

INTRODUCTION TO PROBABILITY

Matt Brems

DSI+

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 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} :
 - Event:
- Experiment: Rolling a single die.
 - Sample Space \mathcal{S} :
 - Event:

DEFINITIONS

- **Set:** An unordered collection of distinct objects.
 - $\{Derek Jeter, \pi, \text{☺}\}$
- **Element:** An object that is a member of a set.
 - Derek Jeter
 - π
 - ☺

SET OPERATIONS

- **Union:** $A \cup B$ = the set of elements in A or B
- **Intersection:** $A \cap B$ = the set of elements in A and B
- Example:
 - A = even numbers between 1 and 10 = $\{2,4,6,8\}$
 - B = prime numbers between 1 and 10 = $\{2,3,5,7\}$
 - $A \cup B$ =
 - $A \cap B$ =

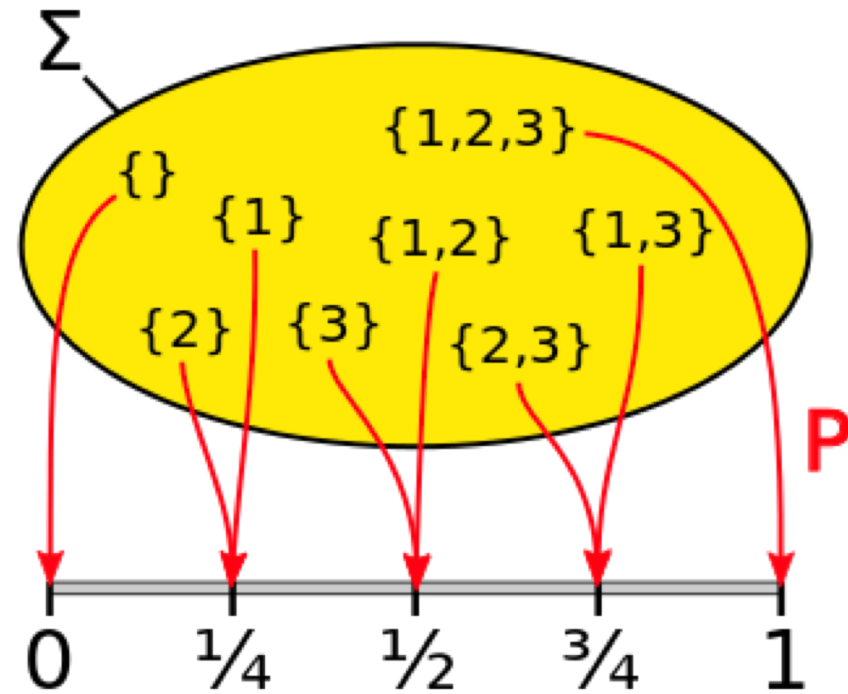
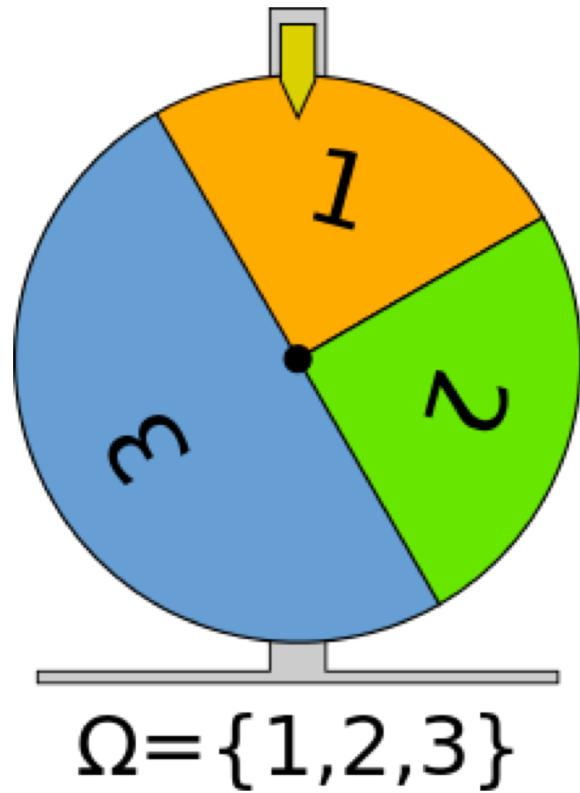
PROBABILITY – PRACTICE

- 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?

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

PROBABILITY BASICS



PROBABILITY RULES

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

PROBABILITY RULES

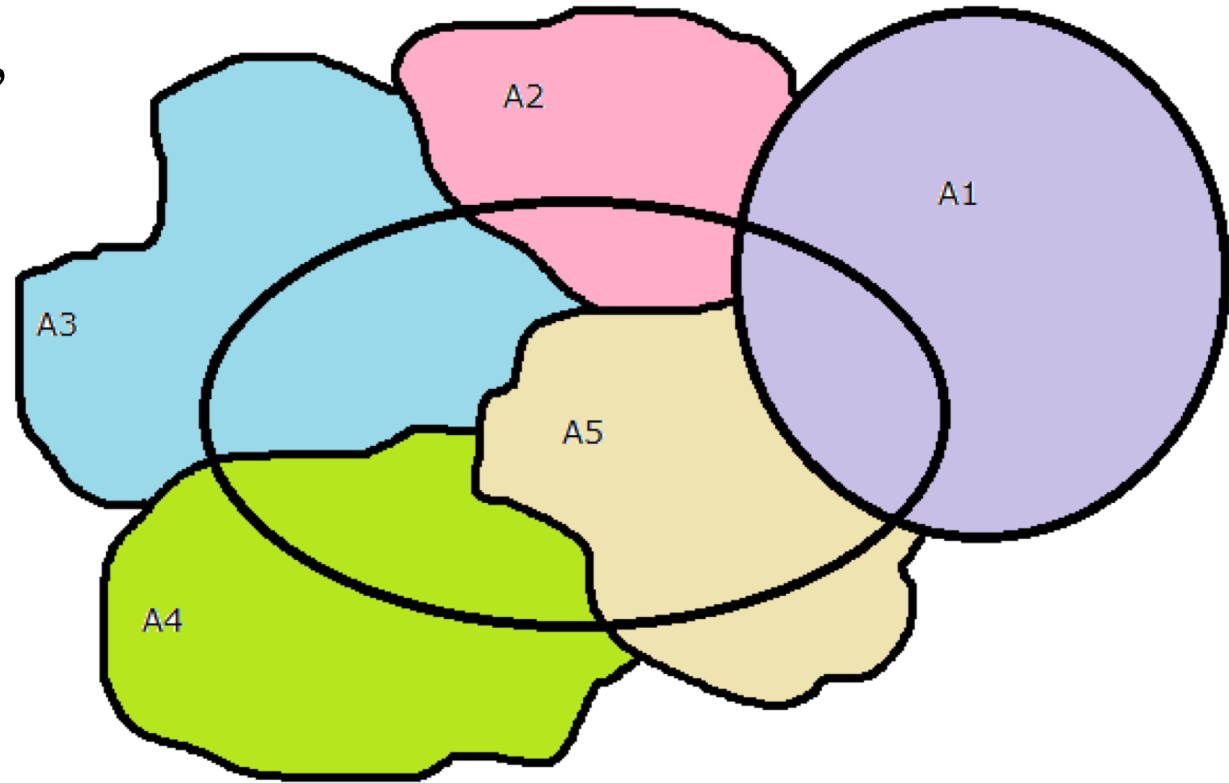
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
 - Note: $A|B$ means “ A given B ” or “ A conditional on the fact that B happens.”

PROBABILITY RULES

- $P(A \cap B) = P(A|B)P(B)$
 - We just took the last rule and multiplied both sides by $P(B)$.
- $P(B \cap A) = P(B|A)P(A)$
 - We can rearrange these, as well!
- 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”



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?

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.
- 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.

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?

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

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

- 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?

$P(A)$ = “How likely is it that A is true?”

$P(heads)$ = “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!