고려대학교

Basic Statistics and Monte-Carlo Method -1

고려대학교
화공생명공학과
강정원

Table of Contents

1. Basic Statistics

- Probability and Statistics
  - Needed to understand for general simulation techniques
  - Acquaintance with notation and symbols
- Probability and Statistics in Simulation Methods
  - Generation of random samples from a distribution
  - Design of simulation experiments
  - Statistical analysis of simulation data
  - Validation of simulation model

1.1 Discrete Random Variables and Their Properties

고려대학교

- The Probability theory says;
  - Experiment – An outcome cannot be predicated with certainty
  - Sample Space (*S*) – All Possible outcome of an experiment

Examples of Sample Space

Tossing a die
$S = \{ 1, 2, 3, 4, 5, 6\}$

Tossing two dice
$S = \{ (1,1), (1,2), \ldots, (6,6)\}$

- A Random Variable X
  - Function or Rule that assigns a real number x
  - Can be discrete or continuous

X	Random Variable
x	Values taken

Examples of Random Variables

Tossing a die
$x = 1$

Tossing two dice
$x = 7 \quad (1,6)$

1.1 Discrete Random Variables and Their Properties

■ Cumulative Distribution Function *F(x)*

- The Probability that a random variable X takes on a value longer than x

$$F(x) = P(X > x) \qquad -\infty < x < \infty$$

P(cond.)
Probability associated with
The given condition

- Properties

i) $0 \le F(x) \le 1$ for all x

ii) $F(x)$ is nondecreasing: if $x_1 < x_2 \rightarrow F(x_1) < F(x_2)$

iii) $F(-\infty) = 0, \quad F(\infty) = 1$

1.1 Discrete Random Variables and Their Properties

고려대학교

- Probability Mass Function *P(x)*
  - The Probability that a random variable X takes on the value x

$$P(x) = P(X = x) \qquad -\infty < x < \infty$$

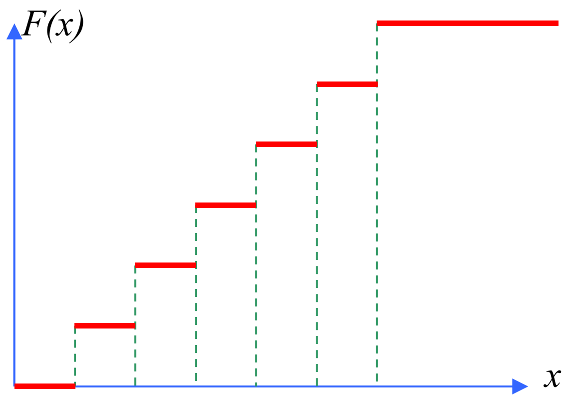
  - Properties

i) $P(x) \geq 0$ for all x

ii) $\sum_x^{\infty} P(x) = 1$

- Cumulative Distribution Function and Probability Mass Function for the outcome of tossing a die

x	*1*	*2*	*3*	*4*	*5*	*6*
P(x)	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*
F(x)	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*



1.1 Discrete Random Variables and Their Properties

- Expected Value : *E(X)*
  - the mean average value

$$E(X) = \sum_x xp(x) = \mu$$

  - Example) The outcome of tossing a die

$$E(X) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3.5$$

1.1 Discrete Random Variables and Their Properties

고려대학교

- **Variance**
  - Expected squared value of deviation of X from the mean value
  - The measure of how values are distributed from the mean value

$$V(x) = E((X-\mu)^2)$$

- **Standard deviation**
  - Square root of the variance

$$\sigma = \sqrt{V(x)}$$

Example

$$V(x) = \frac{1}{6}\left[(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\right] =$$

1.1 Discrete Random Variables and Their Properties

- **Properties**

$$E(c) = \sum cp(x) = c\sum p(x) = c \qquad c: \text{constant}$$

$$E(cg(X)) = \sum cg(x)p(x) = c\sum g(x)p(x) = cE(g(X))$$

c : constant $g(X)$: a function of X

$$E(g_1(X) + g_2(X) + \ldots) = \sum_x (g_1(x) + g_2(x) + \ldots)p(x) =$$

$$\sum_x (g_1(x)p(x) + g_2(x)p(x) + \ldots) = E(g_1(X)) + E(g_2(X)) + \ldots$$

$$V(X) = E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) =$$

$$E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2$$

1.1 Discrete Random Variables and Their Properties

$$V(X+c)=E((X+c)^2)-(E(X+c))^2$$
$$E(X+c)=\mu+c$$
$$E((X+c)^2)=E(X^2)+2\mu c+c^2$$
$$V(X+c)=E(X^2)+2\mu c+c^2-(\mu+c)^2=E(X^2)-\mu^2$$
$$V(X+c)=V(X)$$

A rigid shift in distribution does not change the breadth of the distribution

$$V(cX)=E((cX)^2)-(E(cX))^2$$
$$=c^2E(X^2)-c^2(E(X))^2=c^2V(X)$$

1.2 Continuous Random Variables and Their Properties

고려대학교

- Cumulative Distribution Function of a Continuous Random Variable X

$$F(x) = P(X \geq x) \qquad \text{for} -\infty < x < \infty$$

$$F(-\infty) = 0$$

$$F(\infty) = 1$$

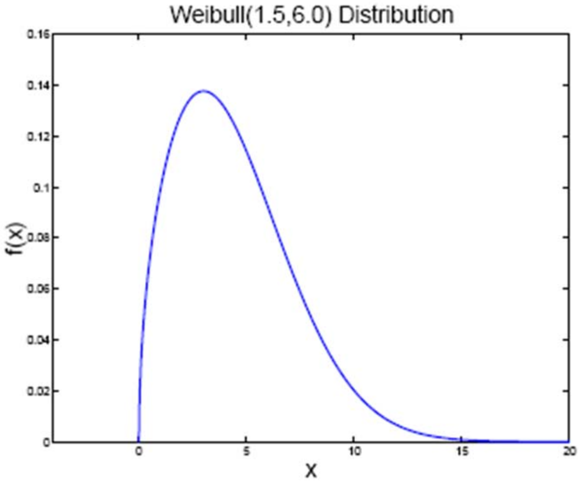
- Probability Density Function

$$f(x) = \frac{dF(x)}{dx}$$

$$dF = f(t)dt$$

$$\int_{-\infty}^{x} dF = \int_{-\infty}^{x} f(t)dt$$

$$F(x) - F(-\infty) = F(x) = \int_{-\infty}^{x} f(t)dt$$



1.2 Continuous Random Variables and Their Properties

Properties of probability distribution function *f(x)*

- F(x) is non-decreasing function → $f(x) = \dfrac{dF}{dx} \geq 0$

- $F(\infty) = 1 \quad \rightarrow \quad \int_{-\infty}^{\infty} f(x)dx = 1$

- Calculation of probability

$$P(a \leq x \leq b) = P(x \leq b) - P(x \leq a) = F(b) - F(a)$$
$$= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx$$

1.2 Continuous Random Variables and Their Properties

고려대학교

- The expected value

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$

- Properties

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$E(c) = c$$

$$E(cg(X)) = cE(g(X))$$

$$E(g_1(X) + g_2(X) + \ldots) = E(g_1(X)) + E(g_2(X)) + \ldots$$

$$V(X) = E(X^2) - \mu^2$$

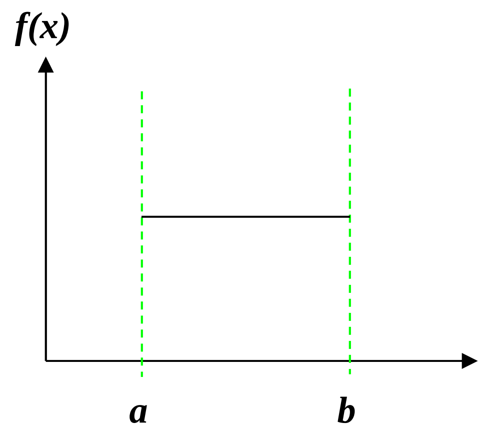
All the same relations as a discrete random variable X

1.2 Continuous Random Variables and Their Properties

고려대학교

Uniform Probability Distribution Functions

- Consider a random variable X , values are distributed uniformly in the interval *[a,b]*



- The probabilities are the same in *[a,b]*

$$P(x_1 \le X \le x_1 + \Delta x) = P(x_2 \le X \le x_2 + \Delta x)$$

$$\int_{x_1}^{x_1+\Delta x} f(x)dx = \int_{x_2}^{x_2+\Delta x} f(x)dx$$

- Requirement for normalization

$$\int_a^b f(x)dx = c\int_a^b dx = c(b-a) = 1$$

$$c = \frac{1}{b-a}$$

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & x \le a \\ 1/(b-a) & a < x < b \\ 0 & x \ge b \end{cases}$$

1.2 Continuous Random Variables and Their Properties

고려대학교

Uniform Probability Distribution Functions

- Cumulative Probability Distribution Function

$$F(x)=\frac{x}{b-a}+c \qquad F(x)=0 \text{ at } x=a \rightarrow c=-\frac{a}{b-a}$$

$$F(x)=\begin{cases} 0 & x\le a \\ (x-a)/(b-a) & a<x<b \\ 0 & x\ge b \end{cases}$$

- Expected Value

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{1}{b-a}\left(\frac{x^2}{2}\right)_a^b=\frac{a+b}{2}$$

- Variance

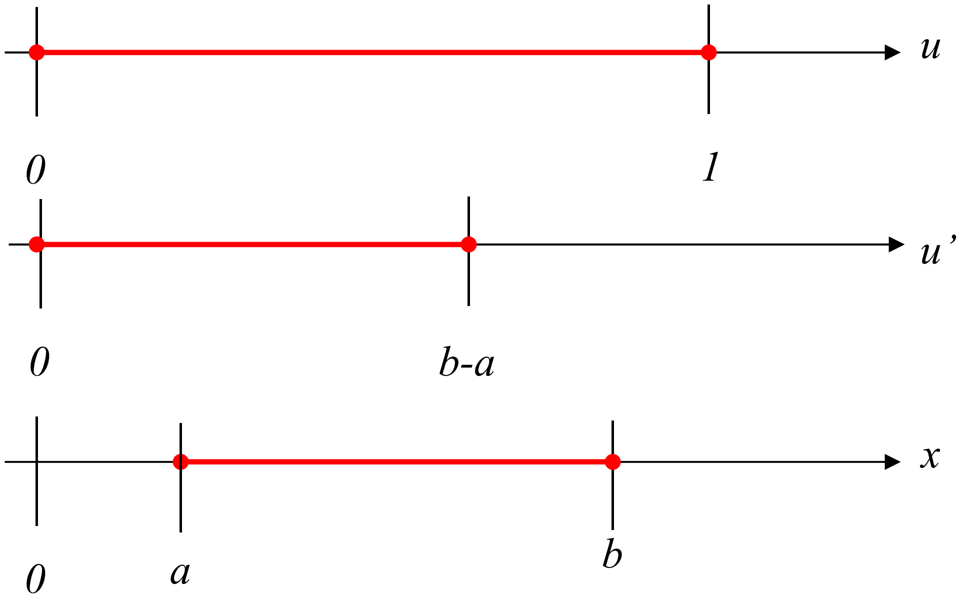
$$E(X^2)=\int_a^b \frac{x^2}{b-a}dx=\frac{a^2+ab+b^2}{3} \qquad V(X)=E(X^2)-(E(X))^2=\frac{(b-a)^2}{12}$$

1.2 Continuous Random Variables and Their Properties

고려대학교

Mapping to Different Domain

- Mapping
  - Pseudo random number generator (u) in $[0,1]$ → x in $[a,b]$



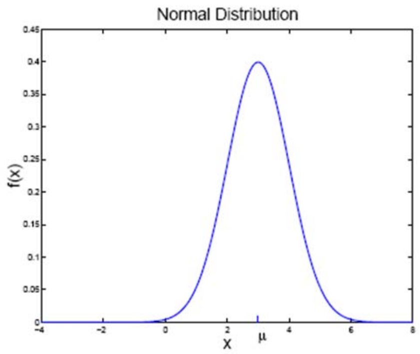
$$u' = (b-a)u$$

$$x = u' + a = (b-a)u + a$$

1.3 Normal Probability Distribution Function

- Normal probability distribution functions

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad -\infty < x < \infty$$



- Parameters : μ and σ
- Normalization

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{t^2}{2}\right]dt = 1$$

$$t = \frac{(x-\mu)}{\sigma}$$

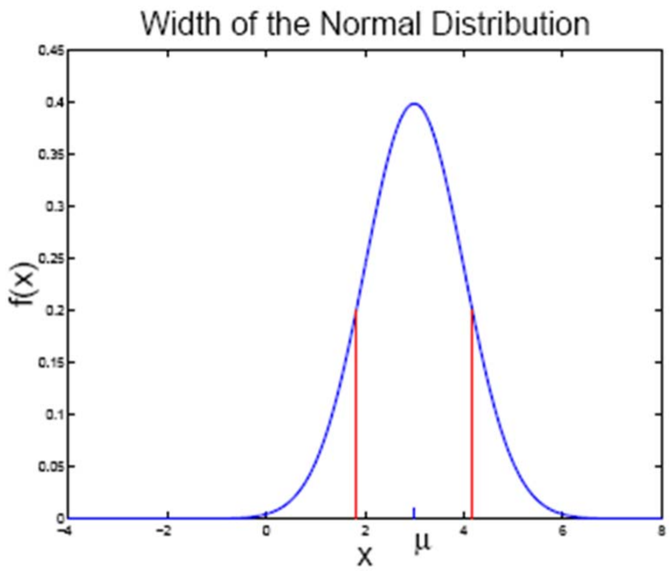
1.3 Normal Probability Distribution Function

- Normal probability distribution function
  - Bell-shaped curve with a single peak at $x = \mu$
  - The value at the peak : $1/\sqrt{2\pi}\sigma$
  - The value of x when the value becomes the half of the peak value

$$x = \mu + h$$

$$\exp\left[-\frac{h^2}{2\sigma^2}\right] = \frac{1}{2}$$

$$h = \pm\sqrt{2\ln 2}\sigma \approx \pm 1.177\sigma$$



1.3 Normal Probability Distribution Function

고려대학교

$$E(X)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}x\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}(t+\mu)\exp\left[-\frac{t^2}{2\sigma^2}\right]dt$$

$$=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}t\exp\left[-\frac{t^2}{2\sigma^2}\right]dt+\frac{\mu}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}\exp\left[-\frac{t^2}{2\sigma^2}\right]dt$$

$$=0+\mu=\mu$$

$$V(X)=E((X-\mu)^2)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}(x-\mu)^2\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx$$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^2\exp\left[-t^2\right]dt$$

⟵ $t=\frac{x-\mu}{\sqrt{2}\sigma}$

$$=\sigma^2$$

1.3 Normal Probability Distribution Function

고려대학교

■ The Probability for finding X having value between x_1 and x_2

$$P(x_1 \le X \le x_2) = \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] dx = \frac{1}{\sqrt{\pi}} \int_{t_1}^{t_2} \exp\left[-t^2 \right] dt$$

$$= \frac{1}{\sqrt{\pi}} \left\{ \int_0^{t_2} - \int_0^{t_1} \right\} \exp(-t^2) dt$$

$$= \frac{1}{2} \left[\Phi(t_2) - \Phi(t_1) \right]$$

$$\Phi(t) = \frac{1}{\sqrt{\pi}} \int_0^t \exp(-x^2) dx$$

← **Error Function**

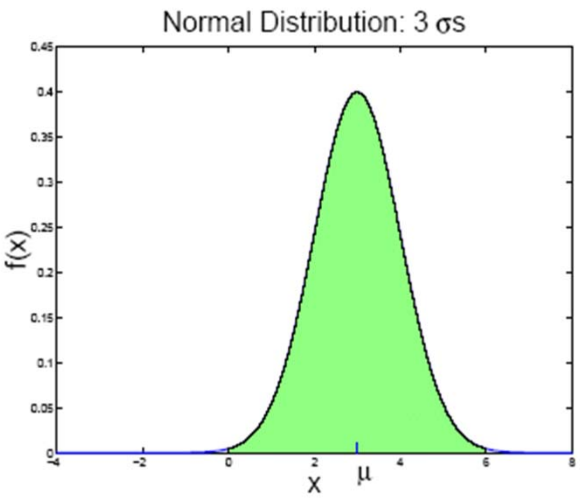
1.3 Normal Probability Distribution Function

고려대학교

The rule of 3σs

- 99.7 % of the trial values fall within the range of $(\mu-3\sigma)$ and $(\mu+3\sigma)$

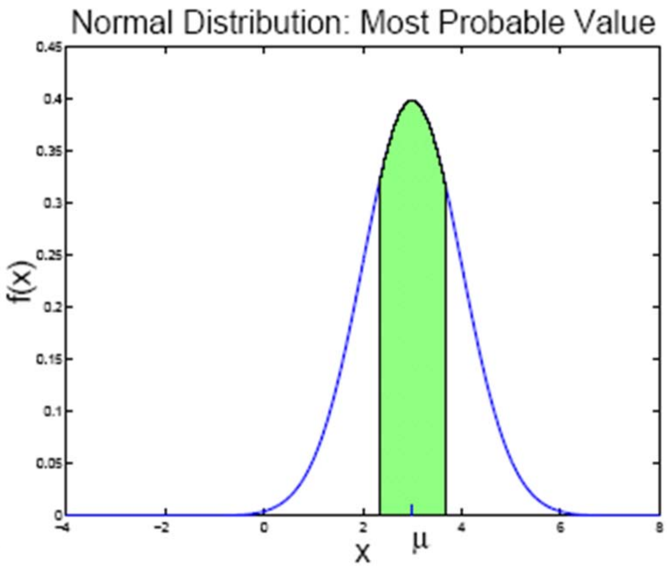
$$P(\mu-3\sigma \le X \le \mu+3\sigma) = \frac{1}{2}\left[\Phi\left(\frac{3}{\sqrt{2}}\right)-\Phi\left(-\frac{3}{\sqrt{2}}\right)\right] = \Phi\left(\frac{3}{\sqrt{2}}\right) \approx 0.9973$$



1.3 Normal Probability Distribution Function

The most probable error

- If there is equal chance that outcome will falls outside or inside of shaded region
  → most probable error



$$P(\mu - r \le X \le \mu + r) = \frac{1}{2}$$

$$\frac{1}{2}\left[\Phi\left(\frac{r}{\sqrt{2}\sigma}\right) - \Phi\left(-\frac{r}{\sqrt{2}\sigma}\right)\right] = \Phi\left(\frac{r}{\sqrt{2}\sigma}\right) = \frac{1}{2}$$

$$r = \sqrt{2}\Phi^{-1}\left(\frac{1}{2}\right)\sigma \approx 0.6745\sigma$$

1.4 Sampling Distributions

- Suppose we are interested in functions of N random variables $X_1, X_2, X_3, \ldots, X_N$
  - $X_1 \ldots X_N$ are independent
  - $X_1 \ldots X_N$ share the same distribution
- Population sample mean

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

- The goodness of fit depends on the behavior of random variable

$$\bar{X} = \frac{1}{N}\sum_{n=1}^{N} X_n$$

1.4 Sampling Distributions

- X_n are independent, normally distributed variable
  - With common mean $E(X_n) = \mu$
  - With common variance $V(X_n) = \sigma^2$

- It can be shown that,

$\bar{X}$ is normally distributed variable

$$E(\bar{X}) = E\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N}\sum_{n=1}^{N} E(X_n) = \frac{1}{N}\sum_{n=1}^{N} \mu = \mu$$

$$V(\bar{X}) = V\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2}\sum_{n=1}^{N} V(X_n) = \frac{1}{N^2}\sum_{n=1}^{N} \sigma^2 = \frac{\sigma}{N}$$

therefore, $\bar{X}$ has a mean $\bar{\mu} = \mu$ and variance $\bar{\sigma}^2 = \sigma/N$

1.5 Central Limit Theorem

고려대학교

- Let ;
  - $X_1 \ldots X_N$ are independent
  - $X_1 \ldots X_N$ share the same distribution
  - Each $X_1 \ldots X_N$ are not normally distributed
  - Mean = μ , variance = σ^2
- In nature, the behavior of variable often depends on the *accumulated effect on large number* of small random factors → behavior is approximately **normal.**
- Central Limit Theorem
  - $\bar{X}$: Normal distribution
  - Mean value $E(\bar{X}) = \mu$
  - Variance $V(\bar{X}) = \sigma^2 / N$

1.6 Central Limit Theorem and Monte Carlo Method

- In Monte Carlo Simulation, we compute a quantity of interest by random sampling population
- The central limit theorem can be applied
- Sampling scheme in MC simulation require reduction of the value $\sigma \rightarrow$ "Variance Reduction Technique"

2. Generating Non-uniform Random Numbers

- Topics
  - Methods for generating random numbers those obey non-uniform probability distributions
    - Discrete random variables
    - The inverse function method
    - The superposition method
    - The rejection method

2.1 Modeling a discrete random variables

x	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

$$\sum_{k=1}^{n} p_n = 1$$

- Method
  - Divide [0,1] interval into n segments with lengths equal to $p_1, p_2, \ldots, p_n$
  - Generate uniform random number u in [0,1]
  - If u reside $p_1 + \ldots + p_{k-1} < u < p_1 + \ldots + p_k$, then choose x_k as the value of x

If u<p_1
$x = x_1$
Else if u < $p_1 + p_2$
$x = x_2$
Else if …
:
:
Else if u<$p_1 + \ldots + p_{n-1}$
$x = x_{n-1}$
Else
$x = x_n$
End if

2.2 The Inverse Function Method

- The Inverse Function Method
  - General scheme for generating non-uniform random numbers
  - The method involves evaluation of indefinite integral
    - cannot be applied to all types of PDF
- Methods
  - Y : uniform random variable in [0,1]
  - transform y→ x
  - x are distributed according to PDF *f(x)*

2.2 The Inverse Function Method

고려대학교

$$P(y \le Y \le y + dy) = dy$$ ← y has uniform distribution in [0,1]

$$P(x \le X \le x + dx) = \int_x^{x+dx} f(t)dt = f(x)dx$$

$$dy = f(x)dx$$

$$y = F(x)$$ ← Cumulative Distribution Function

$$x = F^{-1}(y)$$

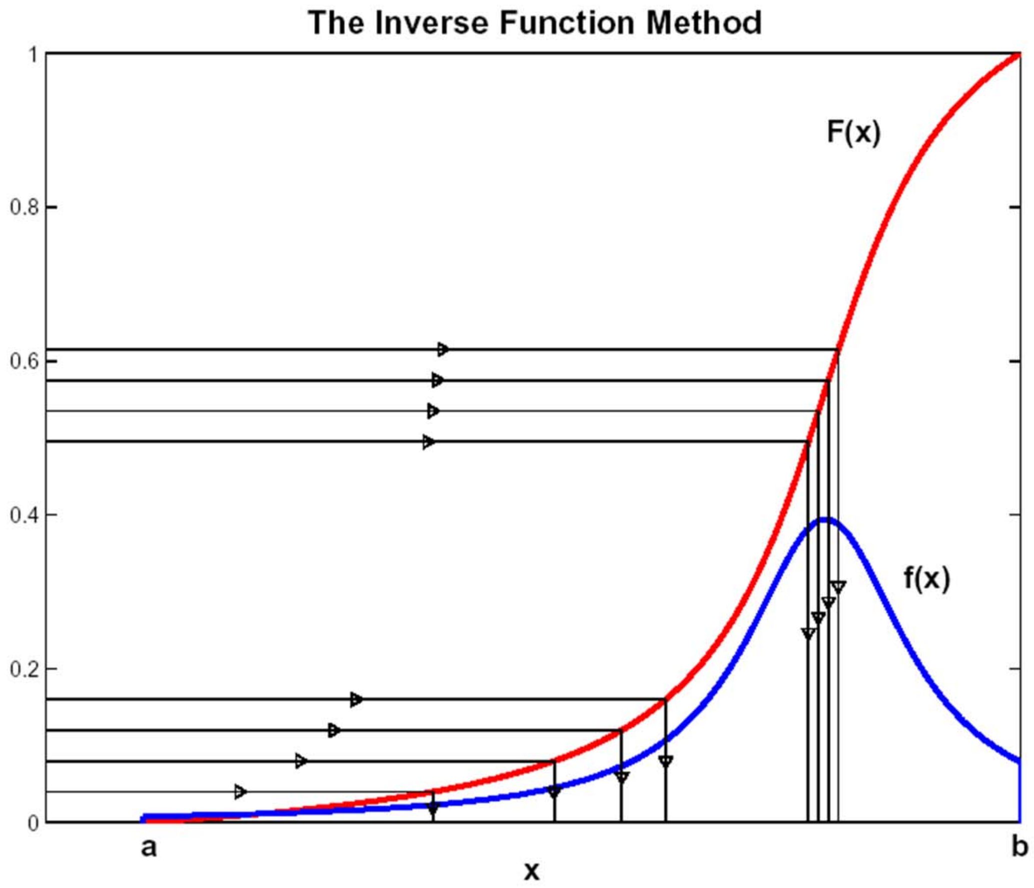
i) First, we have to find CDF, F(x)

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

ii) First, we have to find inverse function, $F^{-1}(y)$

2.2 The Inverse Function Method

Graphical Interpretation of Inverse Function Method



[QUIZ]

- Find inverse function for
  - Uniform distribution , f(x) = c, in [a,b] , otherwise f(x) = 0
  - Exponential distribution , f(x) = a exp(-ax) for x>0 , otherwise, f(x) = 0

2.3 Superposition Method

고려대학교

- If CDF, *F(x)* can be written as a superposition of two or more functions

$$F(x)=\sum_{k=1}^{m} c_k F_k(x)$$

$$c_k > 0$$

$$\sum_{k=1}^{m} c_k = 1$$

- Choice of Fk(x) relies on the generation of discrete random integer variable Q

q	1	2	3	…	k
$p(q)$	c_1	c_2	c_3		c_m

$$P(Q=k)=c_k$$

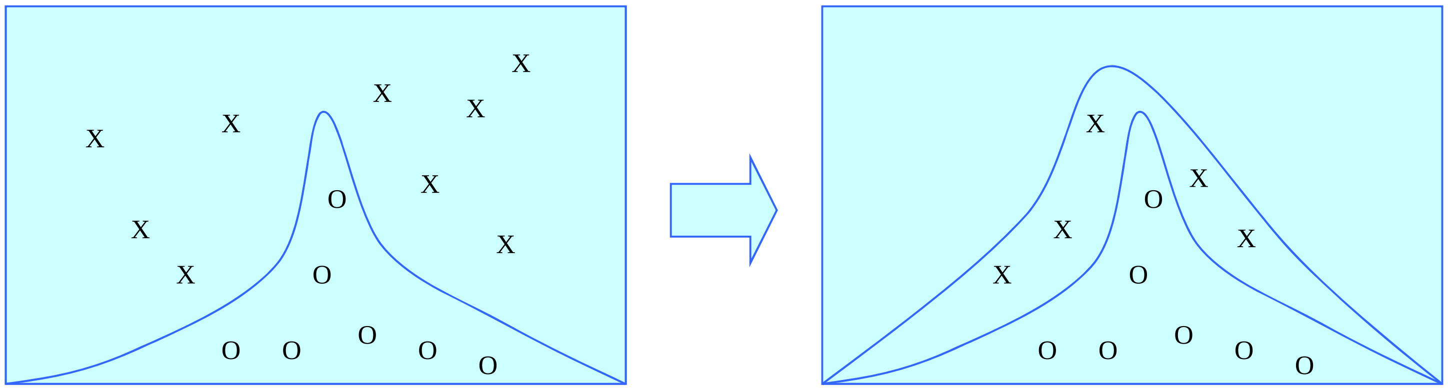
2.3 Superposition Method

- Algorithm
  - Randomly pick an integer u_1 from 1 to m according to $c_1, \ldots, c_k$
    - Use method for discrete random variables
  - Randomly choose a value u_2, in [0,1]
  - $x = F^{-1}_k(x)$

- *Example)* $f(x) = (3/8)(1+x^2)$

2.4 The generalized rejection method

- An inversion of CDF is not an easy task
- The case when the function similar to CDF is available
- Basic idea



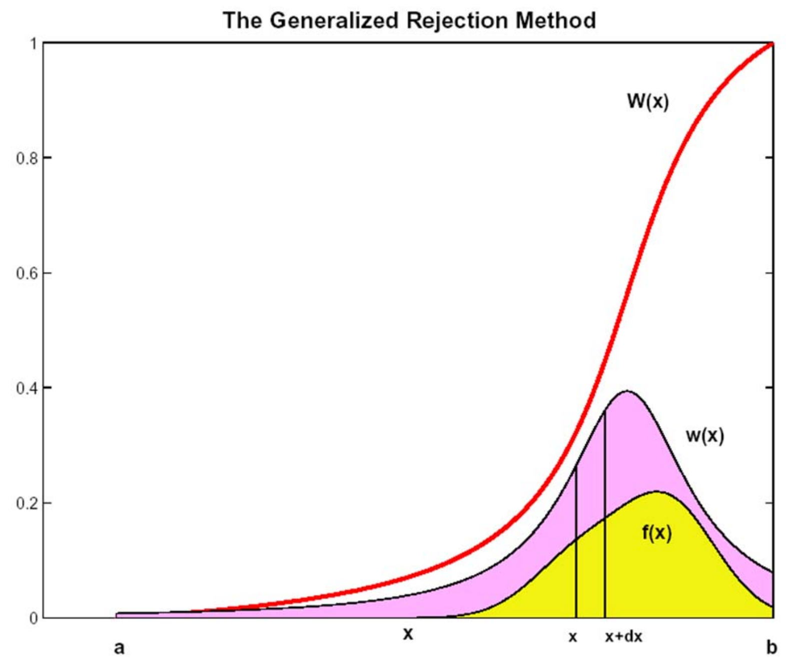
2.4 The generalized rejection method

■ The comparison function *w(x)*

$$w(x) \ge f(x) \qquad \text{for all x}$$

$$W(x) = \int_{-\infty}^{x} w(x)dx \qquad \text{can be calculated analytically}$$

$$\int_{-\infty}^{\infty} w(x)dx = A \qquad \text{not 1}$$



2.4 The generalized rejection method

- Algorithm
  - Generate random number u in [0,1]

    then Au is a random number in [0,A]
  - $x = W^{-1}(Au)$
  - choose a random y in [$0,w(x)$]

    then, (x,y) is uniformly distributed in $w(x)$
  - if $y <= f(x)$ → accept value
  - if $y < f(x)$ → reject value
- Points are distributed according to f(x)

2.4 The generalized rejection method

고려대학교

- The efficiency of generalized rejection method

$$e = \frac{\int_{-\infty}^{\infty} f(x)dx}{\int_{-\infty}^{\infty} w(x)dx} = 1/A$$

- For greater efficiency, A→ 1 (Inversion method)

Next Lecture

- General Monte-Carlo Simulation Method
- Variance Reduction in Monte-Carlo Method