Updated 9/18/2025

Part 1: Information Theory Basics
L02: Basics of Probability (1/2)
[Random experiment, independence, conditional probability, chain rule, Bayes' theorem, random variables]

Wolfgang Gatterbauer

cs7840 Foundations and Applications of Information Theory (fa25)

https://northeastern-datalab.github.io/cs7840/fa25/

9/11/2025

Pre-class conversations

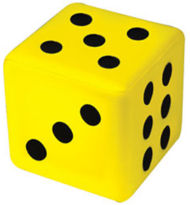
- Last class recapitulation
- Organizational matters: Does every one get Piazza messages?
- First scribe on Piazza. Awesome! I will look over the weekend
- Office hours: Usually right after class, or via email / Teams. Also feel free talk to me regularly during break/ after class about project ideas, or papers/research directions to include towards the end
- New class arrivals

- Today:
  - The basics of probability theory

Basics of probability theory

Following slides are built upon examples by Jay Aslam from earlier editions of this class. For a more extensive cover of the basics, I recommend "Bertsekas, Tsitsiklis. Introduction to Probability, 2008." It's a solid textbook on probability theory and I regularly find myself going back to this book to look up basic concepts. Working by yourself through chapter 1 on "Sample Space and Probability" is a good investment of your time.

Sample space / outcomes / events / probability



PROBABILITY

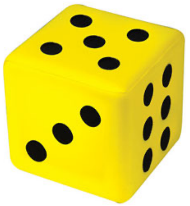
- A random experiment
- Sample space Ω: set of all possible outcomes

EXAMPLE 1:

roll a fair die with 6 sides

$\Omega =$?

Sample space / outcomes / events / probability



PROBABILITY

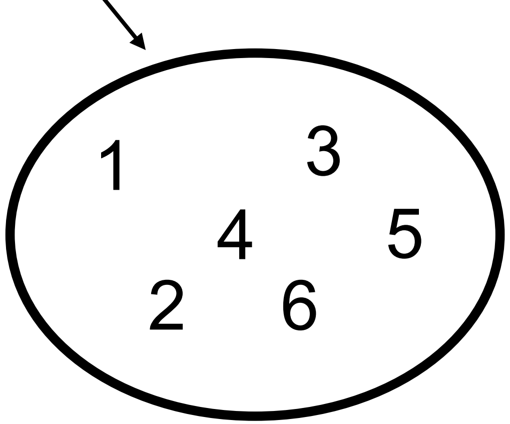
- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.

EXAMPLE 1:

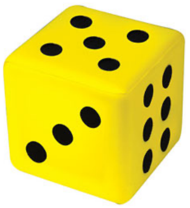
roll a fair die with 6 sides

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

?



Sample space / outcomes / events / probability



PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$

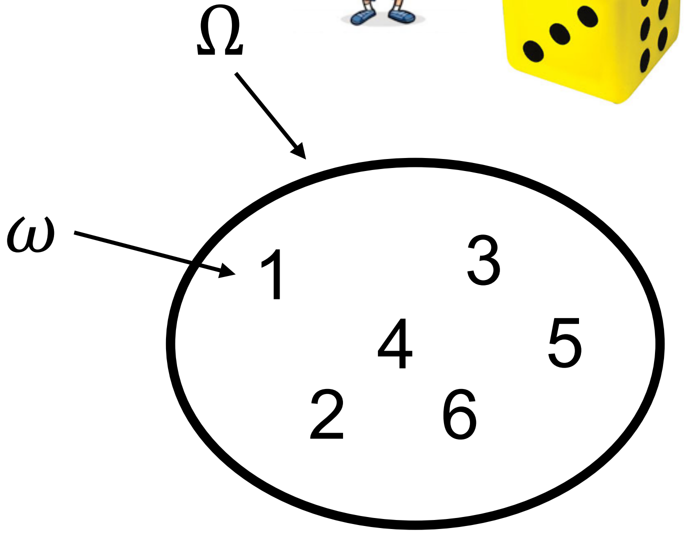
EXAMPLE 1:

roll a fair die with 6 sides

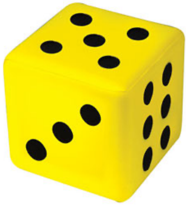
$\Omega = \{1, 2, 3, 4, 5, 6\}$

Say the outcome is a 1

?



Sample space / outcomes / events / probability



PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$

EXAMPLE 1:

roll a fair die with 6 sides

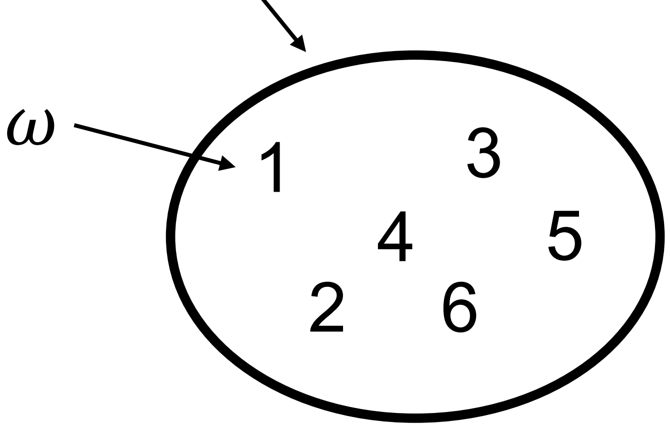
$\Omega = \{1, 2, 3, 4, 5, 6\}$

Say the outcome is a 1

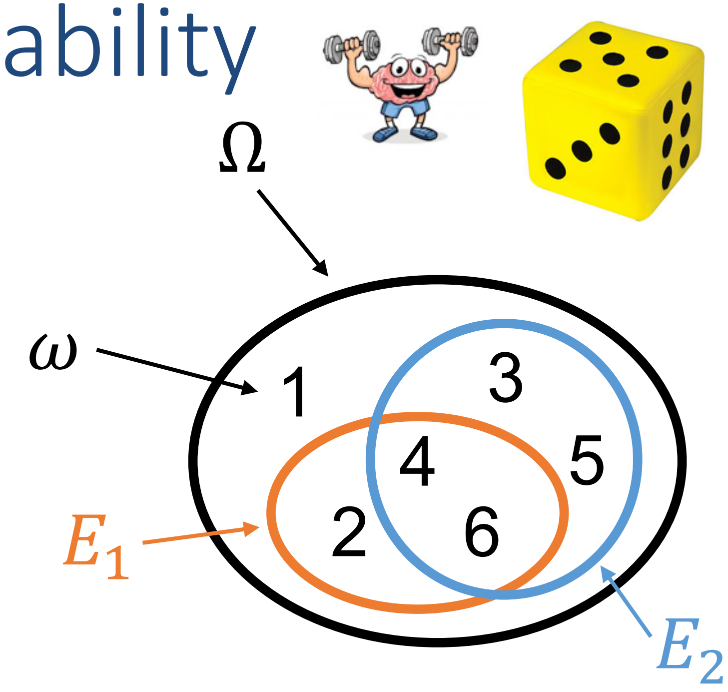
$E_1 = \text{"even"}$

$E_2 = \text{"}{\geq}3\text{"}$

?



Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:

?

EXAMPLE 1:

roll a fair die with 6 sides

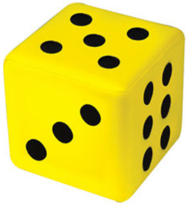
$\Omega = \{1, 2, 3, 4, 5, 6\}$

Say the outcome is a 1

$E_1 = \text{"even"} = \{2, 4, 6\}$

$E_2 = \text{"}{\geq}3\text{"} = \{3, 4, 5, 6\}$

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: ~~$\mathbb{P}: \Omega \rightarrow [0, 1]$~~

?

EXAMPLE 1:

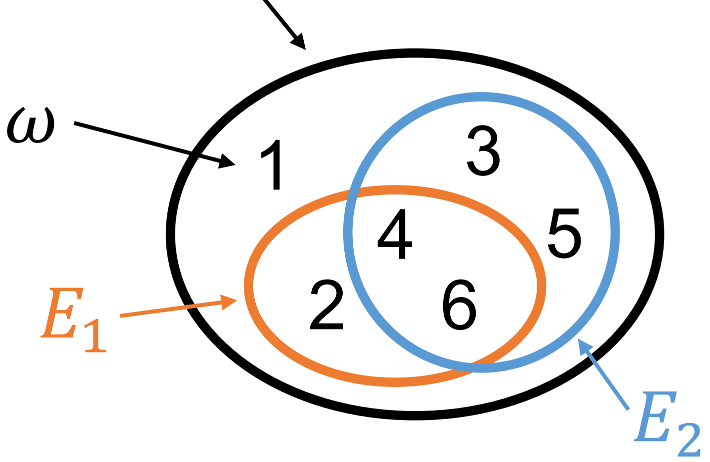
roll a fair die with 6 sides

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

Say the outcome is a 1

$E_1 = \text{"even"} = \{2, 4, 6\}$

$E_2 = \text{"}\geq 3\text{"} = \{3, 4, 5, 6\}$



Sample space / outcomes / events / probability

PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:
  ~~$\mathbb{P}: \Omega \to [0,1]$~~
  $\mathbb{P}: \Sigma \to [0,1]$

EXAMPLE 1:

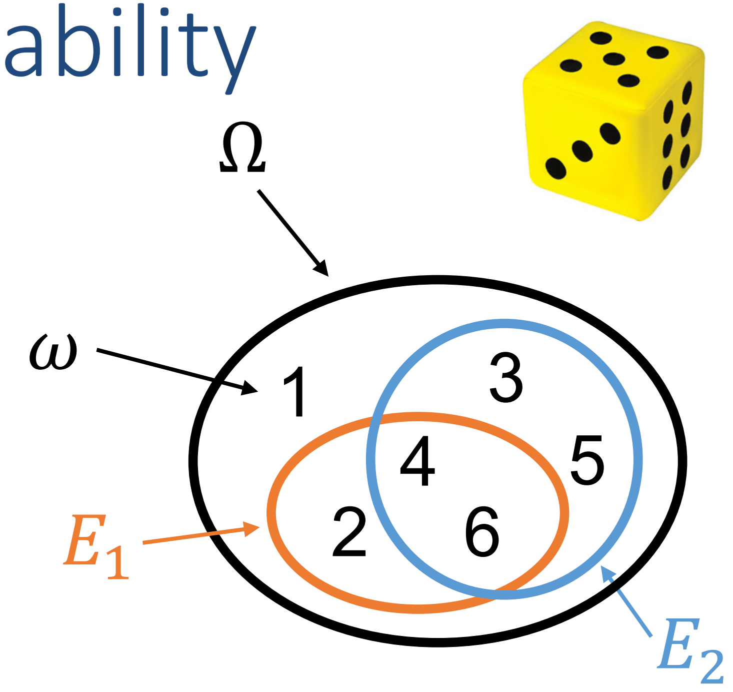
roll a fair die with 6 sides

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

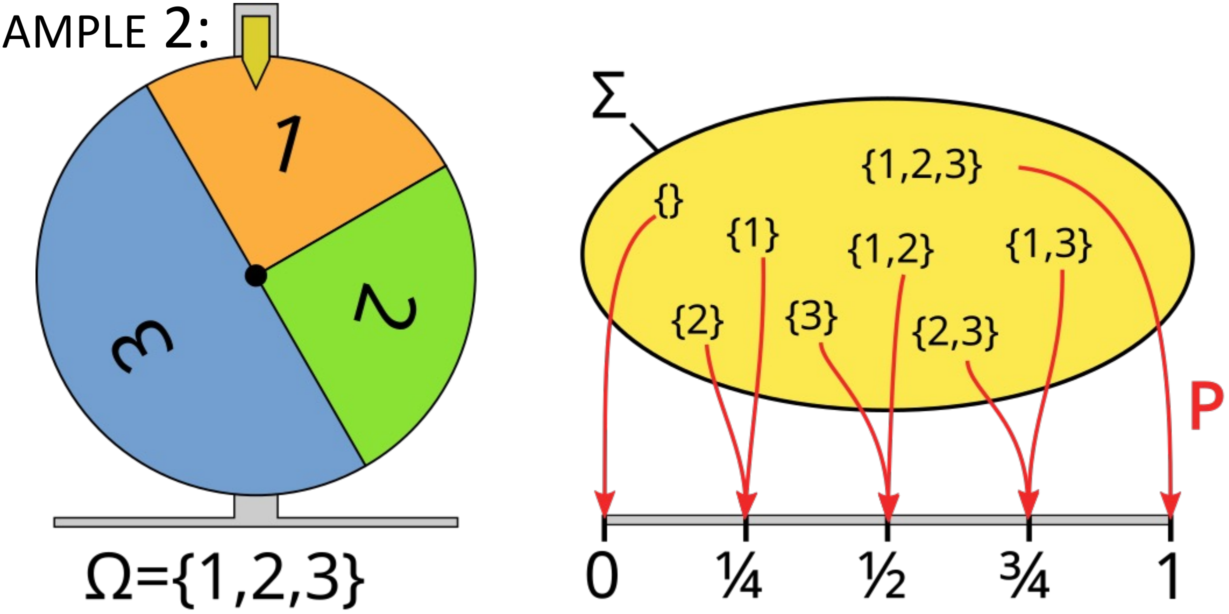
Say the outcome is a 1

$E_1 = \text{"even"} = \{2, 4, 6\}$

$E_2 = \text{"}\geq 3\text{"} = \{3, 4, 5, 6\}$



EXAMPLE 2:



The set Σ of all events E ∈ Σ is a σ-algebra, basically the same as the power set 2^{Ω}, except for some pathological non-measurable sets like the Vitali-Set (https://www.youtube.com/watch?v=hs3eDa3_DzU). See also: https://en.wikipedia.org/wiki/%CE%A3-algebra

Source of figure to the right: https://en.wikipedia.org/wiki/Probability_measure

Sample space / outcomes / events / probability

PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:
  ~~$\mathbb{P}: \Omega \to [0,1]$~~
  $\mathbb{P}: \Sigma \to [0,1]$

also $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$

1) <u>Nonnegativity:</u> $\mathbb{P}(E) \geq 0, \forall E \in \Sigma$
2) <u>Normalization:</u> $\mathbb{P}(\Omega) = 1$
3) <u>Additivity:</u> if E_1 and E_2 are disjoint events, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$,

The set Σ of all events $E \in \Sigma$ is a σ-algebra, basically the same as the power set 2^Ω, except for some pathological non-measurable sets like the Vitali-Set (https://www.youtube.com/watch?v=hs3eDa3_DzU). See also: https://en.wikipedia.org/wiki/%CE%A3-algebra

EXAMPLE 1:

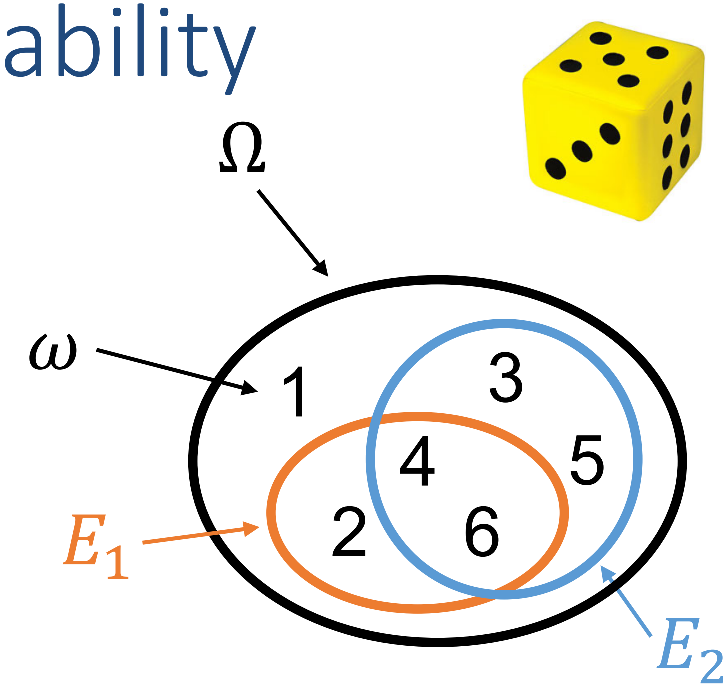
roll a fair die with 6 sides

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

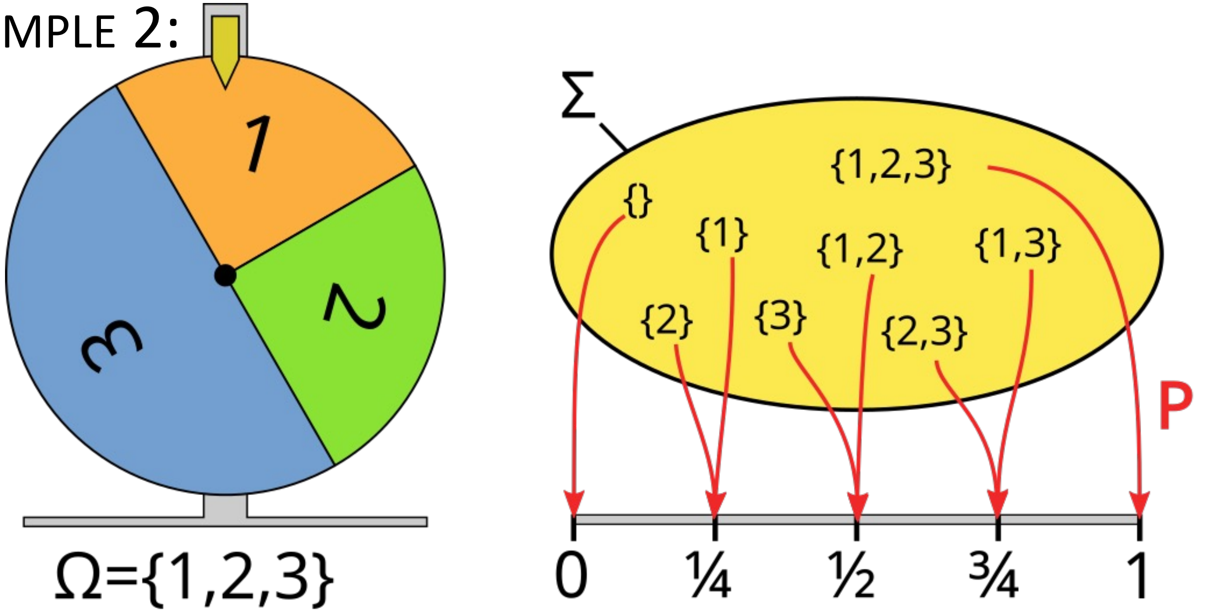
Say the outcome is a 1

E_1 = "even" $= \{2, 4, 6\}$

E_2 = "≥3" $= \{3, 4, 5, 6\}$

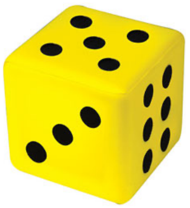


EXAMPLE 2:



Source of figure to the right: https://en.wikipedia.org/wiki/Probability_measure

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:
  ~~$\mathbb{P}: \Omega \to [0,1]$~~
  $\mathbb{P}: \Sigma \to [0,1]$

also $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$

1) Nonnegativity: $\mathbb{P}(E) \geq 0, \forall E \in \Sigma$
2) Normalization: $\mathbb{P}(\Omega) = 1$
3) Additivity: if E_1 and E_2 are disjoint events, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$,

EXAMPLE 1:

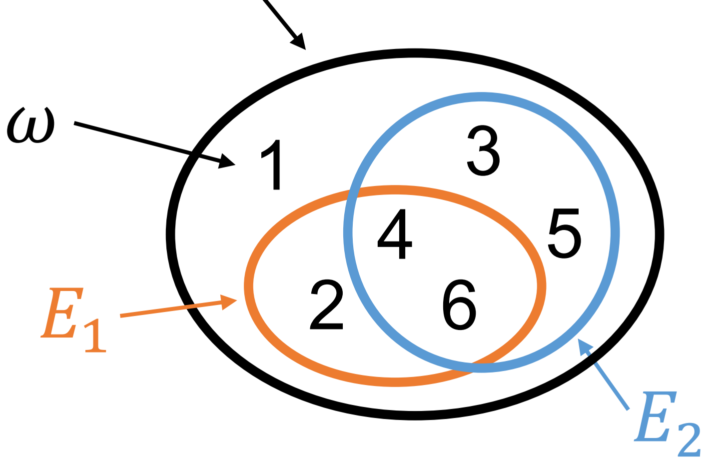
roll a fair die with 6 sides

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

Say the outcome is a 1

$E_1 = \text{"even"} = \{2, 4, 6\}$

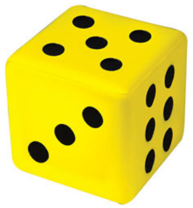
$E_2 = \text{"}{\geq}3\text{"} = \{3, 4, 5, 6\}$



short for $\mathbb{P}(\{\omega\})$

- If $\mathbb{P}(\omega) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$, then $\mathbb{P}(E) =$?

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:
  ~~$\mathbb{P}: \Omega \to [0,1]$~~
  $\mathbb{P}: \Sigma \to [0,1]$
  also $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$
  1) Nonnegativity: $\mathbb{P}(E) \geq 0, \forall E \in \Sigma$
  2) Normalization: $\mathbb{P}(\Omega) = 1$
  3) Additivity: if E_1 and E_2 are disjoint events, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$,

EXAMPLE 1:

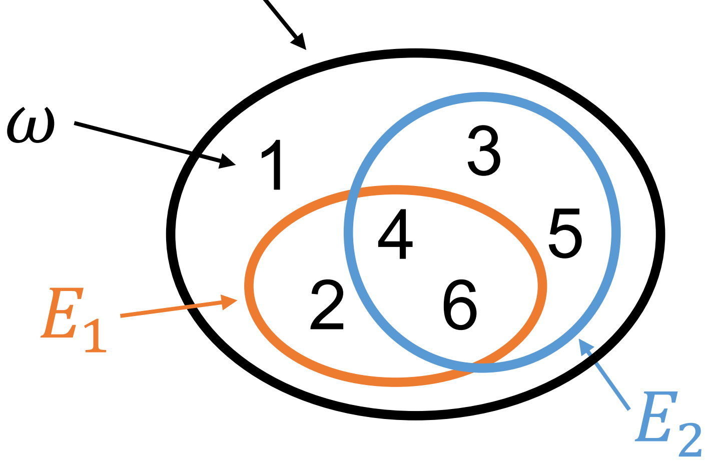
roll a fair die with 6 sides

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

Say the outcome is a 1

$E_1 = \text{"even"} = \{2, 4, 6\}$

$E_2 = \text{"}\geq 3\text{"} = \{3, 4, 5, 6\}$



short for $\mathbb{P}(\{\omega\})$

- If $\mathbb{P}(\omega) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$, then $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$

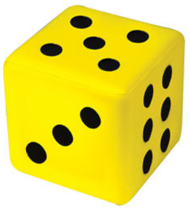
$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}$

$\mathbb{P}(E_1) =$

$\mathbb{P}(E_2) =$

?

Sample space / outcomes / events / probability



PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure:
  ~~$\mathbb{P}: \Omega \to [0,1]$~~
  $\mathbb{P}: \Sigma \to [0,1]$

also $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$

1) <u>Nonnegativity:</u> $\mathbb{P}(E) \geq 0, \forall E \in \Sigma$
2) <u>Normalization:</u> $\mathbb{P}(\Omega) = 1$
3) <u>Additivity:</u> if E_1 and E_2 are disjoint events, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$,

EXAMPLE 1:

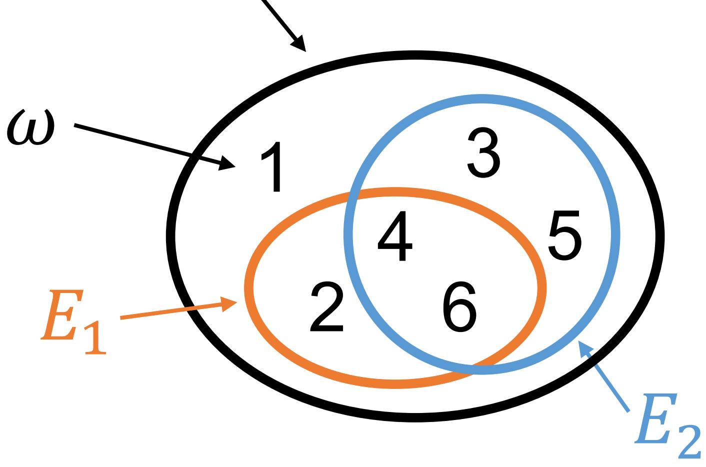
roll a fair die with 6 sides

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

Say the outcome is a 1

E_1 = "even" = $\{2, 4, 6\}$

E_2 = "≥ 3" = $\{3, 4, 5, 6\}$



short for $\mathbb{P}(\{\omega\})$

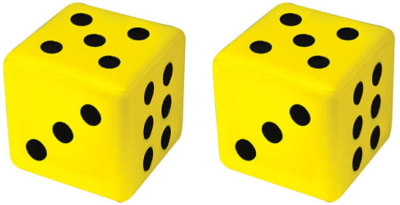
- If $\mathbb{P}(\omega) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$, then $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$

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

$$\mathbb{P}(E_1) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(E_2) = \frac{4}{6} = \frac{2}{3}$$

Sample space / outcomes / events / probability



PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

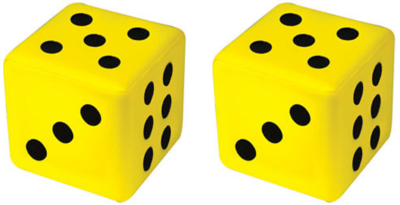
roll two fair dice with 6 sides

$\Omega =$

?

$|\Omega| =$

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

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

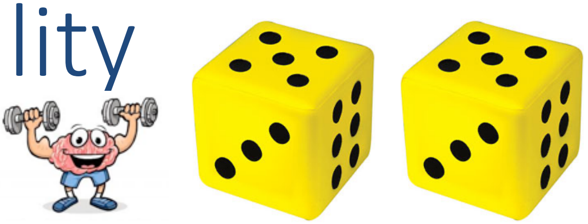
$$|\Omega| = 36$$

second die

first die

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Sample space / outcomes / events / probability



PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \to [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

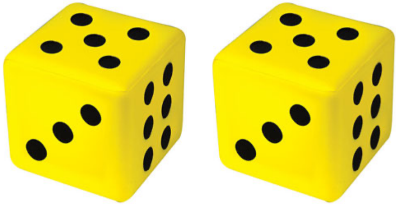
$$E_1 = \text{"sum is 7"}$$

$$|E_1| =$$

$$\mathbb{P}(E_1) =$$

?

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \to [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$
$$= \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$$

$$|\Omega| = 36$$

second die

first die

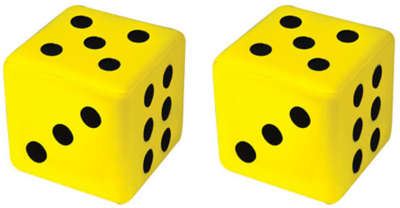
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$E_1 = \text{"sum is 7"}$$
$$= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

$$|E_1| = 6$$

$$\mathbb{P}(E_1) = \frac{|E_1|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \to [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

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

$$|\Omega| = 36$$

second die

first die

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

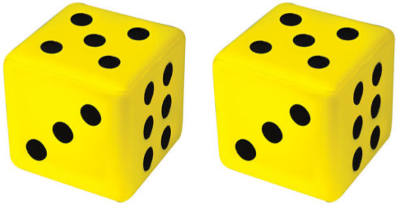
$$E_2 = \text{"total is greater than 8"}$$

$$|E_2| =$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

?

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \to [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$= \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$$

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

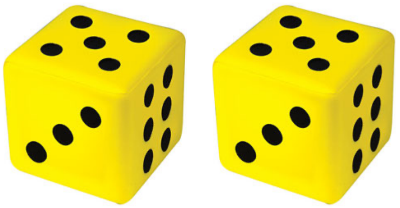
$$E_2 = \text{"total is greater than 8"}$$

$$= \text{"total is 9 or 10 or 11 or 12"}$$

$$|E_2| =$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

Sample space / outcomes / events / probability



PROBABILITY

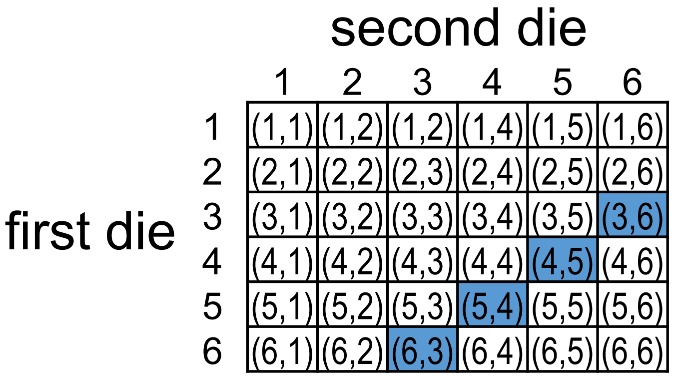
- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$



second die

first die

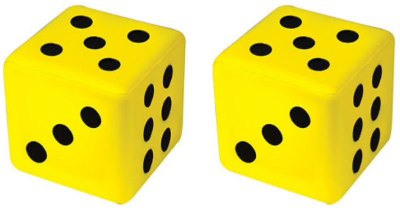
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$E_2 = \text{"total is greater than 8"}$$
$$= \text{"total is } \underset{\substack{\downarrow \\ 4}}{9} \text{ or 10 or 11 or 12"}$$

$$|E_2| = \quad 4$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

Sample space / outcomes / events / probability



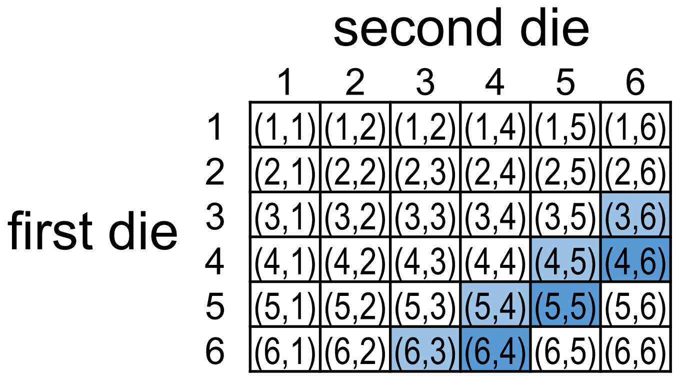
PROBABILITY

- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$$
$$= \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$$
$$|\Omega| = 36$$



second die

first die

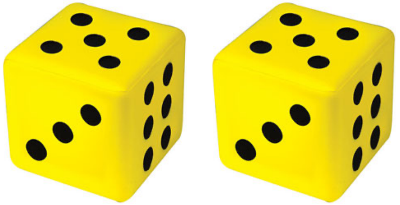
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$E_2 = \text{"total is greater than 8"}$$
$$= \text{"total is 9 or 10 or 11 or 12"}$$

$$|E_2| = \quad 4 + 3$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

Sample space / outcomes / events / probability



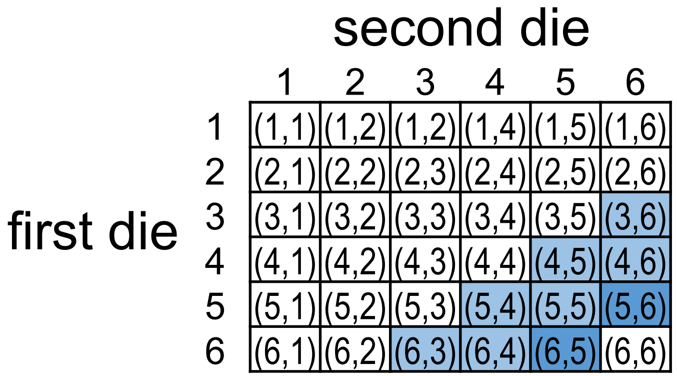
PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \to [0,1]$

EXAMPLE 3:

roll two fair dice with 6 sides

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$$
$$= \{(1,1),(1,2),(1,3)\ldots(1,6),(2,1),(2,2),\ldots(6,6)\}$$
$$|\Omega| = 36$$



second die

first die

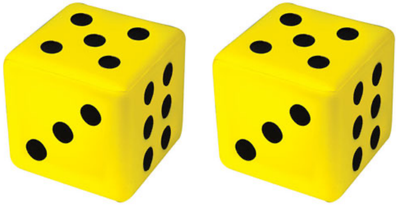
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$E_2 = \text{"total is greater than 8"}$$
$$= \text{"total is 9 or 10 or 11 or 12"}$$

$$|E_2| = \quad \underset{\uparrow}{4} + \underset{\uparrow}{3} + \underset{\uparrow}{2}$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

Sample space / outcomes / events / probability



PROBABILITY

- A random experiment
- Sample space Ω: set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

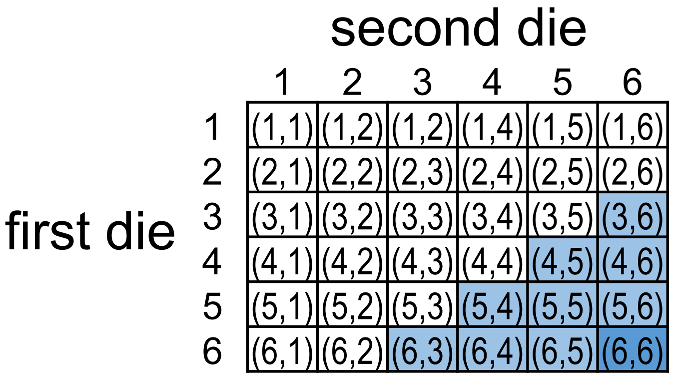
EXAMPLE 3:

roll two fair dice with 6 sides

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

$$= \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$$

$$|\Omega| = 36$$



second die

first die

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

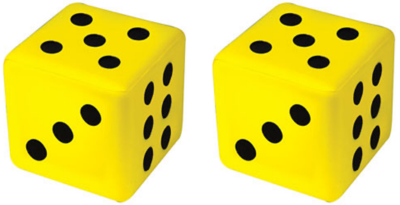
$$E_2 = \text{"total is greater than 8"}$$

$$= \text{"total is 9 or 10 or 11 or 12"}$$

$$|E_2| = 4 + 3 + 2 + 1$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|}$$

Sample space / outcomes / events / probability



PROBABILITY

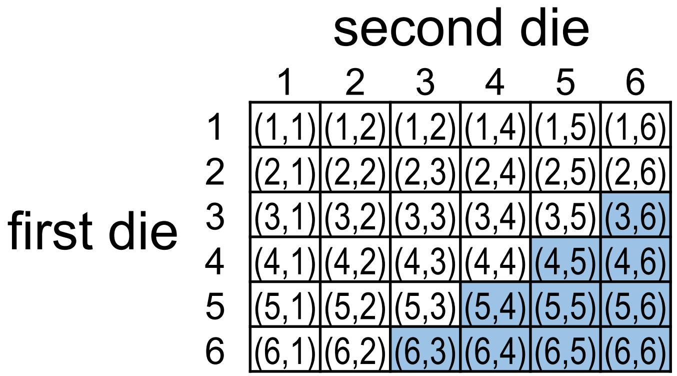
- A <u>random experiment</u>
- <u>Sample space</u> Ω: set of all possible outcomes
- An <u>outcome</u> $\omega \in \Omega$ of the exp.
- <u>Event</u> E: a subset of the sample space $E \subseteq \Omega$
- Probability measure: $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$



second die

first die

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

$$E_2 = \text{"total is greater than 8"}$$
$$= \text{"total is 9 or 10 or 11 or 12"}$$

additivity of disjoint events

$$|E_2| = \quad 4 + 3 + 2 + 1 = 10$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|} = \frac{10}{36} = \frac{5}{18}$$

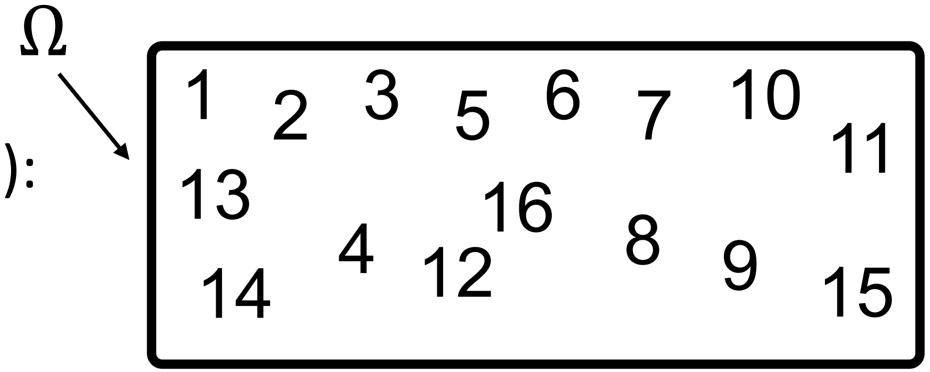
Conditional Probabilities / Independence

Conditional probability

Definition

Conditional probability $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

Example 4
(fair 16-sidedxDie):



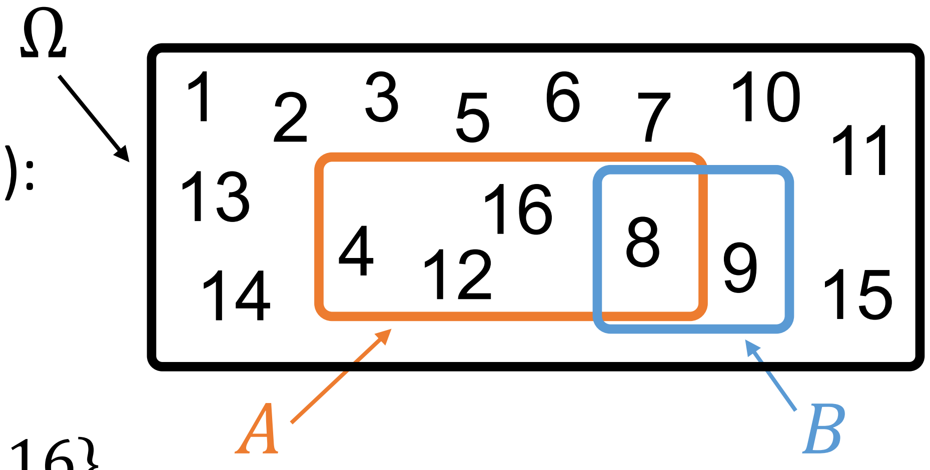
$$\Omega = \{1, 2, \ldots, 16\}$$

Conditional probability

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

EXAMPLE 4
(FAIR 16-SIDEDXDIE):



$$\Omega = \{1, 2, \dots, 16\}$$

$$A = \text{"divisible by 4"} = \{4, 8, 12, 16\}$$

$$B = \text{"center numbers"} = \{8, 9\}$$

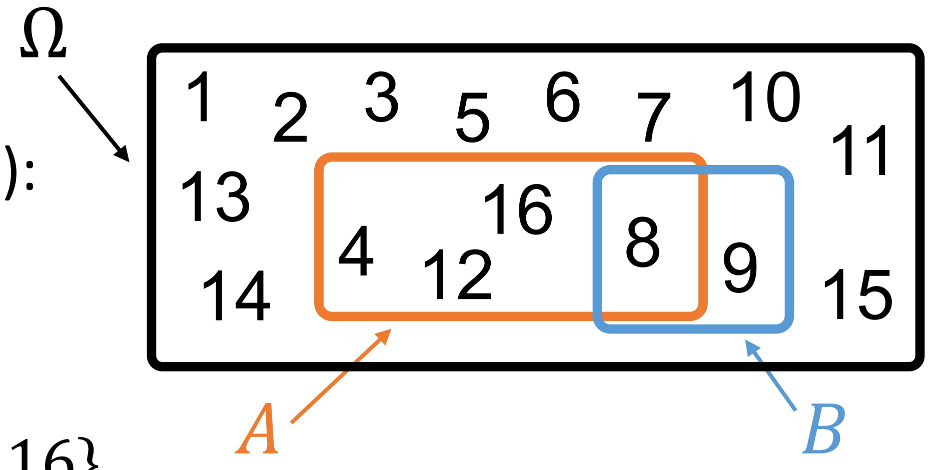
$$\mathbb{P}(A|B) = ?$$

Conditional probability

DEFINITION

Conditional probability $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

EXAMPLE 4
(FAIR 16-SIDEDXDIE):



$$\Omega = \{1, 2, \ldots, 16\}$$

$$A = \text{"divisible by 4"} = \{4, 8, 12, 16\}$$

$$B = \text{"center numbers"} = \{8, 9\}$$

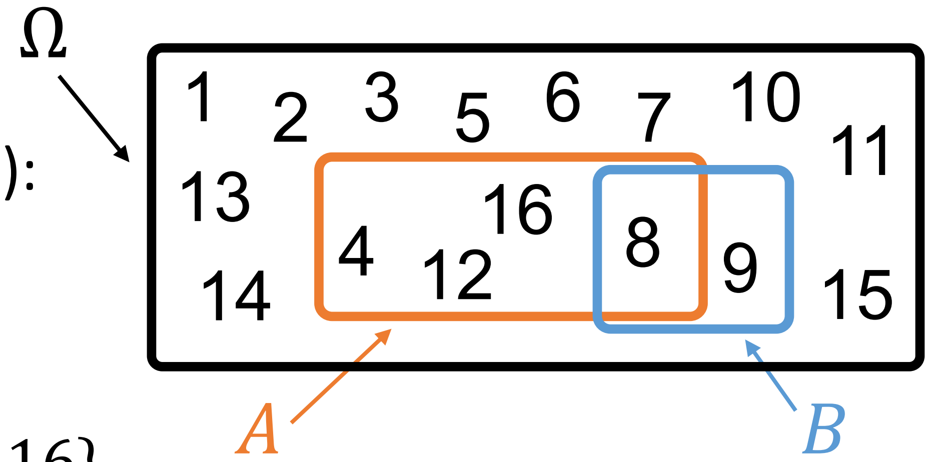
$$\mathbb{P}(A|B) = \frac{1}{2}$$

Conditional probability

DEFINITION

Conditional probability $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

EXAMPLE 4
(FAIR 16-SIDEDXDIE):

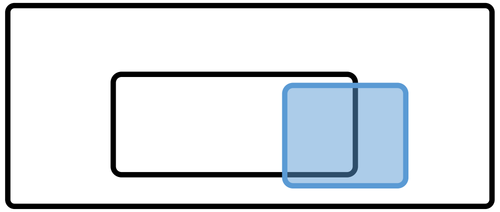


$$\Omega = \{1, 2, \ldots, 16\}$$

$$A = \text{"divisible by 4"} = \{4, 8, 12, 16\}$$

$$B = \text{"center numbers"} = \{8, 9\}$$

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

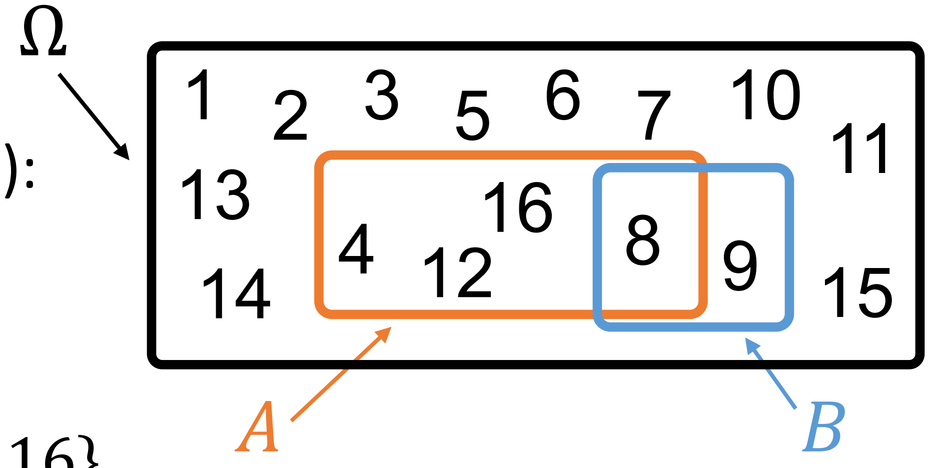


Conditional probability

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

EXAMPLE 4
(FAIR 16-SIDEDXDIE):

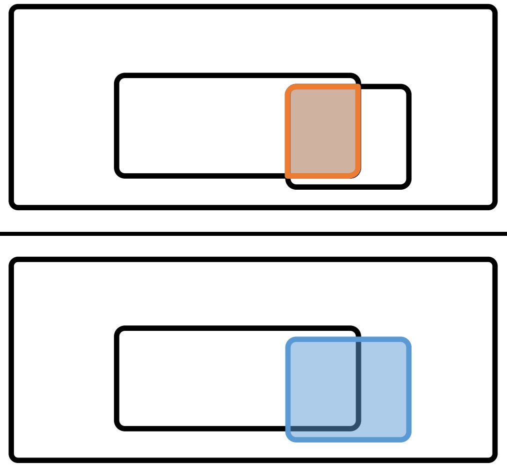


$$\Omega = \{1, 2, \ldots, 16\}$$

$$A = \text{"divisible by 4"} = \{4, 8, 12, 16\}$$

$$B = \text{"center numbers"} = \{8, 9\}$$

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

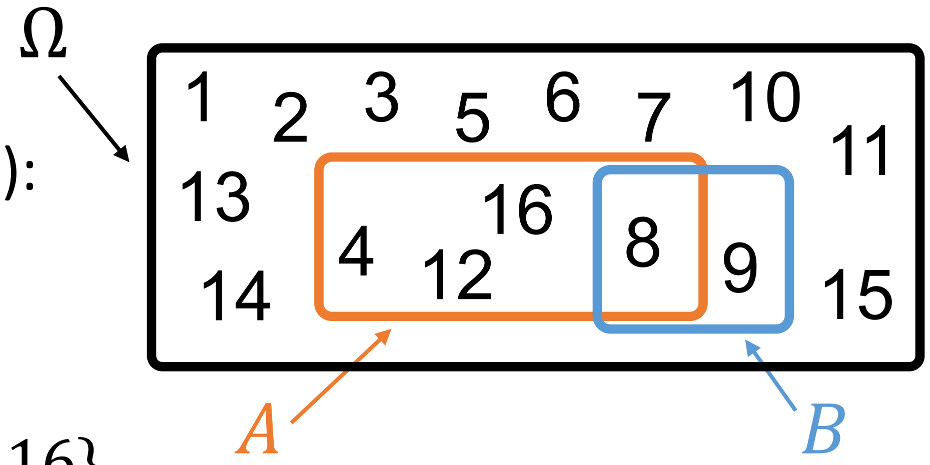


Conditional probability

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

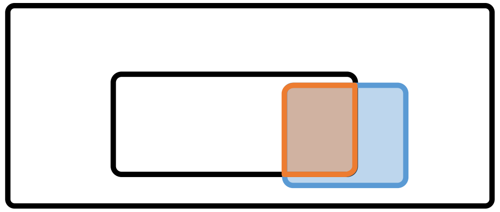
EXAMPLE 4
(FAIR 16-SIDEDXDIE):



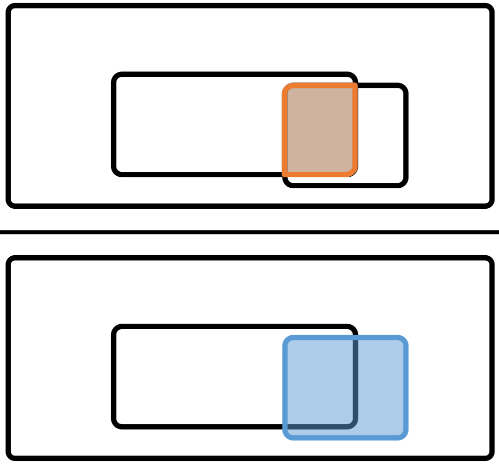
$$\Omega = \{1, 2, \ldots, 16\}$$

$$A = \text{"divisible by 4"} = \{4, 8, 12, 16\}$$

$$B = \text{"center numbers"} = \{8, 9\}$$



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



Conditional probability

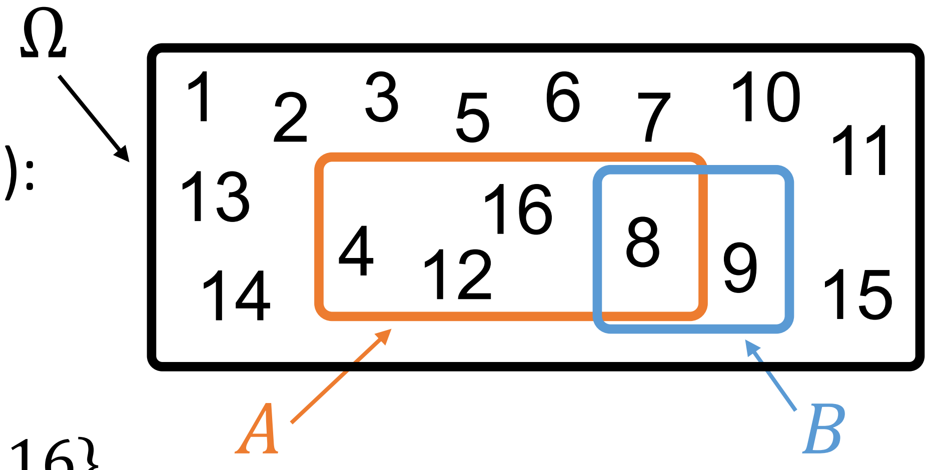
DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred

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

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

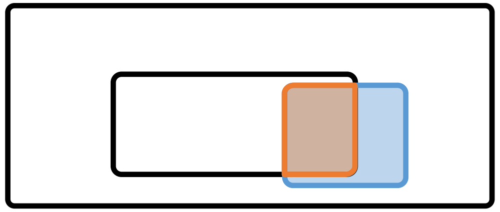
EXAMPLE 4 (FAIR 16-SIDEDXDIE):



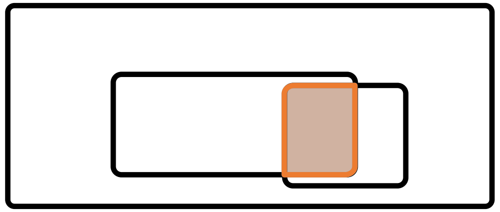
$\Omega = \{1, 2, \dots, 16\}$

A = "divisible by 4" $= \{4, 8, 12, 16\}$

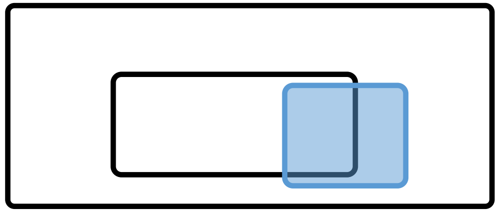
B = "center numbers" $= \{8, 9\}$



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



$\frac{1}{16}$



$\frac{2}{16}$

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

Independence

DEFINITION:

Two events are independent if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad (\text{if } \mathbb{P}(A) \neq 0)$$

Independence

DEFINITION:

Two events are <u>independent</u> if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad (\text{if } \mathbb{P}(A) \neq 0)$$

EXAMPLE 1 (CONTINUED):

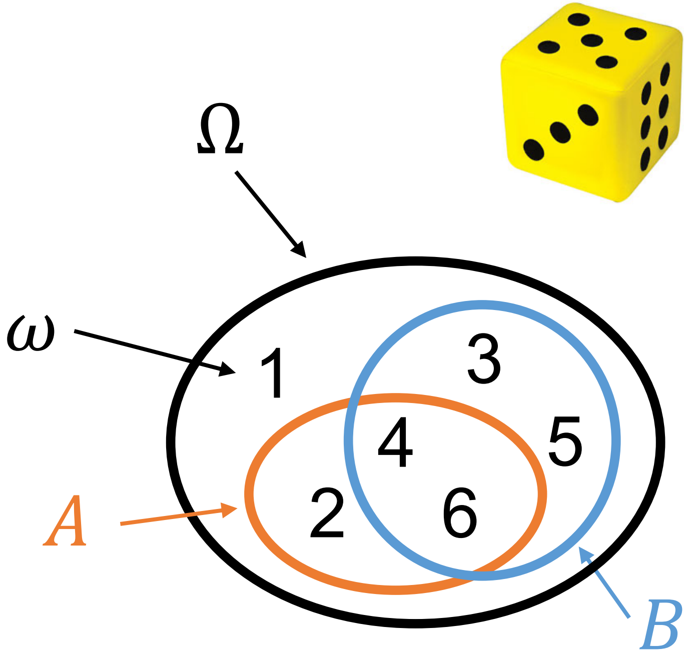
roll a fair die with 6 sides

the outcome is a 1

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

$$A = \text{"even"} = \{2, 4, 6\}$$

$$B = \text{"}\geq 3\text{"} = \{3, 4, 5, 6\}$$



Are A and B independent ?

Independence

DEFINITION:

Two events are <u>independent</u> if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

EXAMPLE 1 (CONTINUED):

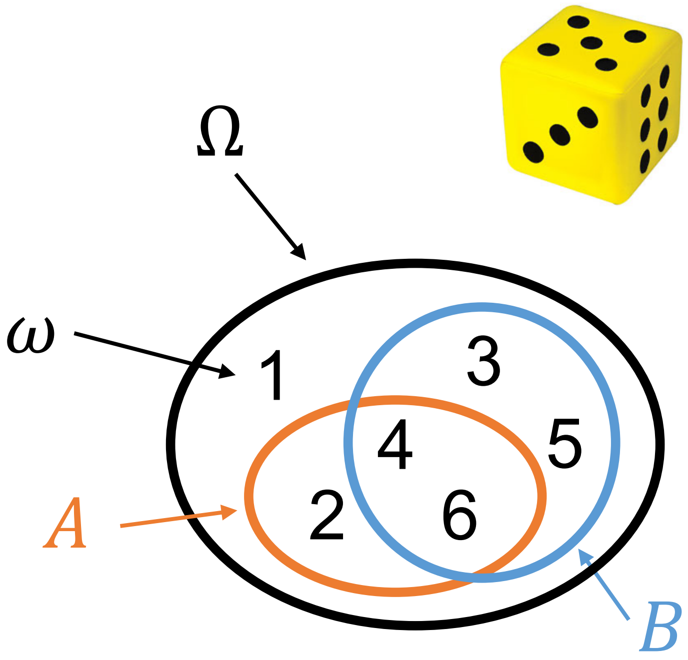
roll a fair die with 6 sides

the outcome is a 1

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

$$A = \text{"even"} = \{2, 4, 6\}$$

$$B = \text{"}\geq 3\text{"} = \{3, 4, 5, 6\}$$



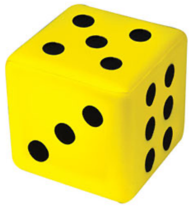
A and B are independent !

$$\mathbb{P}(A) = \frac{1}{2}$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} = \frac{1}{2}$$

1	3	5
2	4	6

Independence



DEFINITION:

Two events are <u>independent</u> if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

EXAMPLE 1 (CONTINUED):

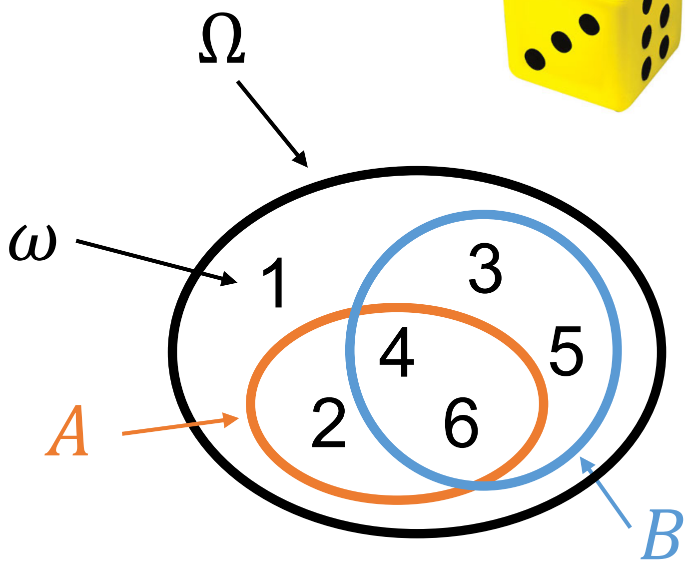
roll a fair die with 6 sides

the outcome is a 1

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

$A = \text{"even"} = \{2, 4, 6\}$

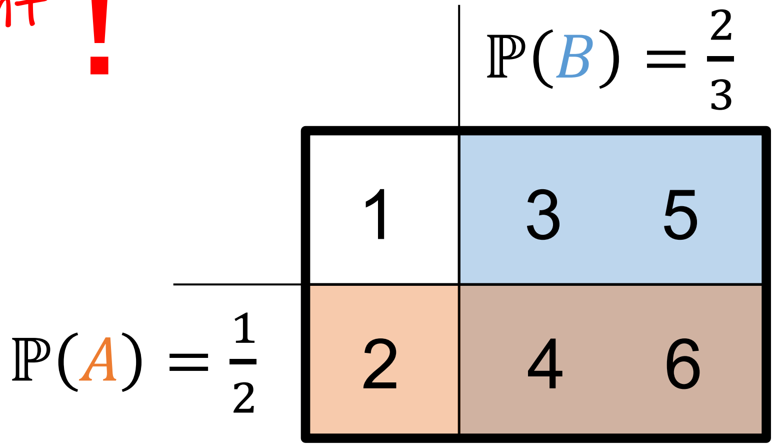
$B = \text{"}{\geq}3\text{"} = \{3, 4, 5, 6\}$



A and B are independent !

$$\mathbb{P}(A) = \frac{1}{2}$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} = \frac{1}{2}$$



$\mathbb{P}(B) = \frac{2}{3}$

$\mathbb{P}(A) = \frac{1}{2}$

1	3	5
2	4	6

Think about a coordinate system:
x and y are independent

Independence

DEFINITION:

Two events are <u>independent</u> if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

Recall the definition of <u>cond. prob.</u>:

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

CLAIM

Two events A and B are <u>independent</u> if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

PROOF (assuming $\mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0$)

?

Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Independence

DEFINITION:

Two events are independent if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

Recall the definition of cond. prob.:

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

CLAIM

Two events A and B are independent if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

PROOF (assuming $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$)

(⇒) If $\mathbb{P}(A|B) = \mathbb{P}(A)$:

?

(⇐) If $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$:

?

Independence

DEFINITION:

Two events are <u>independent</u> if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

Recall the definition of <u>cond. prob.</u>:

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

CLAIM

Two events A and B are <u>independent</u> if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

PROOF (assuming $\mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0$)

($\Rightarrow$) If $\mathbb{P}(A|B) = \mathbb{P}(A)$:

Then $\mathbb{P}(A,B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

($\Leftarrow$) If $\mathbb{P}(A,B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\underbrace{\mathbb{P}(B)}_{\neq 0}} = \frac{\mathbb{P}(A) \cdot \cancel{\mathbb{P}(B)}}{\cancel{\mathbb{P}(B)}} = \mathbb{P}(A)$$

Independence

DEFINITION:

Two events are independent if the probability that one occurred is not affected by knowledge that the other occurred.

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad (\text{if } \mathbb{P}(B) \neq 0)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B) \qquad (\text{if } \mathbb{P}(A) \neq 0)$$

Recall the definition of cond. prob.:

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

Commonly used as definition of independence because it is slightly more general: if one event has probability 0 then the statement holds vacuously (while one conditional probability is undefined). And if both are not 0, then we just proved that to be equivalent to the earlier definition.

DEFINITION:

Two events A and B are independent $\Leftrightarrow$ $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

PROOF (assuming $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$)

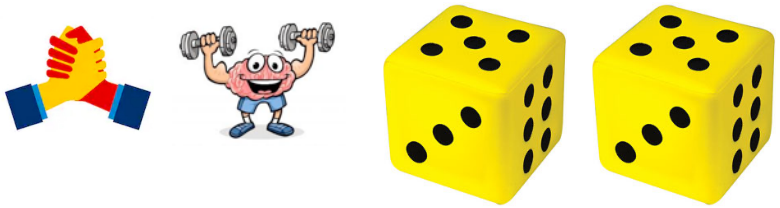
($\Rightarrow$) If $\mathbb{P}(A|B) = \mathbb{P}(A)$:

Then $\mathbb{P}(A,B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

($\Leftarrow$) If $\mathbb{P}(A,B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\underset{\neq 0}{\mathbb{P}(B)}} = \frac{\mathbb{P}(A) \cdot \cancel{\mathbb{P}(B)}}{\cancel{\mathbb{P}(B)}} = \mathbb{P}(A)$$

Independence



DEFINITION:

Two events A and B are independent if
$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

EXAMPLE 3 (CONTINUED):

roll two fair dice with 6 sides

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

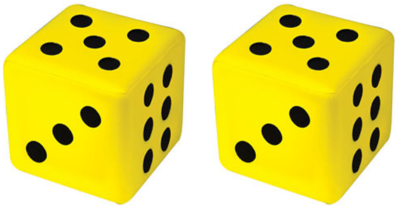
A = "1st roll is 1"

B = "sum is 7"

Are A and B independent? ?

Example adapted from example 1.19 of [Bertsekas, Tsitsiklis. Introduction to probability, 2nd ed, 2008]

Independence



DEFINITION:

Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

EXAMPLE 3 (CONTINUED):

roll two fair dice with 6 sides

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

$$A = \text{"1}^{\text{st}}\text{ roll is 1"}$$

$$B = \text{"sum is 7"}$$

Are A and B independent?

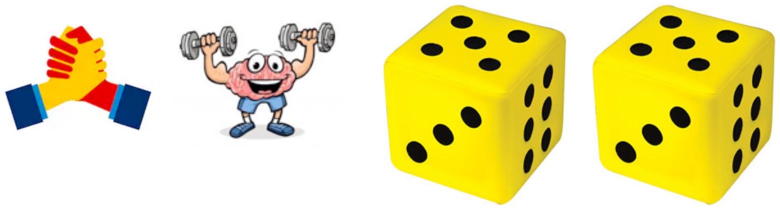
$\mathbb{P}(A) =$?

$\mathbb{P}(B) =$?

$\mathbb{P}(A, B) =$?

Example adapted from example 1.19 of [Bertsekas, Tsitsiklis. Introduction to probability, 2nd ed, 2008]

Independence



DEFINITION:

Two events A and B are <u>independent</u> if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

EXAMPLE 3 (CONTINUED):

roll two fair dice with 6 sides

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

$$A = \text{"1}^{\text{st}}\text{ roll is 1"}$$

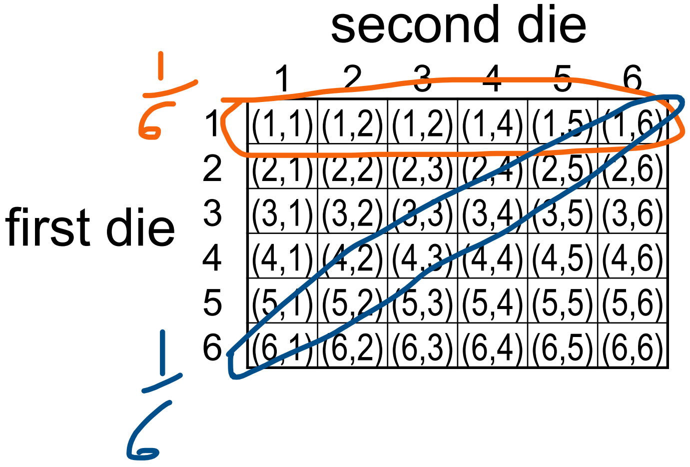
$$B = \text{"sum is 7"}$$

Are A and B independent?

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

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

$$\mathbb{P}(A, B) = \ ?$$



second die

first die

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

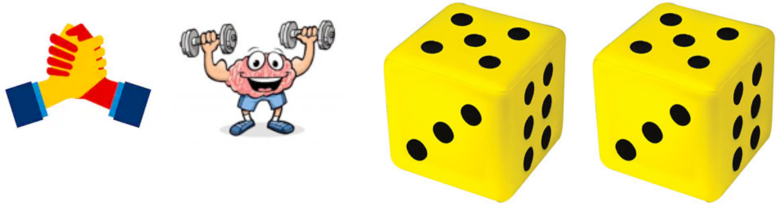
$\frac{1}{6}$

$\frac{1}{6}$

Example adapted from example 1.19 of [Bertsekas, Tsitsiklis. Introduction to probability, 2nd ed, 2008]

Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Independence



DEFINITION:

Two events *A* and *B* are <u>independent</u> if
$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

EXAMPLE 3 (CONTINUED):

roll two fair dice with 6 sides

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

$$A = \text{"}1^{\text{st}}\text{ roll is 1"}$$

$$B = \text{"sum is 7"}$$

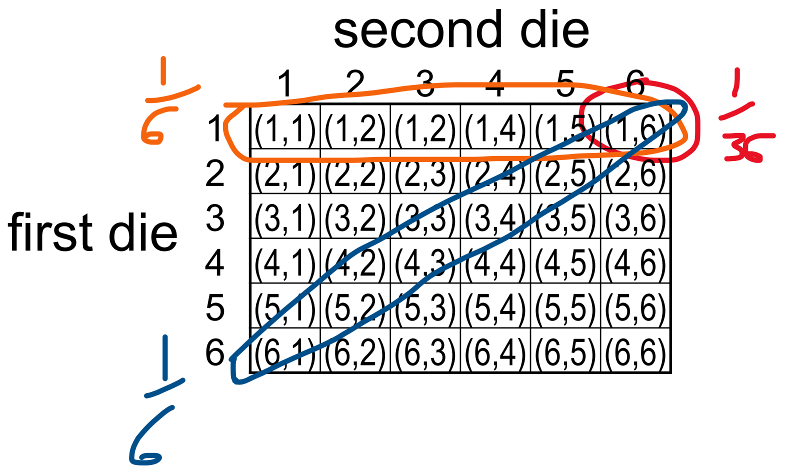
Are *A* and *B* independent?

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

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

$$\mathbb{P}(A, B) = \frac{1}{36}$$

Yes: $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$



second die

first die

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

$\frac{1}{6}$

$\frac{1}{36}$

$\frac{1}{6}$

Example adapted from example 1.19 of [Bertsekas, Tsitsiklis. Introduction to probability, 2nd ed, 2008]

Chain rules

Chain Rule

DEFINITION:

$$\mathbb{P}(A,B) \quad = \mathbb{P}(A) \quad \cdot \mathbb{P}(B|A)$$

$$\mathbb{P}(A,B,C) = \quad ?$$

Recall the definition of cond. prob.:

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

Chain Rule

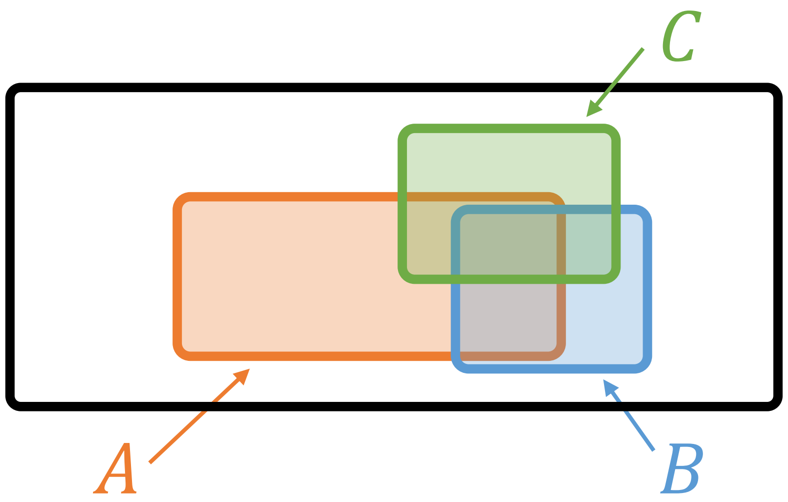
DEFINITION:

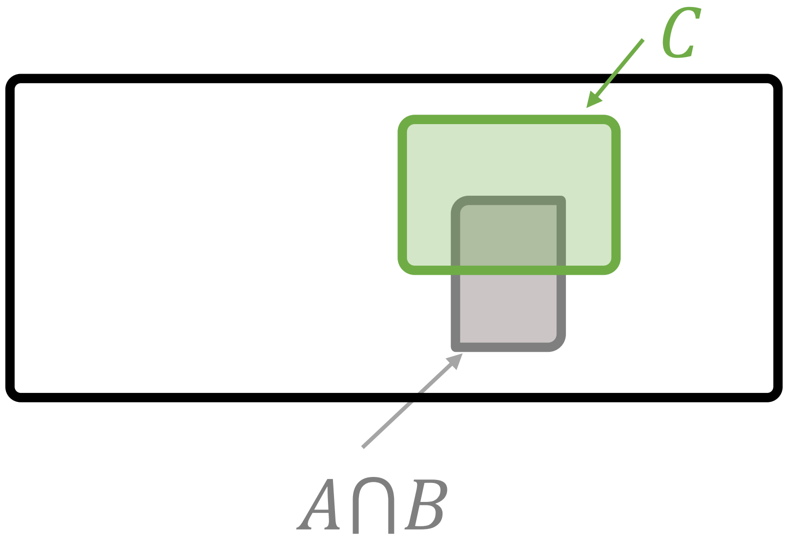
$$\mathbb{P}(A,B) \quad = \mathbb{P}(A) \quad \cdot \mathbb{P}(B|A)$$

$$\mathbb{P}(\underbrace{A,B}_{E},C) = \mathbb{P}(\underbrace{A,B}_{E}) \cdot \mathbb{P}(C|\underbrace{A,B}_{E})$$

Just treat $A \cap B$ as a new event E

$$\boxed{\mathbb{P}(A,B,C) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A,B)}$$





Recall the definition of cond. prob.:

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

Bayes law

Bayes Law

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred (defined if $\mathbb{P}(B) \neq 0$)

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

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

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

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

BAYES LAW

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

Bayes Law

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred (defined if $\mathbb{P}(B) \neq 0$)

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

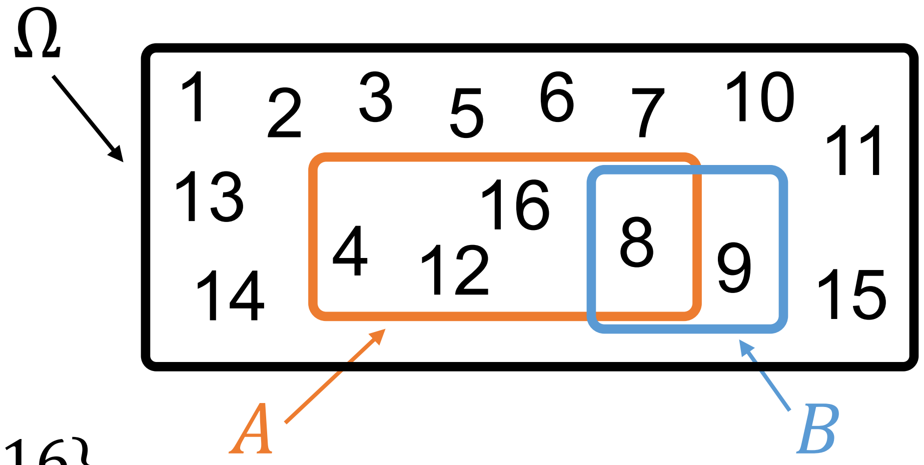
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(A)}$$

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

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

?

EXAMPLE 4:



$$\Omega = \{1, 2, \ldots, 16\}$$

BAYES LAW

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

Bayes Law

DEFINITION

<u>Conditional probability</u> $\mathbb{P}(A|B)$ is the probability that event A occurred, given that event B occurred (defined if $\mathbb{P}(B) \neq 0$)

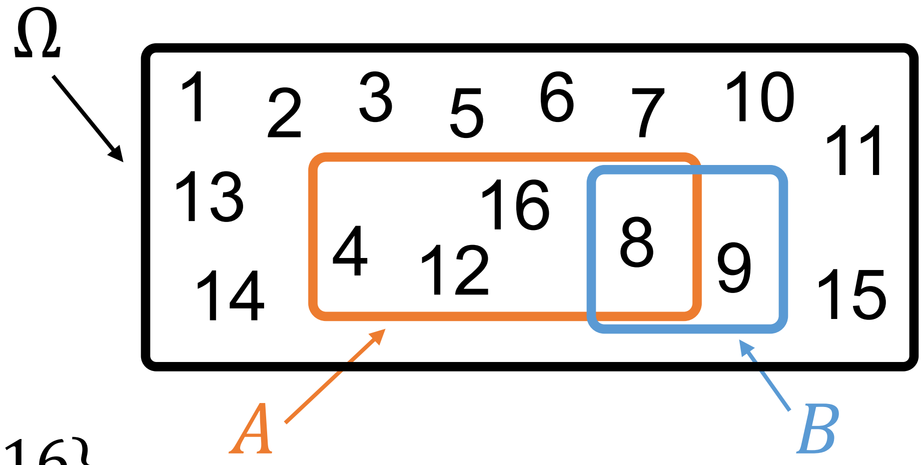
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)}$$

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

$$\mathbb{P}(A,B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$

$$= \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{8} = \frac{1}{16}$$

EXAMPLE 4:



$\Omega = \{1, 2, \ldots, 16\}$

BAYES LAW

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

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k)
- accuracy of blood test is 99%
- A patient has a positive test.
  What is the chance they have Zika?

?

blood test positive (B)

patient has Zika (Z)	B	$\bar{B}$
Z	$\mathbb{P}(B\|Z) = 0.99$	$\mathbb{P}(\bar{B}\|Z) = 0.01$
$\bar{Z}$	$\mathbb{P}(B\|\bar{Z}) = 0.01$	$\mathbb{P}(\bar{B}\|\bar{Z}) = 0.99$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = base rate = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- A patient has a positive test.
  What is the chance they have Zika?

?

		blood test positive (B)	
		B	$\bar{B}$
patient has Zika (Z)	Z	$\mathbb{P}(B\|Z) = 0.99$	$\mathbb{P}(\bar{B}\|Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B\|\bar{Z}) = 0.01$	$\mathbb{P}(\bar{B}\|\bar{Z}) = 0.99$

false negative

false positive

We don't want $\mathbb{P}(B|Z) = 0.99$, but rather $\mathbb{P}(Z|B)$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- A patient has a positive test.
  What is the chance they have Zika?

		blood test positive (B)	
		B	$\bar{B}$
patient has Zika (Z)	Z	$\mathbb{P}(B\|Z) = 0.99$	$\mathbb{P}(\bar{B}\|Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B\|\bar{Z}) = 0.01$	$\mathbb{P}(\bar{B}\|\bar{Z}) = 0.99$

We don't want $\mathbb{P}(B|Z) = 0.99$, but rather $\mathbb{P}(Z|B)$

$$\mathbb{P}(Z|B) = \frac{\mathbb{P}(B,Z)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|Z)\cdot\mathbb{P}(Z)}{?}$$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- A patient has a positive test.
  What is the chance they have Zika?

		blood test positive (B)	
		B	$\bar{B}$
patient has Zika (Z)	Z	$\mathbb{P}(B\|Z) = 0.99$	$\mathbb{P}(\bar{B}\|Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B\|\bar{Z}) = 0.01$	$\mathbb{P}(\bar{B}\|\bar{Z}) = 0.99$

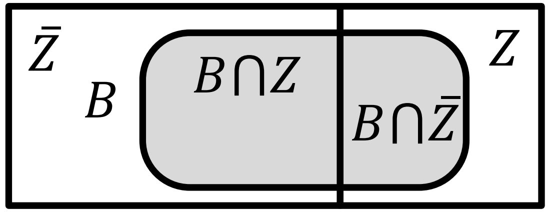
We don't want $\mathbb{P}(B|Z) = 0.99$, but rather $\mathbb{P}(Z|B)$

$$\mathbb{P}(Z|B) = \frac{\mathbb{P}(B,Z)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|Z)\cdot\mathbb{P}(Z)}{\mathbb{P}(B,Z)+\mathbb{P}(B,\bar{Z})}$$

$\mathbb{P}(B|Z)\cdot\mathbb{P}(Z)$

?

total probability theorem



Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- A patient has a positive test.
  What is the chance they have Zika?

		blood test positive (B) B	 $\bar{B}$
patient has Zika (Z)	Z	$\mathbb{P}(B\|Z) = 0.99$	$\mathbb{P}(\bar{B}\|Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B\|\bar{Z}) = 0.01$	$\mathbb{P}(\bar{B}\|\bar{Z}) = 0.99$

We don't want $\mathbb{P}(B|Z) = 0.99$, but rather $\mathbb{P}(Z|B)$

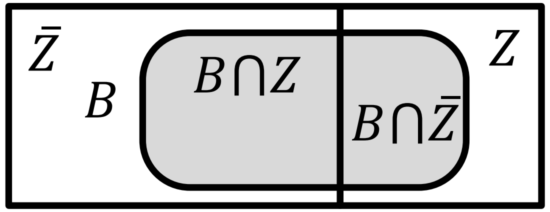
$$\mathbb{P}(Z|B) = \frac{\mathbb{P}(B,Z)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|Z)\cdot\mathbb{P}(Z)}{\mathbb{P}(B,Z)+\mathbb{P}(B,\bar{Z})} = \frac{0.99\cdot 10^{-5}}{0.99\cdot 10^{-5}+0.01\cdot(1-10^{-5})}$$

$\mathbb{P}(B|Z)\cdot\mathbb{P}(Z)$

$\mathbb{P}(B|\bar{Z})\cdot\mathbb{P}(\bar{Z})$

$$\approx \frac{10^{-5}}{10^{-5}+\underset{10^{-2}}{0.01}} \approx \frac{10^{-5}}{10^{-2}} = 10^{-3} \boxed{= 0.1\%}$$

total probability theorem



Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL.
  Complete the numbers! ?

patient has Zika (Z)	blood test positive (B) B	$\bar{B}$
Z	?	?
$\bar{Z}$	?	?

$$\boxed{\mathbb{P}(Z|B) \approx 0.1\%}$$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = base rate = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL.
  Complete the numbers!

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z			100
	$\bar{Z}$			9,999,900

$$\boxed{\mathbb{P}(Z|B) \approx 0.1\%}$$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL.
  Complete the numbers!

$\mathbb{P}(B|Z) = 99\%$

$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$			9,999,900

$$\boxed{\mathbb{P}(Z|B) \approx 0.1\%}$$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = base rate = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

$\mathbb{P}(B|Z) = 99\%$

$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900

$$\boxed{\mathbb{P}(Z|B) \approx 0.1\%}$$

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

$\mathbb{P}(B|Z) = 99\%$

$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

blood test positive (B)

		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900
		100,098	9,899,902	

$\mathbb{P}(Z|B) \approx 0.1\%$

probability that:
a positive test is correct

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = base rate = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

$\mathbb{P}(B|Z) = 99\%$

$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900
		100,098	9,899,902	

$$\mathbb{P}(B|Z) \cdot \mathbb{P}(Z) + \mathbb{P}(\bar{B}|\bar{Z}) \cdot \mathbb{P}(\bar{Z}) = 99\%$$

probability that a random test is correct

$\neq$

$$\mathbb{P}(Z|B) \approx 0.1\%$$

probability that: a positive test is correct

Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = base rate = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

$$\mathbb{P}(B|Z) = 99\%$$
$$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$$

blood test positive (B)

patient has Zika (Z)	B	$\bar{B}$	
Z	99	1	100
$\bar{Z}$	99,999	9,899,901	9,999,900
	100,098	9,899,902	

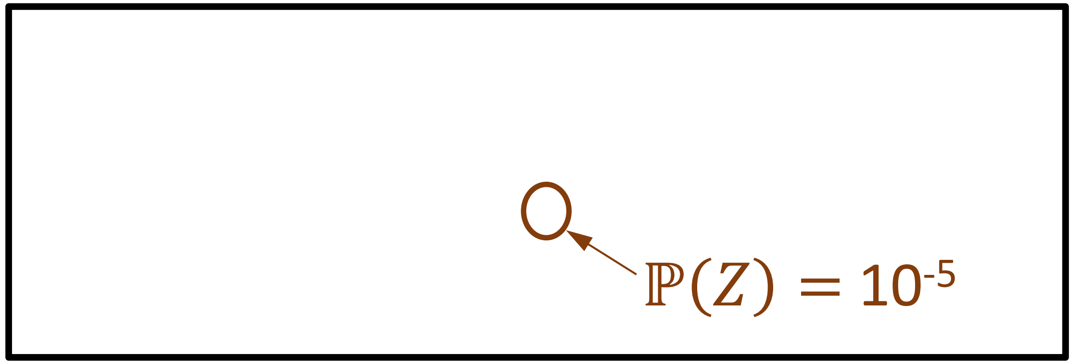
$$\mathbb{P}(B|Z) \cdot \mathbb{P}(Z) + \mathbb{P}(\bar{B}|\bar{Z}) \cdot \mathbb{P}(\bar{Z}) = 99\%$$

probability that a random test is correct

$$\neq$$

$$\mathbb{P}(Z|B) \approx 0.1\%$$

probability that: a positive test is correct



Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

$\mathbb{P}(B|Z) = 99\%$
$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900
		100,098	9,899,902	

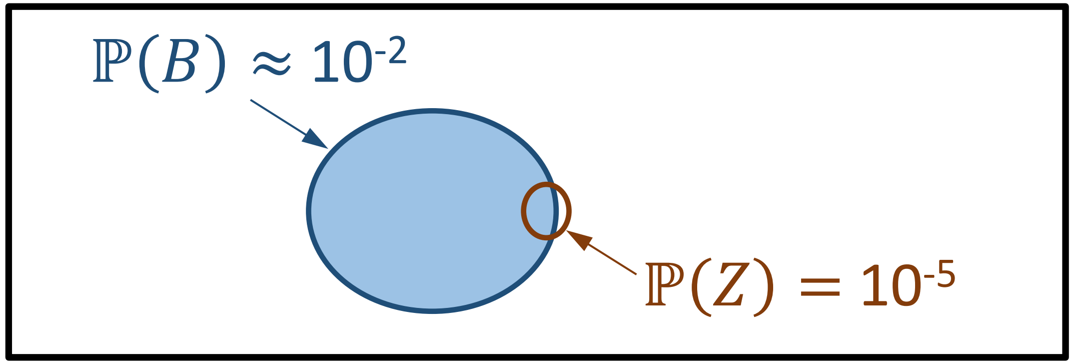
$$\mathbb{P}(B|Z) \cdot \mathbb{P}(Z) + \mathbb{P}(\bar{B}|\bar{Z}) \cdot \mathbb{P}(\bar{Z}) = 99\%$$

probability that a random test is correct

$\neq$

$$\mathbb{P}(Z|B) \approx 0.1\%$$

probability that:
a positive test is correct



Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

Bayes Law ("Base rate fallacy problem")

EXAMPLE 5: Zika in Florida in 2016

- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) $=$ base rate $= \mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL.
  Complete the numbers

$\mathbb{P}(B|Z) = 99\%$
$\mathbb{P}(\bar{B}|\bar{Z}) = 99\%$

		blood test positive (B)		
		B	$\bar{B}$	
patient has Zika (Z)	Z	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900
		100,098	9,899,902	

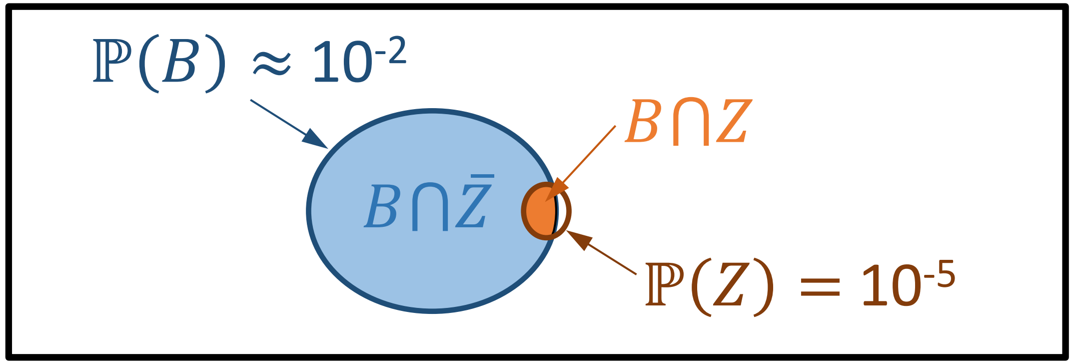
$$\mathbb{P}(B|Z) \cdot \mathbb{P}(Z) + \mathbb{P}(\bar{B}|\bar{Z}) \cdot \mathbb{P}(\bar{Z}) = 99\%$$

probability that a random test is correct

$\neq$

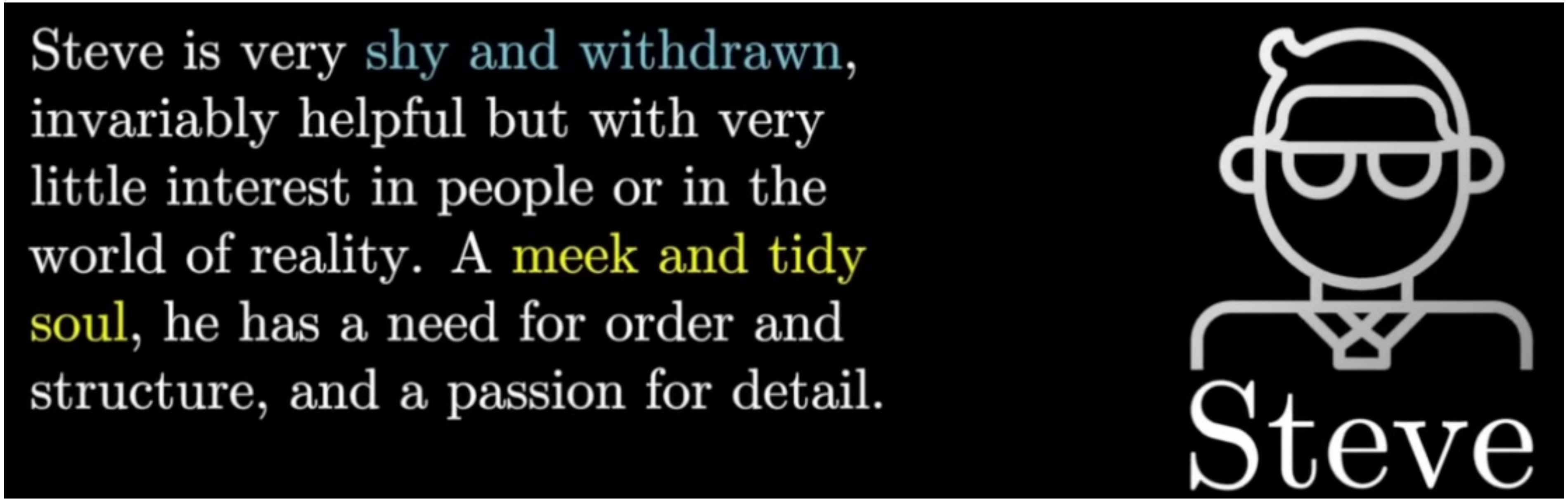
$$\mathbb{P}(Z|B) = 0.1\%$$

probability that:
a positive test is correct



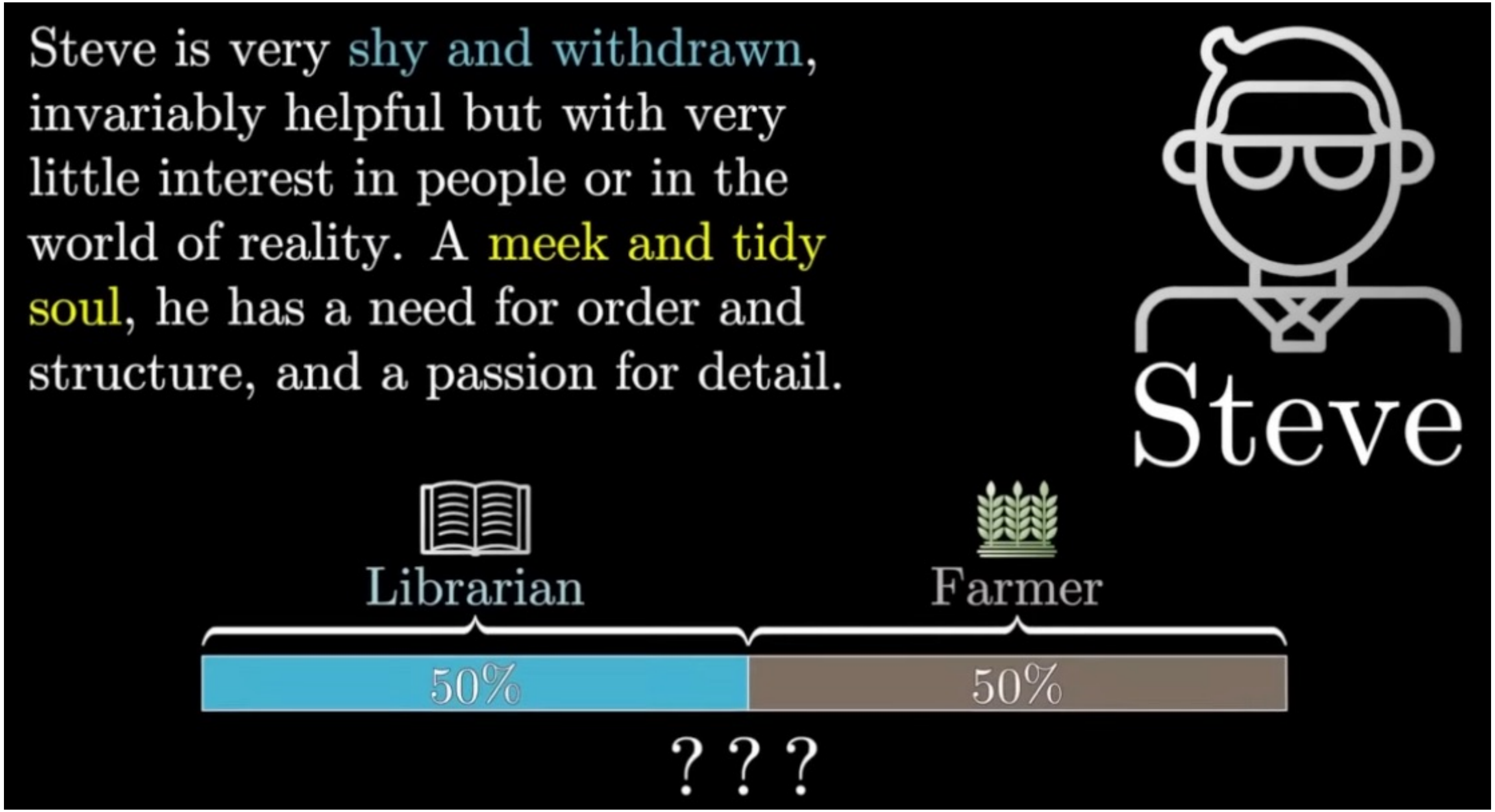
Source: Jay Aslam's course notes from 2024: https://northeastern-datalab.github.io/cs7840/fa24/download/cs7840-L02-3-Probability_Primer.pdf

An unrelated (?) question:



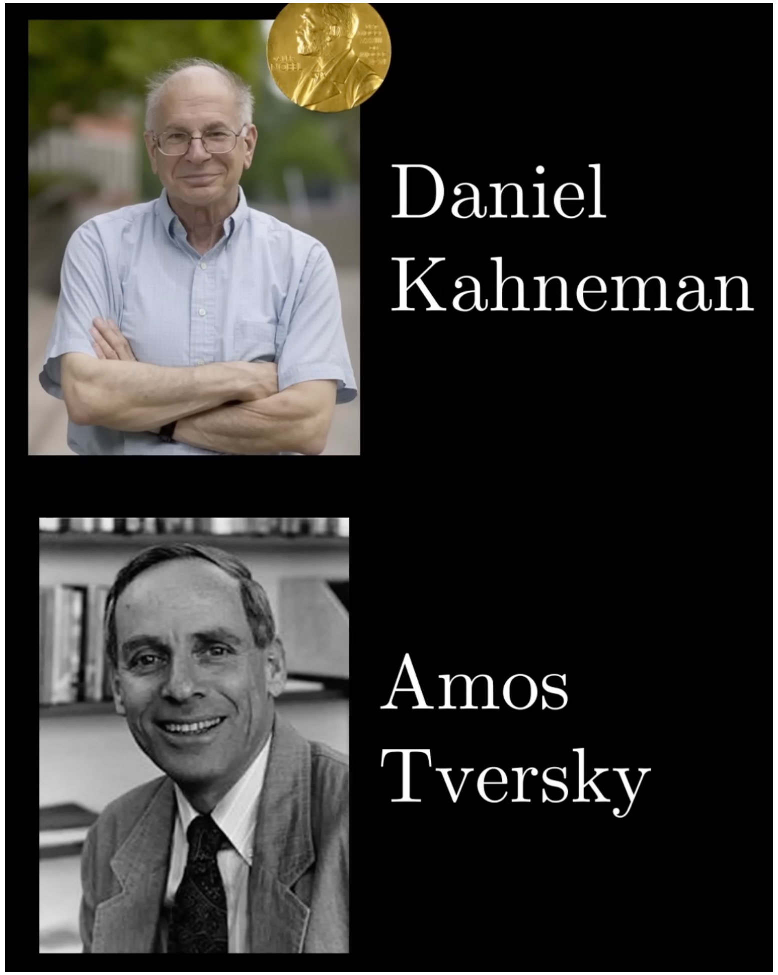
Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

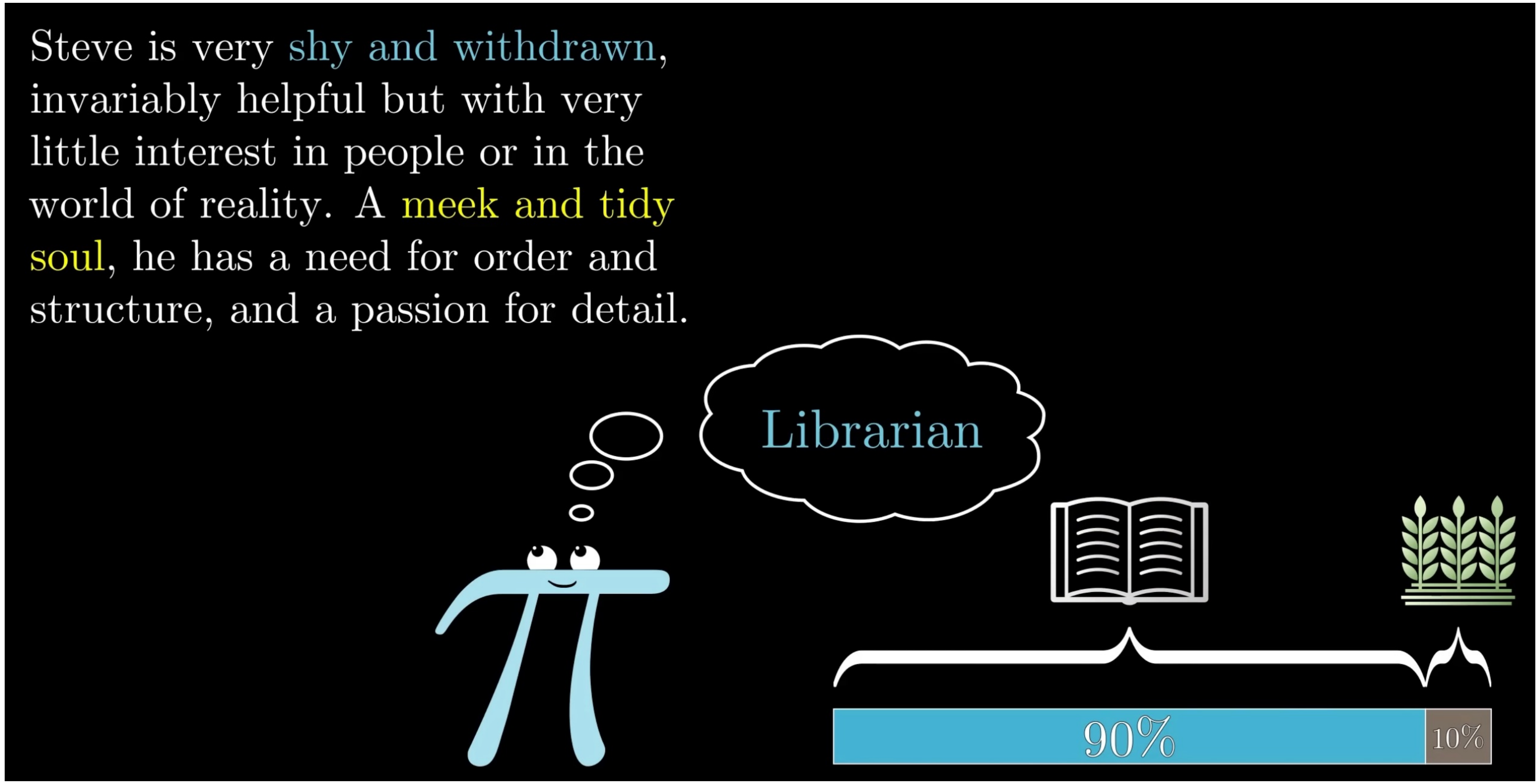
An unrelated (?) question:



Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

Bayes theorem illustrated by 3Blue1Brown

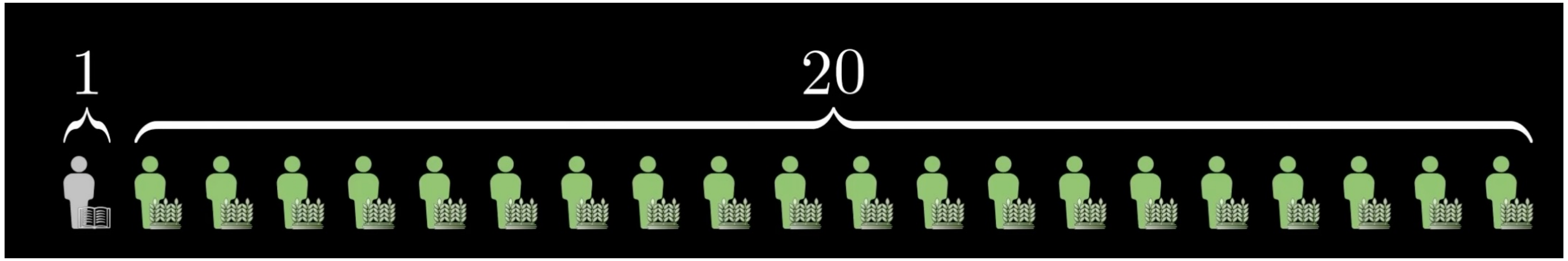




Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

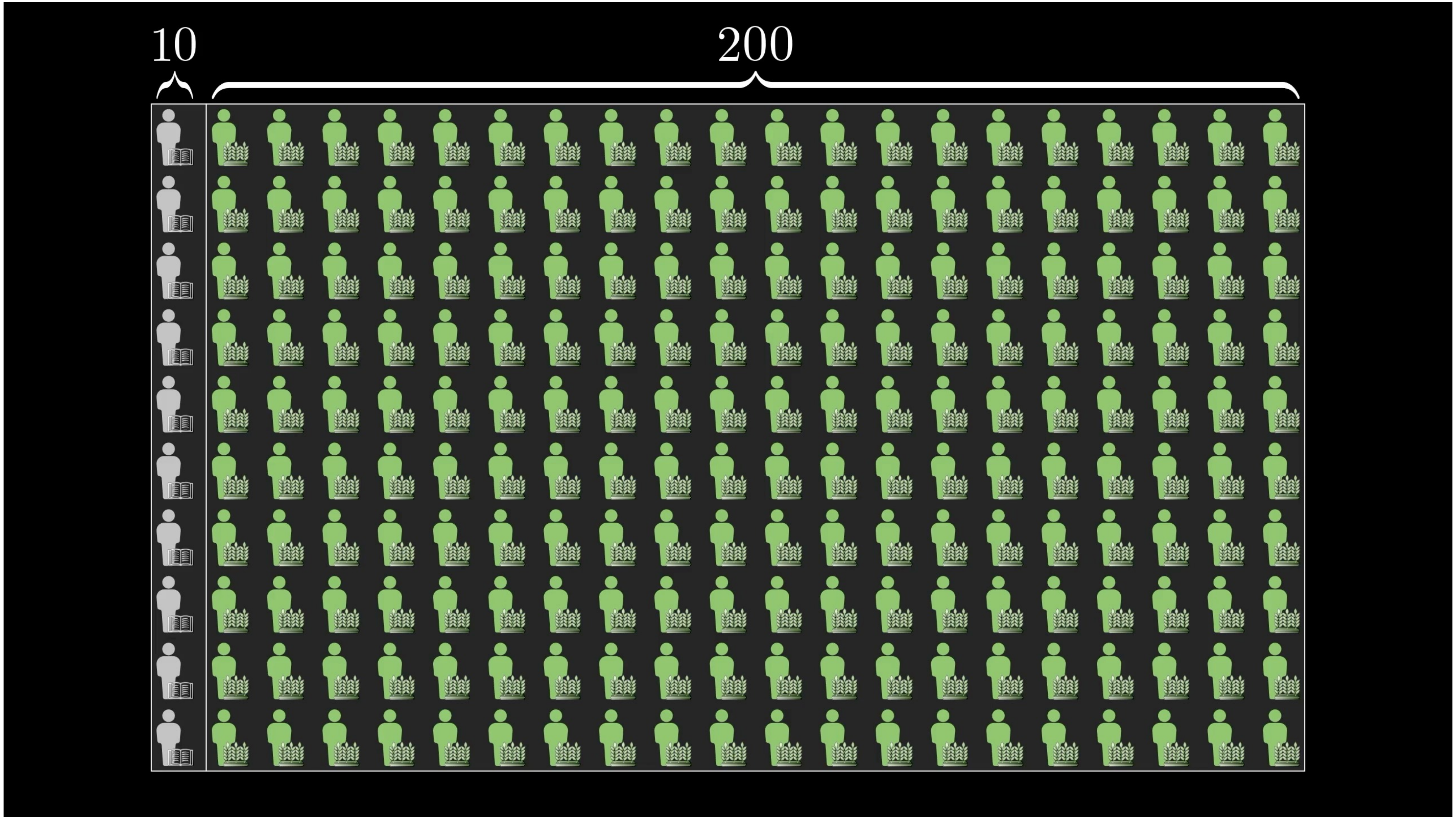
Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes theorem illustrated by 3Blue1Brown



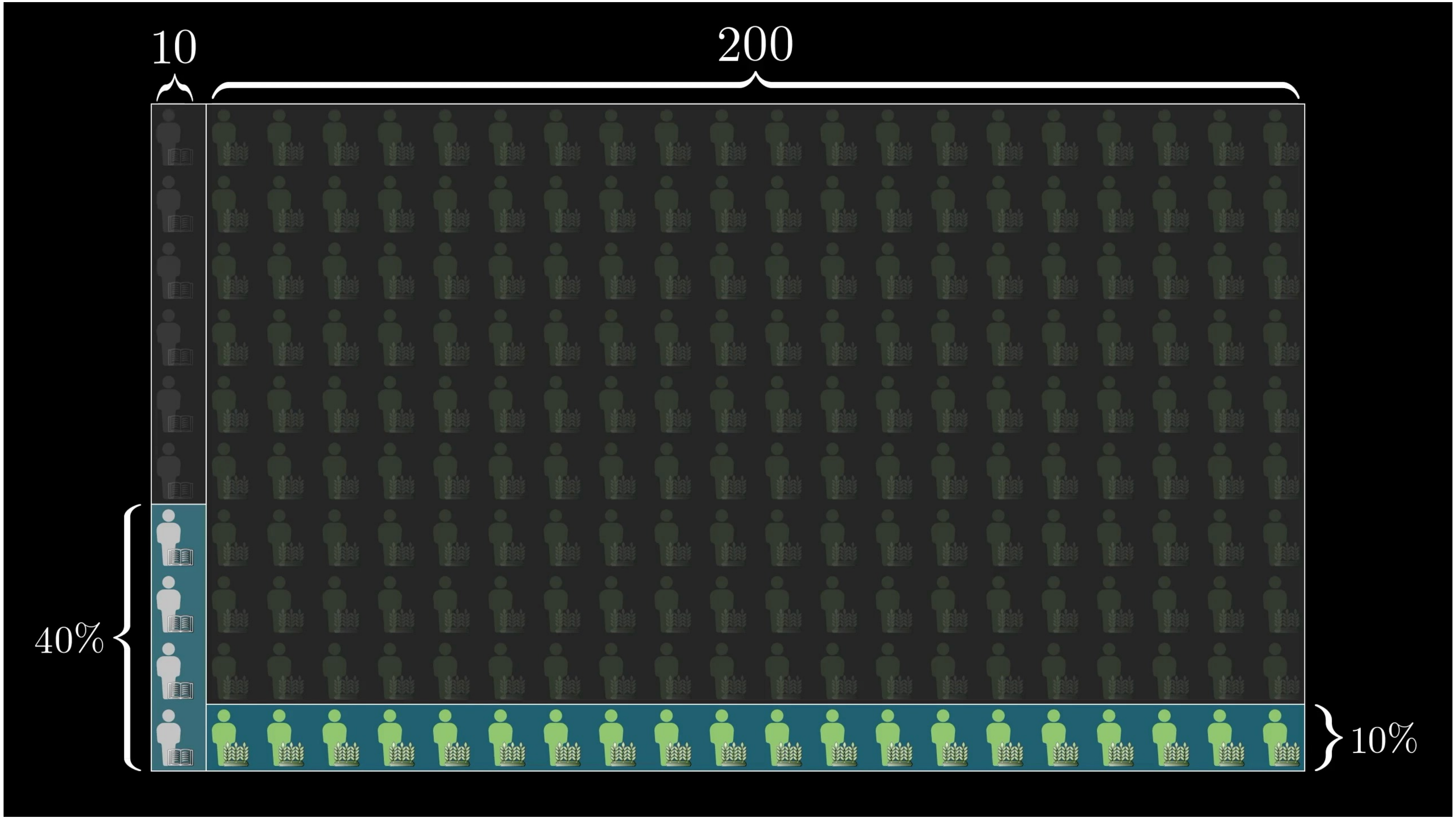
Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM

Bayes theorem illustrated by 3Blue1Brown



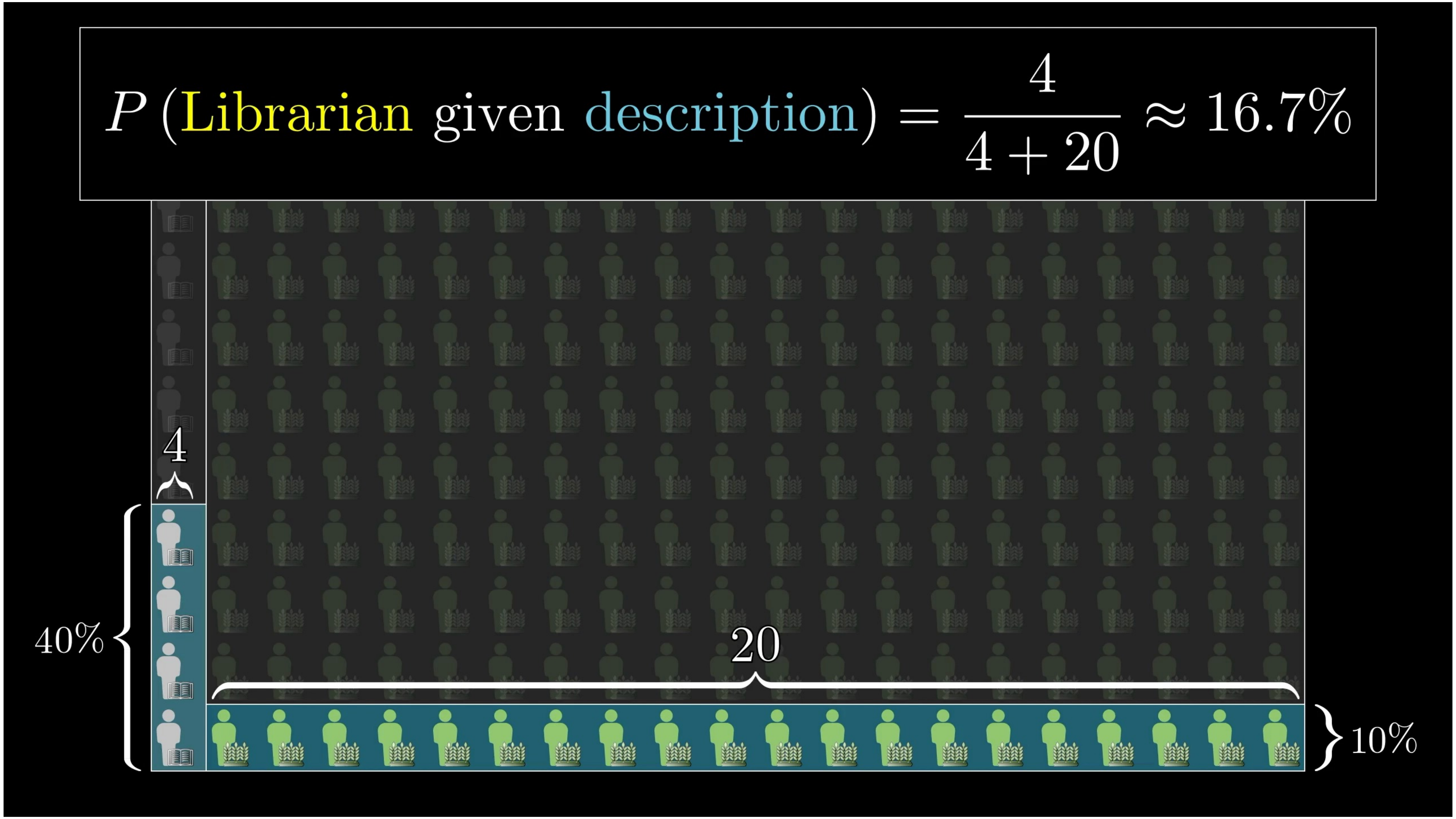
Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes theorem illustrated by 3Blue1Brown



Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

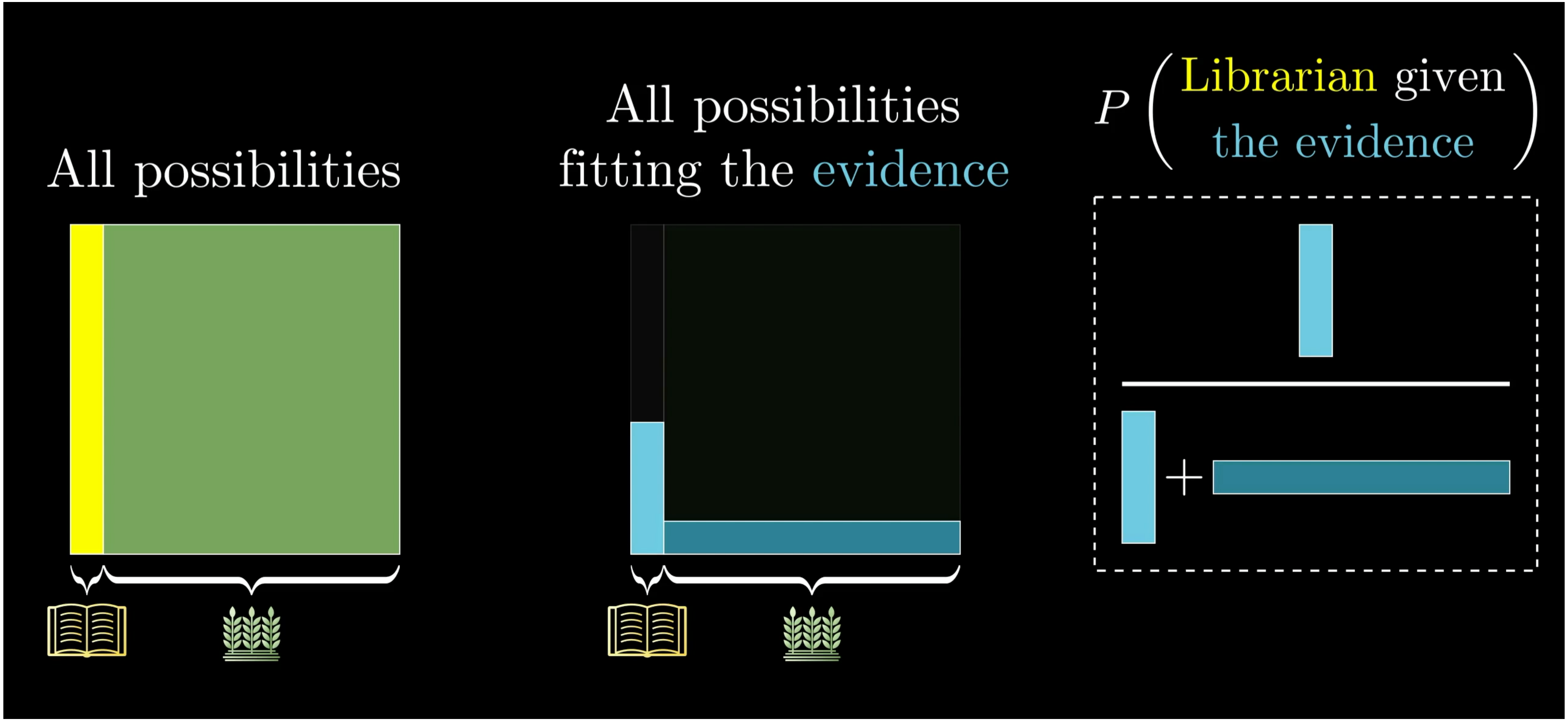
Bayes theorem illustrated by 3Blue1Brown



$$P\,(\text{Librarian given description}) = \frac{4}{4+20} \approx 16.7\%$$

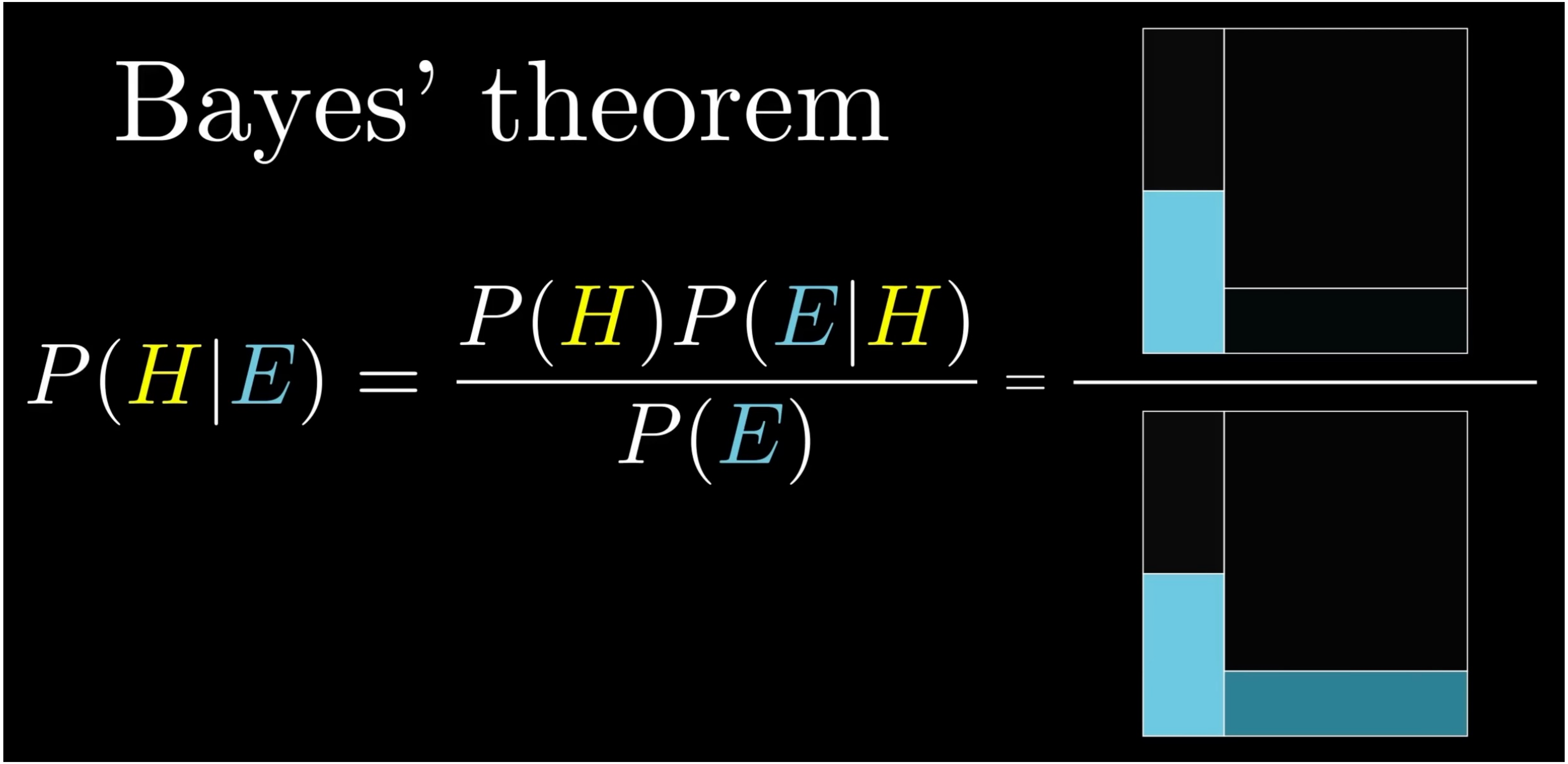
Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes theorem illustrated by 3Blue1Brown



Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes theorem illustrated by 3Blue1Brown



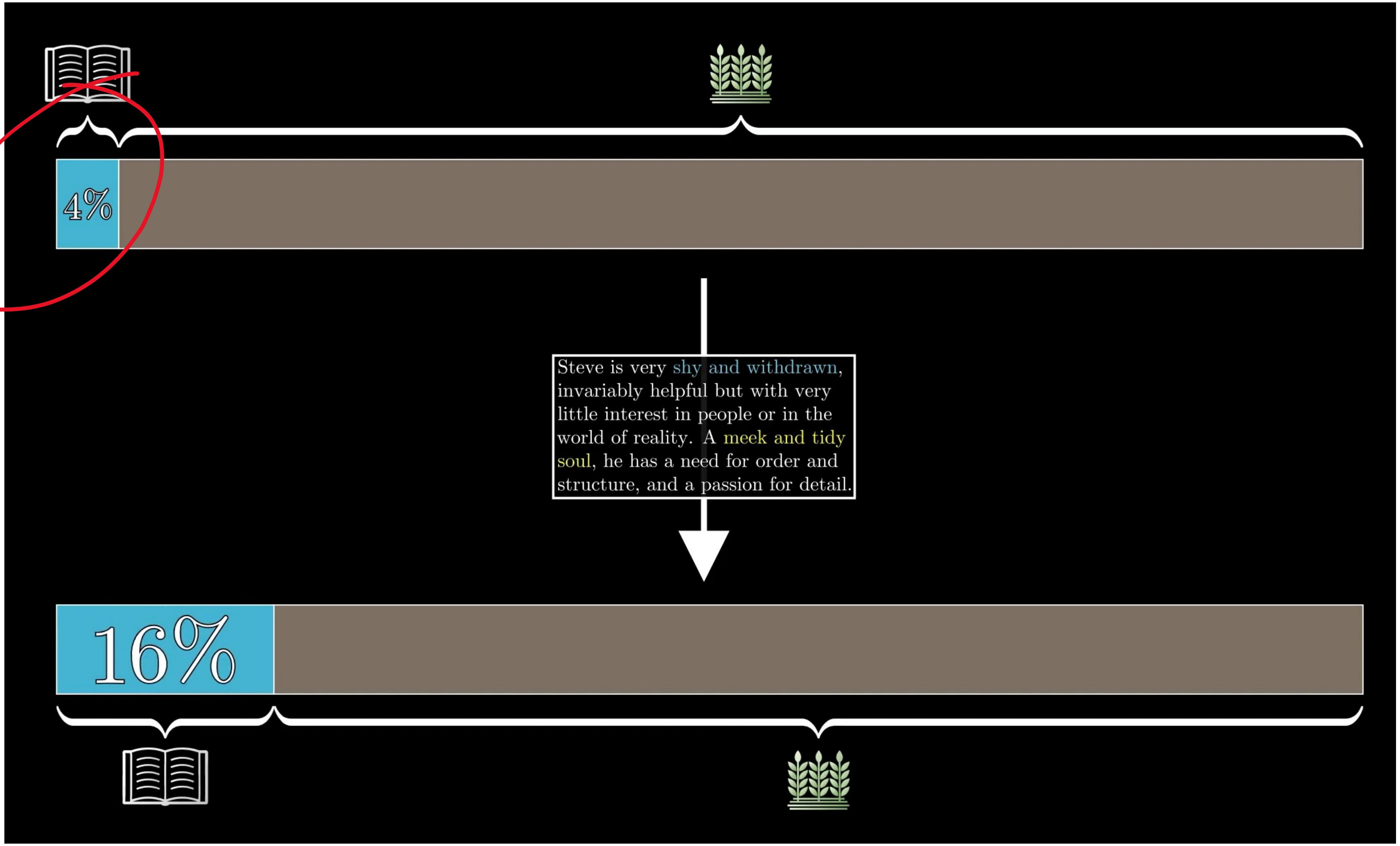
Bayes' theorem

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

Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Bayes theorem illustrated by 3Blue1Brown

New evidence updates prior beliefs! Evidence does not exist in a vaccum



Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

Source: 3Blue1Brown: "Bayes theorem, the geometry of changing beliefs": https://www.3blue1brown.com/lessons/bayes-theorem , https://www.youtube.com/watch?v=HZGCoVF3YvM
Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Updated 9/18/2025

Part 1: Information Theory Basics
L03: Basics of Probability (2/2)
[Random experiment, independence, conditional probability, chain rule, Bayes' theorem, random variables]

Wolfgang Gatterbauer

cs7840 Foundations and Applications of Information Theory (fa25)

https://northeastern-datalab.github.io/cs7840/fa25/

9/15/2025

Pre-class conversations

- Last class recapitulation
- Suggestions on small class projects = "scribes" (e.g. parallelizing compression)
- New class arrivals
- Questions

- Today:
  - The basics of probability theory
  - Start of information theory basics

Random variable, expectation

Random variable: pmf $p_X(x)$

<u>Random variable</u> (RV)
$X\colon\ \Omega \to \mathbb{R}$

Two types:

1. numerical
   e.g. $X(\omega)$ lottery win
2. indicator
   e.g. $X(\omega) = \begin{cases} 1 & \text{if win} \\ 0 & \text{if not} \end{cases}$

X is called a "random variable" (RV) because it depends on the outcome of a <u>random experiment</u>. But the mapping $X\colon\ \Omega \to \mathbb{R}$ is deterministic.

The underlying probability measure $\mathbb{P}\colon \Sigma \to [0, 1]$ induces a pmf p_X (<u>probability mass function</u>) over the range of the RV X: $p_X\colon \mathbb{R} \to [0,1]$

$$p_X(x) = \mathbb{P}(\{X = x\})$$

also written as $p(x) = \mathbb{P}(X = x)$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let X be the sum of the rolls.

Then X is a RV.

What is the pmf ?

Ω		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Random variable: pmf $p_X(x)$

Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Two types:
1. numerical
   e.g. $X(\omega)$ lottery win
2. indicator
   e.g. $X(\omega) = \begin{cases} 1 & \text{if win} \\ 0 & \text{if not} \end{cases}$

X is called a "random variable" (RV) because it depends on the outcome of a random experiment. But the mapping $X\colon\ \Omega \to \mathbb{R}$ is deterministic.

The underlying probability measure $\mathbb{P}\colon \Sigma \to [0,1]$ induces a pmf p_X (probability mass function) over the range of the RV X: $p_X\colon \mathbb{R} \to [0,1]$

$$p_X(x) = \mathbb{P}(\{X = x\})$$

also written as $p(x) = \mathbb{P}(X = x)$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \ldots (1,6), (2,1), (2,2), \ldots (6,6)\}$

Let X be the sum of the rolls.

Then X is a RV.

What is the pmf?

Ω		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X		1	2	3	4	5	6
first die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Random variable: pmf $p_X(x)$

Random variable (RV)
$X\colon \Omega \to \mathbb{R}$

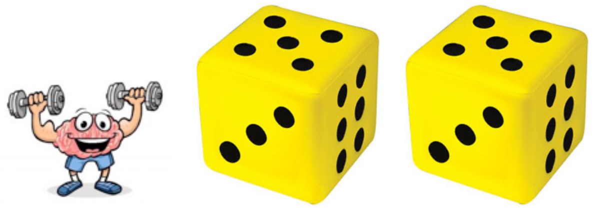
Two types:
1. numerical
   e.g. $X(\omega)$ lottery win
2. indicator
   e.g. $X(\omega) = \begin{cases} 1 & \text{if win} \\ 0 & \text{if not} \end{cases}$

X is called a "random variable" (RV) because it depends on the outcome of a random experiment. But the mapping $X\colon \Omega \to \mathbb{R}$ is deterministic.

The underlying probability measure $\mathbb{P}\colon \Sigma \to [0,1]$ induces a pmf p_X (probability mass function) over the range of the RV X: $p_X\colon \mathbb{R} \to [0,1]$

$$p_X(x) = \mathbb{P}(\{X = x\})$$

also written as $p(x) = \mathbb{P}(X = x)$



EXAMPLE 3 (CONT.):

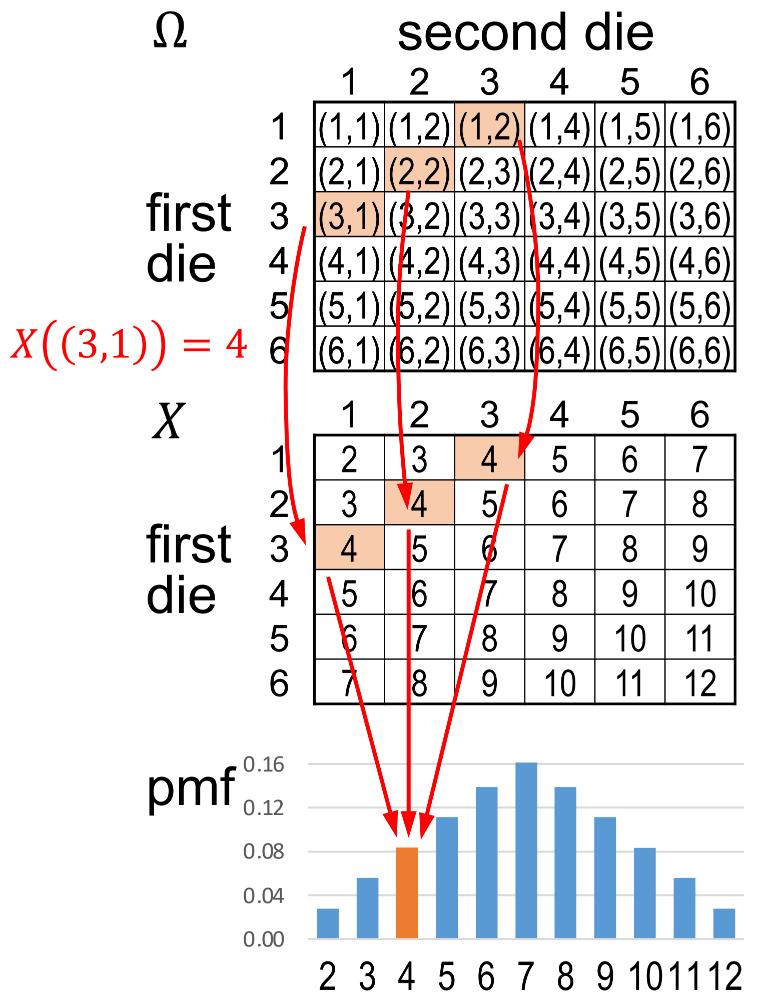
roll two fair dice with 6 sides

$$\Omega = \{(1,1), (1,2), (1,3) \ldots (1,6), (2,1), (2,2), \ldots (6,6)\}$$

Let X be the sum of the rolls.

Then X is a RV.

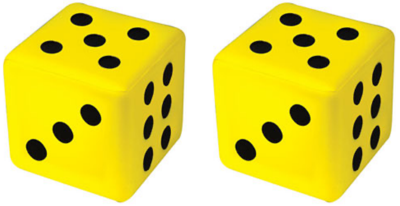
What is the pmf?



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

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

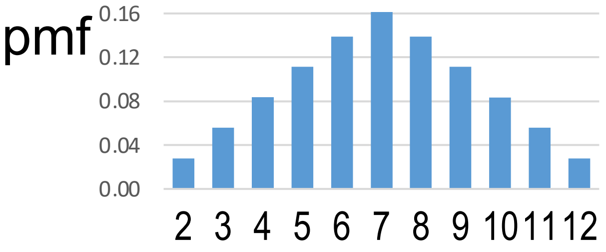
Ω

second die

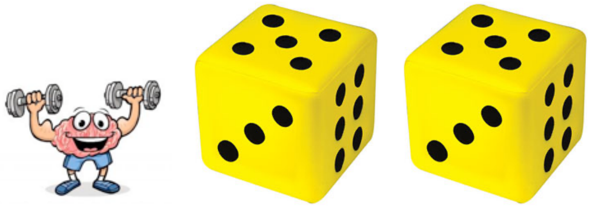
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X

		1	2	3	4	5	6
first die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12



Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

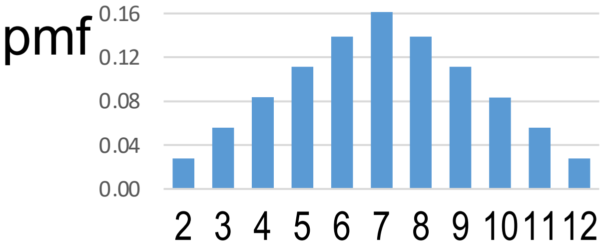
Variant 1:

$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$

?

Ω		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X		1	2	3	4	5	6
first die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12



Random variable: expectation $\mathbb{E}$

Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \ldots (1,6), (2,1), (2,2), \ldots (6,6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

Variant 1:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) \quad \frac{1}{36}$$

$$= \frac{1}{36} \cdot \sum_{\omega \in \Omega} X(\omega)$$

$$= \frac{1}{36} \cdot (\text{"sum of table"})$$

$$= 7$$

Ω second die

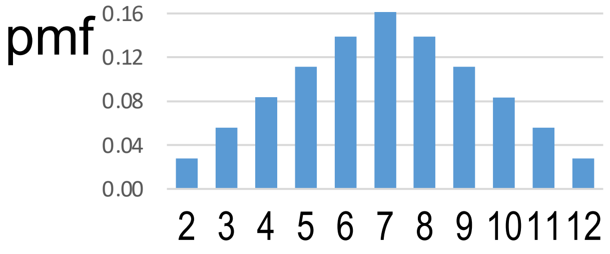
first die

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

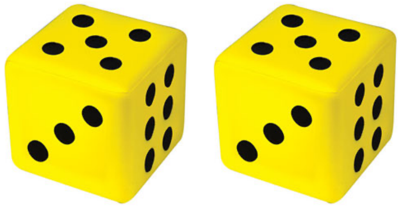
X

first die

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12



Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

Variant 2:

$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}(X = x)$

?

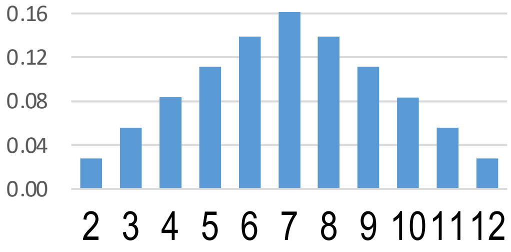
Ω second die

		1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
first die	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X

		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
first die	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

pmf



Random variable: expectation $\mathbb{E}$

<u>Random variable</u> (RV)
$X\colon\ \Omega \to \mathbb{R}$

<u>Expectation</u>: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\cdot\mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x\cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

Variant 2:

$\mathbb{E}[X] = \sum_x x\cdot\mathbb{P}(X = x)$

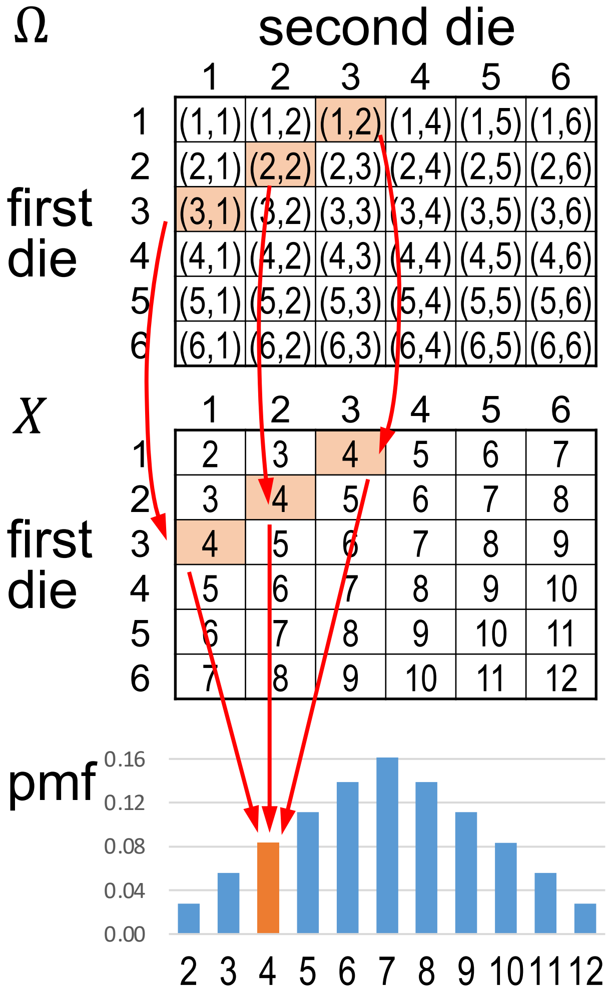
$X = 2 \quad p_X(2) = \frac{1}{36}$

$X = 3 \quad p_X(3) = \frac{2}{36}$

$\ldots \quad \ldots$

$X = 12 \quad p_X(12) = \frac{1}{36}$

$\mathbb{E}[X] = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + \ \ldots + 12\cdot\frac{1}{36} \ = 7$



Ω second die; first die

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

X; first die

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Random variable: expectation $\mathbb{E}$

Random variable (RV)
$X\colon\ \Omega \rightarrow \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \ldots (1,6), (2,1), (2,2), \ldots (6,6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

Variant 3:

RVs for the outcome of the first and second die rolls

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \text{?}$$

Ω second die

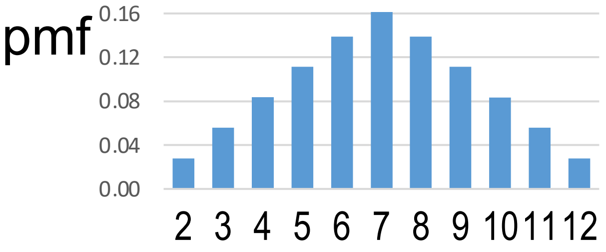
first die

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

X

first die

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12



Random variable: expectation $\mathbb{E}$

Random variable (RV)
$X: \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$

Let X be the sum of the rolls.

What is $\mathbb{E}[X]$?

Variant 3:

RVs for the outcome of the first and second die rolls

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$$

linearity of expectation!

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 3.5$$

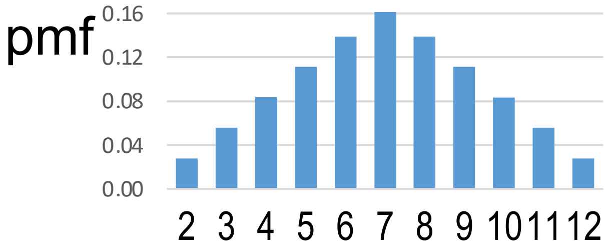
$$\mathbb{E}[X] = 3.5 + 3.5 = 7$$

Ω second die

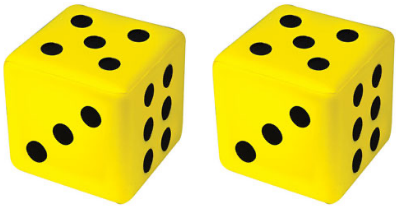
first die		1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X

first die		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12



Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let Y be the product of the rolls.

What is $\mathbb{E}[Y]$?

Ω second die

first die

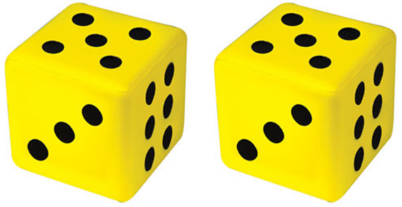
	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Y

first die

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \rightarrow \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let Y be the product of the rolls.

What is $\mathbb{E}[Y]$?

$\mathbb{E}[Y] = \mathbb{E}[X_1 \cdot X_2] = \ \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]$

?

Ω

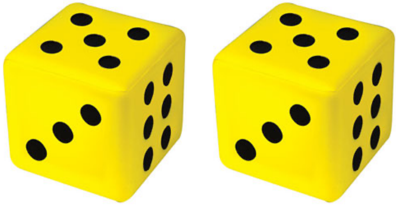
second die

		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Y

		1	2	3	4	5	6
first die	1	1	2	3	4	5	6
	2	2	4	6	8	10	12
	3	3	6	9	12	15	18
	4	4	8	12	16	20	24
	5	5	10	15	20	25	30
	6	6	12	18	24	30	36

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \rightarrow \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \dots (1,6), (2,1), (2,2), \dots (6,6)\}$

Let Y be the product of the rolls.

What is $\mathbb{E}[Y]$?

$$\mathbb{E}[Y] = \mathbb{E}[X_1 \cdot X_2] = \ \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]$$

because the $X_1 \perp X_2$

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 3.5$$

$$\mathbb{E}[X] = 3.5 \cdot 3.5 \quad = 12.25$$

Ω second die

first die

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

Y

first die

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

Random variable: expectation $\mathbb{E}$

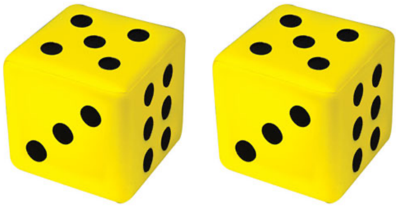
Random variable (RV)
$X\colon\ \Omega \rightarrow \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):



roll two fair dice with 6 sides

$\Omega = \{(1,1), (1,2), (1,3) \ldots (1,6), (2,1), (2,2), \ldots (6,6)\}$

Let Y be the product of the rolls.

Let X be the sum of the rolls.

What is $\mathbb{E}[X+Y]$?

Ω second die

		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

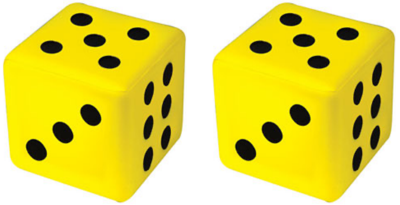
Y

		1	2	3	4	5	6
first die	1	1	2	3	4	5	6
	2	2	4	6	8	10	12
	3	3	6	9	12	15	18
	4	4	8	12	16	20	24
	5	5	10	15	20	25	30
	6	6	12	18	24	30	36

X

		1	2	3	4	5	6
first die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \rightarrow \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\cdot\mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x\cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1,1),(1,2),(1,3)\ \ldots(1,6),(2,1),(2,2),\ldots(6,6)\}$

Let Y be the product of the rolls.

Let X be the sum of the rolls.

What is $\mathbb{E}[X+Y]$?

$$\mathbb{E}[X+Y] = \ \mathbb{E}[X] + \mathbb{E}[Y]$$?

Ω second die

first die

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

Y

first die

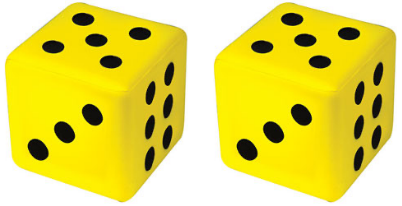
	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

X

first die

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average (in proportion to the probabilities) of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let Y be the product of the rolls.

Let X be the sum of the rolls.

What is $\mathbb{E}[X+Y]$?

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

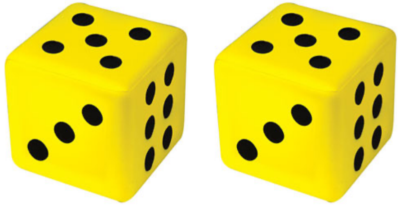
X and Y are clearly dependent ☹

Ω		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,2)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Y		1	2	3	4	5	6
first die	1	1	2	3	4	5	6
	2	2	4	6	8	10	12
	3	3	6	9	12	15	18
	4	4	8	12	16	20	24
	5	5	10	15	20	25	30
	6	6	12	18	24	30	36

X		1	2	3	4	5	6
first die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Random variable: expectation $\mathbb{E}$



Random variable (RV)
$X\colon\ \Omega \to \mathbb{R}$

Expectation: a weighted average
(in proportion to the probabilities)
of the possible values of X

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x)$$

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

$\Omega = \{(1{,}1), (1{,}2), (1{,}3) \ldots (1{,}6), (2{,}1), (2{,}2), \ldots (6{,}6)\}$

Let Y be the product of the rolls.

Let X be the sum of the rolls.

What is $\mathbb{E}[X+Y]$?

$\mathbb{E}[X+Y] = \ \mathbb{E}[X] + \mathbb{E}[Y]$

X and Y are clearly dependent ☹

But linearity of expectation still holds
even if the RVs are dependent ☺

$\mathbb{E}[X] = 7 + 12.25\ = 19.25$

Ω second die

first die

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

Y

first die

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

X

first die

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$\mathbb{E}[X+Y]$ vs. $\mathbb{E}[X\cdot Y]$

$\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$ (linearity of expectation)

holds even if X and Y are not independent

$\mathbb{E}[X\cdot Y]=\mathbb{E}[X]\cdot\mathbb{E}[Y]$, only if $X\perp Y$ (independent)

PROOF

$$\mathbb{E}[X+Y]=\sum_x\sum_y(x+y)\cdot p(x,y)$$

?

$$=\mathbb{E}[X]+\mathbb{E}[Y]$$

PROOF

$$\mathbb{E}[X\cdot Y]=\sum_x\sum_y x\cdot y\cdot p(x,y)$$

?

$$=\mathbb{E}[X]\cdot\mathbb{E}[Y]$$

$\mathbb{E}[X+Y]$ vs. $\mathbb{E}[X \cdot Y]$

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (linearity of expectation)

holds even if X and Y are not independent

PROOF

$$\mathbb{E}[X+Y] = \sum_x \sum_y (x+y) \cdot p(x,y)$$

$$= \sum_x \sum_y x \cdot p(x,y) + \sum_x \sum_y y \cdot p(x,y)$$

$$= \sum_x x \cdot \sum_y p(x,y) + \sum_y y \cdot \sum_x p(x,y)$$

$$= \sum_x x \cdot p(x) + \sum_y y \cdot p(y)$$

$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, only if $X \perp Y$ (independent)

PROOF

$$\mathbb{E}[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot p(x,y)$$

because $X \perp Y$

$$= \sum_x \sum_y x \cdot y \cdot p(x) \cdot p(y)$$

distributive law of multiplication over addition (each cross-term appears exactly once on both sides)

$$= \left(\sum_x x \cdot p(x)\right) \cdot \left(\sum_y x \cdot p(y)\right)$$

$$= \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB} \cdot \overline{CD}$?

?

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?

?

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- What ~~about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$~~?
- No! $\mathbb{E}[xy] = \mathbb{E}[x]\cdot\mathbb{E}[y]$ only holds when the RVs are independent.

  Clearly, $\overline{AB}$ and $\overline{CD}$ are not independent since each digit can only be used once (e.g., if $\overline{AB} = 42$ then we would know that $\overline{CD}$ can only be 13 or 31).

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- ~~What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?~~
- Can we instead get some kind of sum? ?

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- ~~What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?~~
- Can we instead get some kind of sum?

$$\overline{AB}\cdot\overline{CD} = (10\cdot A + B)\cdot(10\cdot C + D) = 100\cdot A\cdot C + 10\cdot A\cdot D + 10\cdot B\cdot C + B\cdot D$$

- Now by linearity of expectation,

?

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- ~~What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?~~
- Can we instead get some kind of sum?

$$\overline{AB}\cdot\overline{CD} = (10\cdot A + B)\cdot(10\cdot C + D) = 100\cdot A\cdot C + 10\cdot A\cdot D + 10\cdot B\cdot C + B\cdot D$$

- Now by linearity of expectation,

$$\mathbb{E}[\overline{AB}\cdot\overline{CD}] = 100\cdot\mathbb{E}[A\cdot C] + 10\cdot\mathbb{E}[A\cdot D] + 10\cdot\mathbb{E}[B\cdot C] + \mathbb{E}[B\cdot D] =$$

?

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- ~~What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?~~
- Can we instead get some kind of sum?

$$\overline{AB}\cdot\overline{CD} = (10\cdot A + B)\cdot(10\cdot C + D) = 100\cdot A\cdot C + 10\cdot A\cdot D + 10\cdot B\cdot C + B\cdot D$$

- Now by linearity of expectation,

$$\mathbb{E}[\overline{AB}\cdot\overline{CD}] = 100\cdot\mathbb{E}[A\cdot C] + 10\cdot\mathbb{E}[A\cdot D] + 10\cdot\mathbb{E}[B\cdot C] + \mathbb{E}[B\cdot D] = 121\cdot\mathbb{E}[A\cdot C] = ?$$

The expected value of all of these products are the same since there is symmetry among A, B, C, D.

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Linearity of Expectation

EXAMPLE 6 (CONT.):

- The digits 1,2,3, and 4 are randomly arranged to form two two-digit numbers, $\overline{AB}$ and $\overline{CD}$. For example, we could have $\overline{AB} = 42$ and $\overline{CD} = 13$.
- What is the expected value of $\overline{AB}\cdot\overline{CD}$?
- ~~What about $\mathbb{E}[\overline{AB}]\cdot\mathbb{E}[\overline{CD}]$?~~
- Can we instead get some kind of sum?

$$\overline{AB}\cdot\overline{CD} = (10\cdot A + B)\cdot(10\cdot C + D) = 100\cdot A\cdot C + 10\cdot A\cdot D + 10\cdot B\cdot C + B\cdot D$$

- Now by linearity of expectation,

$$\mathbb{E}[\overline{AB}\cdot\overline{CD}] = 100\cdot\mathbb{E}[A\cdot C] + 10\cdot\mathbb{E}[A\cdot D] + 10\cdot\mathbb{E}[B\cdot C] + \mathbb{E}[B\cdot D] = 121\cdot\mathbb{E}[A\cdot C] = \frac{4235}{6} = 705.8\dot{3}$$

The expected value of all of these products are the same since there is symmetry among A, B, C, D.

$$\mathbb{E}[A\cdot C] = \frac{1\cdot 2+1\cdot 3+1\cdot 4+2\cdot 3+2\cdot 4+3\cdot 4}{6} = \frac{35}{6}$$

Example from: https://brilliant.org/wiki/linearity-of-expectation/

Variance

Measuring variability

EXAMPLE: average size of mice

case 1:	case 2:	case 3:
{4 cm, 5 cm, 6 cm}	{9 cm, 10 cm, 11 cm}	{8 cm, 10 cm, 12 cm}

What is a reasonable measure of centrality?

Measuring variability

EXAMPLE: average size of mice

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	What is a reasonable measure of centrality?
$\mathbb{E}[X] =$	5	10	10	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$ mean = average = "expected value"

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$
$\mathbb{E}[X] =$	5	10	10	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y]=\sum_y y\cdot p_Y(y)$
$\mathbb{E}[X]=$	5	10	10	$\mathbb{E}[X]=\sum_x x\cdot p_X(x)$
$\mathbb{E}[Y_1]=$	$-1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=0$	$-1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=0$	$-2\cdot\frac{1}{3}+0\cdot\frac{1}{3}+2\cdot\frac{1}{3}$ $=0$ ☹ cancels out	$Y_1=X-\mathbb{E}[X]$ How can we fix that ?

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$
$\mathbb{E}[X] =$	5	10	10	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹ cancels out	$Y_1 = X - \mathbb{E}[X]$
$\mathbb{E}[Y_2] =$	$\lvert -1 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 1 \rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$\lvert -1 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 1 \rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$\lvert -2 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 2 \rvert \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$ ☺ makes more sense	$Y_2 = \lvert X - \mathbb{E}[X] \rvert$ absolute deviation

Anything else ?

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$
$\mathbb{E}[X] =$	5	10	10	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹ cancels out	$Y_1 = X - \mathbb{E}[X]$
$\mathbb{E}[Y_2] =$	$\lvert -1\rvert \cdot \frac{1}{3} + \lvert 0\rvert \cdot \frac{1}{3} + \lvert 1\rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$\lvert -1\rvert \cdot \frac{1}{3} + \lvert 0\rvert \cdot \frac{1}{3} + \lvert 1\rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$\lvert -2\rvert \cdot \frac{1}{3} + \lvert 0\rvert \cdot \frac{1}{3} + \lvert 2\rvert \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$ ☺ makes more sense	$Y_2 = \lvert X - \mathbb{E}[X]\rvert$ absolute deviation
$\mathbb{E}[Y_3] =$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$2^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3}$ $4 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3}$ $= \frac{8}{3}$	$Y_3 = (X - \mathbb{E}[X])^2$ squared error

What are now the "units" of variability ?

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$
$\mathbb{E}[X] =$	5	10	10	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹ cancels out	$Y_1 = X - \mathbb{E}[X]$
$\mathbb{E}[Y_2] =$	$\lvert -1 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 1 \rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}\ \text{cm}$	$\lvert -1 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 1 \rvert \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}\ \text{cm}$	$\lvert -2 \rvert \cdot \frac{1}{3} + \lvert 0 \rvert \cdot \frac{1}{3} + \lvert 2 \rvert \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}\ \text{cm}$ ☺ makes more sense	$Y_2 = \lvert X - \mathbb{E}[X] \rvert$ absolute deviation
$\mathbb{E}[Y_3] =$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}\ \text{cm}^2$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}\ \text{cm}^2$	$2^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3}$ $4 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3}$ $= \frac{8}{3}\ \text{cm}^2$ ☹ ???	$Y_3 = (X - \mathbb{E}[X])^2$ squared error

cm squared is strange. What can we do?

?

Measuring variability

EXAMPLE: average size of mice

What are possible ways to measure expected "variability" Δ around the mean for each point?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}	$\mathbb{E}[Y]=\sum_y y\cdot p_Y(y)$
$\mathbb{E}[X]=$	5	10	10	$\mathbb{E}[X]=\sum_x x\cdot p_X(x)$
$\mathbb{E}[Y_1]=$	$-1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=0$	$-1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=0$	$-2\cdot\frac{1}{3}+0\cdot\frac{1}{3}+2\cdot\frac{1}{3}$ $=0$ ☹ cancels out	$Y_1=X-\mathbb{E}[X]$
$\mathbb{E}[Y_2]=$	$\lvert-1\rvert\cdot\frac{1}{3}+\lvert0\rvert\cdot\frac{1}{3}+\lvert1\rvert\cdot\frac{1}{3}$ $=1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=\frac{2}{3}$ cm	$\lvert-1\rvert\cdot\frac{1}{3}+\lvert0\rvert\cdot\frac{1}{3}+\lvert1\rvert\cdot\frac{1}{3}$ $=1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=\frac{2}{3}$ cm	$\lvert-2\rvert\cdot\frac{1}{3}+\lvert0\rvert\cdot\frac{1}{3}+\lvert2\rvert\cdot\frac{1}{3}$ $=2\cdot\frac{1}{3}+0\cdot\frac{1}{3}+2\cdot\frac{1}{3}$ $=\frac{4}{3}$ cm ☺ makes more sense	$Y_2=\lvert X-\mathbb{E}[X]\rvert$ absolute deviation
$\mathbb{E}[Y_3]=$	$1^2\cdot\frac{1}{3}+0^2\cdot\frac{1}{3}+1^2\cdot\frac{1}{3}$ $1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=\frac{2}{3}$ cm^2	$1^2\cdot\frac{1}{3}+0^2\cdot\frac{1}{3}+1^2\cdot\frac{1}{3}$ $1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$ $=\frac{2}{3}$ cm^2	$2^2\cdot\frac{1}{3}+0^2\cdot\frac{1}{3}+2^2\cdot\frac{1}{3}$ $4\cdot\frac{1}{3}+0\cdot\frac{1}{3}+4\cdot\frac{1}{3}$ $=\frac{8}{3}$ cm^2 ☹ ???	$Y_3=(X-\mathbb{E}[X])^2$ squared error
Just take the square root:	$=\sqrt{\frac{2}{3}}$ cm	$=\sqrt{\frac{2}{3}}$ cm	$=2\sqrt{\frac{2}{3}}$ cm ???	

Looks prety complicated. So why is everyone so excited about squared error instead of absolute error?

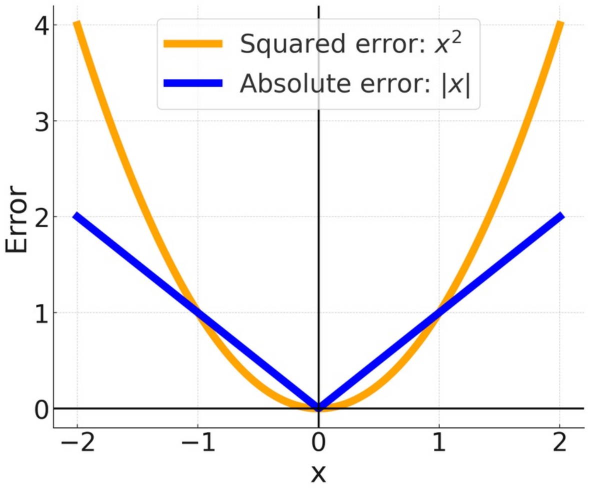
?

Why variance, and not absolute deviation?

But what is the most substantial reason?

?

- Geometric explanation in Euclidean space (distance = square root of squared components): Projection onto a subspace. Variance is literally the "average squared distance" from the mean.
- Variance plays an important role in the Central Limit Theorem. The variance is a natural parameter for the Normal Distribution.
- Algebraic convenience: $f(x) = x^2$ is differentiable, but $g(x) = |x|$ is not.
- Outliers have a bigger influence



$\Delta_2 = |X - \mathbb{E}[X]|$

absolute deviation

$\Delta_3 = (X - \mathbb{E}[X])^2$

squared error

Why variance, and not absolute deviation?

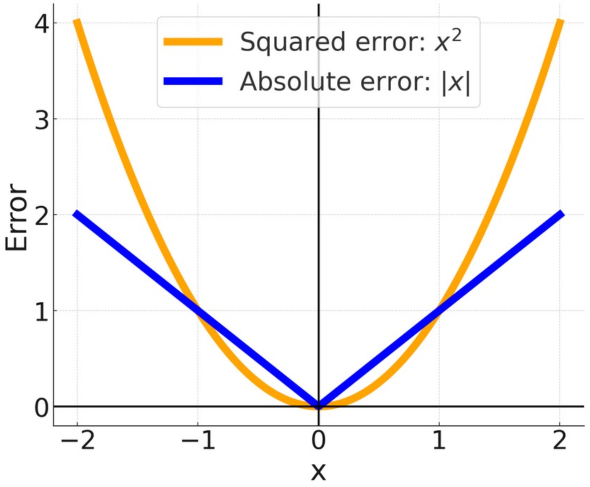
- Statistical properties: The mean minimizes the sum of squared errors, while the median minimizes the sum of absolute errors. (Maximum likelihood estimation leads to minimizing squared errors)

- Geometric explanation in Euclidean space (distance = square root of squared components): Projection onto a subspace. Variance is literally the "average squared distance" from the mean.

- Variance plays an important role in the Central Limit Theorem. The variance is a natural parameter for the Normal Distribution.

- Algebraic convenience: $f(x) = x^2$ is differentiable, but $g(x) = |x|$ is not.

- Outliers have a bigger influence



$\Delta_2 = |X - \mathbb{E}[X]|$

absolute deviation

$\Delta_3 = (X - \mathbb{E}[X])^2$

squared error

Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Mean (minimizes variance), Median (minimizes absolute error)

EXAMPLE (MEAN VS. MEDIAN):

$\text{data} = \{1, 2, 3, 6, 8\}$

$\text{mean} =$?

$\text{median} =$?

Mean (minimizes variance), Median (minimizes absolute error)

EXAMPLE (MEAN VS. MEDIAN):

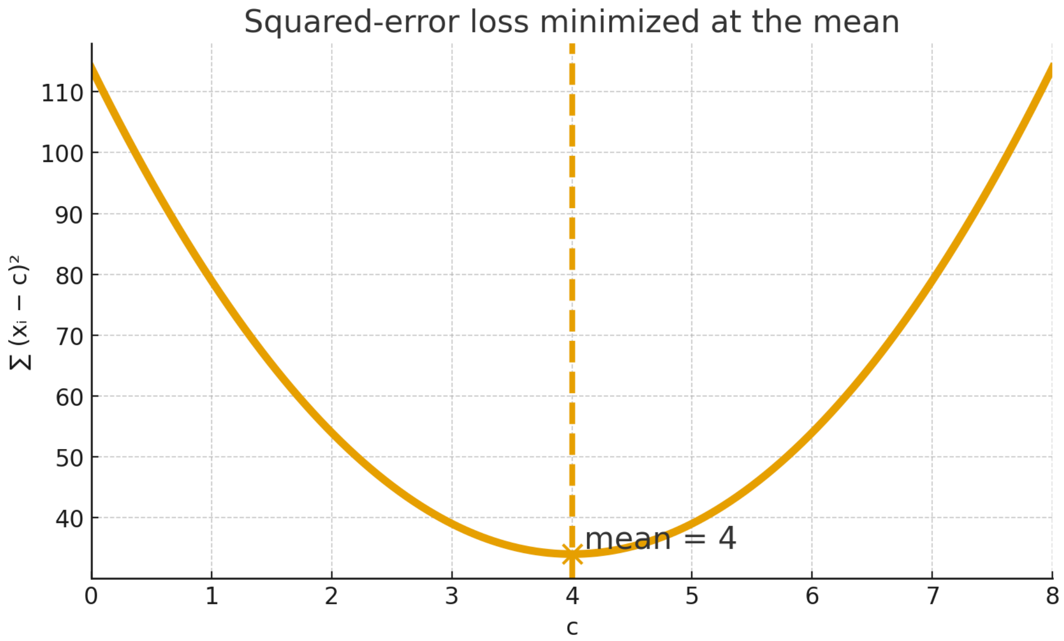
$$\text{data} = \{1, 2, 3, 6, 8\}$$

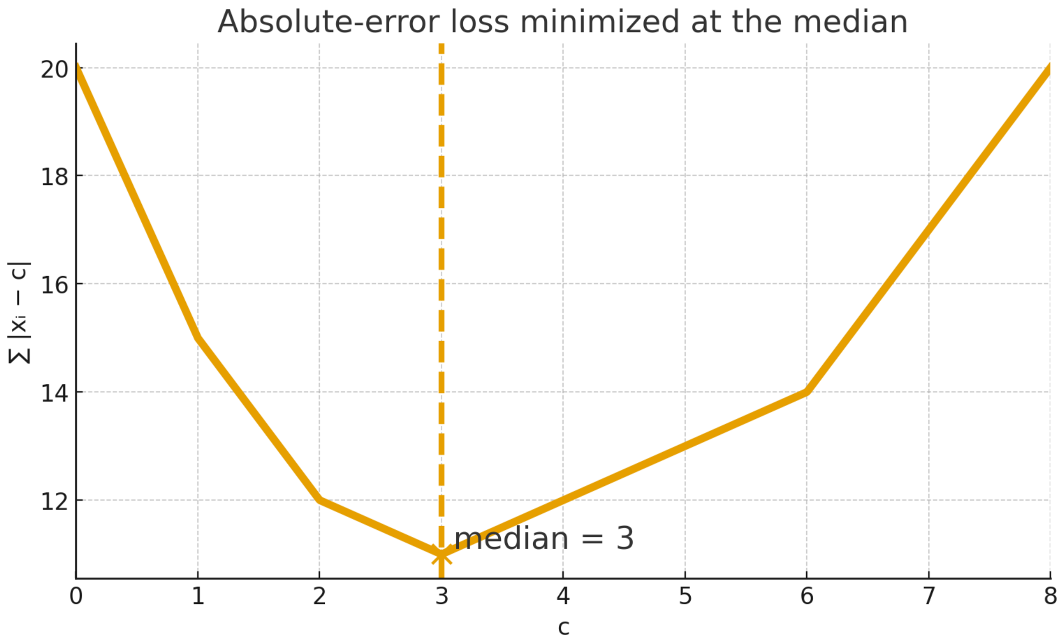
$$\text{mean} = 4$$

$$\text{median} = 3$$

$$\text{mean} = \min_c \sum_i (x_i - c)^2$$

$$\text{median} = \min_c \sum_i |x_i - c|$$





Gatterbauer. Foundations and Applications of Information Theory: https://northeastern-datalab.github.io/cs7840/

Mean (minimizes variance), Median (minimizes absolute error)

EXAMPLE (MEAN VS. MEDIAN):

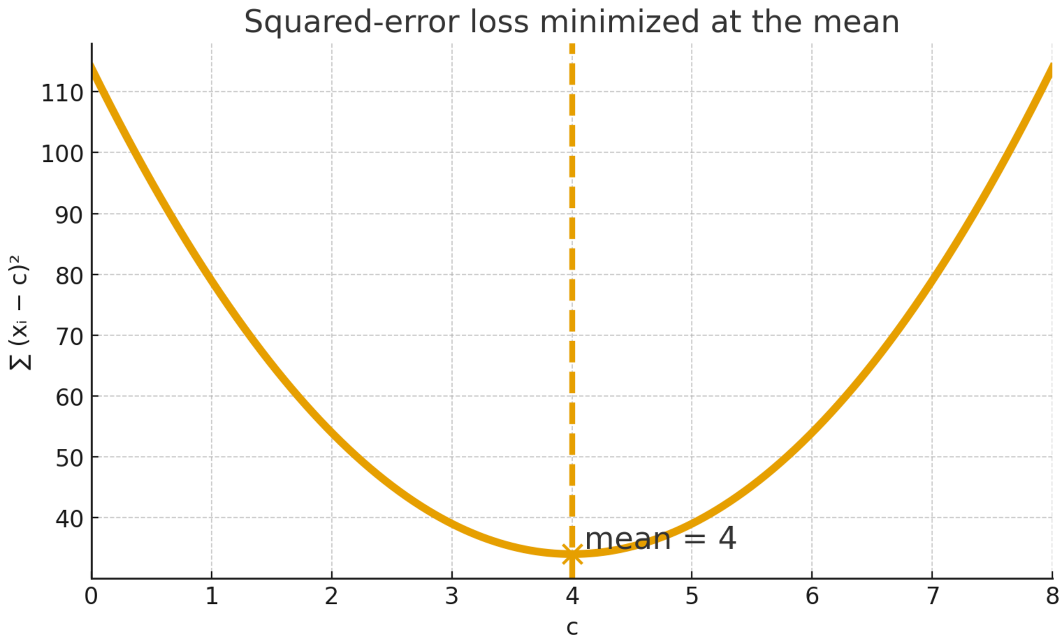
$$\text{data} = \{1, 2, 3, 6, 8\}$$

$$\text{mean} = 4$$

$$\text{mean} = \min_c \sum_i (x_i - c)^2$$

$$\text{mean:} \sum_i x_i - c = 0$$

?



Mean (minimizes variance), Median (minimizes absolute error)

EXAMPLE (MEAN VS. MEDIAN):

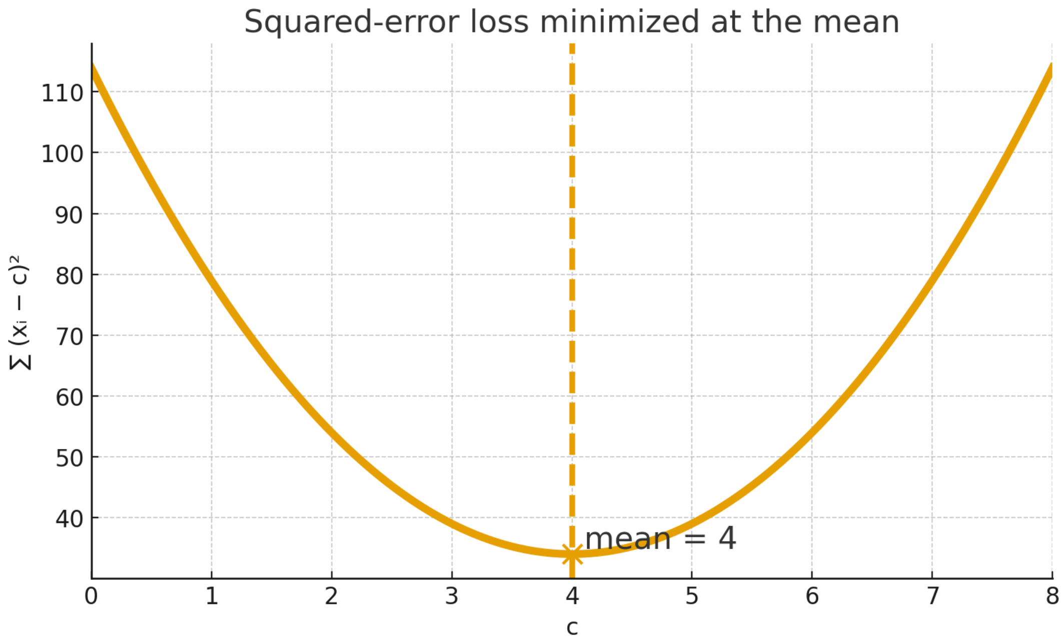
$$\text{data} = \{1, 2, 3, 6, 8\}$$

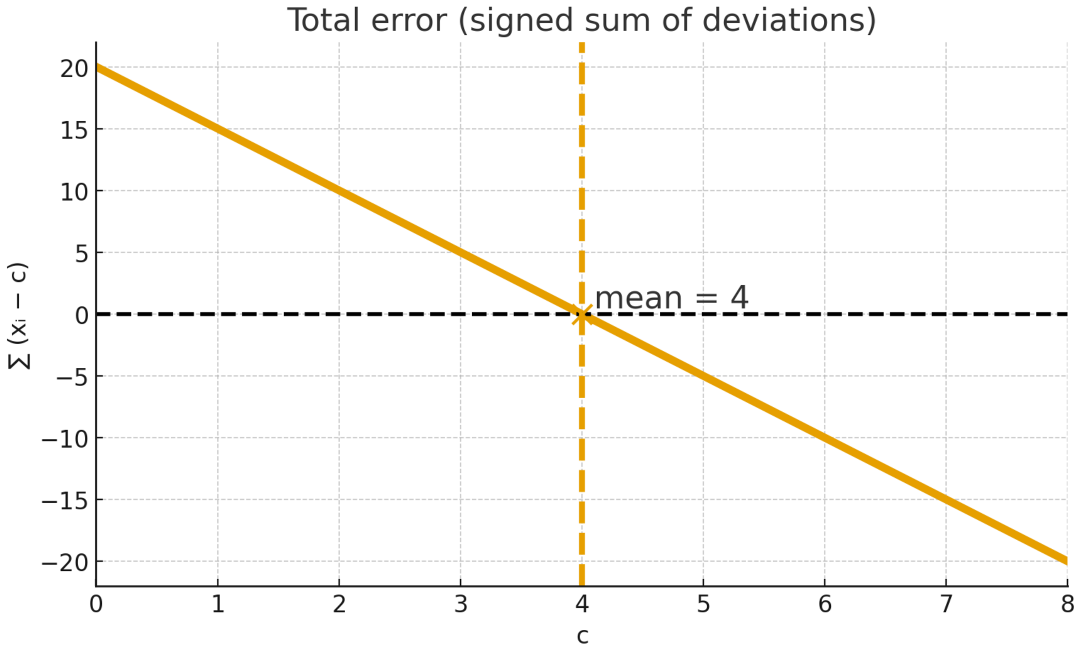
$$\text{mean} = 4$$

$$\text{mean} = \min_c \sum_i (x_i - c)^2$$

This is not an optimization problem, but a constraint problem: the sum of signed errors = 0.

$$\text{mean:} \sum_i x_i - c = 0$$





Alternative formula for variance

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$ variance

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$ standard deviation (back in original units)

ALTERNATIVE FORMULA

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

PROOF

$$\mathbb{E}[(X - \mathbb{E}[X])^2] =$$

?

case 3:
{8 cm, 10 cm, 12 cm}

$$\mathbb{E}[X^2] = \frac{8^2+10^2+12\hat{}2}{3} = \frac{308}{3}$$

$$(\mathbb{E}[X])^2 = 100$$

$$\frac{308}{3} - \frac{300}{3} = 8/3$$

Alternative formula for variance

$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ variance

$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$ standard deviation (back in original units)

ALTERNATIVE FORMULA

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

PROOF

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2 \cdot X \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2]$$

by linearity of expectation

$$= \mathbb{E}[X^2] - \mathbb{E}\big[2 \cdot X \cdot \mathbb{E}[X]\big] + \mathbb{E}[(\mathbb{E}[X])^2]$$

just a constant

$$= \mathbb{E}[X^2] - 2 \cdot \mathbb{E}[X] \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2$$

$$= \mathbb{E}[X^2] - 2 \cdot (\mathbb{E}[X])^2 + (\mathbb{E}[X])^2$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

case 3:
{8 cm, 10 cm, 12 cm}

$$\mathbb{E}[X^2] = \frac{8^2 + 10^2 + 12\text{^}2}{3} = \frac{308}{3}$$

$$(\mathbb{E}[X])^2 = 100$$

$$\frac{308}{3} - \frac{300}{3} = 8/3$$