Lecture 12

Foundations of Probability

DSC 40A

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 🤔

Take a moment to pause and reflect...

If you have any questions please post online to our forms/Q&A site.

Course staff will answer them ASAP!

Overview: Probability and statistics

From Lecture 1: Course overview

Part 1: Learning from Data (Weeks 1 through 6) finding the Best Params

- 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.

→ SPAM Filter

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.

Sample ~ Population

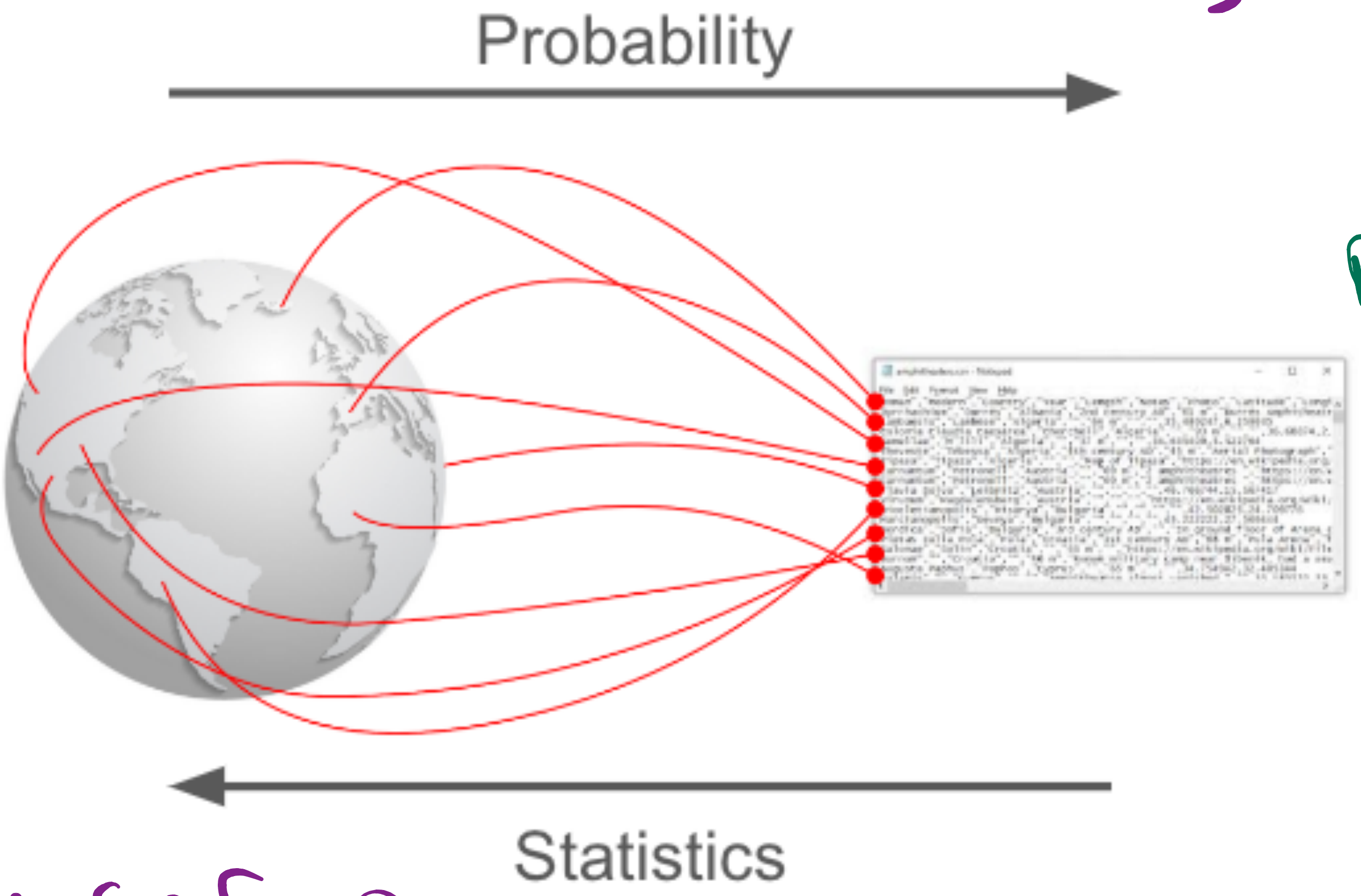
Dataset ~ ϕ

Learn from the Past ~ Predict future

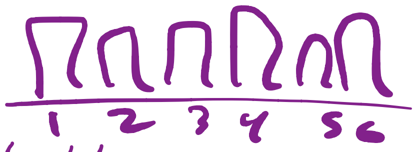
Probability and statistics

given a fair D6, what is the likelihood of rolling a 4?

Data Generating Process



Dataset



given the same fair D6. we roll it 100 times, collect the Data, and analyze the Distribution of outcomes.

Probability and statistics

Population

Subset || Dataset of Population

Draw conclusions about all Youtube comments?

Describe, show, and summarize our data in a meaningful way. Explore then infer!

The plan 🗺️

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

Terminology

$|S| = 2$

$S = \{H, T\}$ "Sample Space"

- 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 **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."

$A = \{1, 2, 5\}$

$|A| = 3$

"D6"

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

$|S| = 6$

$S = \{1, 2, 3, 4, 5, 6\}$ "D6" $|S| = 6$

$E = \{1, 3, 5\}$ prob. of rolling an odd # $|E| = 3$

? 0 X ? 7X

$P(3) \rightarrow \boxed{3} \rightarrow$ 
0: 0.0
1: 0.01
2: 0.05
⋮ 0.70
9: 0.15
$\} = 1$

$0 \le P(3) \le 1$

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:

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

0.5 for H
0.5 for T
$\} = 1$

Rolling a D4

1 $P(1) = 1/4$
2 $P(2) = 1/4$
3 $P(3) = 1/4$
4 $P(4) = 1/4$
S

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

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

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

total Probability of E

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

For each event in E add the probabilities of the outcomes.

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 Approach"

- 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"

$= \frac{1}{2}$

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

S

1 $P(1) = \frac{1}{6}$ s	2 $P(2) = \frac{1}{6}$ s	3 $P(3) = \frac{1}{6}$ s
4 $P(4) = \frac{1}{6}$ s	5 $P(5) = \frac{1}{6}$ s	6 $P(6) = \frac{1}{6}$ s

E = rolling an even num on a D6

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

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

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

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

$S \notin s$

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:

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

only when all outcomes are equally likely !!

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

S = {HHH, HHT, HTT, HTH, THH, THT, TTH, TTT} = 8 outcomes

|S| = 8

|E| = 3

P(s) = 1/8 P(s) = 1/8 . . . P(s) = 1/8

NOT 4 !!!

E = exactly P(2 heads) = {HHT, HTH, THH}

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

Complement, addition, and multiplication rules

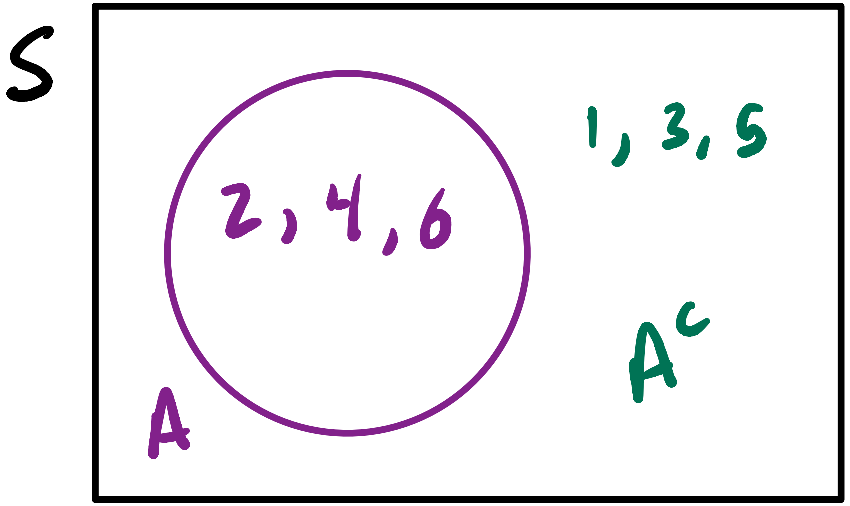
Complement rule

notation: ($\bar{A}$, A^c)

- 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.

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

$s = 1$



P(A) rolling even # on a D6

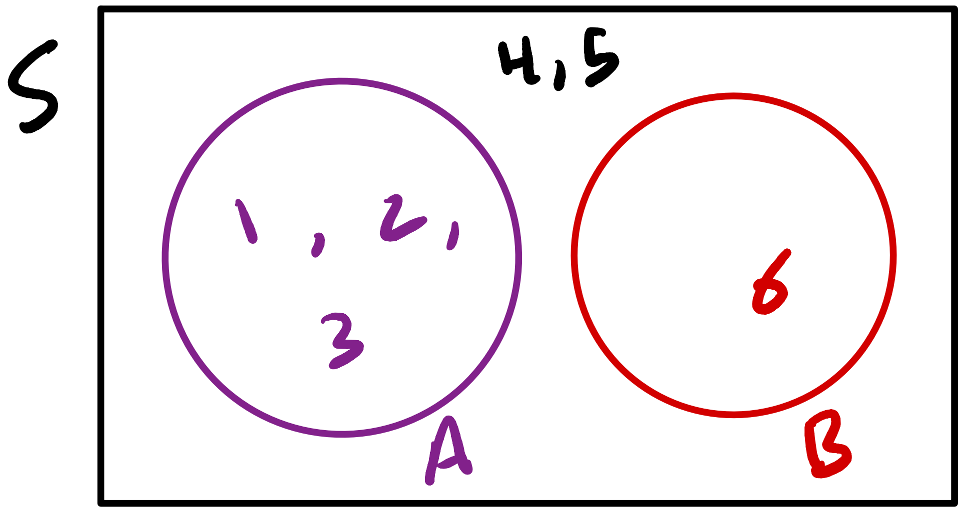
$$P(A) + P(A^c) = 1$$

$$P(A^c) = 1 - P(A)$$

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 & B have no outcomes that overlap!



rolling a D6

A: rolling $\{1, 2, 3\}$

B: rolling $\{6\}$

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

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

$\cup$ is a union!

"or"

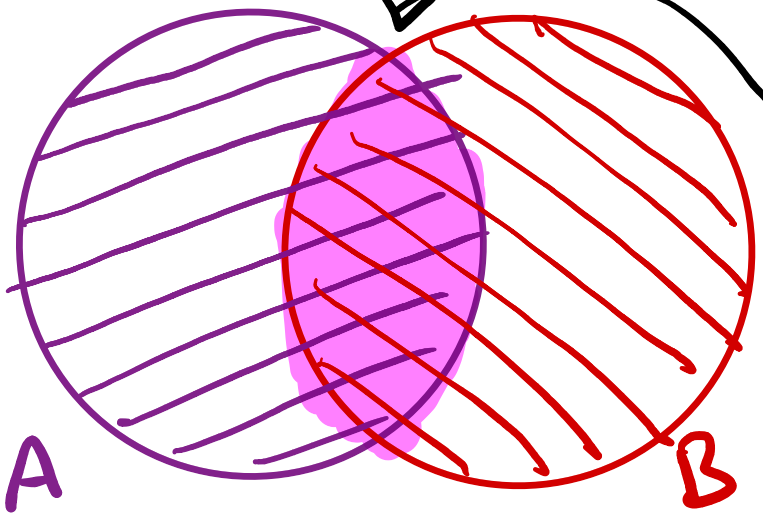
$\cup$ = union "or"

$\cap$ = intersection "and"

Principle of inclusion-exclusion

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

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



Double Counted!

So we remove the double counted area

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

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

S
1 2 3 4

Question 🤔

Pause the video and try to answer the question...

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.

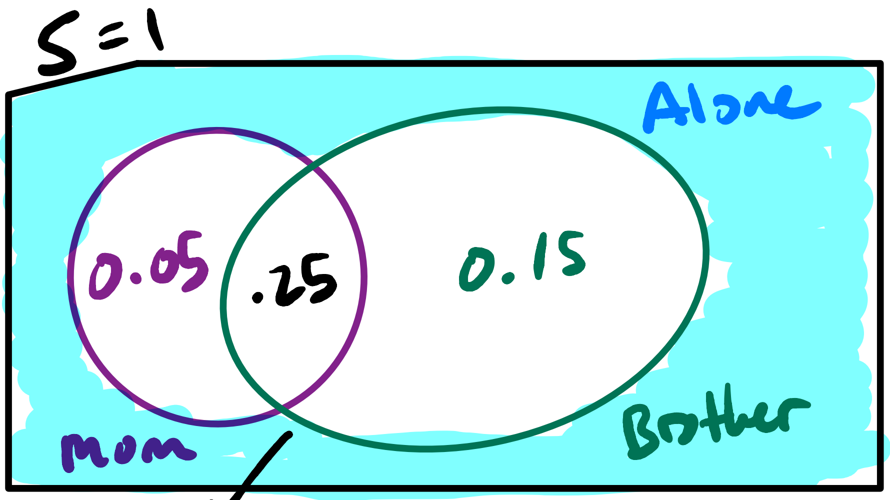
1. your Mom is home, no Bro.
2. your Bro. is home, but no Mom
3. your Bro + Mom is home
4. neither of them are home

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

$P(\text{mom}) = 0.3$ $P(\text{Bro}) = 0.4$

$P(\text{Mom} \cap \text{Bro}) = 0.25$



$1 -$ = Answer

where No one is Home!

$1 - .45 = .55$

$P(\text{Mom} \cup \text{Bro}) = P(M) + P(B) - P(M \cap B)$

$= 0.3 \quad 0.4 - 0.25$

$= \underline{0.45}$

$P(\text{Alone}) = 1 - P(M \cup B) = 1 - 0.45 = \underline{0.55}$

P(B|A) B Knowing A or Assuming

given

Multiplication rule and independence

- The probability that events A and B both happen is

"union" and

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

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

ONLY when A & B are 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	2	3
4	5	6

$P(1) = \frac{1}{6}$

$P(2) = \frac{1}{6}$

~~$= \frac{2}{6}$?? yes no~~

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

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

S

1	2	3
4	5	6

$P(1) = \frac{1}{6}$

$P(2) = \frac{1}{6}$

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

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

Example: Rolling a die

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

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

for 1 trial,
never seeing a 1

$P(\text{not a } 1) = 1 - 5/6 = \frac{1}{6}$

$P(1^{st} \neq 1) \cdot P(2^{nd} \neq 1) \cdot P(3^{rd} \neq 1)$

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

Because events are INDEPENDENT

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

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

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

General form

$1 - (1-p)^n$

Example: rolling a die

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

Key!!

not multiplied by n

$P(1^{st} \in \{1,3,4\}$ and $2^{nd} \in \{1,3,4\}$ and $\ldots n \in \{1,3,4\})$

Because of independence $P(1^{st} \in \{1,3,4\})^n = P(\frac{1}{6}+\frac{1}{6}+\frac{1}{6})^n = (\frac{1}{2})^n$

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

$P(\text{both rolls are Diff}) = 1 - \frac{1}{6} = \frac{5}{6}$

$|S|$ $6 \cdot 6 = 36$ $S = \{(1,1), (1,2), (1,3), \cdots (6,6)\}$

$P(\text{Both same}) = \frac{6}{36} = \frac{1}{6} \longrightarrow P(\text{Both Diff}) = 1 - \frac{1}{6} = \frac{5}{6}$

Conditional probability

Conditional probability

- The probability of an event may **change** if we have additional information about outcomes.
- Starting with the multiplication rule, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$, we have that:

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

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

$S = (A, A^c)$ $\quad S_1 = (A)$

Question 🤔

Pause the video and try to answer the question...

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**. 0.5
- The probability that both pets are dogs given that **at least one of them is a dog**. 0.33

Are these two probabilities equal?

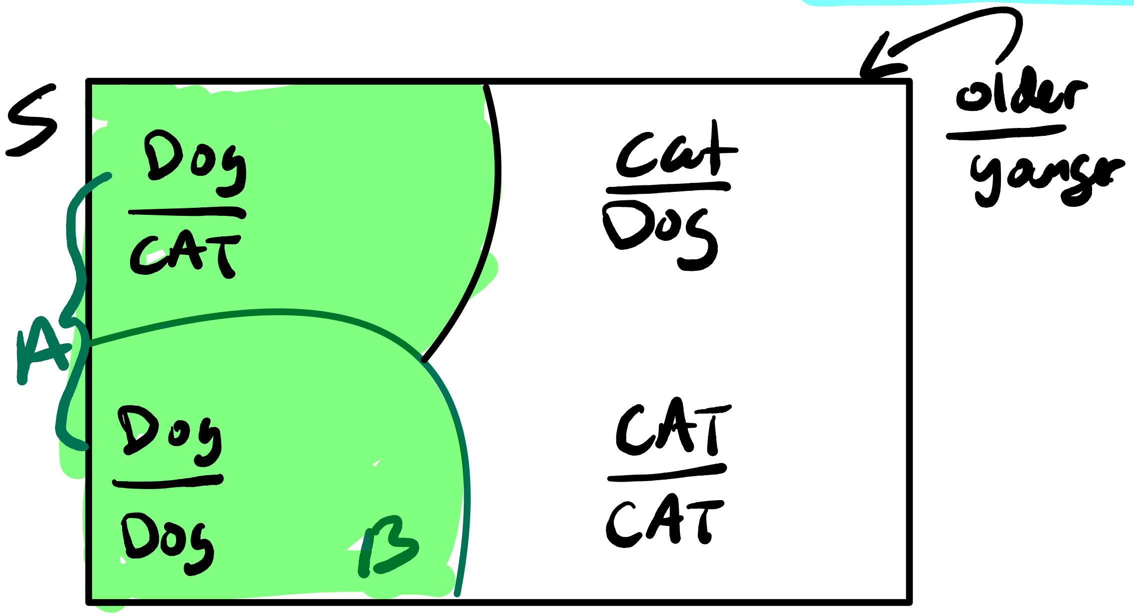
- A. Yes, they're equal.
- ✓ B. No, they're not equal.

Example: Pets

Key !

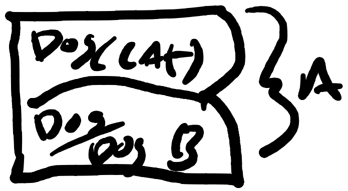
B | A

Let's compute the probability that both pets are dogs given that **the oldest is a dog**.



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

new "S" =



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

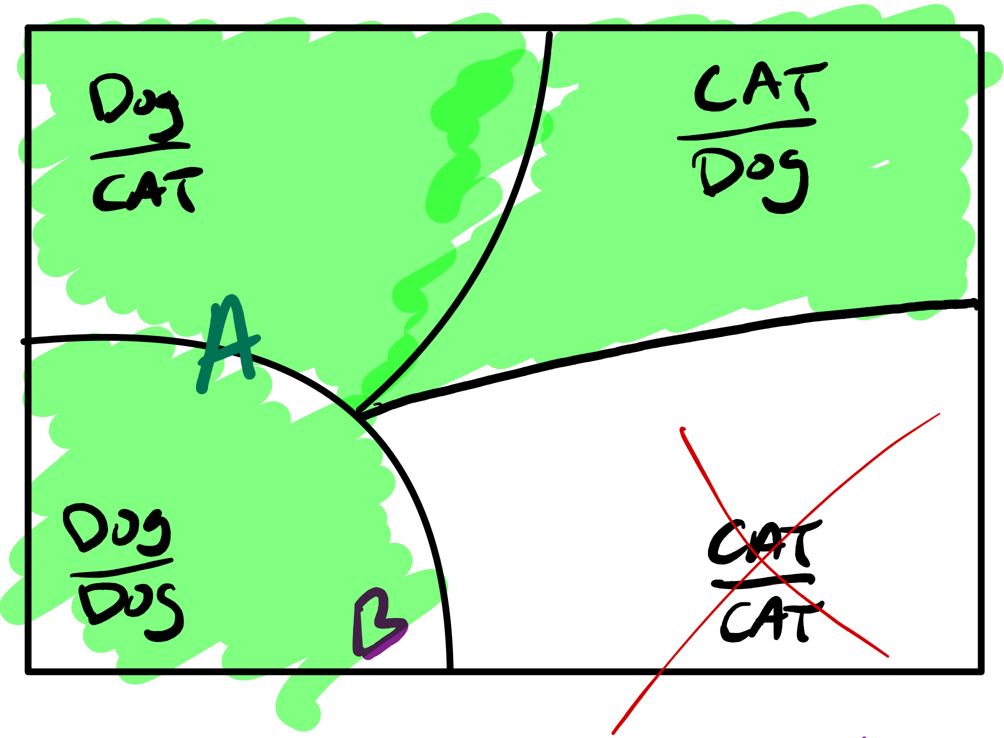
$P(\text{both} = \text{Dogs} \mid \text{oldest} = \text{Dog})$

$$= \frac{P(\text{Both Dogs And oldest is a dog})}{P(\text{oldest being a Dog})}$$

$$= \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} \quad .50$$

Example: Pets

Let's now compute the probability that both pets are dogs given that **at least one of them is a dog**.



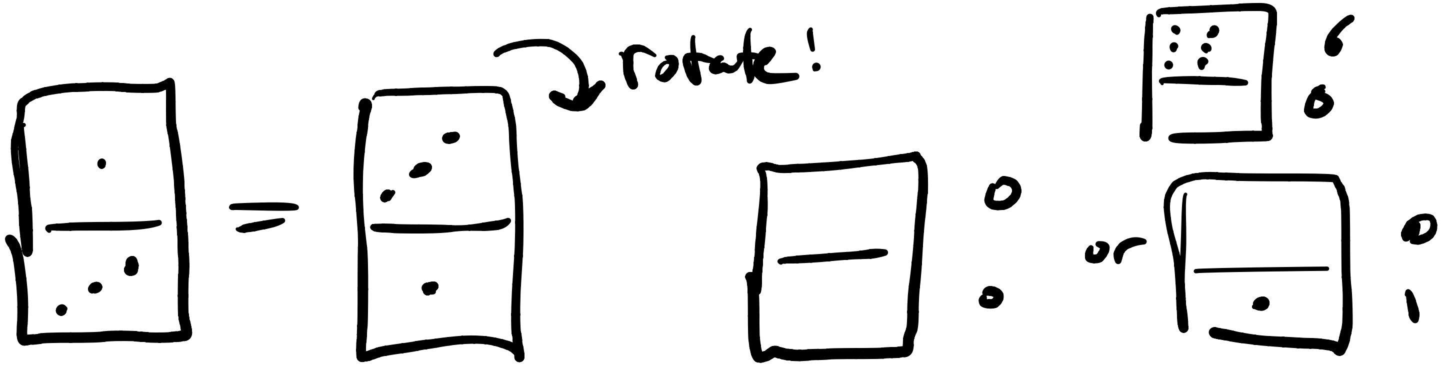
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \quad .33$$

B is not $\frac{1}{2}$ but $\frac{1}{3}$

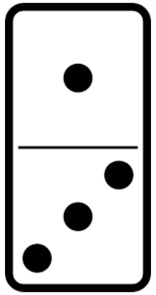
Example: Dominoes

(source: FiveThirtyEight)



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.

(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)



$P(\text{Double}) = \frac{7}{28} \left(\frac{1}{4}\right)$

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**.