GENERAL ASSEMBLY

INTRODUCTION TO PROBABILITY

Matt Brems

DSI+

DATA SCIENCE PROCESS

1. Define problem. real world problem → data science problem
2. Gather data. survey experiment database .csv
3. Explore data. outliers missing data relationships
4. Model with data. linear regression neural network
5. Evaluate model.
6. Answer problem. data science answer → real world answer

DEFINITIONS

- **Experiment**: A procedure that can be repeated infinitely many times and has a well-defined set of outcomes.
- **Event**: Any collection of outcomes of an experiment.
- **Sample Space**: The set of all possible outcomes of an experiment, denoted $\mathcal{S}$.

EXAMPLES

- Experiment: Flip a coin twice.

- Sample Space $\mathcal{S}$:

$S = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$

- Event:

A = flipping at least one H = $\{\{H, H\}, \{H, T\}, \{T, H\}\}$

- Experiment: Rolling a single die. once.

- Sample Space $\mathcal{S}$:

$S = \{1, 2, 3, 4, 5, 6\}$

- Event:

$A = \{2\}$

C = rolling even

B = greater than 3

DEFINITIONS

Python

- **Set**: An unordered collection of distinct objects.
  - $\{Derek\ Jeter, \pi, ☺\}$

↳ unique

list: [ ]

set: { }

dict: {k1 : v1,
k2 : v2,
⋮ }

- **Element**: An object that is a member of a set.
  - Derek Jeter
  - π
  - ☺

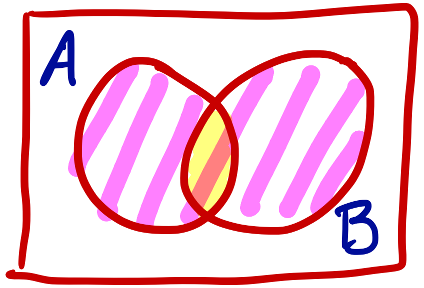
Example taken from Justin Gash's MAT 181 course at Franklin College.

SET OPERATIONS

"A union B"

- **Union**: $A \cup B$ = the set of elements in A or B (or both!)
- **Intersection**: $A \cap B$ = the set of elements in A and B

"A intersect B"



XOR

- Example:
  - A = even numbers between 1 and 10 = {2,4,6,8}
  - B = prime numbers between 1 and 10 = {2,3,5,7}
  - $A \cup B = \{2,4,6,8\} \cup \{2,3,5,7\} = \{2,3,4,5,6,7,8\}$
  - $A \cap B = \{2\}$

PROBABILITY – PRACTICE

- A = “a U.S. birth results in twin females”
- B = “a U.S. birth results in identical twins”
- C = “a U.S. birth results in twins”

- In words, what does $P(A \cap C)$ mean?

the probability that a US birth results in twin females.

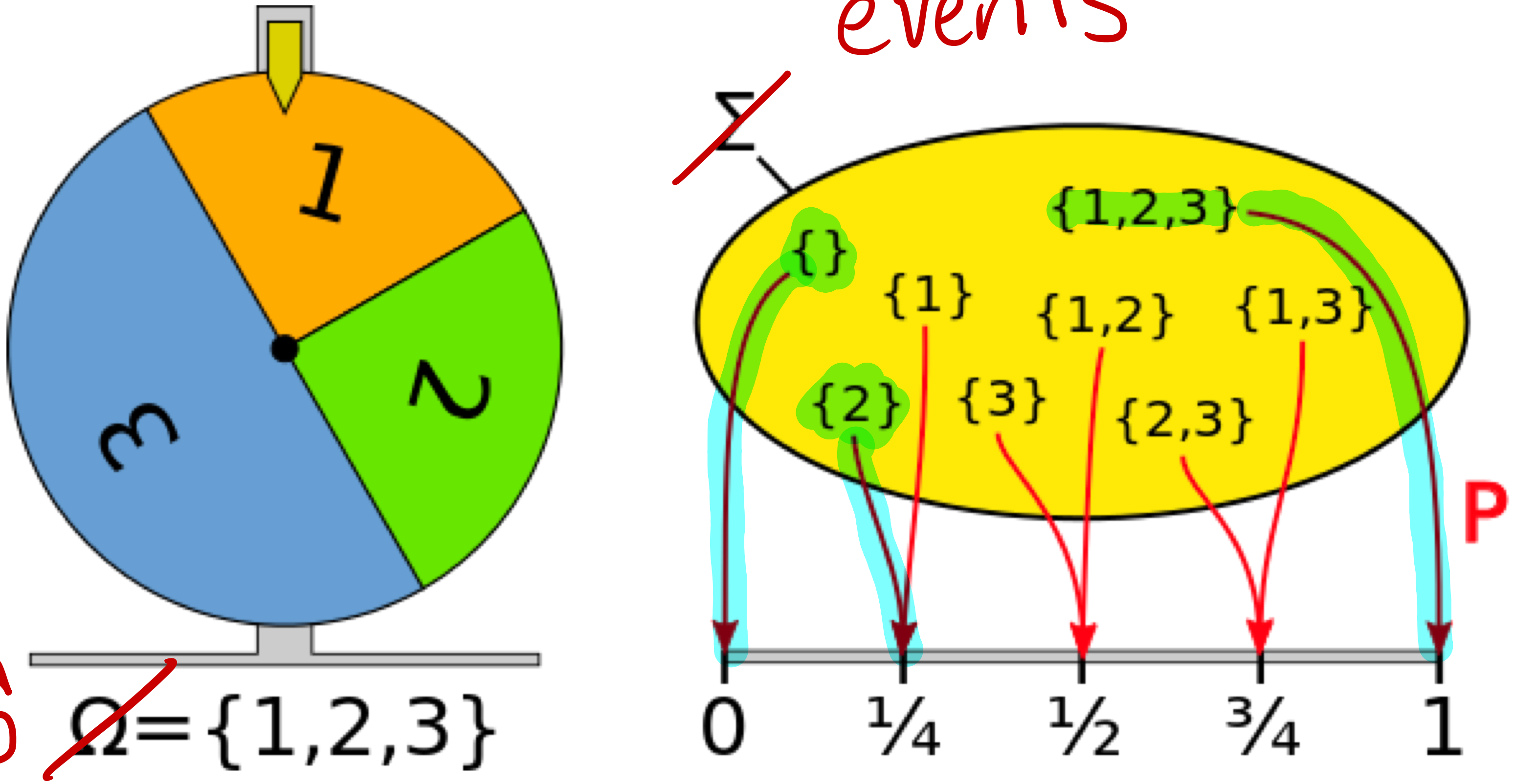
- In words, what does $P(A \cap B \cap C)$ mean?

the probability that a US birth results in identical twin females.

PROBABILITY BASICS

Experiment: spin the spinner

events



S

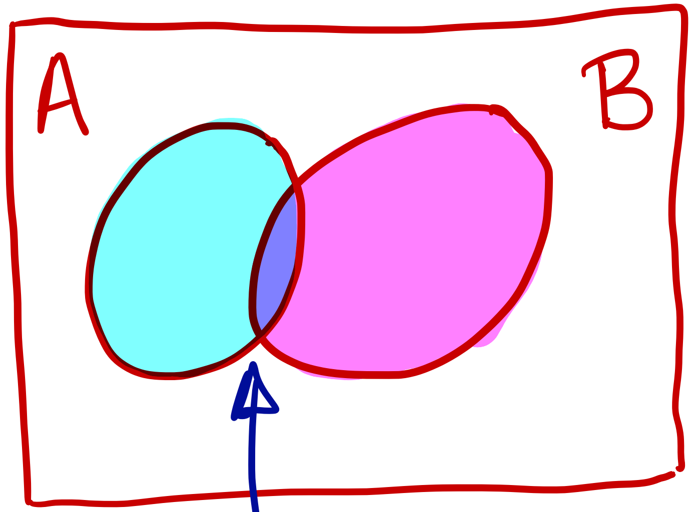
sample space

https://commons.wikimedia.org/wiki/File:Probability-measure.svg

PROBABILITY RULES

- $P(A \cup B) = P(A) + \boxed{P(B)} - \boxed{P(A \cap B)}$
  - Venn diagrams can help to illustrate this – but remember that Venn diagrams are not proofs!
  - If A and B are disjoint, then $P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B)$.

cannot happen together



covered twice!

PROBABILITY RULES

$\frac{1}{3}$ $\frac{1}{6}$

independence: $P(A|B) = P(A)$

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - Note: $A|B$ means "A given B" or "A conditional on the fact that B happens."

A = roll a 2 B = roll even #

$$P(2 \mid \text{even}) = \frac{P(2 \cap \text{even})}{P(\text{even})} = \frac{P(2)}{P(\text{even})} = \frac{1/6}{1/2} = \frac{1}{3}$$

$\{1, 2, 3, 4, 5, 6\}$

PROBABILITY RULES

infant 2 dies | infant 1 dies

infant 1 dies

$P(A \cap B) = P(A)P(B)$

↳ only true if A, B are independent.

- $P(A \cap B) = P(A|B)P(B)$
  - We just took the last rule and multiplied both sides by $P(B)$.

$$P(B) \times P(A|B) = \frac{P(A \cap B)}{\cancel{P(B)}} \times \cancel{P(B)}$$

  - We can rearrange these, as well! $P(B \cap A) = P(B|A)P(A)$

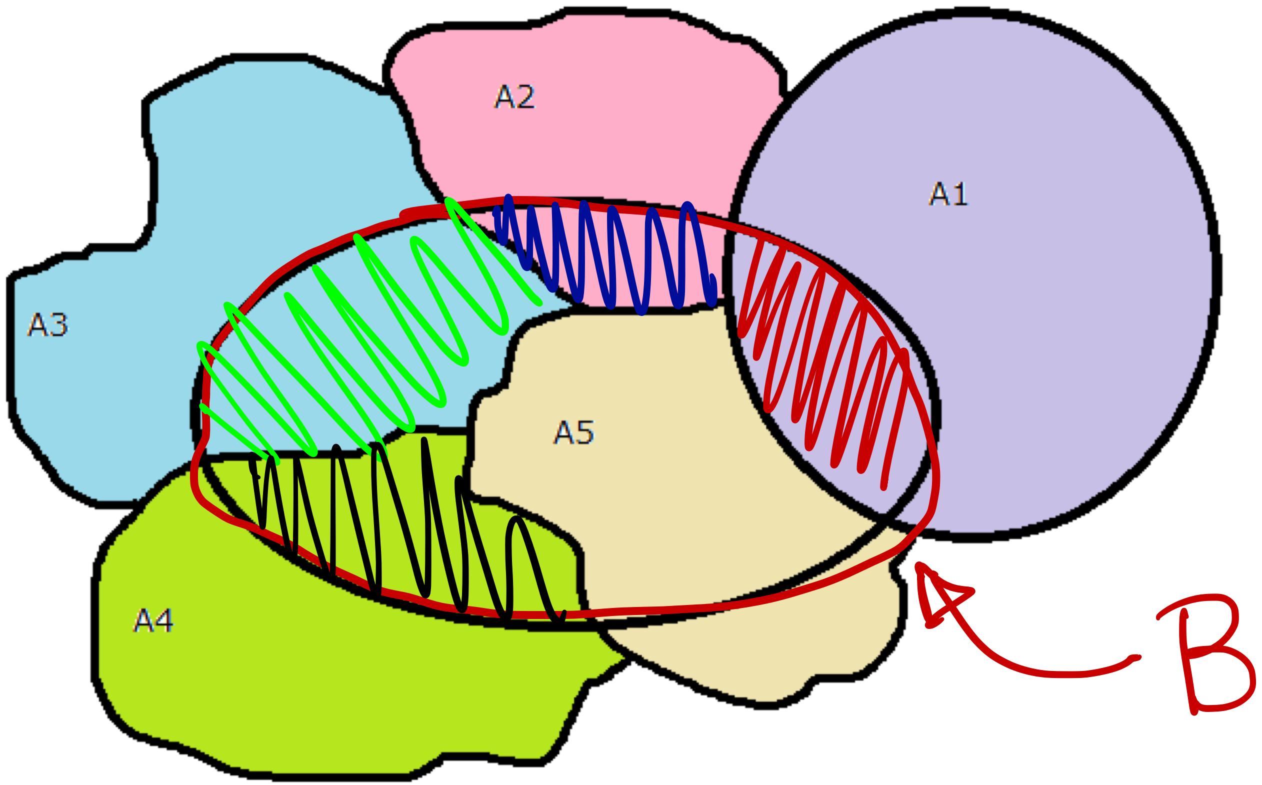
$P(\text{milk} \cap \text{cookies}) = P(\text{cookies} \cap \text{milk})$

  - This isn't limited to two events: $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$

PROBABILITY RULES

- $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$
  - "Law of Total Probability"

$P(B \cap A_1)$
$+ P(B \cap A_2)$
$+ \ldots$



PROBABILITY RULES – SUMMARY

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

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$

- $P(A \cap B) = P(A|B)P(B)$

- $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$

PRACTICE: INTERVIEW QUESTION

- There are 24 balls in a bucket: 12 red and 12 black.
- If you draw one ball, then draw a second ball, what is the probability of drawing two balls of the same color?

without replacing ball 1!

$$P(2 \text{ balls, same color}) = P(R \cap R \cup B \cap B)$$

$$= P(R \cap R) + P(B \cap B) - \cancel{P(R \cap R \cap B \cap B)} \quad 0$$

$$= P(R \cap R) + P(B \cap B)$$

$$= P(R)P(R|R) + P(B)P(B|B)$$

$$= \frac{12}{24} \cdot \frac{11}{23} + \frac{12}{24} \cdot \frac{11}{23}$$

$$\frac{11}{23} \approx 0.478$$

Interview question shared by DSI alum interviewing with Outbound Engine.

WHEN BY HAND IS TOUGH…

- Oftentimes, we won't evaluate probabilities by hand.
  - It's still very important to understand the ideas behind probability – as we move forward, it's critical to:
    - a) know probability's relationship with statistics and machine learning.
    - b) identify potentially bad assumptions. independence

- We often think of probability as how frequently an event occurs.
  - We can use simulations to give us a good approximation of the true probability of some event.

GENERAL ASSEMBLY

SUPPLEMENTAL SECTION

BAYES' THEOREM

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

WHAT IS $P(A)$?

- We've talked a lot about probabilities of certain events, but what does this actually mean?

- There are two broad classes of probabilistic interpretations.

TWO INTERPRETATIONS OF $P(A)$

- In the long run, how many times will A occur relative to how many times we conduct our experiment?

$$P(A) = \lim_{\#\, of\ exp's \to \infty} \frac{\#\, of\ times\ A\ occurs}{\#\, of\ experiments}$$

$$P(heads) = \lim_{\#\, of\ coin\ tosses \to \infty} \frac{\#\, of\ heads}{\#\, of\ coin\ tosses}$$

- This is called the **frequentist** interpretation of probability.

TWO INTERPRETATIONS OF $P(A)$

- What is one's degree of belief in the statement A, possibly given evidence?

$$P(A) = \text{“How likely is it that } A \text{ is true?”}$$

$$P(heads) = \text{“How likely is it that I flip a heads?”}$$

- This is called the **Bayesian** interpretation of probability.

TWO INTERPRETATIONS OF $P(A)$

- Neither interpretation of P(A) is more or less correct.
- However, these different interpretations can give rise to different ways of analyzing our data, as we'll see later!