CSE 312

Foundations of Computing II

Lecture 5: Conditional Probability and Bayes Theorem

Review Probability

Definition. A **sample space** Ω is the set of all possible outcomes of an experiment.

Examples:

- Single coin flip: $\Omega = \{H, T\}$
- Two coin flips: $\Omega = \{HH, HT, TH, TT\}$
- Roll of a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Definition. An **event** $E \subseteq \Omega$ is a subset of possible outcomes.

Examples:

- Getting at least one head in two coin flips: $E = \{HH, HT, TH\}$
- Rolling an even number on a die : $E = \{2, 4, 6\}$

Review Probability space

Either finite or infinite countable (e.g., integers)

Definition. A (discrete) **probability space** is a pair (Ω, P) where:

- Ω is a set called the **sample space**.
- P is the **probability measure,** a function $P: \Omega \to \mathbb{R}$ such that:
  - $P(x) \geq 0$ for all $x \in \Omega$
  - $\sum_{x \in \Omega} P(x) = 1$

Set of possible **outcomess**

$$A \subseteq \Omega:\ P(A) = \sum_{x \in A} P(x)$$

Specify Likelihood (or probability) of each **outcome**

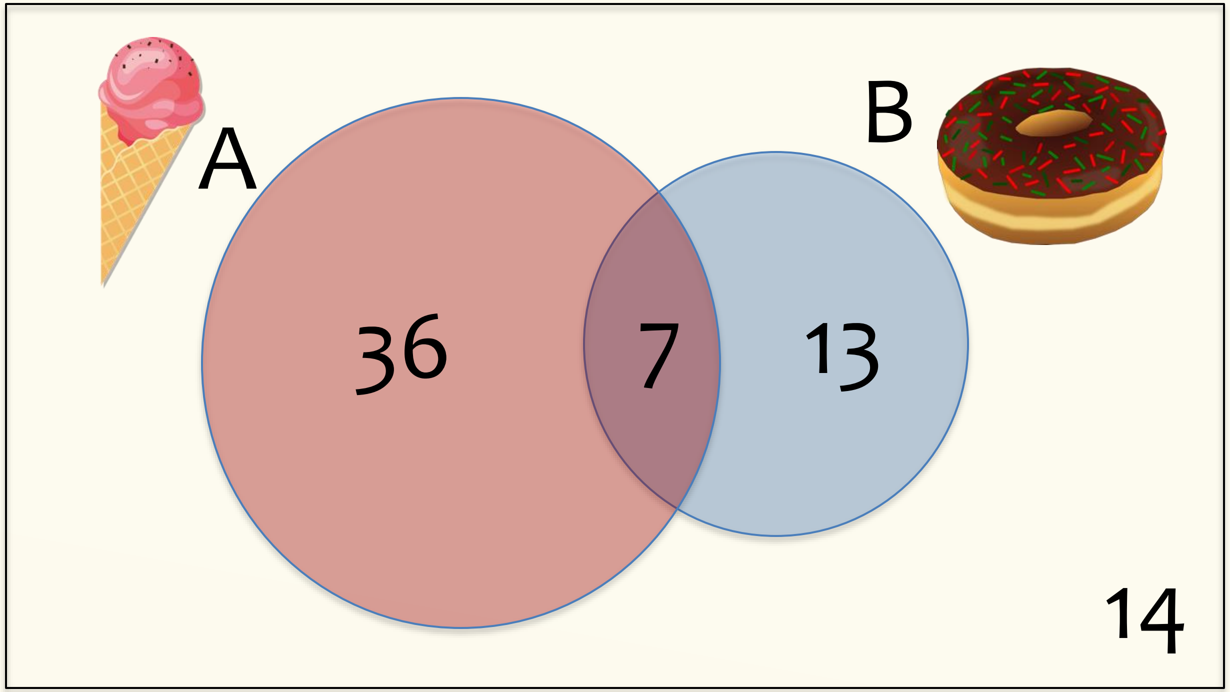
Some outcome must show up

The likelihood (or probability) of each outcome is non-negative.

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- More Examples

Conditional Probability (Idea)



What's the probability that a randomly chosen person (everyone equally likely) likes ice cream **given** they like donuts?

$$\frac{7}{7+13} = \frac{7}{20}$$

Conditional Probability

Definition. The **conditional probability** of event A **given** an event B happened (assuming $P(B) \neq 0$) is

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

A useful formula is

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

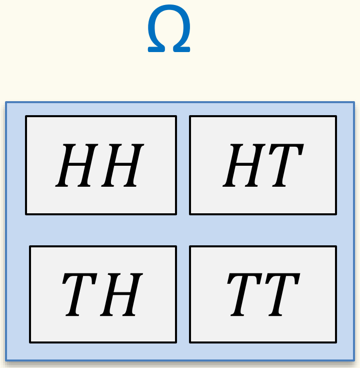
Conditional Probability Examples

Suppose that you flip a fair coin twice.

What is the probability that both flips are heads given that you have at least one head?

Let O be the event that at least *one* flip is heads

Let B be the event that *both* flips are heads



$P(O) = 3/4 \qquad P(B) = 1/4 \qquad P(B \cap O) = 1/4$

$$P(B|O) = \frac{P(B \cap O)}{P(O)} = \frac{1/4}{3/4} = \frac{1}{3}$$

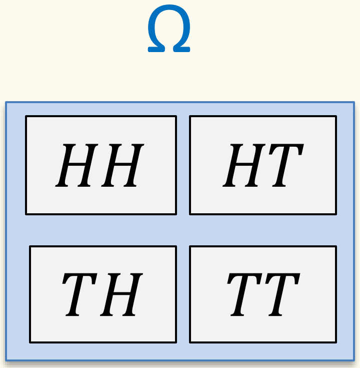
Conditional Probability Examples

Suppose that you flip a fair coin twice.

What is the probability that at least one flip is heads given that at least one flip is tails?

Let H be the event that at least one flip is *heads*

Let T be the event that at least one flip is *tails*



$P(H) = 3/4$ $\quad P(T) = 3/4$ $\quad P(H \cap T) = 1/2$

$$P(H|T) = \frac{P(H \cap T)}{P(T)} = \frac{1/2}{3/4} = \frac{2}{3}$$

Reversing Conditional Probability

Question: Does $P(A|B) = P(B|A)$?

No!

- Let A be the event you are wet
- Let B be the event you are swimming

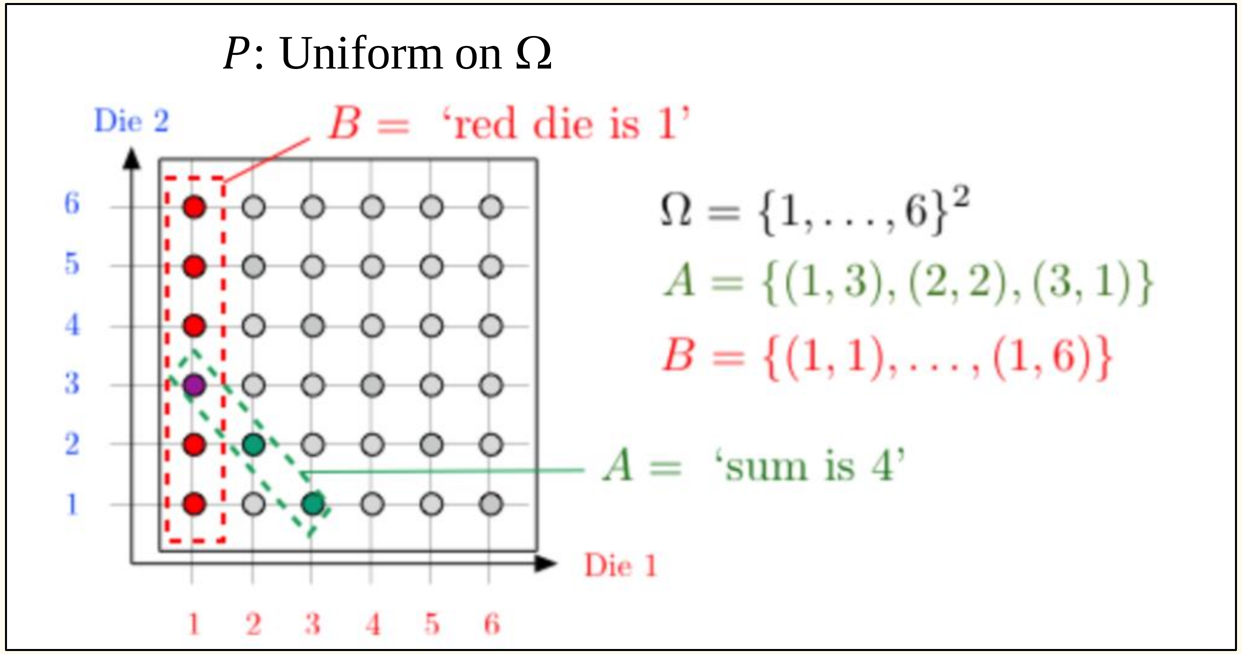
$$P(A|B) = 1$$
$$P(B|A) \neq 1$$

Example with Conditional Probability

pollev.com/stefanotessaro617

Suppose we toss a red die and a blue die: both 6 sided and all outcomes equally likely.

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



	$P(B)$	$P(B\|A)$
a)	1/6	1/6
b)	1/6	1/3
c)	1/6	3/36
d)	1/9	1/3

Gambler's fallacy

Assume we toss **51** fair coins.
Assume we have seen **50** coins, and they are all "tails".
What are the odds the **51st** coin is "heads"?

A = first 50 coins are "tails"

B = first 50 coins are "tails", 51st coin is "heads"

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/2^{51}}{2/2^{51}} = \frac{1}{2}$$

51st coin is independent of outcomes of first 50 tosses!

Gambler's fallacy = Feels like it's time for "heads"!?

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- More Examples

Bayes Theorem

A formula to let us "reverse" the conditional.

Theorem. (Bayes Rule) For events A and B, where $P(A), P(B) > 0$,

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

$P(A)$ is called the **prior** (our belief without knowing anything)
$P(A|B)$ is called the **posterior** (our belief after learning B)

Bayes Theorem Proof

Claim:

$$P(A), P(B) > 0 \quad \Rightarrow \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Theorem Proof

Claim:

$$P(A), P(B) > 0 \quad \Rightarrow \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

By definition of conditional probability

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

Swapping A, B gives

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

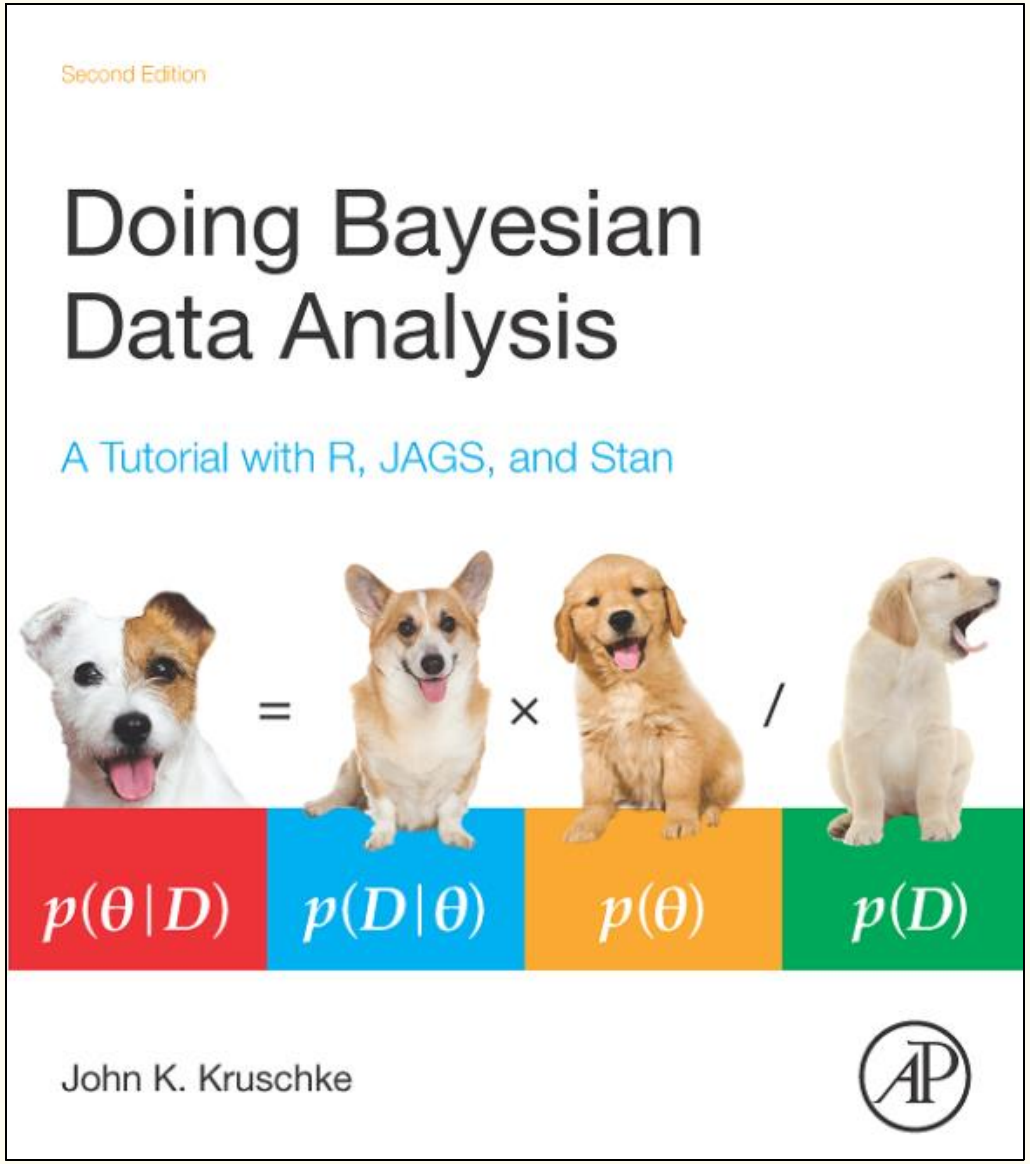
But $P(A \cap B) = P(B \cap A)$, so

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

Dividing both sides by $P(B)$ gives

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

Brain Break



Second Edition
Doing Bayesian Data Analysis
A Tutorial with R, JAGS, and Stan
=
×
/
$p(\theta|D)$
$p(D|\theta)$
$p(\theta)$
$p(D)$
John K. Kruschke

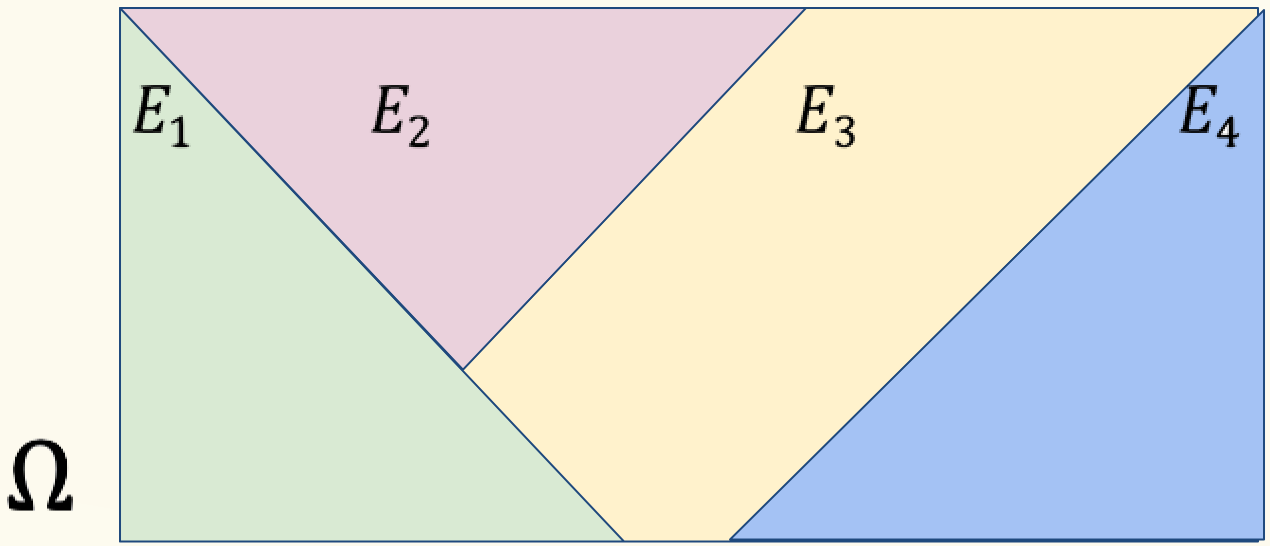
Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- More Examples

Partitions (Idea)

These events **partition** the sample space

1. They "cover" the whole space
2. They don't overlap, i.e., they are mutually exclusive



Partition

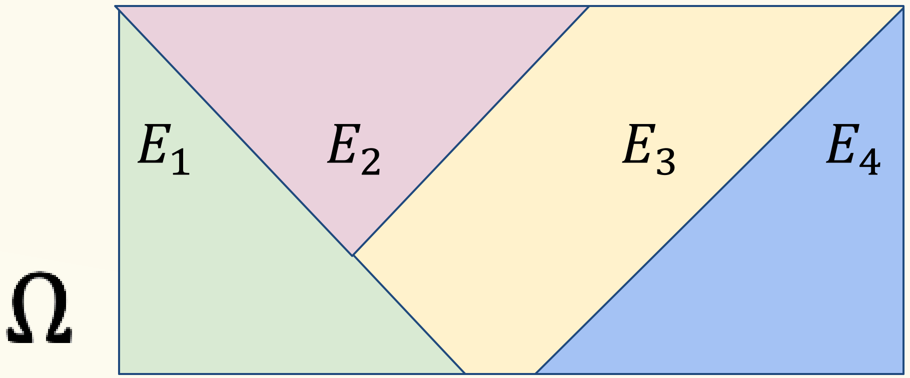
Definition. Non-empty events $E_1, E_2, \ldots, E_n$ **partition** sample space Ω if

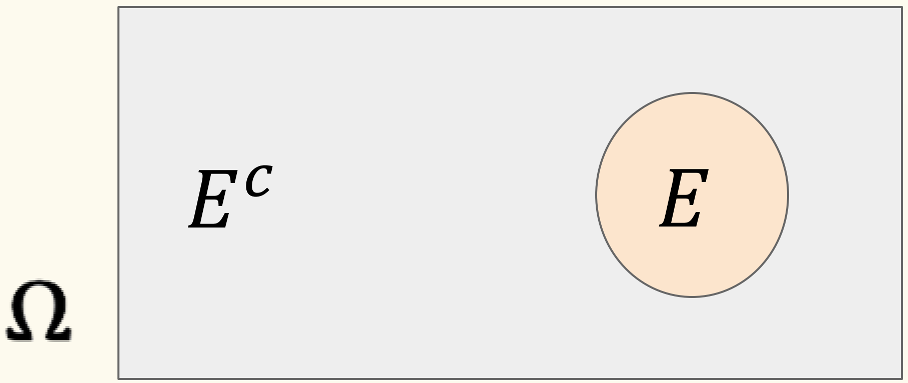
(Exhaustive)

$$E_1 \cup E_2 \cup \cdots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$$

(Pairwise Mutually Exclusive)

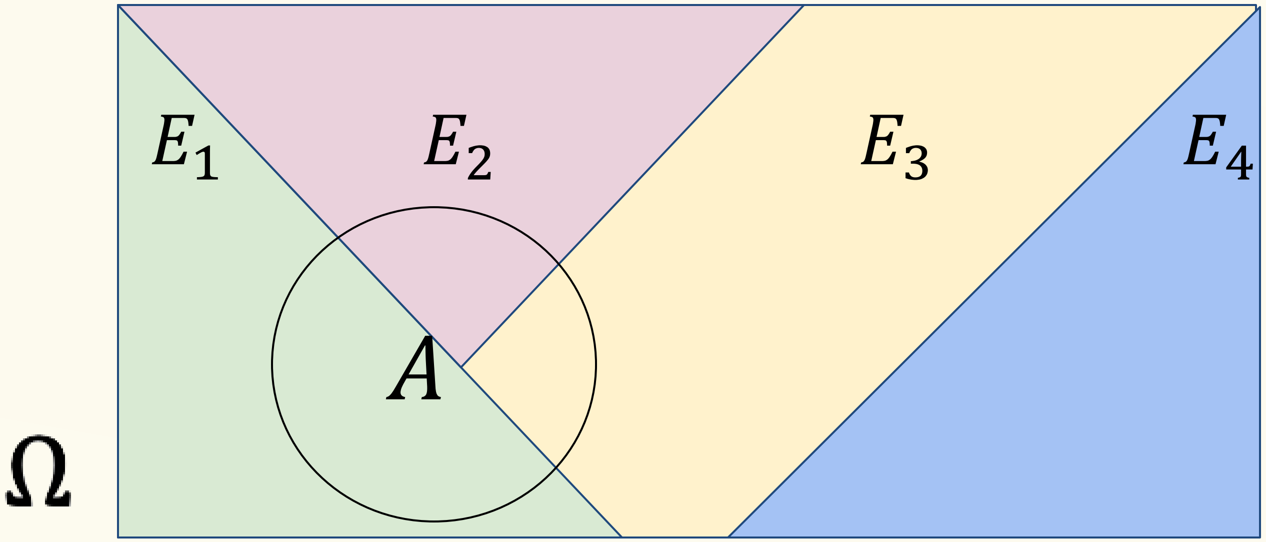
$$\forall_i \forall_{i \neq j} \;\; E_i \cap E_j = \emptyset$$





Law of Total Probability (Idea)

If we know $E_1, E_2, \ldots, E_n$ partition Ω, what can we say about $P(A)$?



Law of Total Probability (LTP)

Theorem. If events $E_1, E_2, \ldots, E_n$ partition the sample space Ω, then for any event $A \subseteq \Omega$

$$P(A) = P(A \cap E_1) + \ \ldots + P(A \cap E_n) = \sum_{i=1}^{n} P(A \cap E_i)$$

Using the definition of conditional probability $P(A \cap E) = P(A|E)P(E)$ we can get the alternate form of this that shows

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$$

Why do we care? Often easier to compute $P(A)$ this way!

LTP Example

Alice has two pockets:

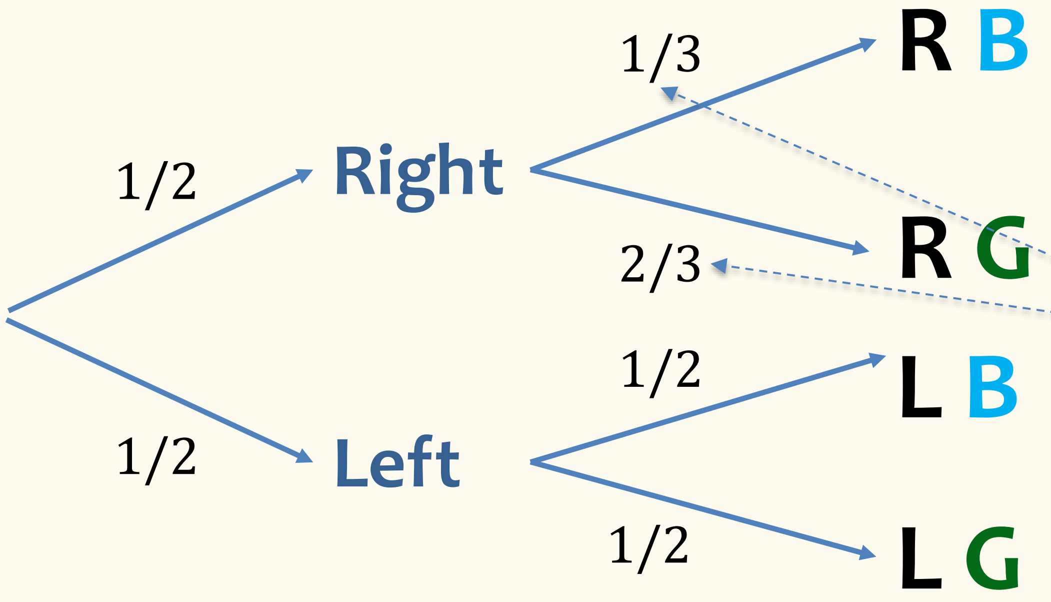
- **Left pocket:** Two blue balls, two green balls
- **Right pocket:** One blue ball, two green balls.

Alice picks a random ball from a random pocket.

[Both pockets equally likely, each ball equally likely.]

LTP Example

Left pocket: 2 blue, 2 green
Right pocket: 1 blue, 2 green



$1/3 = P(\mathbf{B}|\mathbf{Right})$ and $2/3 = P(\mathbf{G}|\mathbf{Right})$

$$P(\mathbf{B}) = P(\mathbf{B} \cap \mathbf{Left}) + P(\mathbf{B} \cap \mathbf{Right})$$ **(Law of total probability)**

$$= P(\mathbf{Left}) \times P(\mathbf{B}|\mathbf{Left}) + P(\mathbf{Right}) \times P(\mathbf{B}|\mathbf{Right})$$

$$= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

Agenda

- Conditional Probability
- Bayes Theorem
- Law of Total Probability
- More Examples

Example – Zika Testing

Usually no or mild symptoms (rash); sometimes severe symptoms (paralysis).

During pregnancy: may cause birth defects.

Suppose you took a Zika test, and it returns "positive", what is the likelihood that you actually have the disease?

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive") $P(T|Z)$
- However, the test may yield a "false positive" 1% of the time $P(T|Z^c)$
- 0.5% of the US population has Zika. $P(Z)$

What is the probability you have Zika (event Z) if you test positive (event T)?

Bayes Theorem $$P(Z|T) = \frac{P(Z) \cdot P(T|Z)}{P(T)} = \frac{0.005 \cdot 0.98}{0.01485} \approx 0.33$$

LTP $$P(T) = P(Z) \cdot P(T|Z) + P(Z^c)P(T|Z^c) = 0.005 \cdot 0.98 + 0.995 \cdot 0.01 = 0.01485$$

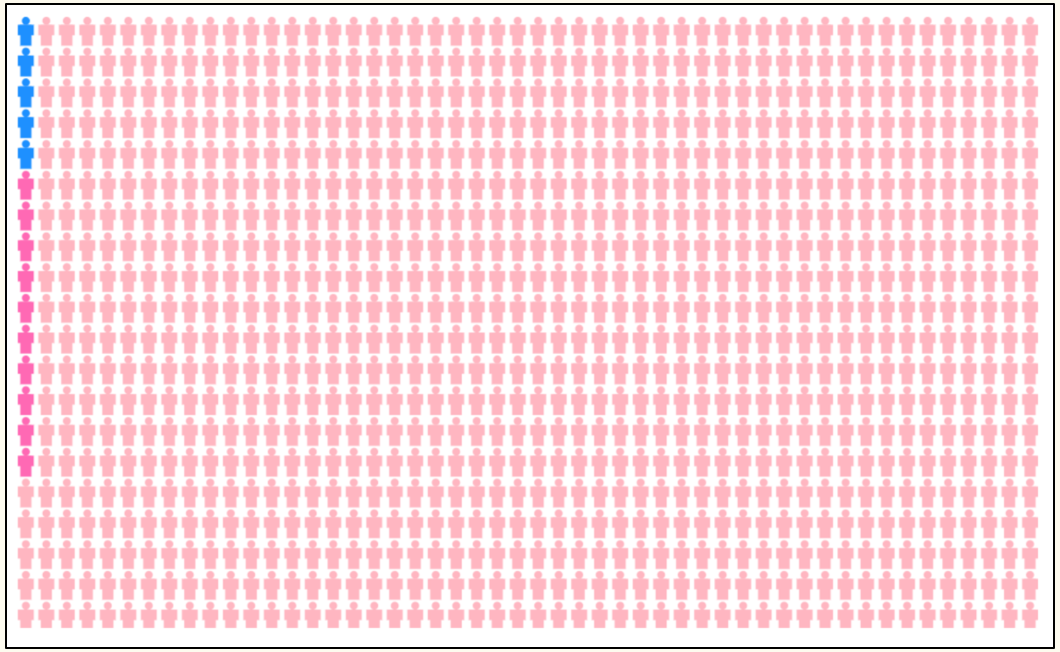
Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika ("true positive") $P(T|Z)$
- However, the test may yield a "false positive" 1% of the time $P(T|Z^c)$
- 0.5% of the US population has Zika. $P(Z)$

500 have Zika (0.5%)
99,500 do not

What is the probability you have Zika (event Z) if you test positive (event T)?



Suppose we had 100,000 people:

- 490 **have Zika** and **test positive**
- 10 **have Zika** and **test negative**
- 995 **do not have Zika** and **test positive**
- 98,505 **do not have Zika** and **test negative**

98% of those with Zika

2% of those with Zika

1% of those without Zika

$$\frac{490}{490 + 995} \approx 0.33$$

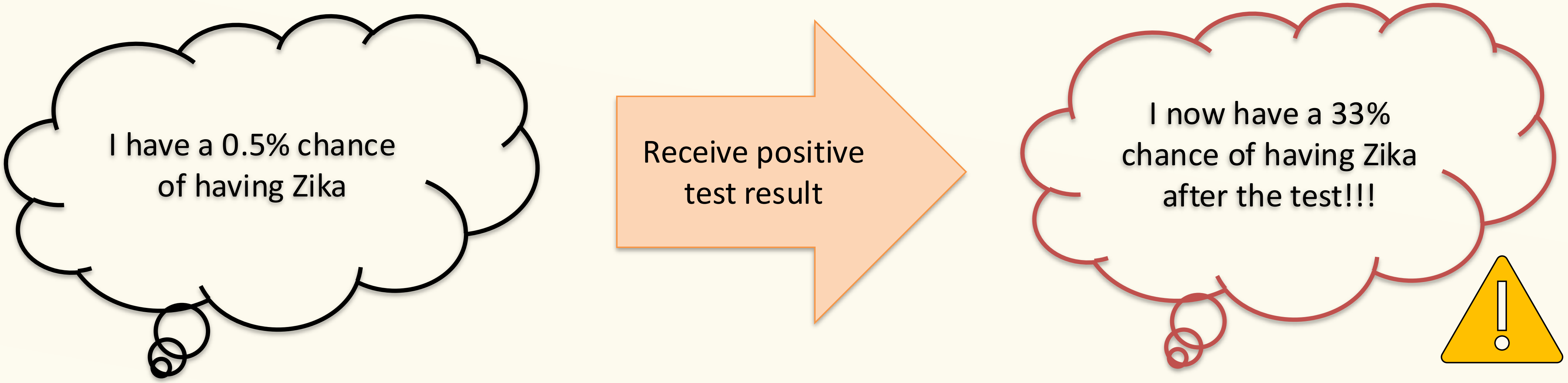
[Demo]

Philosophy – Updating Beliefs

While it's not 98% that you have the disease, your beliefs changed **drastically**

Z = you have Zika

T = you test positive for Zika



Prior: $P(Z)$

Posterior: $P(Z|T)$

Example – Zika Testing

Suppose we know the following Zika stats

- A test is 98% effective at detecting Zika (“true positive”) $P(T|Z)$
- However, the test may yield a “false positive” 1% of the time $P(T|Z^c)$
- 0.5% of the US population has Zika. $P(Z)$

What is the probability you test negative (event T^c) if you have Zika (event Z)?

$$P(T^c|Z) = 1 - P(T|Z) = 0.02$$

Conditional Probability Defines a Probability Space

The probability conditioned on A follows the same properties as (unconditional) probability.

Example. $P(B^c|A) = 1 - P(B|A)$

Formally. (Ω, P) is a probability space and $P(A) > 0$

$\Rightarrow$ $(A, P(\cdot\,|A))$ is a probability space