STAT 88: Lecture 1

Contents

Course resources:

- Course website: http://stat88.org
- Piazza: https://piazza.com/class/kdxsrxzejp77mg
- Gradescope: https://www.gradescope.com/courses/163710

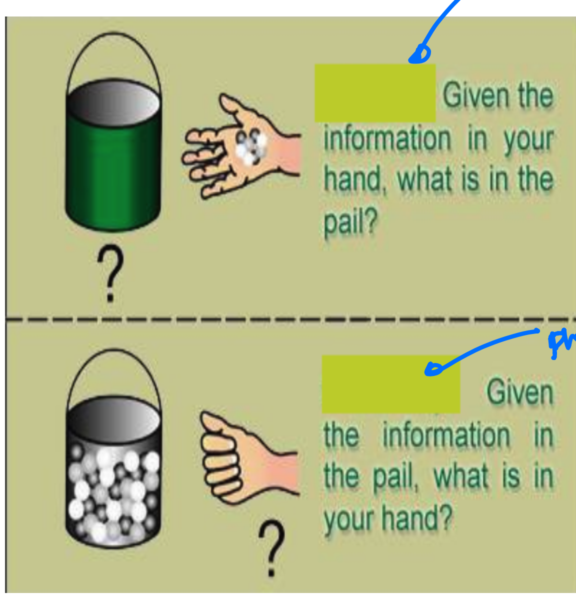
Ice-breaker: Introduce yourself to your classmates (breakout rooms).

Discuss which is probability and which is statistics.

Statistics: observe some events that happened → underlying random process

e.g.) A coin w/ unknown $p(\text{head})=\theta$. Toss it 10 times. You observe # heads $=6$. What is θ?

sample → pop.



pop → sample.

probability: underlying random process → figure out the chance of events

e.g.) You have a fair coin. Toss it 10 times. What is the chance that # head > 6?

1.1. Probabilities as Proportions

Probability is a numerical measure of uncertainty.

Terminology

- Experiment: activity that has a set of possible outcomes and involves chance

  ex Rolling a six-sided die once
- Outcome space or Sample space: the set of all outcomes of an experiment

  ex $\{1,2,3,4,5,6\} \rightarrow$ outcomes
- Event: a subset of outcome space

  ex The die shows a multiple of 3 $= \{3, 6\}$

Outcome space is denoted as Ω and $A \subseteq \Omega$ is an event.

Equally likely outcomes: for any $A \subseteq \Omega$, $P(A) = \frac{\#A}{\#\Omega}$.

Example: What is the chance that a die shows a multiple of 3?

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $A = \{3,6\}$
- $P(A) = \frac{2}{6} = \frac{1}{3}$

  $= \frac{\#A}{\#\Omega}$

Example: Deck of cards:

4 suits : clubs, diamonds, hearts, spades
13 ranks : Ace, 2–10, J, Q, K

52 cards

Suppose a deck of cards is shuffled and the top 2 cards are picked. What is the chance that you get at least one ace among the 2 cards? i.e.

$4*3 + 4*48 + 48*4 = ?$

$$A = \{(\text{ace}, \text{ace}), (\text{ace}, \text{nonace}), (\text{nonace}, \text{ace})\}.$$ What is $P(A)$?

$\Omega =$ any pairs of cards $= \{$(any card, any other card)$\}$

$52 * 51 = 2652.$

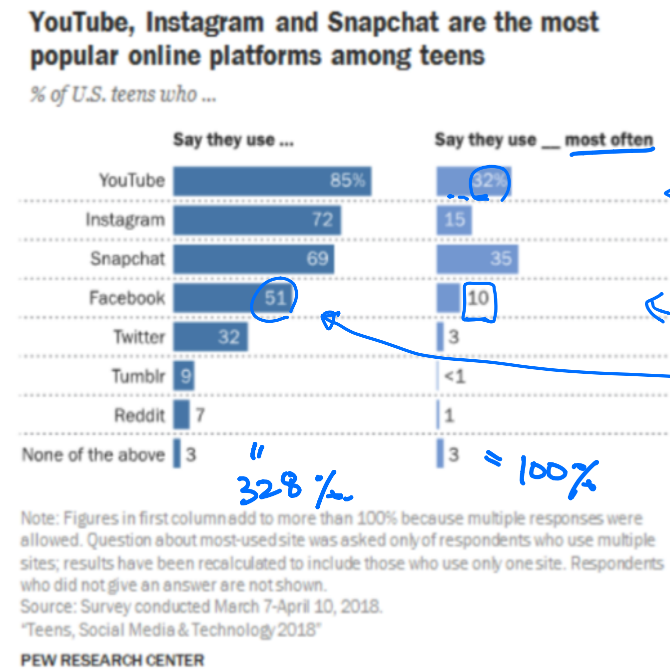
$$p(A) = \frac{\#A}{\#\Omega} = \frac{4*3 + 4*48 + 48*4}{2652.} = 396$$

(Textbook)

Example: Teen use of online formats:

Data

Each teen can pick multiple platform



Each teen picked only one platform

100 students

Wish to understand student usage of online platform such as Facebook or Twitter.

Are these columns distribution? The second is, first isn't b/c it adds up to more than 100%

Why the percents add up to more than 100 in the first column?

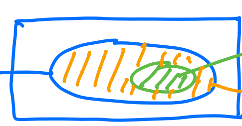
A teen can choose more than one answer.

What is the chance that the chosen teen used FB most often? What is the chance that the chosen teen did not use FB most often? 10%

100 − 10 = 90%

What is the chance that the student used Facebook but used some other platform more often?

FB user



FB most often

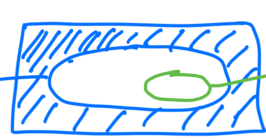
FB user but not FB favorite.

= 51 − 10 = 41 %.

Conditioning We are given a new information: Someone picks a teen at random from the population. We know that the teen used Facebook.

Given this information, what is the chance that the teen used Facebook most often?

FB user



FB most often

chance FB most often given FB user

$= P(\text{FB most often} \mid \text{FB user}) = \dfrac{10}{51} \approx 0.196$

P(FB most often)

P(FB user)

200 students

$\dfrac{200 \cdot 10\%}{200 \cdot 51\%}$

"given" =

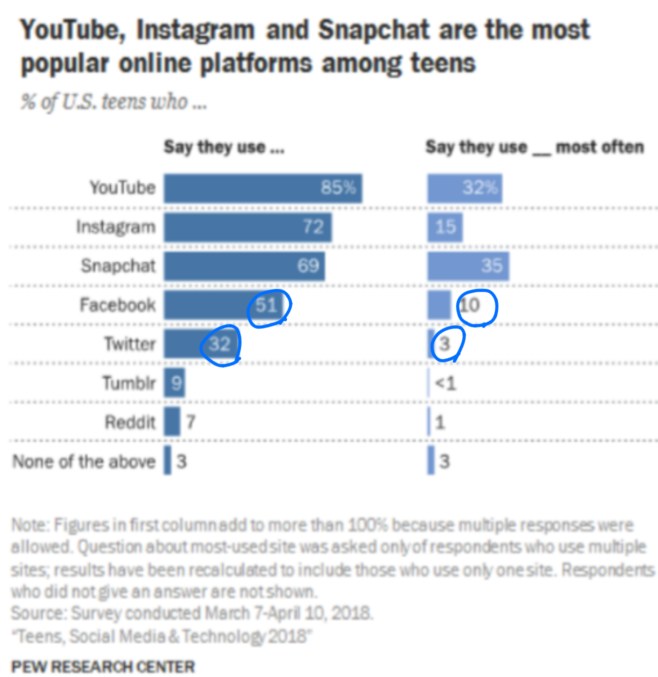
To find a conditional probability:

teens who used FB

- 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

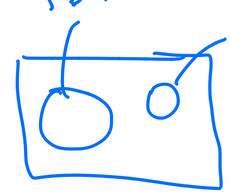
1.2. Exact Calculation or Bound

Sometimes we can only give a bound for probability of an event instead of the exact value.



FB favorite

tw favorite.

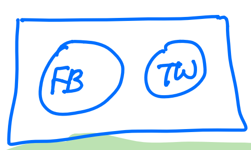


What is the chance that the selected teen used Facebook or Twitter most often?

10 + 3 = 13%

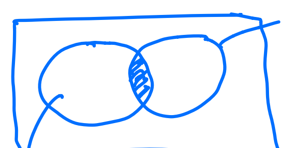
What is the chance that the selected teen used Facebook or Twitter? FB ∪ TW

- Upper bound?
- Lower bound?



51+32

= $\boxed{83\%}$

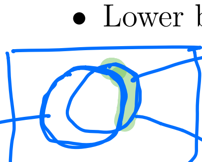


TW user

FB user

Facebook and Twitter

" FB ∩ TW

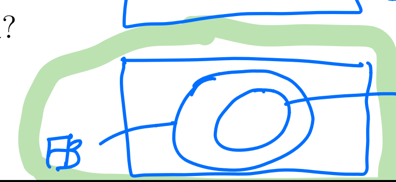


x

FB

" 51%

TW



TW

= $\boxed{51\%}$.

$\boxed{51+x}$ x=0

Extra problem: Two draws are made at random with replacement from the box

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

a) If you draw a 1, what is the chance that the second number is bigger than 2?

b) Find the chance the second number is bigger than twice the first.

(a) $\Omega = \{(1,1), (1,2), (1,3), \ldots, (1,10)\}$ $\quad (\#\Omega = 10)$

$A = \{(1, >2)\}$ $\quad (\#A = 8)$

$P(A) = \frac{\#A}{\#\Omega} = \frac{8}{10} = 80\%$

(b) $\Omega =$ all pairs of numbers $\quad (\#\Omega = 10 * 10 = 100)$

$$A = \left\{ \begin{array}{l} (1, >2) \\ (2, >4) \\ (3, >6) \\ (4, >8) \\ (5, >10) \end{array} \right\} \begin{array}{l} 8 \\ 6 \\ 4 \\ 2 \\ 0 \end{array}$$

$(\#A = 8+6+4+2+0 = 20)$

$P(A) = \frac{\#A}{\#\Omega} = \frac{20}{100} = 20\%$

$\cancel{(1,1), (2,2), (3,3), \ldots, (10,10)}$

$(1,3), (1,4), \ldots, (1,10)$

1.3. Fundamental Rules

Formally, probability is a function on events, $P : A \subseteq \Omega \mapsto [0, 1]$, satisfying the following 3 axioms:

1. $P(A) \geq 0$ for all $A \subseteq \Omega$.
2. $P(\Omega) = 1$.
3. If A and B are mutually exclusive, i.e. $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.