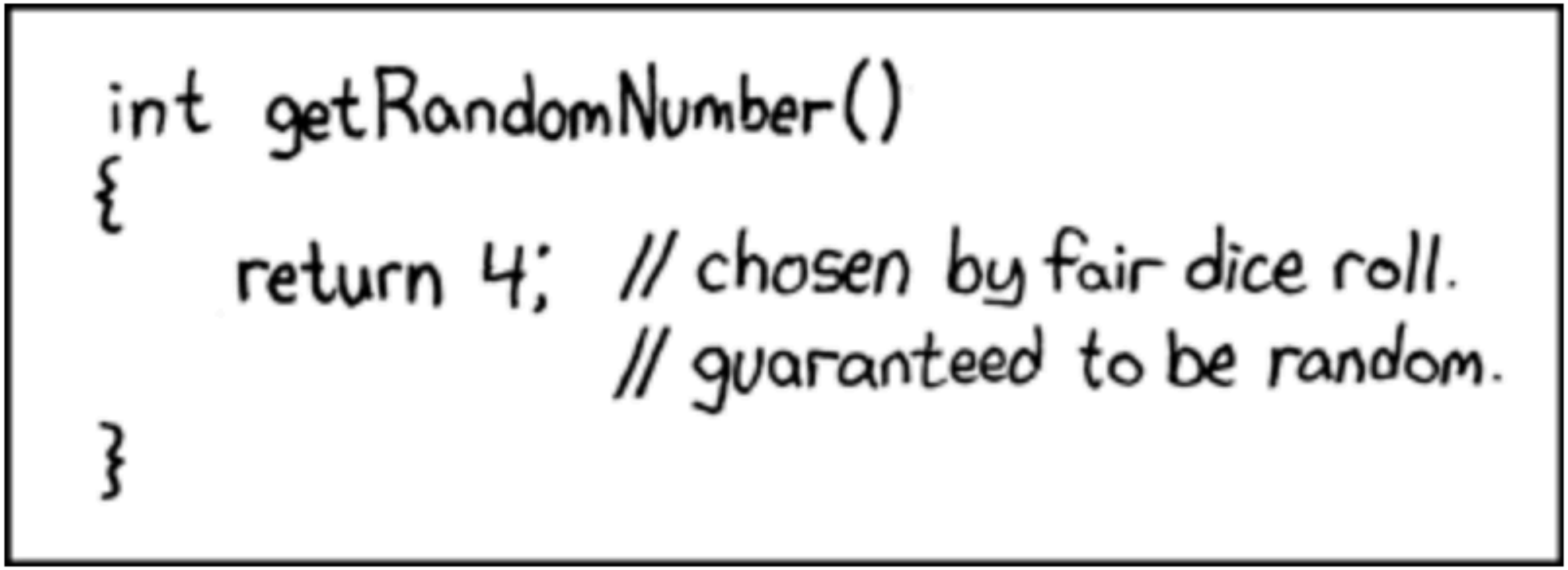
Stat 88: Probability and Statistics in Data Science



https://xkcd.com/221/

Lecture 2: 1/19/2024

Basics, Intersections

1.1, 1.2

Shobhana M. Stoyanov

Agenda

- Quick recap of definitions so far, and continue with terminology
- Idea of probability as a proportion, and assumptions for this
- Social media example from text (not equally likely outcomes)
- Exact calculations of probabilities vs bounds
- If we have time, talk about de Méré's Paradox

Terminology

- **Experiment**: action that results in exactly one of several possible outcomes or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.

- An *event* is a collection of outcomes.

- A collection of all possible outcomes of an action is called a *sample space* or an *outcome space*. Usually denoted by Ω (sometimes also by S).

- An event is always a subset of Ω. Suppose we call the event A, then we write this as $A \subset \Omega$. We denote the probability of A as $P(A)$.

Warm up: Go to pollev.com/shobhana to answer #1

1. If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

   $1/52$, $\frac{1}{13} = 4/52$

2. How did you do this? What were your assumptions?

   Equally likely

3. Say we roll a die. What is Ω?

   Do at home

4. What is the chance that the die shows a multiple of 3? What were your assumptions?

Chance of a particular outcome

- We usually think of the chance of a particular outcome (roll a 6, coin lands heads etc) as the number of ways to get that outcome divided by the total possible number of outcomes.

$$\frac{\#\ of\ particular\ outcomes\ of\ interest}{total\ \#\ of\ outcomes\ possible}$$

means # of outcomes in A

- So if A is an event (subset of Ω), then $P(A) = \dfrac{\#(A)}{\#(\Omega)}$, $A \subseteq \Omega$

$A \subset \Omega$ $\quad$ $A \subset \Omega$ $\quad$ $\subset$ vs $\subseteq$

$A \subseteq \Omega$

- If an experiment has a **finite** number of possible ***equally likely*** outcomes, then the probability of an event is the proportion of outcomes that are included in the event.

$\underline{52} \cdot \underline{51}$

Cards

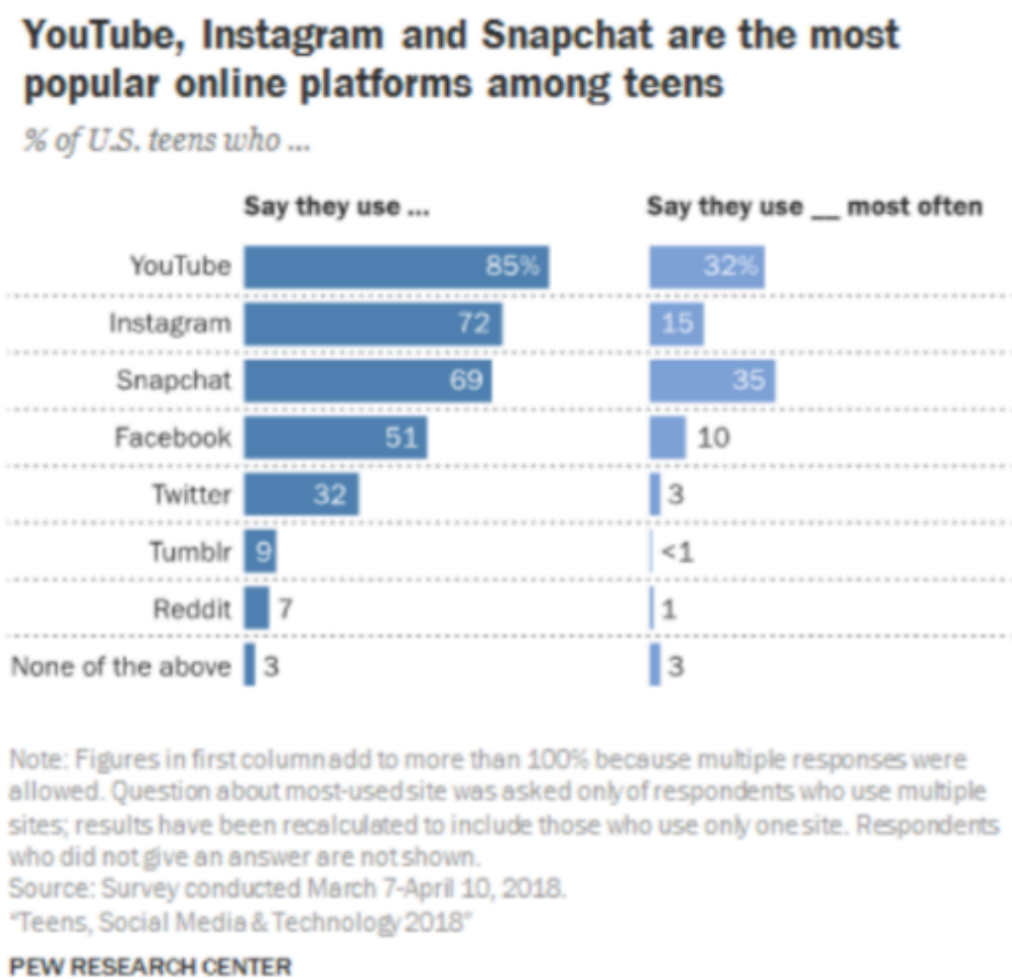
$\#(\Omega) = 52 \cdot 51$

- If you deal 2 cards, what is the chance that at least *one* of them is a queen?

We'll come back to. Meanwhile try at home.

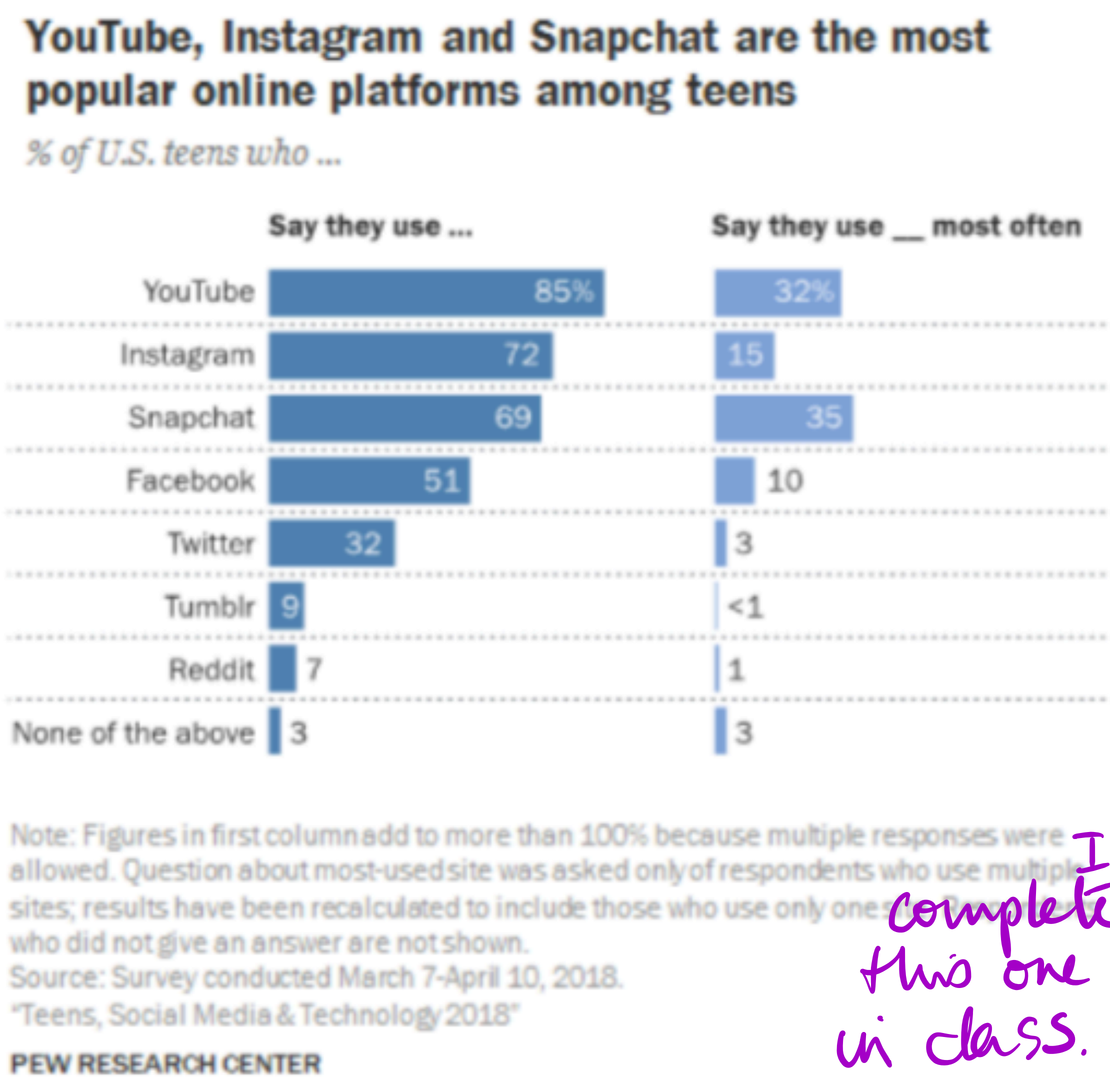
Not equally likely outcomes

- What if our assumptions of equally likely outcomes don't hold (as is often true in life, data is messier than nice examples).
- Here is a graphic from Pew Research displaying the results of a 2018 survey of social media use by US teens.



- What is the difference b/w 2 charts?
- Why do the % add up to more than 100 in the first graph?
- Second graph gives us a ***distribution*** of teens over the different categories

Not equally likely outcomes



1. What is the chance that a randomly picked teen uses FB most often?

10%

2. What is the chance that a randomly picked teen did *not* use FB most often?

90%

complement

3. What is the chance that FB *or* Twitter was their favorite?

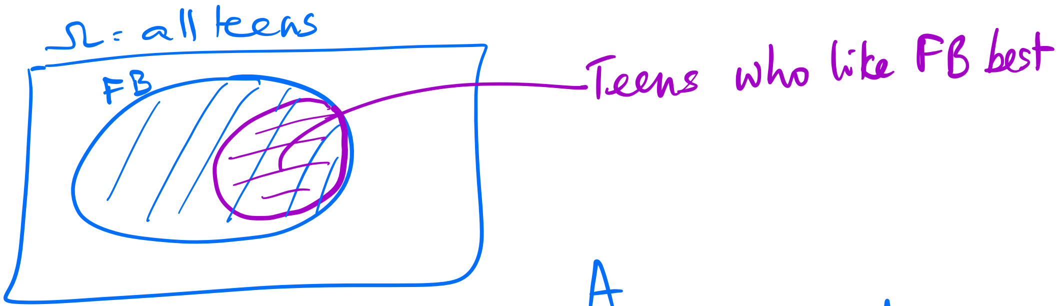
10% + 3% = 13%

4. What is the chance that the teen used FB, just not most often?

I didn't complete this one in class.

51% − 10% = 41%

5. Given that the teen used FB, what is the chance that they used it most often?



Given that the $\overbrace{\underbrace{\text{teen uses FB}}}^{A}$, what is the chance that $\underbrace{\text{FB is their favorite}}_{B}$;

Prob. of B GIVEN A

$$P(B \mid A) = \frac{10}{51} \approx 0.2$$

"Given"

only interested in these outcomes

Conditional Prob of B given A

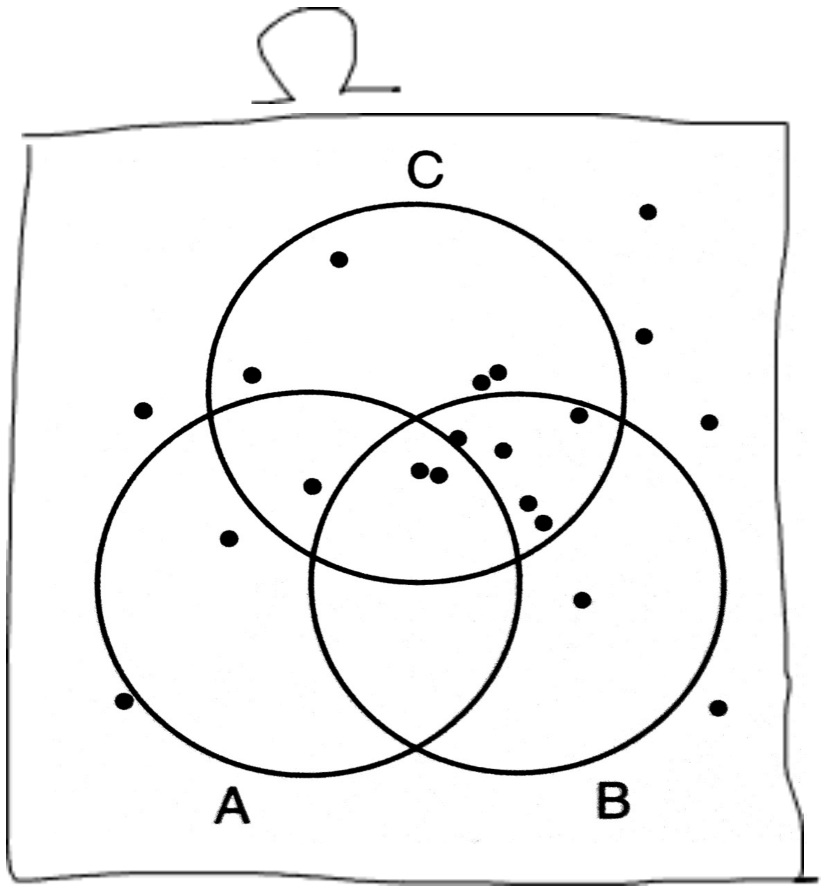
Terminology

- ***Experiment***: action that results in exactly one of several possible outcomes or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.
- An ***event*** is a collection of outcomes.
- A collection of all possible outcomes of an action is called a ***sample space*** or an ***outcome space***. Usually denoted by Ω (sometimes also by *S*).
- An event is *always* a subset of Ω. Suppose we call the event A, then we write this as $A \subset \Omega$

- A ***distribution*** of the outcomes over some categories represents the proportion of outcomes in each category (each outcome appears in one and only one category)

- The ***complement*** of an event A is an event consisting of all the outcomes that are not in A. It is denoted by A^C and we have that $P(A^C) = 1 - P(A)$ (***Complement Rule)***

Venn Diagrams



- You worked on this picture in section today. This kind of drawing is called a ***Venn Diagram.*** The events are usually represented as roughly circular, though they don't have to be, and the outcome space as a rectangle containing it. Often drawing Venn diagrams and other pictures will make your problem more clear.

So far:

- If all the possible outcomes are *equally likely*, then each outcome has probability $1/n$, where $n = \#(\Omega)$
- Let $A \subseteq \Omega$, $P(A) = \dfrac{\#(A)}{\#(\Omega)}$

certain event.
outcome space
$P(\Omega) = 1$

- Probabilities as proportions
- Sum of the probabilities of all the distinct outcomes should add to 1
- $0 \le P(A) \le 1$, $A \subseteq \Omega$
- A *distribution* of the outcomes over different categories is when each outcome appears in one and only one category.
- What would happen if we get some information about the outcome or event whose probability we want to figure out?
- Our outcome space reduces (number of possible outcomes), incorporating that information, so we recompute the probability.

Conditional probability

- In the last question, we used the information that the teen used FB. We were told the teen used FB, and *then* asked to compute the chance that FB was their favorite.
- This is called the *conditional probability that the teen used Facebook most often, given that they used Facebook* and denoted by:

$$P(\text{FB most} \mid \text{FB}) = \frac{10}{51}$$

size of new Ω

Conditional probability

- This probability we computed is called a ***conditional probability***. It puts a condition on the teen, and *changes* (restricts) the universe (the sample space) of the next outcome, a teen who likes FB best.
- To compute a conditional probability:
  - First restrict the set of all outcomes as well as the event to *only* the outcomes that *satisfy* the given **condition**
  - Then calculate proportions accordingly
- How do the probabilities in #1 and #5 compare?

~~Example~~ Exercise for home.

- A six-sided fair die is rolled twice:

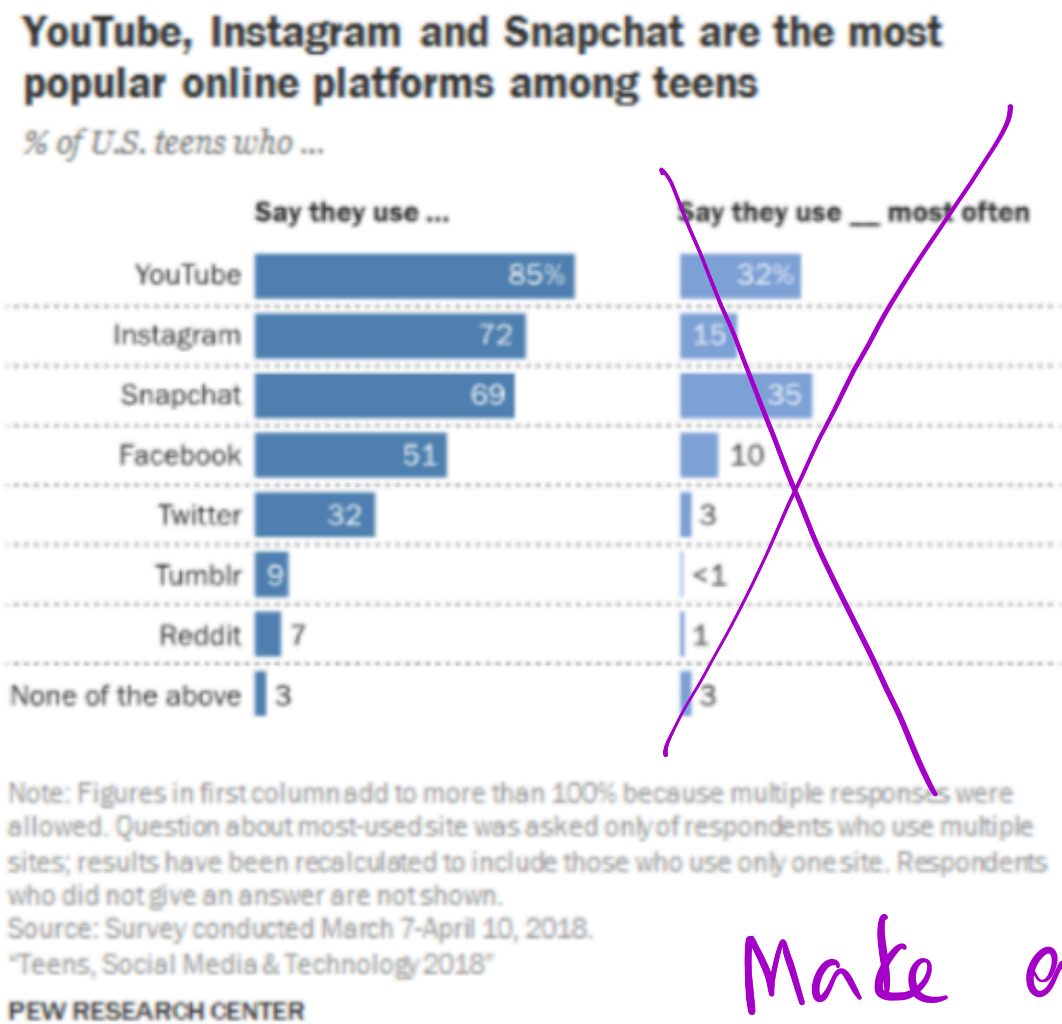
  - If the first roll lands on 1, what is the chance that the second roll lands on a number bigger than 2?

$$\frac{4}{6}$$

$$P(2^{nd}\text{ roll} > 2 \text{ AND } 1^{st}\text{ roll} = 1) = \frac{4}{36}, \quad P(2^{nd}\text{ roll} > 2 \mid 1^{st}\text{ roll} = 6) = \frac{4}{6}$$

Exercise: Find the probability that the second number is greater than the twice the first number.

Section 1.2: Exact Calculations, or Bound?



Recall #3 about FB or Twitter. What was the answer? What can you say about the chance that a randomly selected teen used FB or Twitter (not necessarily most often)?

$P(FB) = 0.51$

$P(T) = 0.32$

$P(FB \text{ or } T)$

Make a Venn diagram