Graduate Econometrics (MSc.)
Review of Probability

Yonas Alem

Department of Economics
University of Gothenburg

November 4, 2014

- Introduction (Ch.1 & Others)
- Review of Probability (Ch.2)
- Linear Regression with One Regressor (Ch.4-7, 17,18)
- Linear Regressions with Matrix Algebra (Ch.4-7, 17,18)
- The Classical Linear Regression Model (Ch.4-7, 17,18)
- Heteroskedasticity - Robust Standard Errors (Ch.4-7, 17,18)

- Multicollinearity - Implications and Solutions (Ch.4-7, 17,18)
- Hypothesis Testing (Ch.4-7, 17,18)
- Large Sample Properties of OLS (Ch.4-7, 17,18)
- IV Estimation (Ch.12)
- Regression with a Binary Dependent Variable (Ch. 11)
- Regression with Panel Data (Ch.10)
- Experiments and Quasi-Experiments (Optional) (Ch.13)

- Five labs: four are graded, 10 points each = 40 points
- Final exam: Friday, Jan 16, 08:00 - 13:00 = 60 points
- Office Hours: Monday, Tuesday, 16:00-17:00 & by appointment

Economics Questions and Econometrics

- **Econometrics** is a the science and art of using economic theory and statistical techniques to analyze economic data
  - Used in many branches of economics: e.g., microeconomics, macroeconomics, labor economics, fiance, marketing etc..
- This course introduces you the key tools used by econometricians to answer a number of interesting research questions
- Examples of economic questions:
  1. Does reducing class size improve elementary school education?
  2. Is there racial discrimination in the market for home loans?
  3. How much do cigarette taxes reduce smoking?
  4. What will the rate of inflation be next year?
  5. Do international remittances reduce poverty in developing countries?

Economics Questions and Econometrics cont.

- The first three questions concern causal relationships among variables
  - Causality: a specific action leads to a specific measurable consequence
- How does one estimate causal effects?
  - Randomized controlled experiments
    - A setting in which individuals or groups of individuals are randomly assigned to the treatment and comparison groups
- In this course, the **causal effect** is defined to be the effect of an outcome of a given action or treatment as measured in an ideal randomized controlled experiment.
  - The only systematic reason for differences in outcomes between the treatment and control groups is the treatment itself
- How does one answer the above research questions using an RCT?

Estimation of Causal Effects

- The concept of an ideal RCT allows one to make a causal effect
- However, experiments are rare in econometrics because often they are unethical, impossible to execute satisfactorily, or prohibitively expensive
- The concept of RCT does however, provide a theoretical benchmark for an econometric analysis of causal effects using actual data

Forecasting and Causality

- Q.4, i.e., forecasting inflation however does not concern causal effects
- One does not need to know causal relationship to make a good forecast
- A good way to "forecast" if it is raining is to observe whether pedestrians are using umbrellas, but the act of using an umbrella does not cause it to rain
- Historical relationship suggested by economic theory can be used to check relationships between variables has been stable over time, to make quantitative forecasts about he future and to assess the accuracy of those forecasts

Data Sources and Types

1. Experimental Data
   - Come from experiments designed to evaluate a treatment or policy or to investigate a causal effect
   - E.g., the Tennessee class size experiment in the 1980's
     - Thousands of students were randomly assigned to classes of different sizes for several years and were given annual standard tests
     - Costed millions of dollars and required ongoing cooperation of many administrators, parents, and teachers over several years
   - Experiments are difficult to implement on human subjects $\Longrightarrow$ have some flaws
   - Can be unethical
   - As a result, economists mainly rely on observational data
2. Observational Data: data obtained by observing actual behavior outside an experimental setting

Data Sources and Types

- Observational Data Types:
  1. Cross-Sectional Data: data on different entities - workers, consumers, firms, governmental units etc..
  2. Time Series Data: data for a single entity (person, firm, country) collected at multiple time periods
  3. Panel (Longitudinal) Data: data for multiple entities in which each entity is observed at two or more time periods

- **Probabilities:** an event which has an element of chance or randomness (e.g., the number of times your computer will crash while you are writing your term paper)
- **Outcomes:** the mutually exclusive potential results of a random process (e.g., your computer might never crash, it might crash once, twice and so on
  - The outcomes are mutually exclusive
  - The outcomes need not be equally likely
- The **probability** of an outcome is the proportion of the time that the outcome occurs in the long run
- The **sample space:** the set of all possible outcomes
- An **event**: is a subset of the sample space (a set of one or more outcomes)
  - The event "my computer will crash no more than once" is the set consisting of two outcomes: "no crashes" and "one crash"

- A **random variable**: a numerical summary of a random outcome
  - The number of times your computer crashes while you are writing a term paper is random and takes on a numerical value $\Longrightarrow$ it is a random variable
- A **discrete random variable:** takes on only a discrete set of values, like 0, 1, 2,...,
- A **continuous random variable:** takes a continuum of possible values
- The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur (these probabilities sum to 1)
- See table 2.1 and fig 2.1

- INSERT TABLE AND FIGURE 2.1 HERE

- **Probabilities of events:** can be computed from the probability distribution
  - The probability of the the event of one or two crashes is the sum of the probabilities of the constituent outcomes
    - $Pr(M = 1 or M = 2) = Pr(M = 1) + Pr(M = 2) = 0.10 + 0.06 = 0.16 or 16\%$
- **Cumulative probability distribution:** the probability that the random variable is less than or equal to a particular value
  - The probability at most one crash $Pr(M \leq 1)$ is 90%
- A cumulative probability distribution is also referred to as a **cumulative distribution function (a cdf)** or a **cumulative distribution**

- The **Bernoulli distribution**: when the random variable is binary (the outcomes are 0 or 1)
  - A binary random variable is called a **Bernoulli random variable** in honor of the seventeenth-century Swiss mathematician and scientist Jacob Bernoulli), and its probability distribution is called the **Bernoulli distribution**
- Let G be the gender of the next new person you meet: $G = 0$ indicates that the person is male and $G = 1$ indicates that she is female
- The outcomes of G and their probabilities thus are:

$$G = \begin{cases} 1 & \text{with probability p} \\ 0, & \text{with probability 1-p,} \end{cases} \tag{1}$$

- The probability distribution in (1) is called the Bernoulli distribution

A Continuous Random Variable

- **Cumulative probability distribution:** defined just as it is for a discrete random variable, i.e., the probability that the continuous random variable is less than or equal to a particular value
- See fig 2.2 plotting commuting time of a student
- **Probability density function:** summarizes the probability of a continuous variable
  - Also called **p.d.f**, a **density function**, or simply a **density**
- The probability density function and the cumulative probability distribution show the same information in different formats

A Continuous Random Variable

- INSERT FIG 2.2A & B here

The Expected Value of a Random Variable

- The Expected Value of a random variable Y, denoted $E(Y)$ is the long-run average value of the random variable over many repeated trials or occurrences
  - Computed as the weighted average of the possible outcomes of that random variable, where the weights are the probabilities of that outcome
  - Also called the **expectation** of Y and is denoted μ_Y
- Suppose the random variable Y takes on k possible values, $y_1, ..., y_k$ and the probability that Y takes on y_1 is p_1, the probability that Y takes on y_2 is p_2, and so forth
- The expected value of Y, denoted $E(Y)$ is

$$E(Y) = y_1 p_1 + y_2 p_2 + ... + y_k p_k = \sum_{i=1}^{k} y_i p_i \qquad (2)$$

The Expected Value of a Random Variable cont.

- **Expected value of a Bernoulli random variable:** Let G be the Bernoulli random variable with the probability distribution in Eq. (1). The expected value of G is

$$E(G) = 1Xp + 0X(1-p) = p \tag{3}$$

- The expected value of a Bernoulli random variable is p, the probability that it takes on the value "1"
- **Expected value of a continuous random variable:** is the probability-weighted average of the possible outcomes of the random variable

The Standard Deviation and Variance

- Both measure the dispersion or the “spread” of a probability distribution
- The variance of the discrete random variable Y, denoted σ_Y^2, is

$$\sigma_Y^2 = var(Y) = E[(Y - \mu_Y)^2] = \sum_{i=1}^{k} (y_i - \mu_Y)^2 p_i \qquad (4)$$

- Because the variance involves the square of Y; the units of the variance are the units of the square of Y (makes its interpretation difficult)
- The **standard deviation** is the square root of the variance and is denoted σ_Y and has the same units as Y

The Standard Deviation and Variance Cont.

- **Variance of a Bernoulli random variable**: we saw that the mean of the Bernoulli variable $\mu_G = p$. So the variance is

$$var(G) = \sigma_G^2 = (0-p)^2 \times (1-p) + (1-p)^2 \times p = p(1-p) \quad (5)$$

- $\Longrightarrow$ the **standard deviation (SD)** of a Bernoulli random variable is $\sigma_G = \sqrt{p(1-P})$

A Linear Function

- Consider:

$$Y = 2000 + 0.8X \tag{6}$$

- Where Y = after-tax earnings, X = pre-tax earning
- Suppose an individual's pre-tax earnings next year are a random variable with mean μ_X and variance σ_X^2
- Pre-tax earnings are random $\Longrightarrow$ after-tax earnings are random
- The expected value of after-tax earnings is

$$E(Y) = \mu_Y = 2000 + 0.8\mu_X \tag{7}$$

- The variance of after-tax earnings: $E[(Y - \mu_Y)^2] = 0.64E[(X - \mu_X)^2]$
- I.e., $var(Y) = 0.64var(X)$. Can you show this?

A Linear Function Cont.

- The SD of Y is therefore:

$$\sigma_Y = 0.8\sigma_X \tag{8}$$

- The SD of the distribution of after-tax earnings is 80% of the SD of the distribution of pre-tax earnings
- Generalizing the analysis: let Y depend on X with an intercept a and slope b, so that:

$$Y = a + bX \tag{9}$$

- Then the mean, variance and SD are:

$$\mu_Y = a + b\mu_X \tag{10}$$

$$\sigma_Y^2 = b^2\sigma_X^2 \quad \textit{and} \quad \sigma_Y = b\sigma_X \tag{11}$$

- The **skewness** of the distribution of a random variable Y, which measures how much a distribution deviates from symmetry is given by

$$\frac{E[(Y-\mu_Y)^3]}{\sigma_Y^3} \tag{12}$$

- Skewness $= 0, -ve \;\; and \;\; +ve$ for a symmetric, long left-tailed, and long right-tailed distributions respectively
- The **kurtosis** of a distribution is a measure of how much of the variance of Y arises from extreme values **(outliers)**. Is measured by

$$\frac{E[(Y-\mu_Y)^4]}{\sigma_Y^4} \tag{13}$$

- $(Y-\mu_Y)^4$ cannot be negative, $\implies$ the kurtosis cannot be negative

- The kurtosis of a normally distributed random variable is 3
- A distribution with $kurtosis > 3$ is called **leptokurtic** (or heavy-tailed)
- Like skewness, the kurtosis is unit free
- **Moments**:
  - The mean of Y, $E(Y)$ is also called the first moment of Y
  - $E(Y^2)$ is the second moment of Y
  - In general, the expected value of Y^r is called the r^{th} **moment** of the random variable Y

Joint Probability Distributions

- **Joint probability distribution** of two discrete random variable say X and Y: the probability that the random variables simultaneously take on certain values, say x and y. The probabilities of all possible (x, y) combinations sum to 1.
  - The joint probability distribution can be written as the function $Pr(X = x, Y = y)$
- E.g., weather conditions (whether or not it is raining) affect commuting time of a student. Let Y be a binary random variable that equals 1 if the commute is short (less than 20 minutes) and equals 0 otherwise and let X be a binary random variable that equals 0 if it is raining 1 if not.
  - Four possible joint outcomes: $(X = 0, Y = 0), (X = 0, Y = 1), (X = 1, Y = 0)$, and $(X = 1, Y = 1)$
- The joint probability distribution is the frequency with which each of these four outcomes occurs over many repeated

Joint Probability Distributions Cont.

- Here is a tabular representation

Table: 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain(X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- What are the values of $Pr(X = 0, Y = 0), Pr(X = 0, Y = 1), Pr(X = 1, Y = 0)$, and $Pr(X = 1, Y = 1)$?

Marginal Probability Distribution

- The **marginal probability** distribution of a random variable Y is just another name for its probability distributions
- It shows the distributions of Y alone (the marginal distribution) from the joint distribution of Y and another random variable
- Can be computed from the joint distribution of X and Y by adding up the probabilities of all possible outcomes for which Y takes on a specified value
- If X can take on l different values $x_1, ..., x_l$ then the marginal probability that Y takes on the value y is

$$Pr(Y = y) = \sum_{i=1}^{l} Pr(X = x_i, Y = y) \tag{14}$$

- Given in the final column and row of table 2.2 respectively.

Conditional Distributions

- The distribution of a random variable Y conditional on another random variable X taking on a specific value is called the **conditional distribution** of Y given X.
- The conditional probability that Y takes on the value y when X takes on the value x is written $Pr(Y = y|X = x)$

$$Pr(Y = y|X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)} \tag{15}$$

Conditional Expectation

- The **conditional expectation** of Y given X, also called the **conditional mean** of Y given X, is the mean of the conditional distribution of Y given X.
- It is therefore the expected value of Y, computed using the conditional distribution of Y given X
- If Y takes on k values $y_1, ..., y_k$, then the conditional mean of Y given X=x is

$$E(Y|X=x) = \sum_{i=1}^{k} y_i Pr(Y = y_i|X = x) \quad (16)$$

The Law of Iterated Expectations

- The mean of Y is the weighted average of the conditional expectation of Y given X, weighted by the probability distribution of X
  - For e.g., the mean height of adults is the weighted average of the mean height of men and the mean height of women, weighted by the proportions of men and women.
- Mathematically, if X takes on the l values $x_1, ..., x_l$, then

$$E(Y) = \sum_{i=1}^{l} E(Y|X = x_i)Pr(X = x_i) \qquad (17)$$

- Eq(17) follows from Equations (16) and (15)
- In other words, the expectation of Y is the expectation of the conditional expectation of Y given X

$$E(Y) = E[E(Y)|X)] \qquad (18)$$

The Law of Iterated Expectations cont.

- Eq (18) is knows as the **law of iterated expectations**
- **Conditional variance of Y given X** is the variance of Y conditional on X which is the variance of the conditional distribution of Y given X.

$$var(Y|X = x) = \sum_{i=1}^{k}[y_i - E(Y|X = x)]^2 Pr(Y = y_i|X = x) \tag{19}$$

Independence

- Two random variables X adn Y are **independently distributed** or simply, **independent**, if knowing the value of one of the variables provides no information about the other

$$Pr(Y = y|X = x) = Pr(Y = y) \quad (independence \;\; of \;\; X \;\; and \;\; Y) \tag{20}$$

- If X and Y are independent, then

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y) \tag{21}$$

- I.e., the joint distribution of two independent variables is the product of their marginal distributions

Covariance

- **Covariance**: of two random variables X and Y denoted, $cov(X,Y)$ or σ_{XY} measures the extent to which two random variables move together
- The $cov(X,Y)$ is the expected value $E[(X-\mu_X)(Y-\mu_Y)]$
- If X can take on l values and Y can take on k values, then the covariance is given by the formula

$$cov(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] \quad (22)$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j-\mu_X)(y_i-\mu_Y)Pr(X=x_j, Y=y_i) \quad (23)$$

- Could be $+$ve (positive relationship), -ve (negative relationship) or 0 (no relationship)

Correlation

- **Correlation**: the covariance between X and Y divided by their standard deviations

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{24}$$

- The covariance is the product of X and Y, deviated from their means, its units are awkwardly, the units of X multiplied by the Units of Y. This "units" problem can make numerical values of the covariance difficult to interpret
- the correlation is a better measure in that respect, because the units of the numerator and the denominator are the same in the $corr(X,Y)$ above

$$-1 \leq corr(X,Y) \leq 1 \tag{25}$$

Correlation cont.

- **Correlation and conditional mean:** If the conditional mean of Y does not depend on X, then Y and X are uncorrelated. I.e.,

$$if \;\; E(Y|X) = \mu_Y, \;\; then \;\; cov(Y,X) = 0 \;\; and \;\; corr(Y,X) = 0 \tag{26}$$

- The mean of the sum of two random variables, X and Y, is the sum of their means

$$E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y \tag{27}$$

The Normal Distribution

- The probability density function of a normally distributed random variable (the **normal p.d.f**):

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{y-\mu}{\sigma})^2] \quad (28)$$

- Has the familiar bell-shaped probability density
- Symmetric around its mean and has 95% of its probability between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$
- The normal distribution with mean μ and variance σ^2 is expressed concisely as "$N(\mu, \sigma^2)$"
- $N(0, 1)$ is known as the **standard normal distribution**

The Normal Distribution cont.

- Random variables that have a $N(0,1)$ are often denoted Z
- The standard normal cumulative distribution function is denoted by the Greek letter Φ; accordingly, $Pr(Z \leq c) = \Phi(c)$ where c is a constant
- To compute the probabilities for a normal variable with a general mean and variance, one must **standardize the variable** by first subtracting the mean, then by dividing the result by the standard deviation
- E.g., suppose $Y \sim N(1,4)$, what is the probability that $Y \leq 2$?
- The standard version of Y is Y minus its mean, divided by its standard deviation, that is, $(Y-1)/(\sqrt{4} = \frac{1}{2}(Y-1)$

The Normal Distribution cont.

- Accordingly the random variable $\frac{1}{2}(Y-1)$ is normally distributed with mean zero and variance one and has the standard normal distribution
- Now, $Y \leq 2 \equiv \frac{1}{2}(Y-1) \leq \frac{1}{2}(2-1)$ i.e., $\frac{1}{2}(Y-1) \leq \frac{1}{2}$. Thus,

$$Pr(Y \leq 2) = Pr[\frac{1}{2}(Y-1) \leq \frac{1}{2}] = Pr(Z \leq \frac{1}{2}) = \Phi(0.5) = 0.691 \tag{29}$$

- Where the value 0.691 is taken from the table of $\Phi(Z)$
- The normal distribution is symmetric, so its skewness is zero, and its kurtosis is 3
- The **multivariate normal distribution**: used to describe the joint distribution of a set of random variables
- The **bivariate normal distribution**: the joint distribution of

The Chi-Squared Distribution

- The **chi-squared distribution**: the distribution of the sum of m squared independent standard normal random variables
- the distribution depends on m which is called the degrees of freedom of the chi-squared distribution
- E.g., let Z_1, Z_2 and Z_3 be independent standard normal random variables. Then Z_1^2, Z_2^2 and Z_3^2 has a chi-squared distribution with 3 degrees of freedom
- A chi-squared distribution with m degrees of freedom is denoted χ_m^2

The student t Distribution

- The **Student t distribution** with m degrees of freedom is defined to be the distribution of the ratio of a standard normal random variable, divided by the square root of an independently distributed chi-squared random variable with m degrees of freedom divided by m
- I.e., let Z be a standard normal random variable, let W be a random variable with a chi-squared distribution with m degrees of freedom, and let Z and W be independently distributed
  - Then, the random variable $Z/\sqrt{W/m}$ has a student t distribution (also called the **t distribution**) with m degrees of freedom
  - Selected percentiles of the Student t distribution are given in the Appendix

The student t Distribution

- The F distribution with m and n degrees of freedom, denoted $F_{m,n}$ is defined to be the distribution of the ratio of a chi-squared random variable with degrees of freedom m, divided by m, to an independently distributed chi-squared random variable with degrees of freedom n, divided by n.
- The 90th, 95th, and 99th percentiles of the $F_{m,n}$ distribution are given in the appendix