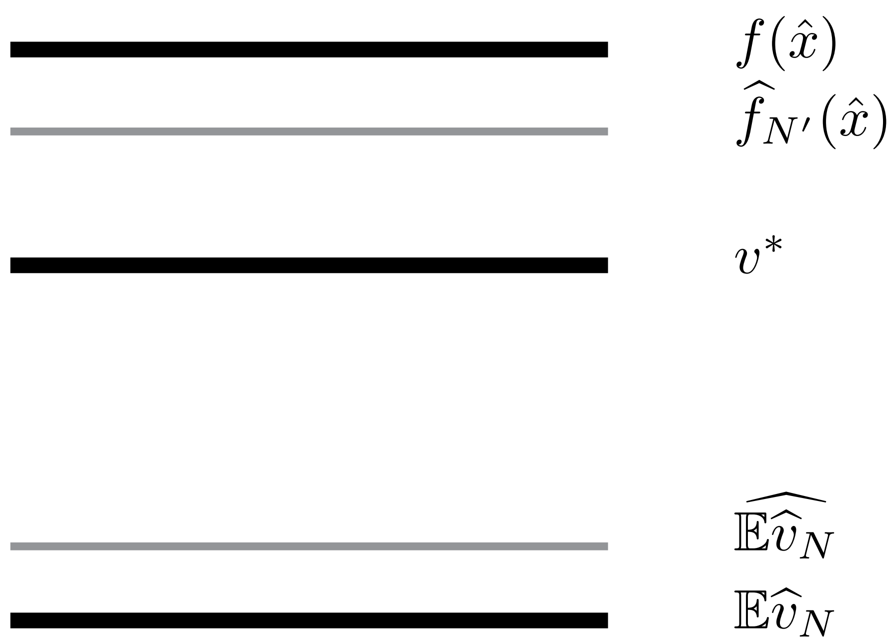
gbd Convergence. Monte Carlo Sampling



gbd Convergence. Latin Hypercube Sampling



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

Next Time

- Go over homeworks
- Convergence of Optimal Solution Values
- Stochastic Decomposition