# Introduction to Probability

*Matt Brems*

GENERAL ASSEMBLY

# Data Science Process

1. Define problem.
2. Gather data.
3. Explore data.
4. Model with data.
5. Evaluate model.
6. Answer problem.

# Definitions

- **Experiment**: A procedure that can be repeated an infinite number of times and has a well-defined set of outcomes.

- **Sample Space**: The set of all possible outcomes of an experiment, usually denoted $S$.

- **Event**: Any collection of outcomes of an experiment.

# Examples

- **Experiment**: Flip a coin twice.
- **Sample Space**:
- **Event**:

- **Experiment**: Roll one die.
- **Sample Space**:
- **Event**:

# Definitions

- **Set**: An unordered collection of distinct objects.
  - **{ Caroline, π, sweaterdresses }**

- **Element**: An object that is a member of a set.
  - **Caroline**
  - **π**
  - **sweaterdresses**

# Set Operations

- **Intersection**: A ∩ B = the set of elements in set A **and** set B
- **Union**: A ∪ B = the set of elements in set A **or** in set B

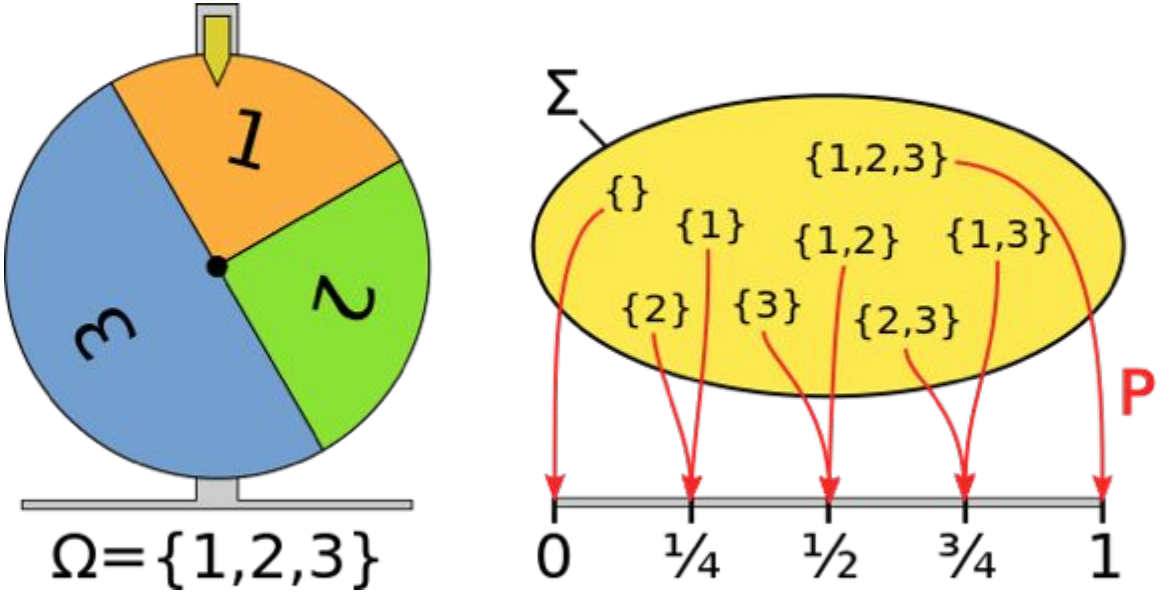
Example:

- A = {2, 4, 6, 8}
- B = {2, 3, 5, 7}
- A ∩ B =
- A ∪ B =

 | 

# Probability

We are often interested in the probability of some event(s) occurring.




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

We write P(A) to mean the probability that event A occurs.

# 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 $P(A \cup B)$ mean?

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

Example adapted from *Statistical Inference*, 2nd Ed., Casella & Berger

# Probability Rules

When trying to find the probability of a complex event, it's not straightforward.

- There are 12 red and 12 black balls. If you draw one ball, then a second ball without replacing the first, what is the probability they are the same color?
- Suppose you roll three dice. What is the probability that the three dice are rolled in increasing order?
- You call 3 friends of yours in Seattle and ask each independently if it's raining. Each of your friends tells you the truth ⅔ of the time. All 3 friends tell you it is raining. Based on historical evidence, it rains ¼ of the time in Seattle. What is the probability that it's actually raining in Seattle right now?

There are three probability rules that come in handy.

# Probability Rule 1: P(A ∪ B)

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

- Venn diagrams can help *illustrate* this, but Venn diagrams are not formal proofs!

# Probability Rule 2: P(A | B)

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

- A | B means "A given B" or "A conditional on the fact that B happens."

# Probability Rule 3: $P(A \cap B)$

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

- We just took the last rule and multiplied both sides by P(B).
- We can rearrange these as well: $P(B \cap A) = P(B \mid A) * P(A)$
- This isn't limited to two events: $P(A \cap B \cap C) = P(A \mid B \cap C) * P(B \mid C) * P(C)$

# Probability Rule 4: A special case of $P(A \cap B)$

- **When $P(A|B) = P(A)$**, we say that events A and B are **independent** of each other.
  - Put another way, whether or not B happens does not affect the probability that A happens!

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

# Probability Practice

- There are 12 red and 12 black balls. If you draw one ball, then a second ball without replacing the first, what is the probability that they are the same color?

# When by hand is tough...

- Oftentimes, it’s challenging to evaluate probabilities by hand.

- But it’s important to understand the ideas behind probability!
  - For example: when we want to build models, are two events independent of one another?

- We often think of probability as how frequently an event occurs.
  - We can use computer simulations to give us a good approximation of the true probability of some event.
  - This lets us “check our work” or to tackle harder probability problems!