

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 (fa26)

<https://northeastern-datalab.github.io/cs7840/fa26/>

9/14/2026

Pre-class conversations

- Last class recapitulation
- Organizational matters:
 - Does every one receive Piazza messages?
 - Preliminary posting of slides internally on Piazza ("unedited version")
- 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. The notation follows mostly the one from "*Bertsekas, Tsitsiklis. Introduction to Probability, 2008.* <http://www.athenasc.com/probbook.html> " This is a solid textbook on probability theory and I regularly find myself going back to this book to look up basic concepts. Working 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

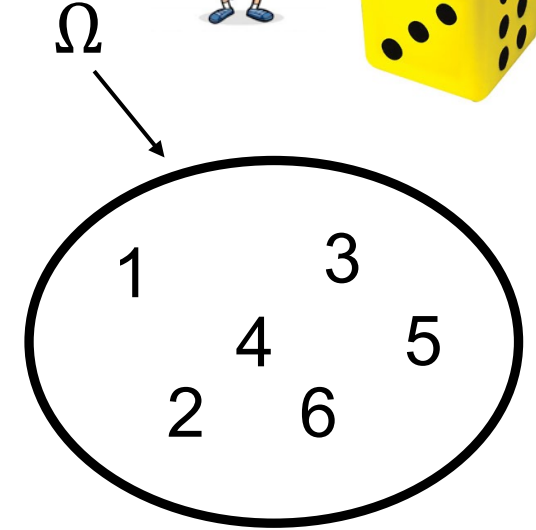
- A random experiment
- Sample space Ω : set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.

EXAMPLE 1:

roll a fair die with 6 sides

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

?



Sample space / outcomes / events / probability



PROBABILITY

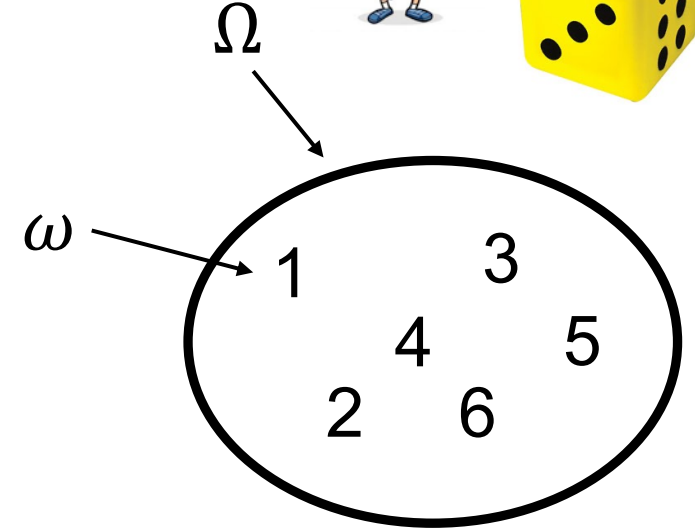
- A random experiment
- Sample space Ω : set of all possible outcomes
- An outcome $\omega \in \Omega$ of the exp.
- Event E : a subset of the sample space $E \subseteq \Omega, E \in \Sigma$

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 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 $E \subseteq \Omega, E \in \Sigma$

EXAMPLE 1:

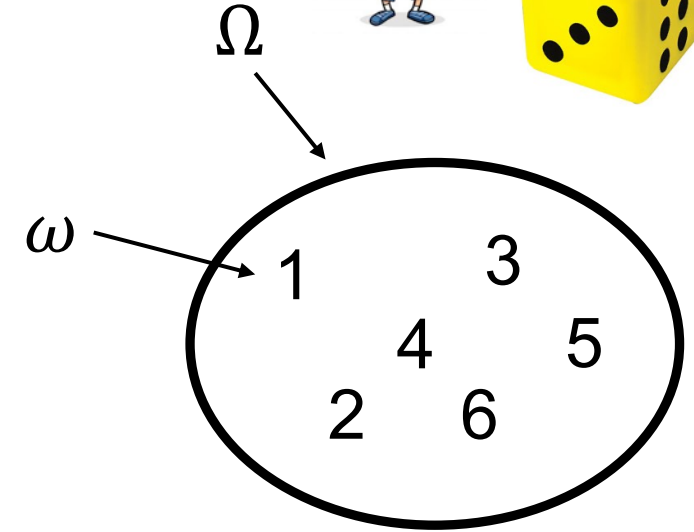
roll a fair die with 6 sides

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:

?

EXAMPLE 1:

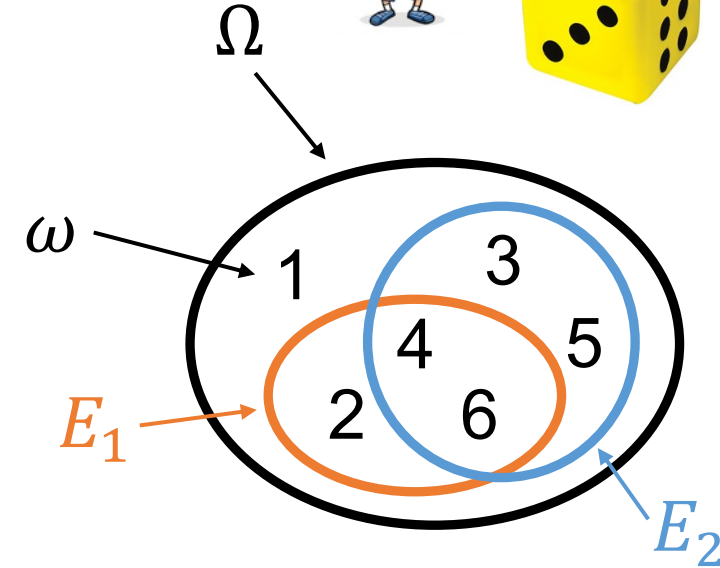
roll a fair die with 6 sides

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 ~~$\mathbb{P}: \Omega \rightarrow [0, 1]$~~

?

EXAMPLE 1:

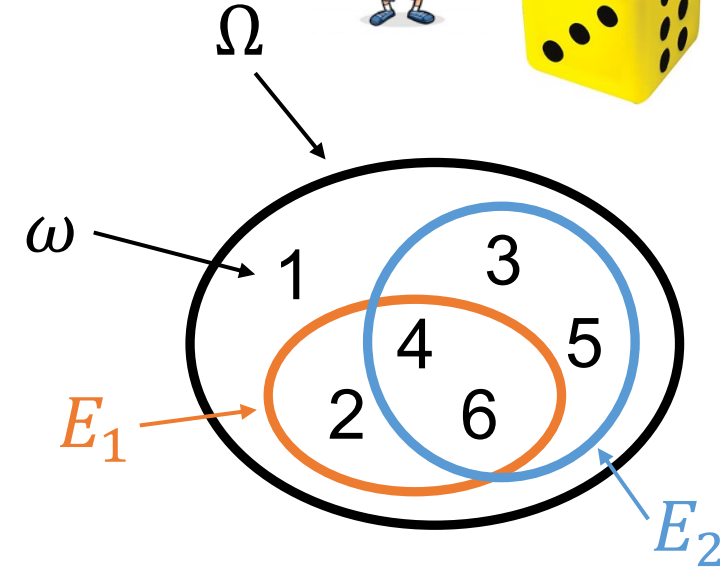
roll a fair die with 6 sides

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

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 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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:

~~$\mathbb{P}: \Omega \rightarrow [0, 1]$~~
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

A probability measure assigns probabilities to events (sets of outcomes), instead of individual outcomes. Thus, its domain is Σ , the collection of events (not Ω , the collection of outcomes: $\Sigma = \mathcal{P}(\Omega)$). Intuitive reason: continuous random outcomes. The probability of each outcome is 0, and we need to be able to reason about intervals.

The set Σ of all events $E \in \Sigma$ is a σ -algebra, basically a subset of 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

Gatterbauer. Foundations and Applications of Information Theory: <https://northeastern-datalab.github.io/cs7840/fa26/>

EXAMPLE 1:

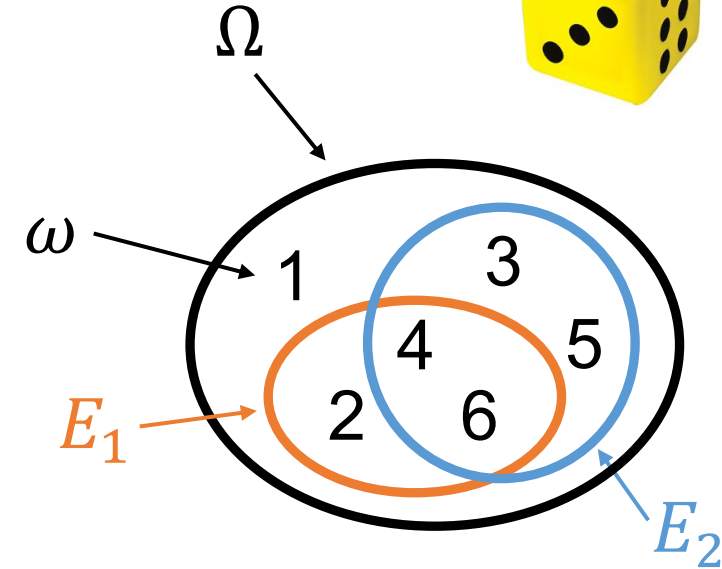
roll a fair die with 6 sides

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

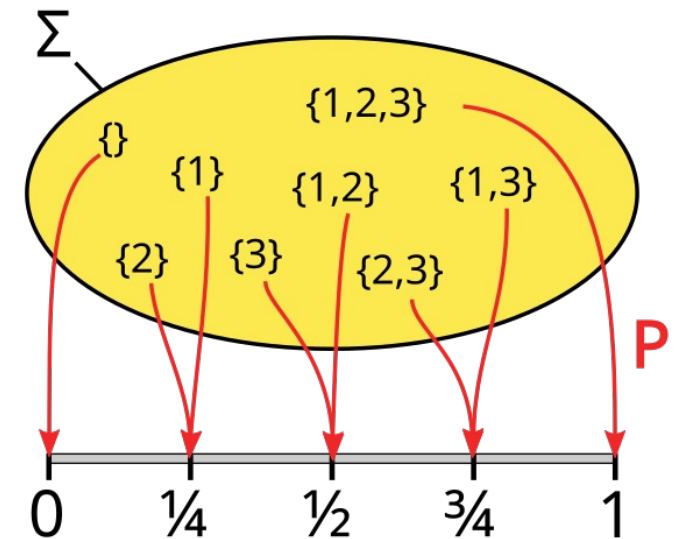
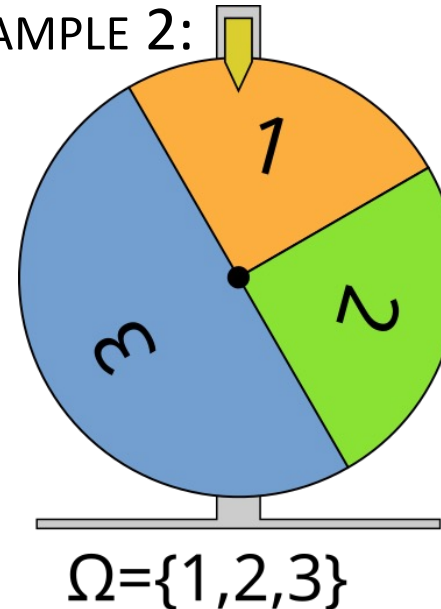
Say the outcome is a 1

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

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



EXAMPLE 2:



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 $E \subseteq \Omega, E \in \Sigma$

- Probability measure:

~~$$\mathbb{P}: \Omega \rightarrow [0, 1]$$~~

$$\mathbb{P}: \Sigma \rightarrow [0, 1]$$

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

What counts as a valid probability measure.

From that follows: $\mathbb{P}(\emptyset) = 0, \mathbb{P}(E^c) = 1 - \mathbb{P}(E)$

EXAMPLE 1:

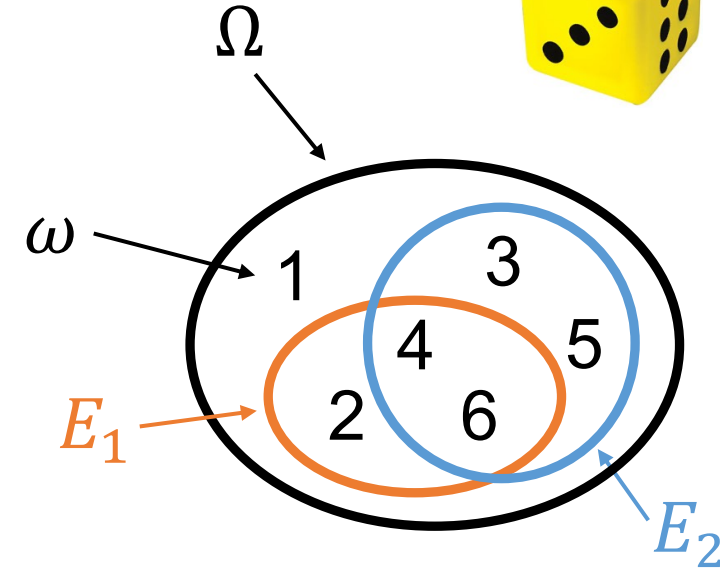
roll a fair die with 6 sides

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

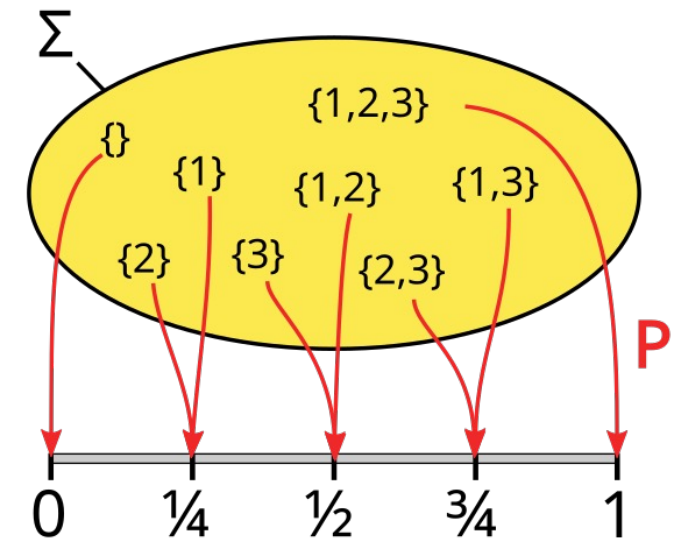
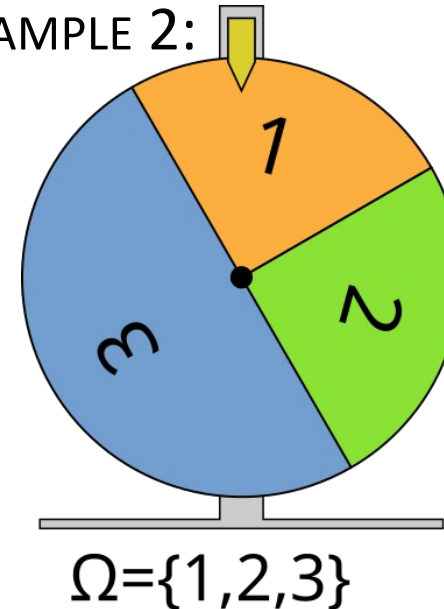
Say the outcome is a 1

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

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



EXAMPLE 2:



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

Gatterbauer. Foundations and Applications of Information Theory: <https://northeastern-datalab.github.io/cs7840/fa26/>

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:

~~$\mathbb{P}: \Omega \rightarrow [0, 1]$~~
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

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

What counts as a valid probability measure.

From that follows: $\mathbb{P}(\emptyset) = 0, \mathbb{P}(E^c) = 1 - \mathbb{P}(E)$

EXAMPLE 1:

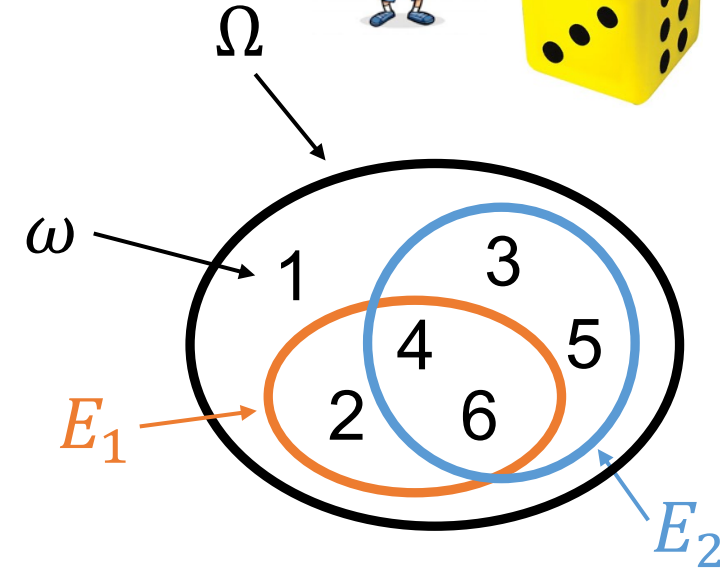
roll a fair die with 6 sides

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:

$$\cancel{\mathbb{P}: \Omega \rightarrow [0, 1]}$$
$$\mathbb{P}: \Sigma \rightarrow [0, 1]$$

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

What counts as a valid probability measure.

EXAMPLE 1:

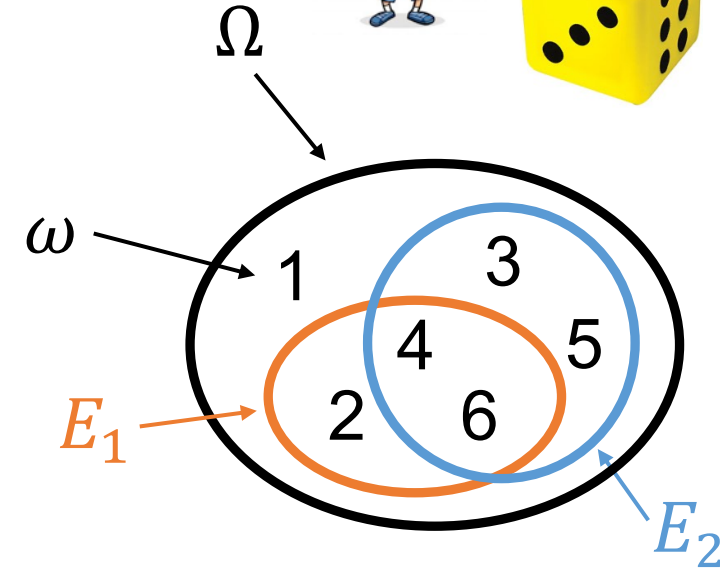
roll a fair die with 6 sides

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

Say the outcome is a 1

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

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



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

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

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

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

$$\mathbb{P}(E_2) =$$

?

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:

$$\cancel{\mathbb{P}: \Omega \rightarrow [0, 1]}$$
$$\mathbb{P}: \Sigma \rightarrow [0, 1]$$

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

What counts as a valid probability measure.

EXAMPLE 1:

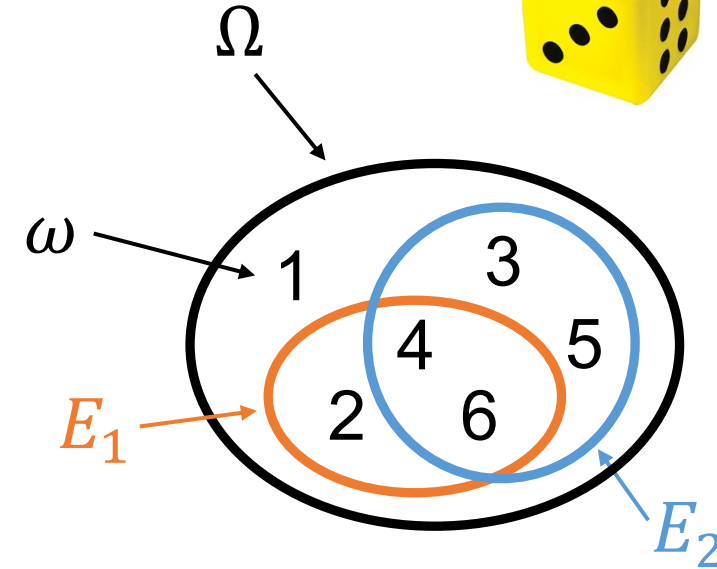
roll a fair die with 6 sides

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

Say the outcome is a 1

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

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



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

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

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

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

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

Sample space / outcomes / events / probability



PROBABILITY

- A 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 $E \subseteq \Omega, E \in \Sigma$
- 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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1						
	2						
	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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,3)	(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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

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

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

$$|E_1| = 6$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,3)	(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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

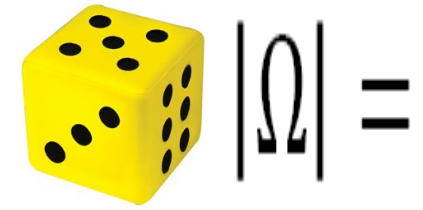
		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

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

$$|E_2| =$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,3)	(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 = "total is greater than 8"

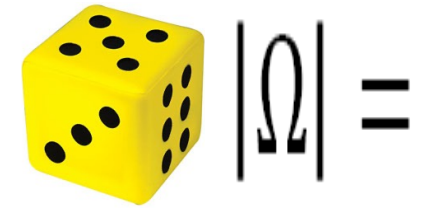
= "total is 9 or 10 or 11 or 12"

↓

$$|E_2| = 4$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

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

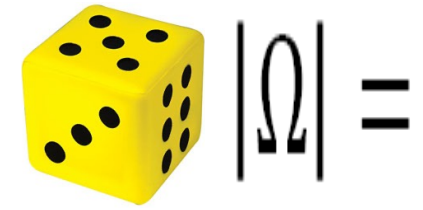
E_2 = "total is greater than 8"

= "total is 9 or 10 or 11 or 12"

$$|E_2| = \begin{array}{c} \downarrow \quad \downarrow \\ 4 + 3 \end{array}$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

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

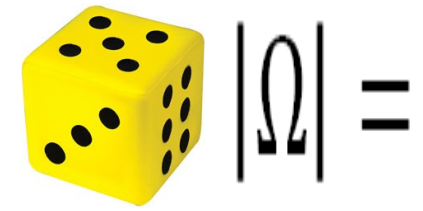
E_2 = "total is greater than 8"

= "total is 9 or 10 or 11 or 12"

$$|E_2| = \begin{array}{c} \downarrow \quad \downarrow \quad \downarrow \\ 4 + 3 + 2 \end{array}$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

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

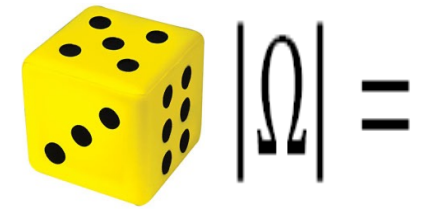
E_2 = "total is greater than 8"

= "total is 9 or 10 or 11 or 12"

$$|E_2| = \begin{array}{ccccccc} & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & & 4 & + & 3 & + & 2 & + & 1 \end{array}$$

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

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 $E \subseteq \Omega, E \in \Sigma$
- Probability measure:
 $\mathbb{P}: \Sigma \rightarrow [0, 1]$

EXAMPLE 3:

roll two fair dice with 6 sides

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

$$|\Omega| = 36$$

		second die					
		1	2	3	4	5	6
first die	1	(1,1)	(1,2)	(1,3)	(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 = "total is greater than 8"

= "total is 9 or 10 or 11 or 12"

additivity of
disjoint events

$$|E_2| = \begin{array}{ccccccc} & \downarrow & \downarrow & \downarrow & \downarrow & & \\ & 4 & + & 3 & + & 2 & + & 1 & = & 10 \end{array}$$

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

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-SIDED DIE):

Ω



1	2	3	5	6	7	10		11
13				16				
14	4	12		8	9		15	

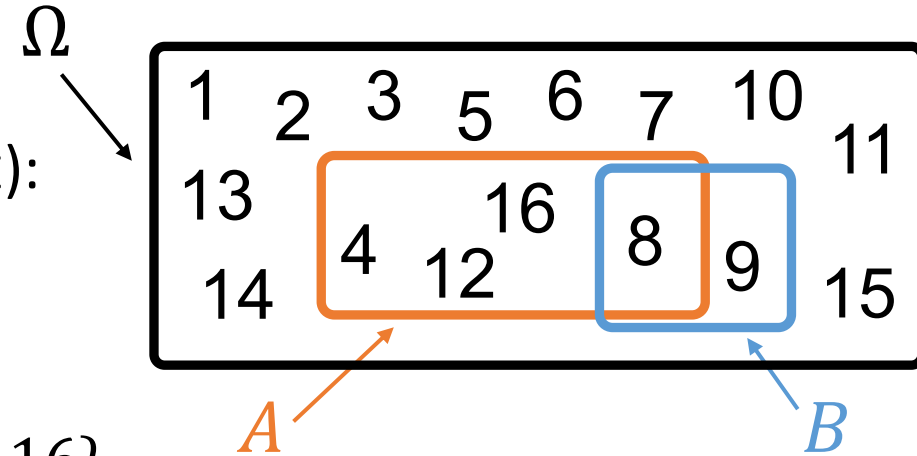
$$\Omega = \{1, 2, \dots, 16\}$$

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-SIDED DIE):



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

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

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

Our colorful hands represent "team exercises" If we are online, please make a screenshot!



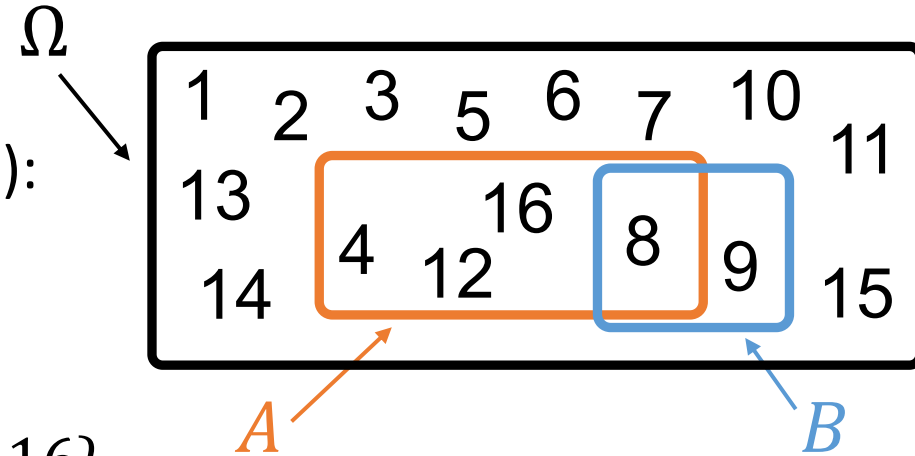
$$\mathbb{P}(A|B) = ?$$

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-SIDED DIE):



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

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

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

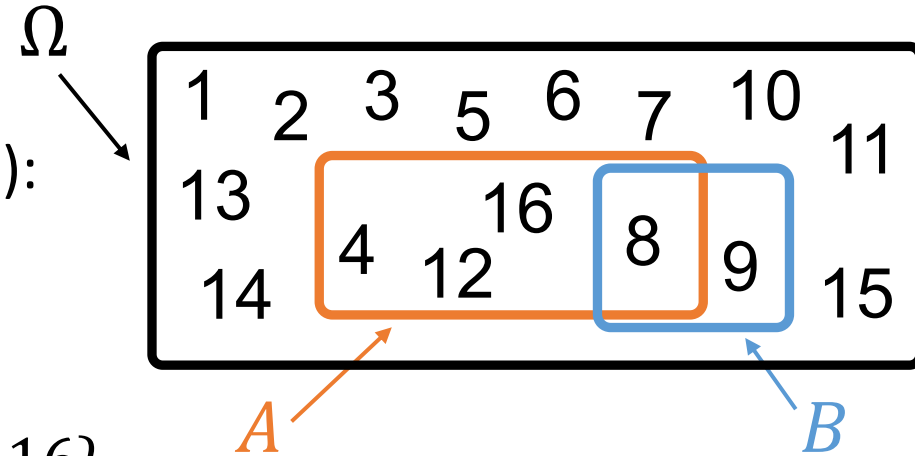
$$\mathbb{P}(A|B) = \frac{1}{2}$$

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-SIDED DIE):



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

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

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

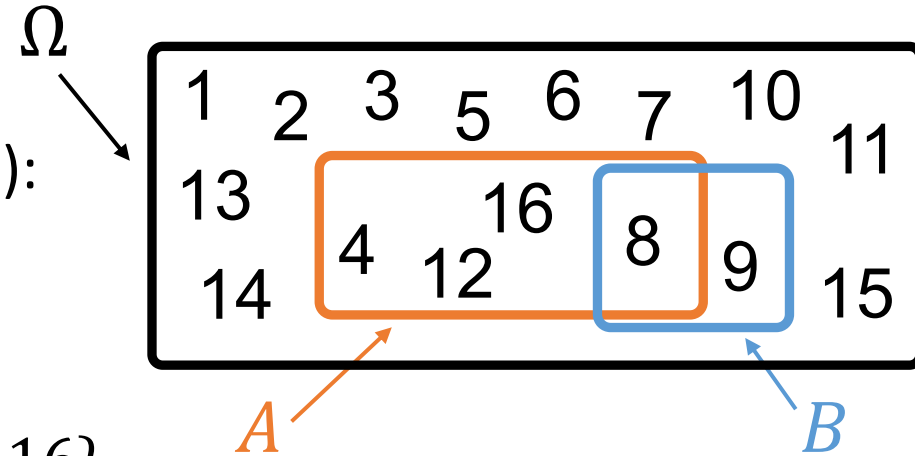
$$\mathbb{P}(A|B) = \frac{\text{Number of outcomes in } A \cap B}{\text{Number of outcomes in } B}$$

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-SIDED DIE):



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

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

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

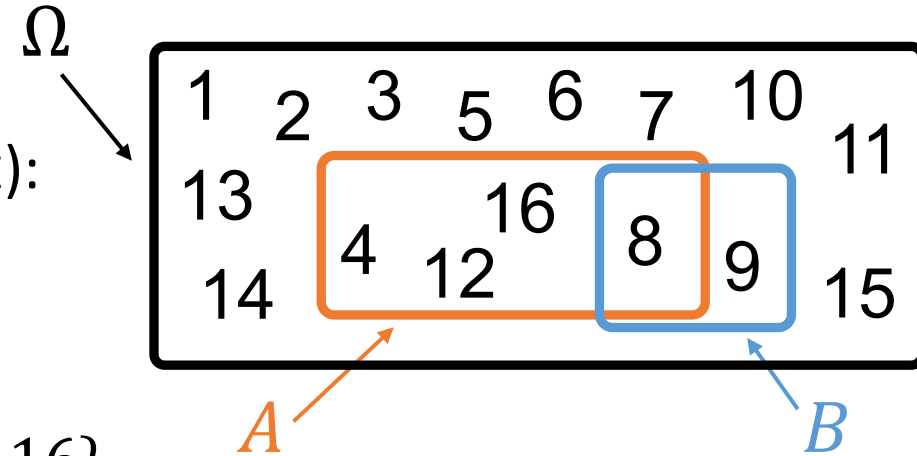
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

Conditional probability

DEFINITION

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

EXAMPLE 4
(FAIR 16-SIDED DIE):



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

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

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

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

Conditional probability

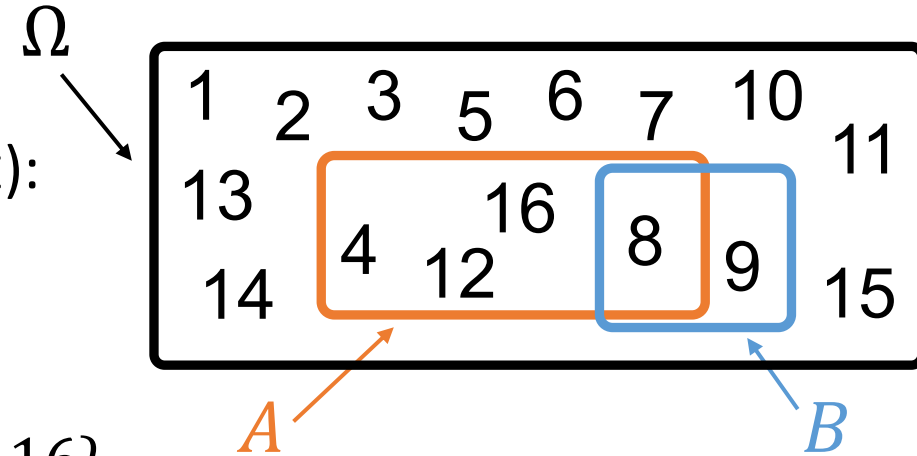
DEFINITION

Conditional probability $\mathbb{P}(A|B)$ 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-SIDED DIE):



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

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

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

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

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

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

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

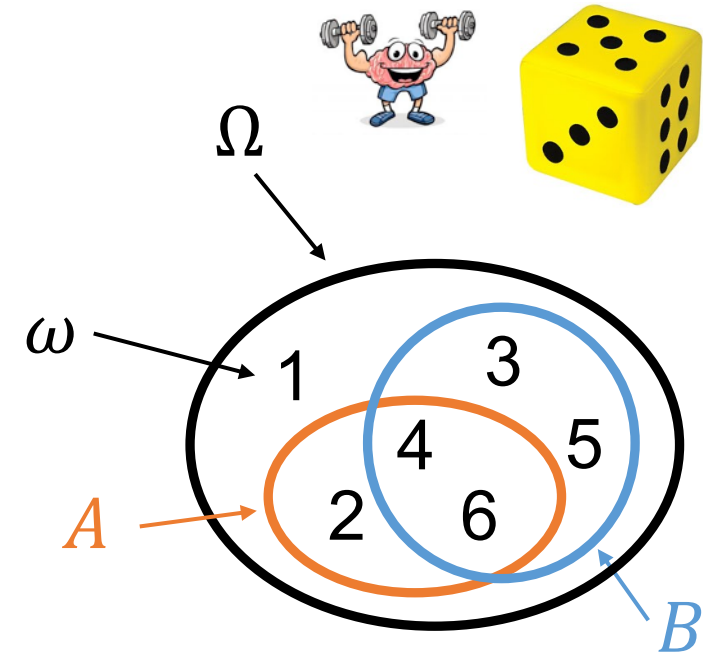
EXAMPLE 1 (CONTINUED):

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

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

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

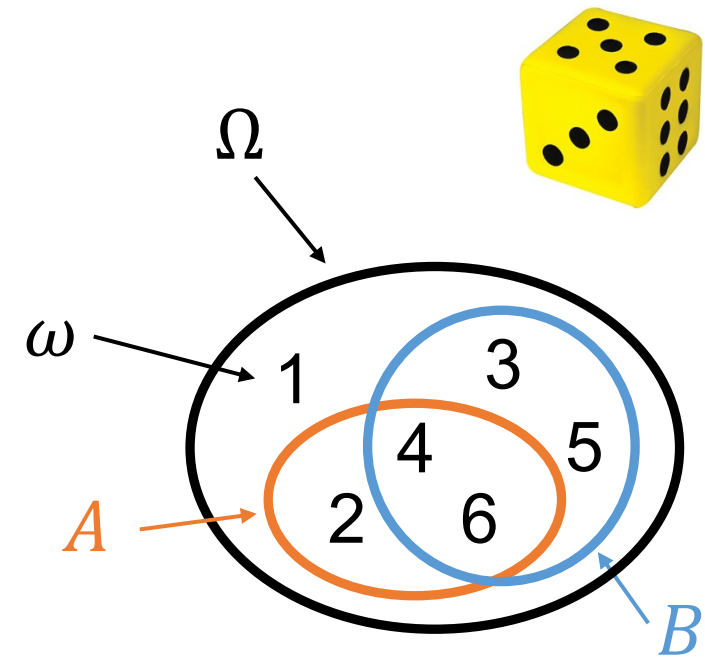
EXAMPLE 1 (CONTINUED):

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 !

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

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

Independence

DEFINITION:

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

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

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

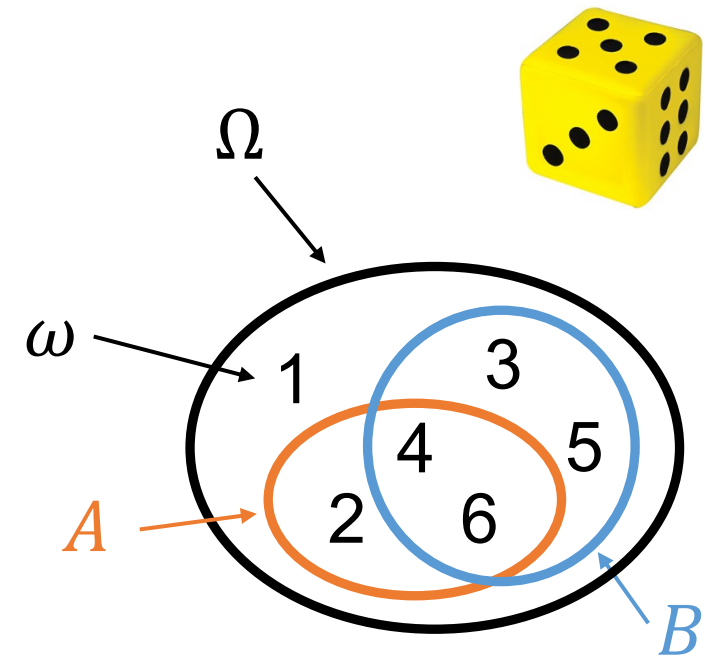
EXAMPLE 1 (CONTINUED):

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

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

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

Also see orthogonality in CS: "Orthogonality in programming language design is the ability to use various language features in arbitrary combinations with consistent results. This usage was introduced by Van Wijngaarden in the design of Algol 68: 'The number of independent primitive concepts has been minimized in order that the language be easy to describe, to learn, and to implement. On the other hand, these concepts have been applied "orthogonally" in order to maximize the expressive power of the language while trying to avoid deleterious superfluities.'"

https://en.wikipedia.org/wiki/Orthogonality#Computer_science

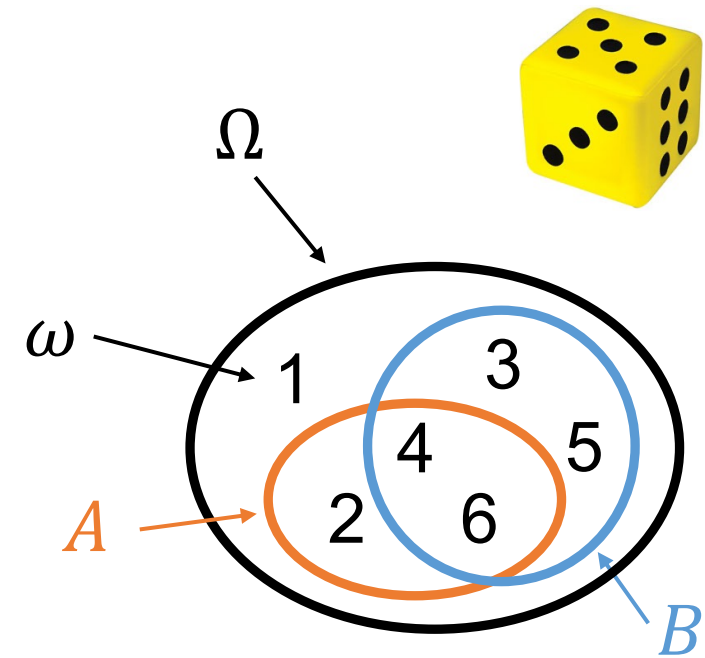
EXAMPLE 1 (CONTINUED):

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 !

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

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

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

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

Think about a coordinate system: x and y are independent

Independence

DEFINITION:

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

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

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

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



Independence

DEFINITION:

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

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

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

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

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

?

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

?

Independence

DEFINITION:

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

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

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

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

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

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

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

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

$\nearrow \neq 0$

Independence

DEFINITION:

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

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

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

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
DEFINITION: that to be equivalent to the earlier definition.

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

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

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

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

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

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

$\nearrow$
 $\neq 0$

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):

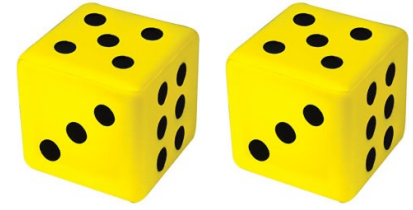
roll two fair dice with 6 sides

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

A = "1st roll is 1"

B = "sum is 7"

Are A and B independent? ?



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"

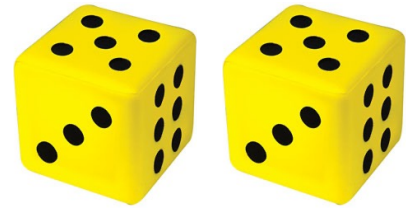
B = "sum is 7"

Are A and B independent?

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

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

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



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"

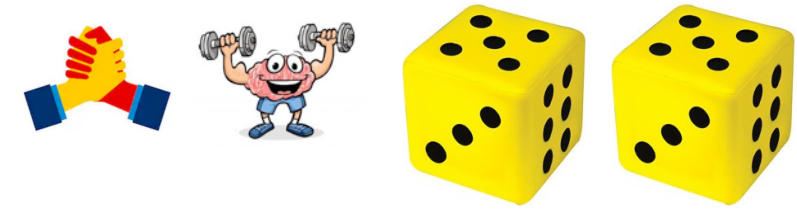
B = "sum is 7"

Are A and B independent?

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

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

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



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

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"

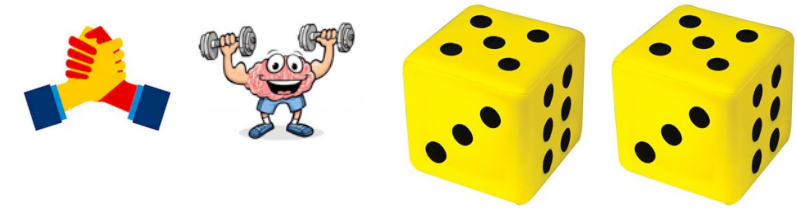
Are A and B independent?

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

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

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

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



second die

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

first die

Handwritten annotations: An orange circle highlights the first row (1,1) through (1,6). A blue circle highlights the first column (1,1) through (6,1). A red circle highlights the cell (1,6). A blue diagonal line runs from (1,1) to (6,6). Handwritten fractions: 1/6 next to the first row, 1/6 next to the first column, and 1/36 next to the cell (1,6).

Chain rules

Chain Rule

DEFINITION:

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

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

Recall the definition of cond. prob.:

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

Chain Rule

DEFINITION:

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

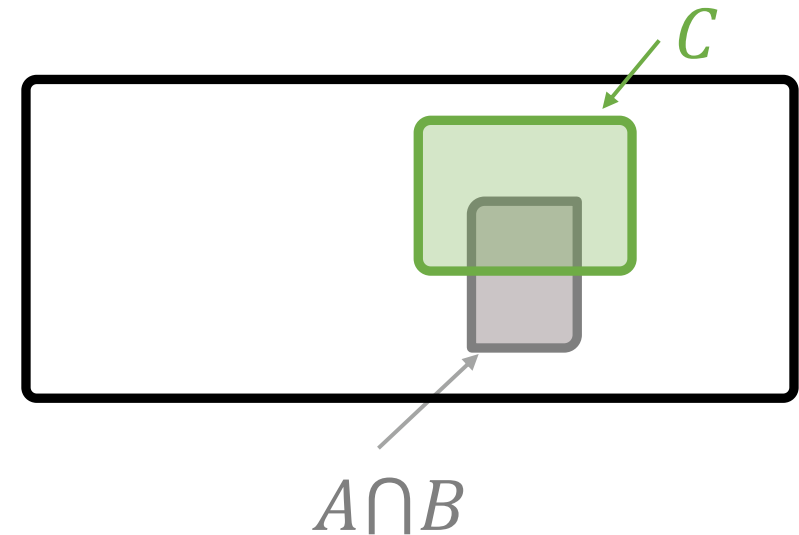
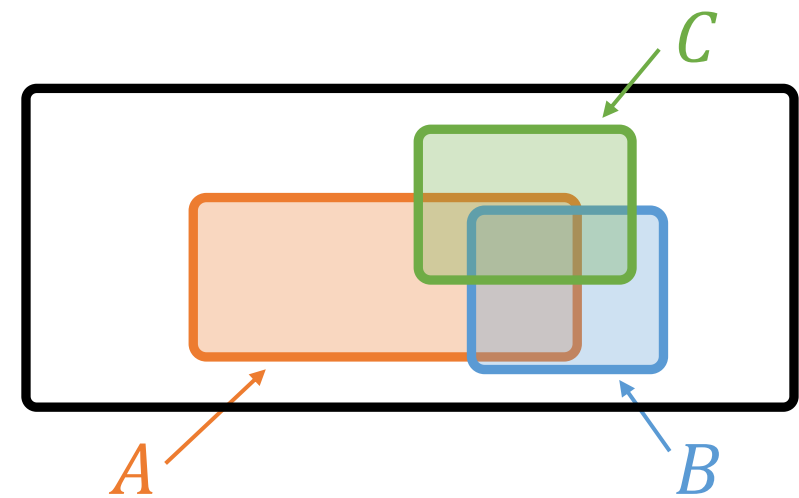
$$\mathbb{P}(\underbrace{A, B}_E, C) = \mathbb{P}(\underbrace{A, B}_E) \cdot \mathbb{P}(C|\underbrace{A, B}_E)$$

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

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

Recall the definition of cond. prob.:

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



Bayes' law

Bayes' Law

DEFINITION

Conditional probability $\mathbb{P}(A|B)$ 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

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

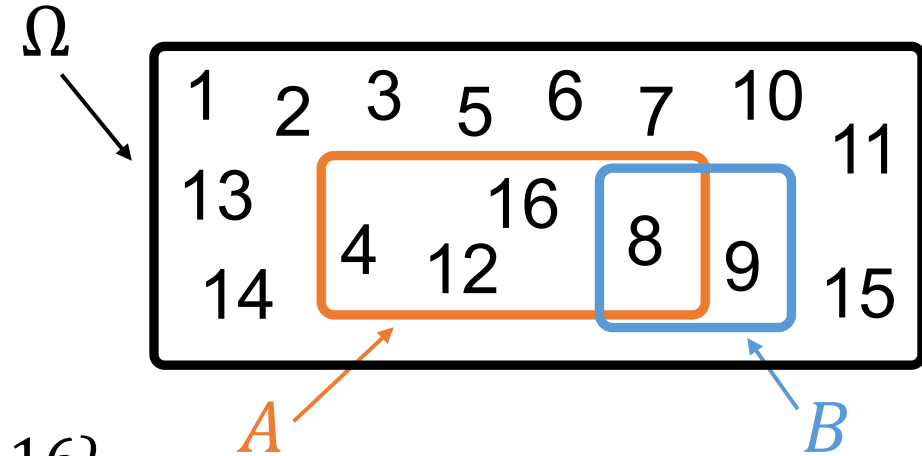
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

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

$$\begin{aligned}\mathbb{P}(A, B) &= \mathbb{P}(B|A) \cdot \mathbb{P}(A) \\ &= \mathbb{P}(A|B) \cdot \mathbb{P}(B)\end{aligned}$$

EXAMPLE 4:

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



BAYES' LAW

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

?

Bayes' Law

DEFINITION

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

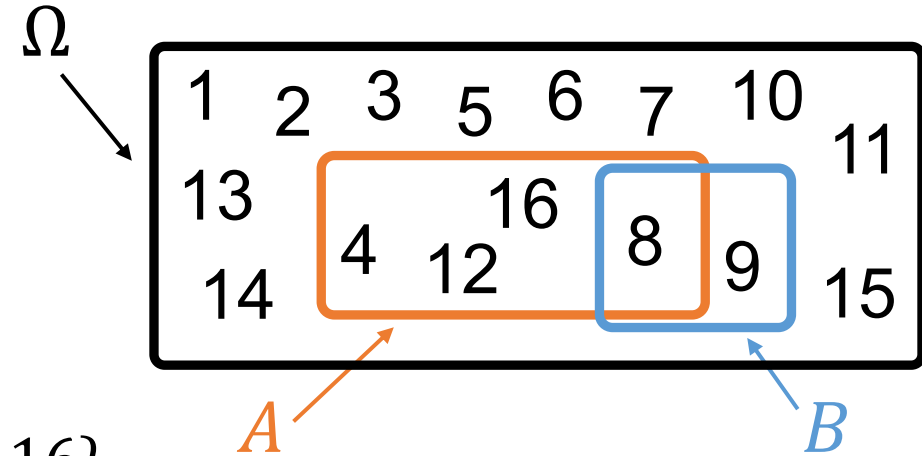
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

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

$$\begin{aligned}\mathbb{P}(A, B) &= \mathbb{P}(B|A) \cdot \mathbb{P}(A) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} \\ &= \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{8} = \frac{1}{16}\end{aligned}$$

EXAMPLE 4:

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



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?

but first: what are A and B



Bayes' Law ("Base rate fallacy problem")



EXAMPLE 5: Zika in Florida in 2016

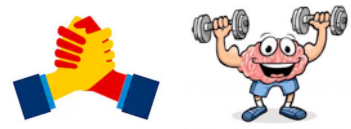
- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k)
- accuracy of blood test is 99%
- A patient has a positive test.

What is the chance they have Zika?

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



Bayes' Law ("Base rate fallacy problem")



EXAMPLE 5: Zika in Florida in 2016

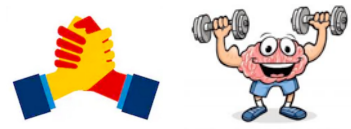
- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k)
- accuracy of blood test is 99%
- A patient has a positive test.

What is the chance they have Zika?



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

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?



patient has Z
Zika (Z)

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

false negative

false positive

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

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

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



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	$\mathbb{P}(B Z) = 0.99$	$\mathbb{P}(\bar{B} Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B \bar{Z}) = 0.01$	$\mathbb{P}(\bar{B} \bar{Z}) = 0.99$

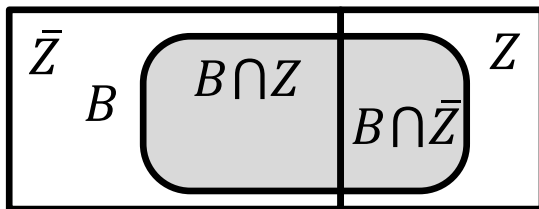
We don't want $\mathbb{P}(B|Z) = 0.99$, but rather $\mathbb{P}(Z|B)$

but weighted

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

?

total probability theorem



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

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?

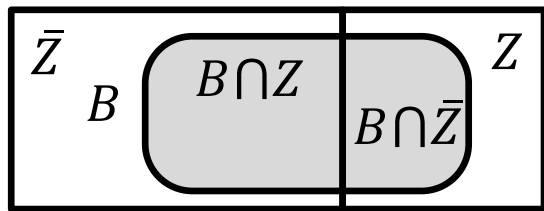
		blood test positive (B)	
		B	$\bar{B}$
patient has Zika (Z)	Z	$\mathbb{P}(B Z) = 0.99$	$\mathbb{P}(\bar{B} Z) = 0.01$
	$\bar{Z}$	$\mathbb{P}(B \bar{Z}) = 0.01$	$\mathbb{P}(\bar{B} \bar{Z}) = 0.99$

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

but weighted

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

total probability theorem



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

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

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

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

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

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

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

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	99	1	100
	$\bar{Z}$			9,999,900

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

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	99	1	100
	$\bar{Z}$	99,999	9,899,901	9,999,900

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

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	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

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!

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

blood test positive (B)		
B	$\bar{B}$	
99	1	100
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

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	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

$\bigcirc \rightarrow \mathbb{P}(Z) = 10^{-5}$

Bayes' Law ("Base rate fallacy problem")



EXAMPLE 5: Zika in Florida in 2016

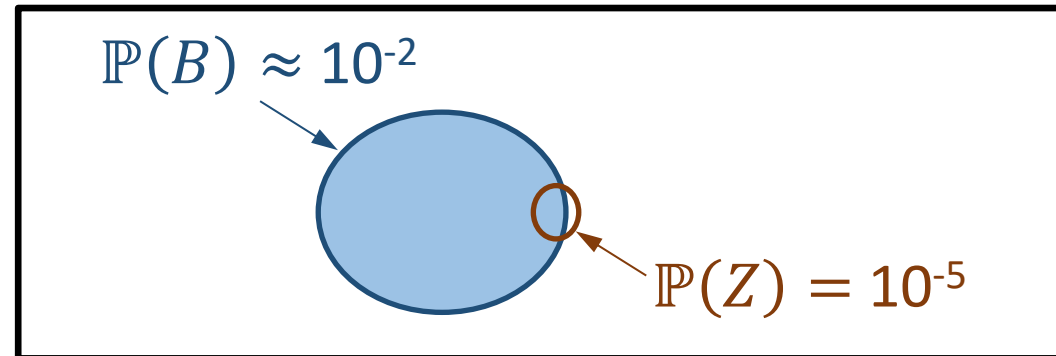
- prevalence of Zika (Z) in Florida is 10^{-5} (1 in 100k) = **base rate** = $\mathbb{P}(Z)$
- accuracy of blood test is 99%
- Assume 10M people in FL. Complete the numbers!

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



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

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

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

patient has Z
Zika (Z) $\bar{Z}$

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

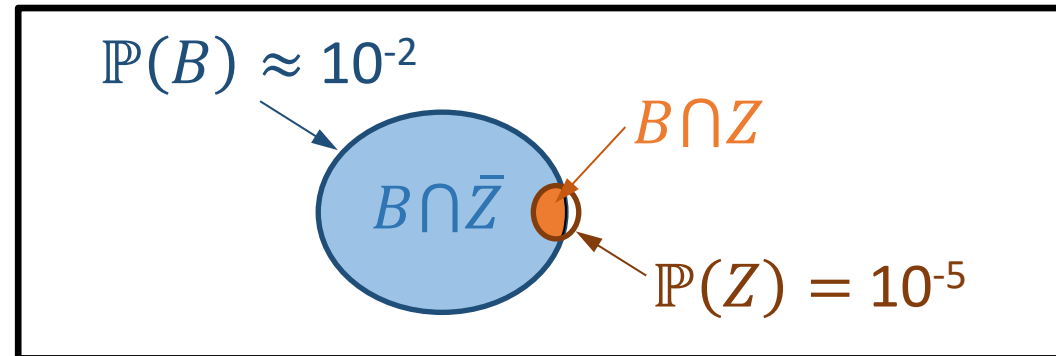
$$\mathbb{P}(B|Z) \cdot \mathbb{P}(Z) + \mathbb{P}(\bar{B}|\bar{Z}) \cdot \mathbb{P}(\bar{Z}) = 99\%$$

probability that a random test is correct

$\neq$

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

probability that: a positive test is correct



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 (fa26)

<https://northeastern-datalab.github.io/cs7840/fa26/>

9/17/2026

Pre-class conversations

- Last class recapitulation
- First mini-projects on Canvas: I added comments, you can make another pass
- Pointing out errors: Thanks! Feel free to use non-private messages
- Procedure: Preliminary posting of slides internally on Piazza ("unedited version"). We will keep that, thanks for the earlier post
- Today:
 - A lot of fun with basic probability theory 😊
 - Start of information theory basics

Deterministic DNNF Is Not Polynomially Closed Under Negation

Guy Van den Broeck*
Computer Science Department
University of California, Los Angeles

...

*This paper builds on contributions by Oliver Broadrick. He has not yet agreed to be an author, but has agreed to the circulation of this note.

¹This proof was developed with substantial AI assistance. At the time of writing, scientific standards for AI-assisted theoretical results are not yet settled. We have chosen to take three steps:

1. We have carefully checked the claims and proofs in this paper and take full responsibility for their correctness.
2. We have presented the proof as a detailed tutorial requiring little background beyond undergraduate mathematics. Footnotes throughout the proof highlight connections to more advanced topics. The tutorial proof aims to make the reasoning and the ideas behind the construction accessible to human readers, motivated by Tao's view that "*A proof that no human can properly explain should be viewed as incomplete, even if it has been formally verified*" [74, Section 8].
3. Readers can check the nonclosure claim by inspecting a short formal trust boundary: the Lean definitions and theorem statement explained in Appendix A. If these express the intended claim, the machine-checked proof establishes it, assuming the correctness of Lean's kernel. Readers who also want to formally check that the theorem holds for the reasons given in the tutorial proof can compare the two proofs: each statement links to its Lean formalization.

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.

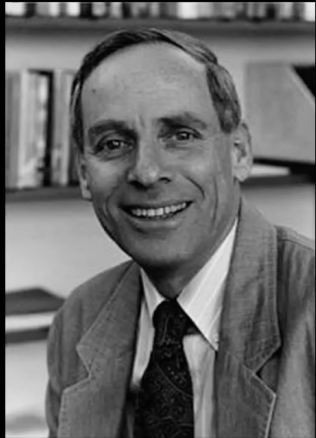


Bayes theorem illustrated by 3Blue1Brown

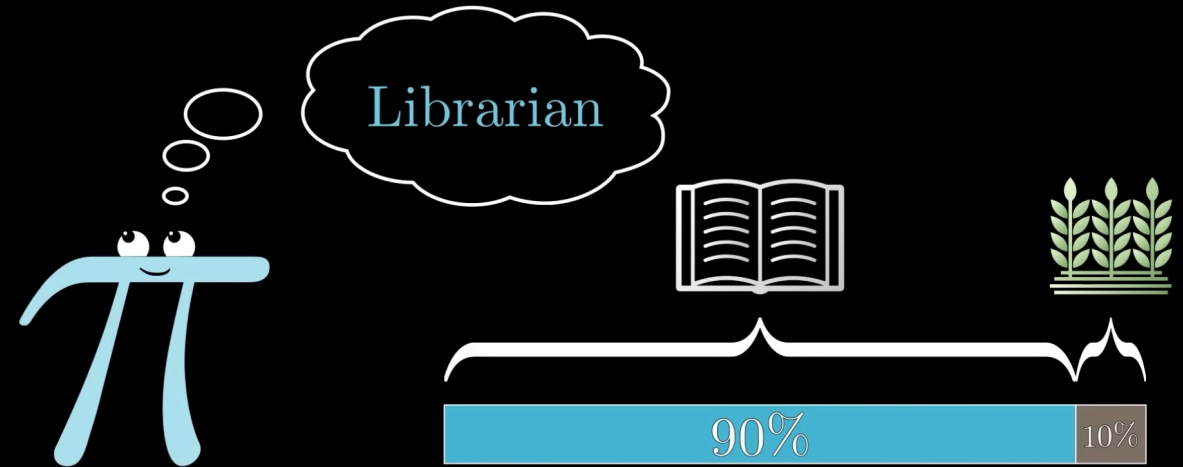


Daniel
Kahneman

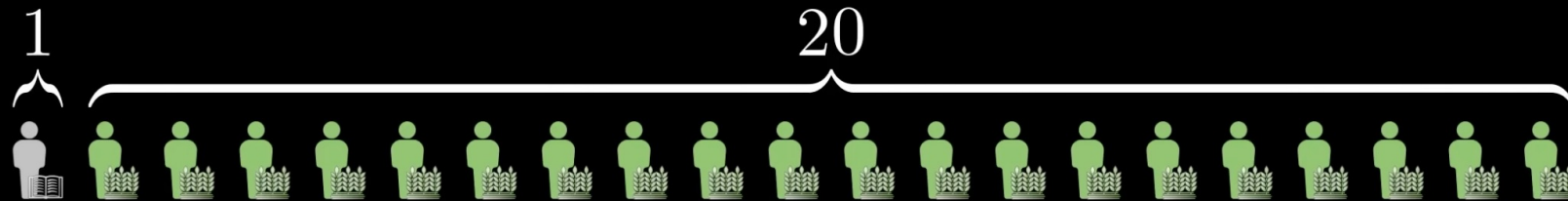
Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.



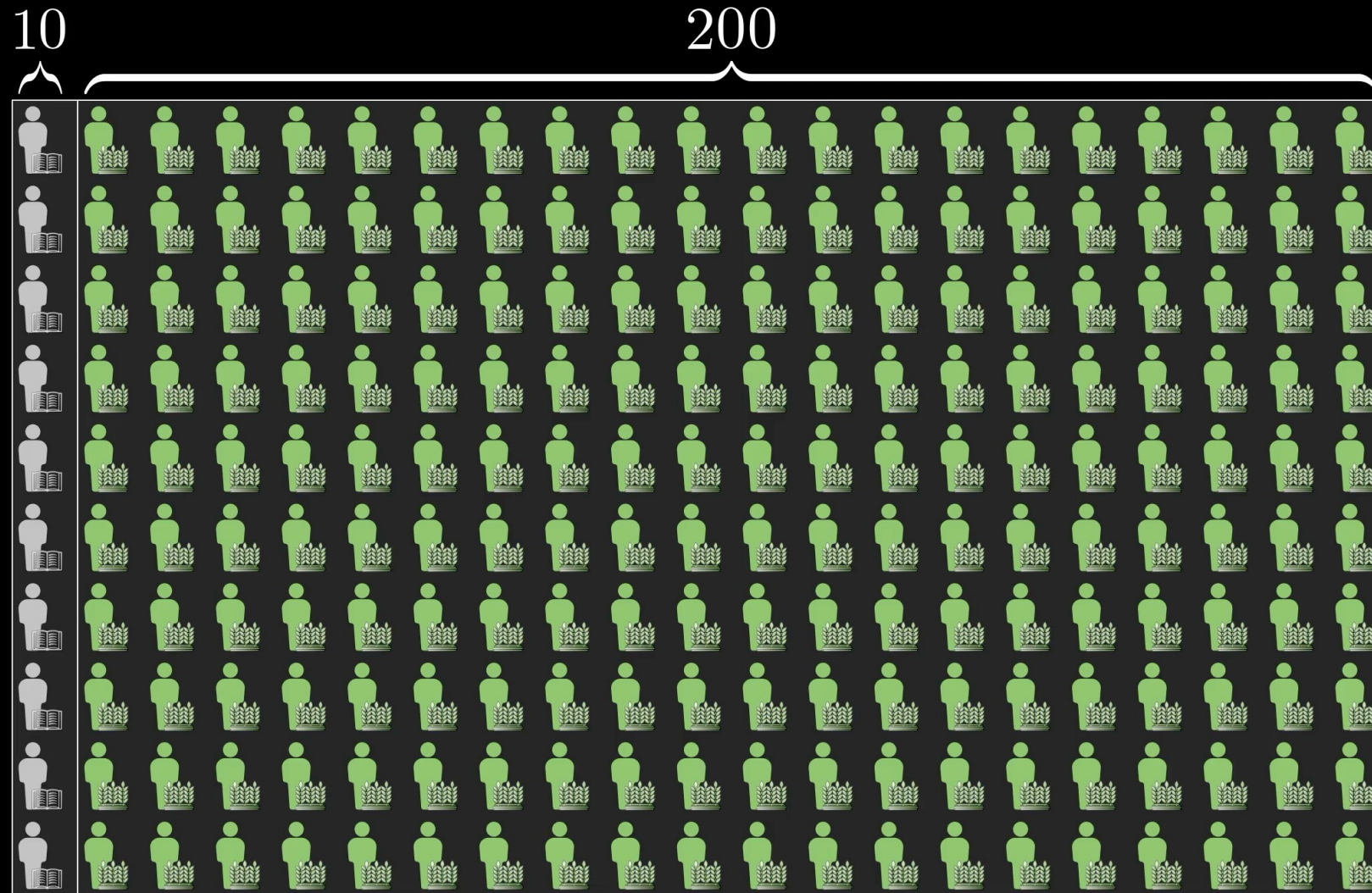
Amos
Tversky



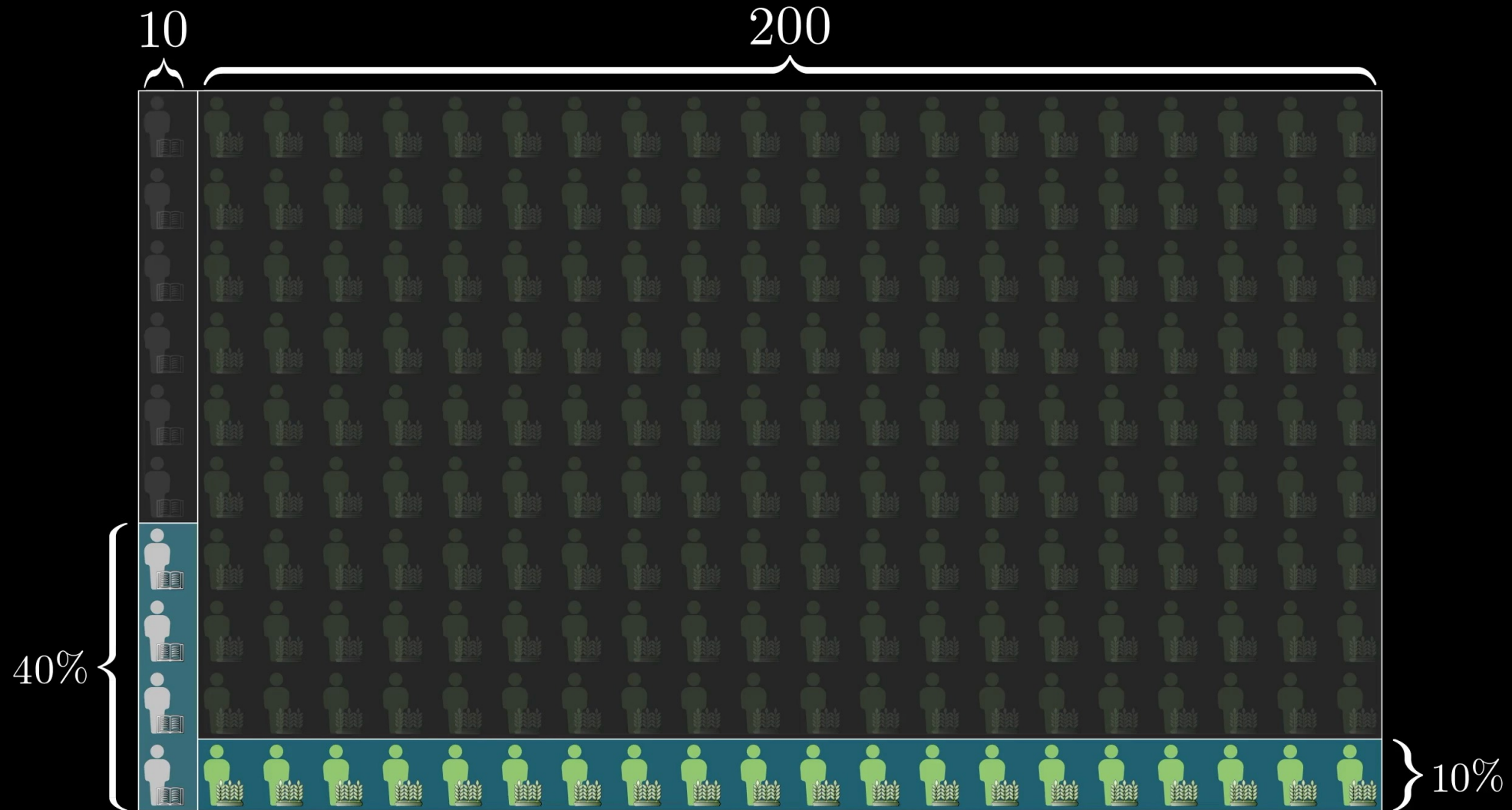
Bayes theorem illustrated by 3Blue1Brown



Bayes theorem illustrated by 3Blue1Brown

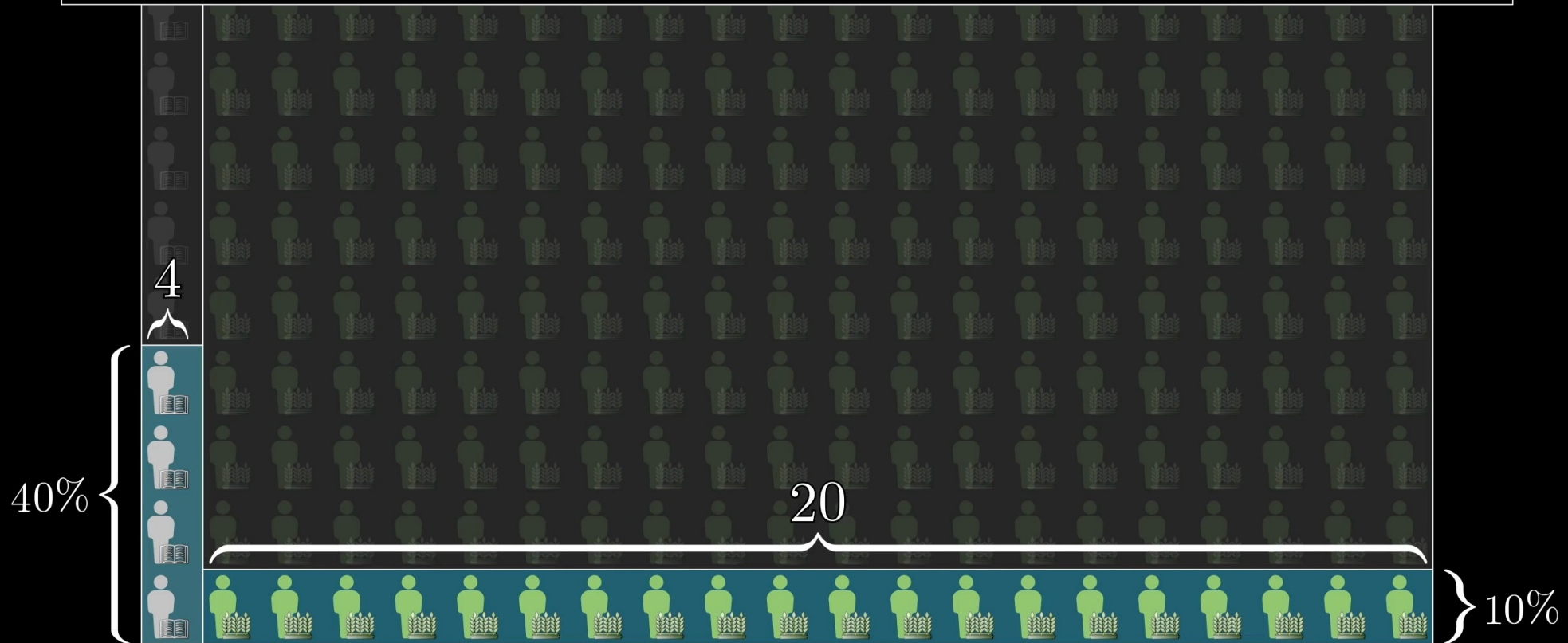


Bayes theorem illustrated by 3Blue1Brown



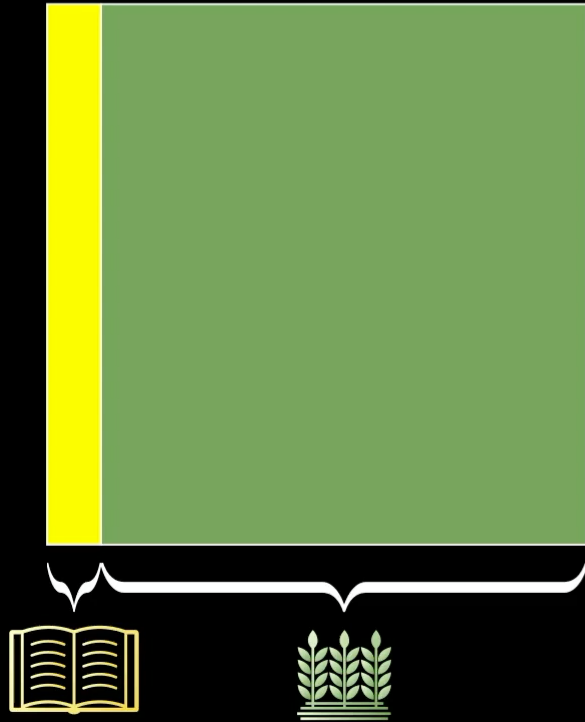
Bayes theorem illustrated by 3Blue1Brown

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

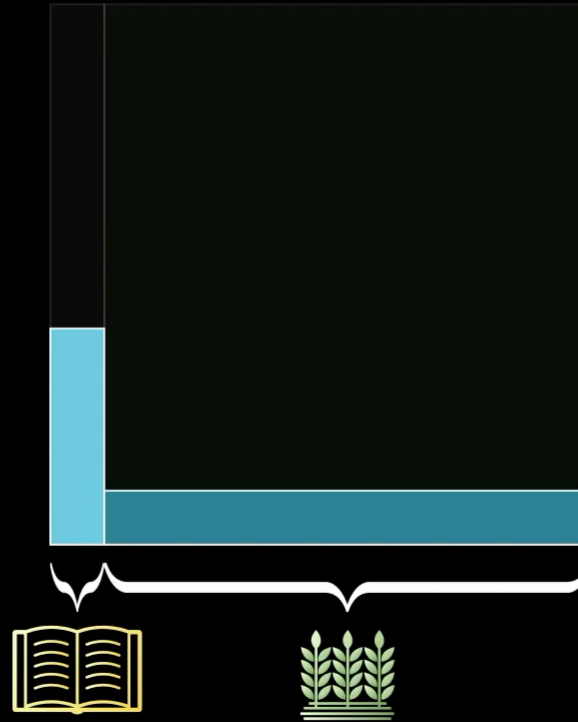


Bayes theorem illustrated by 3Blue1Brown

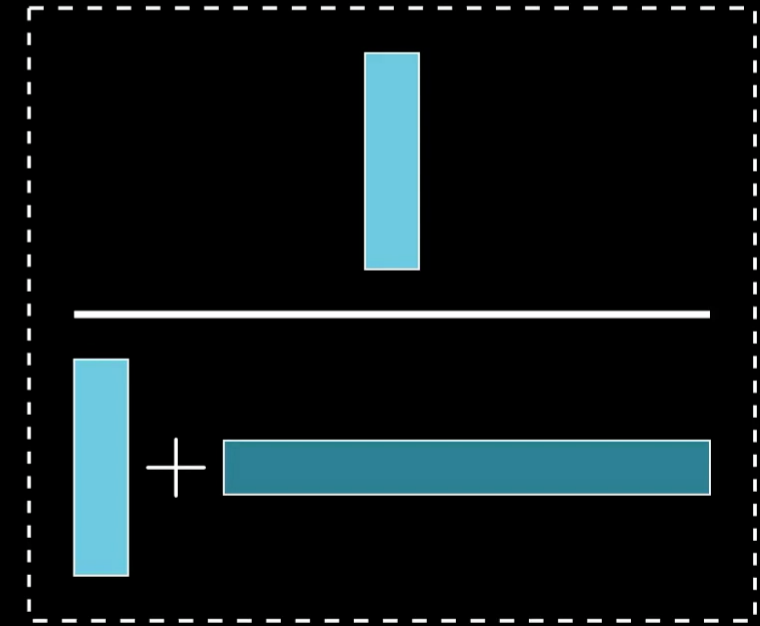
All possibilities



All possibilities
fitting the evidence



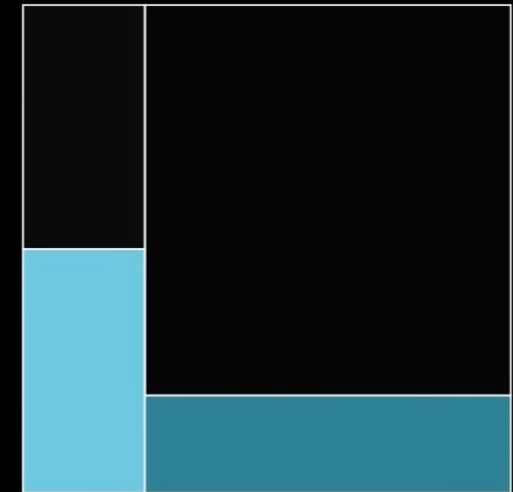
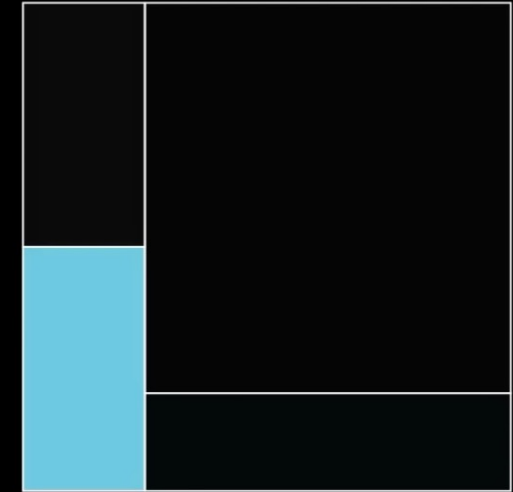
$P\left(\text{Librarian given the evidence}\right)$



Bayes theorem illustrated by 3Blue1Brown

Bayes' theorem

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} = \frac{\text{[Diagram 1]}}{\text{[Diagram 2]}}$$



Sample space / outcomes / events / probability

$$P(H|E) = \frac{\text{1 person in blue column}}{\text{1 person in blue column} + \text{20 people in yellow box}} = \frac{\overbrace{(\# \text{ person}) P(H) P(E|H)}^4}{\underbrace{(\# \text{ person}) P(H) P(E|H) + (\# \text{ person}) P(\neg H) P(E|\neg H)}_{20}}$$

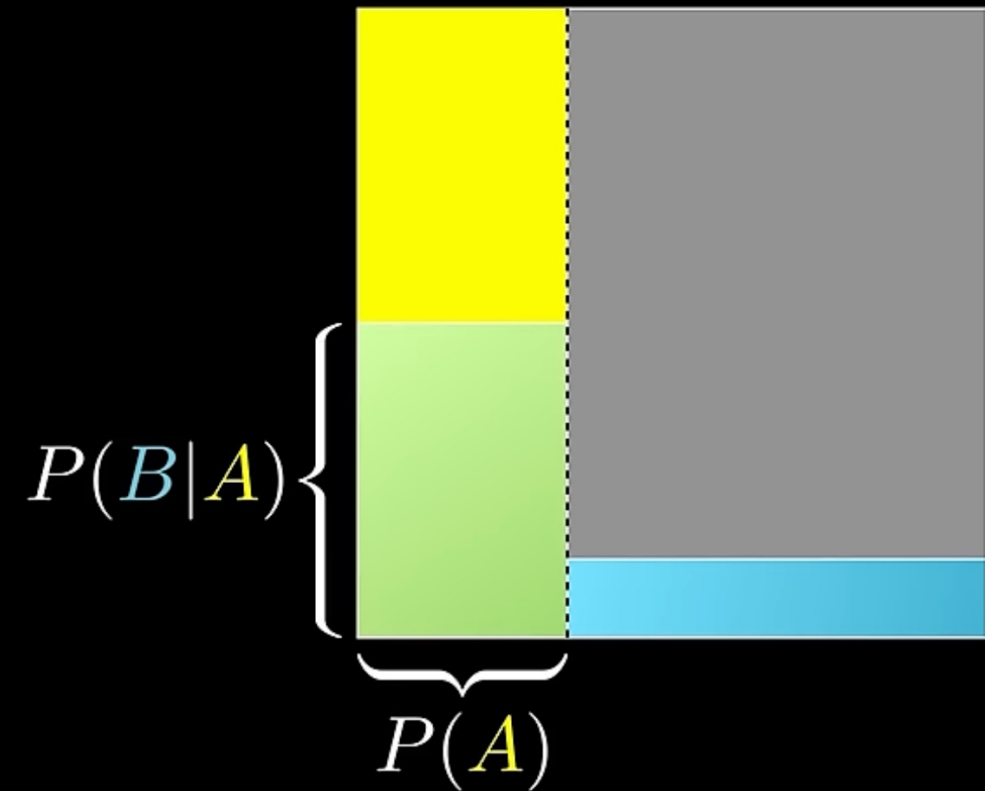
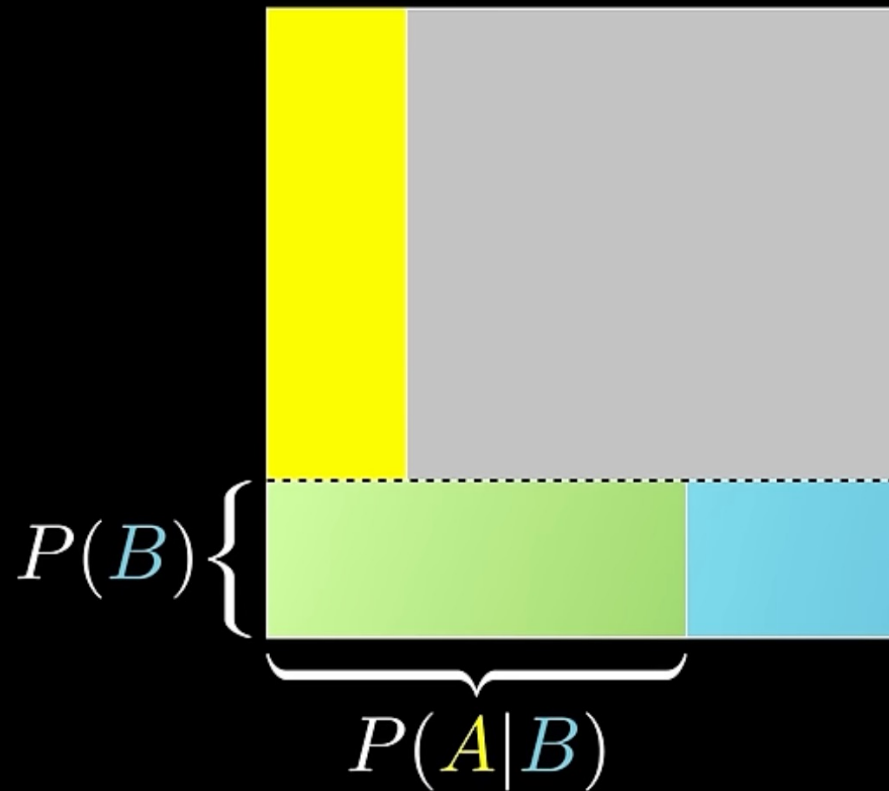
$$P(\neg H) = 20/21$$

“Likelihood”

$$P(E|H) = 0.4 \left\{ \begin{array}{l} \text{4 people in blue column} \\ \text{20 people in yellow box} \end{array} \right\} P(E|\neg H) = 0.1$$

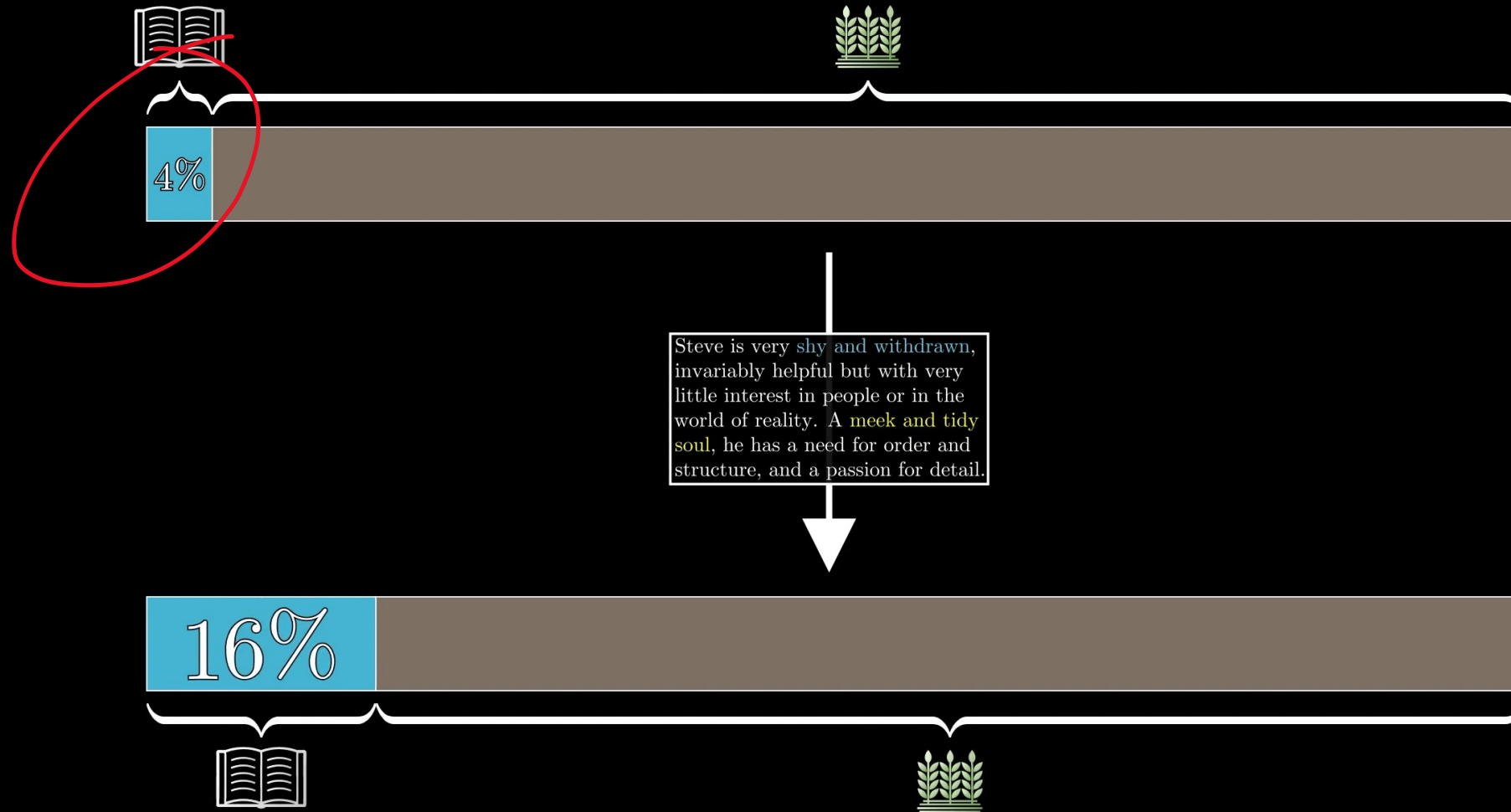
Sample space w/ dependent variables: $P(B) \neq P(B|A)$

$$P(B)P(A|B) = P(A \text{ and } B) = P(A)P(B|A)$$



Bayes theorem illustrated by 3Blue1Brown

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



Random variable, expectation

Random variable

Random variable (RV)



Random variable

Random variable (RV)

$$X: \Omega \rightarrow \mathbb{R}$$



Random variable

Random variable (RV)

$X: \Omega \rightarrow \mathbb{R}$

sample space of random experiment

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

Two types:



Random variable

Random variable (RV)

$$X: \Omega \rightarrow \mathbb{R}$$

sample space of random experiment

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

Two types:

1. numerical

e.g. $X(\omega)$ lottery win

2. indicator

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

Random variable: pmf $p_X(x)$

Random variable (RV)

$$X: \Omega \rightarrow \mathbb{R}$$

sample space of random experiment

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

Two types:

1. numerical

e.g. $X(\omega)$ lottery win

2. indicator

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

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

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

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

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

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

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)	(1,3)	(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)

$$X: \Omega \rightarrow \mathbb{R}$$

sample space of random experiment

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

Two types:

1. numerical

e.g. $X(\omega)$ lottery win

2. indicator

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

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

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

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

EXAMPLE 3 (CONT.):

roll two fair dice with 6 sides

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

Let X be the sum of the rolls.

Then X is a RV.

What is the pmf?



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

$$X: \Omega \rightarrow \mathbb{R}$$

sample space of random experiment

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

Two types:

1. numerical

e.g. $X(\omega)$ lottery win

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

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

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

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

EXAMPLE 3 (CONT.):

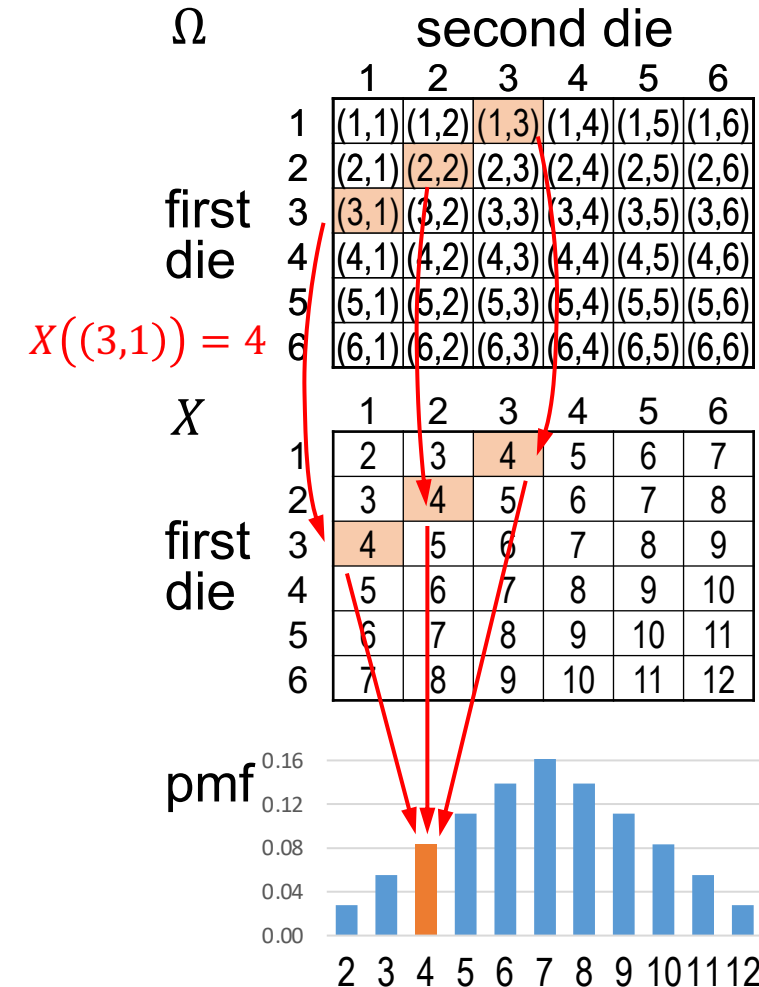
roll two fair dice with 6 sides

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

Let X be the sum of the rolls.

Then X is a RV.

What is the pmf?



Random variable: expectation $\mathbb{E}$

Random variable (RV)

$X: \Omega \rightarrow \mathbb{R}$

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

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

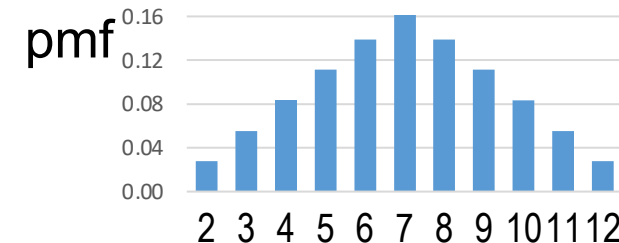
Let X be the sum of the rolls.



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

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

$$X: \Omega \rightarrow \mathbb{R}$$

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 X be the sum of the rolls.



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

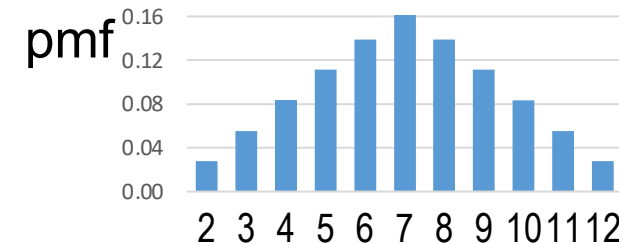
Variant 1:

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

?

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

$X: \Omega \rightarrow \mathbb{R}$

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 X be the sum of the rolls.



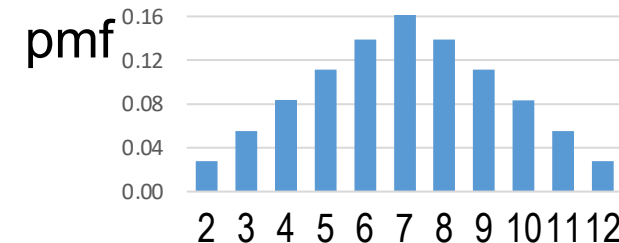
What is $\mathbb{E}[X]$?

Variant 1:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) \\ &= \frac{1}{36} \cdot \sum_{\omega \in \Omega} X(\omega) \\ &= \frac{1}{36} \cdot (\text{"sum of table"}) \\ &= 7 \end{aligned}$$

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

$$X: \Omega \rightarrow \mathbb{R}$$

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 X be the sum of the rolls.



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

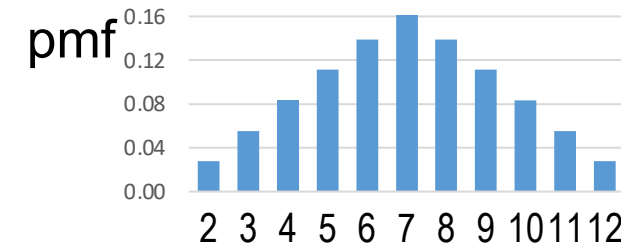
Variant 2:

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

?

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

$$X: \Omega \rightarrow \mathbb{R}$$

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 X be the sum of the rolls.



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

Variant 2:

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

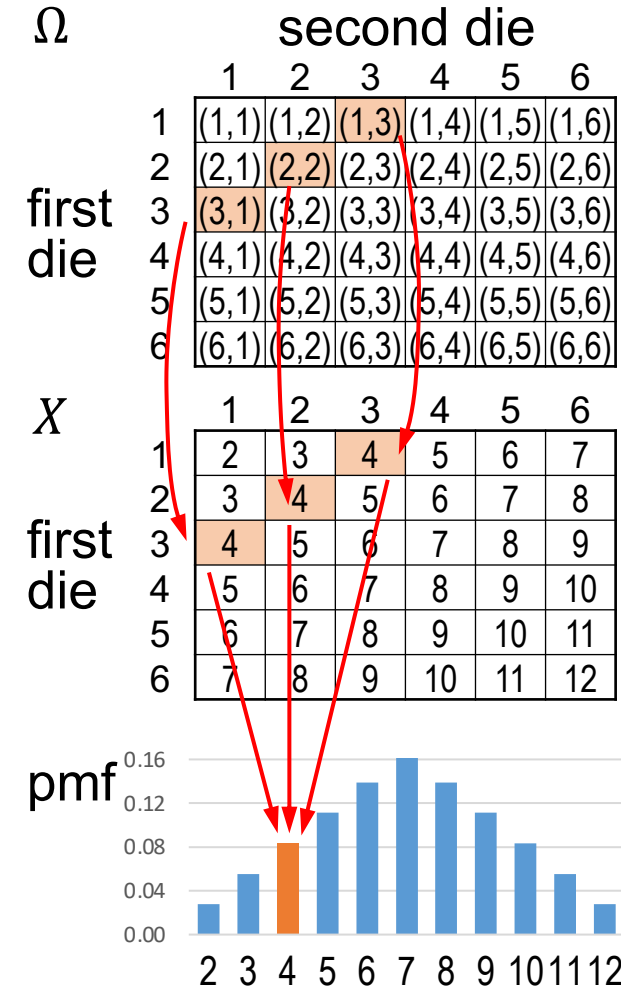
$$X = 2 \quad p_X(2) = \frac{1}{36}$$

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

...

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

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



Random variable: expectation $\mathbb{E}$

Random variable (RV)

$$X: \Omega \rightarrow \mathbb{R}$$

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

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

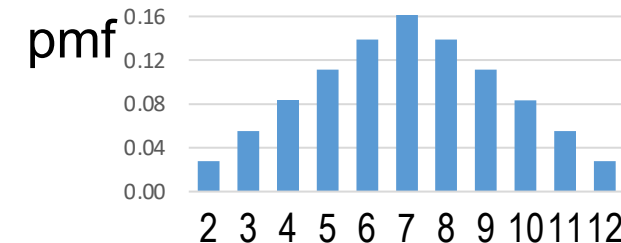
Let X be the sum of the rolls.



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

Variant 3: $\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = ?$
RVs for the outcome of the first and second die rolls

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

$$X: \Omega \rightarrow \mathbb{R}$$

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 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*

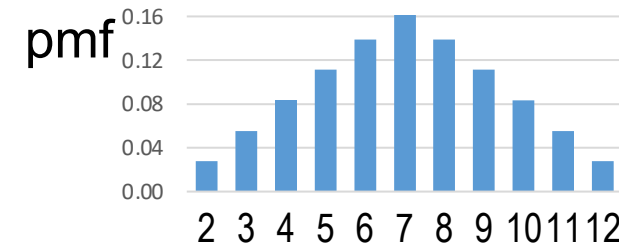
$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$$

linearity of expectation!

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

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

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

$X: \Omega \rightarrow \mathbb{R}$

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 $\mathbb{E}[Y]$?

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

$X: \Omega \rightarrow \mathbb{R}$

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 $\mathbb{E}[Y]$?

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

?



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

$$X: \Omega \rightarrow \mathbb{R}$$

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 $X_1 \perp X_2$

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

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

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

$X: \Omega \rightarrow \mathbb{R}$

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.

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)	(1,3)	(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)

$X: \Omega \rightarrow \mathbb{R}$

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.

Let X be the sum of the rolls.

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

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



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

$X: \Omega \rightarrow \mathbb{R}$

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.

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)	(1,3)	(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)

$$X: \Omega \rightarrow \mathbb{R}$$

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.

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 😊

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



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

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



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

holds even if X and Y are not independent

PROOF

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

?

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

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

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

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

holds even if X and Y are not independent

PROOF

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

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

PROOF

$$\begin{aligned}\mathbb{E}[X \cdot Y] &= \sum_x \sum_y x \cdot y \cdot p(x, y) \\ &= \sum_x \sum_y x \cdot y \cdot p(x) \cdot p(y) \\ &= \left(\sum_x x \cdot p(x) \right) \cdot \left(\sum_y y \cdot p(y) \right) \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y]\end{aligned}$$

because $X \perp Y$

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

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 $\overline{AB} \cdot \overline{CD}$?

?

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 $\mathbb{E}[\overline{AB}] \cdot \mathbb{E}[\overline{CD}]$?



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).

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?



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,

?

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,

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

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 .

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 .

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

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 .

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

Variance

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?* ?

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?

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$



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

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10

Measuring variability

EXAMPLE: average size of mice

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

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$

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

$$Y_1 = X - \mathbb{E}[X]$$

How can we fix that ?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ = 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

Measuring variability

EXAMPLE: average size of mice

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

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$

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

$$Y_1 = X - \mathbb{E}[X]$$

$$Y_2 = |X - \mathbb{E}[X]|$$

absolute deviation

Anything else ?

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹️ <i>cancel out</i>
$\mathbb{E}[Y_2] =$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$ -2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$ 😊 <i>makes more sense</i>

Measuring variability

EXAMPLE: average size of mice

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

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$

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

$$Y_1 = X - \mathbb{E}[X]$$

$$Y_2 = |X - \mathbb{E}[X]|$$

absolute deviation

$$Y_3 = (X - \mathbb{E}[X])^2$$

squared error

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$
$\mathbb{E}[Y_2] =$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$ -2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$
$\mathbb{E}[Y_3] =$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$	$2^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3}$ $4 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3}$ $= \frac{8}{3}$

☹️ cancels out

😊 makes more sense

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?

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$

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

$$Y_1 = X - \mathbb{E}[X]$$

$$Y_2 = |X - \mathbb{E}[X]|$$

absolute deviation

$$Y_3 = (X - \mathbb{E}[X])^2$$

squared error

	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹️ <i>cancels out</i>
$\mathbb{E}[Y_2] =$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ cm	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ cm	$ -2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$ cm 😊 <i>makes more sense</i>
$\mathbb{E}[Y_3] =$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ 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 ☹️ ???

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?

$$\mathbb{E}[Y] = \sum_y y \cdot p_Y(y)$$

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

$$Y_1 = X - \mathbb{E}[X]$$

$$Y_2 = |X - \mathbb{E}[X]|$$

absolute deviation

$$Y_3 = (X - \mathbb{E}[X])^2$$

squared error

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

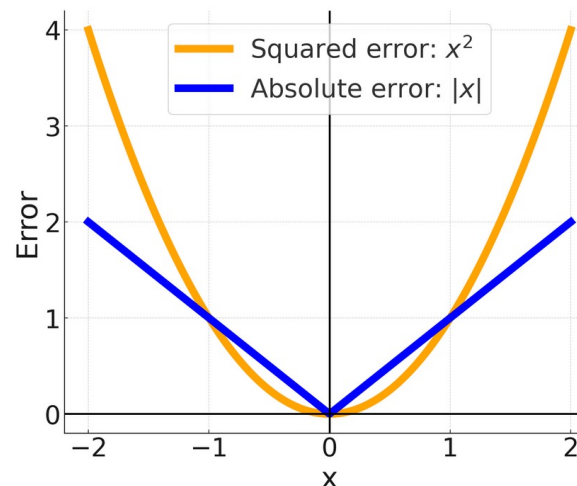
	case 1: {4 cm, 5 cm, 6 cm}	case 2: {9 cm, 10 cm, 11 cm}	case 3: {8 cm, 10 cm, 12 cm}
$\mathbb{E}[X] =$	5	10	10
$\mathbb{E}[Y_1] =$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 0$	$-2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 0$ ☹️ <i>cancels out</i>
$\mathbb{E}[Y_2] =$	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ cm	$ -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ cm	$ -2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= 2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$ $= \frac{4}{3}$ cm 😊 <i>makes more sense</i>
$\mathbb{E}[Y_3] =$	$1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3}$ $1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$ $= \frac{2}{3}$ 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 ☹️ ???
Just take the square root:	$= \sqrt{\frac{2}{3}}$ cm	$= \sqrt{\frac{2}{3}}$ cm	$= 2\sqrt{\frac{2}{3}}$ cm ???

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



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

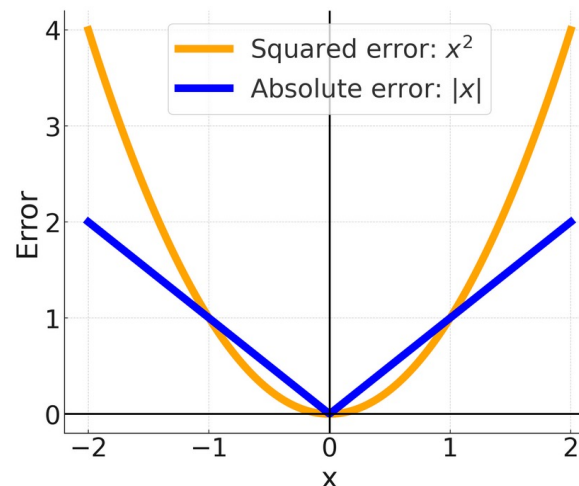
absolute deviation

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

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

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

EXAMPLE (MEAN VS. MEDIAN):

data = {1, 2, 3, 6, 8}

mean = ?

median = ?

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

EXAMPLE (MEAN VS. MEDIAN):

data = {1, 2, 3, 6, 8}

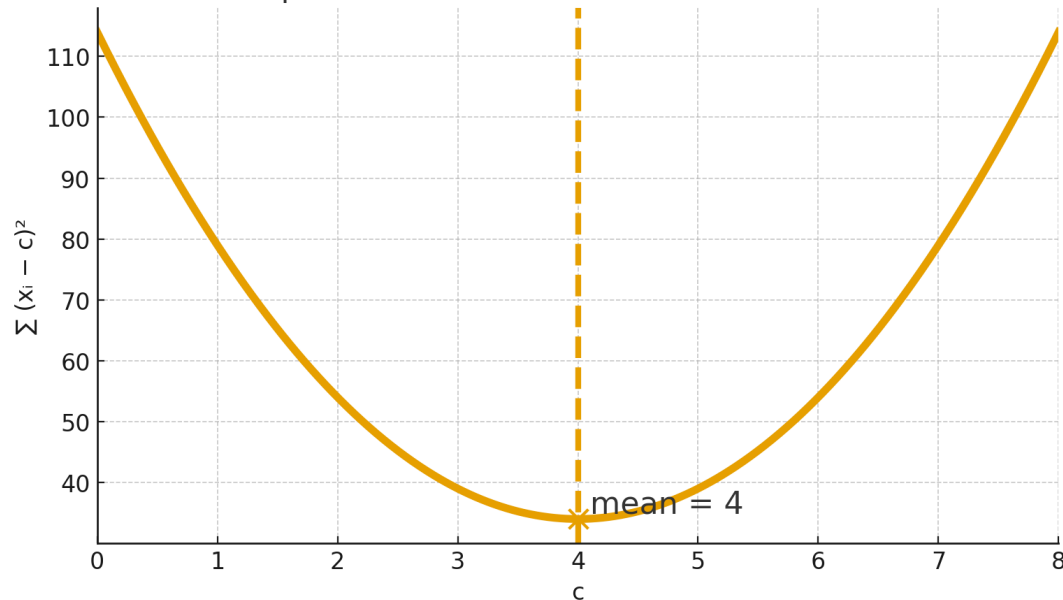
mean = 4

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

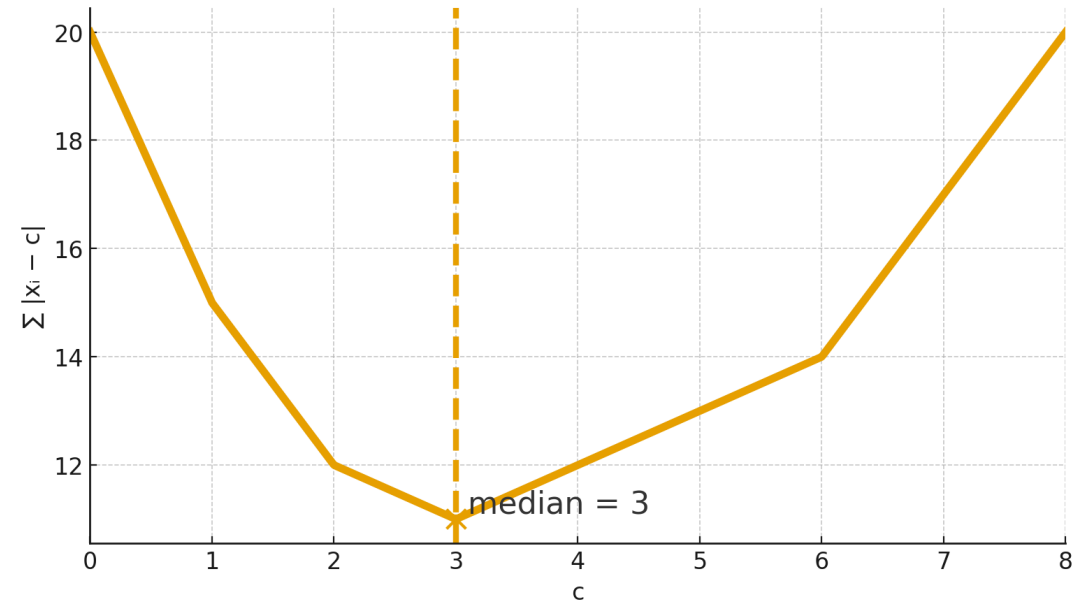
median = 3

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

Squared-error loss minimized at the mean



Absolute-error loss minimized at the median



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

EXAMPLE (MEAN VS. MEDIAN):

data = {1, 2, 3, 6, 8}

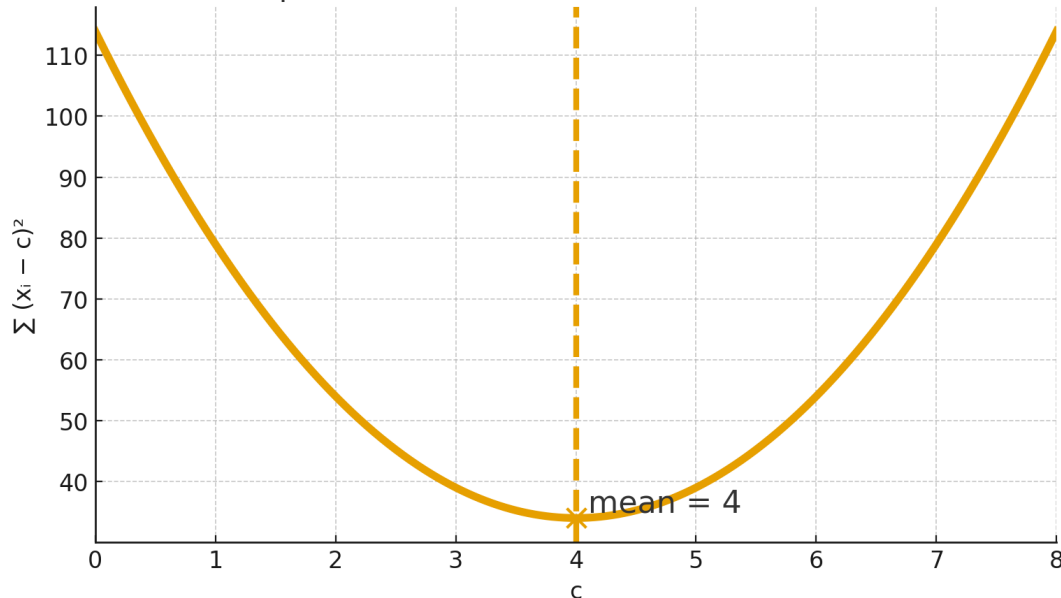
mean = 4

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

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

?

Squared-error loss minimized at the mean



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

EXAMPLE (MEAN VS. MEDIAN):

data = {1, 2, 3, 6, 8}

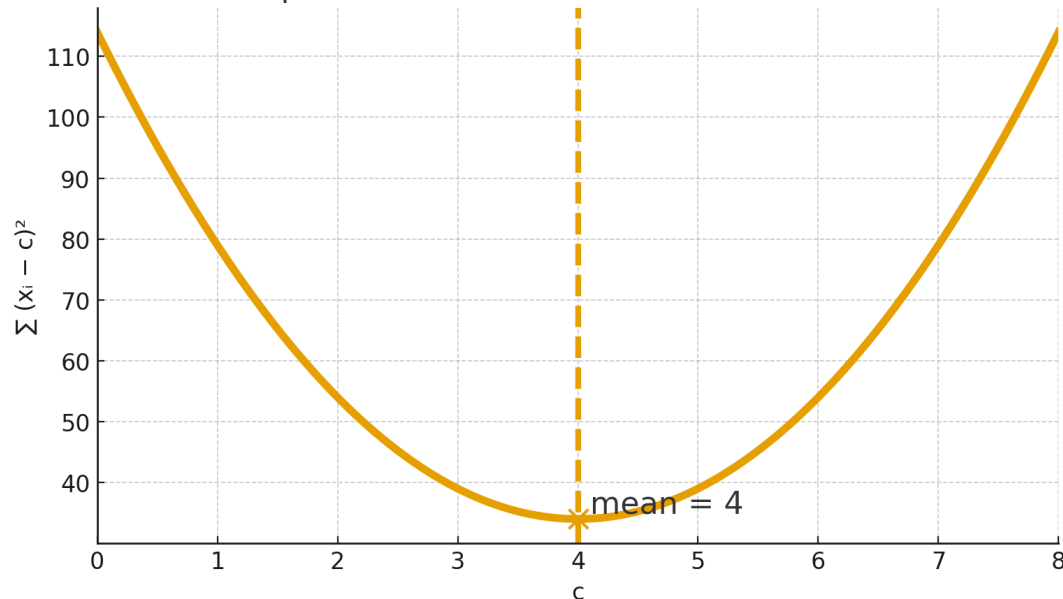
mean = 4

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

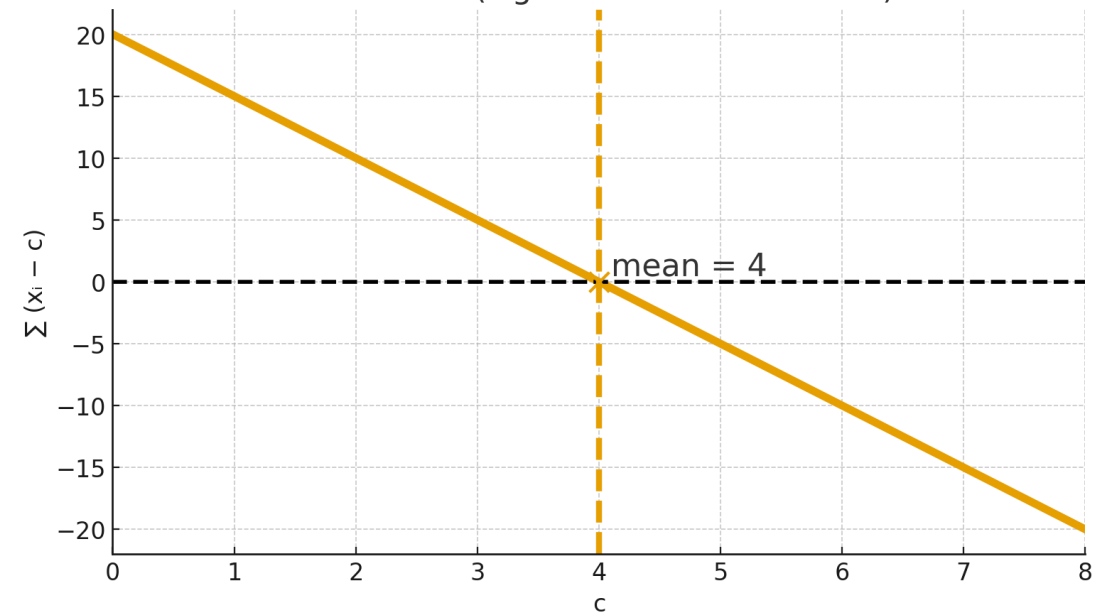
This is not an optimization problem, but a constraint problem: the sum of signed errors = 0.

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

Squared-error loss minimized at the mean



Total error (signed sum of deviations)



Alternative formula for variance

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

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

ALTERNATIVE FORMULA

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

PROOF

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

case 3:

{8 cm, 10 cm, 12 cm}

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

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

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

Alternative formula for variance

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

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

ALTERNATIVE FORMULA

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

PROOF

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2 \cdot X \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2] \\ &\quad \text{by linearity of expectation} \\ &= \mathbb{E}[X^2] - \mathbb{E}[2 \cdot X \cdot \mathbb{E}[X]] + \mathbb{E}[(\mathbb{E}[X])^2] \\ &\quad \text{just a constant} \\ &= \mathbb{E}[X^2] - 2 \cdot \mathbb{E}[X] \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - 2 \cdot (\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

case 3:

{8 cm, 10 cm, 12 cm}

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

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

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

Markov chains (later)