18-660: Numerical Methods for Engineering Design and Optimization

Xin Li
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213

Electrical & Computer ENGINEERING

Overview

- Monte Carlo Analysis
  - Latin hypercube sampling
  - Importance sampling

Latin Hypercube Sampling (LHS)

- A great number of samples are typically required in traditional Monte Carlo to achieve good accuracy

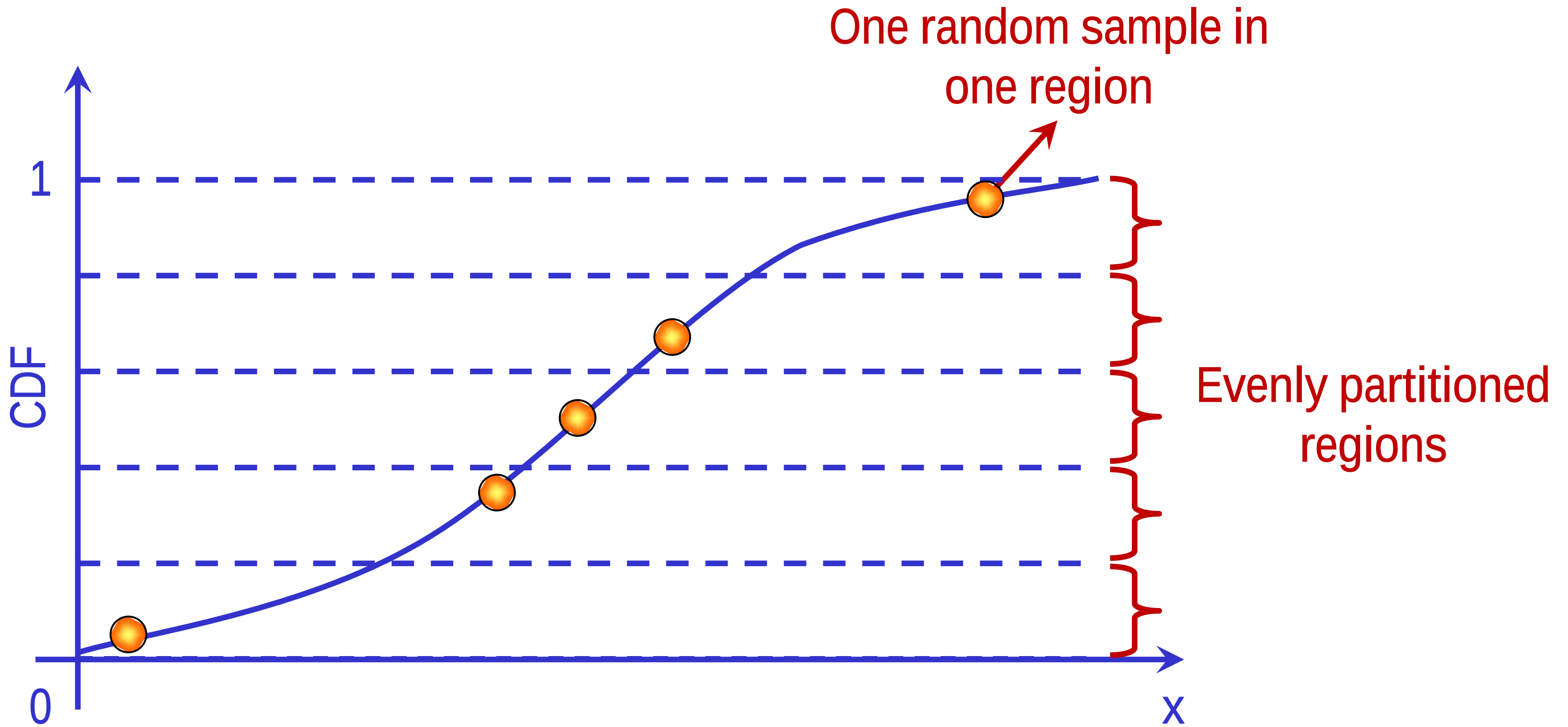
- Various techniques exist to improve Monte Carlo accuracy

- Controlling sampling points is the key
  - Latin hypercube sampling is a widely-used method to generate controlled random samples
  - The basic idea is to make sampling point distribution close to probability density function (PDF)

M. Mckay, R. Beckman and W. Conover, “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,” Technometrics, vol. 21, no. 2, pp. 239-245, May. 1979

Latin Hypercube Sampling (LHS)

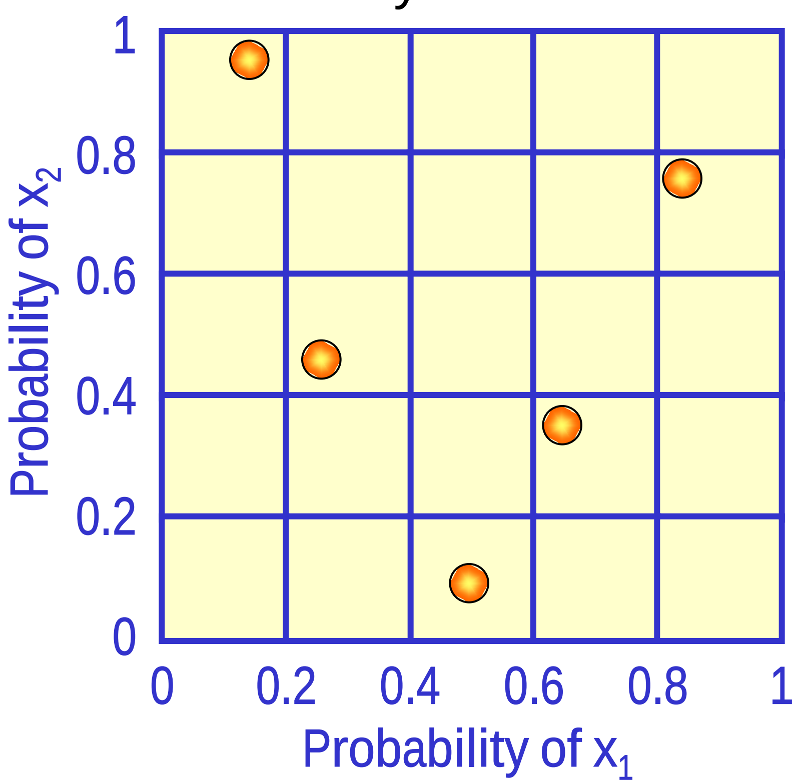
- One dimensional Latin hypercube sampling
  - Evenly partition CDF into N regions
  - Randomly pick up one sampling point in each region



Avoid the probability that all sampling points come from the same local region

Latin Hypercube Sampling (LHS)

- Two dimensional Latin hypercube sampling
  - x_1 and x_2 must be independent
  - Generate one-dimensional LHS samples for x_1
  - Generate one-dimensional LHS samples for x_2
  - Randomly combine the LHS samples to two-dimensional pairs



- One sample in each row and each column
- Sampling is random in each grid
- Higher-dimensional LHS samples can be similarly generated

Latin Hypercube Sampling (LHS)

- Matlab code for LHS sampling of independent standard Normal distributions

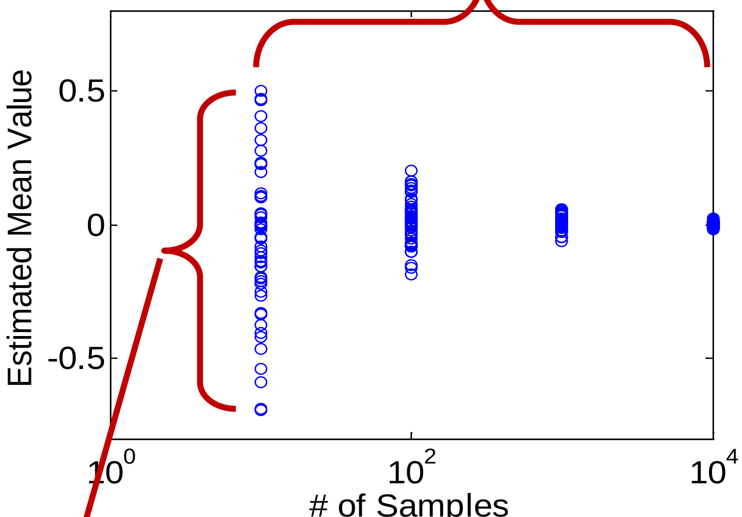
```
data = rand(NSample,NVar);
for i = 1:NVar
    index = randperm(NSample);
    prob = (index'-data(:,i))/NSample;
    data(:,i) = sqrt(2)*erfinv(2*prob-1);
end;
```

  - NVar: # of random variables
  - NSample: # of samples
  - data: LHS sampling points

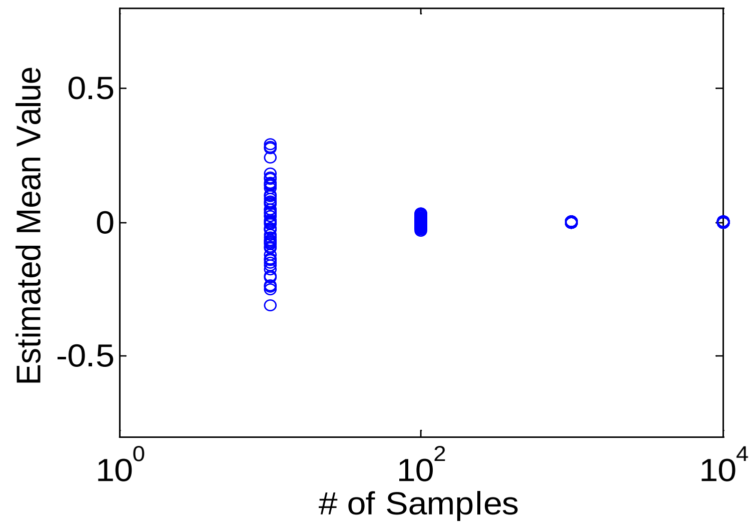
Latin Hypercube Sampling (LHS)

- Compare Monte Carlo accuracy for a simple example
  - x ~ N(0,1) (standard Normal distribution)
  - Repeatedly estimate the mean value by Monte Carlo analysis

Accuracy is improved by increasing sampling #



Random sampling



Latin hypercube sampling

Monte Carlo is not deterministic

Importance Sampling

- Even with Latin hypercube sampling, Monte Carlo analysis requires a HUGE number of sampling points

- Example: rare event estimation

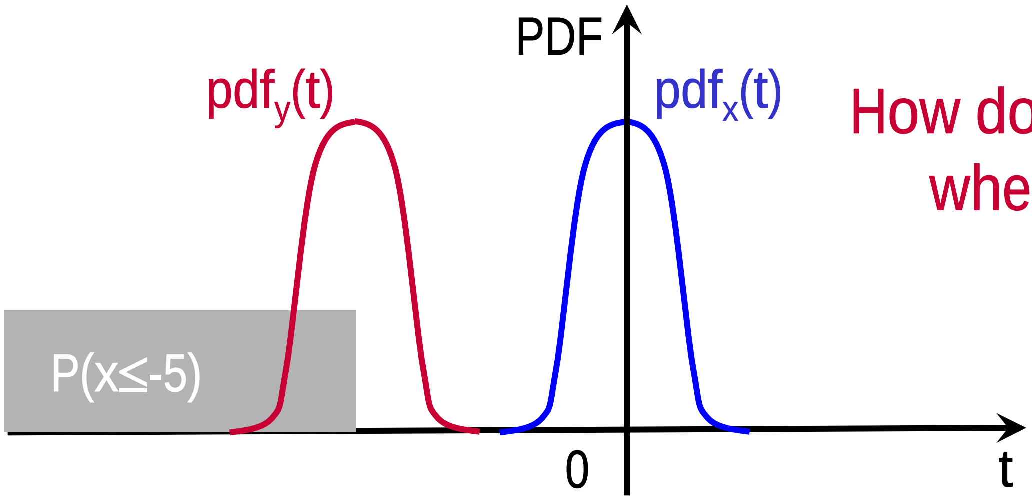
$$x \sim N(0,1) \qquad \text{Standard Normal distribution}$$

$$\text{Estimate} \qquad P(x \leq -5) = ???$$

  - The theoretical answer for $P(x \leq -5)$ is equal to 2.87×10^{-7}
  - ~100M sampling points are required if we attempt to estimate this probability by random sampling or LHS

Importance Sampling

- Key idea:
  - Do not generate random samples from $pdf_x(t)$
  - Instead, find a good distorted $pdf_y(t)$ to improve Monte Carlo sampling accuracy
- Example: if $x \sim N(0,1)$, what is $P(x \le -5)$?
  - Intuitively, if we draw sampling points based on $pdf_y(t)$, more samples will fall into the grey area



How do we calculate $P(x \le -5)$ when sampling $pdf_y(t)$?

Importance Sampling

- Assume that we want to estimate the following expected value

$$E\left[f(x)\right] = \int_{-\infty}^{+\infty} f(t) \cdot pdf_x(t) \cdot dt$$

- Example: if we want to estimate P(x ≤ -5), then

$$f(x) = \begin{cases} 1 & (x \le -5) \\ 0 & (x > -5) \end{cases}$$

$$E\left[f(x)\right] = \int_{-\infty}^{-5} 1 \cdot pdf_x(t) \cdot dt + \int_{-5}^{+\infty} 0 \cdot pdf_x(t) \cdot dt = P(x \le -5)$$

Importance Sampling

- Estimate E[f(x)] where x ~ $pdf_x(t)$ by importance sampling

$$E[f(x)] = \int_{-\infty}^{+\infty} f(t) \cdot pdf_x(t) \cdot dt = \int_{-\infty}^{+\infty} f(t) \cdot pdf_x(t) \cdot \frac{pdf_y(t)}{pdf_y(t)} \cdot dt$$

$$\Downarrow$$

$$\boxed{E[f(x)] = \int_{-\infty}^{+\infty} \left[f(t) \cdot \frac{pdf_x(t)}{pdf_y(t)} \right] \cdot pdf_y(t) \cdot dt = \int_{-\infty}^{+\infty} g(t) \cdot pdf_y(t) \cdot dt = E[g(x)]}$$

Importance Sampling

- Estimate $E[f(x)]$ where $x \sim pdf_x(t)$ by traditional sampling
  - Step 1: draw M random samples $\{t_1,t_2,...,t_M\}$ based on $pdf_x(t)$
  - Step 2: calculate $f_m = f(t_m)$ at each sampling point $m = 1,2,...,M$
  - Step 3: calculate $E[f] \approx (f_1+f_2+...+f_M)/M$

- Estimate $E[f(x)]$ where $x \sim pdf_x(t)$ by importance sampling
  - Step 1: draw M random samples $\{t_1,t_2,...,t_M\}$ based on $pdf_y(t)$
  - Step 2: calculate $g_m = f(t_m) \cdot pdf_x(t_m)/pdf_y(t_m)$ at each sampling point $m = 1,2,...,M$
  - Step 3: calculate $E[f] \approx (g_1+g_2+...+g_M)/M$

How do we decide the optimal $pdf_y(t)$ to achieve minimal Monte Carlo analysis error?

Importance Sampling

- Determine optimal $pdf_y(t)$ for importance sampling

$$E[f] \approx \underbrace{\mu_f}_{\text{Estimator}} = \frac{1}{M} \cdot \sum_{m=1}^{M} f(t_m) \cdot \frac{pdf_x(t_m)}{pdf_y(t_m)}$$

- The accuracy of an estimator can be quantitatively measured by its variance

$$Error \sim VAR[\mu_f]$$

  - To improve Monte Carlo analysis accuracy, we should minimize $VAR[\mu_f]$

Importance Sampling

- Determine optimal $pdf_y(t)$ for importance sampling

$$E[f] \approx \mu_f = \frac{1}{M} \cdot \sum_{m=1}^{M} f(t_m) \cdot \frac{pdf_x(t_m)}{pdf_y(t_m)}$$

- We achieve the minimal VAR[μ_f] = 0 if

$$f(t) \cdot \frac{pdf_x(t)}{pdf_y(t)} = k \quad \text{(Constant)}$$

$f(t_m) \cdot pdf_x(t_m)/pdf_y(t_m)$ is equal to the same constant for all sampling points

$$\boxed{pdf_y(t) = \frac{f(t) \cdot pdf_x(t)}{k} \quad \text{(Optimal PDF)}}$$

Importance Sampling

$$pdf_y(t) = \frac{f(t) \cdot pdf_x(t)}{k}$$

- How do we decide the value k?

- K cannot be arbitrarily selected
  - $pdf_y(t)$ must be a valid PDF that satisfies the following condition

$$\int_{-\infty}^{+\infty} pdf_y(t) \cdot dt = \int_{-\infty}^{+\infty} \frac{f(t) \cdot pdf_x(t)}{k} \cdot dt = 1$$

$$\Downarrow$$

$$\boxed{k = \int_{-\infty}^{+\infty} f(t) \cdot pdf_x(t) \cdot dt = E[f]}$$

Finding the optimal $pdf_y(t)$ requires to know E[f], i.e., the answer of our Monte Carlo analysis!!!

Importance Sampling

- In practice, such an optimal $pdf_y(t)$ cannot be easily applied

- Instead, we typically look for a sub-optimal solution that satisfies the following constraints
  - Easy to construct – we do not have to know k = E[f]
  - Easy to sample – not all random distributions can be easily sampled by a random number generator
  - Minimal estimator variance – the sub-optimal $pdf_y(t)$ is close to the optimal case as much as possible

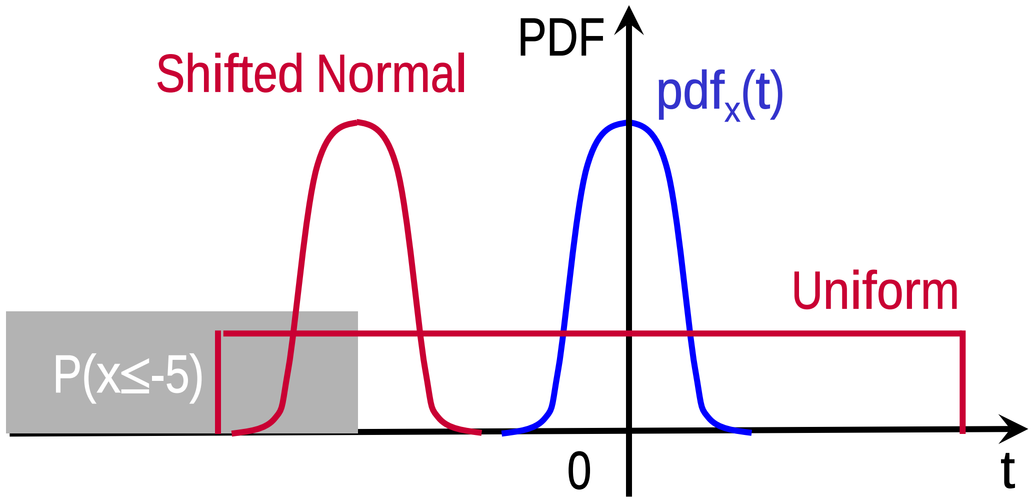
Importance Sampling

- Finding the right $pdf_y(t)$ is nontrivial for practical problems
  - No magic equation exists in general
  - Engineering approach is based on heuristics
  - Sometimes require a lot of human experience and numerical optimization

- The criterion to choose $pdf_y(t)$ is also application-dependent

Importance Sampling

- Example: if x ~ N(0,1), what is $P(x \leq -5)$?



Several possible choices for $pdf_y(t)$

Summary

- Monte Carlo analysis
  - Latin hypercube sampling
  - Importance sampling