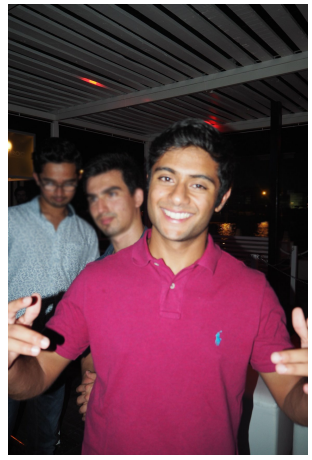
Welcome to Data 100 Fall 2019!

Raguvir Kunani

September 4, 2019

Challenge: Meet someone next to you before section begins!

Also, please fill out this form: tinyurl.com/raguvirTAform



Raguvir Kunani

I'm a junior EECS major from Bakersfield, CA. This is my second time as a TA and I'm really excited to teach you all!

- **Email**: rkunani@berkeley.edu - feel free to email about anything!
- **Discussion**: Wed 11-12, Evans 332
- **Lab**: Mon 11-12, Evans B6
- **OH**: Tues 2-3, Evans 426 (room might change)

Introductions

If you haven't already, meet the person behind, left, and right of you!

Suggested talking points:

- What classes you're taking
- What you did this summer
- One thing you're looking forward to today or in the near future

After you're done talking, fill out this form: tinyurl.com/raguvirTAform

A Word on Prerequisites

- Official Prerequisites: CS61A, Math 54, Data 8
- IMO, you can do well in this class without taking Data 8.
- You'll need to be comfortable programming in Python and know basic linear algebra concepts like matrix-vector multiplication, inverses, invertibility, rank, and projections.

Discussion Logistics & Ground Rules

- The discussions are intentionally long, so I likely won't be able to get through all the problems
- I'll make heavy use of group work, so it's in your best interest to talk the people around you
- If I ask a question, I'll wait until someone answers *no matter how long it takes* . It's ok if you answer incorrectly, that's how you learn.
- Be respectful to your fellow classmates. Everyone is here to learn.

Lecture Recap: Sampling

What is sampling and why do we need it?

Simple random sample (SRS) : sampling from the population uniformly at random *without replacement*

- Each element of the population has an equal chance of being included in the sample

Lecture Recap: Sampling Terminology

Target Population : complete collection of individuals we want to generalize to

Sampling Frame : the set of individuals that could potentially get chosen for the sample

Sample : the individuals that were actually chosen by the sampling method

> Note: The target population, sampling frame, and sample could overlap in certain cases.

Lecture Recap: Random Variables

A **random variable** (RV) is a variable that can take on many different values (as any variable would). A random variable *also* has a certain probability of taking on each of its values.

For example, let X be a RV denoting the outcome of a roll of a fair, six-sided die. Then, we can say that X can take on values $1, 2, 3, 4, 5, 6$ each with probability $\frac{1}{6}$ since the die is fair.

The fancy math way of saying this is $X \in \{1, 2, 3, 4, 5, 6\}$ and $P(X = x) = \frac{1}{6}$ for all $x \in X$.

Expectation and Variance

These are the key elements of RVs you need to know. The formulas for them can be hard to grasp at first, but if we look at each formula piece by piece, they aren't so bad.

$$E[X] = \sum_{x \in X} x \cdot P(X = x)$$

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Lecture Recap: Data Summarization

Statistic : a function or characteristic of a sample (e.g. mean)

Loss Function : a function that describes how "good" a certain model describes the observed data

- In lecture you saw the average squared loss $\sum_{i=1}^{n} (x_i - c)^2$

Risk : expected loss on the population (why is there an expectation?)

Empirical Risk : Average loss on the observed data

Discussion Worksheet

Probability Distributions

A **probability distribution** of a RV specifies the probability of each value a RV can take on. You can think of a probability distribution as a table with 2 columns: value and probability of the value.

In our die example, this is the probability distribution of X:

Value	Probability
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Feedback Form

This *anonymous* form is for me to learn what I can do to ensure you all get the most of discussion and lab. This form will be open all semester, and I'll be checking it regularly. Be as ruthless as you want, I promise my feelings won't get hurt.

Feedback Form: tinyurl.com/raguvirTAfeedback

Fill out this form so I can get to know you! → tinyurl.com/raguvirTAform