Department of Economics
Seoul National University
September 27, 2017
ECON 212.703 *Studies in Econometrics*

Probability and Statistics
: Basic Concept Review

1 Distribution of Random Vectors

1.1 Joint Distribution and Expectation

- **Joint density of a random vector**
  : Let $X = (X_1, X_2)$ be a random vector. For given values x_1 and x_2,
  $f_X(x_1, x_2) =$ the probability (density) that $X_1 = x_1, X_2 = x_2$.

- **Expectation**
  :

$$E[X] := \begin{bmatrix} EX_1 \\ EX_2 \end{bmatrix}$$

, where

$$EX_1 := \int\int x_1 f(x_1, x_2) dx_1 dx_2$$

and so forth.

- Note that EX is not a random vector any more, but is a **constant**.

- **Mean, Variance, and Covariance**
  : For $X = (X_1, ..., X_k)'$,

$$EX := (EX_1, ..., EX_k)'$$

$$Var(X) := Cov(X_i, X_j) = E[(X - EX)(X - EX)'].$$

: For $X = (X_1, ..., X_k)'$ and $Y = (Y_1, ..., Y_q)'$,

$$Cov(X, Y) := Cov(X_i, Y_j) = E[(X - EX)(Y - EY)']$$

- **Properties of Mean and Variance**
  : Let X, Y, Z be random vectors, and A, B be non-random matrices with appropriate dimensions.

  - $E[X'] = (EX)'$ for a random matrix X.
  - $E[AX + BY] = AE[X] + BE[Y]$; i.e. $E[\cdot]$ is linear.
  - If $X \geq Y$, then $EX \geq EY$; i.e. $E[\cdot]$ is monotonic.
  - $Cov(Y, X) = Cov(X, Y)'$
  - $Cov(AX + BY, Z) = ACov(X, Z) + BCov(Y, Z)$
  - $Var(X + Y) = Var(X) + Cov(X, Y) + Cov(Y, X) + Var(Y)$
  - $Var(AX + B) = AVar(X)A'$
  - $Var(X)$ is symmertic and positive semidefinite.

- **Correlation Coefficient**
  : For two random variables X_1 and X_2,

$$\rho_{1,2} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}}$$

.

1.2 Conditional Distribution

In this section, let X, Z be random variables, a, b be (non-random) constants, and g be a function.

- **Conditional density and conditional expectation**

$$f_{X|Z}(x|z) := \frac{f_{X,Z}(x,z)}{f_Z(z)}$$

$$E[X|Z=z] := \int x f_{X|Z}(x,z)dx.$$

- Note that $E[X|Z=z]$ is a (non-random) $function$ of z, while $E[X|Z]$ is a $random\ variable$.

- **Properties of conditional expectation**
  - $E[aX+b|Z=z] = aE[X|Z=z]+b$
  - $E[g(X,Z)|Z=z] = E[g(X,z)|Z=z]$
  - $E[g(Z)X|Z=z] = g(z)E[X|Z=z]$
  - The Law of Iterated Expectations: $E[X] = E[E[X|Z]]$

2 Independence

- **Independence of two random variables**

$$X_1 \perp X_2 \;\text{ iff }\; f_{1,2}(x_1, x_2) = f_1(x_1) f_2(x_2) \;\text{ iff }\; f(x_1) = f(x_1|x_2),\ \forall x_1, x_2.$$

- **Consequences of Independence**

$$X_1 \perp X_2 \Rightarrow \begin{bmatrix} g(X_1) \perp h(X_2) \\ E[X_1 X_2] = E[X_1]E[X_2] \\ Var(X_1 + X_2) = Var(X_1) + Var(X_2) \end{bmatrix}$$

- **Independence, Mean Independence, Zero Correlation**

$$u_i \perp X_i \ \textit{and}\ E[u_i] = 0 \Rightarrow E[u_i|X_i] = 0 \Rightarrow Cov(u_i, X_i) = 0.$$

3 Moment Generating Function

- **Def.** Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The *moment generating function(m.g.f.)* of X is defined to be the function $M(t) = E(e^{tX})$, for $-h < t < h$.

- **Property.**
  - $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ iff $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$.

4 Important Inequalities

4.1 Markov's Inequality

: Let $u(X)$ be a nonnegative function of the random variable X. If $E[u(X)]$ exists, then for every positive constant c,

$$P[u(X) \geq c] \leq \frac{E[u(X)]}{c}.$$

4.2 Chebyshev's Inequality

: Let the random variable X have a distribution of probability about which we assume only that there is a finite variance σ^2. Then for every $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2},$$

or, equivalently,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

4.3 Jensen's Inequality

: If ϕ is convex on an open interval I and X is a random variable whose support is contained in I and has finite expectation, then

$$\phi[E(X)] \leq E[\phi(X)].$$

If ϕ is strictly convex, then the inequality is strict, unless X is a constant random variable.

4.4 Cauchy-Schwartz Inequality

$$(E(XY))^2 \leq (E(X^2))(E(Y^2)).$$

- If the *Cauchy-Schwartz Inequality* is applied to the two random variables $X - \mu x$ and $Y - \mu y$ centered around their means, then

$$cov(X,Y)^2 \leq var(X)var(Y).$$

5 Some Distribution Families

5.1 Normal Distribution

- **Probability Density Function**
  Assuming Σ is $(k \times k)$ non-singular matrix, $X \sim N(\mu, \Sigma)$ has pdf of

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^k}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)).$$

- **Representational Definition**

$$X \overset{d}{\equiv} \mu + AZ \quad \text{with} \quad AA' = \Sigma, \quad Z_1, ..., Z_k \sim iid\ N(0,1)$$

.

- **Properties**
  - $X \sim N(\mu, \Sigma) \Rightarrow AX + b \sim N(A\mu + b, A\Sigma A')$
  - If $X \sim N(\mu, \Sigma)$, then $AX \perp BX$ iff $Cov(AX, BX) = 0$.

5.2 Distributions Related to Normal

- **Chi-squared Distribution**

$$Y \overset{d}{\equiv} \sum_{i=1}^{k} Z_i^2 \text{ with } Z_i \sim iid\ N(0,1) \ \Rightarrow\ Y \sim \chi_k^2.$$

- **Students' t-distribution**

$$Y \overset{d}{\equiv} \frac{Z}{\sqrt{\frac{V}{k}}} \text{ with } Z \sim N(0,1),\ V \sim \chi_k^2,\ Z \perp V,\ Y \sim t_k$$

- **F-distribution**

$$y \overset{d}{\equiv} \frac{\frac{V}{k}}{\frac{W}{m}} \text{ with } V \sim \chi_k^2,\ W \sim \chi_m^2,\ V \perp W \Rightarrow Y \sim F_{k,m}$$

6 Inference and Hypothesis Testing

6.1 Inference

- **Population and Sample**
  An econometric model of *population*:

$$y_i = \alpha + \beta x_i + u_i, \qquad E[u_i|x_i] = 0.$$

An estimator based on *sample*:

$$\hat{\beta} := \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}.$$

- **Confidence Interval**
  If

$$Pr_\theta[l(X) \leq \theta \leq u(X)] = \alpha\%, \quad \forall \theta$$

, then we call (l, u) an confidence interval with confidence level $\alpha\%$, for the parameter θ.

- **Example**
  Suppose we have a model

$$X_i \sim N(\mu, 1^2), \quad X_i \text{ are independent.}$$

Q. Based on the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{16} X_i$, construct a 5% Confidence interval for μ.

6.2 Hypothesis Testing

- **Test Procedure**
  - For $\Theta = \Theta_0 \cup \Theta_1$, $H_0 : \theta \in \Theta_0$ is tested against $H_1 : \theta \in \Theta_1$.
  - Construct a test statistic $T(X)$ and a critical region C.
  - Compute T based on the sample you draw, and reject H_0 if $T \in C$.

- **P-value and Significance**
  Suppose we are dealing with regression model (with good assumptions)

$$y_i = \alpha + \beta x_i + u_i$$

. Here we intend to test $H_0 : \beta = 0$ v.s. $H_1 : \beta \neq 0$.
Now suppose that we know somehow that $\hat{\beta} \equiv^d Z \sim N(0,1)$ under $H_0(X \equiv^d Z$ means X has the same
Suppose further that we have already obtained b as the realized value of $\hat{\beta}$.
In this context, we define

$$\text{P-value} := Pr[|Z| \geq |b|]$$

, i.e., *the probability that the test statistic takes more extreme values than our currently realized one*. By saying "β is significantly different from zero (or alternatively, (the variable) x is statistically significant)", we mean

$$\text{P-value} \leq \text{significance level (typically 0.01, 0.05, or 0.10)}$$

, i.e., we can reject the $H_0 : \beta = 0$

7 Basic Concepts for Asymptotic Theory

7.1 Pointwise Convergence

- **Def.** Consider a sequence $\{Y_n : \ n \geq 1\}$. We say Y_n converges point wise to Y, or

$$\lim_{n\to\infty} Y_n = Y$$

if $\forall\ \epsilon > 0, \exists n_\epsilon \geq 1$, such that

$$|Y_n - Y| \leq \epsilon \ \ \forall \ \ n \geq n_\epsilon.$$

7.2 Convergence in Probability

- **Def.** Let $\{X_n\}$ be a sequence of random variables. We say that X_n *converges in probability* to X if for all $a > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0, \quad \text{or} \quad X_n \xrightarrow{P} X.$$

- Convergence in probability does not imply pointwise convergence.

- **Example**.
  Law of Large Number(LLN) Let $\{X_n\}$ be a sequence of *iid* random variables having common mean μ and variance $\sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then,

$$\bar{X}_n \xrightarrow{P} \mu.$$

- **Properties**
  - Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then, $X_n + Y_n \xrightarrow{P} X + Y$.

  - Suppose $X_n \xrightarrow{P} X$ and a is a constant. Then, $aX_n \xrightarrow{P} aX$.

- Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then, $X_n Y_n \xrightarrow{P} XY$.

- Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then, $X_n/Y_n \xrightarrow{P} X/Y$, where $Y \neq 0$.

- **Example** Suppose $X_n, \cdot, X_n$ are *iid* with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $s^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$. Then,
$$s^2 \xrightarrow{p} \sigma^2.$$

7.3 L^2 Convergence

- **Def.** We say Y_n converges in mean square to Y, or

$$Y_n \xrightarrow{L^2} Y$$

if

$$\lim_{n\to\infty} E(Y_n - Y)^2 = 0.$$

- Mean-squared(L^2) convergence implies convergence in probability.

- **Mean-Squared Error**

$$\begin{aligned}(\boldsymbol{MSE}) &= E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E\hat{\theta})^2 + (E\hat{\theta} - \theta)^2 \\ &= \text{Variance} + \text{Bias}^2.\end{aligned}$$

7.4 Convergence in Distribution

- **Def.** Let $X_1,\ X-2,\ \cdots$ be a sequence of random variables. We say X_n converges in distribution to a random variable X, or

$$X_n \xrightarrow{d} X$$

if

$$F_n(x) = P(X_n \leq x) \to F(x) = P(X \leq x) \ \text{ as } \ n \to \infty$$

$\forall x$ where $F(\cdot)$ is continuous at y.

- We can use the convergence in distribution result to approximate the distribution of Y_n by that of Y.

- **Example.** (**Central Limit Theorem: CLT**) Suppose $X_1,\ X_2,\ \cdots,\ X_n$ are *iid* random vector with $E \parallel X_i \parallel^2 < \infty$. Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \xrightarrow{d} N(0, \Sigma) \ \text{ as } \ n \to \infty$$

, where $\Sigma = E(X_i - EX_i)(X_i - EX_i)'$.

- **Properties.**
  - (**Slutsky's Theorem**) Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$
    , where c is a constant. Then,

$$
\begin{aligned}
(a) &\quad X_n + Y_n \xrightarrow{d} X + c \\
(b) &\quad Y_n X_n \xrightarrow{d} cX \\
(c) &\quad X_n / Y_n \xrightarrow{d} X/c, \quad \text{provided} \quad c \neq 0.
\end{aligned}
$$

  - $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y \nRightarrow X_n + Y_n \xrightarrow{d} X + Y$.
  - $X_n \xrightarrow{d} c \Leftrightarrow Y_n \xrightarrow{p} c$, where c is a constant.
  - (**Continuous Mapping Theorem**) Suppose $Y_n \xrightarrow{d} Y$. Then, $g(Y_n) \xrightarrow{d} g(Y)$, where $g(\cdot)$ is a continuous function.

7.5 Delta(Δ)-Method

- **Def.** Let $\{X_n\}$ be a sequence of random variables such that

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2).$$

Suppose the function $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$. Then,

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2(g'(\theta))^2).$$