Conditional Probability
LTP & Independence

slides mostly by Alex Tsun

Agenda

- conditional probability
- Bayes Theorem
- Law of Total Probability (LTP)
- Bayes Theorem + LTP

Definitions

Sample Space: The set Ω of all possible outcomes of an experiment.

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

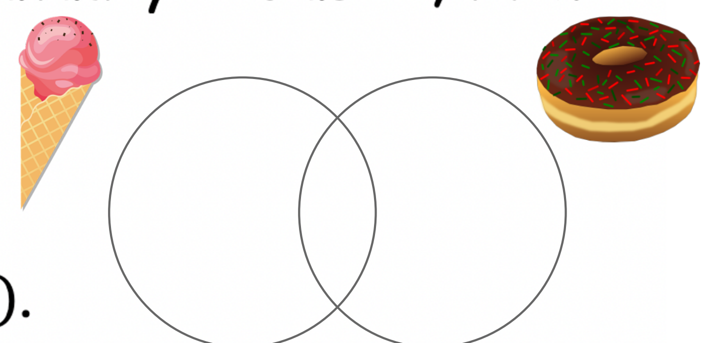
Event: Any subset $E \subseteq \Omega$.

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

Axioms of Probability & Their Consequences

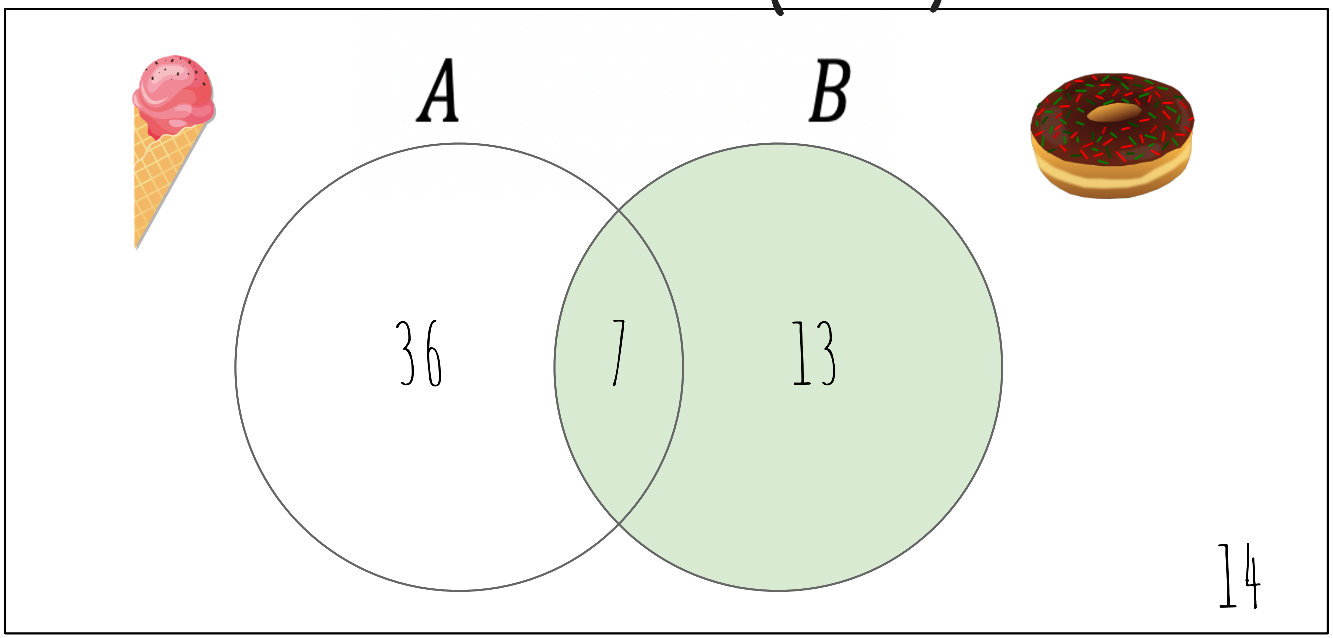
Let Ω denote the sample space and $E, F \subseteq \Omega$ be events.

Axiom 1 (Nonnegativity): $P(E) \geq 0$.
Axiom 2 (Normalization): $P(\Omega) = 1$.
Axiom 3 (Countable Additivity) If E and F are mutually exclusive, then $P(E \cup F) = P(E) + P(F)$.



Corollary 1 (Complementation): $P(E^C) = 1 - P(E)$.
Corollary 2 (Monotonicity): If $E \subseteq F$, $P(E) \leq P(F)$.
Corollary 3 (Inclusion-Exclusion): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Conditional Probability (idea)



What's the probability that someone likes ice cream **given** they like donuts?

$$P(A|B) = \frac{7}{20} = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability

Conditional Probability: The (conditional) probability of A given an event B happened is

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

An equivalent and useful formula is $P(A \cap B) = P(A|B)P(B)$.

Conditional Probability (Reversal)

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

Conditional Probability (Intuition)

Does $P(A|B) = P(B|A)$? No!!

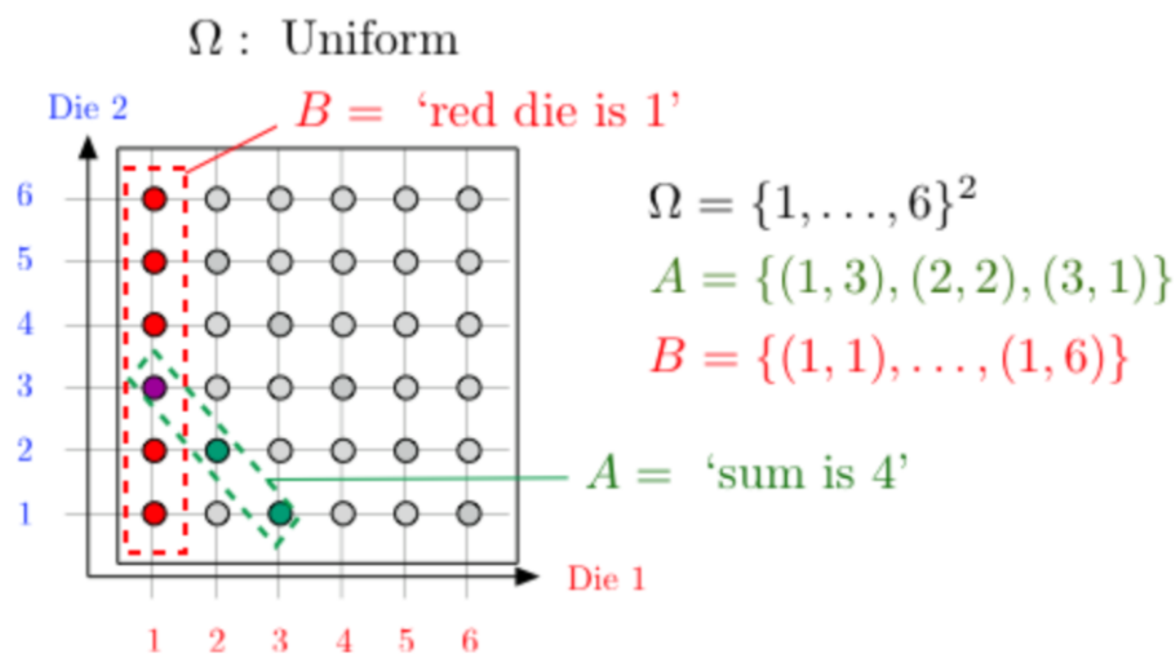
Let A be the event you are wet.
Let B be the event you are swimming.

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

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

Fun with conditional probability

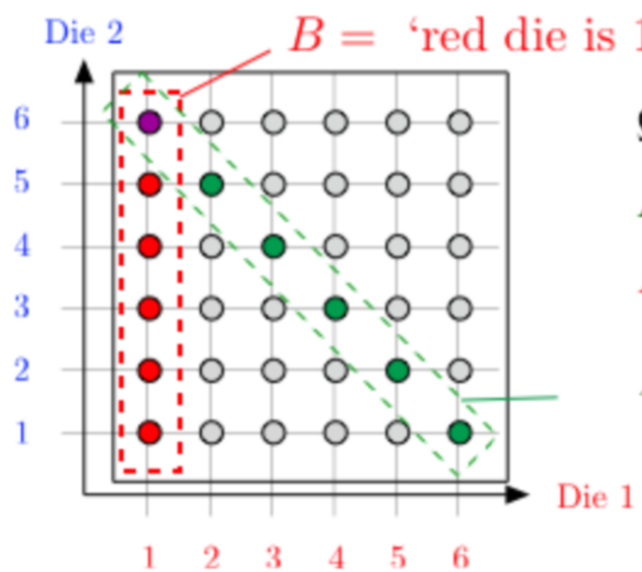
- Toss a red die and a blue die. All outcomes equally likely. What is Pr(B | A)? What is Pr(B)?



Fun with conditional probability

- Toss a red die and a blue die. All outcomes equally likely. What is Pr(B | A)?

Ω : Uniform



$\Omega = \{1, \ldots, 6\}^2$

$A = \{(1, 6), \ldots, (6,1)\}$

$B = \{(1, 1), \ldots, (1, 6)\}$

A = 'sum is 7'

Gambler's fallacy

- Flip a fair coin 51 times. All outcomes equally likely.
- A = "first 50 flips are heads"
- B = "the 51^{st} flip is heads"

- Pr (B | A) = ?

Bayes Theorem

Bayes Theorem: Let A, B be events with nonzero probability. Then,

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

Allows us to "reverse" the conditioning!

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

Bayes Theorem (Proof)

Bayes Theorem (Proof)

By definition of conditional probability,

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

Swapping A, B gives

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

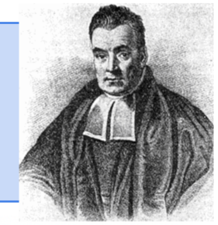
But $P(A \cap B) = P(B \cap A)$, so

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

Dividing both sides by $P(B)$ gives

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

Bayes Theorem



Bayes Theorem: Let A, B be events with nonzero probability. Then,

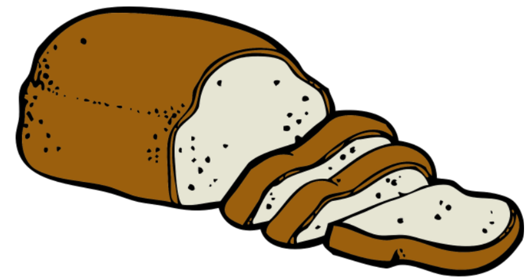
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Allows us to "reverse" the conditioning!

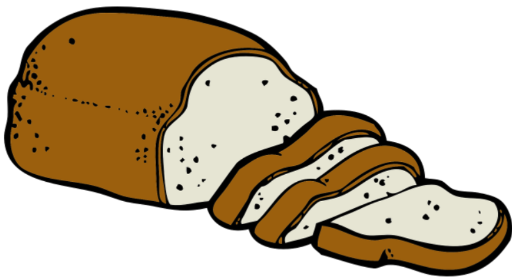
$P(A)$ is called the **prior** (our belief without knowing anything), and
$P(A|B)$ is called the **posterior** (our belief after learning B).

Random Picture

Cutting up a Sample Space

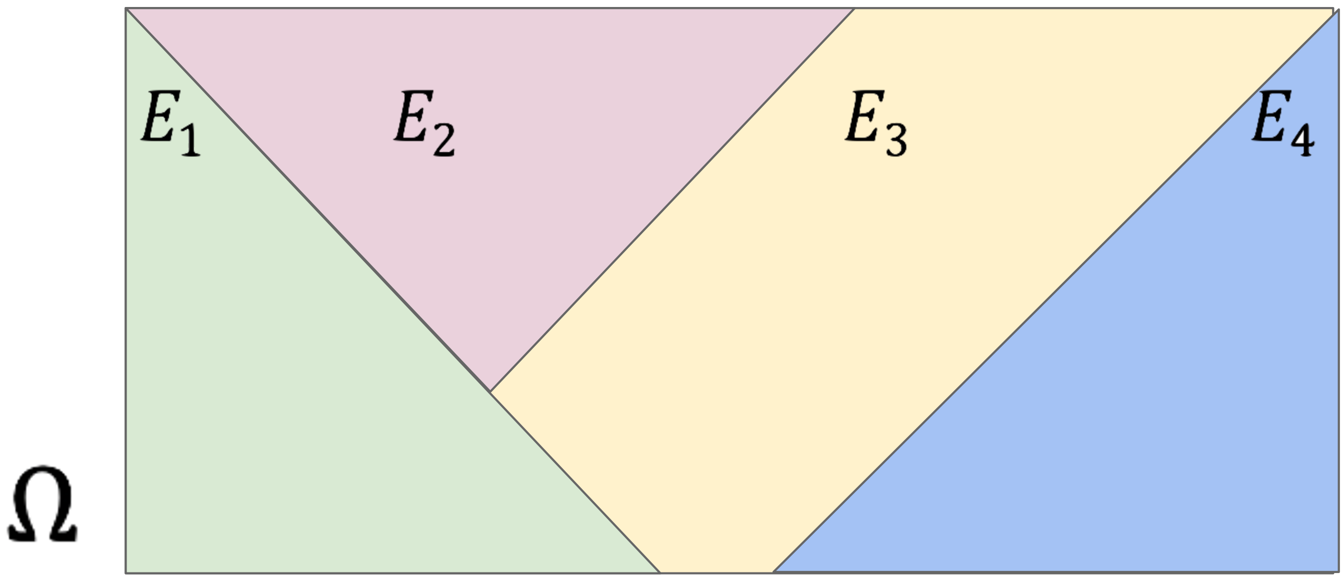


Cutting up a Sample Space



These events **partition** the sample space i.e.,

1. They "cover" the whole space.
2. They don't overlap.

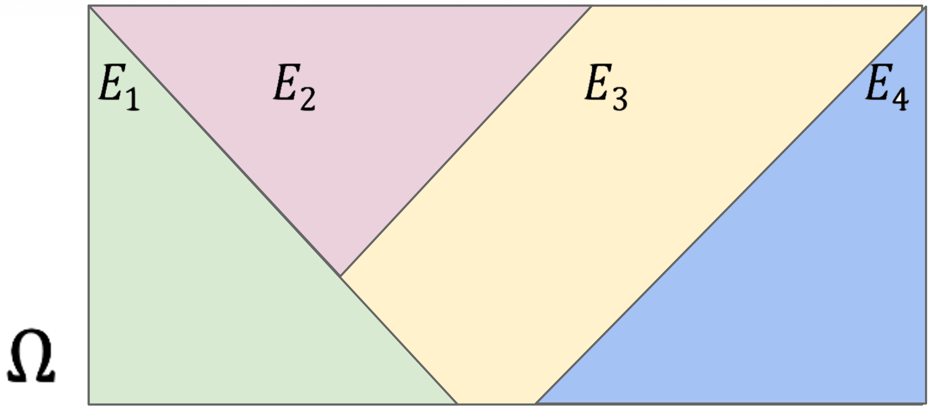


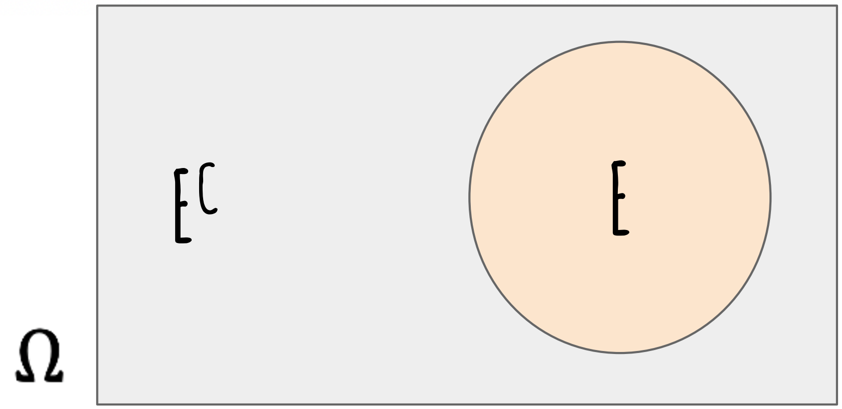
Partitions

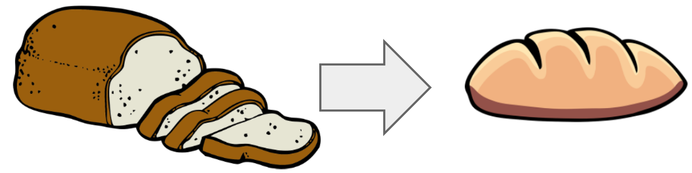
Partition: Non-empty events $E_1, \ldots, E_n$ partition the sample space Ω if

- (**Exhaustive**) $E_1 \cup E_2 \cup \ldots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$.
- (**Pairwise Mutually Exclusive**) For all $i \neq j$, $E_i \cap E_j = \emptyset$.

Notice for any event E: E and E^C always partition Ω.

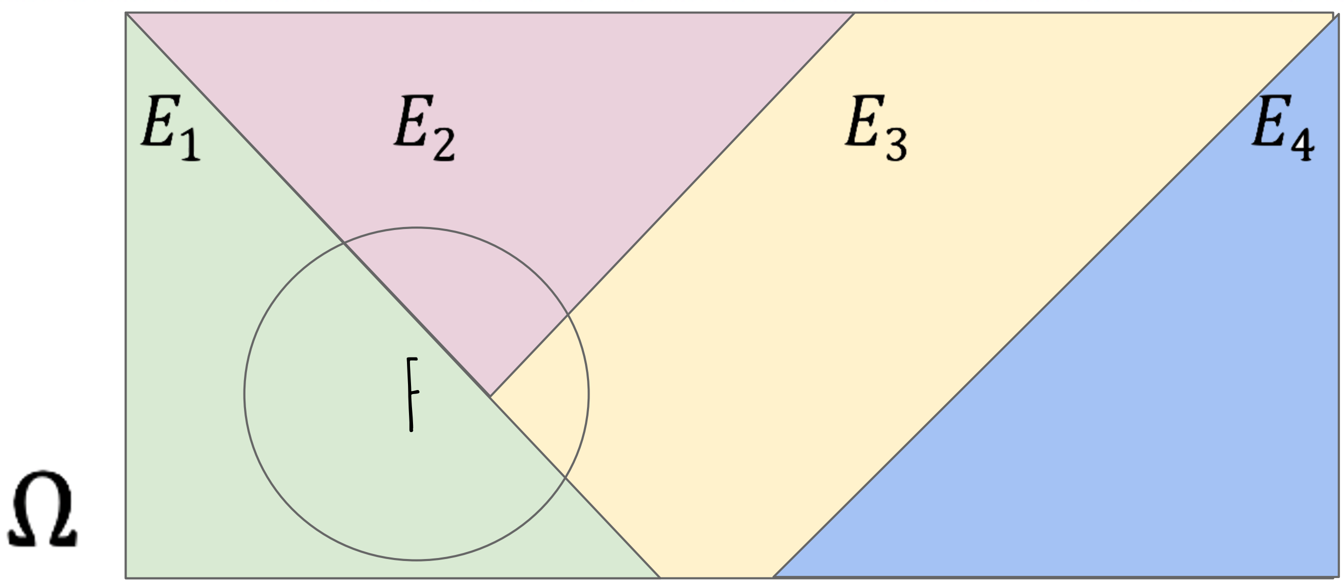




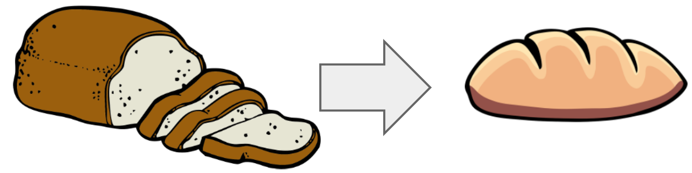
(The Picture) Law of total probability 

Back to the old picture. How can we decompose event **F**?

$P(F) =$

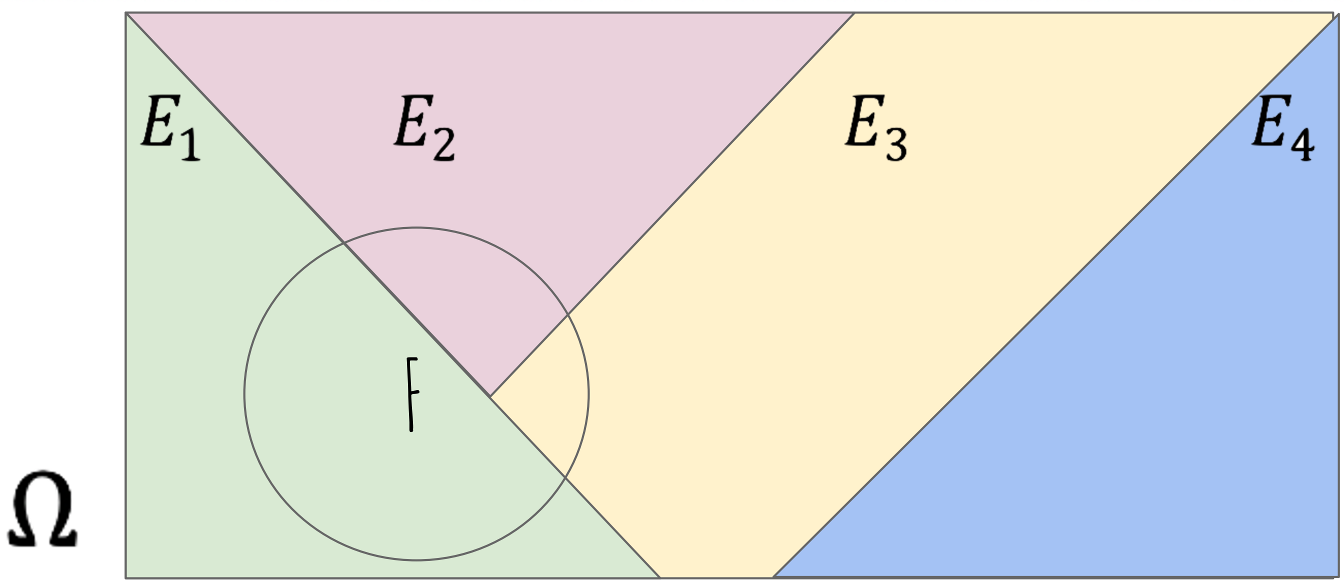


(The Picture) Law of total probability

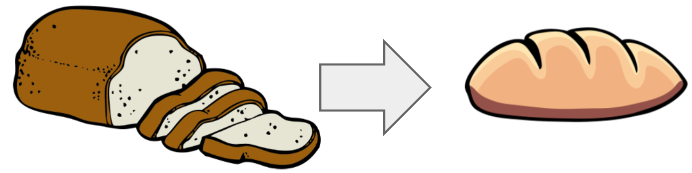


Back to the old picture. How can we decompose event **F**?

$$P(F) = P(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3)$$

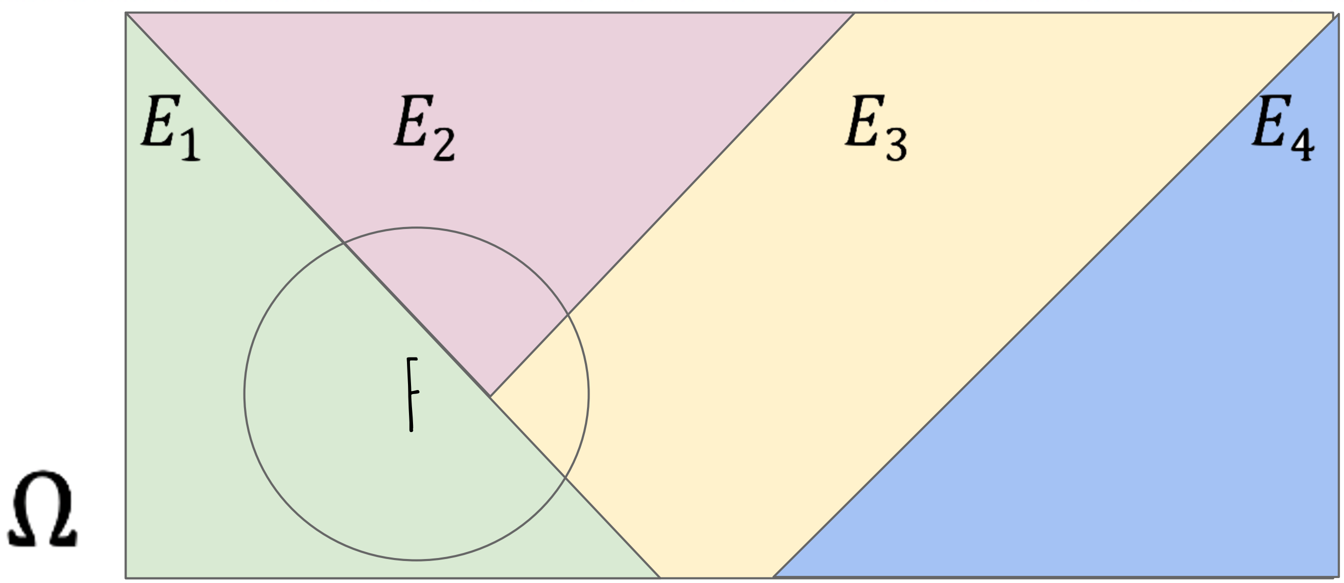


(The Picture) Law of total probability



Back to the old picture. How can we decompose event F?

$$P(F) = P(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3) + P(F \cap E_4)$$



For Completion

Law of Total Probability (LTP)

Law of Total Probability: If events $E_1, \dots, E_n$ partition Ω, then for any event F,

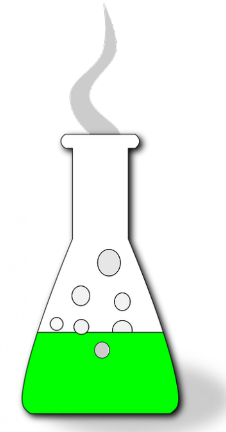
$$P(F) = P(F \cap E_1) + \cdots + P(F \cap E_n) = \sum_{i=1}^{n} P(F \cap E_i)$$

Using the definition of conditional probability ($P(F \cap E_i) = P(F|E_i)P(E_i)$), we get an alternate (more useful) form

$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$

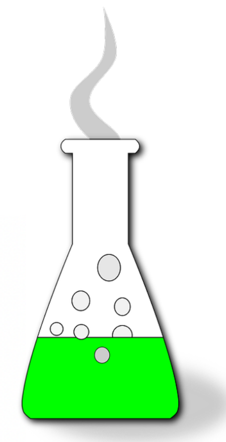
Intuition (LTP)

$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$



Intuition (LTP)

$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$



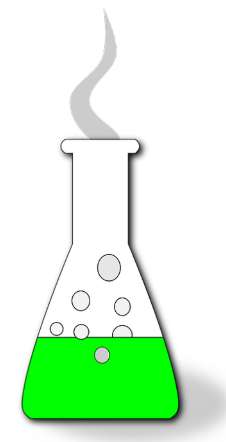
- You want to know the probability company you invested in FAILS to produce a successful vaccine
- You chose randomly which company to invest in

First, compute the probability of failure for each of 3 companies

Then, weight those by the probability of investing in that company

Example (LTP)

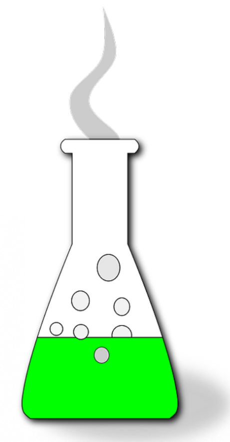
$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$



	AstraZeneca (E_1)	Merck (E_2)	Moderna (E_3)
Probability invest in this company	$6/8$	$1/8$	$1/8$
Probability this company's vaccine fails to work	1	0	$1/2$

Example (LTP)

$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$

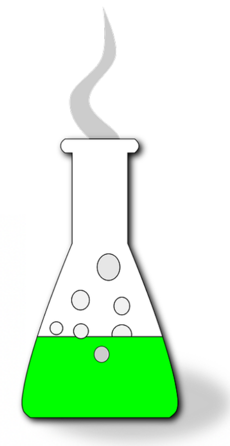


	AstraZeneca (E_1)	Merck (E_2)	Moderna (E_3)
Probability invest in this company	$6/8$	$1/8$	$1/8$
Probability this company's vaccine fails to work	1	0	$1/2$

$$P(F) = P(F|E_1)P(E_1) + P(F|E_2)P(E_2) + P(F|E_3)P(E_3) = 1 \cdot \frac{6}{8} + 0 \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{8} = \frac{13}{16}$$

Example (LTP)

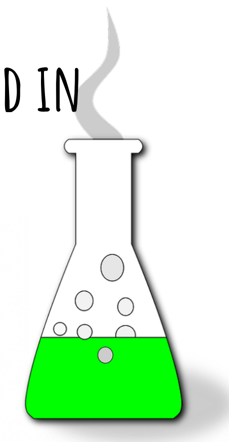
$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$



	AstraZeneca (E_1)	Merck (E_2)	Moderna (E_3)
Probability invest in this company	6/8	1/8	1/8
Probability this company's vaccine fails to work	1	0	1/2

What is the probability invested in Moderna given investment did not pay off (vaccine of company you invested in fails to work)?

What's the probability that you invested in Moderna, given that the company you invested in fails to produce successful vaccine? Need LTP for denominator...



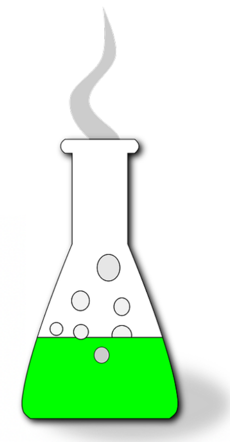
$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$

		Moderna (E_3)
Probability invest in this company	$P(E_3|F) =$	$1/8$
Probability this company's vaccine fails to work		$1/2$

$$P(F) = P(F|E_1)P(E_1) + P(F|E_2)P(E_2) + P(F|E_3)P(E_3) = 1 \cdot \frac{6}{8} + 0 \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{8} = \frac{13}{16}$$

Example (LTP)

$$P(F) = P(F|E_1)P(E_1) + \cdots + P(F|E_n)P(E_n) = \sum_{i=1}^{n} P(F|E_i)P(E_i)$$



		Moderna (E_3)
Probability invest in this company	$P(E_3|F) = \frac{P(F|E_3)P(E_3)}{P(F)} = \frac{\frac{1}{2}\cdot\frac{1}{8}}{\frac{13}{16}} = \frac{1}{13}$	$1/8$
Probability this company's vaccine fails to work		$1/2$

$$P(F) = P(F|E_1)P(E_1) + P(F|E_2)P(E_2) + P(F|E_3)P(E_3) = 1\cdot\frac{6}{8} + 0\cdot\frac{1}{8} + \frac{1}{2}\cdot\frac{1}{8} = \frac{13}{16}$$

Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, \ldots, E_n$ be a partition of the sample space, and F an event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{P(F|E_1)P(E_1)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

(Simple Partition) In particular, if E is an event with nonzero probability, then

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$

Another example

- 1% of people have a certain genetic disorder.
- 90% of time people with disorder test positive (true positives)
- 9.6% of the time, people that don’t have the disorder test negative (false positives).
- If a person gets a positive test result, what is the probability that they actually have the disorder?

G: have the disorder

P: positive test result

Another example

- 1% of people have a certain genetic disorder.
- 90% of time people with disorder test positive (true positives)
- 9.6% of the time, people that don’t have the disorder test negative (false positives).
- If a person gets a positive test result, what is the probability that they actually have the disorder?

G: have the disorder
P: positive test result

Probability

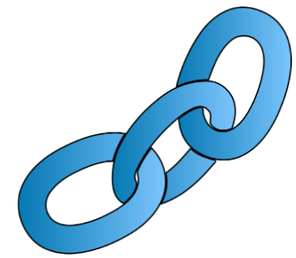
2.3 Independence

Most slides by Alex Tsun

Agenda

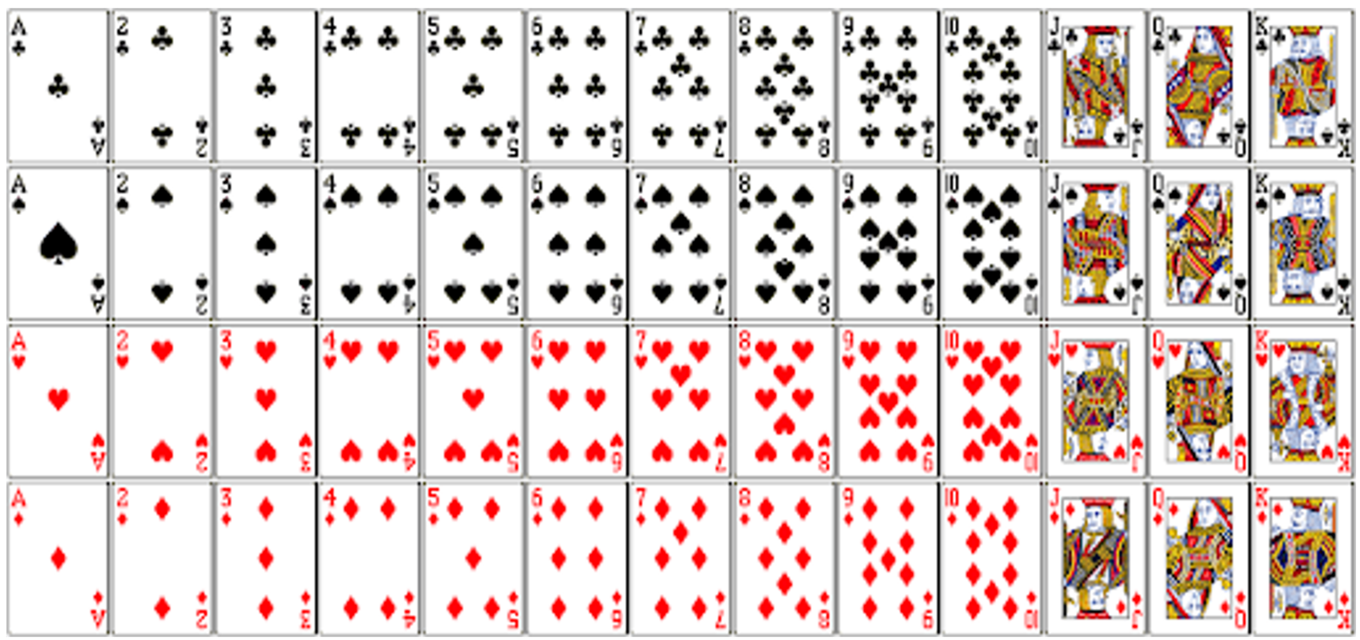
- Chain Rule
- Independence
- Conditional Independence

Chain Rule (Idea)

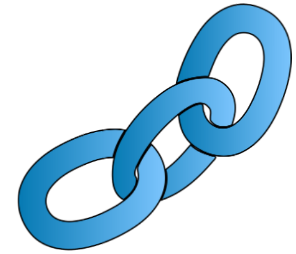


Have a Standard 52-Card Deck.

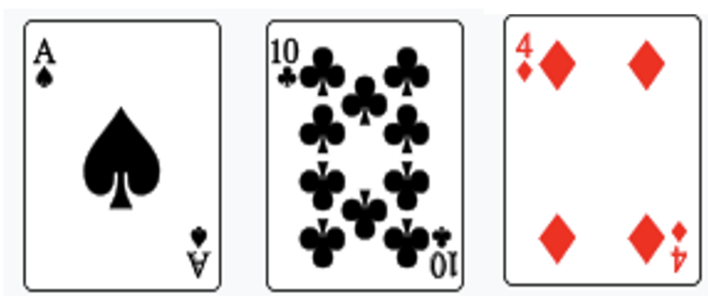
- 4 Suits (Clubs, Diamonds, Hearts, Spades)
- 13 ranks (A, 2, 3, ..., 9, 10, J, Q, K)



Chain Rule (Idea)



Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards.

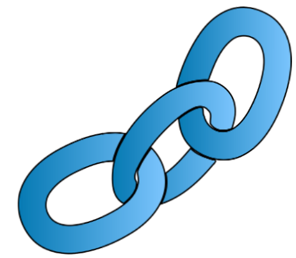
What is P () = P(**A**, **B**, **C**)?

A: Ace of Spades First

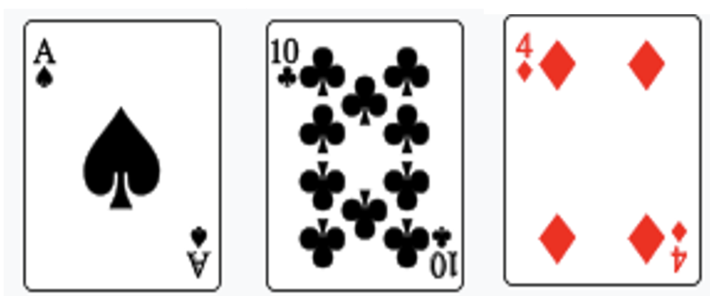
B: 10 of Clubs Second

C: 4 of Diamonds Third

Chain Rule (Idea)



Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards.

What is P () = P(A, B, C)?

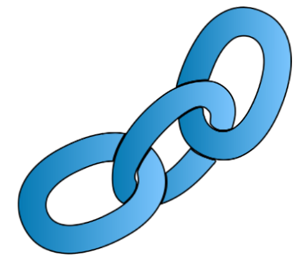
$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

Chain Rule (Idea)

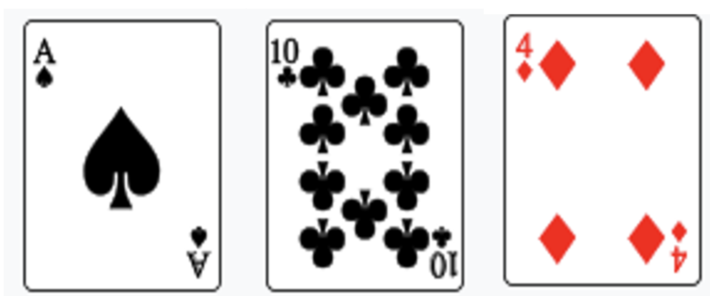


Have a Standard 52-Card Deck. Shuffle It, and draw the top 3 cards. (uniform probability space).

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

What is P (

) = P(A, B, C)?

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

$$P(A) \cdot P(B|A) \cdot P(C|A,B)$$

Chain Rule

Chain Rule: Let $A_1, \dots, A_n$ be events with nonzero probability. Then,

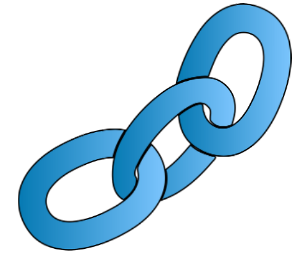
$$P(A_1, \dots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \dots P(A_n|A_1, \dots, A_{n-1})$$

In the case of two events A, B,

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

An easy way to remember this formula: we need to do n tasks, so we can perform them one at a time, conditioning on what we've done so far.

Chain Rule (Idea)

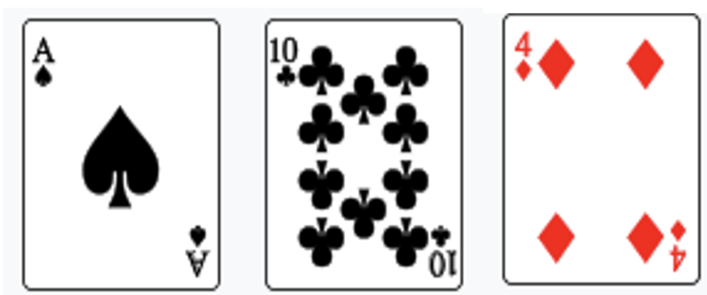


Have a standard 52-card deck. Shuffle it, and draw the top 3 cards. (uniform probability space).

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

What is P () = P(A, B, C)?

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

$$P(A) \cdot P(B|A) \cdot P(C|A,B)$$

Fun with cards

- Two people, A and B, are playing the following game.
- A 6-sided die is thrown and each time it’s thrown, regardless of the history, it is equally likely to show any of the six numbers
- If it shows 5, A wins.
- If it shows 1, 2 or 6, B wins.
- Otherwise, they play a second round and so on.

- What is Pr(A wins on 4^{th} round)?

Fun with cards

Two people, A and B, are playing the following game.
A 6-sided die is thrown and each time it’s thrown, regardless of the history, it is equally likely to show any of the six numbers
If it shows 5, A wins.
If it shows 1, 2 or 6, B wins.
Otherwise, they play a second round and so on.

What is Pr(A wins on 4^{th} round)?

The need for independence

Quick question: In general, is

$$P(A, B) = P(A)P(B)?$$

The need for independence

Quick question: In *general*, is

$$P(A, B) = P(A)P(B)?$$

The chain rule says

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

So no, unless the special case when $P(B|A) = P(B)$. This case is so important it has a name.

Independence

Independence: Events A, B are independent if any of the three equivalent conditions hold:

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A, B) = P(A)P(B)$

Independence

- Toss a coin 3 times. Each of 8 outcomes equally likely.

Define

- A = {at most one T} = {HHH, HHT, HTH, THH}
- B = {at most 2 Heads}= $\{HHH\}^c$

- Are A and B independent?

Using independence to define a probabilistic model

- We can **define** our probability model via independence.
- Example: suppose a biased coin comes up heads with probability 2/3, independent of other flips.
- Sample space: sequences of 3 coin tosses.

- Pr (HHH)=?
- Pr (TTT) = ?
- Pr (HHT) = ?
- Pr (2 heads) = ?