# 통계분석
# Statistical Analysis

# Expectation (기대값)

## Discrete Random Variable

Expectation of the function u(x) defined on discrete random variable

$$E[u(X)] = \sum_{x \in D} u(x)p(x)$$

$p(x) = P(X = x)$ : probability distribution for discrete random variable $X$

$u(x)$ : some function defined on the discrete random variable $X$

# Expectation (기대값)

## Continuous Random Variable

Expectation of the function u(x) defined on continuous random variable

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

- ***f(x)*** : probability density distribution for continuous random variable *X*
- ***h(x)*** : some function defined on the continuous random variable X

# Expected Value: Mean

- Discrete random variable

$$E[u(X)] = \sum_{x \in D} u(x)p(x) = \sum_{x \in D} xp(x)$$

- Continuous random variable

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

# Variance and Standard Deviation

- Variance of random variable X

$$\sigma_X^2 = V(x) = \mathrm{Var}[X] = E[(X-\mu)^2]$$
$$= \sum_x p(x)(x-\mu)^2 \ \text{[discrete]}$$
$$= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x) dx \ \text{[continuous]}$$

- Standard deviation of random variable X

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}$$

# Properties of Mean / Variances

- For constant numbers a and b

$$E[aX + b] = a \cdot E[X] + b$$

mean

- For constant numbers a and b

$$\mathrm{Var}[aX + b] = a^2 \cdot \mathrm{Var}[X]$$

- Variance in terms of Mean Values

$$\mathrm{Var}[X] = E(X^2) - [E(X)]^2$$

# Higher Moments

- the 1st moment around zero

$$E[X] = \sum xp(x)$$

- the 2nd moment around E[X]

$$\text{Var}[X] = \sum (x - \mu)^2 p(x) = E[X^2] - E[X]^2$$

- The rth moment around b

$$E[(X - b)^r] = \sum (x - b)^r p(x)$$

- skewness and kurtosis are related to 3rd and 4th moments

# Bernoulli Distribution

Tossing an unfair coin

- **Head (success)** $P(X = 1) = p$
- **Tail (failure)** $P(X = 0) = 1 - p$

$$E[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

$$\mathrm{Var}[X] = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p)$$

p = 0 or 1 : zero variance (no fluctuation, every case is fixed to be one case (head or tail)
p = 0.5 : Max variance