Carnegie Mellon

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**
  - Random variable
  - Probability distribution
  - Random sampling

Random Variables

- A **random variable** is a real-valued function of the outcome of the experiment

x (random variable) =
- +1.0 (Experiment 1)
- −0.3 (Experiment 2)
- ...
- −2.4 (Experiment N)

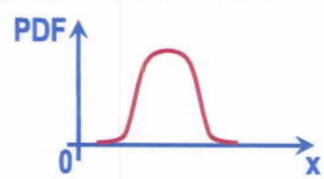
We get different results from different experiments (i.e., the output is random)

Probability Distribution

- A continuous random variable x is defined by its probability distribution function

- Probability density function (PDF)
  - $pdf_x(t_x)$ denotes the probability per unit length near $x = t_x$

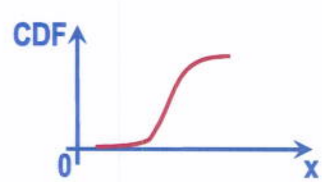
$\int_a^b pdf_x(\zeta)\,d\zeta = P[a \le x \le b]$



- Cumulative distribution function (CDF)
  - $cdf_x(t_x)$ equals the probability of $x \le t_x$

$cdf(t_x) = p(x \le t_x)$

$= \int_{-\infty}^{t_x} pdf_x(\zeta)\,d\zeta$



Expectation

- Given a random variable x and a function f(x), the expectation of f(x) is the weighted average of the possible values of f(x)

$$E[f(x)] = \int_{-\infty}^{+\infty} f(y) \cdot pdf_x(y)\,dy$$

- A useful equation for expected value calculation

$$E[f(x) + g(x)] = E[f(x)] + E[g(x)]$$

Mean, Variance and Standard Deviation

- Mean

$$E[x]$$

- Variance

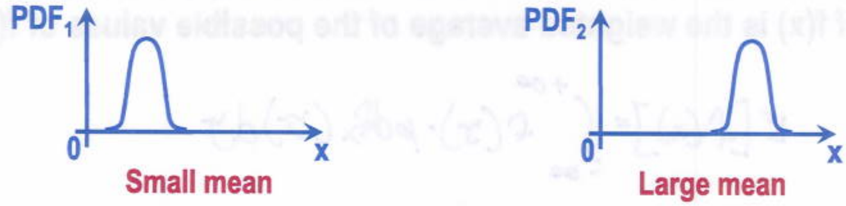
$$VAR[x] = E[(x - E[x])^2]$$

- Standard deviation

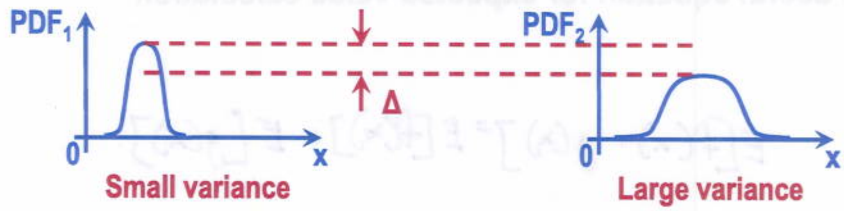
$$STD[x] = \sqrt{VAR(x)}$$

Mean, Variance and Standard Deviation

- Mean measures the "average position" of x



- Variance measures the "spread" of the distribution



$$\int_{-\infty}^{+\infty} pdf_x(y)\,dy = 1$$

Moments and Central Moments

- k-th order moment

$$E[x^k]$$

  - Mean is the first order moment $(k=1)$

- k-th order central moments

$$E[(x-E[x])^k]$$

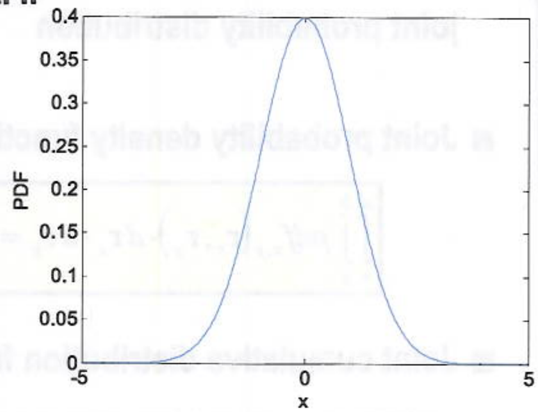
  - Variance is the second order central moment $(k=2)$

Normal Distribution

- A random variable x is Normal if

$$pdf_x(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

  - μ: mean
  - σ: standard deviation
  - Denoted as N(μ, σ²)



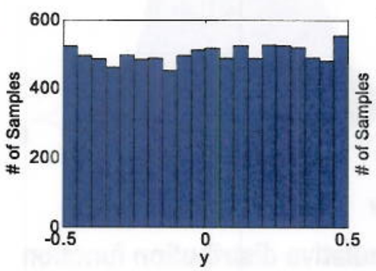
Standard Normal distribution

- If μ = 0 and σ = 1, it is called standard Normal distribution

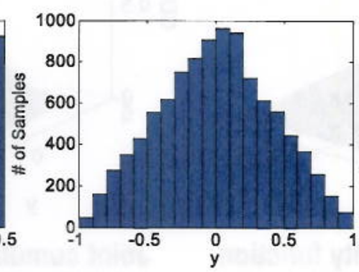
Why is Normal distribution important to us?

Normal Distribution

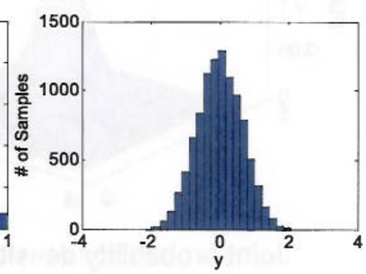
- Many physical variations are Normal
- Central limit theorem: the variation caused by a large number of independent random factors is "almost" Normal



$y = x_1$



$y = x_1 + x_2$



$y = x_1 + x_2 + x_3 + x_4 + x_5$

Assume that all x_i's are independent and have the same uniform distribution

Multiple Random Variables

- Two continuous random variables x and y are defined by their joint probability distribution

- Joint probability density function

$$\boxed{\int_c^d \int_a^b pdf_{x,y}(\tau_x,\tau_y)\cdot d\tau_x \cdot d\tau_y = P(a \le x \le b, c \le y \le d)}$$

$x \in [a, b]$
$y \in [c, d]$

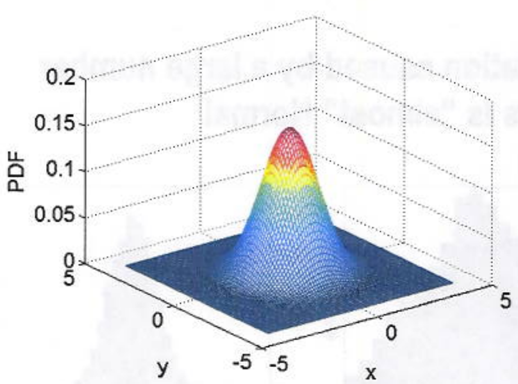
- Joint cumulative distribution function

$$\boxed{cdf_{x,y}(t_x,t_y) = P(x \le t_x, y \le t_y) = \int_{-\infty}^{t_y} \int_{-\infty}^{t_x} pdf_{x,y}(\tau_x,\tau_y)\cdot d\tau_x \cdot d\tau_y}$$

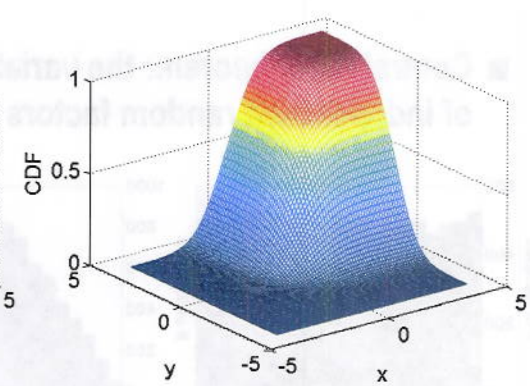
Applicable to more than two random variables

Joint Probability Distribution

- Example: bivariate Normal distribution



Joint probability density function



Joint cumulative distribution function

Marginal Distribution Function

- **Marginal probability density function**

$$pdf_x(t_x) = \int_{-\infty}^{+\infty} pdf_{x,y}(t_x, \tau_y) \cdot d\tau_y$$

$$pdf_y(t_y) = \int_{-\infty}^{+\infty} pdf_{x,y}(\tau_x, t_y) \cdot d\tau_x$$

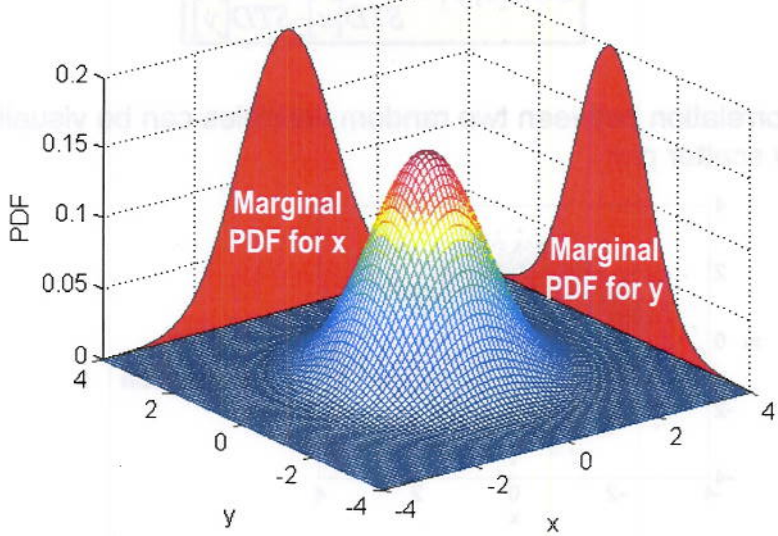
- **Marginal cumulative distribution function**

$$cdf_x(t_x) = P(x \le t_x, y \le +\infty) = \lim_{t_y \to +\infty} cdf_{x,y}(t_x, t_y)$$

$$cdf_y(t_y) = P(x \le +\infty, y \le t_y) = \lim_{t_x \to +\infty} cdf_{x,y}(t_x, t_y)$$

Marginal Distribution Function

- **Example: bivariate Normal distribution**



Covariance and Correlation

- **Covariance**

$$Cov[x,y] = E[(x-E[x])\cdot(y-E[y])]$$

  - If COV[x,y] = 0, then x and y are uncorrelated

- **Covariance matrix**

$$\Sigma = \begin{bmatrix} Cov[x,x] & Cov[x,y] \\ Cov[x,y] & Cov[y,y] \end{bmatrix}$$

$$Cov[y,y] \downarrow Var[y]$$

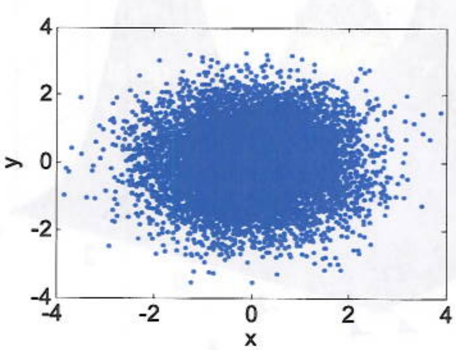
  - Σ is always symmetric
  - Diagonal components are corresponding to variance values
  - Σ is diagonal if x and y are uncorrelated

Covariance and Correlation

- **Correlation (normalized covariance)**

$$\boxed{COR[x,y] = \frac{COV[x,y]}{STD[x]\cdot STD[y]}}$$

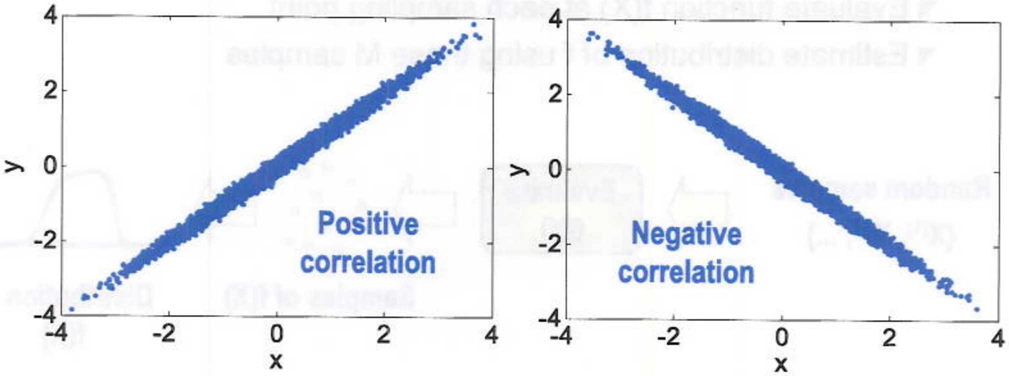
  - Correlation between two random variables can be visualized by scatter plot



Zero correlation

Covariance and Correlation

- **Example: correlated random variables**



Monte Carlo Analysis

- **Problem definition**
  - Find probability distribution and/or moments of

$$f(X)$$

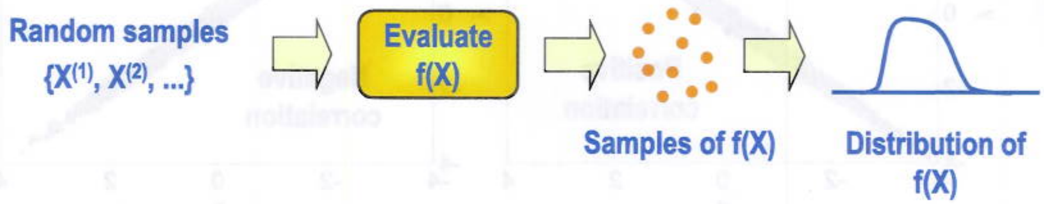
Function of interest

Random variable with known distribution

- **In general, the distribution and/or moments of f cannot be calculated analytically, because**
  - f(X) is nonlinear
  - f(X) may not have closed-form expression (we can only numerically calculate f for a given X value)

Monte Carlo Analysis

- **Monte Carlo analysis for f(X)**
  - Randomly select M samples for X
  - Evaluate function f(X) at each sampling point
  - Estimate distribution of f using these M samples



Monte Carlo Analysis Example

- **Example: estimate the probability distribution of**

$$y = \exp(x)$$

  - x ~ N(0,1) (standard Normal distribution)

Monte Carlo Analysis Example

- Step 1: draw random samples for x

Samples	1	2	3	4	5	6	...
x	-0.4326	-1.6656	0.1253	0.2877	-1.1465	1.1909	...

M random samples for x

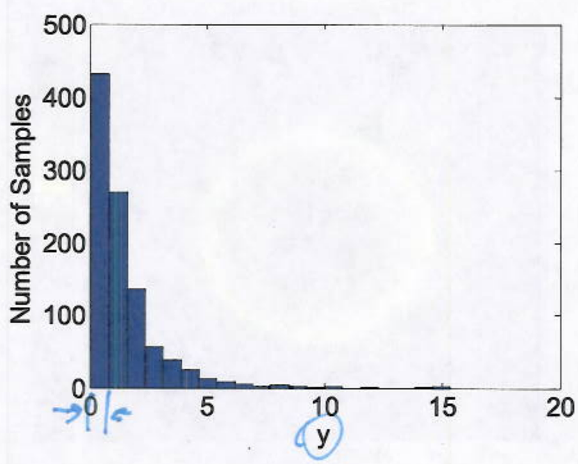
- Step 2: calculate y at each sampling point

Samples	1	2	3	4	5	6	...
y	0.6488	0.1891	1.1335	1.3333	0.3178	3.2901	...

M random samples for y

Monte Carlo Analysis Result

- Monte Carlo result is typically represented by a histogram
  - A big table of data is not intuitive

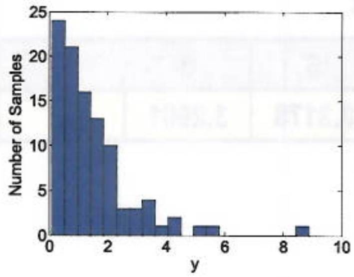


Histogram of y based on 1000 random samples

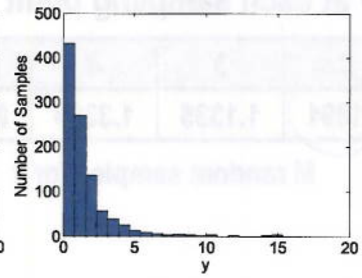
QUESTION: how accurate is Monte Carlo analysis?

Monte Carlo Analysis Accuracy

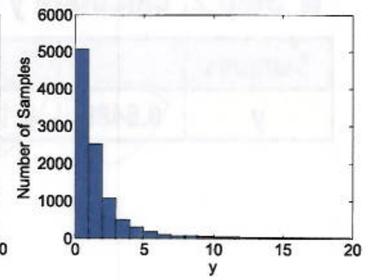
- **Monte Carlo analysis is not deterministic**
  - We cannot get identical results when running MC twice
  - The analysis error is not deterministic
- **Monte Carlo accuracy depends on the number of samples**
  - Examples: histogram of y



100 samples



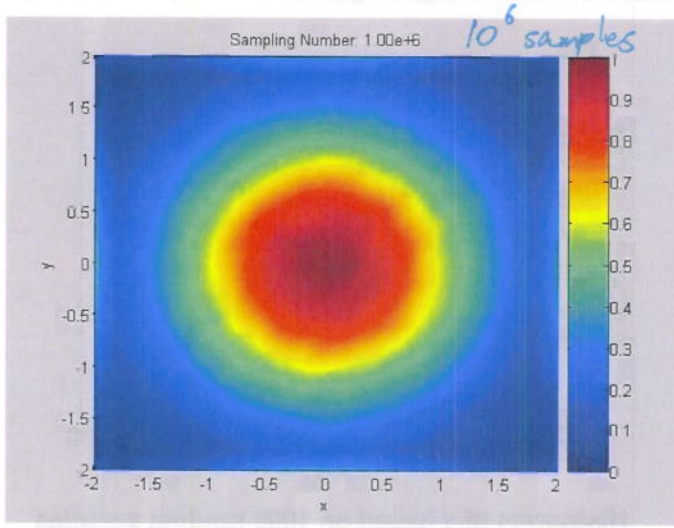
1000 samples



10000 samples

Monte Carlo Analysis Accuracy

- **Example: bivariate Normal distribution**
  - x and y are independent and jointly standard Normal



10^6 samples

Monte Carlo Analysis Accuracy

- **Statistical methods exist to analyze Monte Carlo accuracy**

- **Example: Monte Carlo accuracy analysis**

$$x \sim N(0,1) \quad \textbf{Standard Normal distribution}$$

  - Estimate the mean value μ_x by Monte Carlo analysis
  - Our question: how accurate is the estimated μ_x (dependent on the number of Monte Carlo samples)?

Monte Carlo Analysis Accuracy

- **Monte Carlo analysis for the mean value μ_x**
  - Randomly draw M sampling points $\{x^{(1)}, x^{(2)}, \ldots, x^{(M)}\}$
  - Estimate μ_x by the following equation

$$\mu_x = \frac{x^{(1)} + x^{(2)} + \ldots\ldots + x^{(m)}}{M}$$

Estimator

- **Assumptions in our accuracy analysis**
  - Each $x^{(i)}$ is random and satisfies standard Normal distribution – it is randomly created for $x \sim N(0,1)$
  - All $x^{(i)}$'s are mutually independent – samples from a good random number generator should be independent

- **μ_x is a function of $\{x^{(1)}, x^{(2)}, \ldots, x^{(M)}\}$, which is a random variable**

Monte Carlo Analysis Accuracy

- **Mean of μ_x**

$$E[\mu_x] = E\left[\frac{x^{(1)}+x^{(2)}+\dots+x^{(M)}}{M}\right]$$
$$= \frac{1}{M}\left(E[x^{(1)}]+E[x^{(2)}]+\dots+E[x^{(M)}]\right)$$
$$= \frac{1}{M}[0+0+0+\dots+0] = 0$$

- **Variance of μ_x**

$$E[\mu_x^2] = E\left[\left(\frac{x^{(1)}+x^{(2)}+\dots+x^{(M)}}{M}\right)^2\right]$$

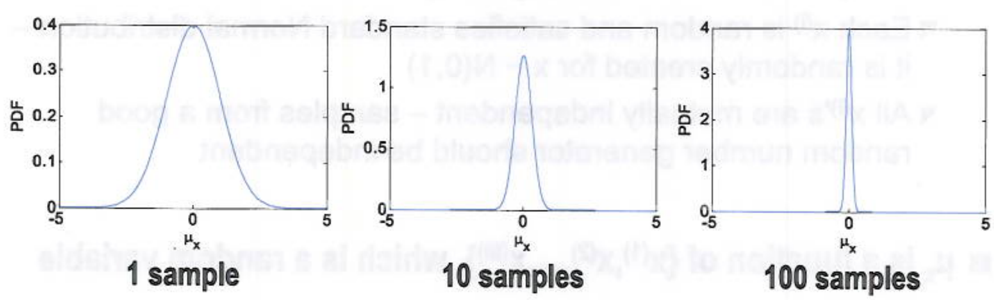
$$E[x^{(i)}x^{(j)}] = E[x^{(i)}]E[x^{(j)}] = 0 \quad (i \neq j)$$

$$= E\left\{\frac{x^{(1)2}+x^{(2)2}+\dots+x^{(M)2}}{M^2}\right\}$$

$$= \frac{1+1+\dots+1}{M^2} = \frac{M}{M^2} = \frac{1}{M}$$

Monte Carlo Analysis Accuracy

- **$E\{\mu_x\} = 0$**
  - u_x is an unbiased estimator
  - Otherwise, if the estimator mean is not equal to the actual mean, it is called a biased estimator
- **$E\{\mu_x^2\} = 1/M$**
  - Variance decreases as M increases
  - Distributions of μ_x for different M values



μ_x is a Normal distribution N(0, 1/M)

Monte Carlo Analysis Accuracy

- "Average" estimation accuracy is better when using larger M

- In this μ_x example
  - If we require that ± 3 sigma of μ_x is within [-0.1, 0.1]

$\sqrt{\frac{1}{M}}$

$$\frac{3}{\sqrt{M}} \le 0.1 \Rightarrow M \ge 900$$

  - If we require that ± 3 sigma of μ_x is within [-0.01, 0.01]

$$\frac{3}{\sqrt{M}} \le 0.01 \Rightarrow M \ge 90000$$

Monte Carlo Analysis Accuracy

- Accuracy is improved by **10x** if the number of samples is increased by **100x**

- 1K ~ 10K sampling points are typically required to achieve reasonable accuracy

- However, even if you use 10K sampling points, an accurate result is not guaranteed!
  - Monte Carlo analysis is random, and you can be unlucky (e.g., going beyond ± 3 sigma range)

Summary

- **Monte Carlo analysis**
  - Random variable
  - Probability distribution
  - Random sampling

Carnegie Mellon

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

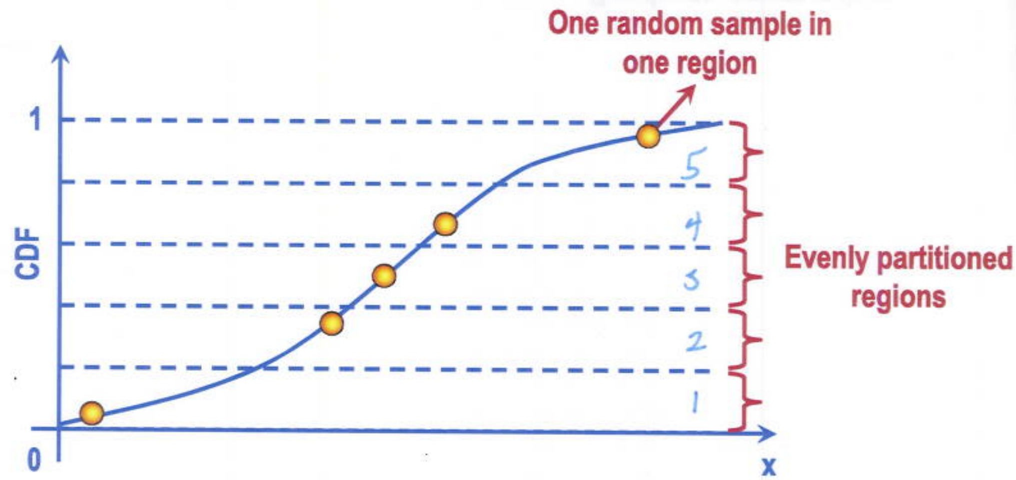
Latin Hypercube Sampling (LHS)

- **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)

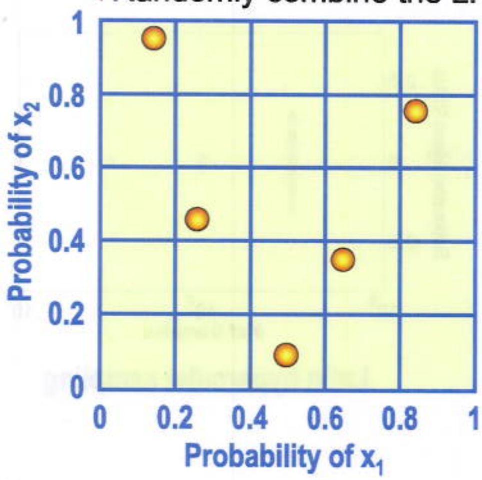
- **One dimensional Latin hypercube sampling**
  - Evenly partition CDF into N regions
  - Randomly pick up one sampling point in each region



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**

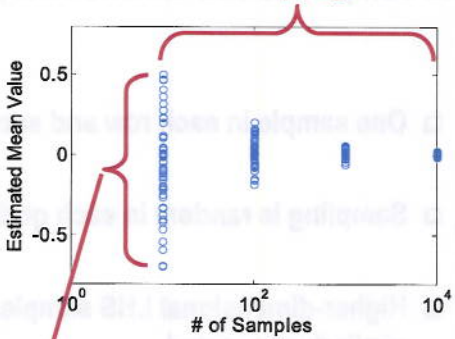
```
data = rand(NSample,NVar);
for i = 1:NVar
   index = randperm(NSample);
   prob = (index'-data(:,i))/NSample;
   data(:,i) = sqrt(2)*erfinv(2*prob-1);
end;
```

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

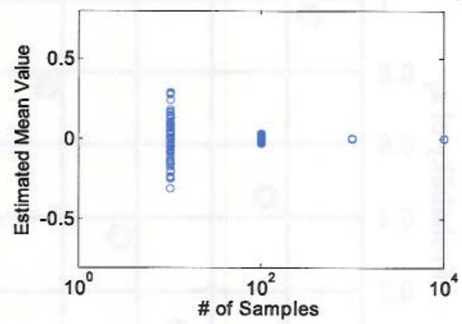
Latin Hypercube Sampling (LHS)

- **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**

$$x \sim N(0,1) \quad \textbf{Standard Normal distribution}$$

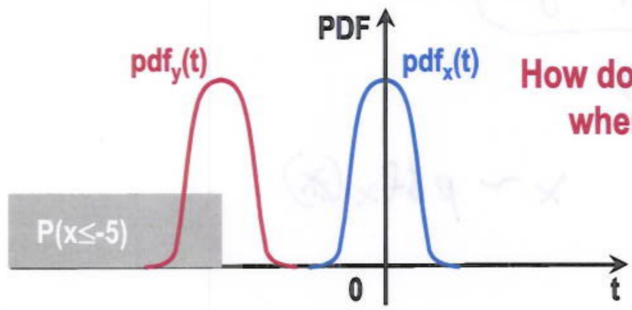
$$\textbf{Estimate} \quad P(x \le -5) = ???$$

  - The theoretical answer for $P(x \le -5)$ 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**

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

- **Example: if we want to estimate $P(x \le -5)$, then**

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

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

$$= \int_{-\infty}^{-5} \underbrace{f(t)}_{1} \cdot pdf_x(t) dt + \int_{-5}^{+\infty} \underbrace{f(t)}_{0} pdf_x(t) dt$$

$$= P(x \le -5)$$

Importance Sampling

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

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

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

$$= \int_{-\infty}^{+\infty} \underbrace{f(t) \cdot \frac{pdf_x(t)}{pdf_y(t)}}_{g(t)} \cdot pdf_y(t) dt$$

$$= E[g(x)] \qquad x \sim pdf_x(*)$$

Importance Sampling

- **Estimate E[f(x)] where x ~ $pdf_x(t)$ 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 ~ $pdf_x(t)$ 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**

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

μ_f: Estimator

- **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**

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

- **We achieve the minimal $VAR[\mu_f] = 0$ if**

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

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

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

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

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

must know $E[f]$!!!

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

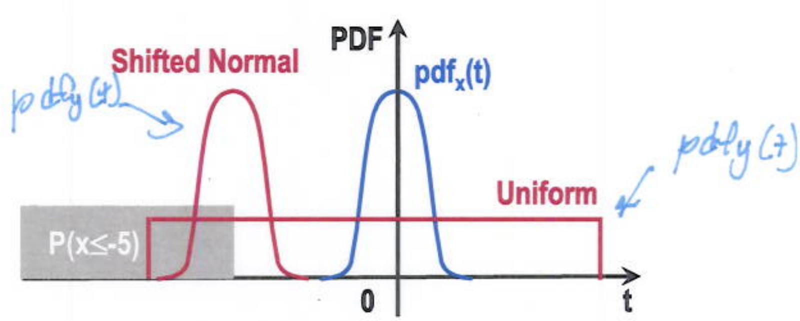
Importance Sampling

- **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 ≤ -5)?**



Several possible choices for $pdf_y(t)$

Summary

- **Monte Carlo analysis**
  - Latin hypercube sampling
  - Importance sampling