ECO585: Applied Econometrics
Probability Referesher
Lecture 1

Prof. Mary Kaltenberg

Pace University

January 28, 2021

Overview

The goal of this course is to provide a solid understanding of OLS and expand this knowledge into other frequently applied econometric techniques.
We will cover:

- Properties of OLS (finite sample theory)
- Estimation, Inference, and Interpretation of OLS parameters (ideally, this is review)
- Asymptotics (large sample theory)
- Generalized Method of Moments (two staged least squares/Instrumental Variables)
- Panel Data Methods: Difference and Difference, POLS, Random and Fixed Effects
- Logit/Probit (If we have time)

Overview

Key concepts that we will learn are:

- Biasedness
- Efficiency
- Consistency
- Endogeneity/Exogeneity
- Convergence
- Identification
- Heteroskedasticity

Goal: Think through these concepts and how it applies to your particular data and/or estimation technique. If you can identify the problem, you can fix it more easily, or at the very least, theorize the limitations (or limits of limitations) in your quantitative research.

Random Variable

A random variable, X (usually denoted in capital letters), makes it possible for us to quantify the outcomes of a random process.

Random variable can be either **Discrete** or **Continuous**

- **Discrete** is when you can list the numbers, like the outcome of a 4-faced dice
- **Continuous** is when there are infinite possible numbers, such as estimating the number of stars.

Random Variable

Why use Random Variables?

We want to find out implications of random processes and . . .

- we don't assign a specific value to them,

- so that a variable can take on many values allowing us to discuss probability

- and using mathematical notation on the outcome can simplify our expressions.

Example

If we roll 10 four-faced dice, what is the probability that the sum will be equal or greater than 25 can be simplified to:

$$P(X \geq 25) \tag{1}$$

Discrete Random Variables

Bernoulli or binary random variable is a type of discrete random variable.
It can take on either a 1 or 0 and has interesting properties.
An example of a binary random variable is:

$$P(X = 1) = \theta \quad (2)$$

$$P(X = 0) = 1 - \theta \quad (3)$$

We are often trying to estimate θ. Since the sum of the probability of any events will always be equal to 1, we can think of the probability an event taking place to be θ and the event not taking place to be $1 - \theta$.

Probability Density Function

We often look at what is the probability for any given set of outcomes, this is essentially what a **probability density function** (pdf) does.

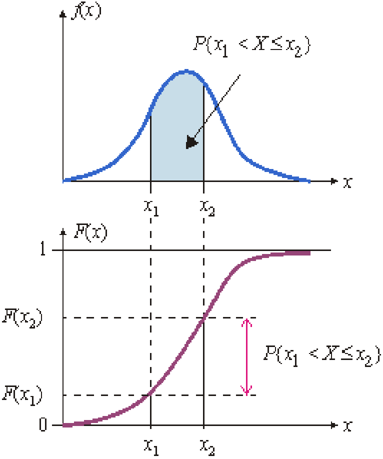
$$f(x_j) = p_j, j = 1, 2, \dots k \tag{4}$$

For any real number x, f(x) is the probability that the random variable X takes on the particular value x.
For a continuous variable, the pdf is used for events that involve a range of values. Whereas the binary variable is in the probability an outcome will occur.

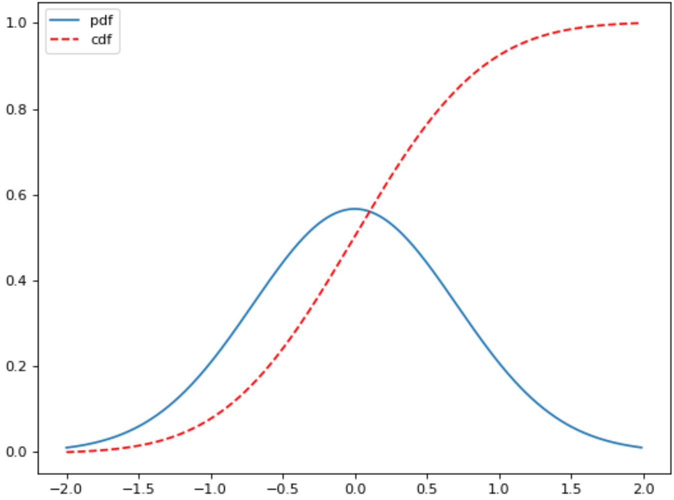
Cumulative Density Function

Cumulative density functions (cdf) is the probability that that an event will take a value less than or equal to some value, x.

In a continuous variable, a cdf is the area under the pdf between two points. Simply, it's the integral to the pdf.

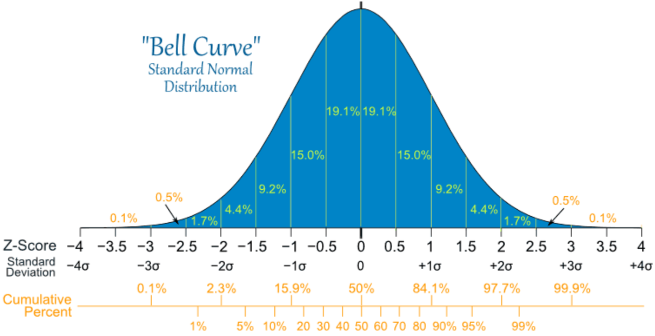


Probability Density Function



Normal Distribution

There are many types of distributions, but the most well known is the normal distribution.



- is symmetrical
- has no outliers
- mean $=$ median
- requires that the total area under the curve $= 1$

Normal Distribution

The Normal Distribution pdf is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp[-(x-\mu)^2/2\sigma^2], -\infty < x < \infty \tag{5}$$

where $\mu = E[X]$ and $\sigma^2 = Var(X)$

Sometimes this is called a Gaussian distribution

Joint & Conditional Distributions

joint distribution is when we calculate the probability of two random variable occurring at the same time.

$$f_{X,Y}(x, y) = P(X = x, Y = y) \tag{6}$$

An important consideration is whether X and Y are **independent**. Independence between two random variables means that their outcomes do not depends on one another.
If the two variables are independent, a nice feature is that their joint pdf is the product of the individual pdfs

Joint & Conditional Distributions

conditional distributions is when we calculate the probability of two dependent random variables (or more). In other words, we want to know how X may affect Y. We can summarize this as the conditional probability density function.

$$f_{Y|X}(y|x) = f_{x,y}(x, y)/f_x(x) \tag{7}$$

We'll often see it like this:

$$f_{Y|X}(y|x) = P(Y = y|X = x) \tag{8}$$

Expected Value

Take a random variable X, the **expected value** of X (E(X) is the weighted average of all possible values of X.
This is basically what econometrics is all about. Expected values.
Sometimes it's called a population mean, especially if we are relating it to represent some specific population.
The expected value of X is the weighted average:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) \cdots + x_1 k(x_k) = \sum_{j=1}^{k} x_j f(x_j) \qquad (9)$$

Expected Value

There are some properties in expected values that are useful to know (especially when we are deriving estimators, etc.)
Property 1 For any constant c, $E(c) = c$
Property 2 For any constant a and b, $E(aX + b) = aE(X) + b$
Property 3 If $a_1, a_2 ... a_n$ are constant and $X_1, X_2 ..., X_3$ are random variables then
$E(a_1X_1 + a_2X_2 + ... + a_nX_n) = aE(x_1) + a_2E(X_2) + ...a_nE(X_n)$

Expected Values

Since we work with random values, we know that the actual value can vary and we can never know the exact answer.

However, we can estimate what the value should be because of probability. We know that there is a 50/50 chance to get tails when I flip a coin. The law of large numbers shows us that if we repeat this experiment enough times, the value we get should be .5.

Thus, the **Expected value** of the probability to get tails is .5.

Expected Values

Similarly, we have an expected value that the mean sampling distribution of a random value to be approximately equal to its population mean.

$$E[X] = \mu \tag{10}$$

1. When a statistic meets this property it is an **unbiased estimator**.
2. It is always better to use an unbiased estimator.

Remember: that statistics are never 'exact', they are approximate

Variance & Standard Deviation

Variance is the expected distance to its mean

$$Var(X) = E(X - \mu)^2 \tag{11}$$

Variance & Standard Deviation

Standard deviation is simply the positive square root of the variation. This makes it way easier to work with.

$$sd(X) = \sqrt{Var(X)} \tag{12}$$

It has some properties:

- for any constant c, $sd(c) = 0$
- $sd(aX + b) = |a| sd(X)$

Standardizing a Random Variable

$$Z = \frac{X - \mu}{\sigma} \tag{13}$$

Where the mean is 0 and the standard deviation is 1

Standardizing a Random Variable: Normal Distribution

Connecting this statistic to our normal distribution, we can see a special case called the standard normal distribution when the mean is 0 and the standard deviation is 1.

You'll see this often as:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp(-z^2/2), -\infty < z < \infty \tag{14}$$

Higher Order Moments

Technically, each of these statistics is a 'moment' . The moments are:

1. mean
2. variance
3. skewness
4. kurtosis

skewness is defined as:

$$E(X^3) = E[(X - \mu)^3]/\sigma^3 \tag{15}$$

If skewness is around 0, it means it's perfectly symmetrical.
Kurtosis is defined as:

$$E(X^3) = E[(X - \mu)^4]/\sigma^4 \tag{16}$$

Larger value suggest the distribution has 'fatter' tails. It also describes how 'peaked' a distribution is.

Measures of Association: Covariance

Sometimes, it's useful to characterize the relationship between any two random variables: covariance and correlation do this job well.

Covariance is defined as:

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_Y)] \tag{17}$$

Covariance measures the amount of *linear* dependence between two variables. Positive values mean the random variables move together, negative values mean they move in opposite directions.

Measures of Association: Covariance

There are some useful properties of covariance:

1. If X and Y are **independent**, then $Cov(X, Y) = 0$
   (the opposite is not true)
2. $Cov(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 Cov(X, Y)$
   covariance can be altered by multiplying one or both of the random variables by a constant
   This basically implies that measurement units matters‘
3. $|Cov(X, Y)| \leq sd(X)sd(Y)$
   The absolute value is bounded by the product of their standard deviations

Measures of Association: Correlation

Given Cov property 2, a measure that doesn't take into account units is desirable, and why the correlation coefficient is used so frequently.
correlation coefficient is defined as:

$$Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)\cdot sd(Y)} = \frac{\sigma_{XY}}{\sigma_x \sigma_y} \tag{18}$$

This also has some useful properties:

1. $-1 \leq Corr(X,Y) \leq 1$
   if correlation is 0, there is no relationship (but, this does **not** imply independence)
2. $Corr(a_1 X + b_1, a_2 Y + b_2) = Corr(X,Y)$ if $a_1 a_2 > 0$
   What's nice, is that the correlation is invariant to measurement units

Properties of Variance

Property:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y) \quad (19)$$

It follows that if Cov(X,Y) = 0, then:
$Var(X + Y) = Var(X) + Var(Y)$
$Var(X + Y) = Var(X) - Var(Y)$

This is called **pairwise uncorrelated random variables**
If $\{X_1, \dots X_n\}$ are pairwise uncorrelated random variables and a_i are constants then,
Property:

$$Var(a_1 X_1 \dots X_1) = a_i^2 Var(X_1) + \cdots + a_n^2 Var(X_i) \quad (20)$$

Conditional Expectation

This is *very* important for econometrics - basically it's the premise for everything we will be doing.
We may want to summarize the conditional probability density function (ie we want to know the cdf of Y given X).
Suppose X has taken some value, x, then, we can compute the expected value of Y, given that we know what the outcome of X is.

This is basically:

$$E(Y|X = x) \tag{21}$$

Sometimes, you'll see it called, the **conditional expectation function**. For a continuous variable it's this:

$$E[Y|X = x] = \int t f_y(t|X = x)\, dt \tag{22}$$

It is the weighted average of possible values of Y. It tells us how the expected value of Y varies with x.

Properties of Conditional Expectation

Property 1: For any function c(X):

$$E[c(X)|X] = c(X) \tag{23}$$

This means that functions of X behaves constant when we calculate conditional expectations conditional on X. For example, if we know X, then we also know X^2

Property 2: For functions of a(X) and b(X):

$$E[a(X)Y + b(X)|X] = a(X)E(Y|X) + b(X) \tag{24}$$

Property 3: If X and Y are independent, then:

$$E(Y|X) = E(Y) \tag{25}$$

If they are independent, the expected value of Y given X does not depend on X. Therefore, $E(Y|X)$ always equals the unconditional expected value of Y.

Properties of Conditional Expectation

The next two properties describe the **Law of Iterated Expectations**, which is that the expected value of $\mu(X)$ is equal to the expected value of Y

Property 4:

$$E[E(Y|X)] = E(Y) \tag{26}$$

We can find the expected value of Y conditional in X if we first find $E(Y|X,Z)$ for any other random variable Z and then find the expected value of $E(Y|X,Z)$ conditional on X.

Property 4':

$$E(Y|X) = E[E(Y|X,Z)|X] \tag{27}$$

Properties of Conditional Expectation

Property 5:
if $E(Y|X) = E(Y)$, then $Cov(X, Y) = 0$

If knowledge of X does not change the expected value of Y, then X and Y *must* be uncorrelated (similarly, if they are correlated, then $E(Y|X)$ must be correlated.)

Property 6:
If $E(Y^2) < \infty$ and $E[g(X)^2] < \infty$ for some function g, then
$E\{[Y - \mu(X)^2|X\} \leq E\{[Y - g(X)]^2|X\}$ and
$E\{[Y - \mu(X)]^2\} \leq E\{[Y - g(X)]2^{\}}$

This is useful for prediction/forecasting. The first inequality says that if we measure prediction inaccuracy as the expected squared prediction error, conditional on X, then the conditional mean is better than any other function of X for predicting Y.

Conditional Variance

Conditional Variance is very useful for econometrics. Ittells us how much variance is left if we use $E(Y|X)$ to "predict" Y. It is defined by:

$$E(Y|X = x) = E(Y^2|x) - [E(Y|x)]^2 \tag{28}$$

One useful property of it is that the distribution if Y given X does not depend on X:
If X and Y are **independent**, then $Var(Y|X) = Var(Y)$

Chi-Square Distribution

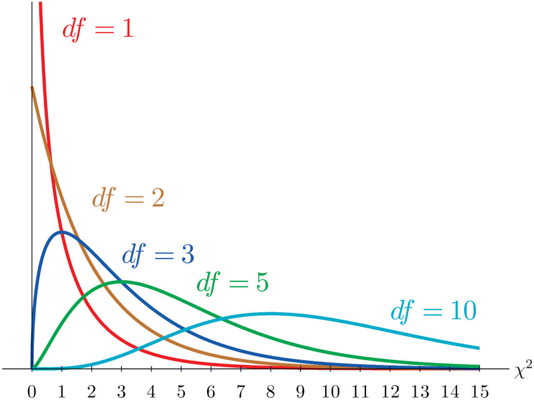
Besides the normal distributions, there are a few other distributions that will be useful for us throughout the semester.
The **Chi-sqaure distribution** is directly obtained from independent, standard normal random variables.

$$X = \sum_{i=1}^{n} Z_i^2 \tag{29}$$

X has a **chi-squared distribution** with n degrees of freedom (df). df corresponds to the number of terms in the sum - this is important for future analysis.

Chi-Square Distribution

The pdf looks like:



It is non-negative and is not symmetric. The expected value of X is n, and the variance of Z is 2n

t Distribution

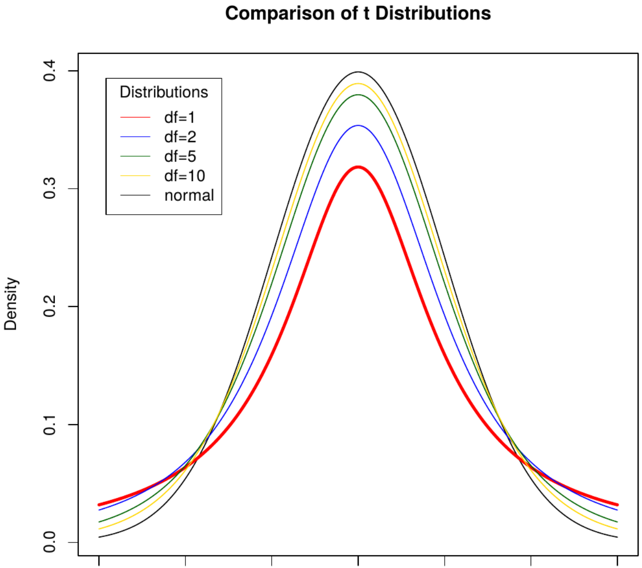
The t-distribution is the soul of econometrics.
Let Z have a standard normal distribution and X have a chi-squared distribution with n degrees of freedom. Assume Z and X are independent.

$$T = \frac{Z}{\sqrt{X/n}} \tag{30}$$

the t-distribution has n degrees of freedom. It's expected value exists only for $n > 1$, and the variance is $n/(n-2)$

t Distribution

The pdf looks like:



F Distribution

This distribution will be your friend especially when you are testing multiple variables.
Assume X_1 and X_2 are independent

$$F = \frac{X_1/k_1}{X_2/k_2} \tag{31}$$

has an F distribution with (k_1, k_2) degrees of freedom.

F Distribution

The pdf looks like: