Econ 312: Introduction to Econometrics

Review of Probability II

Sang-Yeob Lee

February 5, 2009

Two random variables

Why do we care?

- The most of the interesting questions in economics involve two or more variables.
  - How does the distribution of income for women compare to that for men?
  - Are college graduates more likely to have a job than nongraduates?
- Answering such questions require an understanding of the concepts of joint, marginal, and conditional probability distributions.

Joint and Marginal Distribution

- Joint probability distribution: $\Pr(X = x, Y = y)$
- probability that the random variables simultaneously take on certain values,say x and$y$
- Marginal probability distribution (another name for its probability distribution)
- If Y takes on k different values, then marginal probability that X takes on the values x is

$$\Pr(X = x) = \sum_{i=1}^{k} (X = x, Y = y_i)$$

Conditional Distribution

- The conditional probability that X takes on the value x when Y takes on the values y is $\Pr(X = x|Y = y)$
- In general,

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

Conditional Expectation

- Conditional Expectation (or Conditional mean) of Y given X is the mean of of the conditional distribution of Y given X.
- If Y takes on k values, then the conditional mean of Y given $X = x$ is

$$E(Y|X = x) = \sum_{i=1}^{k} y_i \Pr(Y|X = x).$$

- Suppose that the expected value of **WAGE** given **EDUC** is the linear function

$$E(WAGE|EDUC) = 1.05 + 0.45 EDUC.$$

  - If this relationship holds, the average wage for people with 16 years of education is $8.25.

Conditional Variance

- The variance of Y conditional on X is the variance of conditional distribution of Y given X.
-
$$\text{Var}(Y|X=x) = \sum_{i=1}^{k} [y_i - E(Y|X=x)]^2 \Pr(Y=y_i|X=x).$$

Independence

- Two random variable X and$Y$ are independently distributed or independent if knowing value of one of the variables provides no information about the other.
- X and Y are independent if, for all values of x and y, $\Pr(Y = y|X = x) = \Pr(Y = y)$
- If X and Y are independent, $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$

Covariance

a measure of the linear relationship between two random variables.

- One measure of the extent to which two random variable move together is their covariance.
- $Cov(X, Y) = E(X - \mu_x)(Y - \mu_y)$
- $Cov(X, y) = \sigma_{XY} = E(XY) - \mu_x \mu_y$
- If X takes on l values and Y take on k values, then the covariance is given by the formula $Cov(X, Y) = \sum_{j=1}^{l} \sum_{i=1}^{k} (x_j - \mu_x)(y_i - \mu_y) \Pr(X = x_i, Y = y_i)$
- If X and Y are independent, $Cov(X, Y) = 0$. (but, not vice versa!)

Correlation Coefficient

is defined in terms of the covariance.

- The correlation does not depend on units of measurement.
- $$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \rho$$
- $-1 \leq Cov(X,Y) \leq 1$ (correlation inequality)
  - $Corr(X,Y) = 1$ means perfect positive linear relationship
  - $Corr(X,Y) = -1$ means perfect negative linear relationship
  - $Corr(X,Y) = 0$ means no linear relationship.

Means, Variance, and Covariances of Sums of Random Variables

- Let X,Y, and Z be random variable, and let a,b,c be constant. Then,
  1. $E(a + bX + cY) = a + b\mu_x + c\mu_y$
  2. $Var(aX + BY) = a^2Var(X) + 2abCov(X, Y) + b^2Var(Y)$
  3. $Cov(a + bX + cZ, Y) = bCov(X, Y) + cCov(Z, Y)$
  4. $E(XY) = Cov(X, Y) - E(X)E(Y)$
  5. $|Corr(X, Y)| \leq 1$ and $|Corr(X, Y)| \leq \sqrt{\sigma_X^2 \sigma_Y^2}$ (correlation inequality)