Computer Science

CSC380: Principles of Data Science

Probability 2

Xinchen Yu

Summary: calculating probabilities

- If we know that all outcomes are equally likely, we can use

We will use combinatorics to do counting

$$P(E) = \frac{|E|}{|S|}$$

Number of elements in event set (|E|)

Number of possible outcomes (e.g. 36) (|S|)

- If |E| is hard to calculate directly, we can try
  - the rules of probability
  - the Law of Total Probability, using an appropriate partition of sample space S

Rules of probability

- To recap and summarize:

Rules of Probability

1. **Non-negativity:** All probabilities are between 0 and 1 (inclusive)
2. **Unity of the sample space:** $P(S) = 1$
3. **Complement Rule:** $P(E^C) = 1 - P(E)$
4. **Probability of Unions:**
   (a) *In general,* $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
   (b) *If* E and F are disjoint, then $P(E \cup F) = P(E) + P(F)$

Overview

- Conditional probability
- Probabilistic reasoning
  - contingency table
  - probability trees
- Bayes rule
- Independence of events
- Probability and combinatorics

Conditional Probability

Every Probability is a Conditional Probability

- We can consider the original probabilities to be conditioned on the event S: at first what we know is that "something in S" occurs. E.g.

$$P(B) = P(B|S)$$
$$P(B \cap C) = P(B \cap C|S)$$

- $P(B|S)$ in words: what proportion of S does B happen?
- If we then learn that A occurs, A becomes our restricted sample space.

  $P(B|A)$ in words: what proportion of A does B happen?

Joint Probability and Conditional Probability

- We can rearrange $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$ and derive:

The "Chain Rule" of Probability

For any events, A and B, the joint probability $P(A \cap B)$ can be computed as

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

Or, since $P(A \cap B) = P(B \cap A)$

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

Terminology

When we have two events A and B…

- Conditional probability: $P(A|B), P(A^c|B), P(B|A)$ etc.

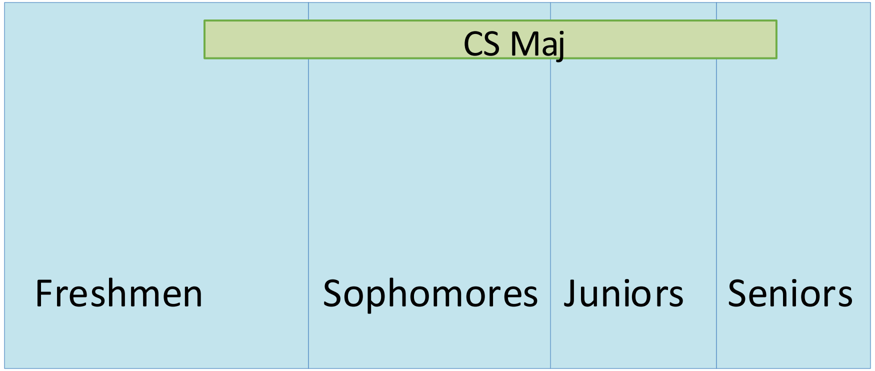
- Joint probability: $P(A, B)$ or $P(A^c, B)$ or ...

- Marginal probability: $P(A)$ or $P(A^c)$

Law of Total Probability, revisited

Law of Total Probability Suppose $B_1, \ldots, B_n$ *form a partition* of the sample space S. Then,

$$P(A) = P(A, B_1) + \cdots + P(A, B_n)$$



Law of Total Probability, revisited

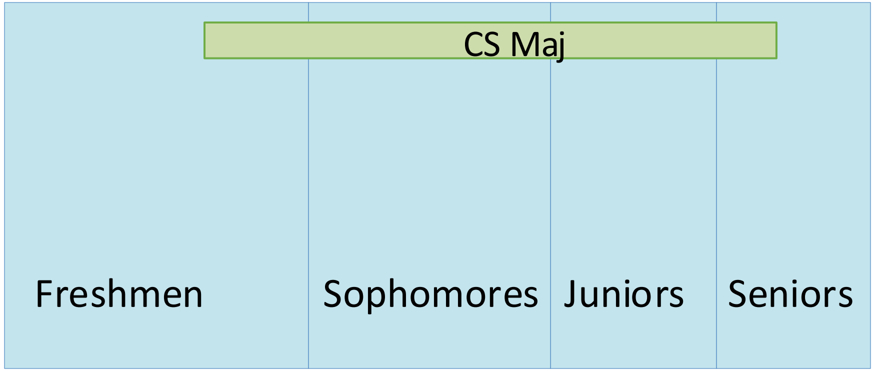
Expanding each $P(A, B_i) = P(A \mid B_i)P(B_i)$, we have:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

A: student in CS major

B_i : student in class year i

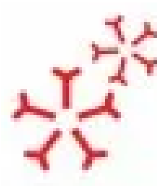
$P(A \mid B_i)$ The fraction of CS major in class year i



Probabilistic reasoning

Probabilistic reasoning

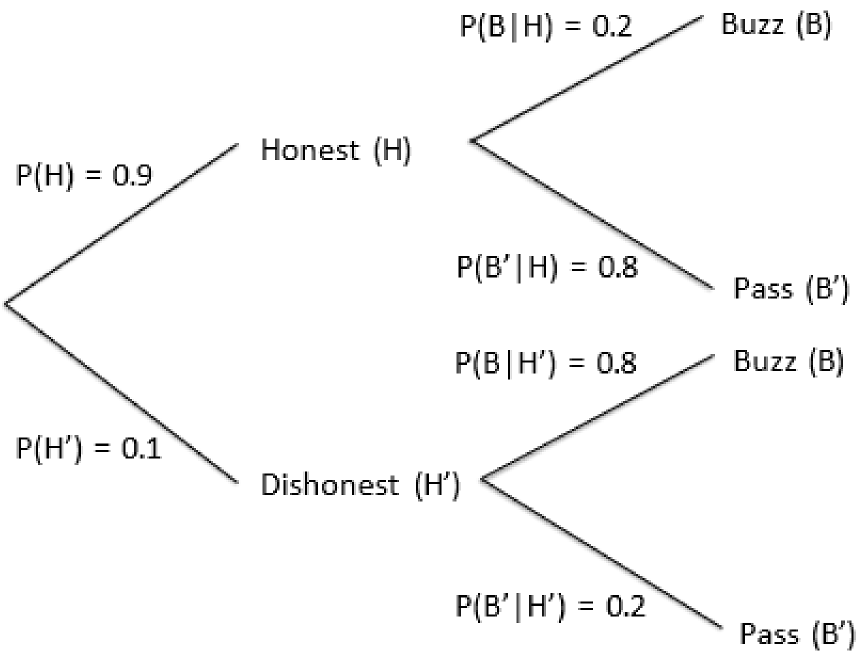
- We have some prior belief of an event A happening
  - $P(A)$, prior probability
  - e.g. me infected by COVID



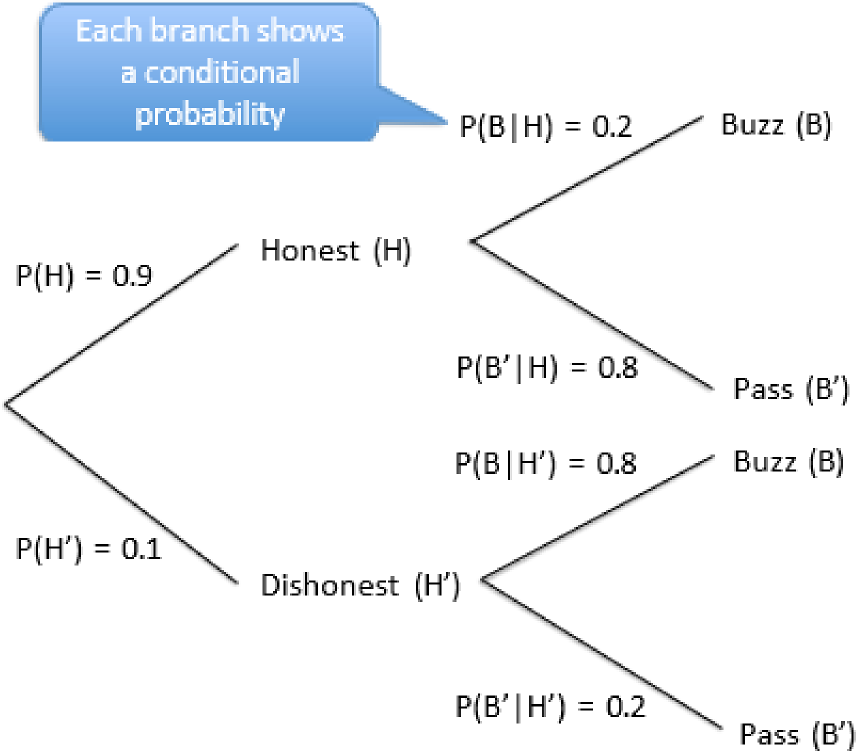
- We see some new evidence B
  - e.g. I test COVID positive

- How does seeing B affect our belief about A?
  - $P(A \mid B)$, posterior probability

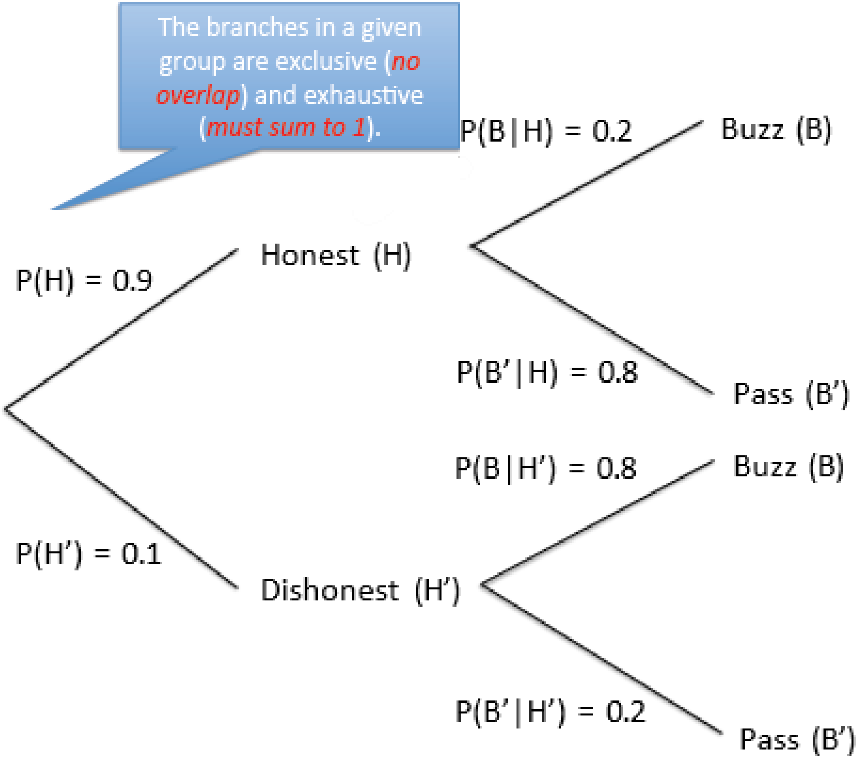
Probability trees: another useful tool



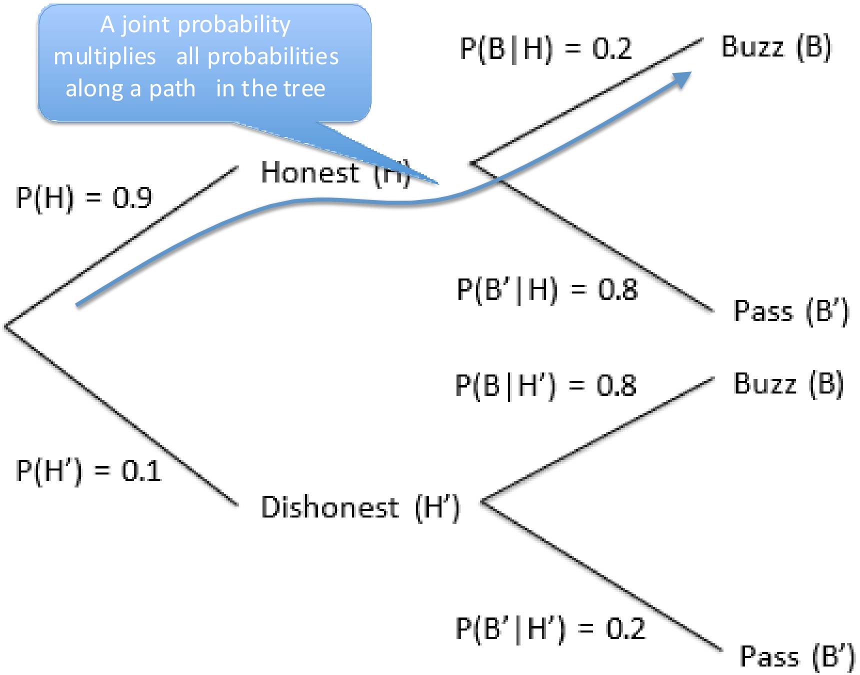
Probability trees: another useful tool



Probability trees: another useful tool



Probability trees: another useful tool



Conditional probability: additional note

- The rules of probability also applies to the rules of conditional probability
- Just replace $P(E), P(F)$ with $P(E|A), P(F|A)$
  - But, need to condition on the same A in the same equation

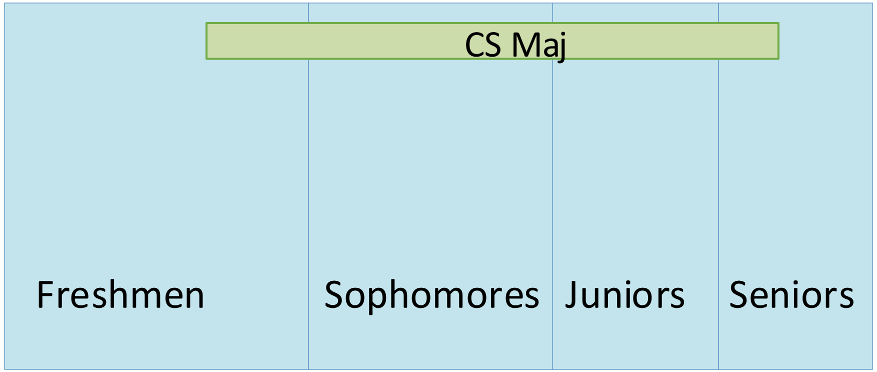
Rules of Probability

1. **Non-negativity:** All probabilities are between 0 and 1 (inclusive)
2. **Unity of the sample space:** $P(S) = 1$
3. **Complement Rule:** $P(E^C) = 1 - P(E)$
4. **Probability of Unions:**
   (a) *In general,* $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
   (b) If E and F are disjoint, then $P(E \cup F) = P(E) + P(F)$

Some examples

- $P(S|A) = 1$
- $P(E|A) + P(E^C|A) = 1$
- $P(E|A) + P(F|A) = P(E \cup F|A)$

A: CS major



Bayes rule

Reversing conditional probabilities

- Is $P(A \mid B) = P(B \mid A)$ in general?
- Let's see..

$$P(A, B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

  - Equal only when $P(A)$ and $P(B)$ are equal

- Let's take a look at a real-world example when they are unequal…

Reversing conditional probabilities

Event A: A person is from France.

Event B: A person speaks English with a French accent.

- In a diverse city, only 5% of people are from France
- Of those from France, 80% speak English with a French accent
- Of those not from France, only 2% speak English with what sounds like a French accent (maybe due to schooling, mimicry, or neighboring countries)

What is $P(A)$ and $P(B)$?

Reversing conditional probabilities

What is $P(A)$ and $P(B)$?

- $P(A) = 0.05$
- $P(B) = P(A, B) + P(A^c, B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) = 0.8 \cdot 0.05 + 0.02 \cdot 0.95 = 0.04 + 0.019 = 0.059$
- $P(A \mid B) = P(A, B)/P(B) = 0.04/0.059 \approx 0.678$

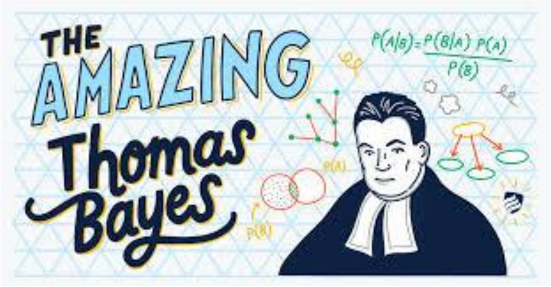
So $P(A) \neq \mathrm{P(B)}$, also hearing a French accent doesn’t guarantee someone is French: a ~68% chance

Bayes rule

Bayes rule For events A, B,

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)}$$

- Very easy to derive from the chain rule, so remember that first.
- Named after Thomas Bayes (1701-1761), English philosopher & pastor



Bayes rule

Bayes rule For events A, B,

Prior probability　Support of evidence

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)}$$

Posterior probability

Examples:

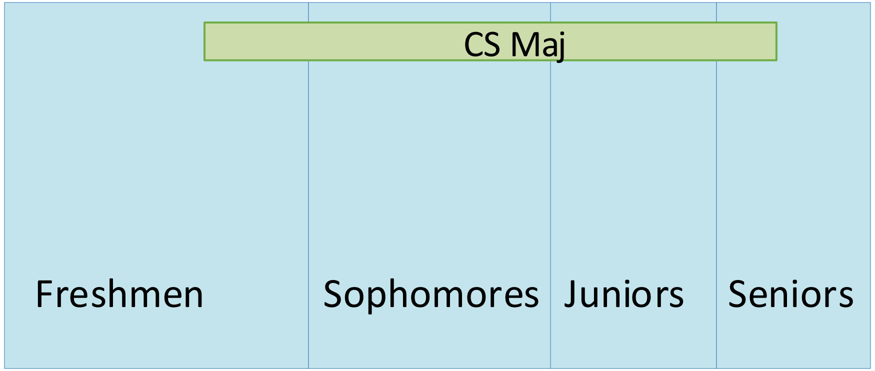
- A: I have COVID, B: my test shows positive
- A: employee lies B: the lie detector buzzes
- A: student is CS major B: student is a senior

Bayes rule and Law of Total Probability

Bayes rule (equivalent form) For event A and $B_1, \ldots, B_n$ forming a partition of S,

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j) \cdot P(B_j)}$$

← $P(A)$



....

Independence

Probabilistic Independence

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

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

- If the employee is dishonest, what's the probability that it will rain tomorrow?
- Seems like independence is symmetric. Is it?

Probabilistic Independence

- If A is independent of B, then $P(A \mid B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?

- Using Bayes' rule, we have

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

- So independence is indeed a symmetric notion

Independence: equivalent statement

- If A, B are independent, then their joint probability has a simple form:

$$P(A,B) = P(A \mid B)P(B) \\ = P(A) \cdot P(B)$$

- This is an equivalent characterization of independence

Independence (version 2)

If A and B are independent events, then

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

Recap: conditional probability

- Conditional prob $P(B \mid A)$
$$= \frac{P(A \cap B)}{P(A)}$$

The "Chain Rule" of Probability

For any events, A and B, the joint probability $P(A \cap B)$ can be computed as

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

Or, since $P(A \cap B) = P(B \cap A)$

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

Extension: chain rule for conditional probability

- If we deal with more than 3 events happening together, we can apply the chain rule of probability repeatedly:

$$P(A, B, C) = P(\, A \mid B, C \,)\, P(B, C)$$

$$= P(\, A \mid B, C \,)\, P(\, B \mid C \,)\, P(C)$$

Recap: probability independence

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

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

Independence (version 2)

If A and B are independent events, then

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

Independence of several events

- We can generalize the notion of independence from two events to more than two.
  - E.g. A: employee is honest; B: rain tomorrow, C: stock price up

- Events $A_1, \dots, A_n$ are independent if for any subsets $A_{i_1}, \dots, A_{i_j}$,

$$P\left(A_{i_1}, \dots, A_{i_j}\right) = P\left(A_{i_1}\right) \cdot \dots \cdot P\left(A_{i_j}\right)$$

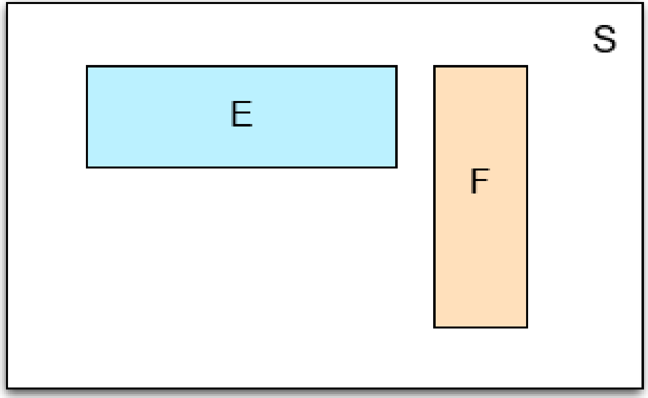
Independence of several events

- E.g. if events A, B, C are independent, then
  - $P(A, B, C) = P(A) \cdot P(B) \cdot P(C)$
  - $P(A, C) = P(A) \cdot P(C)$
  - $P(B, C) = P(B) \cdot P(C)$
  - …

Independent vs. Disjoint Events

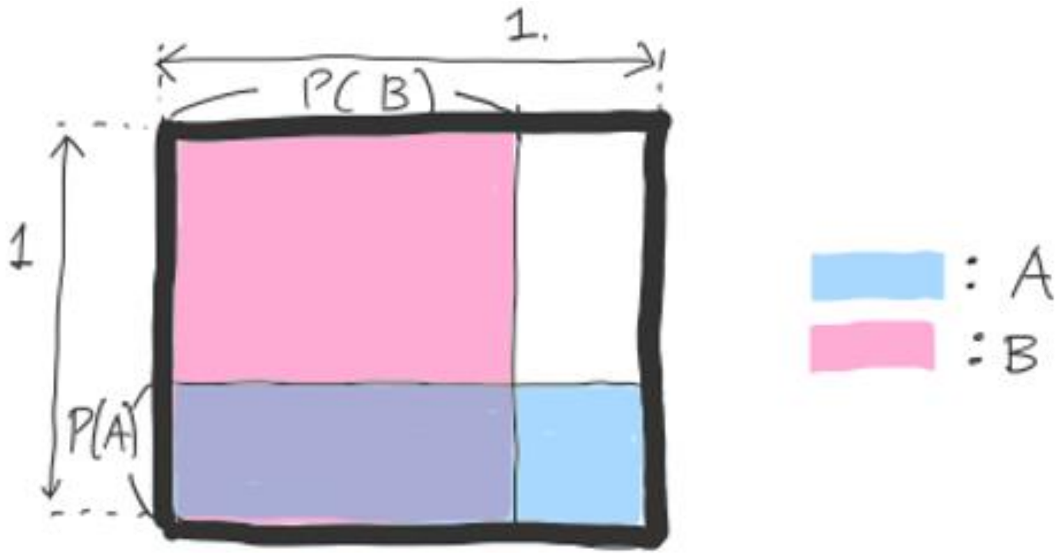
- Many people confuse independence with disjointness.
- They are very different!
- What does it mean for two events to be disjoint?
- If A and B are disjoint, then they cannot occur simultaneously; they are *mutually exclusive*; their intersection is the empty set.

- What does the Venn diagram look like?



Independent vs. Disjoint Events

- If A and B are independent, then P(B|A) = P(B).
- What does the Venn Diagram look like?



Independent vs. Disjoint Events

- If A and B are disjoint, what is P(B|A)?

$$P(\, B \mid A \,) = \frac{P(A,B)}{P(A)} = 0!$$

- Disjointness is practically the opposite of independence: if A occurs, you have all the information about whether B will occur.
  - Specifically, B doesn't occur

Independent vs. Disjoint Events

- Defining property of independent events:
$$P(A \cap B) = P(A)P(B)$$

- Defining property of disjoint events:
$$P(A \cap B) = 0$$

Summary

Conditional Probability Summary

- Representing conditional probabilities using contingency tables, Venn diagrams, and probability trees.
- The chain rule
- Bayes rule
- The law of total probability
- Independent events
- Disjoint events

Probability and Combinatorics

Probability and Combinatorics

- Combinatorics (in CSc144) are useful in calculating probabilities
- Recall: when all outcomes are equally likely:

We will use combinatorics to do counting

$$P(E) = \frac{|E|}{|S|}$$

Number of elements in event set ← $|E|$

Number of possible outcomes (e.g. 36) ← $|S|$

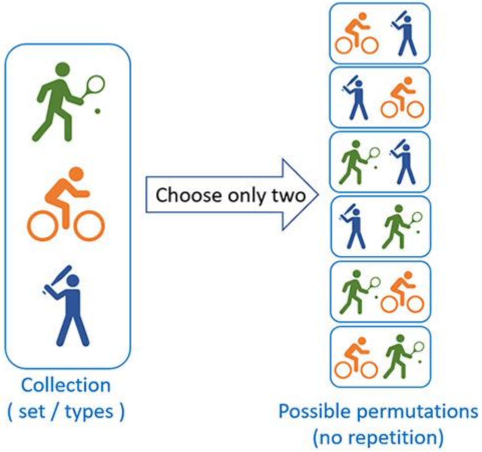
- We will also see its another usage in a popular example: repeated independent trials (Bernoulli trials)

Permutation number

- If *ordered* selection of k items out of n is done without replacement, there are

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

outcomes



Combination number

- If *unordered* selection of k items out of n is done without replacement, there are

$$\frac{n!}{(n-k)!\ k!} =: \binom{n}{k}$$

outcomes