PROBABILITY

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 $\mathcal{S}$.

DEFINITIONS

- Experiment: A procedure that can be repeated infinitely many times and has a well-defined set of outcomes.

  what is the weather today?

  flip a coin

- Event: Any collection of outcomes of an experiment.

  Weather: Rainy, Snowy, Clear

  "not rainy" ≡ snowy or clear

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

$$S = \{\text{Rainy}, \text{Snowy}, \text{Clear}\} \qquad S = \{H, T\}$$

EXAMPLES

- Experiment: Flip a coin twice.
- Sample Space $\mathcal{S}$:
  $\{\{H,T\}, \{T,H\}, \{H,H\}, \{T,T\}\}$
- Event:

flip no heads $\equiv \{T,T\}$

do not flip the same

$\equiv \{\{H,T\}, \{T,H\}\}$

- Experiment: Rolling a single die.
- Sample Space $\mathcal{S}$: $\{1,2,3,4,5,6\}$
- Event:

even $\equiv \{2,4,6\}$

$3 \equiv \{3\}$

prime $\equiv \{2,3,5\}$

less than 4

$\downarrow$

$\{1,2,3\}$

DEFINITIONS

- Set: A well-defined collection of distinct objects.
  - $\{Derek\ Jeter, \pi, ☺\}$
    - (Standing on the shoulders of Justin Gash for this one.)

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

SET OPERATIONS

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

SET OPERATIONS

"A union B"

\cup

Chi

LaTeX

math typesetting

- Union: $A \cup B$ = the set of elements in A or B
- Intersection: $A \cap B$ = the set of elements in A and B

\cap

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

$A = \{2, 4, 6, 8\}$ $\qquad$ $B = \{2, 3, 5, 7\}$

$A \cup B = \{2, 3, 4, 5, 6, 7, 8\}$

$A \cap B = \{2\}$

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?

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

Adapted from "Statistical Inference" by Casella & Berger.

PROBABILITY – PRACTICE

the probability that a U.S. birth results in twin females

- 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}

"the probability of"

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

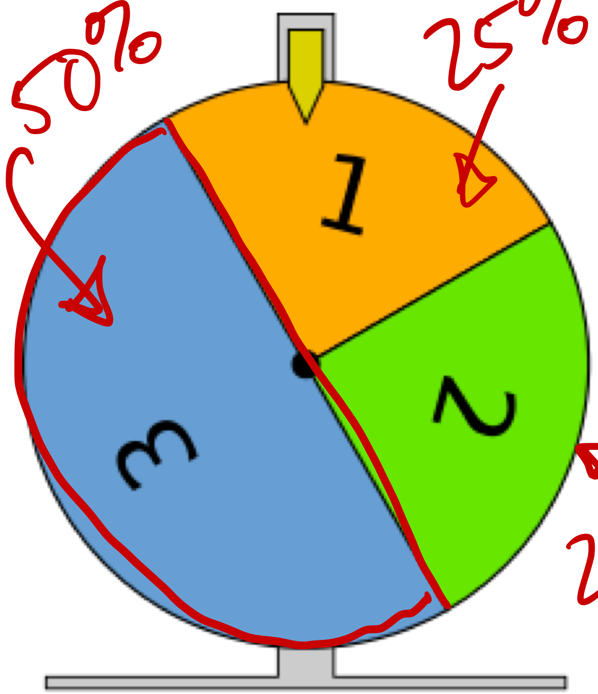
the probability that a U.S. birth results in twins and twin females.

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

the probability that a U.S. birth results in identical twin females.

PROBABILITY BASICS

Experiment: Spinning a spinner

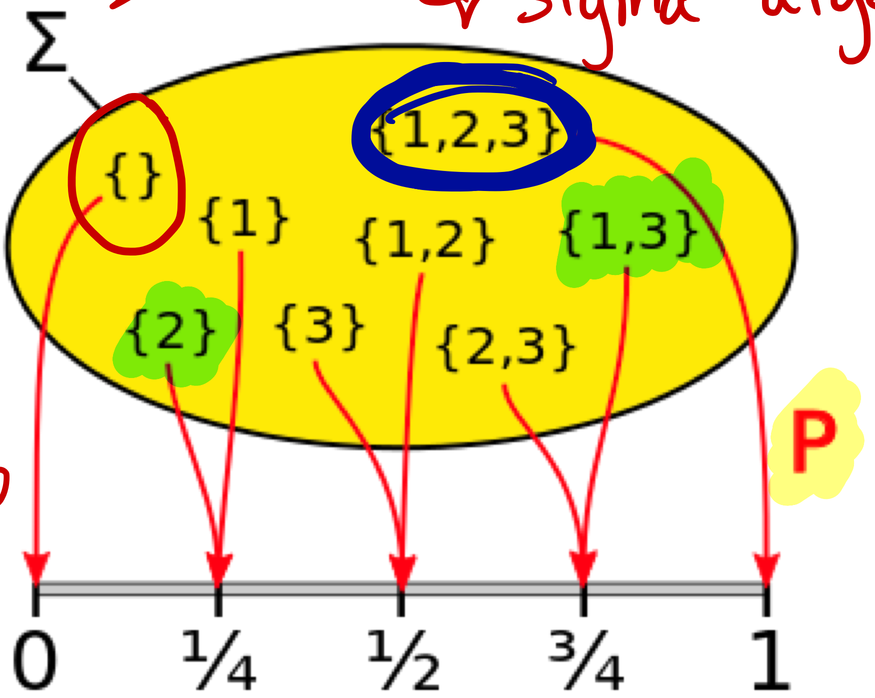


$\Omega = \{1,2,3\}$

sample space S

Event Space

↳ sigma algebra / Borel field



$P(\{\}) = 0$

$P(\{1,2,3\}) = 1$

PROBABILITY RULES

- $P(\emptyset) = 0$
  - Note: $\emptyset$ indicates the "empty set," or the event containing zero outcomes from the experiment.

$\phi = \{\}$

the probability of getting no outcome from our sample space is zero.

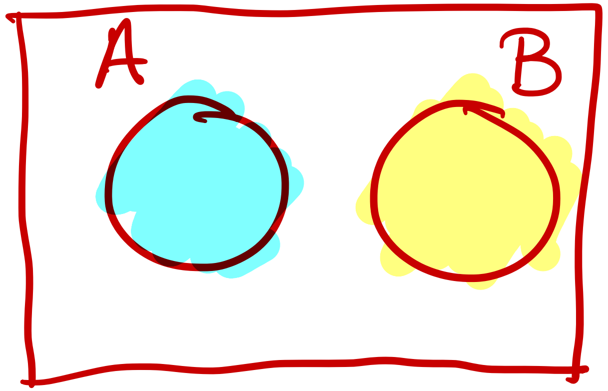
PROBABILITY RULES

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

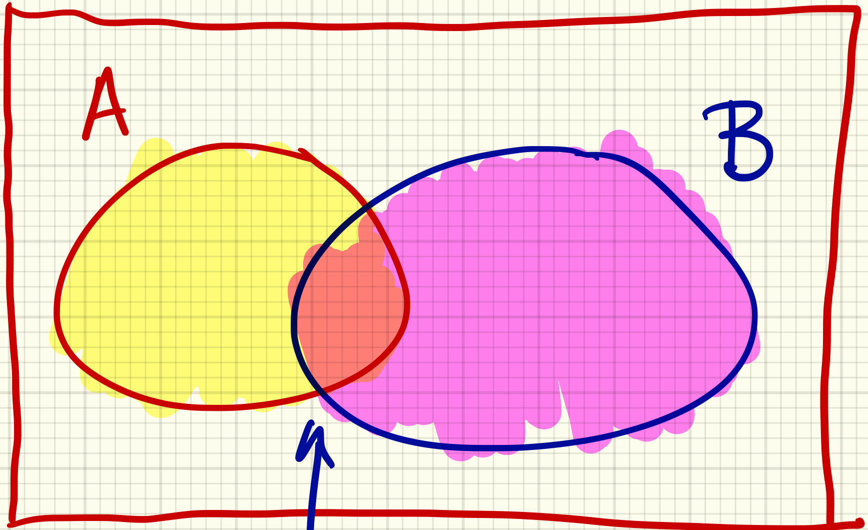
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Venn diagrams can help to illustrate this – but remember that Venn diagrams are not proofs!
  - If A and B are disjoint, then $P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B)$.

↳ no overlap



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

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



overlap is highlighted twice

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

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

A = roll a 2

B = roll an even number

$$P(A|B) = P(2|\text{even}) = \frac{P(2 \text{ and even})}{P(\text{even})} = \frac{1/6}{1/2} = \frac{1}{3}$$

PROBABILITY RULES

- $P(A \cap B) = P(A|B)P(B)$
  - We took the last rule, multiplied both sides of $P(B)$, and voila!
  - We can rearrange these, as well! $P(B \cap A) = P(B|A)P(A)$
  - $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$

PROBABILITY RULES

- $P(A \cap B) = P(A|B)P(B)$
  - We took the last rule, multiplied both sides of $P(B)$, and voila!

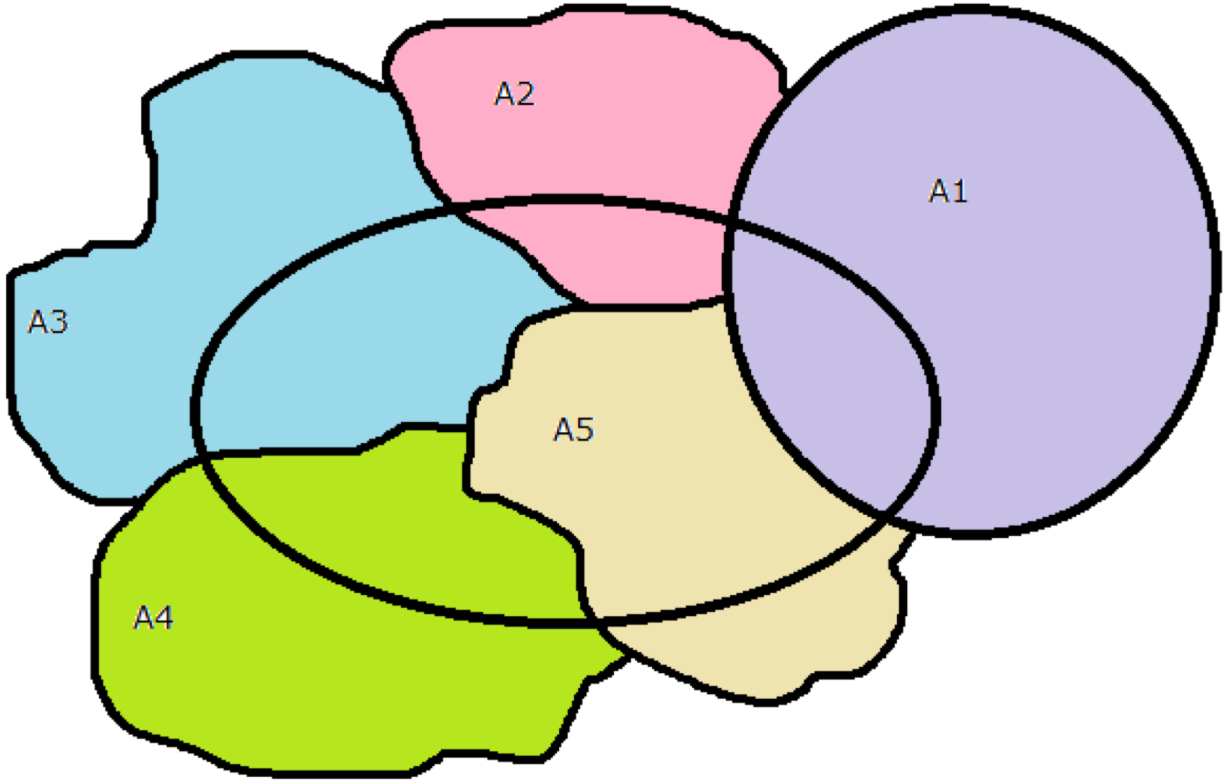
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B)P(B) = P(A \cap B)$$

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

  - $P(A \cap B \cap C) = P(A|\underline{B \cap C})P(B|\underline{C})P(C)$ Naive Bayes

PROBABILITY RULES

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

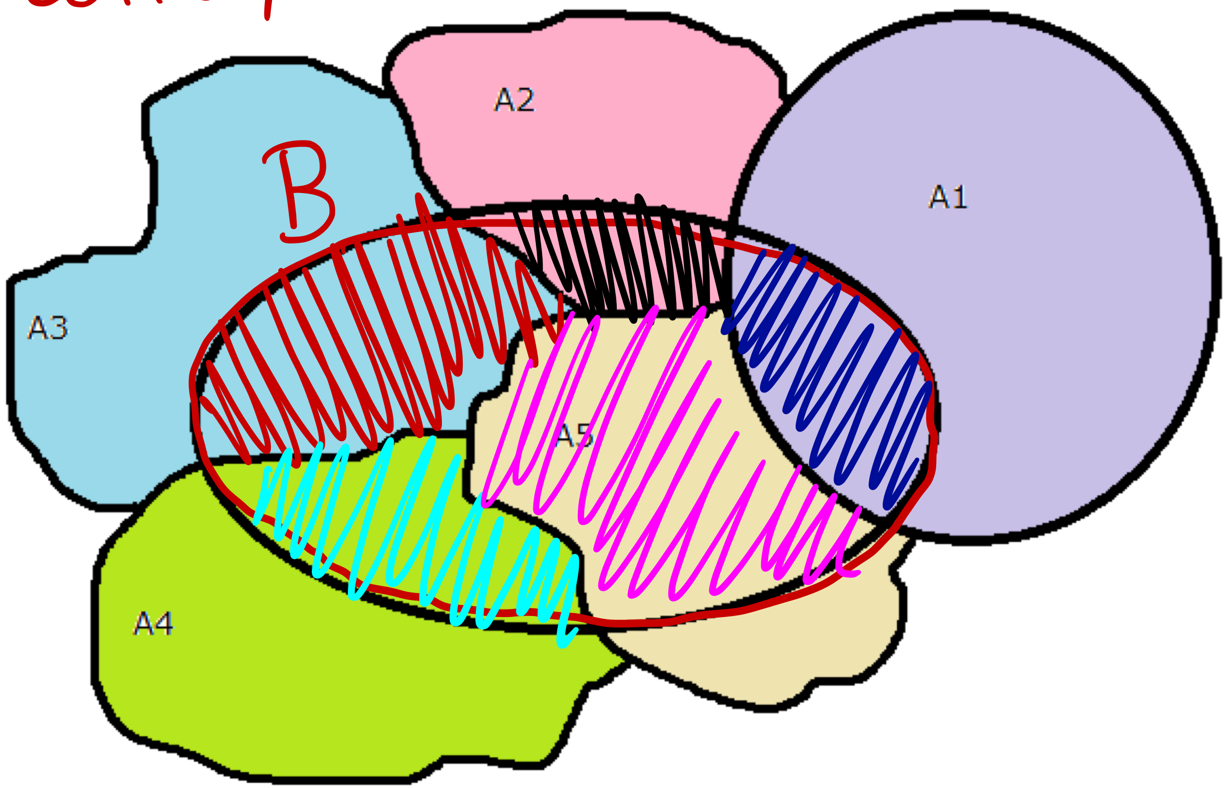


PROBABILITY RULES

no overlap / "mutually exclusive"
collectively exhaustive w.r.t. B

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

Bayesian statistics



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

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

BAYES' THEOREM

P(eating at rest. A | satisfied)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \frac{P(\text{satisfied}\,|\,\text{rest. A.})P(\text{rest A})}{P(\text{satisfied})}$$

A = ate at restaurant A

B = satisfied

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!