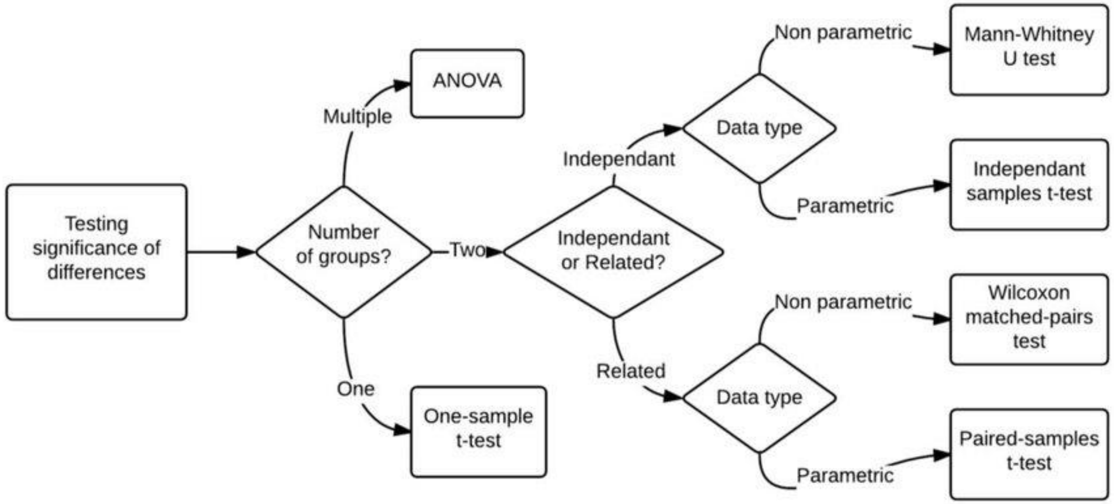
확률 이론 Review



부산대학교 정보·의생명공학대학

정보컴퓨터공학부

Basic Axioms, Theorems, Definitions related to Probability

❖ Axioms

- For any event $\boldsymbol{A}$, $\mathbf{Pr}(\boldsymbol{A}) \geq \mathbf{0}$
  - A negative probability does not make sense
- Probability of the sample space $\boldsymbol{S}$ is $\mathbf{Pr}(\boldsymbol{S}) = \mathbf{1}$
- If $\boldsymbol{A} \cap \boldsymbol{B} = \emptyset$, then $\mathbf{Pr}(\boldsymbol{A} \cup \boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B})$

❖ Theorems

- $\boldsymbol{Pr}(\bar{A}) = \mathbf{1} - \mathbf{Pr}(\mathbf{A})$
- $\mathbf{Pr}(\mathbf{A} \cup \boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B}) - \mathbf{Pr}(\mathbf{A} \cap \mathbf{B})$
- For any events A and B such that $\mathrm{Pr}(B) \neq 0$,

$$\boldsymbol{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{Pr}(\boldsymbol{B}|\boldsymbol{A}) \cdot \boldsymbol{Pr}(\boldsymbol{A})}{\boldsymbol{Pr}(\boldsymbol{B})}$$

  - Known as **"Bayes's Theorem"**

❖ Definitions

- The **Joint Probability** of the set A and B

$$\mathbf{Pr}(\mathbf{A} \cap \boldsymbol{B}) \equiv \mathbf{Pr}(\mathbf{A}, \mathbf{B})$$

- The **Conditional Probability** of A conditioned on knowing that B has occurred is

$$\boldsymbol{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{Pr}(\mathbf{A}, \boldsymbol{B})}{\boldsymbol{Pr}(\boldsymbol{B})}$$

- Two events are statistically **independent** if and only if

$$\mathbf{Pr}(\boldsymbol{A}, \boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{A}) \cdot \boldsymbol{Pr}(\boldsymbol{B})$$

Random Variables

❖ Random Variable

- Described informally as a variable whose values depend on outcomes of a random phenomenon
- a real-valued function of the elements of a sample space, *S*.

❖ Two Types

- Discrete Random Variables (이산확률변수)
  - Discrete Sample Space, Ex) $\mathbb{Z}$:Integer Set
  - 확률 변수가 이산적인 값 만을 가질 경우
  - Ex) Binomial (이항), Geometric (기하), Poisson (포아송)
- Continuous Random Variables (연속확률변수)
  - Continuous Sample Space, Ex) $\mathbb{R}$: Real Number Set
  - 확률 변수가 연속적인 값을 가질 경우, Ex) 실수
  - Ex) Uniform(항등), Exponential(지수), Gaussian (정규)
- 확률 변수가 가질 수 있는 값의 분포 → Distribution
  - 확률 변수/분포를 표현하기 위한 함수가 필요

❖ Description of Random Variables

❖ PMF – Probability Mass Function

- $P_X(k) = \Pr(X = k)$

❖ CDF – Cumulative Distribution Function

- $F_X(x) = \Pr(X \le x)$

❖ PDF – Probability Density Function

- $f_X(x) = \frac{d}{dx} F_X(x)$

	PMF	CDF	PDF
Discrete	O	O	
Continuous		O	O

Properties of PMF, CDF, PDF

❖ PMF Definition and Properties

$$\boldsymbol{P_X(k) = Pr(\mathbf{X} = \mathbf{k})}$$

- For all k, $0 \le P_X(k) \le 1$
- $\sum_{k=-\infty}^{\infty} P_X(k) = 1$
- For any set $A \subset S$, $\Pr(k \in A) = \sum_{k \in A} P_X(k)$

❖ CDF Definition and Properties

$$\boldsymbol{F_X(x) = \Pr(X \le x)}$$

- $F_X(-\infty) = 0,\ F_X(\infty) = 1$
- $0 \le F_X(x) \le 1$
- $For\ \ x_1 < x_2,\ F_X(x_1) \le F_X(x_2)$
    - *Increasing function*
- $For\ \ x_1 < x_2,\ \Pr(x_1 \le X \le x_2) = F_X(x_2) - F_X(x_1)$

❖ PDF Properties

$$f_X(x) = \lim_{\epsilon \to 0} \frac{\Pr(x \le X < x + \epsilon)}{\epsilon}$$
$$= \lim_{\epsilon \to 0} \frac{F_X(x+\epsilon) - F_X(x)}{\epsilon} = \boldsymbol{\frac{dF_X(x)}{dx}}$$

- $f_X(x) \ge 0$
- $f_X(x) = \frac{dF_X(x)}{dx}$
- $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $\int_a^b f_X(x)dx = \Pr(a \le X < b)$

❖ PMF, CDF, PDF 함수가 주어지면 확률 값을 계산 할 수 있다.

Some Discrete Random Variables

❖ Bernoulli Random Variables

- Experiments with 2 possible outcomes

$$P_X(k) = \begin{cases} 1-p & k=0 \\ p & k=1 \end{cases}$$

❖ Binomial Random Variables

- **Repeating a Bernoulli trial n times**, Where the outcome of each trial is **independent**
- **Notation : $X \sim B(n,p)$**

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \cdots, n$$

❖ Geometric Random Variables

- Repeating a Bernoulli trial **until the first occurrence** of the outcome ξ_1

$$P_X(k) = (1-p)^{k-1} \cdot p, \qquad k = 1, 2, 3, \cdots$$

  - Or until the first occurrence of the outcome ξ_0
    - $P_X(k) = (1-p) \cdot p^{k-1}, \quad k = 1, 2, 3, \cdots$

❖ Poisson Random Variables

- A limiting case of the Binomial random variable for very large n with $np = \lambda$.
- Usually used to model the number of events that occur with a known **average rate (λ)** and **independently**.
- Notation : $X \sim Poisson(\lambda)$

$$P_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \cdots$$

Some Continuous Random Variables (1)

❖ Uniform Random Variables

- Events occur equally likely on the interval [a,b]
- Notation : $X \sim Uniform(a, b)$

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x < b \\ 0 & elsewhere \end{cases}$$

❖ Exponential Random Variables

- The time between events in a Poisson process, i.e., a process in which events occur continuously and independently at a constant average rate λ
- Notation : $X \sim Exp(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \ge 0$$

❖ Gaussian Random Variables

- The Gaussian Random Variable is referred to as a “normal” random variable/distribution
- μ=mean, σ=standard deviation
- Notation : $X \sim N(\mu, \sigma^2)$.
  - $\mu = 0, \sigma = 1$ → standard normal

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

$$x \in (-\infty, \infty)$$

Some Continuous Random Variables (2)

❖ Gamma Random Variables

- The Gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-square distribution are special cases of the gamma distribution. There are two different parameterizations in common use:

1. With a shape parameter k and a scale parameter θ.
2. With a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter.

$$f_X(x) = \frac{1}{\Gamma(k) \cdot \theta^k} x^{k-1} \cdot e^{-\frac{x}{\theta}}, \qquad x \geq 0$$

- The Gamma Function - $\Gamma(k)$ - is one commonly used extension of the factorial function to complex numbers.
- For any positive integers,

$$\Gamma(z+1) = z!$$

- For complex numbers with a positive real part,

$$\Gamma(z) = \int_0^\infty t^{z-1} \cdot e^{-t} dt$$

❖ Student's T Random Variables

- A family of continuous probability distributions that arise when estimating the mean of a normally-distributed population in situations where the sample size is small and the population's standard deviation is unknown.
- if we take a sample of n observations from a normal distribution, then the T-distribution with $\nu = n - 1$ degrees of freedom can be defined as the distribution of the location of the sample mean relative to the true mean, divided by the sample standard deviation, after multiplying by the standardizing term $\sqrt{n}$. In this way, the T-distribution can be used to construct a confidence interval for the true mean

$$f_X(x, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \cdot \nu}} \cdot \frac{\left(1+x^2\right)^{-\frac{(\nu+1)}{2}}}{\nu}, \quad x \in (-\infty, \infty), \nu > 0$$

부산대학교 PUSAN NATIONAL UNIVERSITY

Expected Value of a Random Variable

- Expected Value of a Random Variable X
  - Mean, average, expectation
  - For Continuous Random Variables

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx \rightarrow \mu_X$$

  - For Discrete Random Variables

$$E[X] = \sum_k k \cdot P_X(k)$$

- Expected Values for various functions $g(X)$
  - For Continuous Random Variables

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

  - For Discrete Random Variables

$$E[g(X)] = \sum_k g(k) \cdot P_X(k)$$

- Theorem

$$E[aX + b] = aE[X] + b,$$

$$E\left[\sum_{k=1}^{N} g_k(X)\right] = \sum_{k=1}^{N} E[g_k(X)]$$

Expected Values of Various Functions of Random Variables

Name	Functions of X	Expected value, notation
Mean, Average, Expected Value, Expectation, First Moment	$g(x) = x$	$\mu_X = \bar{X} = E[X]$
nth Moment	$g(x) = x^n$	$\overline{X^n} = E[X^n]$
nth Central Moment	$g(x) = (x - \mu_X)^n$	$\overline{(X - \mu_X)^n} = E[(X - \mu_X)^n]$
Variance	$g(x) = (x - \mu_X)^2$	$\sigma_X^2 = E[(X - \mu_X)^2]$
Coefficient of Skewness	$g(x) = \left(\frac{x - \mu_X}{\sigma_X}\right)^3$	$c_s = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right]$
Coefficients of Kurtosis	$g(x) = \left(\frac{x - \mu_X}{\sigma_X}\right)^4$	$c_k = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^4\right]$
Characteristic function	$g(x) = e^{i\omega x}$	$\Phi_X(\omega) = E\left[e^{i\omega X}\right]$
Moment Generating Function	$g(x) = e^{sx}$	$M_X(\omega) = E[e^{sX}]$
Probability Generating Function	$g(x) = z^x$	$H_X(z) = E[z^X]$

Moments and Central Moments

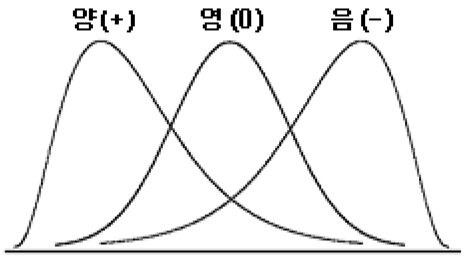
❖ n th Moment

$$E[X^n]$$

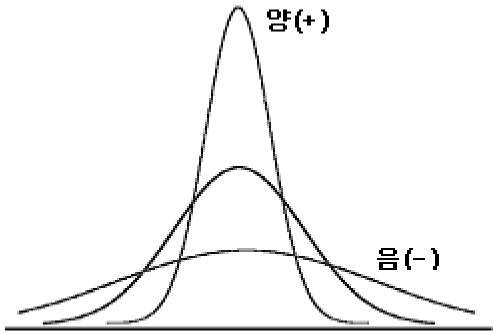
❖ n th Central Moment

$$E[(X-\mu_X)^n]$$

- The 2nd Central Moment : Variance (분산(分散),
  - A Measure of the Width of the PDF
  - Standard Deviation
- The 3rd Central Moment : Skewness (왜도(歪度), 비대칭도)
  - A Measure of the Symmetry of the PDF
- The 4th Central Moment : Kurtosis (첨도(尖度), 꼬리부분의 길이와 중앙부분의 뾰족함에 대한 정보 제공)
  - A Measure of the peakedness of a random variable near the mean



[그림 1-2] 왜도



[그림 1-3] 첨도

Sequence of Random Variables

- Random Process $X(t), X_k$
  - **An indexed collection of random variables**
  - is a random function of time; is an indexed collection of random variables; Often used to represent the evolution of some random value, or system, over time.
  - Index Space : Discrete, Continuous,
  - Sample Space : Discrete, Continuous
    - Discrete Time Discrete Random Process,
    - Discrete Time Continuous Random Process,
    - Continuous Time Discrete Random Process,
    - Continuous Time Continuous Random Process

- Sequence of IID Random Variables
  - ***IID - Independent and Identically Distributed***
  - Sequence of Random Variables : X_k, a kind of random process where the index (time) is discrete.

$\mu_X, \ \hat{\mu}, \ \mu$?

Population
모집단(母集團)

X

$E[X] = \mu_X$
$Var[X] = \sigma_X^2$

Sample set Instance

Independent & Identical

X_1	X_2	X_3	$\cdots$	X_n	IID Random **Variables**
x_1	x_2	x_3		x_n	Sample **Instance/Values**

$$\hat{\mu} = \frac{1}{n} \cdot (x_1 + x_2 + \ \ldots + x_n)$$

Mean (Expected **value**) of a sample set instance

$$\mu = \frac{1}{n} \cdot (X_1 + X_2 + \ \ldots + X_n)$$

A random **variable** for a mean of sample

$E[\mu] = ?, \qquad Var[\mu] = ?$

Ex) $X\sim$ 동전 던지기 $(n = 2)$

$$\mu_X = \frac{1}{2}, \sigma_X^2 = \left(\frac{1}{2}\right)^2$$

	X_1	X_2
S_1	0	0
S_2	0	1
S_3	1	0
S_4	1	1

$\hat{\mu}_1 = 0 \qquad \Pr(\mu = \hat{\mu}_1) = 1/4$

$\hat{\mu}_2 = 0.5 \qquad \Pr(\mu = \hat{\mu}_2) = 1/4$

$\hat{\mu}_3 = 0.5 \qquad \Pr(\mu = \hat{\mu}_3) = 1/4$

$\hat{\mu}_4 = 1 \qquad \Pr(\mu = \hat{\mu}_4) = 1/4$

Ex) $X\sim$ 동전 던지기 $(n = 3)$

$$\mu_X = \frac{1}{2}, \sigma_X^2 = \left(\frac{1}{2}\right)^2$$

	X_1	X_2	X_2
S_1	0	0	0
S_2	0	0	1
S_3	0	1	0
S_4	1	0	0
S_5	0	1	1
S_6	1	0	1
S_7	1	1	0
S_8	1	1	1

$\hat{\mu}_1 = \qquad \Pr(\mu = \hat{\mu}_1) =$

$\hat{\mu}_2 = \qquad \Pr(\mu = \hat{\mu}_2) =$

$\hat{\mu}_3 = \qquad \Pr(\mu = \hat{\mu}_3) =$

$\hat{\mu}_4 = \qquad \Pr(\mu = \hat{\mu}_4) =$

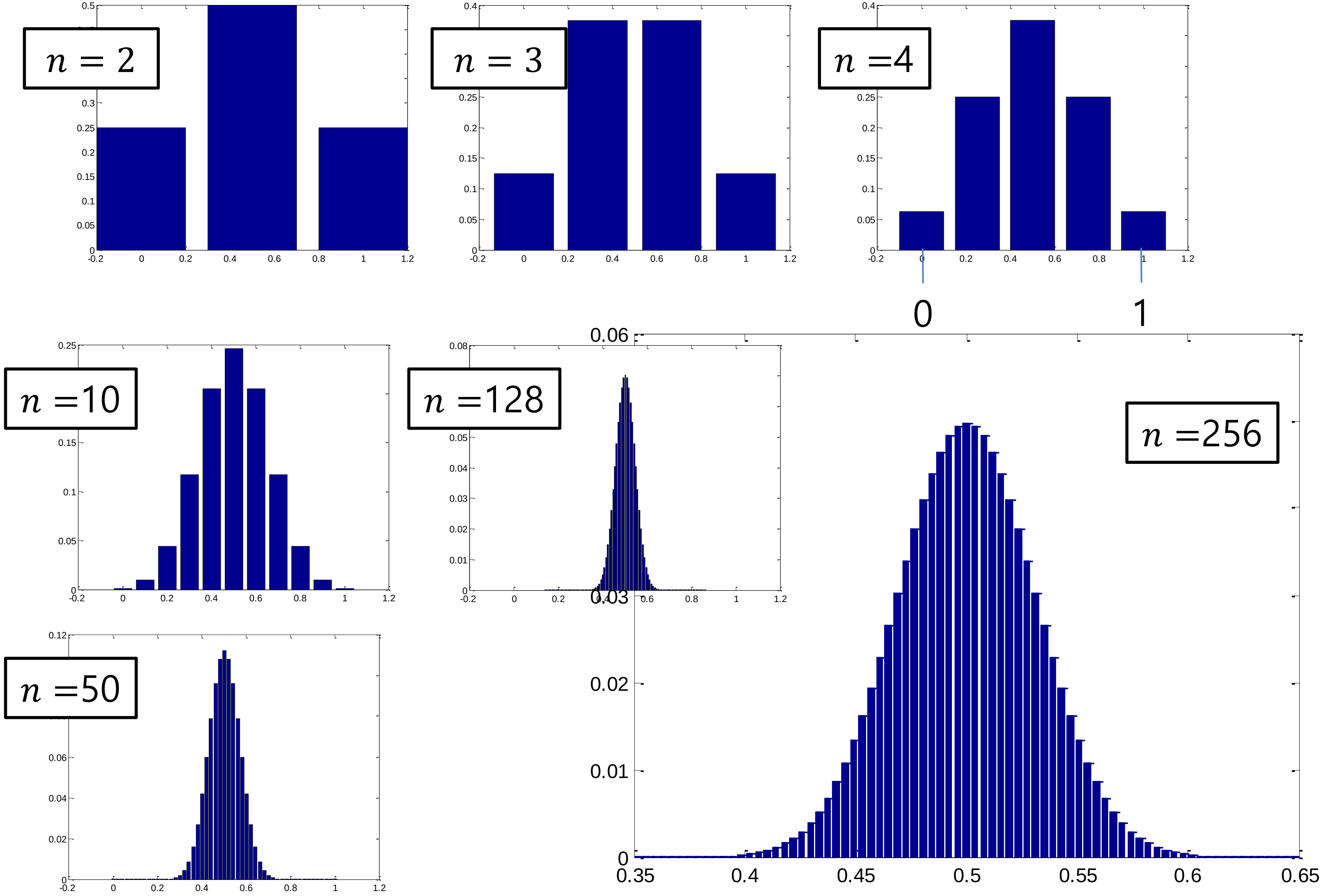
$\hat{\mu}_5 = \qquad \Pr(\mu = \hat{\mu}_5) =$

$\hat{\mu}_6 = \qquad \Pr(\mu = \hat{\mu}_6) =$

$\hat{\mu}_7 = \qquad \Pr(\mu = \hat{\mu}_7) =$

$\hat{\mu}_8 = \qquad \Pr(\mu = \hat{\mu}_8) =$

The Distribution of μ



The Law of Large Numbers

❖ Theorem (The Weak Law of Large Numbers) :

Let S_n be the sample mean computed from n IID random variables, $X_1, X_2, \dots, X_n$. The sequence of sample means, S_n, converges in probability to the true mean of the underlying distribution, $F_X(x)$.

- Proof : Textbook 참조, Chebyshev's Inequality 활용

- Implication : We can estimate the mean of random variable with any amount of precision with arbitrary probability if we use a sufficiently large number of samples.

The Central Limit Theorem

❖ Theorem (The Central Limit Theorem) :

Let X_i be a sequence of IID random variables with mean μ_X and variance σ_X^2.

Define a new random variable, Z, as a (shifted and scaled) sum of the X_i :

$$\boldsymbol{Z = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \frac{X_i - \mu_X}{\sigma_X}}$$

- Note that Z has been constructed such that $E[Z] = 0$ and $Var(Z) = 1$.
  In the limit as n approaches infinity, the random variable Z converges in distribution to a standard normal random variable.

❖ Remarks

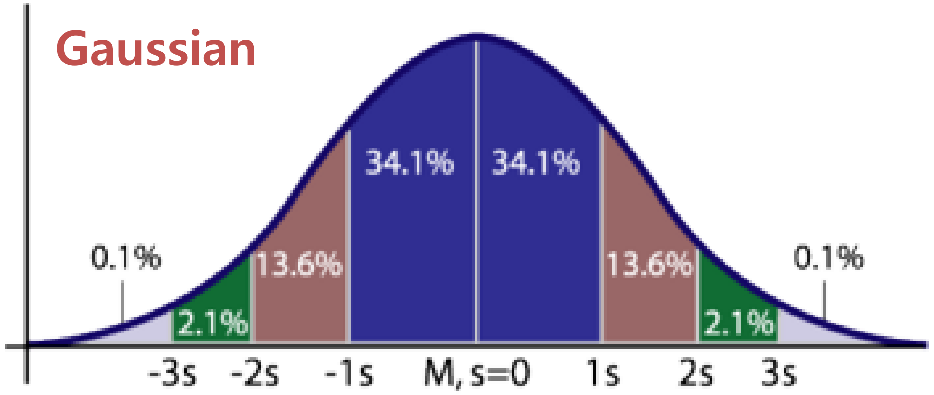
- No restrictions were put on the distribution of the X_i.
  Any infinite sum of IID random variables, regardless of the distribution.
- In many cases, even if X_i are not IID, the central limit theorem may be applied.
- From a practical standpoint, the central limit theorem implies that for the sum of a sufficiently (but finite) number of random variables, the sum is approximately Gaussian distributed.

The Central Limit Theorem

$$\mu = \frac{1}{n} \cdot (X_1 + X_2 + \ \dots + X_n) \quad \Rightarrow \quad E[\mu] = \mu_X \qquad Var[\mu] = \frac{\sigma_X^2}{n}$$

If n is Large Enough

$$\Rightarrow \quad \mu \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) = N\left(\mu_X, \left(\frac{\sigma_x}{\sqrt{n}}\right)^2\right)$$

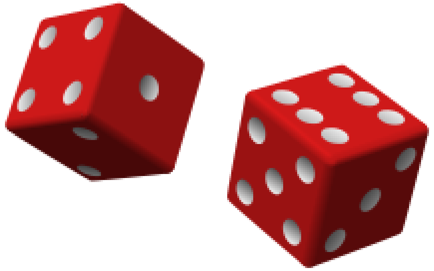


$$Pr\left(|\mu - u_X| < 2 \cdot \frac{\sigma_X}{\sqrt{n}}\right) = ?$$

Random Number Generator

- RNG (Random Number Generator)
  - http://en.wikipedia.org/wiki/Random_number_generator - Computational Method
    - Most computer generated random numbers use **Pseudo random number generators** (PRNGs) which are algorithms that can automatically create long runs of numbers with good random properties but eventually the sequence repeats (or the memory usage grows without bound). This kind of random numbers are fine in many situations but are not as random as numbers generated from electromagnetic atmospheric noise used as a source of entropy. The series of values generated by such algorithms is generally determined by a fixed number called a seed. One of the most common PRNG is the linear congruential generator, which uses the recurrence
  - A simple PRNG : Linear Congruential Generator
    - $X_{n+1} = (a \cdot X_n + b) \bmod m$; Ex) $m = 2^{32}, a = 1664525, c = 1013904223$

- Application ?
  - RNGs have applications in **gambling, statistical sampling, computer simulation, cryptography**, completely randomized design, and other areas where producing an unpredictable result is desirable.



Monte-Carlo Method

- Monte Carlo methods (or Monte Carlo experiments)
  - are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.
  - Their essential idea is using randomness to solve problems that might be deterministic in principle. They are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use other approaches. Monte Carlo methods are mainly used in three problem classes: optimization, numerical integration, and generating draws from a probability distribution.
  - From Wikipedia - http://en.wikipedia.org/wiki/Monte_Carlo_methods

Estimating π using the Monte-Carlo Method

❖ The idea

- is to simulate random (x, y) points in a 2-D plane with domain as a square of side 1 unit. Imagine a circle inside the same domain with same diameter and inscribed into the square. We then calculate the ratio of number points that lied inside the circle and total number of generated points.
- $\boldsymbol{X \sim \mathbf{Uniform}(0, 1), Y \sim \mathbf{Uniform}(0, 1)}$

$$\frac{\text{area of the circular sector}}{\text{area of the square}} = \frac{\boldsymbol{\pi/4}}{\mathbf{1}} = \frac{\text{\# of points inside the circular sector}}{\text{total \# of points inside the square}}$$

$$\boldsymbol{\pi} = \frac{\text{\# of points inside the circular sector}}{\text{total \# of points inside the square}}$$