Lecture 12

Foundations of Probability

DSC 40A, Spring 2024

Announcements

- Midterm Exam scores are available on Gradescope, and regrade requests are due **tonight**.
- Homework 5 is due on **Thursday at 11:59PM**. longer survey
- We're updating the office hours calendar frequently – see dsc40a.com/calendar for the latest.

Agenda

- Overview: Probability and statistics.
- Complement, addition, and multiplication rules.
- Conditional probability.

Note: There are no more DSC 40A-specific readings, but we've posted **many** probability resources on the resources tab of the course website. These will come in handy! Specific resources you should look at:

- The DSC 40A probability roadmap, written by Janine Tiefenbruck.
- The textbook Theory Meets Data, which explains many of the same ideas and contains more practice problems.

Question 🤔

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "🤔 Lecture Questions" link in the top right corner of dsc40a.com.

Overview: Probability and statistics

finding parameters to make the best predictions

From Lecture 1: Course overview

Part 1: Learning from Data (Weeks 1 through 6)

- Summary statistics and loss functions; empirical risk minimization.
- Linear regression (including multiple variables); linear algebra.
- ~~Clustering.~~

Part 2: Probability (Weeks 7 through 10)

- Set theory and combinatorics; probability fundamentals.
- Conditional probability and independence.
- The Naïve Bayes classifier.

Why do we need probability?

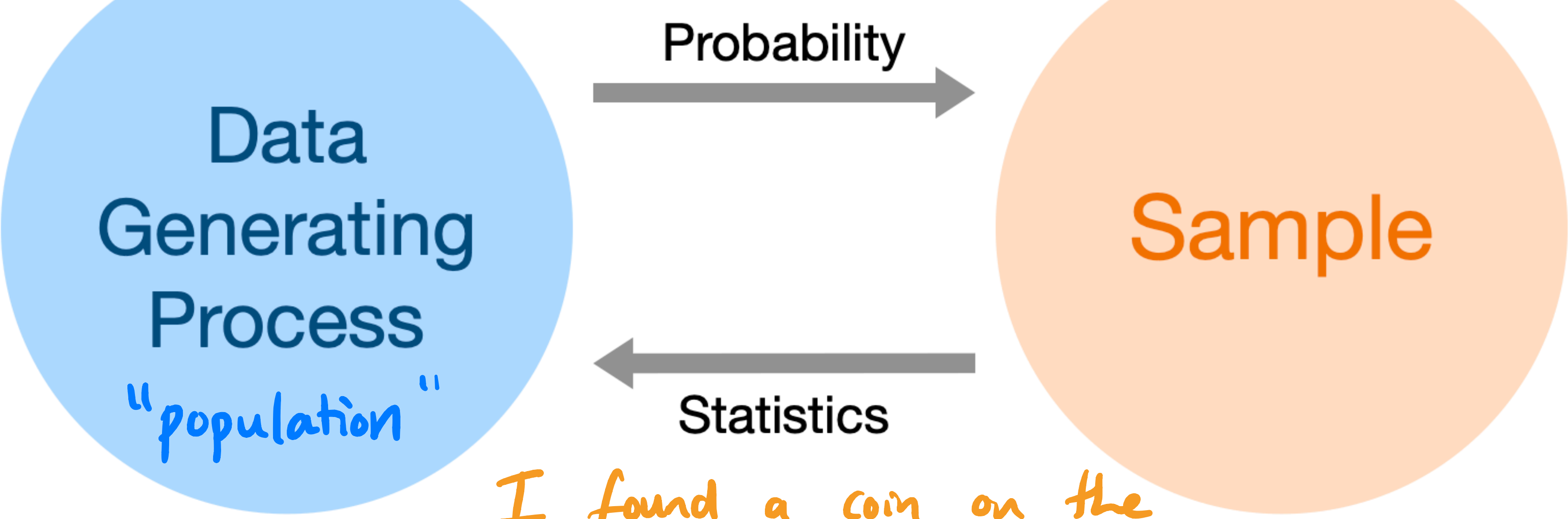
- So far in this class, we have made predictions based on a dataset.
- This dataset can be thought of as a **sample** of some **population**.
- For a hypothesis function to be useful in the future, the sample that was used to create the hypothesis function needs to look similar to samples that we'll see in the future.

future $\approx$ past

Probability and statistics

$\boxed{\text{I have a fair coin.}}$

What's the probability of flipping it 5 times and seeing 5 heads?



I found a coin on the ground and flipped it 5 times, and saw 5 heads.

$\boxed{\Rightarrow \text{Is the coin fair?}}$

The plan 🗺️

- Lecture 12 (today): Key rules of probability.
- Lectures 13-15: Combinatorics.
- Lectures 15-17: Conditional independence and the Naïve Bayes classifier.

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

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

Terminology

- An **experiment** is some process whose outcome is random (e.g. flipping a coin, rolling a die).
- A **set** is an unordered collection of items.
  - Sets are usually denoted with $\{$ curly brackets $\}$.
  - $|A|$ denotes the number of elements in set A.

$A = \{2, 5, 8, 12\}$
$= \{5, 8, 12, 2\}$
$|A| = 4$

- A **sample space**, S, is the set of all possible outcomes of an experiment.
  - Could be finite or infinite!
- An **event** is a subset of the sample space, or a set of outcomes.
  - $E \subseteq S$ means "E is a subset of S."

what you're finding the probability of

$S = \{1, 2, 3, 4, 5, 6\}$
E = rolling an even number
$= \{2, 4, 6\}$

a function that takes in an outcome, s, and outputs its probability

Probability distributions

- A probability distribution, p, describes the **probability** of each outcome s in a sample space S.
  - The probability of each outcome must be between 0 and 1:

$$0 \leq p(s) \leq 1$$

  - The sum of the probabilities of each outcome must be exactly 1:

$$\boxed{\sum_{s \in S} p(s) = 1}$$

- The probability of an **event** is the sum of the probabilities of the outcomes in the event.

the probability of event E

$$\mathbb{P}(E) = \sum_{s \in E} p(s)$$

add the probability of every outcome in event E

Example: Rolling a fair, 6-sided die

$p(1) = \frac{1}{6}$ $\quad$ $p(2) = \frac{1}{6}$ $\quad$ $p(3) = \frac{1}{6}$

• 1 $\quad$ • 2 $\quad$ • 3

$p(4) = \frac{1}{6}$ $\quad$ $p(5) = \frac{1}{6}$ $\quad$ $p(6) = \frac{1}{6}$

• 4 $\quad$ • 5 $\quad$ • 6

S

What do probabilities *mean*?

- One interpretation: if $\mathbb{P}(E) = p$, then if we repeat our experiment infinitely many times, the proportion of repetitions in which event E occurs is p.
  - If p is large, event E occurs very frequently.

"Frequentist" interpretation

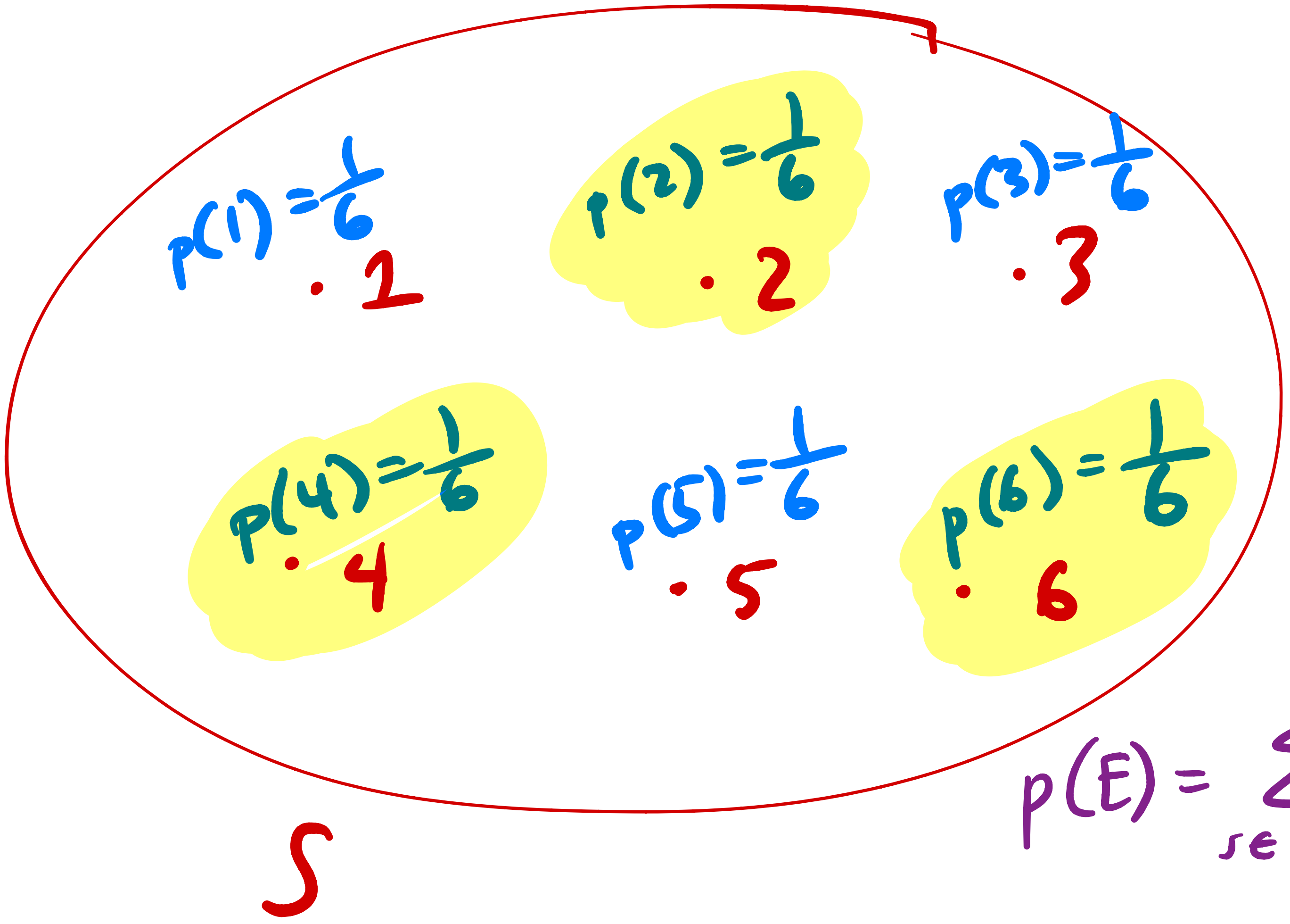
frequency

- Another interpretation: $\mathbb{P}(E) = p$ represents our "degree of belief" in the event E.
  - If p is large, we are pretty sure event E is going to happen when we perform our experiment.

"Bayesian" interpretation

for now, interpretation doesn't change calculations

Example: Probability of rolling an even number on a six-sided die



$E =$ rolling an even number

$= \{2, 4, 6\}$

$$p(E) = \sum_{s \in E} p(s) = p(2) + p(4) + p(6)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

$$= \frac{3}{6} = \boxed{\frac{1}{2}}$$

Equally-likely outcomes

- If S is a sample space with n possible outcomes, and all outcomes are equally-likely, then the probability of any one outcome occurring is $\frac{1}{n}$.
- The probability of an event E, then, is: → a set of possible outcomes

$$\mathbb{P}(E) = \underbrace{\frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}}_{|E| \text{ times}} = \frac{\text{\# of outcomes in E}}{\text{\# of outcomes in S}} = \boxed{\frac{|E|}{|S|}}$$

only when all outcomes are equally likely!!!

$|S| = n$

- **Example**: Flipping a coin three times. → P(exactly 2 heads)?

$S = \{HHH, HHT, HTH, \underbrace{HTT}_{\text{prob: } \frac{1}{8}}, THH, THT, TTH, TTT\}$ → 8 outcomes, $|S| = 8$

E = exactly 2 heads = $\{HHT, HTH, THH\}$ → $|E| = 3$

$P(E) = \frac{|E|}{|S|} = \boxed{\frac{3}{8}}$

Complement, addition, and multiplication rules

Complement rule

- Let A be an event with probability $\mathbb{P}(A)$.
- Then, the event $\bar{A}$ is the **complement** of the event A. It contains the set of all outcomes in the sample space that are **not** in A.

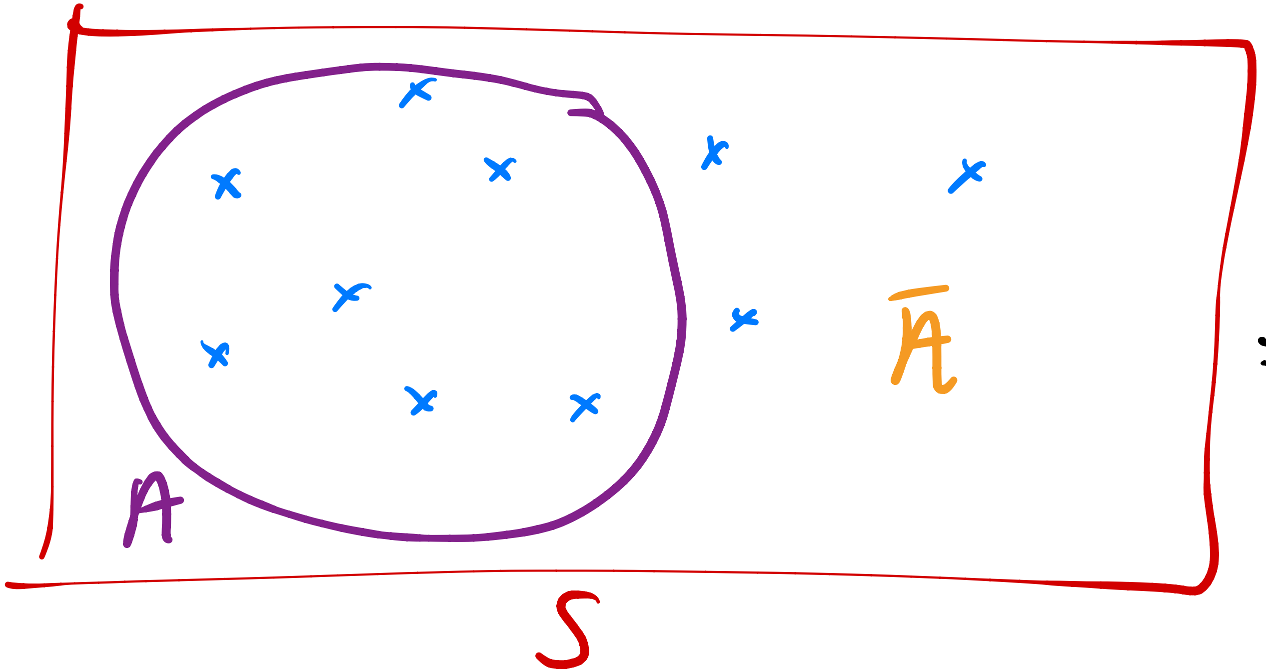
$\bar{A}$, or A^c

$$\boxed{\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)}$$

Everything in S is either in A or $\bar{A}$.

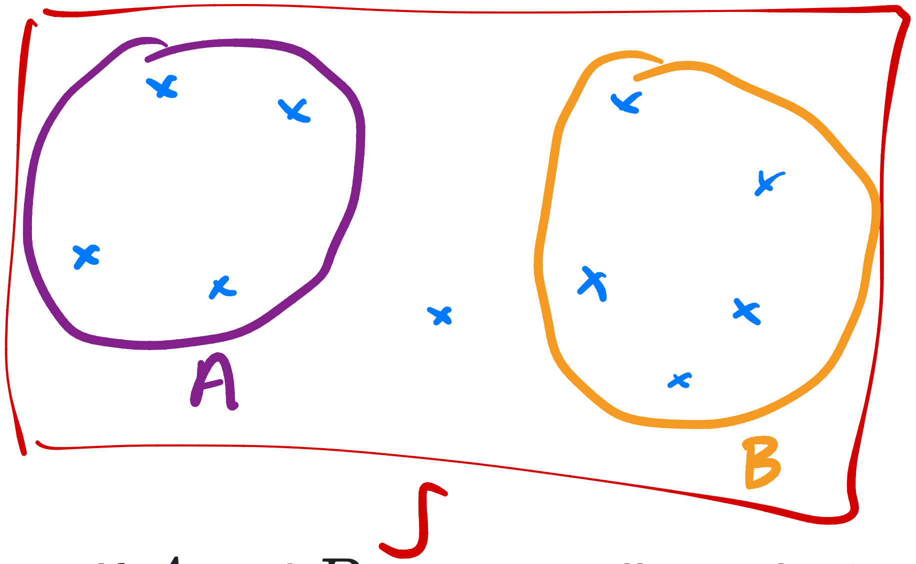
$P(A) + P(\bar{A}) = 1$

$\Rightarrow P(\bar{A}) = 1 - P(A)$



Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).



A and B share no outcomes!

$\Rightarrow$ no overlap!

$\Rightarrow$ mutually exclusive!

e.g. when rolling a fair, 6-sided die

A: rolling an even # $\{2,4,6\}$

B: rolling a 5 $\{5\}$

no overlap!

- If A and B are mutually exclusive, then the probability that A or B happens is:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

"union" of sets A, B $\Rightarrow$ symbol for "or"

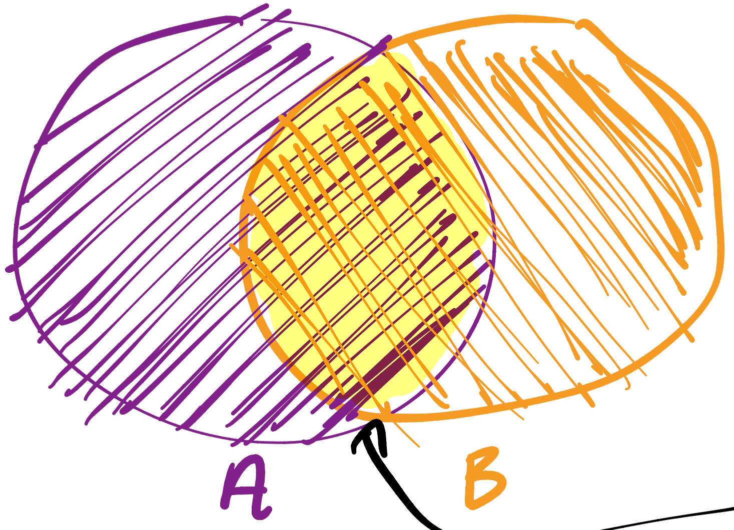
Principle of inclusion-exclusion

- If events A and B are not mutually exclusive, then the addition rule becomes more complicated.

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

"intersection" or "overlap" of two sets

→ "and"



overlap was double counted! :(

- In general, if A and B are any two events, then:

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$$

Question 🤔

either

1 nobody is home

2 mom or brother is home

Answer at q.dsc40a.com

Each day when you get home from school, there is a:

- 0.3 chance your mom is at home.
- 0.4 chance your brother is at home.
- 0.25 chance that both your mom and brother are at home.

When you get home from school today, what is the chance that **neither** your mom nor your brother are at home?

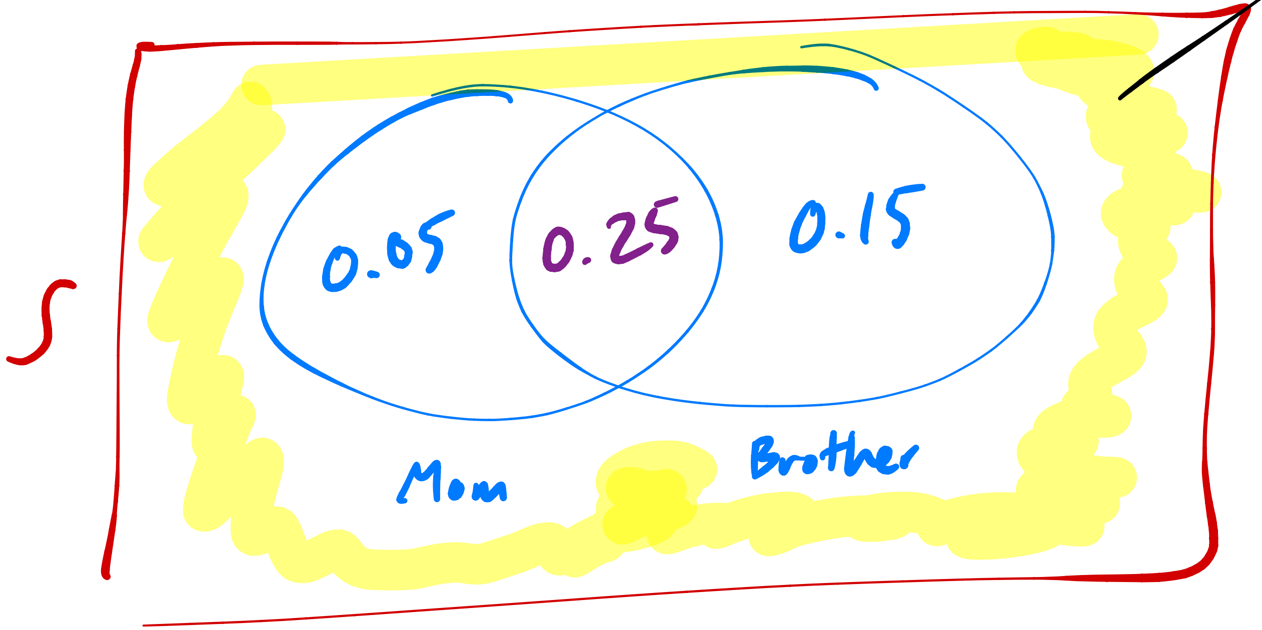
A. 0.3 B. 0.45 C. 0.55 D. 0.7 E. 0.75

$P(\text{Mom}) = 0.3, \quad P(\text{Brother}) = 0.4, \quad P(\text{Mom} \cap \text{Brother}) = 0.25$

$$P(\text{Mom} \cup \text{Brother}) = P(\text{Mom}) + P(\text{Brother}) - P(\text{Mom} \cap \text{Brother})$$
$$= 0.3 + 0.4 - 0.25$$
$$= 0.45$$

$$P(\text{Alone}) = 1 - P(\text{Mom} \cup \text{Brother}) = 1 - 0.45 = \boxed{0.55}$$

$$1 - 0.05 - 0.25 - 0.15 = \boxed{0.55}$$



Multiplication rule and independence

- The probability that events A and B both happen is

"given that"

$$\underbrace{\mathbb{P}(A \cap B)}_{\text{"and"}} = \mathbb{P}(A) \times \mathbb{P}(B|A)$$

- $\mathbb{P}(B|A)$ means "the probability that B happens, given that A happened." It is a **conditional probability**.

  knowing that / assuming that

  - More on this soon!

- If $\mathbb{P}(B|A) = \mathbb{P}(B)$, we say A and B are **independent**.
  - Intuitively, A and B are independent if knowing that A happened gives you no additional information about event B, and vice versa.
  - For two independent events, $\underbrace{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)}_{\text{only when } A, B \text{ independent}}$.

Example: Rolling a die

Let's consider rolling a fair 6-sided die. The results of each die roll are independent from one another.

- Suppose we roll the die once. What is the probability of seeing both 1 and 2?

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

$P(1 \text{ and } 2) = 0$

can only see one number at a time!

- Suppose we roll the die once. What is the probability of seeing 1 or 2?

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

$P(1 \text{ or } 2) = \frac{1}{6} + \frac{1}{6} - 0$

$= \frac{2}{6} = \boxed{\frac{1}{3}}$

$P(\text{first} \neq 1) = 1 - \frac{1}{6} = \frac{5}{6}$

Example: Rolling a die

- Suppose we roll the die 3 times. What is the probability of never seeing a 1 in any of the rolls?

$$P(\text{never } 1) = P(\text{first} \neq 1 \text{ and } 2^{nd} \neq 1 \text{ and } 3^{rd} \neq 1)$$

$$= P(\text{first} \neq 1) \cdot P(2^{nd} \neq 1 \mid \cancel{\text{first} \neq 1}) \cdot P(3^{rd} \neq 1 \mid \cancel{\text{first} \neq 1 \text{ and second} \neq 1})$$

$$= \left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) = \left(\frac{5}{6}\right)^3$$

independent

- Suppose we roll the die 3 times. What is the probability of seeing a 1 at least once?

$$P(\text{at least one } 1) = 1 - P(\text{never } 1)$$

$$= 1 - \left(1 - \frac{1}{6}\right)^3 = 1 - \left(\frac{5}{6}\right)^3$$

previous answer

General form: $1 - (1-p)^n$

DSC 10: Grandma problem

exponent, not multiplication!
answer ≤ 1

Example: rolling a die

- Suppose we roll the die n times. What is the probability of only seeing the numbers 1, 3, and 4?

$$p(\text{first} \in \{1,3,4\} \text{ AND second} \in \{1,3,4\} \text{ AND} \ldots) \underset{\text{independent}}{=} p(\text{first} \in \{1,3,4\})^n$$

$$= \left(\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right)^n = \boxed{\left(\frac{1}{2}\right)^n}$$

- Suppose we roll the die 2 times. What is the probability that the two rolls are different?

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

$$|S| = 6 \cdot 6 = 36$$

\# outcomes both rolls same : 6

another interpretation : 6 possibilities for roll two, 5 are different from first

$\Rightarrow p(\text{both diff}) = 5/6$

$$p(\text{both same}) = \frac{6}{36} = \frac{1}{6} \longrightarrow p(\text{both diff}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Conditional probability

Conditional probability

- The probability of an event may **change** if we have additional information about outcomes.

"and"

- Starting with the multiplication rule, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$, we have that:

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

definition of conditional probability

assuming that $\mathbb{P}(A) > 0$.

Question 🤔

Answer at q.dsc40a.com

Suppose a family has two pets. Assume that it is equally likely that each pet is a dog or a cat. **Consider the following two probabilities**:

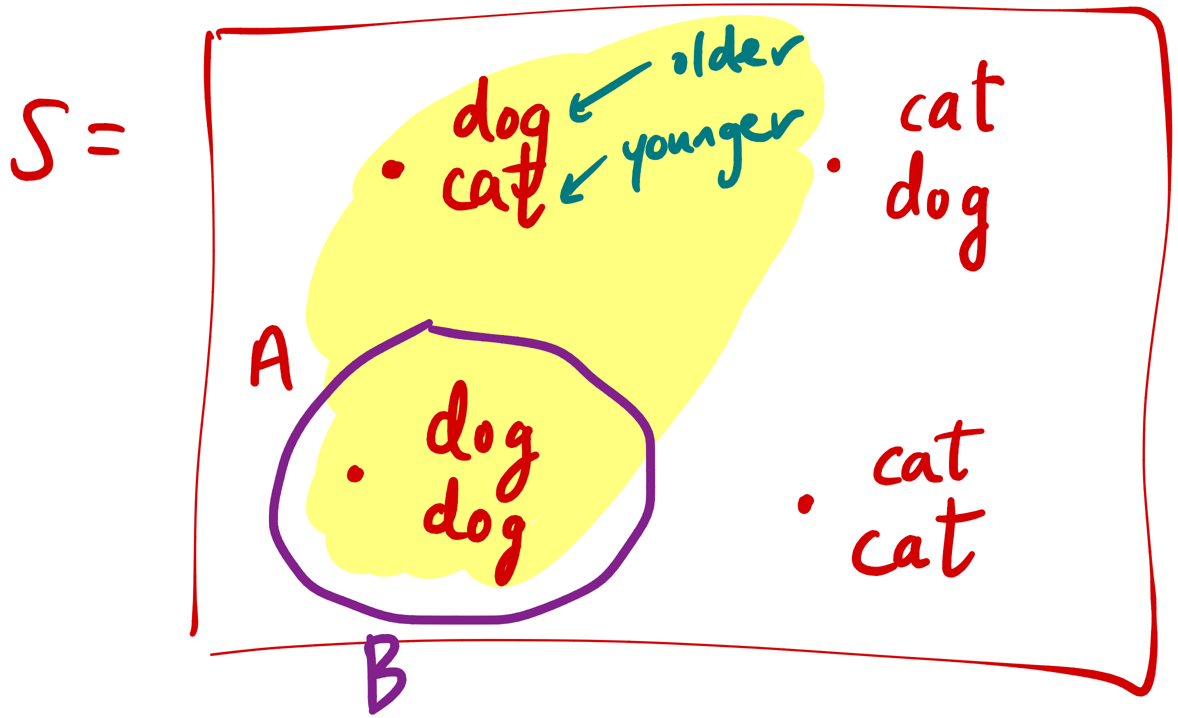
- The probability that both pets are dogs given that **the oldest is a dog**.
- The probability that both pets are dogs given that **at least one of them is a dog**.

Are these two probabilities equal?

- A. Yes, they're equal.
- B. No, they're not equal.

Example: Pets

Let's compute the probability that $\overbrace{\text{both pets are dogs}}^{B}$ given that $\overbrace{\textbf{the oldest is a dog}}^{A}$.



$$P(\text{oldest is dog}) = \frac{2}{4} = \frac{1}{2}$$

$$P(\text{both dogs} \mid \text{oldest is dog})$$

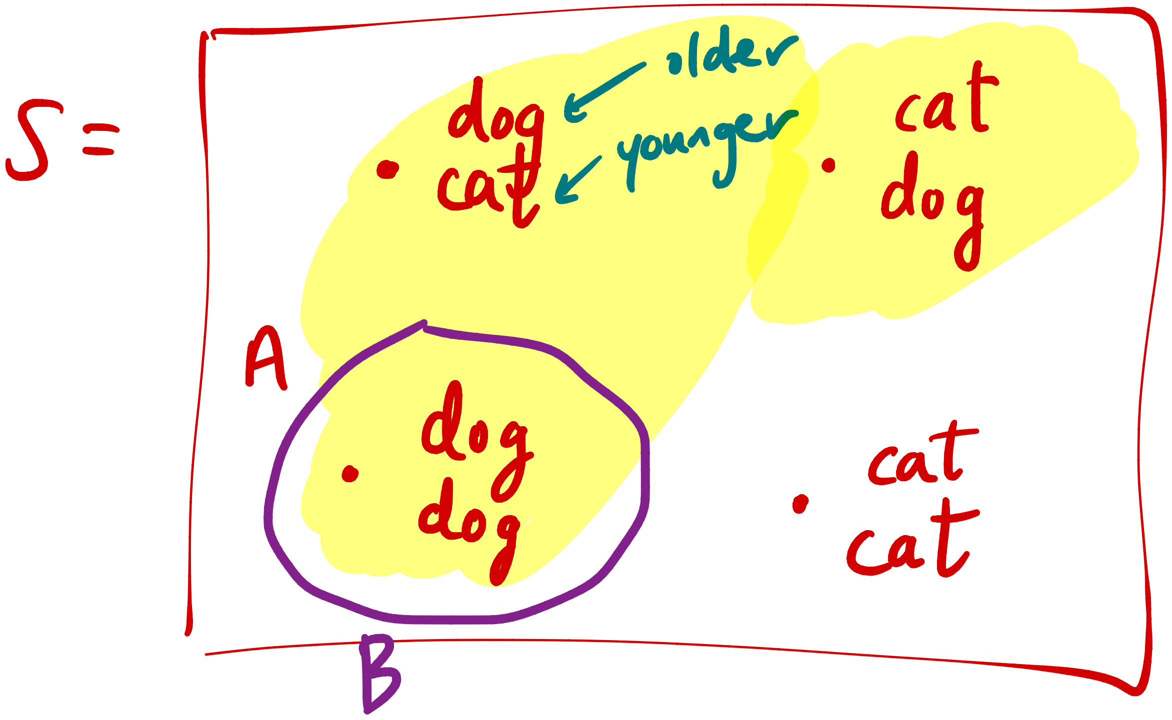
$$= \frac{P(\text{both dogs AND oldest is dog})}{P(\text{oldest is dog})}$$

$$= \frac{\frac{1}{4}}{\frac{2}{4}} = \boxed{\frac{1}{2}}$$

intuitively, event B is $\frac{1}{2}$ of event A

Example: Pets

Let's now compute the probability that both $\overbrace{\text{pets are dogs}}^{B}$ given that $\underbrace{\textbf{at least one of them}}_{A}$ **is a dog**.



$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \boxed{\frac{1}{3}}$$

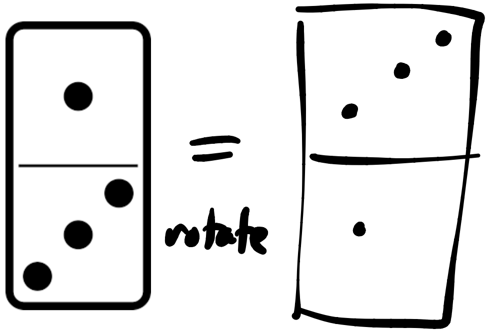
intuitively, B is 1/3 of A

Example: Dominoes

(source: 538)

In a set of dominoes, each tile has two sides with a number of dots on each side: 0, 1, 2, 3, 4, 5, or 6. There are 28 total tiles, with each number of dots appearing alongside each other number (including itself) on a single tile.

7	6	5	4	3	2	1
(0,0)						
(0,1)	(1,1)					
(0,2)	(1,2)	(2,2)				
(0,3)	(1,3)	(2,3)	(3,3)			
(0,4)	(1,4)	(2,4)	(3,4)	(4,4)		
(0,5)	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	
(0,6)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)



$$1+2+\dots+7 = \frac{7(8)}{2} = 28$$

Example: Dominoes

Question 1: What is the probability of drawing a "double" from a set of dominoes – that is, a tile with the same number on both sides?

$$P(\text{double}) = \frac{7}{28} = \boxed{\frac{1}{4}}$$

(0,0)

(0,1) (1,1)

(0,2) (1,2) (2,2)

(0,3) (1,3) (2,3) (3,3)

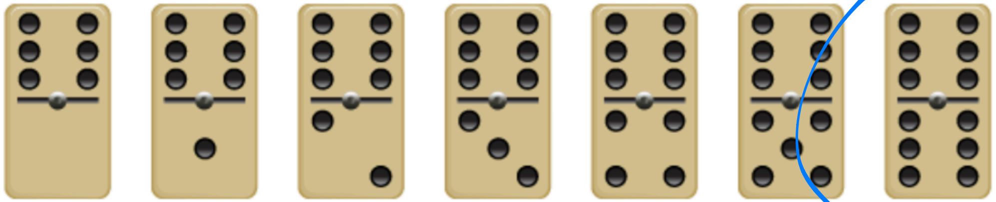
(0,4) (1,4) (2,4) (3,4) (4,4)

(0,5) (1,5) (2,5) (3,5) (4,5) (5,5)

(0,6) (1,6) (2,6) (3,6) (4,6) (5,6) (6,6)

Example: Dominoes

Question 2: Now your friend picks a random tile from the set and tells you that at least one of the sides is a 6. What is the probability that your friend's tile is a double, with 6 on both sides?



one of the 7 possible tiles w/ at least one 6

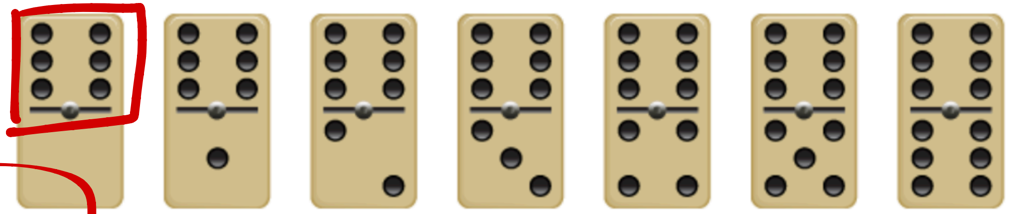
(0,0)
(0,1) (1,1)
(0,2) (1,2) (2,2)
(0,3) (1,3) (2,3) (3,3)
(0,4) (1,4) (2,4) (3,4) (4,4)
(0,5) (1,5) (2,5) (3,5) (4,5) (5,5)
(0,6) (1,6) (2,6) (3,6) (4,6) (5,6) (6,6)

$$P(\text{double} \mid \text{at least one } 6) = \frac{P(\text{double AND at least one } 6)}{P(\text{at least one } 6)}$$

$$= \frac{1/28}{7/28} = \boxed{1/7}$$

Example: Dominoes

Question 3: Now you pick a random tile from the set and uncover only one side, revealing that it has 6 dots. What is the probability that this tile is a double, with 6 on both sides?



Answer: $\boxed{\frac{1}{4}}$

See 538's explanation here.

Summary

- If A is an event, then the complement of A, denoted $\bar{A}$, is the event that A does not happen, and $\boxed{P(\bar{A}) = 1 - P(A)}$.
- Two events A and B are mutually exclusive if they share no outcomes (no overlap). In this case, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- More generally, for any two events, $\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$.
- The probability that events A and B both happen is $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)}$.
- $\mathbb{P}(B|A)$ is the **conditional probability** of B occurring, given that A occurs:

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

  - If $\mathbb{P}(B|A) = \mathbb{P}(B)$, then events A and B are **independent**.