Basic Statistics 2

Class 2

Wonmun Shin

(wonmun.shin@sejong.ac.kr)

Department of Economics, Sejong University

* This lecture note is written based on Professor Chang Sik Kim's lecture note .

Probability Distributions

Probability Distributions

- We have discussed properties of discrete and continuous random variables. Now, we are going to consider some important examples of discrete and continuous random variables.

- **Discrete probability distribution**: Bernoulli distribution, Binomial distribution, Hypergeometric distribution, Poisson distribution

- **Continuous probability distribution**
  - Uniform distribution
  - Exponential distribution
  - Normal distribution
  - χ^2 (chi-square) distribution
  - t distribution
  - F distribution

Normal Distribution

- The most important continuous distribution is the normal distribution which plays a central role in a very large body of statistical analysis. The pdf of normal distribution peaks at the mean and tails off at its extremities, which suggests a bell shape curve.
- Probability density function (pdf): bell-shaped and symmetric about its mean

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We write

$$X \sim N\left(\mu, \sigma^2\right)$$

when the random variable X is normally distributed with mean μ and variance σ^2.

- Some properties of $X \sim N(\mu, \sigma^2)$
  - $E(X) = \mu$ indicates the central location of the pdf.
  - $Var(X) = \sigma^2$ indicates the dispersion of the pdf.
  - It takes its maximum at μ.
- **Standard normal distribution**: the normal distribution with mean 0 and variance 1

$$X \sim N(\mu, \sigma^2)$$
$$\Longrightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

  - Standard normal distribution is very useful because the probability can be easily computed based on the *standard normal table*!

χ^2 Distribution

$$V \sim \chi^2(d)$$

- Random variable V has the χ^2 (chi-square) distribution with d degrees of freedom.
- $E(V) = d$
- $Var(V) = 2d$
- If $Z_i \sim N(0,1)$ and Z_i's are independent, then

$$V = Z_1^2 + Z_2^2 + \cdots + Z_d^2 \sim \chi^2(d)$$

t Distribution

$$T \sim t(d)$$

- Random variable T has the t distribution with d degrees of freedom.
- The pdf is symmetric around 0 and bell-shaped (but it is more dispersed and has thicker tails than the pdf of $N(0,1)$).
- As the degree of freedom increases, the t distribution converges to the $N(0,1)$.
- If $Z \sim N(0,1)$, $V \sim \chi^2(d)$, and Z and V are independent, then

$$T = \frac{Z}{\sqrt{V/d}} \sim t(d)$$

F Distribution

$$F \sim F(d_1, d_2)$$

- Random variable F has the F distribution with degrees of freedom (d_1, d_2)
- The pdf is positively skewed and located over the range of positive numbers.
- If $V_1 \sim \chi^2(d_1)$, $V_2 \sim \chi^2(d_2)$, and V_1 and V_2 are independent, then

$$F = \frac{V_1 / d_1}{V_2 / d_2} \sim F(d_1, d_2)$$

Central Limit Theorem (CLT)

- The *sum* or *average* of **a large number of** *independent* random variables has a distribution **close to normal**, *regardless of the distribution* of the random variables.
- It allows us to make inferences about the true population parameters (for example, mean or variance) on the basis of the sample without knowing the true information about the population.
- Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed ($i.i.d.$) random variables.
  - The CLT says that, *regardless of the common distribution* of X_i's, the distribution of average of X ($i.e. \bar{X}$) follows the normal distribution **when n becomes large**.
  - Again, the CLT do not require X_i's to be normal variables!

Parameter and Estimator

Population and Sample

- We want to know the properties of a large group of objects, given information on a relatively small subset of them.
  - The larger parent group is referred to as a **population**.
  - The subset of population is called a **sample**.
- *"A sample should be representative of the population"*: **Random sampling**
  - A random sampling procedure is to select a sample of n objects from a population.
  - Random sample $\{X_1, X_2, \cdots, X_n\}$ from a population is a collection of random variables that are chosen from the same population distribution and are statistically independent.
  - *i.e.* X_i's are *i.i.d.* for $i = 1, 2, \cdots, n$

Parameter and Estimator

- **Parameter** θ : The unknown characteristic about which information is required
  - eg. $\theta = \mu$
- **Estimator** $\hat{\theta}$: A sample statistic whose realization provide approximation to the population parameter
  - eg. $\hat{\theta} = \bar{X}$
- **Estimate**: A specific numerical realization
  - eg. Let random sample: $\{X_1, X_2, X_3\}$, realization: $\{18, 19, 20\}$
  - μ: population parameter
  - $\bar{X} = \frac{X_1+X_2+X_3}{3}$: (point) estimator of μ
  - $\bar{X} = \frac{18+19+20}{3} = 19$: (point) estimate of μ

Properties of Estimator

Properties of Estimator: Unbiasedness

1. Unbiasedness

- An estimator $\hat{\theta}$ is an unbiased estimator of θ if the mean of the sampling distribution is θ. *i.e.*

$$E\left(\hat{\theta}\right) = \theta$$

- Measure of unbiasedness: $Bias\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$
- $\therefore$ The bias of an unbiased estimator must be zero
- Example: $E\left(\bar{X}\right) = \mu$
  - why?

$$E\left(\bar{X}\right) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} \cdot E\left(\sum X_i\right) = \frac{1}{n} \cdot E\left(X_1 + X_2 + \cdots + X_n\right)$$
$$= \frac{1}{n} \cdot n \cdot \mu = \mu$$

  - $Bias\left(\bar{X}\right) = E\left(\bar{X}\right) - \mu = \mu - \mu = 0$

2. Efficiency

- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators. Then, $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$$

- Example: $Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$, $Var\left(X_{Med}\right) = \frac{\pi}{2} \cdot \frac{\sigma^2}{n}$
  - $\bar{X}$ is a more efficient estimator than X_{Med} since

$$Var\left(\bar{X}\right) = \frac{\sigma^2}{n} < \frac{\pi}{2} \cdot \frac{\sigma^2}{n} = Var\left(X_{Med}\right)$$

Mean Squared Error (MSE)

- When the unbiased estimator is unsatisfactory because its variance is too big, we may sacrifice a little bias for the smaller variance.
- One particular measure of expected closeness is **Mean Squared Error (MSE)**.
- The expectation of the squared difference between the estimator and the parameter:

$$MSE\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var\left(\hat{\theta}\right) + \left[Bias\left(\hat{\theta}\right)\right]^2$$

Properties of Estimator: Consistency

3. Consistency: *a large sample property*

- An estimator $\hat{\theta}$ is a consistent estimator of θ if

$$\lim_{n \to \infty} P\left[\left|\hat{\theta} - \theta\right| < \epsilon\right] = 1 \text{ for any } \epsilon > 0$$

- That is, the larger the sample is, the higher probability that $\hat{\theta}$ lies close to θ (or, $\hat{\theta}$ converges in probability to θ).
- In other expression,

$$\hat{\theta} \xrightarrow{p} \theta \quad \text{or} \quad p \lim_{n \to \infty} \hat{\theta} = \theta$$

- Note Sufficient condition of consistent estimator

$$p \lim_{n \to \infty} MSE\left(\hat{\theta}\right) = 0$$

Properties of Estimator: Consistency

$$p \lim_{n \to \infty} MSE\left(\hat{\theta}\right) = 0$$

- $\left[Bias\left(\hat{\theta}\right)\right]^2 > 0, \; Var\left(\hat{\theta}\right) > 0$
- $\therefore MSE\left(\hat{\theta}\right) \to 0$ iff $Bias\left(\hat{\theta}\right) \to 0, \; Var\left(\hat{\theta}\right) \to 0.$
- Example
    - $Bias\left(\bar{X}\right) = E\left(\bar{X}\right) - \mu = 0$
    - $Var(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$
    - $\therefore \bar{X}$ is a consistent estimator of μ

Hypothesis Testing

Hypothesis

- We are often interested in using data in accessing the validity of some conjecture, or **hypothesis**.
- Generally, hypotheses are formed about the population parameter (θ).
- Null hypothesis (H_0)
  - A hypothesis that is held to be true unless sufficient evidence to the contrary is obtained.
  - Therefore, we hold the null hypothesis to be true unless there is strong evidence that it is not.
  - (e.g.) $H_0 : \theta = 70$
- Alternative Hypothesis (H_1)
  - A hypothesis against which the null hypothesis is tested and which is held to be true if the null is incorrect.
  - One-sided alternative: (e.g.) $H_1 : \theta < 70$
  - Two-sided alternative: (e.g.) $H_1 : \theta \neq 70$

Type I and Type II Error

- **Type I error**: Error of rejecting H_0 when H_0 is true.

$$\begin{aligned} P\left(\text{Type I error}\right) &= \text{size of test} \\ &= \text{significance level} \\ &= \alpha \end{aligned}$$

- **Type II error**: Error of not rejecting H_0 when H_0 is false.

$$P\left(\text{Type II error}\right) = \beta$$

  - Power:

$$\begin{aligned} 1 - P\left(\text{Type II error}\right) &= P\left(\text{reject } H_0 \text{ when } H_0 \text{ is false}\right) \\ &= \text{Power of test} \\ &= 1 - \beta \end{aligned}$$

Type I and Type II Error [cont'd]

- Summary
  - α: Size of test, Significance level
  - $1-\beta$: Power

	H_0 is True	H_0 is False
Don't reject (Accept) H_0	Okay $(1-\alpha)$	Type II error (β)
Reject H_0	Type I error (α, **Size**)	Okay ($1-\beta$, **Power**)

- α and β are inversely related: the smaller α we choose, the larger β will be.
- The only way of lowering both error would be to obtain more information about the true parameter.
- Conventionally, α is chosen to a small number ($\alpha = 0.10, 0.05$, or 0.01).

Steps of Hypothesis Testing

- **[Step 1]** Set the hypothesis
  - Specify H_0 and H_1.
- **[Step 2]** Test statistic
  - Figure out a distribution of estimator.
  - Compute a test statistic.
- **[Step 3]** Set the rejection region
  - Choose a significance level.
  - Compute the critical value.
  - Then, set the rejection area.
- **[Step 4]** Decision
  - Decide whether to reject or do not reject H_0.