# Econometrics Midterm TA session

## Hyunwoo Roh

### Random Variable $X$

- Discrete
  - $P(X = x) = f_X(x)$
  - $P(X \mid Y) = f_{X|Y}(x|y)$
- Continuous
  - $P(a < X < b) = \int_a^b f_X(x)dx$
  - $P(X \in A \mid Y) = \int_A f_{X|Y}(x|y)dx$

### Expectation

- $E[X]$
  - $\sum_{i=1}^n x_i f_X(x_i)$ if X is discrete
  - $\int_{-\infty}^{\infty} x f_X(x)dx$ if X is continuous
- $E[g(X)]$
  - $\sum_{i=1}^n g(x_i) f_X(x_i)$ if X is discrete
  - $\int_{-\infty}^{\infty} g(x) f_X(x)dx$ if X is continuous
- Expectation is a one-number summary of the distribution
- Linearity
  - If $X_1, \ldots, X_n$ are RV that are identically distributed and $a_1, \ldots, a_n$ are constants, then

$$\begin{aligned} E[\sum_i a_i X_i] &= \sum_i a_i E[X_i] \\ &= \sum_i a_i E[X] \end{aligned}$$

- I.I.D
  - If $X_1, \ldots, X_n$ are i.i.d, then

$$\begin{aligned} E[\prod_{i=1}^n X_i] &= \prod_{i=1} E[X_i] \\ &= \prod_{i=1} E[X] \end{aligned}$$

- Let $X$ be a random variable and let $A \subset R$. Then, $I\{x \in A\}$ is a function, so $I\{X \in A\}$ is a random variable. Moreover,

$$E[I\{X \in A\}] = P\{X \in A\}$$

- [Ex1] $Y$ is binary $\{0, 1\}$ with $Y = \beta_0 + \beta_1 X + U$ and $E[U|X] = 0$

$$P\{Y = 1 \mid X\} = E[I\{Y = 1\} \mid X]$$

**Variance**

- $E[X - E[X]]^2$
- $V[X] = E[X^2] - E[X]^2$
- If $a$ and $b$ are constants, then
$$V[aX + b] = a^2 V[X]$$
- $V[b] = 0$

**Covariance**

- $Cov[X, Y] = E[(X - E[X])(Y - E[Y]) = E[XY] - E[X]E[Y]$
- $V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$
- $V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$

**Conditional Expectation**

- $E[X|Y]$
  - $\sum x f_{X|Y}(x|y)$
  - $\int x f_{X|Y}(x|y)dx$
- $E[Y|X]$
  - $\sum y f_{Y|X}(y|x)$
  - $\int y f_{Y|X}(y|x)dy$
  - Whereas $E[X]$ is a number, $E[X|Y = y]$ is a function of $y$ . Before we observe $Y$, we don't know the value of $E[X|Y]$ so it is a random variable denoted as $E[X|Y]$. In other words, $E[X|Y]$ is the random variable whose value is $E[X|Y = y]$ when $Y = y$. So whenever the $y$ is realized, we can calculate $E[X|Y]$.
- [Ex 1 Cont]
$$\begin{aligned}
&E[I\{Y = 1\} \mid X] \\
=&P\{Y = 1 \mid X\} \\
=&1 \cdot P\{Y = 1 \mid X\} + 0 \cdot P\{Y = 0 \mid X\} \\
=&\sum y f_{Y|X}(y|x) \\
=&E[Y|X]
\end{aligned}$$
- [Ex 2]
  - $X \sim U[0, 1]$ and $Y|X \sim U[0, x]$
  - We expect $E[Y|X = x] = \frac{x}{2}$ , which is a function of $x$
  - $E[Y|X] = \frac{X}{2}$ is a random variable whose value is the number $E[Y|X = x]$ once $x$ is observed (realized)
  - This is not mean independent.

- For any random vector $(X, Y)$ and well defined function $g$ and $h$

$$E[g(X) + h(X)Y \mid X] = g(X) + h(X)E[Y|X]$$

- Law of Iterated Expectations
  - $E[E[Y|X]] = E[Y]$ and $E[E[X|Y]] = E[X]$
- [Ex 2 Cont]
  - $E[E[Y|X]] = E[Y] = \frac{1}{2}$

**Conditional Variance**

- $V[Y|X] = E[(Y - E[Y|X])^2 \mid X] = E[Y^2|X] - (E[Y|X])^2$
- Similarly, it is a random variable whose value is number $V[Y|X = x]$ once $x$ is observed
- [Ex 1 Cont]

$$\begin{aligned} V[U|X] &= E[(U - E[U|X])^2 |X] \\ &= E[U^2|X] \\ &= E[(Y - E[Y|X])^2 |X] \\ &= V[Y|X] \end{aligned}$$

Therefore, $U|X$ is a function of $X = x$

**Estimator for a parameter $\theta$**

It is a rule for using the random sample to construct an estimate,

$$\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)$$

is a function from $X_1, ..., X_n$ to $R$.

- Sample mean : $\theta = E[X]$

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

- Sample Variance : $\theta = V[X]$

$$\hat{\theta}_n = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 = \hat{\sigma}_X^2$$

- Sample CDF : $\theta = P\{X \le x\} = E[I\{X \le x\}]$

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\} = \hat{F}_X(x)$$

Since an estimator of the parameter is a random variable because it depends on random sample, we know that it will typically not be equal to the value of the parameter of interest. We hope to conclude that they get close each other as we have more observations.

**Asymptotic**

- Consistency: as we get more and more data, we eventually know the truth
- Asymptotic normality: s we get more and more data, averages of random variables behave like normally distributed random variables
- The main statistical tool for establishing consistency of estimators is the Law of Large Numbers(LLN) and often used with the technique of Continuous Mapping Theorem (CMT).
- The main tool for establishing asymptotic normality is the Central Limit Theorem (CLT) and often used with the technique of .

**Consistency**

- An estimator $\hat{\theta}_n$ of a parameter $\theta$ is said to be consistent if

  $$\hat{\theta}_n \xrightarrow{p} \theta$$

  We read it as $\hat{\theta}_n$ converges to $\theta$ in probability and this is true if $\forall \epsilon > 0$,

  $$P(\mid \hat{\theta}_n - \theta \mid > \epsilon) \to 0$$

  as $n \to \infty$
- Chebychev
  - Let $A$ be any random variable.

    $$P(\mid A \mid > \epsilon) \leq \frac{E[A^2]}{\epsilon^2}$$

  - When, $A = \hat{\theta}_n - E[\hat{\theta}_n]$,

    $$P(\mid A \mid > \epsilon) \leq \frac{V[\hat{\theta}_n]}{\epsilon^2}$$

**LLN**

- Let $X_1, ..., X_n$ be $i.i.d$ sample of size $n$ from $X$. We assume there exists variance. Then,

  $$\bar{X}_n \xrightarrow{p} E[X]$$

- Using Chebyshev,

  $$P(\mid \bar{X}_n - E[X] \mid > \epsilon) \leq \frac{V[\bar{X}_n]}{\epsilon^2} = \frac{\sigma}{n\epsilon^2}$$

**CMT**

- If $A_n, n \geq 1$ and $B_n, n \geq 1$ are sequence of random variables and $a$ and $b$ are constants that satisfy

  $$A_n \xrightarrow{p} a$$
  $$B_n \xrightarrow{p} b$$

  and $g : R^2 \to R$ is continuous at $(a, b)$, then

  $$g(A_n, B_n) \xrightarrow{p} g(a, b)$$

- [Ex 3]

$$\bar{X}_n \xrightarrow{p} E[X]$$
$$\bar{Y}_n \xrightarrow{p} E[Y]$$

with function $g(s,t) = st$ and by CMT,

$$\bar{X}_n \bar{Y}_n \xrightarrow{p} E[X]E[Y]$$

**Asymptotic Normality**

- A consistent estimator $\hat{\theta}_n$ is asymptotically normally distributed if

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\theta^2)$$

where $\sigma_\theta^2$ is called the asymptotic variance of the estimate. Asymptotic normality says that the estimator not only converges to the unknown parameter, but it converges fast enough at a rate $\frac{1}{\sqrt{n}}$

- Converges in Distribution
  - Let $A_n$ be a sequence of random variables and let $A$ be a continuous random variable. We say that $A_n$ converges in distribution to $A$ if their CDFs converge, that is

$$P\{A_n \leq t\} \to P\{A \leq t\}$$

for every $t \in R$. In this case, we write

$$A_n \xrightarrow{d} A$$

**Central Limit Theorem**

- Let $X_1, ..., X_n$ be i.i.d sample size $n$ from $X$ with mean $E[X_n] = \mu$ and the variance $V[X_n] = \sigma^2$
Then, these are all equivalent statments

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right) \xrightarrow{d} N(0, \sigma_X^2)$$
$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n \cdot \mu_X\right) \xrightarrow{d} N(0, \sigma_X^2)$$
$$\sqrt{n}(\bar{X}_n - \mu_X) \xrightarrow{d} N(0, \sigma_X^2)$$
$$X_n \xrightarrow{d} N(\mu_X, \frac{\sigma_X^2}{n})$$

- If in addition, $V[X] > 0$, then

$$\frac{\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]}{\sqrt{V\left[\sum_{i=1}^{n} X_i\right]}} \xrightarrow{d} N(0,1)$$
$$\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma_X} \xrightarrow{d} N(0,1)$$

In other words, for every $t \in R$,

$$P\left\{\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma_X} \leq t\right\} \to \Phi(t)$$

where $\Phi(t)$ is the CDF of the standard normal distribution.

- [Ex 4] For Bernoulli trials $X_1, ..., X_n$, $\mu = p$ and $\sigma^2 = p(1-p)$. Thus for large enough nm

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

For 100 tosses of a fair coin,

$$\frac{10(\bar{X} - 1/2)}{1/2} = 20\bar{X} - 10 \approx Z_n$$

If we are interested in the probability $\bar{X} \leq 0.4$

$$\begin{aligned} P[\bar{X} \leq 0.4] &= P[20\bar{X} - 10 \leq -2] \\ &\approx P[Z \leq -2] \\ &= \Phi(-2) \end{aligned}$$

**Slutsky Theorem**

- Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ , $a$ is constant. Then
  - $Y_n X_n \xrightarrow{d} aX$
  - $X_n + Y_n \xrightarrow{d} X + a$
- [Ex 5] By LLN, the sample variance

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2$$
$$\frac{\hat{\sigma}^2}{\sigma^2} \xrightarrow{p} 1$$

By Slutsky, the t-statistic

$$\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\hat{\sigma}} = \frac{\sigma}{\hat{\sigma}} \frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma} \xrightarrow{d} N(0,1)$$

- [Ex 6] CLT tells much more than the LLN. It could not tell us less because the CLT implies the LLN.

$$\bar{X}_n - \mu = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\bar{X}_n - \mu)$$

,by Slutsky,

$$\xrightarrow{d} 0 \cdot N(0, \sigma^2) = 0$$

**Two sided hypothesis and C.I**

- $H_0 : E[X] = \mu_0 \qquad H_1 : E[X] \neq \mu_0$
- The alternative hypothesis is said to be two-sided because it allows $E[X]$ to be both $>$ and $<$ compared to $\mu_0$
- Test Statistics

$$T_n = \mid \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}} \mid$$

- To see how to choose $c$ for this choice of $T_n$, note that the probability of rejecting the null hypothesis is simply

$$P\left\{ \mid \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}} \mid > c \right\}$$

- We would like to choose $c$ such that this is at least approximately equal to significance level $\alpha$ when the null hypothesis is true. Formally,

$$P\left\{\mid \frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\hat{\sigma}} \mid > c\right\} \to \alpha$$

- Using the result from [Ex 5],

$$\begin{aligned}
&P\left\{\mid \frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\hat{\sigma}} \mid > c\right\} \\
=&P\left\{\frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\hat{\sigma}} > c\right\} + P\left\{\frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\hat{\sigma}} < -c\right\} \\
=&1-\Phi(c)+\Phi(-c) \\
=&1-\Phi(c)+1-\Phi(c) \\
=&2(1-\Phi(c))
\end{aligned}$$

so it suffices to choose c so that

$$\begin{aligned}
2(1-\Phi(c)) &= \alpha \\
\iff \Phi(c) &= 1-\frac{\alpha}{2}
\end{aligned}$$

Let $c_{1-\alpha/2}$ denote the that stasitfies the above. For example if $\alpha=0.05$,then $c_{0.975}=1.96$ satisfies this equation.

**Confidence Interval**

We want to construct a set that contains $E[X]$ with some pre-specified probability $\alpha$, that is, we want to set $C_n$ such that

$$P\{E[X]\in C_n\}\to 1-\alpha$$

To carry this idea, we would include any $\mu_0$ in our confidence interval that satisfies

$$\mid \frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\hat{\sigma}} \mid \leq c_{1-\alpha/2}$$

which is same as the set $\mu_0$ values

$$\bar{X}_n-\frac{\hat{\sigma}}{\sqrt{n}}c_{1-\alpha/2}\leq \mu_0 \leq \bar{X}_n+\frac{\hat{\sigma}}{\sqrt{n}}c_{1-\alpha/2}$$

Thus, the interval

$$C_n=\left[\bar{X}_n-\frac{\hat{\sigma}}{\sqrt{n}}c_{1-\alpha/2}, \quad \bar{X}_n+\frac{\hat{\sigma}}{\sqrt{n}}c_{1-\alpha/2}\right]$$

is the set we desire.

**Basic Derivation of OLS using moment conditions**

- Moment conditions
  - $E[U]=0$
  - $E[XU]=0$
- Solve for $\beta_0$ and $\beta_1$

- From the first condition

$$
\begin{aligned}
E[U] =& E[Y - \beta_0 - \beta_1 X] \\
=& E[Y] - \beta_0 - \beta_1 E[X] \\
=& 0
\end{aligned}
$$

Therefore,

$$\beta_0 = E[Y] - \beta_1 E[X]$$

Using expression for $\beta_0$ and $E[XU] = 0$

$$
\begin{aligned}
E[XU] =& E[X\left(Y - \beta_0 - \beta_1 X\right)] \\
=& E[X\left(Y - E[Y] - \beta_1 E[X] - \beta_1 X\right)] \\
=& E[X(Y - E[Y]) - \beta_1 X(X - E[X])] \\
=& E[X(Y - E[Y])] - \beta_1 E[X(X - E[X])] \\
=& Cov[X, Y] - \beta_1 V[X] \\
=& 0
\end{aligned}
$$

Hence, we have

$$\beta_1 = \frac{Cov[X, Y]}{V[X]}$$

$$\beta_0 = E[Y] - \frac{Cov[X, Y]}{V[X]} E[X]$$