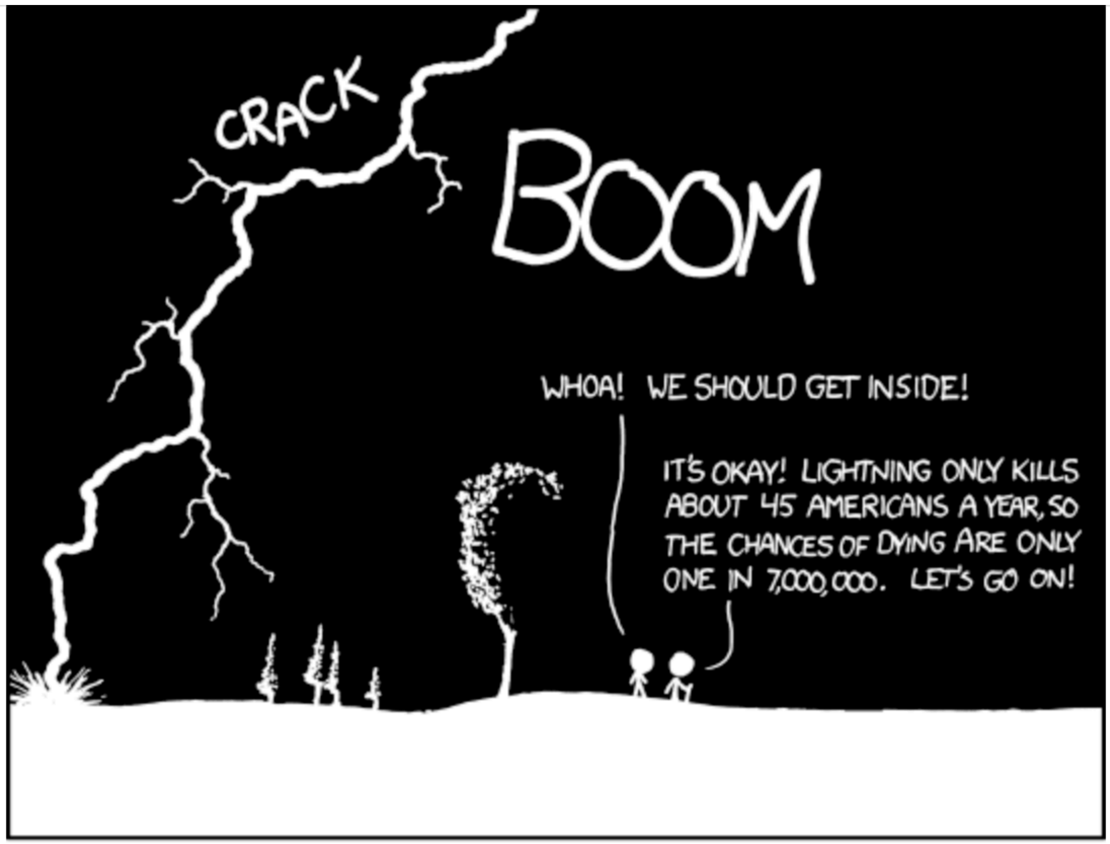
Stat 88: Probability and Statistics in Data Science



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

https://xkcd.com/795/

~~Lecture 3: 1/25/2022~~ Lec 3 : 1/22/2024

~~Axioms of Probability, Intersections,~~

Sections ~~1.3, 2.1~~ 1.2, 1.3

Shobhana M. Stoyanov

Agenda

Quick recap of terms

Section 1.2: Exact calculations or bounds

- unions vs intersections
- addition rule

Section 1.3: Fundamental Rules (the Axioms of Probability)

- Notation
- Axioms
- Consequences of the axioms
- De Morgan's Law

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.

- 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*). Ω : certain event $\emptyset = \{\ \} =$ "impossible event"

- An ***event*** A is a collection of outcomes and $A \subset \Omega$

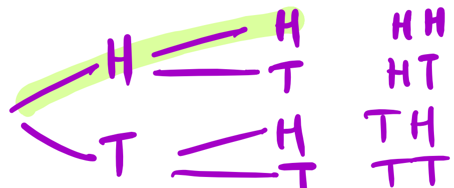
- A *distribution* of the outcomes over some categories represents the proportion of outcomes in each category (each outcome appears in one and only one category)

- 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)*

Terminology & rules

- $P(A) = \dfrac{\#(A)}{\#(\Omega)}$, every outcome is equally likely

ex. Toss a coin twice



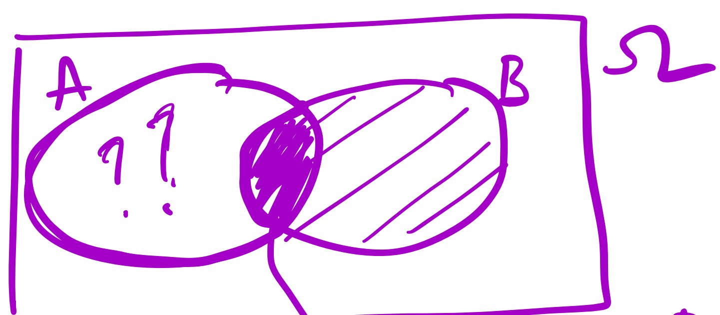
$\Omega = \{HH, HT, TH, TT\}$

- Multiplication: If an experiment is in k stages, and each stage i results in n_i outcomes, then the total number of outcomes is $n_1 \times n_2 \times \ldots n_k$

- 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)

$P(A^c) = \dfrac{\#(A^c)}{\#(\Omega)} = \dfrac{\#(\Omega) - \#(A)}{\#(\Omega)}$
$= 1 - P(A)$

- $\boxed{P(A \mid B) = \dfrac{\#(A \text{ and } B)}{\#(B)}}$ (The conditional probability of A given B)



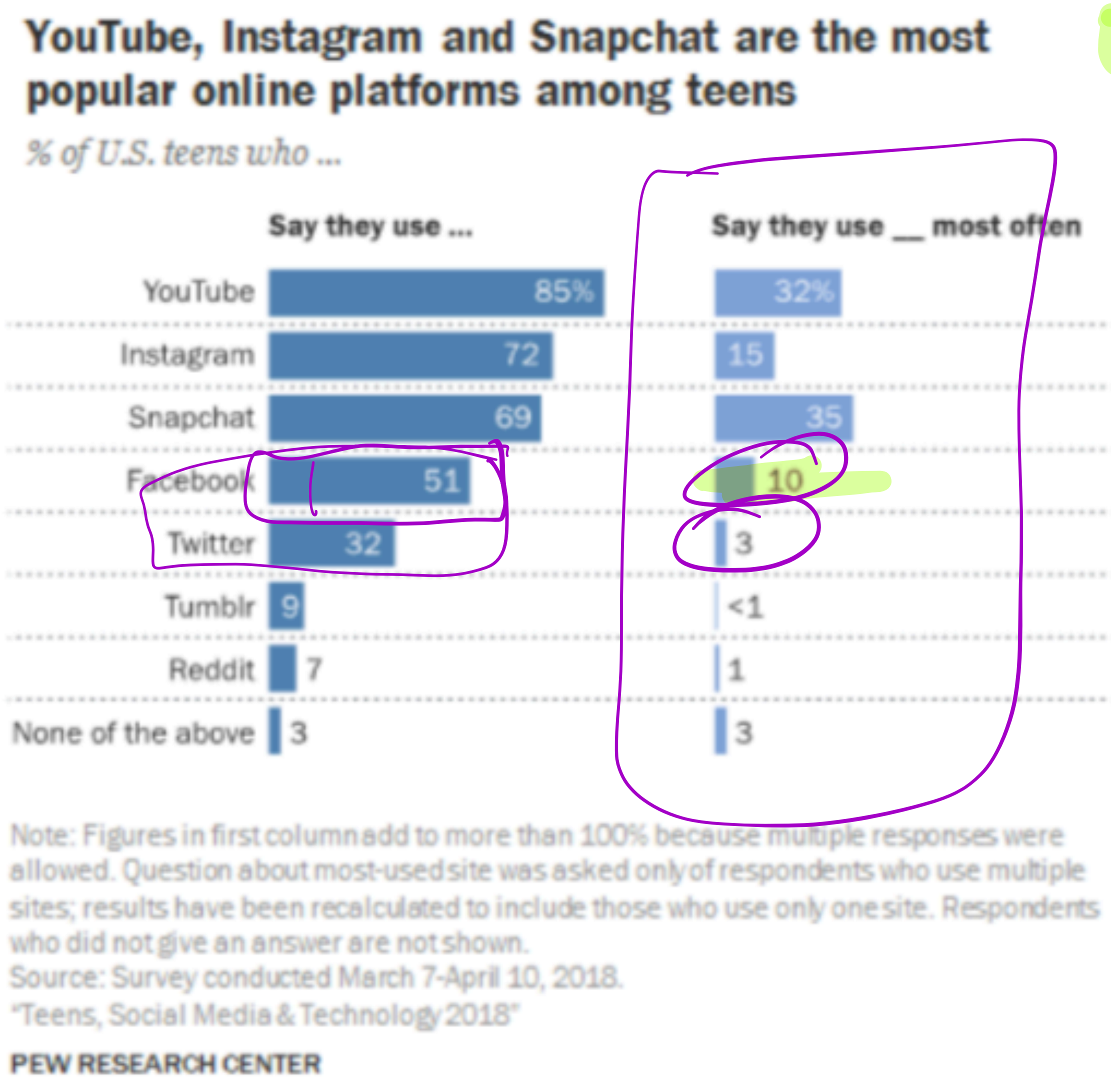
outcomes in BOTH A&B

- $P(B \mid B) = 1$

$P(A) \cap P(B)$ $P(A) \cancel{\le} P(\Omega)$

$A \subset \Omega$ $\quad P(A) \le P(\Omega)$ $\quad = A \cap B$ or AB

From Friday: Not equally likely outcomes



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

0.1

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

0.9 $= 1 - 0.1$

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

0.1+0.03 = 0.13 ✓

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

0.51-0.1 = 0.41

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

B

A

10/51 = 0.1/0.51 ≈ 0.2

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)} = \frac{10}{51} \approx 0.2$$

Notation: Intersections and Unions

When two events A ***and*** B ***both*** happen, we call this the ***intersection*** of A and B and write it as

$$A \text{ } and \text{ } B = A \cap B \text{ (also written as } AB\text{)}$$

When either A ***or*** B happens, we call this the ***union*** of A and B and write it as

$$A \text{ } or \text{ } B = A \cup B$$

If two events A and B ***cannot both occur*** at the same time, we say that they are *mutually exclusive* or *disjoint*.

$$A \cap B = \varnothing$$

Exercise from Friday

$\rightarrow$ 1,1   1,2 $\boxed{1,3 \quad 1,4 \quad 1,5 \quad 1,6}$

A six-sided fair die is rolled twice:

B $\rightarrow$ - If the first roll lands on 1, what is the chance that the second roll lands on a number bigger than 2? } A

$$P(A|B) = \frac{\#(A \& B)}{\#(B)} = \frac{4}{6}$$

$$P(A) = \frac{\#(2^{nd}\text{ roll} > 2)}{\#(\Omega)} = \frac{24}{36}, \quad P(A\&B) = \frac{\#(A \cap B)}{\#(\Omega)} = \frac{4}{36}$$

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

Try at home

Go for glory!

Post on Ed.

Die 2

	11	12	13	14	15	16
Die 1	21	22	23	24	25	26
	⋮					
	61	62	63	64	65	66

Rules that we used: Addition rule

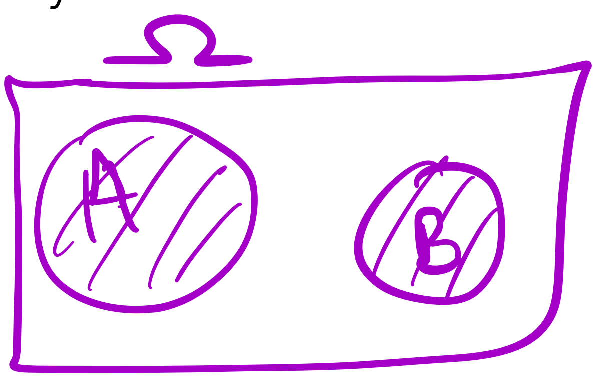
If all the possible outcomes are *equally likely*, then each outcome has probability *1/n*, where *n* = number of possible outcomes.

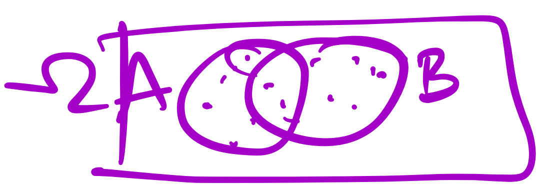
If an event A contains k possible outcomes,

then $P(A) = k/n$. $= \dfrac{\#(A)}{\#(\Omega)}$

Probabilities are between 0 and 1 (b/c $P(A)$ is a proportion)

If two events A and B don't overlap, then the probability of A or B = P(A) + P(B) (since we can just add the number of outcomes in one and the other, and divide by the number of outcomes in Ω)





$$P(A \text{ or } B) = P(A \cup B)$$
$$= P(A) + P(B)$$
$$= \frac{\#(A)}{\#(\Omega)} + \frac{\#(B)}{\#(\Omega)}$$

Rules of probability

Let's think about what rules we can lay down, based on what we have seen so far.

Let Ω be outcome space, $A \subset \Omega$

1) $P(A) \le P(\Omega) = 1$

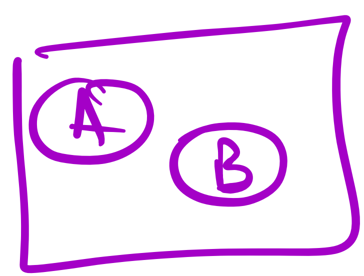
2) $P(A) = \dfrac{\#(A)}{\#(\Omega)} = \dfrac{k}{n}$ if all outcomes are eq. likely

3) $0 \le P(A)$

The intersection of A & B is empty

4) if A & B don't overlap $(\widetilde{AB} = \emptyset)$

then $\underbrace{P(A \text{ or } B) = P(A) + P(B)}_{\text{Addition Rule}}$



5) $P(A^c) = 1 - P(A)$

$(A^c \Leftrightarrow A' \Leftrightarrow \text{not } A)$

Origins of probability: de Méré's paradox

Questions that arose from gambling with dice.

Antoine Gombaud,
Chevalier de Méré

Blaise Pascal

Pierre de Fermat

The dice players
Georges de La Tour
(17th century)

De Méré's Paradox: in section tomorrow

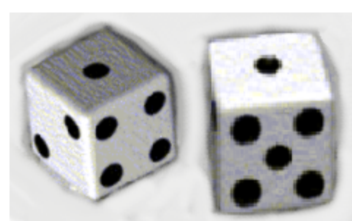
We can think about probability as a numerical measure of uncertainty, and we will define some basic principles for computing these numbers.

These basic computational principles have been known for a long time, and in fact, gamblers thought about these ideas a lot. Then mathematicians investigated the principles.

Famous problem: will the probability of **at least one six** in **four** throws of a die be equal to prob of **at least a double six** in 24 throws of a pair of dice.

de Méré thought that both these were the same $= \frac{2}{3}$

Note: single = die, plural = dice:



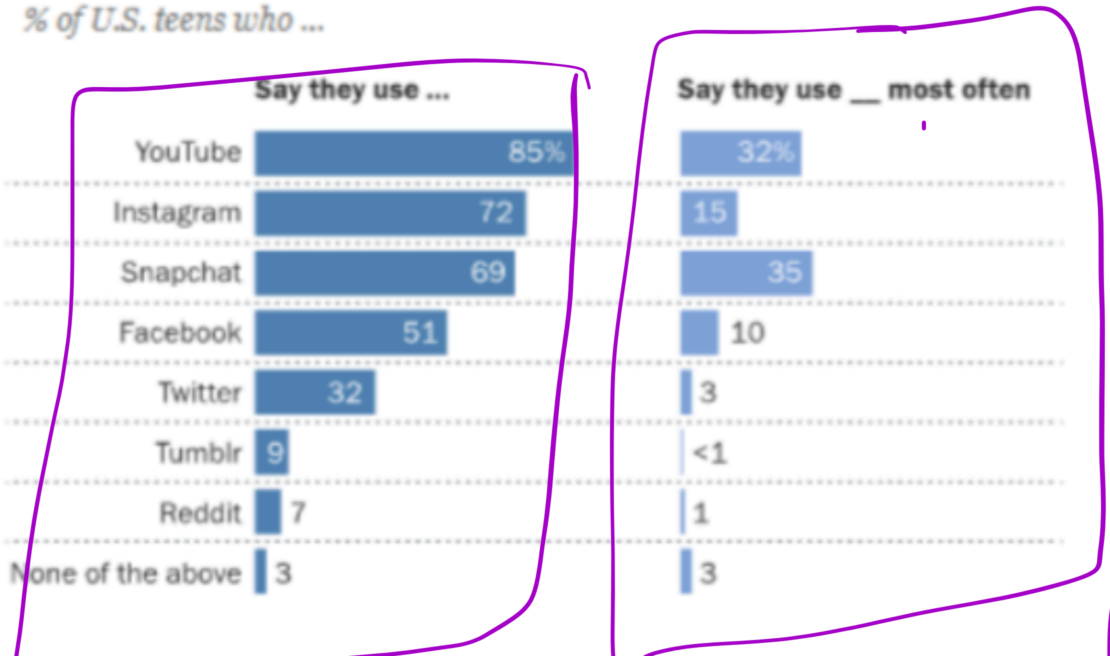
1) How?

2) What are the correct probs?

Section 1.2: Exact Calculations, or Bound?

YouTube, Instagram and Snapchat are the most popular online platforms among teens

% of U.S. teens who ...



Note: Figures in first column add to more than 100% because multiple responses were allowed. Question about most-used site was asked only of respondents who use multiple sites; results have been recalculated to include those who use only one site. Respondents who did not give an answer are not shown.
Source: Survey conducted March 7-April 10, 2018.
"Teens, Social Media & Technology 2018"

PEW RESEARCH CENTER

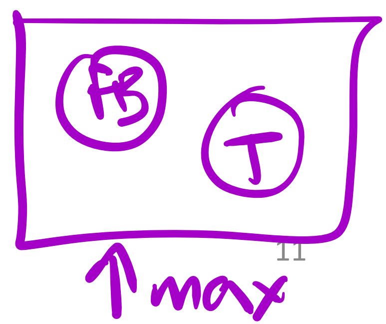
Recall #3 about FB or Twitter. What was the answer? What can you say about the chance that a randomly selected teen used FB or Twitter (not necessarily most often)?

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

$= 0.51 + 0.32 = 0.83$?!

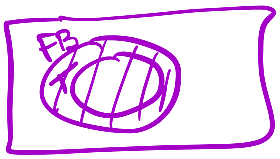
$P(FB \text{ or } T) \leq 0.83$

The prob of FB or T = 0.83 if $(FB) \cap (T) = \emptyset$

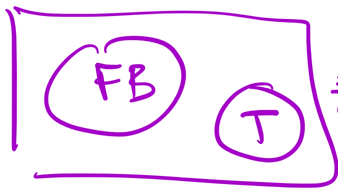


$$0.32 \le \underbrace{P(FB \text{ or } T)} \le 0.83$$ —

min

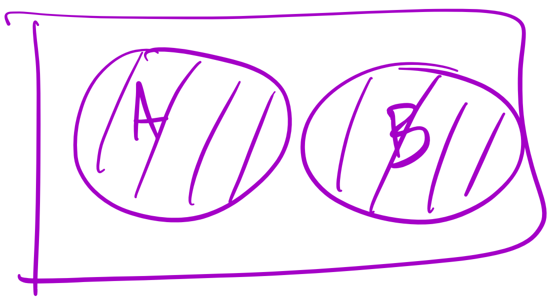


$$0 \le P(FB \cap T) \le P(T)$$

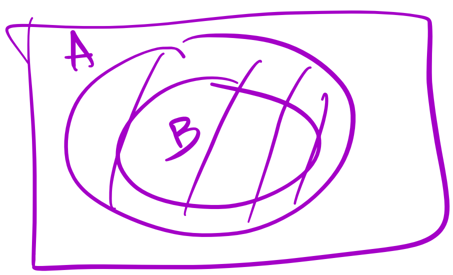


→ $FB \cap T = \emptyset$

⇐ Biggest overlap



$A \cup B$ is biggest



$A \cup B$ is smallest

Bounds

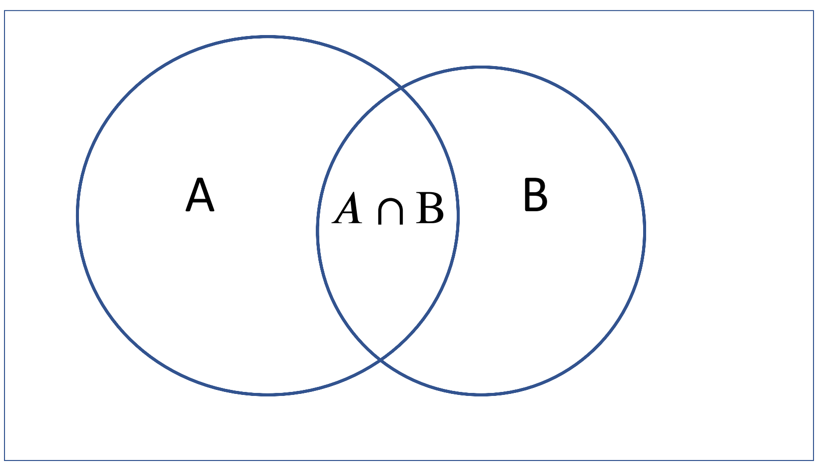
Overlap = Intersection

Biggest Overlap $\longrightarrow$ largest Intersection & smallest union.

When we get some information about the outcome or event whose probability we want to figure out, our outcome space reduces, incorporating that information.

$P(A \cup B)$ for mutually exclusive events

Bounds on probabilities of unions and intersections when events are ***not*** mutually exclusive.



$P(A) = 0.7,\ P(B) = 0.5$

$\_\_\_ \leq P(A \cup B) \leq \_\_\_$

$\_\_\_ \leq P(A \cap B) \leq \_\_\_$

Example with bounds

Let A be the event that you catch the bus to class instead of walking, P(A) = 70%

Let B be the event that it rains, P(B) = 50%

What is the chance of **at least** one of these two events happening?

What is the chance of **both** of them happening?

Exercise: what about if we have 3 events?

Let A be the event that you catch the bus to class instead of walking, P(A) = 70%

Let B be the event that it rains, P(B) = 50%

Let C be the event that you are on time to class, P(C) = 10%

What is the chance of **at least** one of these three events happening?

What is the chance of **all three** of them happening?