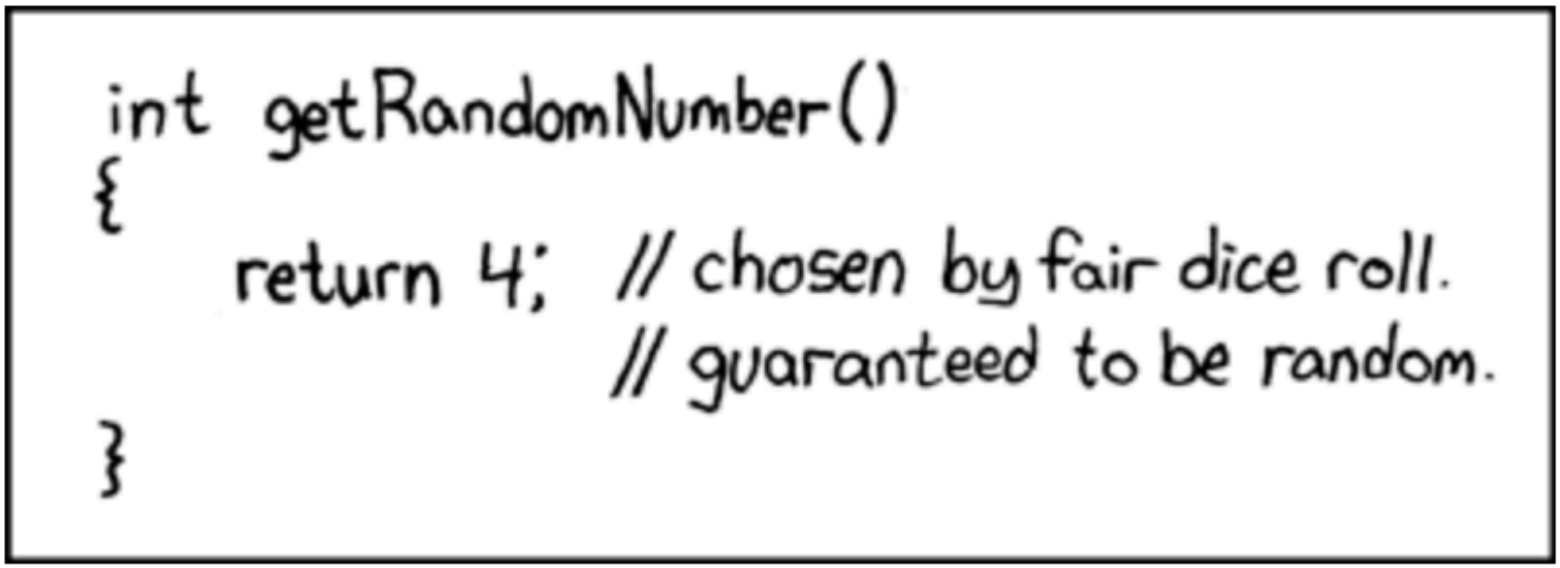
Stat 88: Probability and Mathematical Statistics in Data Science



https://xkcd.com/221/

Lecture 2: 1/22/2021

Axioms of Probability

Shobhana M. Stoyanov

Agenda

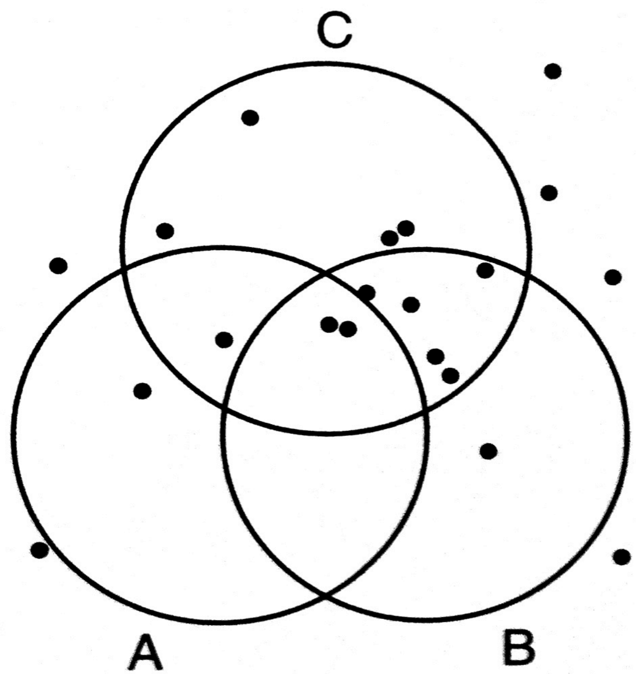
- The Basics:
  - Review of 1.1 + FB question, and conditional probability
  - Extra problems
  - Section 1.2: Exact Calculation or Bound
  - Section 1.3: Fundamental Rules

So far:

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

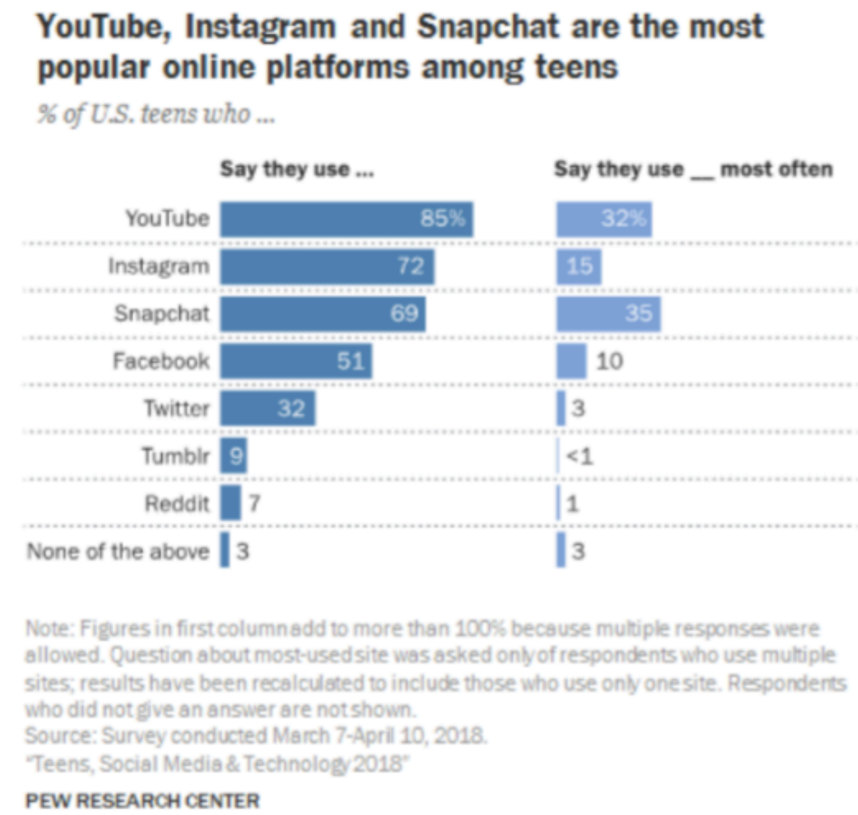
- When we get some information about the outcome or event whose probability we want to figure out, our outcome space reduces, incorporating that information.

Extra problem 1

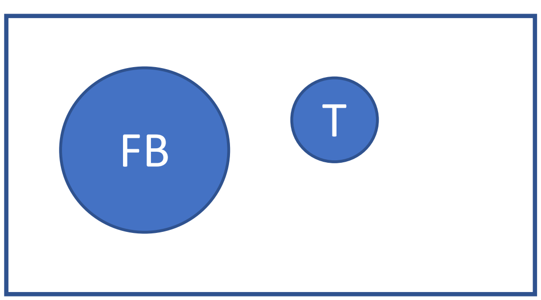


Consider the Venn diagram above. (The sample space consists of all the dots.) What is the probability of A? What about A or B? A or B or C?

Not equally likely outcomes



3. Venn diagram:



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%

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?
51%-10% = 41% of all teens used FB but not most often

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

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?

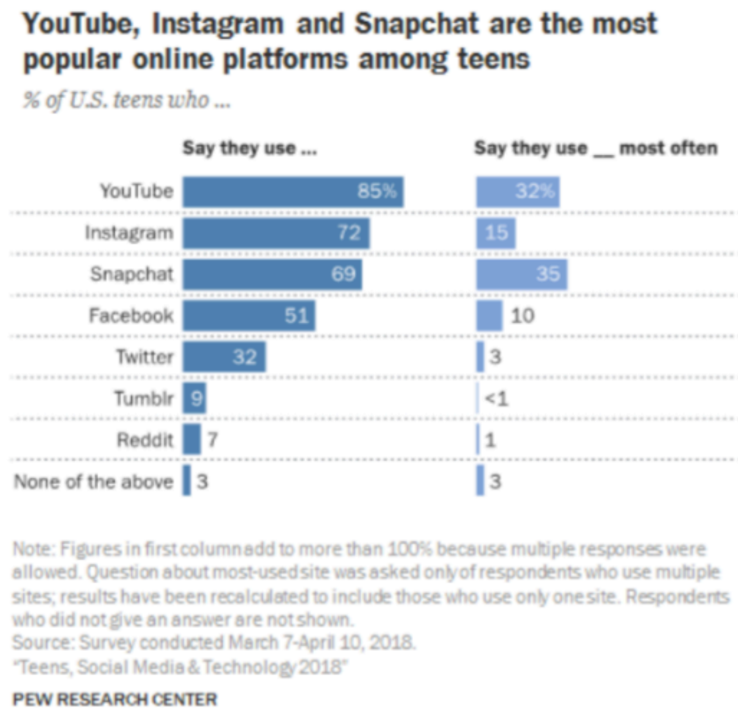
Extra problem 2

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

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?

Section 1.3: Fundamental Rules

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

- <u>First, some notation:</u>
- Events are denoted (usually) by $\mathrm{A}, \mathrm{B}, \mathrm{C} \ldots$

- 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

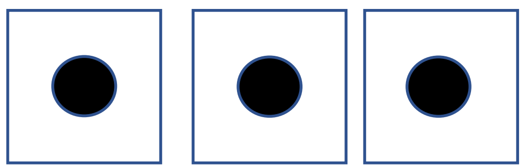
$$A \; or \; B \;=\; A \;\cup B$$

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

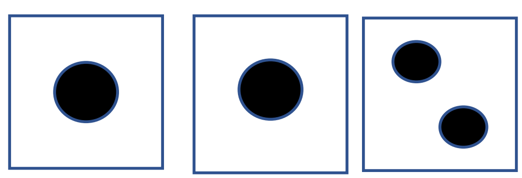
$$\mathrm{A} \cap B = \emptyset$$

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 put in an event A, and P(A) is a number between 0 and 1 satisfying the axioms below.

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

1. For every event $A \subseteq \Omega$, we have $P(A) \geq 0$
2. The outcome space is certain, that is: $P(\Omega) = 1$
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 infinitely many events that are *mutually exclusive* (no pair overlap), then the chance of their union is the sum of their probabilities.

Consequences of the axioms

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

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.

De Morgan’s Laws

- Try to show these using Venn diagrams and shading:

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

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