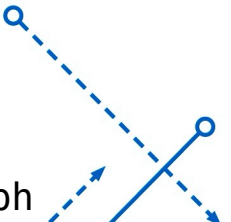
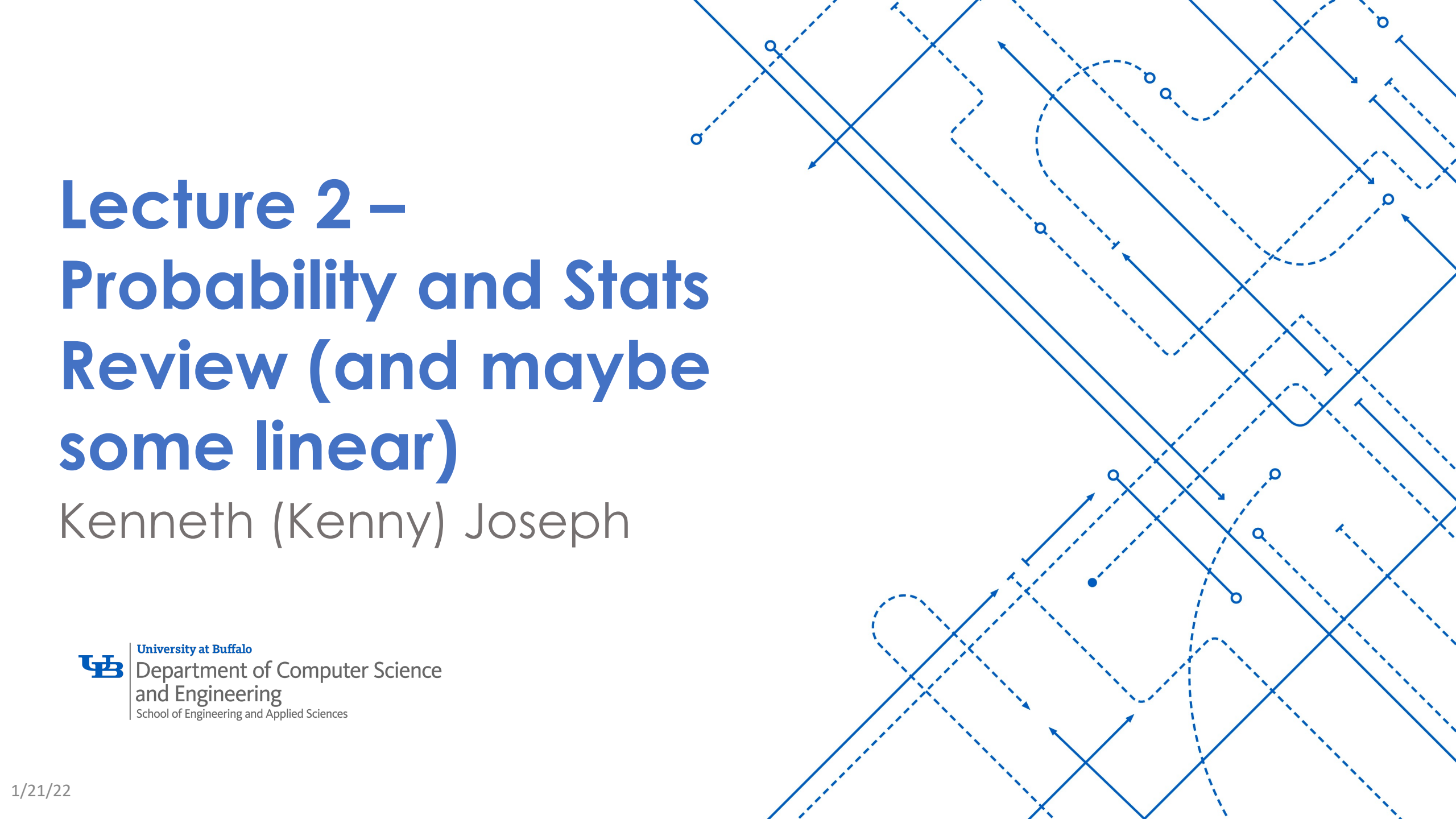


<Insert similar meme about probability and statistics>





Lecture 2 – Probability and Stats Review (and maybe some linear)

Kenneth (Kenny) Joseph

Announcements

- Materials section of website
- Not everyone on Piazza
- Programming Assignment 1 out Sunday night/Monday AM
- Weekly Quiz 0 due Monday
- Weekly Quiz 1 out Tuesday at midnight
- Questions?!

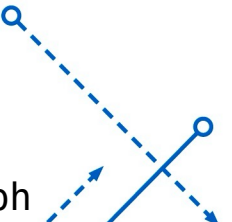
Answers to questions you might not have

- Why are you covering stuff in the pre-reqs?
 - You can never learn about these things enough times
 - You might not have understood some of it the first time
 - (I didn't, and my whole Intro ML course went over my head)
 - It's good setup material

Are there things you expect us to know that you are *not* going to review?

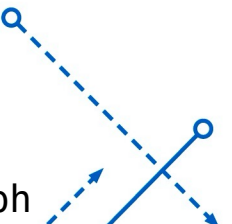
- **Basic calculus.**
 - We will in general aim to avoid the use of any exceptionally difficult calculus, but some of it is unavoidable
- **Basic python.**
 - I do not expect significant prior experience with scientific computing tools, but you should have basic knowledge of python
- **Counting** <- **Seeing Theory!**
 - You should be familiar with combinations and permutations
- **Basic linear algebra** <- **Three Blue one Brown (also as needed)**
 - Matrices, Vectors, multiplication and addition, transpose/etc

What concerns do you have?



Note: you may at times today say to yourself, “what does this have to do with machine learning”?

I’ll try to be clear on this today, but also, you’ll see soon enough!



Random Variables



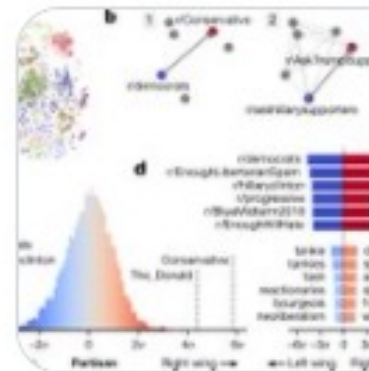
Ashton Anderson @ashton1anderson · Dec 1, 2021

Thrilled to share our paper, out today in @nature, which proposes a new paradigm for the analysis of online platforms and applies it to study political polarization over the complete history of Reddit

w/ @isaacwaller

Paper: [nature.com/articles/s4158...](https://www.nature.com/articles/s41586-021-03415-8)

1/n



nature.com

Quantifying social organization and political polariz...

Nature - A new method quantifies the social makeup of online communities, and applying it to 1...

18

391

1K



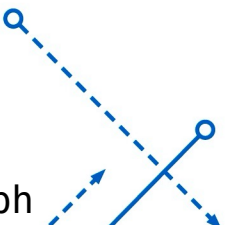
Random Variables

- A map from a sample space to real numbers
- Example – what is the sample space for two consecutively rolled dice?
- What is a random variable we could construct from that sample space?

- KAHOOT

Exercise Define the random variable X to be the number of heads on 3 coin tosses.

1. Enumerate for each point in the sample space the value of X . (pg. 28)
2. Define $P_X(X = x)$ based on the induced probability function
3. What is $P_X(X = 1)$?



Continuous vs discrete

- KAHOOT

Distributions of random variables

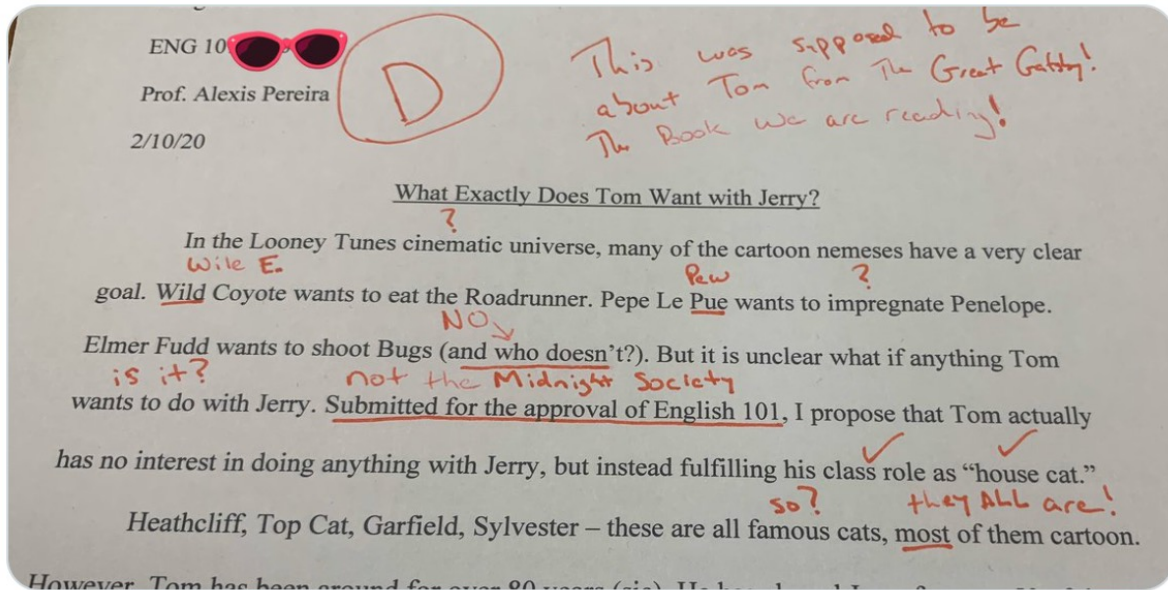
- CDF, PMF, PDF
- **Exercise:**
 - Draw the cdf of X when X is the number of heads observed after tossing 3 fair coins.

What have we said so far?



Alexis Pereira
@MrAlexisPereira

Teaching my first English course this semester has been rewarding but I don't know what to do with this student

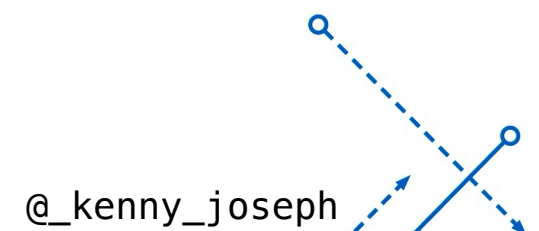


12:57 PM · Feb 10, 2020 · Twitter for iPhone

130.4K Retweets 19.8K Quote Tweets 787.8K Likes

We have a way to talk about the probability that we see this tweet be retweeted, say, at least x number of times. It is as an (unknown) cdf.

One way to think about ML – we want to learn that CDF.

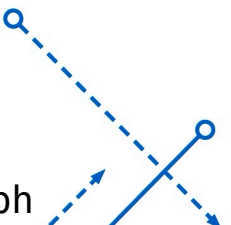


Expected value and Variance

- Mathematically
- Seeing theory + Kahoot
- Linearity assumption
- What does $\text{Var}[X] = 0$ mean?

Probability Distributions

- Why are they useful?
 - Give the functional form for a bunch of naturally arising phenomenon
 - You only need a few numbers (**distribution parameters**) to summarize the whole probability distribution!



Give us functional form for natural things

<http://www.wzchen.com/probability-cheatsheet>

Kahoot.

Hypergeometric Distribution

Let us say that X is distributed $HGeom(w, b, n)$. We know the following:

Story In a population of w desired objects and b undesired objects, X is the number of “successes” we will have in a draw of n objects, without replacement. The draw of n objects is assumed to be a **simple random sample** (all sets of n objects are equally likely).

Examples Here are some $HGeom$ examples.

- Let's say that we have only b Weedles (failure) and w Pikachus (success) in Viridian Forest. We encounter n Pokemon in the forest, and X is the number of Pikachus in our encounters.
- The number of Aces in a 5 card hand.
- You have w white balls and b black balls, and you draw n balls. You will draw X white balls.
- You have w white balls and b black balls, and you draw n balls without replacement. The number of white balls in your sample is $HGeom(w, b, n)$; the number of black balls is $HGeom(b, w, n)$.
- **Capture-recapture** A forest has N elk, you capture n of them, tag them, and release them. Then you recapture a new sample of size m . How many tagged elk are now in the new sample?
 $HGeom(n, N - n, m)$



You only need to know model parameters

- For example, for the normal distribution, if we know the mean and variance, we can compute $F(X=x)$ for any x !
- Seeing Theory
- Code demo

What have we said so far?

- If we have a problem where we can apply a particular probability distribution, if we can learn the parameter(s) of the distribution, we're done!
- ... hang onto this one for a few minutes.

Multiple Random Variables

- Conditional vs Marginal Probability
- Correlation
- Independence

Conditional Probability Practice

You are off to soccer, and want to be the Goalkeeper, but that depends who is the Coach today:

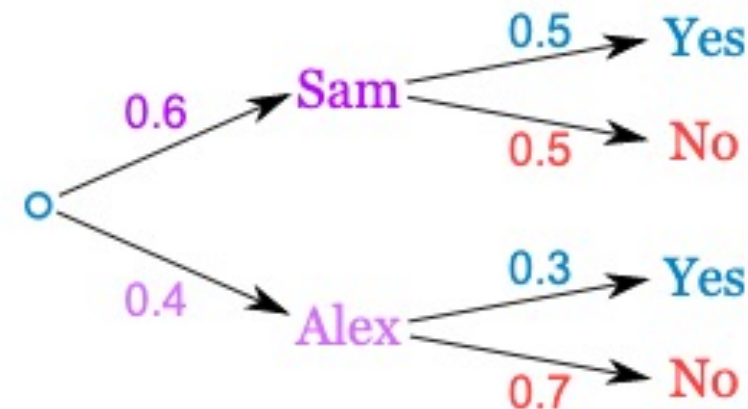
- with Coach Sam the probability of being Goalkeeper is **0.5**
- with Coach Alex the probability of being Goalkeeper is **0.3**

Sam is Coach more often ... about 6 out of every 10 games (a probability of **0.6**).

So, what is the probability you will be a Goalkeeper today?

<https://www.mathsisfun.com/data/probability-events-conditional.html>

- $P(\text{Goalie})$
- $P(\text{Goalie and Sam})$
- $P(\text{Goalie} \mid \text{Alex})$



Correlation

- Most common correlation metric: Pearson...
- Example
 - Pearson = .9?
 - Pearson = 0?
 - Pearson = -.9?
- Be careful with the correlation coefficient

Independence

- Mathematical definition
- Does Pearson correlation being 0 mean that two variables are independent?

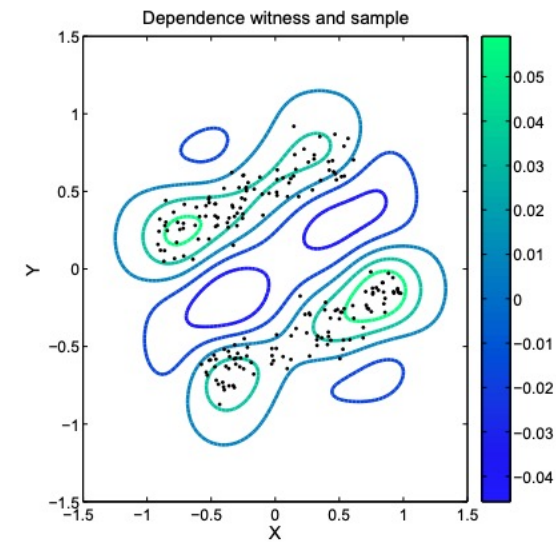
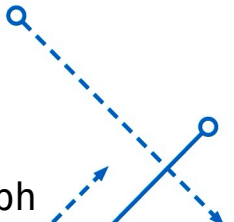


Figure 4: Illustration of the function maximizing the mean discrepancy when MMD is used as a measure of dependence. A sample from dependent random variables x and y is shown in black, and the associated function \hat{f}^* that witnesses the MMD is plotted as a contour. The latter was computed empirically on the basis of 200 samples, using a Gaussian kernel with $\sigma = 0.2$.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1), 723-773.



What have we said so far?

- With one RV: in ML, our goal is to learn the cdf of the RV
- What about supervised learning, where we have outcome RV Y and (many) input RV(s) X ?

- A more complete view ... machine learning is just conditional probability estimation