Probability and Statistics

Sanjit K. Kaul

April 26, 2022

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

What do you think this course is going to be about?

How many of you would rather not take the course?

Our brains are just not wired to do probability problems very well - Persi Diaconis

From the book Randomness by Deborah J. Bennett

Randomness

Uncertainty

Chance

Our Probabilistic World

- Chance, its quantifications, pervade our world

- Not just in a casino or when playing a game of cards or when solving a math puzzle

We All Must Grapple With Chance

- A fraction of natural births are identical twins

We All Must Grapple With Chance

- A fraction of natural births are identical twins

- Tests screening for disease result in false positives and false negatives
  - Your test is negative. Should the doctor put you on medication?

We All Must Grapple With Chance

- A fraction of natural births are identical twins
- Tests screening for disease result in false positives and false negatives
  - Your test is negative. Should the doctor put you on medication?
- Vaccine efficacy
  - You are safe, rather safer, probabilistically
  - Your odds of a severe infection are a fraction of that of one not vaccinated

We All Must Grapple With Chance

- The odds of becoming a lightning victim in the U.S. in any one year is 1 in 7,00,000 (NatGeo)
- The odds of being struck in your lifetime is 1 in 3,000 (NatGeo)
- Are commercial flights more dangerous than lightning?

We All Must Grapple With Chance

- The odds of even being in a plane crash are one in 11 million.

- You end up in a plane crash. What are the odds that you will survive?
  - 95%

Chance As a Decision Maker

Draw of lots to allocate flats

Delhi Development Authority has launched 1354 flats on offer on January 2, 2021. Around 33000 applications have been received till 16 February 2021. Out of these 33000 applications around 22,500 applicants have made payment... *On 10th March 2021, **the draw of lots** for the allotment of flats will be held.*

Chance As a Decision Maker

- Random Elimination using *"Akkad bakkad bumbe bo, assi nabbe pure so, sow mein nikala dhaaga, chor nikal ke bhaaga."*
  - *Random?*

Chance As a Decision Maker

- Assignment of grades via a draw of lots?

Did the Ancients Use Chance?

"Chance mechanisms have been used since antiquity: to divide property, delegate civic probabilities or privileges, settle disputes..."

- One die dated around 2750 BC from excavations of ancient Mesopotamia in Northern Iraq
- Dice found at the Mohenjo-Daro excavation in the Indus valley

From Randomness by Deborah J. Bennett

Probability

- It’s a number

- By convention in [0,1]

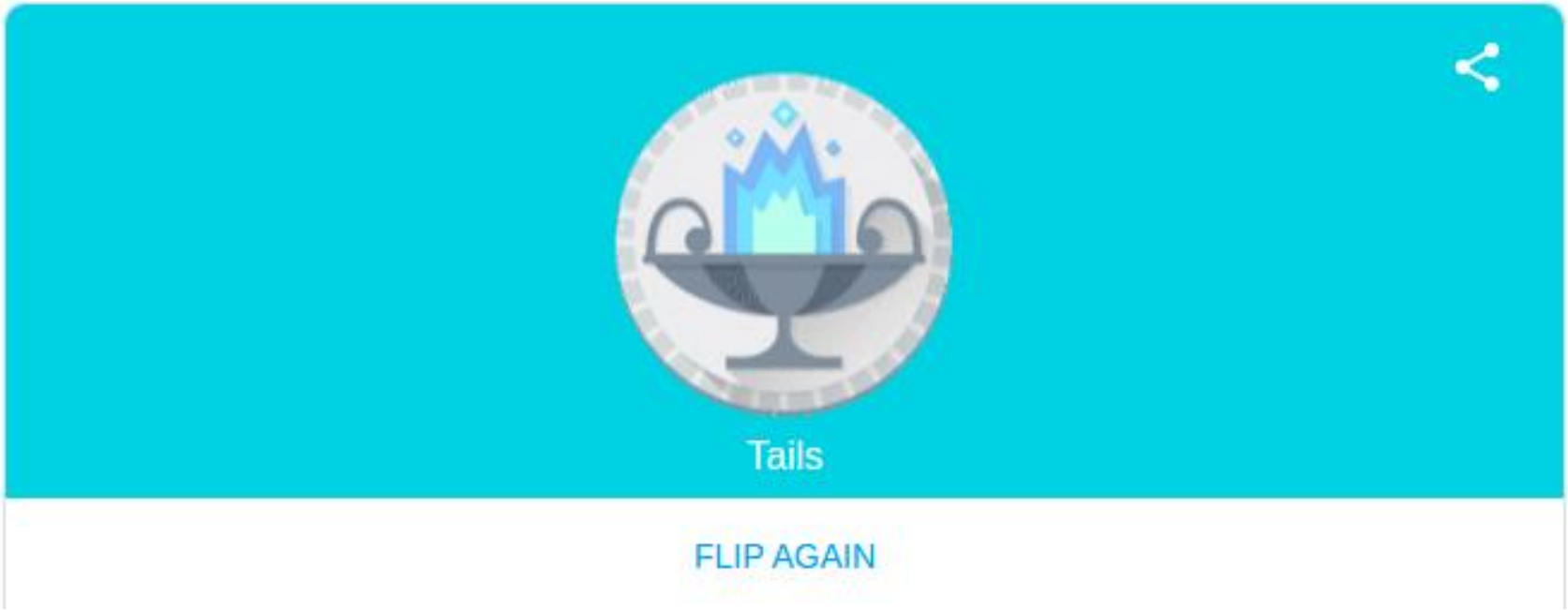
Probability

- A number that quantifies belief in the occurrence of something

- A number that, for our purposes, maybe arrived at as a result of experiment

Let’s Toss a Coin (In class exercise)

https://www.google.com/search?q=toss+a+coin&oq=toss+a+coin



Measurements and Uncertainty

- You perform an experiment multiple times and obtain different results/measurements/outcomes
- What do you do?
  - Resign to such uncertainty?
- You capture the uncertainty using a probability model

Measurements and Uncertainty

- You could perform the experiment many times
- Associate with each of the different outcomes their frequency of outcome
  - Assign frequency of occurrence to the probability
- You have a probability model
  - Captures in a deterministic manner the uncertainty you saw in your experiments

Measurements and Uncertainty

- You could perform the experiment many times
- Associate with each of the different outcomes their frequency of outcome
  - Assign frequency of occurrence to the probability
- You have a probability model
  - Captures in a deterministic manner the uncertainty you saw in your experiments

- Often, in real world experiments, many repeats later you see that the uncertainty can be captured in such a manner

Probability Models

- Many commonly occurring probability models
  - They define the set of all possible outcomes
  - Associate probabilities with outcomes
- You can create your own
  - Must satisfy some commonsense rules
- Different models can be thought of as different randomizers/ methods of drawing lots

Why Probability Models?

- We don’t toss coins and roll dice often

- Maybe you don't play games that require tossing coins or rolling dice

Why Probability Models?

- We don’t toss coins and roll dice often
- Maybe you don't play games that require tossing coins or rolling dice
- For us coins and dice are models that we understand well
- Importantly, they often help model randomness in the real world

Why Probability Models?

- To exemplify, a fair coin can model the "natural process" that determines whether genetic sex of a newborn is male or female

Why Probability Models?

- To exemplify, a fair coin can model the "natural process" that determines whether genetic sex of a newborn is male or female

- A communications channel can be modelled as a coin too – a biased one however

Why Probability Models?

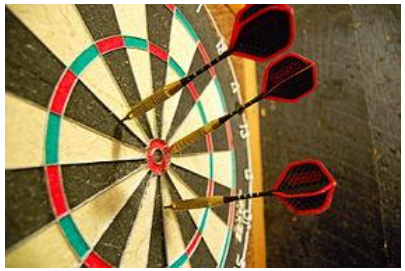
- To exemplify, a fair coin can model the "natural process" that determines whether genetic sex of a newborn is male or female
- A communications channel can be modelled as a coin too – a biased one however
- Of course, not everything is discrete
- Probability models may be more complex than a coin or dice

Why Think in Terms of Probability?

- Some observed phenomena are random (non-deterministic)
  - Thermal Noise
    - Impossible to predict the value of noise at a given instant
    - However, its average is 0 and deterministic
  - Gas contained in a closed vessel
    - Atoms have random trajectories and apply random pressures to the walls of the vessel.

Why Think in Terms of Probability?

- Phenomena may be deterministic but too complex to model deterministically
  - Coin tossing
  - Game of darts
    - Where will the dart hit?
  - Where will a missile hit?



Why Think in Terms of Probability?

- Phenomena may be deterministic but appear chaotic in nature
  - Very small changes in initial conditions, states can lead to very different observations (Butterfly Effect)
  - Your model is a differential equation.
  - However, predicting an output for a given input or state of the system is impossible
    - Small errors in input or state measurement can lead to very different measurements

Probability and Statistics

- Probability models help characterize experimental outcomes

- In the probability part of the course we will assume that a probability model is given/ known/ implicit
  - It *governs* your experiment's outcomes, how the lots are drawn

Probability and Statistics

- Probability models have parameters
  - Just as a line has the parameters slope and y-intercept
- You may not know the parameters in the real world
- You may have only observations/ experiment outcomes governed by the probability model

Probability and Statistics

- In the statistics part of the course, you will begin with measurements/outcomes of an experiment

- You will use the measurements to calculate statistics

Probability and Statistics

- Each parameter of a model *corresponds to* a *statistic* of a collection of numerical outcomes of the experiment

  - Parameter: Expected value, Statistic: *Mean* of outcomes (empirically calculated, comes with uncertainty).

Probability and Statistics

- Probability models will become hypotheses

- Given measurements, you will decide which hypothesis, from a set of hypotheses, it came from

Brief Detour

Textbooks, Grading and etc

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Textbooks

- Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers by Roy D. Yates (2nd Edition) [Indian Edition].
- Introduction to Probability and Statistics for Scientists and Engineers By Sheldon Ross (5^{th} Edition)
- Most homework will be assigned from these books.
- Topics covered in both are very similar. However, organization of topics is different

A Few Dos and Don'ts

- Do not allow coursework to pile up
- Do not come to a class without revising what was taught in the previous class
- Do not read the book as if it were a novel
- Do not learn by rote, you will fail to learn with probability 1

A Few Dos and Don'ts

- The authors RY&DG make very important observations (read "Message to Students from the Authors" in the Preface)
  - Being able to use formulae is not enough
  - You must be able to solve *word* problems
- There is only one way of getting better at the probabilistic way of thinking and that is to keep working on word problems
- Work out problems on a sheet of paper, not in air, not in your head, not in your imagination or dreams
- Use proper notation (as in book by RY)

A Few Dos and Don'ts

- There is only one way of getting better at the probabilistic way of thinking and that is to keep working on word problems

- Work out problems on a sheet of paper, not in air, not in your head, not in your imagination or dreams

- Use proper **notation** (as in book by RY)

Gadgets Policy in Class

- No phones, laptops, and such like.
- Only pen/pencil and paper

On Copying

- **Axiom 0** If you copy from others you cannot learn the subject

- **If you want to learn** and are in a position where you think that copying is the only way out
  - Do not copy
  - Meet me/TA in person and we will work on your situation

On Copying

- **If you do not want** to learn, feel free to copy.

- **Penalties will apply**
  - **At least as harsh as the institute policy**
    - Copying and/or helping another student during an exam leads to an F as per institute policy.

On Attendance

- Must enter classroom within 10 minutes of start
  - This is non-negotiable
  - Important to maintain decorum for the 50 or so minutes we have
- If you leave, you will be marked absent
- If you leave after giving attendance, you will be awarded the F grade.
- Must submit in class exercise sheet (always one sheet, two sides) at end of lecture
  - Put your name and roll number on it
  - Will be graded with a non-zero probability

Grading (Most Likely Distribution)

- Exam based evaluations (85%)
  - Midsem (20%), Endsem (25%), Three quizzes (40%)
- Assignments (0 - 10%)
  - They are for you to learn well topics covered during lectures
  - While the percentage is just 10%, how sincerely you do your assignments will determine your score in the exams
- In class exercise submissions (0 - 5%)
- Coding component (0 – 15%)

More on Grading

- Show your steps, else you get a 0.
- Correct answers are important. The approach is more important

Keeping You on Track

- Assignments
- Exams
- Tutorials (please attend them)

Homework/Assignments

- Discussion is fine
  - Please give credit
- **Submitted solution must be yours**
- Will also feed into out of class interaction
  - Evidence of effort at your end will make me more likely to give you my time
- Solutions will be posted post deadline
- Please don't ignore assignments

Tutorials

- We will solve problems
- Your suggestions
  - Please suggest at least a few days in advance
- Some tutorial slots maybe used to conduct the quizzes

Notation

- Always use the notation we use in class
  - As in book by RY
  - It is not okay to write mathematics using arbitrary notation
    - Define any new notation that you may need to use. Else stick to RY
- You will not get marks for what is mathematically incorrect

Notation

- When you read the book, read the mathematics carefully

- The language of mathematics leaves nothing to chance
  - There should be no ambiguity in a mathematical statement
  - If you feel otherwise when you read, you may want to read more carefully

Plan

- Experiments, outcomes, sets, events, MECE, Bayes' rule, law of total probability, counting
  - Weeks 1 – 2
- Discrete RV
  - Weeks 2 – 3
- Continuous RV
  - Week 3 – 4
- Pairs/ Vectors of RV(s)
  - Weeks 4 – 6
- Chebyshev's Inequality, Sample mean, CLT, Weak Law, …
  - Weeks 7 – 8
- Estimation
  - Weeks 9 – 10
- Hypothesis Testing
  - Weeks 11 - 12

Post Conditions

- Students are able to derive first and second order statistics, conditional PDF(s) and PMF(s), of common random variables and functions of the same.
- Students are able to apply commonly found discrete and continuous state probabilistic models to various applications, as captured by word problems
- Students are able to derive properties of the sample mean and variance of samples from a normal population
- Students are able to derive estimates of the parameters of common probability distributions and test hypothesis from random samples
- Students are able to perform MATLAB (or equivalent) experiments to present data, generate data from distributions, calculate various statistics, and demonstrate basic hypothesis testing

You and Your Instructor

- You are free (in my class) to ask questions
  - Please be polite. Your politeness encourages me to pay attention
- Questions on what is being taught are welcome anytime during the class
  - You do not want me to go to slide 10 if you have questions regarding slide 9
- You will be graded based on what you do in exams and assignments. Not based on what you say or don't

You and Your Instructor

- Other things you may want to say are fine too
  - However, I would encourage you to keep general comments and observations, grievances, course feedback to office hours.

- Please remember that I can choose to not allow a certain interaction at any given time during the class or during office hours
  - We don’t have infinite time

You and Your Instructor

- You are not allowed to beg for marks
  - By doing so, you will be wasting my time. I will penalize you for the same.
- You are not entitled to marks. You must earn them.

You and Your TA(s)

- Please be polite (and expect the same in return)
- Respect the TA's time
  - Come prepared with questions you may have
  - Bring your attempt

Course Website

- On Classroom
  - Code: **idddndm**
  - Probability And Statistics Section B
- All messaging over Classroom
  - **In general, I won't respond to emails**
- Encourage you to share your questions with all
  - No question is a stupid question
  - It is stupid, however, to not ask a question
  - Encourage you to try working out the answer before you ask
  - ***Do not share entire solutions on Classroom***

Enjoy The Course

- Not a "math course"
- We will learn via fun toy problems
- We will learn to think probabilistically
- What we will learn has very broad applicability
- Don't fret, Have fun!