

Intro to Probability, Random Variables and Matrix Algebra

Venki

PhD student

Public Policy and Political Economy, School of EPPS

Start recording!

Note to self

Setting expectations

Objective

Primarily, to discuss the solutions for the worksheets.

Secondarily, to rehash the basics. This will primarily happen in the pre-midterm workshops.

What to expect

1. Worksheet solutions
2. Basic stats/math

What I expect from you

Whenever possible, your participation in the workshop

Whenever you ask for (coding) help, follow the syllabus guideline: “you must provide code in .Rmd file, dataset that is used with this code, Full dialogue with (any) LLM”

Outline

All the basic stats you should (hopefully already!) know...

Probability theory

- Event
- Sample space
- Probability function
- Kolmogorov Axioms

Random variables

- Function
- RVs to probabilities

Densities and Distributions

- Probability Mass, Density and Distribution functions
- Realizations, Support
- Uniform and normal distributions
- Expectation values and moments
- iid

Several Random Variables and Bayes Theorem

- Joint density
- Marginal density
- Conditional density

Estimation, Estimators and Estimands

- Estimation
- Estimator
- Estimand

Linear Algebra for Regression

- Basic matrix operations
- OLS regression in matrix form

Question to ponder...

Say you are writing a proposal to get funding for a field experiment (say RCT). **What are two things you will *have to* say about the experiment in the abstract of the proposal?**

The dependent and independent variables!

But lets first build a definition for the mathematical object that is a “variable”...

Sets of Events

Events

A possible outcome of an experiment is called an event.

In the absence of experiments one can conceive of the idea of a data generating process and its outcome would also be an event.

Sample space

A set containing all events/outcomes from an experiment is called a sample space.

$$\Omega = \{A, B, \dots\}$$

Since it contains all the possible outcomes the events of the sample space are said to be “collectively exhaustive”.

1. Probability theory

Probability v “Impact” of an event

Consider this question...

Say I am building a variable to denote election outcomes. Say its an election with two outcomes - Candidate A won or Candidate B won. So, sample space is,

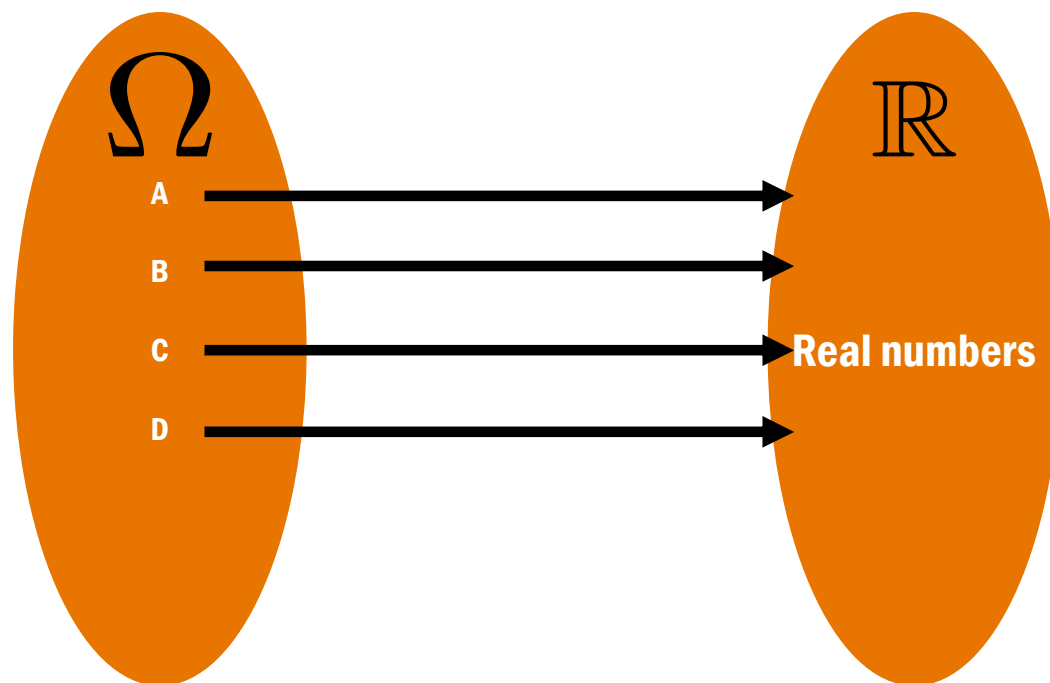
$$\Omega = \{A, B\}$$

Now if I just write $A > B$, what would it mean?

Choose one of these answers:

1. *Candidate A winning is more probable than Candidate B winning*
2. *Candidate A winning is socially more beneficial than Candidate B winning*
3. *It could mean both 1 & 2. This notation is ambiguous!*
4. *I don't understand the question.*

Probability function

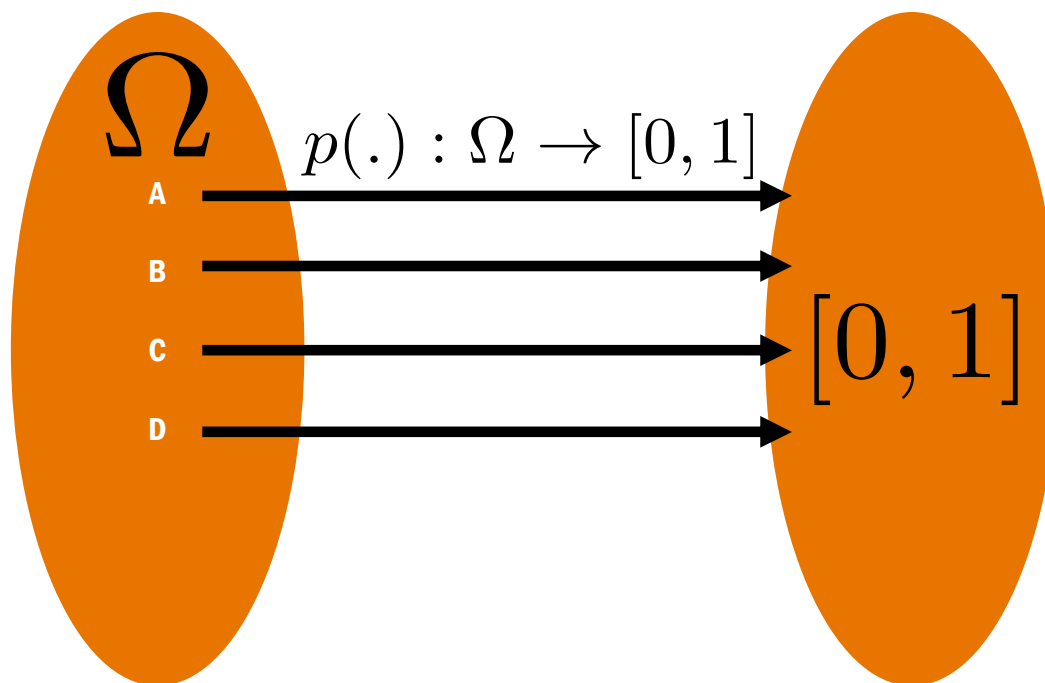


But can we allow all real numbers?

A sensible Probability function!

Kolmogorov axioms

1. Non-negativity: $p(A) \geq 0 \forall A \in \Omega$
2. Normalization to 1: $p(\Omega) = 1$.
3. Additivity: $p(A \cup B) = p(A) + p(B)$
if events $A, B \in \Omega$ are
mutually exclusive (disjoint).



2. Random Variables

Random Variable (another function!)

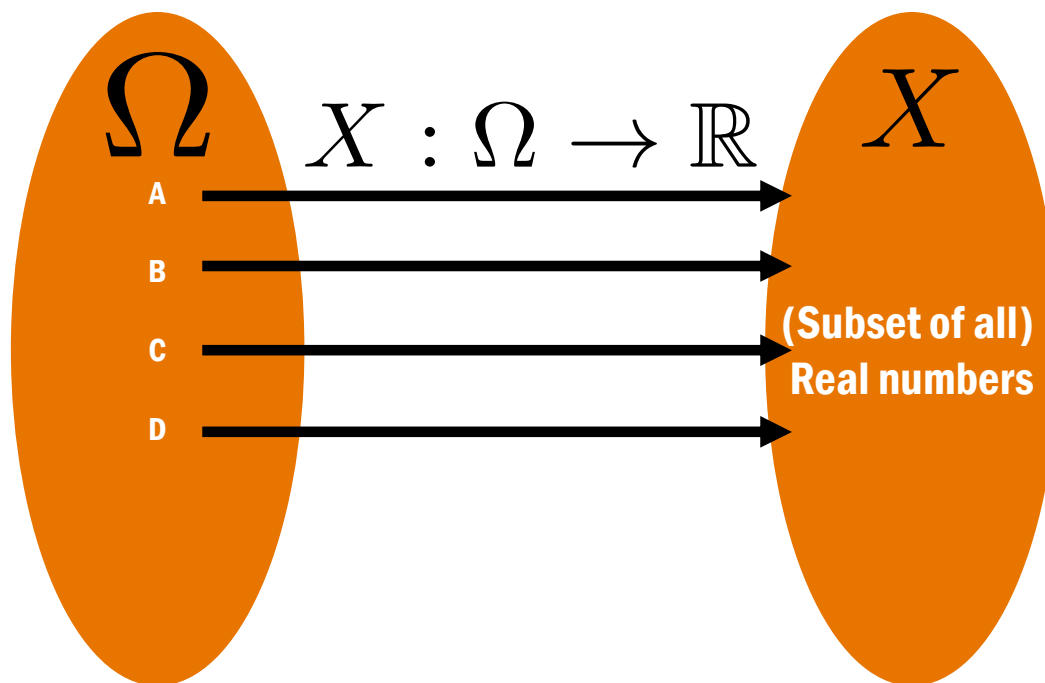
Denoted as

Reconsider the example

$$X = \begin{cases} 0 & \text{if Candidate A wins} \\ 1 & \text{if Candidate B wins} \end{cases}$$

*Is it “Random” or a
“Variable”?*

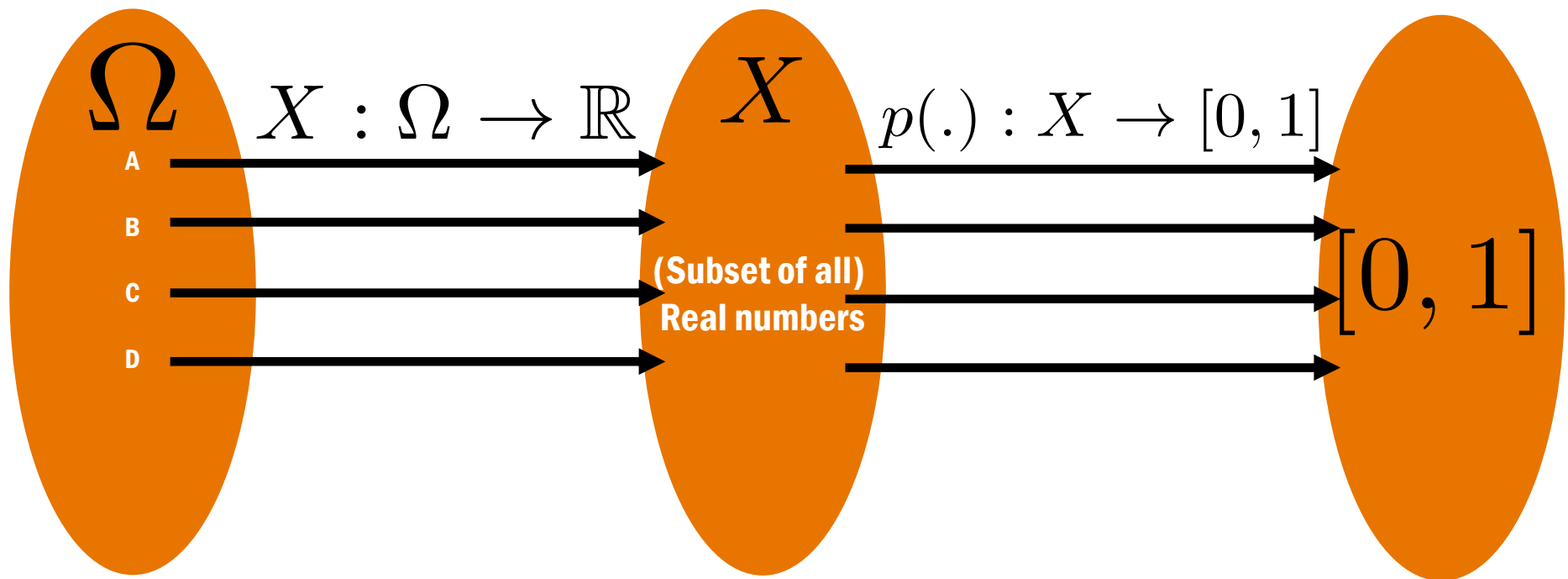
Neither really! Just a function.



*Again, here, can we allow all real numbers?
Unlike last time, YES!*

2. Random Variables

Both functions together (with a small change)



Flashback to earlier...

3. Densities and Distributions

Probability Mass, Density and Distribution Functions

Probability Mass Functions (PMF) must still obey the Kolmogorov axioms.

$$\sum_x p(X = x) = 1$$

Discrete RV

vs.

Continuous RV

If Ω is countable, then the RV is discrete. Otherwise, it is continuous.

If you don't know what "countability" means watch [this video](#) (after workshop!)

For a continuous RV, we instead have a Probability Density Function (PDF) which again must obey the axioms.

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

Other related terms

Realizations

The number assigned to an event is called a realization. x (from previous slide) is used to denote a realization. Remember that *only the realizations can be observed*. I will come back to this point later...

Support

The set of all possible realizations is called support.

Degenerate Random Variable

A Random Variable for a sample space containing only one event is said to be degenerate.

Example

$$Pr(0 \leq X \leq 10) = \int_0^{10} p(x)dx$$

What is X ?

What is, say, $x = 5$?

What is $[0, 10]$?

Continuous or Discrete RV?

What is $p(\cdot)$?

What is $Pr(\cdot)$?

Imagine a sample space for this.

Couple of common densities

Density function of Uniform distribution

$$p(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Density function of Normal distribution

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Expectation values & moments

For discrete RV

$$\mathbb{E}[X] = \sum_x x p(x)$$

For continuous RV

$$\mathbb{E}[X] = \int x p(x) dx$$

Examples

1. Consider a 2 sided coin.

$$\begin{aligned}\mathbb{E}[X_{coin}] &= 0 \times p(T) + 1 \times p(H) \\ &= 0 \times 0.5 + 1 \times 0.5 \\ &= 0.5\end{aligned}$$

2. Consider a 6 sided dice.

$$\begin{aligned}\mathbb{E}[X_{dice}] &= 1 \times p(1) + 2 \times p(2) + 3 \times p(3) + \dots \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + \dots \\ &= 3.5\end{aligned}$$

*Expectation operator
has several useful
properties*

Check out – [Algebra of
Random Variables](#)

Generalizing,

$$\text{Moment}_i = \mathbb{E}[X^i] = \sum_x x^i p(x)$$

$$\text{Var}[X] = \mathbb{E}[(X - E[X])^2] = \sum_x (x - \mu)^2 p(x)$$

What does i.i.d. mean?

Repeat an experiment, say, 6 times

$X_1, X_2, X_3, X_4, X_5, X_6$

If the distribution of realizations across experiments and of each experiment is the same, then the 6 RVs are said to be i.i.d. Such RVs have the property that,
 $E[X] = E[X_1] = E[X_2] = \dots$

i.i.d = independently and identically distributed random variables

You can make something not i.i.d by breaking either the independence or the “identical” or both.

Breaking independence

In a time series, tomorrow depends on today

Breaking “identical”

Distribution of taco consumption on Taco Tuesday v other days. Here tuesday is simply not identical to other days!

Outline

Halfway done...

Probability theory

- Event
- Sample space
- Probability function
- Kolmogorov Axioms

Random variables

- Function
- RVs to probabilities

Densities and Distributions

- Probability Mass, Density and Distribution functions
- Realizations, Support
- Uniform and normal distributions
- Expectation values and moments
- iid

Several Random Variables and Bayes Theorem

- Joint density
- Marginal density
- Conditional density

Estimation, Estimators and Estimands

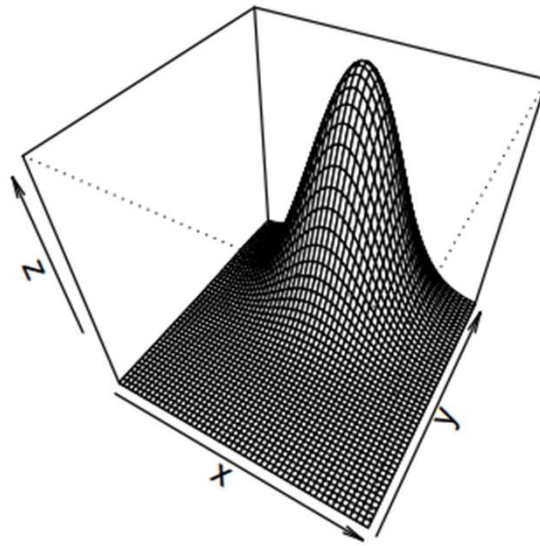
- Estimation
- Estimator
- Estimand

Linear Algebra for Regression

- Basic matrix operations
- OLS regression in matrix form

4. Several Random Variables and Bayes theorem

Joint density

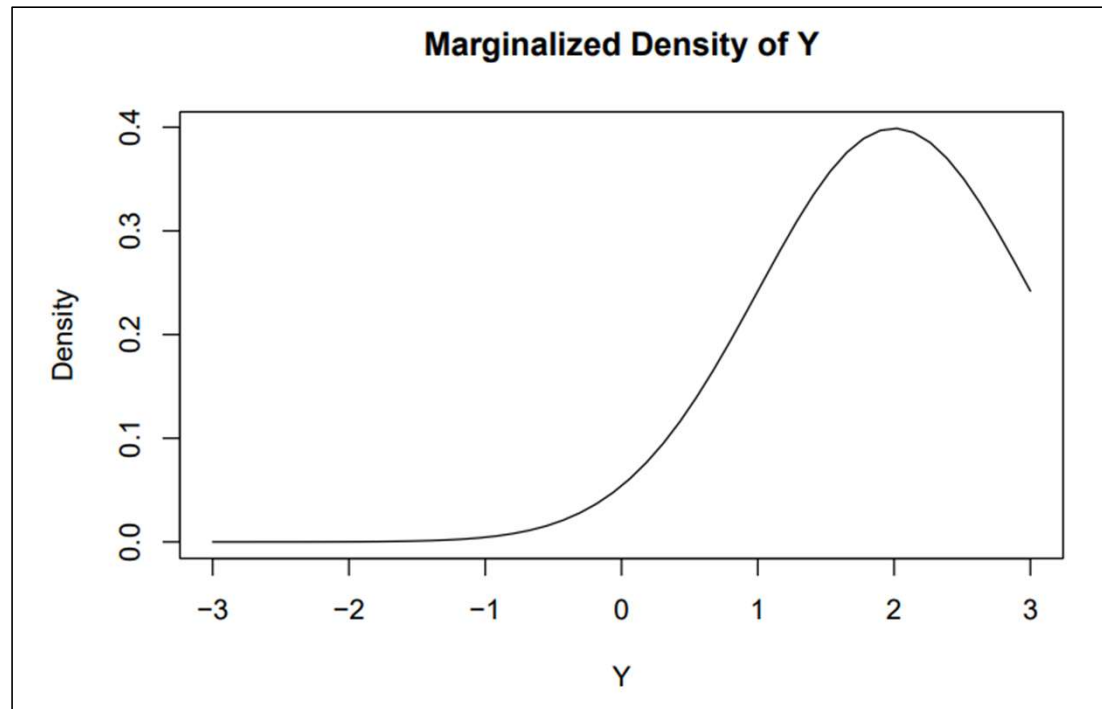


Continuous: $z \equiv \int_y dy \int_x p(x, y) dx = 1$

Discrete: $z \equiv \sum_y \sum_x p(x, y) = 1$

4. Several Random Variables and Bayes theorem

Marginal density



$$p(y) = \sum_x p(x, y)$$

$$p(y) = \int_x p(x, y) dx$$

4. Several Random Variables and Bayes theorem

Conditional density

Conditional Distribution: $p(y|x) = \frac{p(x, y)}{p(x)}$

Bayes Law: $p(x|y) = \frac{p(y|x) p(x)}{p(y)}$

Conditional Expectation (Discrete) : $\mathbb{E}[X|Y] = \sum_x x p(x|y)$

Independence of RVs: $p(y|x) = \frac{p(x, y)}{p(x)}$
 $= \frac{p(x) p(y)}{p(x)}$
 $= p(y)$

$$E[Y|X] = E[Y]$$
$$E[XY] = E[X]E[Y]$$

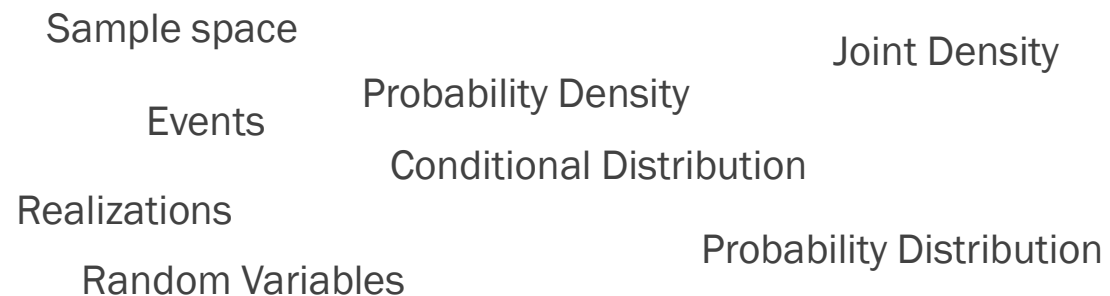
5. Estimation, Estimators and Estimands

The real world...

*When you go to a field
and perform an RCT
and collect data what
exactly are you
getting?*

Options are on the right.

Type responses in the chat.



Flashback to earlier...

5. Estimation, Estimators and Estimands

Estimation of Estimands is done with Estimators

Estimation

Generalizing a “quantity of interest” from the sample (using realizations) to the population is called estimation.

Quantity of interest could be any of the n^{th} order moments – Expectation Value, Variance, Skew...

Estimand

The “quantity of interest” is called estimand.

In this course - ATE, ATT and rarely ATC

When you propose your project on Week 8, tell us which estimand you are planning to get from your analysis.

$$E[X]$$

Estimator

The mathematical function (or numerical algorithm) used to do estimation of an estimand is called an estimator.

There are different type of estimators to find the same estimand! Choose wisely!

$$\hat{E}[X]$$

Thats' all folks!

Oh wait!

Matrices!

I saved the best for last...



Basic matrix operations

Multiplying Conformable matrices

You can't multiply any two matrices. They need to be conformable.

Conformable: $A_{3 \times 5} \times B_{5 \times 10} = AB_{3 \times 10}$

Not Conformable: $B_{5 \times 10} \times A_{3 \times 5}$

Note that matrices can have several dimensions.

Transpose

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 7 \\ 6 & 4 & 9 \\ 3 & 5 & 1 \end{bmatrix}$$

$$\mathbf{A}^T = \begin{bmatrix} 1 & 6 & 3 \\ 3 & 4 & 5 \\ 7 & 9 & 1 \end{bmatrix}$$

Properties:

$$(\mathbf{A}^T)^T = \mathbf{A}$$

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$

$$(c\mathbf{A})^T = c\mathbf{A}^T$$

Inverse

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbb{I}_n$$

$$\text{where, } \mathbb{I}_n = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}_{n \times n}$$

Properties:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

$$(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$$

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$

OLS regression in matrix notation

Regression model: $Y = \mathbf{X}\beta + \epsilon$

We estimate this with:

$$Y = \mathbf{X}\hat{\beta} + e$$

$$e = Y - \mathbf{X}\hat{\beta}$$

OLS means minimizing this:

$$e^{\mathsf{T}}e = (Y - \mathbf{X}\hat{\beta})^{\mathsf{T}}(Y - \mathbf{X}\hat{\beta})$$

Outline

We are done!

1

Probability theory ✓

2

Random Variables ✓

3

Densities and Distributions ✓

4

Several Random Variables and Bayes theorem ✓

5

Estimation, Estimators and Estimands ✓

6

Linear Algebra for Regression ✓

Everything all at once

Summary

Probability theory

- Event
- Sample space
- Probability function
- Kolmogorov Axioms

Random variables

- Function
- RVs to probabilities

Densities and Distributions

- Probability Mass, Density and Distribution functions
- Realizations, Support
- Uniform and normal distributions
- Expectation values and moments
- iid

Several Random Variables and Bayes Theorem

- Joint density
- Marginal density
- Conditional density

Estimation, Estimators and Estimands

- Estimation
- Estimator
- Estimand

Linear Algebra for Regression

- Basic matrix operations
- OLS regression in matrix form



For real this time!
Questions? Comments? Rants?

Venkatesh Subramanian

Venkatesh.Subramanian@utdallas.edu

See you later in class!