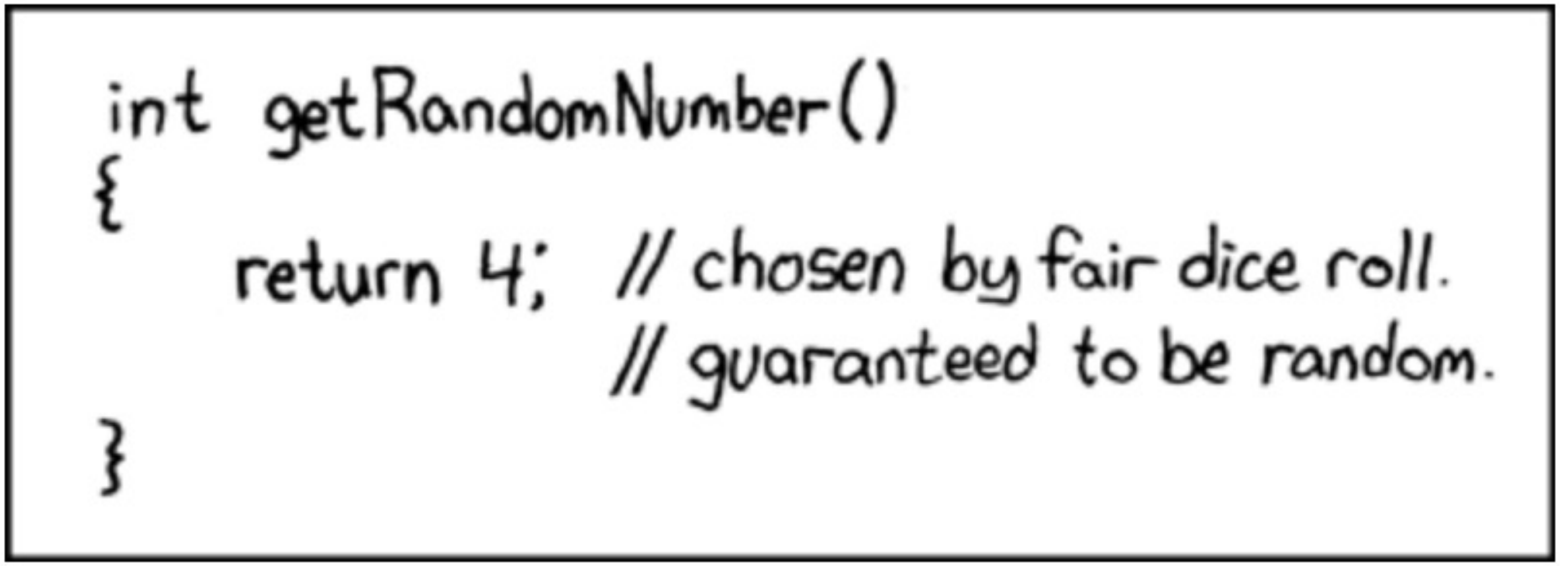
Stat 88: Probability and Statistics in Data Science



https://xkcd.com/221/

Lecture 2: 1/20/2022

Basics, Axioms of Probability, Intersections

1.2, 1.3, 2.1

Shobhana M. Stoyanov

Agenda

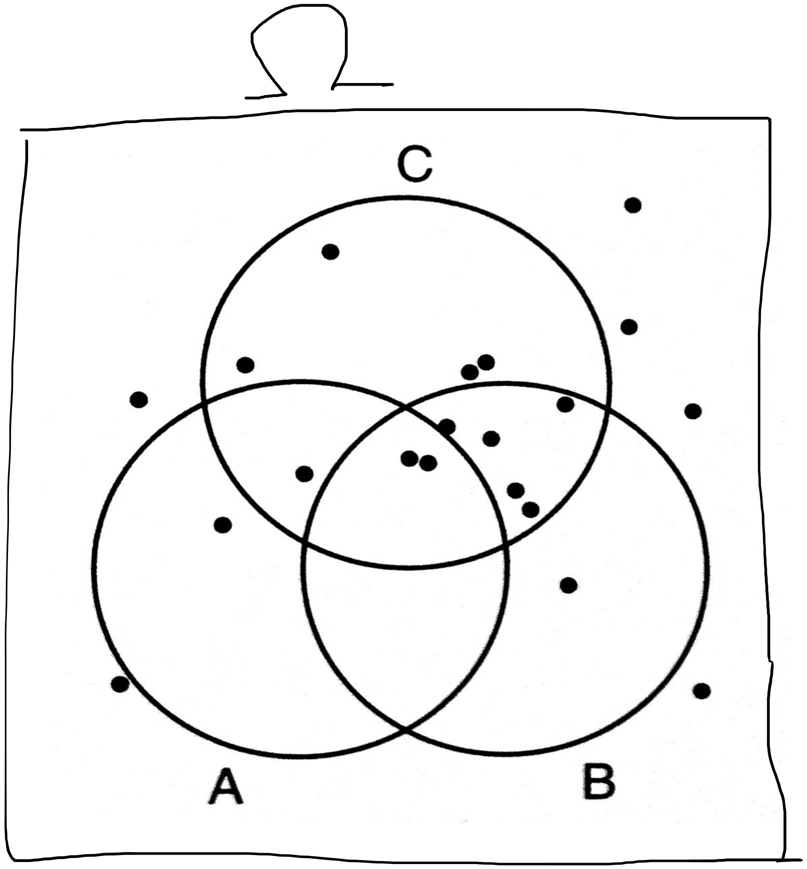
- Review the basics (probabilities as proportions)
- Go over exercises from Tuesday
- De Méré's paradox
- Section 1.2: Exact Calculation or Bound (go over the FB example from the text)
- Section 1.3: Fundamental Rules (Axioms)
- Section 2.1: The chance of an intersection

So far:

- Defined ***random experiments***, and their ***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 a collection of outcomes, so a subset of Ω.

- If all the possible outcomes are ***equally likely***, then each outcome has probability $1/n$, where $n = \#(\Omega)$ (number of outcomes in the sample space)
- Let $A \subseteq \Omega$, $P(A) = \frac{\#(A)}{\#(\Omega)}$
- Defined probabilities as ***proportions***

Venn Diagrams



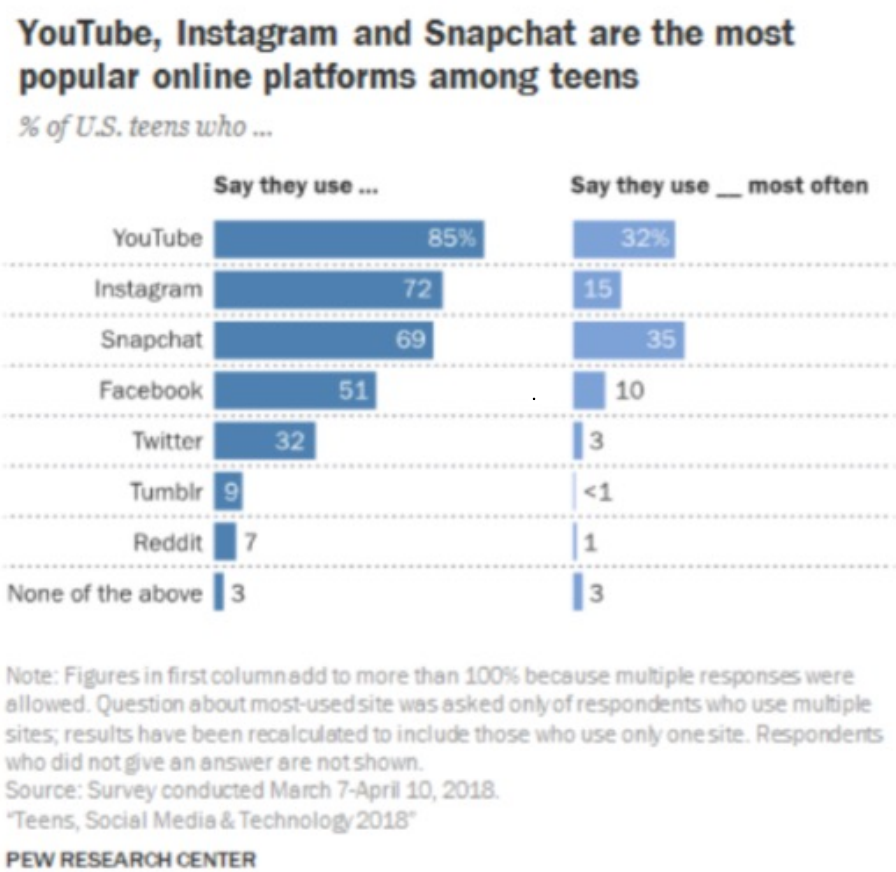
Consider the Venn diagram above. (The sample space consists of all the dots.) What is the probability of A? What about A or B? A or B or C?

Exercises assigned last lecture

1. Write out Ω if the action is rolling a pair of dice.
2. Write out Ω if the action is tossing 3 coins.
3. 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)?
4. If you deal 2 cards, what is the chance that at least *one* of them is a queen?
5. De Méré's paradox: Find the probability of at least 1 six in 4 throws of a fair die, and at least a double six in 24 throws of a pair of dice. (*postpone computation for a bit, but why would he think it should be the same?)*

Not equally likely outcomes (example from text)

- What if our assumptions of equally likely outcomes don't hold (as is often true in life, data are 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.

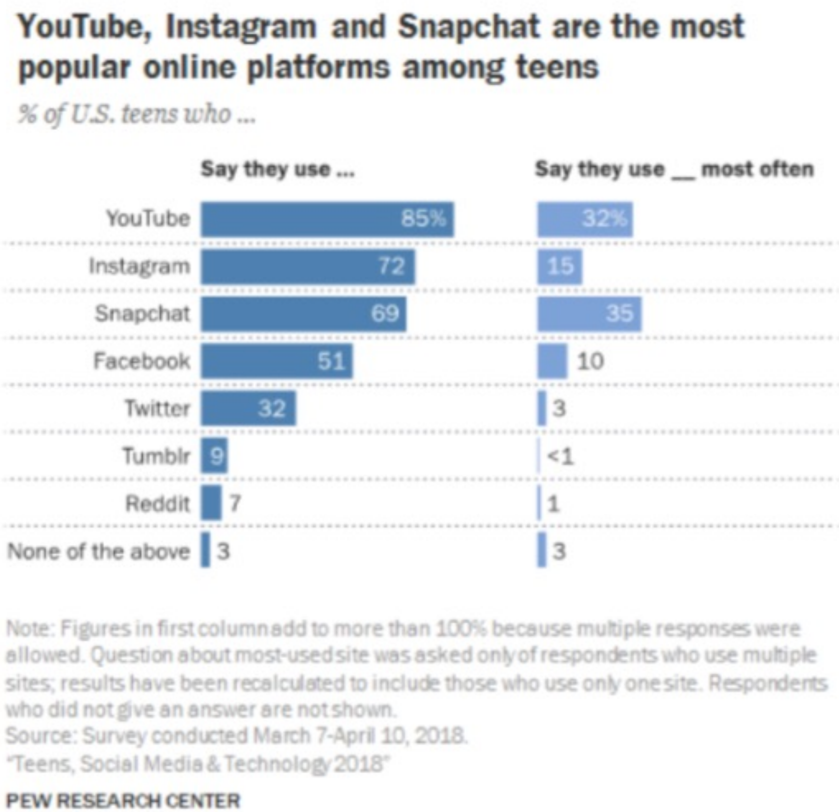


- What is the difference between the 2 charts?

- 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?

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

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

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

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

Recap:

- Sum of the probabilities of all the distinct outcomes should add to 1
- $0 \leq P(A) \leq 1, \mathrm{A} \subseteq \Omega$
- A *distribution* of the outcomes over different categories is when each outcome appears in one and only one category.
- Venn diagrams

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

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:

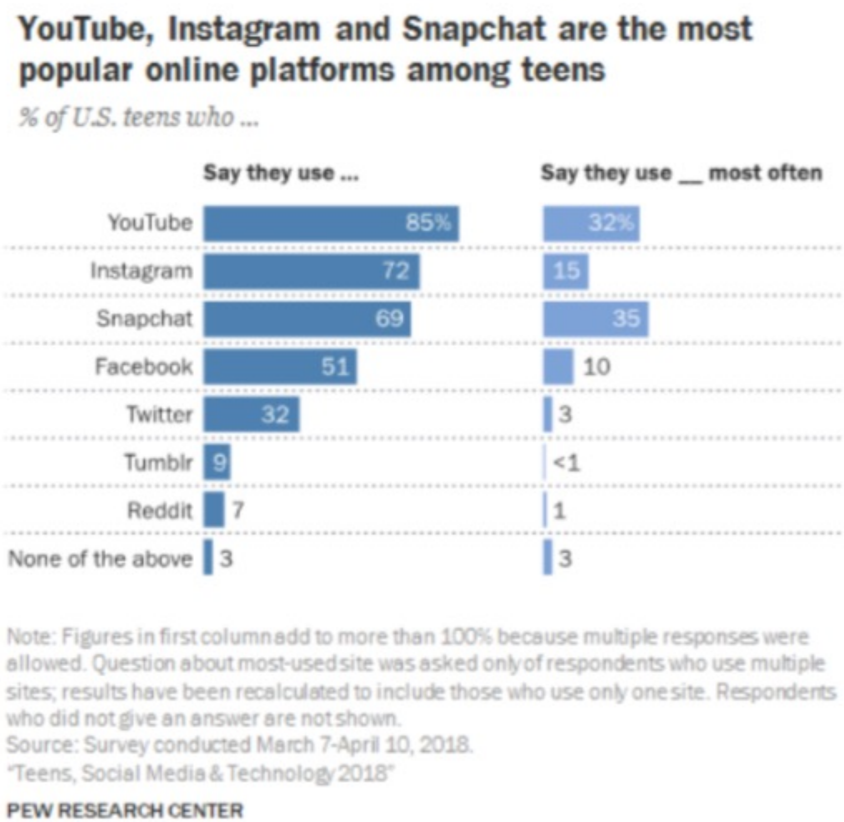
Conditional probability

- 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?

Exercise

- A ten-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 1?
  - 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 is the chance that FB *or* Twitter was a randomly selected teen's favorite?) What was the answer? What can you say about the chance that a randomly selected teen ***used*** FB or Twitter (not necessarily their favorite)?

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%
- 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?

Rules that we used:

- 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*.

- Probabilities are between 0 and 1

- 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)

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:

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?

Section 1.3: Fundamental Rules

- Also called "Axioms of probability", first laid out by Andrey Kolmogorov in 1933
- Recall Ω, the outcome space. Note that Ω can be finite or infinite.

- First, some notation:
- Events are denoted (usually) by A, B, C ...

- Note that Ω is itself an event (called the ***certain*** event) and so is the empty set (denoted $\emptyset$, and called the ***impossible*** event or the *empty set*)

- The ***complement*** of an event A is everything ***else*** in the outcome space (all the outcomes that are *not* in A). It is called "not A", or the complement of A, and denoted by A^c

Intersections and Unions

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

$$A\ and\ B = \mathrm{A} \cap B$$

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

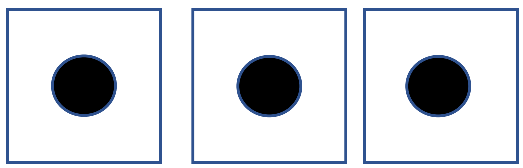
$$A\ or\ B\ =\ A\ \cup B$$

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

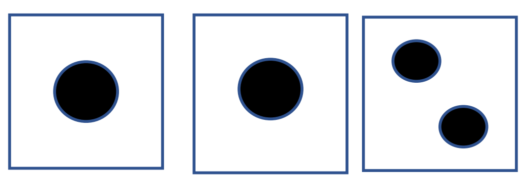
$$\mathrm{A} \cap B = \emptyset$$

Example of complements

- Roll a die 3 times, let A be the event that we roll an ace **each** time.
- A^C = **not** A, or not ***all*** aces. It is **not equal** to "never an ace".
- $A =$



- What about "not A"? Here is an example of an outcome in that set.



The Axioms of Probability

Think about probability as a ***function*** on ***events***, so input an event A, and output a number between 0 and 1, denoted by $P(A)$, satisfying the "**axioms**" below.

"subset of" "is in"

Formally: $A \subseteq \Omega, P(A) \in [0,1]$ such that

1. For every event $A \subseteq \Omega$, we have $0 \leq P(A) \leq 1$

2. The outcome space is certain, that is: $P(\Omega) = 1$

The Axioms of Probability

3. Addition rule: If two events are mutually exclusive, then the probability of their union is the sum of their probabilities:

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

The third axiom is actually more general and says:

If we have many* events that are *mutually exclusive* (no pair overlap), then the probability of their union is the sum of their probabilities.

* Possibly infinitely many

Example

- Toss a fair coin twice, and write out Ω. What is the chance of *both* coins landing the *same*?

Consequences of the axioms

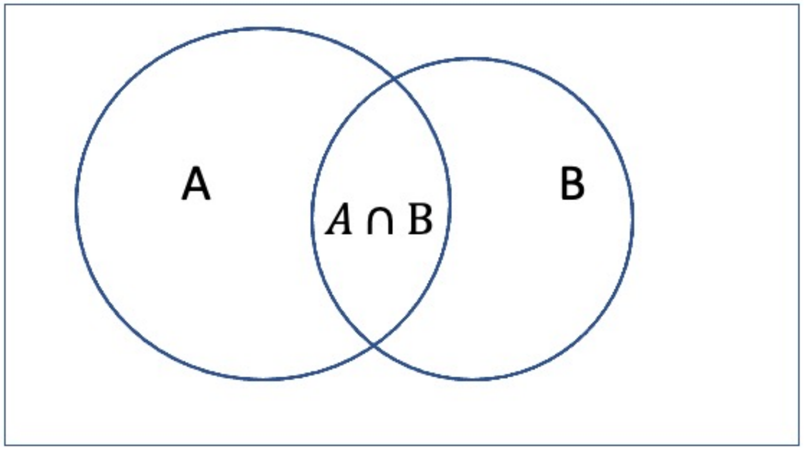
1. **Complement rule**: $P(A^c) = 1 - P(A)$

What is the probability of **not** rolling a pair of sixes in a roll of a pair of dice?

2. **Difference rule**: If $B \subseteq A$, then $P(A \setminus B) = P(A) - P(B)$ where $A \setminus B$ refers to the *set difference between* A *and* B*, that* is, all the outcomes that are A but not in B.

3. **Boole's (and Bonferroni's) inequality:** generalization of the fact that the probability of the union of A and B is at most the sum of the probabilities.

Exercise



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

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

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

De Morgan's Laws

- Exercise: Try to show these using Venn diagrams and shading:

1. $(A \cap B)^c = A^c \cup B^c$

2. $(A \cup B)^c = A^c \cap B^c$

§ 2.1: Probability of an intersection

- Say we have three colored balls in an urn (red, blue, green), and we draw two balls without replacement.
- Find the probability that the first ball is red, and the second is blue
- Write down the outcome space and compute the probability

- We can also write it down in sequence: P(first red, then blue) = P(first drawing a red ball)P(second ball is blue, given 1st was red)

Conditional probability and the multiplication rule

- Conditional probability written as $\boldsymbol{P(B|A)}$, read as "the probability of the event B, given that the event A has occurred"
- Chance that two things will both happen is the chance that the first happens, *multiplied* by the chance that the second will happen *given* that the first has happened.

- Let $A, B \subseteq \Omega, P(A) > 0, P(B) > 0$

- Multiplication rule:

$$P(AB) = P(A|B) \times P(B)$$

$$P(AB) = P(BA) = P(B) \times P(A|B)$$

Multiplication rule

$$P(AB) = P(A|B) \times P(B)$$

- Ex.: Draw a card at random, from a standard deck of 52
  - P(King of hearts) =?
- Draw **2** cards one by one, **without** replacement.
  - P(1st card is K of hearts)=
  - P(2nd card is Q of hearts| 1st is K of hearts) =
  - P(1st card is K of hearts AND 2nd is Q of hearts) =

De Méré's paradox:

Find the probability of at least 1 six in 4 throws of a fair die, and at least a double six in 24 throws of a pair of dice.

Addition rule:

- Addition rule: If *A* and *B* are ***mutually exclusive*** events, then the probability that **at least one** of the events will occur is the sum of their probabilities:

$$P(A \cup B) = P(A) + P(B)$$

- If they are not mutually exclusive, does this still hold?
- How do we write the event that "at least one of the events *A* or *B* will occur? How do we draw it?