Introduction to Probability

Grinnell College

October 7, 2024

Review

A couple of weeks ago we spent some time making tables and using them to answer questions.

- What percent of Titanic passengers survived?
- What percent of Florida voters supported citizenship pathways for immigrants who entered the US illegally.
- What percent of 30-50 year olds had low job satisfaction?

Today's Outline

- continue to use tables of data.
- introduce probabilities
- probability math

What is Probability?

Probabilities are numbers between 0 and 1 that represent how likely (or unlikely) an <u>event</u> is to happen.

- closer to zero = more unlikely
- closer to one = more likely

When multiple events are equally likely, probability can be thought of as a fraction

$$\frac{\text{\# of ways an event can happen}}{\text{\# of all possible outcomes}}$$

Examples:

Flipping a coin: 1 heads out of 2 possibilities $\rightarrow$ prob. heads $= 1/2 = 0.5$
Probability of rolling an odd number on a 20-sided die?

Types of Probability

Subjective Probability:

- How likely an event is to happen based on someone's personal belief / experience / feelings
- Most likely different answers from different people
- Ex: prob. of a sports team winning their next game?

Types of Probability

Theoretical Probability:

- How likely an event is to happen based on formulas or assumptions about the event
- Common assumption: events are equally likely to happen
  - coin flips
  - dice rolling

Example: Suppose there are 20 marbles in a bag. 2 marbles are red, 6 are blue, and 12 are green.

- prob. of pulling red marble?
- prob. of blue?
- prob. of green?

Types of Probability

Empirical Probability:

- How likely an event is to happen based on collected data
- Sometimes we estimate the probability with data in the form of a table
- Ex: flip a coin 1000 times and find the 'empirical' probability of getting a Heads

Law of Large Numbers:
If you repeat trials a whole bunch (and they don't affect each other) then the empirical probability will converge to the "true" probability

Empirical Examples

A report published in 1988 summarizes results of a Harvard Medical School clinical trial determining effectiveness of asprin in preventing heart attacks in middle-aged male physicians

	Heart Attack		
Treatment	Attack	No Attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

Probability a randomly selected participant has a heart attack?

(Conditional) Probability a randomly selected participant in the placebo group has a heart attack?

(Conditional) Of those who had a heart attack, what would be the probability of a randomly selected participant being in the placebo group?

Notation

To save our selves some time, we often use some shorthand notation

P() is used to denote the probability of something, capital letters are quick ways to write down events

- P(patient having a heart attack) $\rightarrow$ P(heart attack) $\rightarrow$ P(H)
- read as "probability of patient having a heart attack"

Often times we may think of an event in terms of "success" (it happened) or "not success" (it did not happen)

Did patient have a heart attack?

- Yes = Success (unfortunate terminology)
- No = Failure

Probability Definitions

Marginal Probability – the probability of a single event

- P(H) = P(Heart attack)
- name comes from using the margins (totals) of a table

Union – Scenario where one event happens **or** another event happens (or both)

- We will always use 'inclusive or' meaning both events happening is allowed
- denoted P(A or B), $P(A \cup B)$

Intersection – Scenario where two events happen at the same time

- denoted P(A and B), $P(A \cap B)$

Probability Definitions

Conditional Probability – probability of event A occuring if event B has already happened

- P(A if B), P(A given B), P(A|B)
- ex: P(Heart attack if patient was given a placebo) = P(HA if placebo)
- we look at the 'given' variable first before calculating our probability

	Heart Attack		
Treatment	Attack	No Attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

Independence – When one event happening does not affect another

- P(A if B) = P(A)
- are Attack and Placebo independent?

Probability Definitions

Complements – when the event doesn't happen

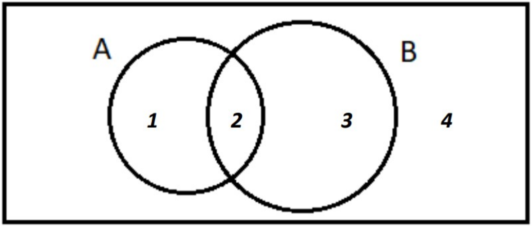
- P(A does not happen) = P(not A) = $P(A^C)$
- ex: P(no heart attack)

Disjoint Events – Events that cannot both happen

- ex: events "Attack" and "No Attack" are disjoint
- ex: events "Placebo" and "Aspirin" are disjoint
- ex: events "Placebo" and "Attack" are not disjoint
    - there were 189 instances of this happening

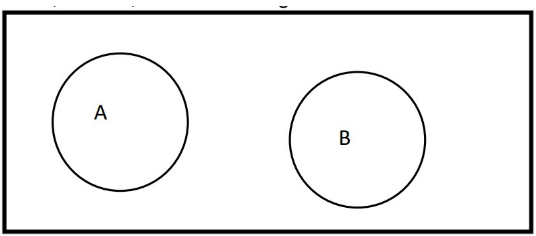
Venn Diagrams

Venn diagrams can be used as a way to help us think about these probabilities.



1. Which portion(s) of the Venn Diagram show above is the intersection of A and B?
2. Which portion(s) of the Venn Diagram show above are the union of A and B?
3. Which portion(s) of the Venn Diagram show above is the complement of A?

Venn Diagrams



Disjoint Events can also be thought of as events that do not overlap

- P(A and B) $= P(A \cap B) = 0$

Probability Rules / Formulas

Complement Rule

- P(not A) = 1 - P(A)

Additive Rule

- P(A or B) = P(A) + P(B) - P(A and B)
- Special: A and B are disjoint $\rightarrow$ P(A or B) = P(A) + P(B)

Multiplicative Rule

- P(A and B) = P(A if B)$\times$P(B) = P(B if A)$\times$P(A)
- Special: A and B are independent $\rightarrow$ P(A and B) = P(A)$\times$P(B)

Conditional Probabilities

- probability of A occurring if B has occurred
- P(A if B) = $\frac{\textbf{P(A and B)}}{\textbf{P(B)}}$