Review - Mathematical Tools & Probability

Caio Vigo

The University of Kansas
Department of Economics

Spring 2019

These slides were based on *Introductory Econometrics* by Jeffrey M. Wooldridge (2015)

Topics

Summation Operator

It is a shorthand for manipulating expressions involving sums.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

Summation Operator - Properties

Property 1: For any constant c,

$$\sum_{i=1}^{n} c = nc$$

Property 2: For any constant c,

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$$

Summation Operator - Properties

Property 3: If $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is a set of n pairs of numbers, and a and b are constants, then:

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

Average

Given n numbers $\{x_1, x_2, \ldots, x_n\}$, their **average** or *(sample) mean* is given by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Summation Operator - Properties

Property 4: The sum of deviations from the average is **always** equal to 0, i.e.:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Property 5:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x})$$

Property 6:

$$\begin{aligned} \textstyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \textstyle\sum_{i=1}^{n} x_i(y_i - \bar{y}) \\ &= \textstyle\sum_{i=1}^{n} y_i(x_i - \bar{x}) \end{aligned}$$

Common Mistakes

Notice that the following does not hold:

$$\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i \right)^2$$

- Most important nonlinear function in econometrics

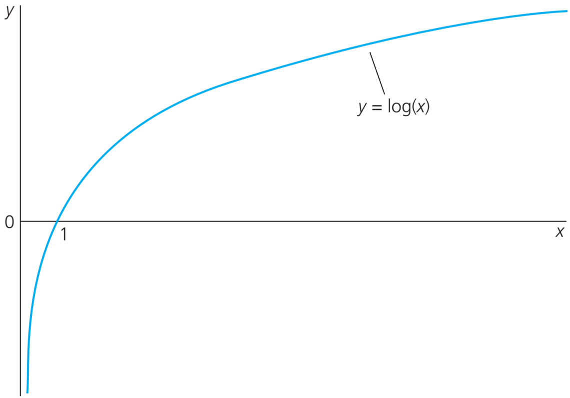
Natural Logarithm

$$y = log(x)$$

Other notations: $ln(x)$, $log_e(x)$

The Natural Logarithm

Figure: Graph of $y = log(x)$

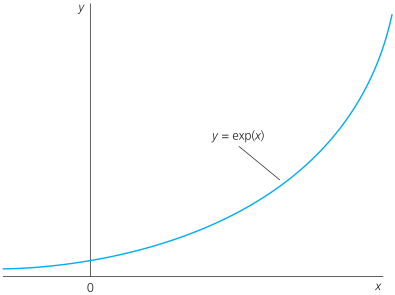


Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.

The Exponential Function

$$exp(0) = 1$$
$$exp(1) = 2.7183$$

Figure: Graph of $y = exp(x)$ (or $y = e^x$)



Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.

The Natural Logarithm

- Things to know about the Natural Logarithm $y = log(x)$:
  - is defined only for $x > 0$
  - the relationship between y and x displays diminishing marginal returns
  - $log(x) < 0$, for $0 < x < 1$
  - $log(x) > 0$, for $x > 1$
  - $log(1) = 0$
  - **Property 1:** $log(x_1 x_2) = log(x_1) + log(x_2)$, $x_1, x_2 > 0$
  - **Property 2:** $log(x_1/x_2) = log(x_1) - log(x_2)$, $x_1, x_2 > 0$
  - **Property 3:** $log(x^c) = c.log(x)$, for any c
  - **Approximation:** $log(1 + x) \approx x$, for $x \approx 0$

Topics

Random Variable

- A **random variable (r.v.)** is one that takes on numerical values and has an outcome that is determined by an experiment.
- Precisely, an r.v. is a function of a **sample space** Ω to the Real numbers.
- Points ω in Ω are called sample **outcomes, realizations, or elements**.
- Subsets of Ω are called **Events**.

Random Variable

- Therefore, X is a r.v. if $X : \Omega \to \mathbb{R}$

- Random variables are always defined to take on numerical values, even when they describe qualitative events.

Example

- Flip a coin, where $\Omega = \{\text{head, tail}\}$

Discrete Random Variable

Probability Function

X is a **discrete** r.v. if takes on only a finite or countably infinite number of values.

We define the **probability function** or **probability mass function** for X by
$f_X(x) = \mathbb{P}(X = x)$

Probability Density Function (pdf)

- A random variable X is **continuous** if there exists a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and for every $a \leq b$,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx$$

The function f_X is called the **probability density function** (pdf). We have that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

and $f_X(x) = F'_X(x)$ at all points x at which F_X is differentiable.

Joint Distributions and Independence

- We are usually interested in the occurrence of events involving more than one r.v.

Example

- Conditional on a person being a student at KU, what is the probability that s/he attended at least one basketball game in Allen Fieldhouse?

Joint Distributions and Independence

Joint Probability Density Function

- Let X and Y be discrete r.v. Then, (X, Y) have a **joint distribution**, which is fully described by the **joint probability density function** of (X, Y):

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

where the right-hand side is the probability that $X = x$ and $Y = y$.

Independence

- Let X and Y be two **discrete r.v.**. Then, X and Y are independent (i.e. $A \perp\!\!\!\perp B$), if:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

- Let X and Y be two **continuous r.v.**. Then, X and Y are independent (i.e. $A \perp\!\!\!\perp B$), if:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

for all x and y, where f_X is the marginal (probability) density function of X and f_Y is the marginal (probability) density function of Y

- In econometrics, we are usually interested in how one random variable, call it Y, is related to one or more other variables.

Conditional Probability

- Let X and Y be two **discrete r.v.**. Then, the conditional probability that $Y = y$ given that $X = x$ is given by:

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

- Let X and Y be two **continuous r.v.**. Then, the conditional distribution of Y give X is given by:

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

Conditional Probability & Independence

- If $X \perp\!\!\!\perp Y$, then:

$$f_{Y|X}(y|x) = f_Y(y)$$

and,

$$f_{X|Y}(x|y) = f_X(x)$$

Features of Probability Distributions

- We are interest in three characteristics of a distribution of a r.v. They are:
  1. measures of central tendency
  2. measures of variability (or spread)
  3. measures of association between two r.v.

Measure of Central Tendency (1): The Expected Value

Expected Value

- The **expected value** of a r.v. X is given by:

$$E(X) = \begin{cases} \sum_{x \in X} x f(x) & \text{, if } X \text{ is discrete} \\ \int_{x \in X} x f(x) d(x) & \text{, if } X \text{ is continuous} \end{cases}$$

- Also called as **first moment**, or *population mean*, or simply **mean**
- **Notation:** the expected value of a r.v. X is denoted as $E(X)$, or μ_X

Properties of Expected Values

Property 1: For any constant c, $E(c) = c$

Property 2: For any constants a and b, $E(aX + b) = aE(X) + b$

Property 3: If $\{a_1, a_2, \ldots, a_n\}$ are constants and $\{X_1, X_2, \ldots, X_n\}$ are r.vs. Then,

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

- **Example:** (on white board) If $X \sim$ Binomial(n, θ), where $X = Y_1, Y_2, \ldots, Y_n$ and $Y_i \sim$ Bernoulli(θ). Find $E(X)$.

Measure of Central Tendency (2): The Median

Median

The **median** is the value separating the higher half from the lower half of a data sample.

For a **continuous** r.v., the median is the value such that one-half of the area under the pdf is to the left of it, and one-half of the area is to the right of it.

For a **discrete** r.v., the median is obtained by ordering the possibles values and then selecting the value in the “middle”.

Measure of Central Tendency (3): The Mode

Mode

The **mode** of a set of data values is the value that **appears most often.**

It is the value of a r.v. X at which its p.d.f. takes its maximum value.

It is the value that is **most likely to be sampled.**

Measure of Central Tendency (2): The Median

- $E(X)$ and $Med(X)$ are both valid ways to measure the center of the distribution of X
- In general, $E(X) \neq Med(X)$
- However, if X has a **symmetric distribution** about the value μ, then:

$$Med(X) = E(X) = \mu$$

Measure of Variability (1): Variance

Variance

Let X be a r.v. with mean μ_X. Then, the **variance** of X is given by:

$$\text{Var}(X) = E\left[(X - \mu_X)^2\right]$$

Properties of Variance

- Let X be a r.v. with a well defined variance, then:

Property 1: $\text{Var}(X) = E(X^2) - \mu_X^2$

Property 2: If a and b are constants, then: $\text{Var}(aX + b) = a^2\text{Var}(X)$

Property 3: If $\{X_1, X_2, \ldots, X_n\}$ are independents r.vs. Then:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right)$$

Measure of Variability (2): Standard Deviation

Standard Deviation

The **standard deviation** of a r.v. X is simply the positive square root of the Variance, i.e.

$$\text{sd}(X) = \sqrt{\text{Var}(X)}$$

among the notations for the standard deviation we have: $\text{sd}(X)$, σ_X, or simply σ.

Property: For any constant c, $\text{sd}(c) = 0$

- **Example:** (on white board) Sample with the weights. What is $Var(X)$ and $sd(X)$?

Measure of Association (1): Covariance

- **Motivation:** (on white board)

Covariance

Let X and Y be two r.v. with mean μ_X and μ_Y respectively. Then, the **covariance** between X and Y is given by:

$$\begin{aligned} \text{Cov}(X,Y) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= E\left(XY\right) - E\left(X\right)E\left(Y\right) \\ &= E\left(XY\right) - \mu_X\mu_Y \end{aligned}$$

Notation: $\sigma_{X,Y}$

- Covariance measures the amount of linear dependence between two r.v.
- If $\text{Cov}(X,Y) > 0$, then X and Y moves in the same direction.

Properties of Covariance

Property 1: If X and Y are independents, then $(\Rightarrow)$ Cov$(X, Y) = 0$

Property 2: If Cov$(X, Y) = 0$, this does NOT imply $(\nRightarrow)$ that X and Y are independents.

Measure of Association (2): Correlation

- **Goal:** A measure of association between r.v.s that is not impacted by changes in the unit of measurement (e.g., income in dollars or thousands of dollars)

Correlation

Let X and Y be two r.v., the **correlation** between X and Y is given by:

$$\text{Corr}(X,Y) = \frac{Cov(X,Y)}{sd(X)sd(Y)} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Notation: $\rho_{X,Y}$

- $\text{Cov}(X,Y)$ and $\text{Corr}(X,Y)$ always have the same sign (because denominator is always positive)
- $\text{Corr}(X,Y) = 0$ if, and only if $\text{Cov}(X,Y) = 0$

Properties of Correlation

Property:

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

- If $\text{Cov}(X, Y) = 0$, then $\text{Corr}(X, Y) = 0$. So, we say that X, Y are **uncorrelated** r.v.
- If $\text{Corr}(X, Y) = 1$, then X, Y have a **perfect POSITIVE linear relationship.**
- If $\text{Corr}(X, Y) = -1$, then X, Y have a **perfect NEGATIVE linear relationship.**

Variance of Sums of Random Variables

Property Variance of Sums of Random Variable: For any constants a and b,

$$\text{Var}\,(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

- **Example:** (on white board) [Let $X \sim$ Binomial(n, θ) and consider $X = Y_1 + Y_2 + \ldots + Y_n$, where each Y_i are independent Bernoulli(θ)]

Goal:

- Want to explain one variable, called Y, in terms of another variable, X

- We can summarize this relationship between Y and X looking at the **conditional expectation** of Y given X, i.e., $E(Y|x)$

- $E(Y|x)$ is just a function of x, giving us how the expected value of Y varies with x.

Conditional Expectation

Conditional Expectation

- If Y is a **discrete** r.v.

$$E(Y|x) = \sum_{j=1}^{m} y_j f_{Y|X}(y_j|x)$$

- If Y is a **continuous** r.v.

$$E(Y|x) = \int_{y \in Y} y f_{Y|X}(y|x).dy$$

Properties of Conditional Expectation

Property 1:

$$E[c(X)|X] = c(X)$$

for any function $c(X)$

Property 2: For any functions $a(X)$ and $b(X)$

$$E[a(X)Y + b(X)|X] = a(X)E(Y|X) + b(X)$$

for any function $c(X)$

Property 3: If $Y \perp\!\!\!\perp X$, then:

$$E[E(Y|X)] = E(Y)$$

Distributions - The Normal Distribution

- The most widely used distribution in Statistics and econometrics.

Normal distribution (Gaussian distribution)

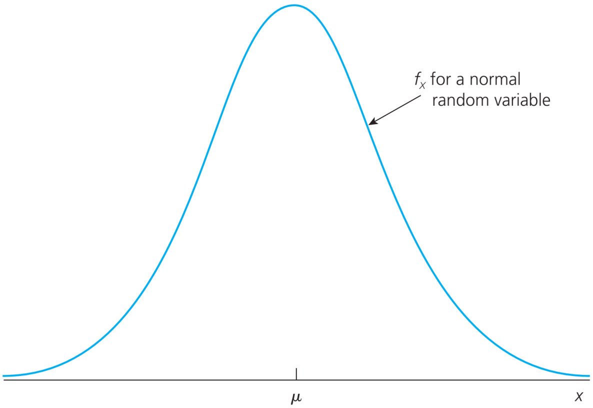
If a r.v. $X \sim N(\mu, \sigma^2)$, then we say it has a **standard normal distribution**. The pdf of X is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty$$

where $f(x)$ denotes the pdf of X.

Property: If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim \mathrm{N}(0, 1)$

Figure: Normal Distribution



Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.

Standard Normal distribution

If a r.v. $Z \sim N(0,1)$, then we say it has a **standard normal distribution**. The pdf of Z is given by:

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

where $\phi(z)$ denotes the pdf of Z.

Distributions - The Chi-Square Distribution

Chi-Square distribution

Let $Z_i, i = 1, 2, \ldots, n$ be independent r.v., where each $Z_i \sim N(0, 1)$. Then,

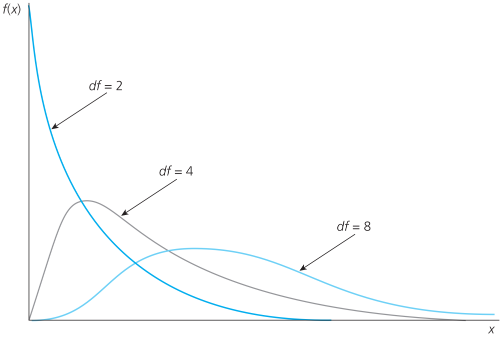
$$X = \sum_{i=1}^{n} Z_i^2$$

has a Chi-Square distribution with n degrees of freedom.

- **Notation:** $X \sim \chi_n^2$

- If $X \sim \chi_n^2$, then $X \geq 0$
- The Chi-square distribution is not symmetric about any point.

Figure: **Chi-Square Distribution**



Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.

- The t-distribution plays a role in a number of widely used statistical analyses, including:
  1. Student's t-test for assessing the statistical significance of the difference between two sample means,
  2. construction of confidence intervals for the difference between two population means,
  3. and in linear regression analysis.

Distributions - The t Distribution

t distribution

Let $Z \sim N(0,1)$ and $X \sim \chi^2_n$, and assume Z and X are independents. Then, the random variable:

$$t = \frac{Z}{\sqrt{X/n}}$$

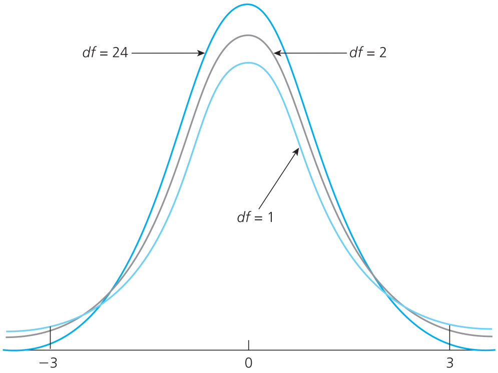
has a t distribution with n degrees of freedom.

- **Notation:** $t \sim t_n$

History:

- The distribution takes its name from William Sealy Gosset's 1908 paper in Biometrika under the pseudonym "Student".
- Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples. For example, the chemical properties of barley where sample sizes might be as few as 3.

Figure: The t distribution



Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.

Distributions - The F Distribution

- Important for testing hypothesis in the context of multiple regression analysis

F distribution

Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$, and assume X_1 and X_2 are independents. Then, the random variable:

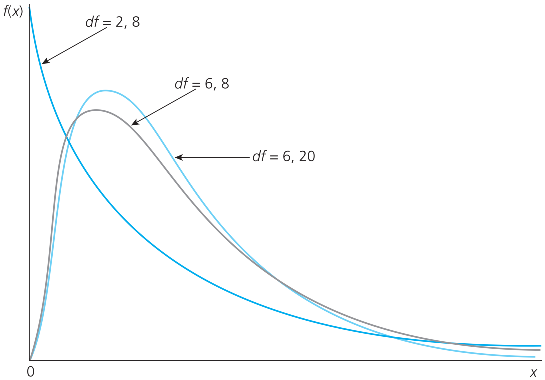
$$F = \frac{(X_1/k_1)}{(X_2/k_2)}$$

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

- **Notation:** $F \sim F_{k_1,k_2}$
  - k_1 : numerator degrees of freedom
  - k_2 : denominator degrees of freedom

Distributions - The F Distribution

Figure: The F_{k_1,k_2} distribution



Source: Wooldridge, Jeffrey M. (2015). Introductory Econometrics: A Modern Approach.