Am I going too fast in lecture? **Your responses are anonymous.**

https://pollev.com/ annakarlin185

A. Yes, you are going **way too fast.**

B. Yes, you are going a **little bit too fast.**

C. No, your pace is **about right.**

D. No, your pace is **too slow.**

The Birthday Paradox, Conditional Probability and more

CSE 312 Spring 26
Lecture 5

Many of the slides in this lecture were created by Mor Harchol-Balter for her book: "Introduction to Probability for Computing", Harchol-Balter '24

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

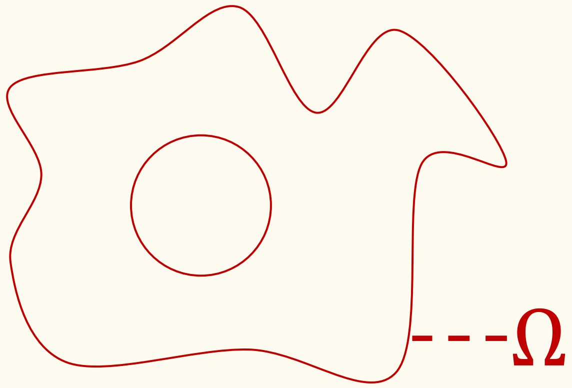
Probability space - more

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

Definition. A (discrete) **probability space** is a pair $(\Omega, \mathbb{P})$ where:

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

Set of possible **elementary outcomes**



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

Some outcome must show up

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

Review Axioms of Probability

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

Axiom 1 (Non-negativity): $P(E) \geq 0$
Axiom 2 (Normalization): $P(\Omega) = 1$
Axiom 3 (Countable Additivity): If E and F are mutually exclusive events,
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): $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Review: Equally Likely Outcomes

If (Ω, P) is a **uniform** probability space, then for any event $E \subseteq \Omega$, then

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

Examples:

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

Events:

- Getting at least one head in two coin flips: $E = \{HH, HT, TH\}$

$$P(E) = \frac{3}{4}$$

- Rolling an even number on a die : $E = \{2, 4, 6\}$

$$P(E) = \frac{3}{6} = \frac{1}{2}$$

Example: Birthday "Paradox"

Suppose we have a collection of n people in a room. What is the probability that at least 2 people share a birthday? Assume that there are exactly 365 possible birthdays, and each possibility is equally likely.

Say the people's names are 1, 2, 3,..n

Sample space: all possible assignments of birthdays to these people =

$$\Omega = \{(b_1, b_2, \ldots, b_n) \ : \ b_i \in \{1, \ldots, 365\}\}$$

(b_i = 25 means person i was born on the 25th day of the year)

$|\Omega| =$

Uniform probability space: $\mathbb{P}(\omega) =$

Example: Birthday "Paradox"

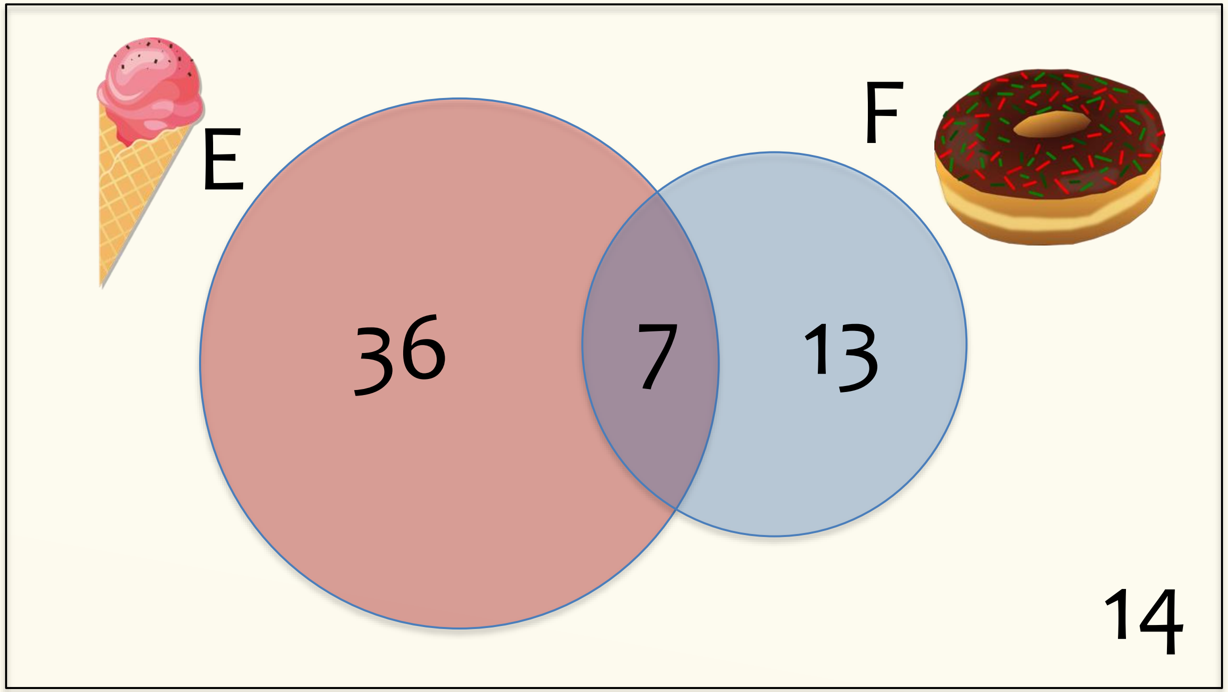
E = { at least 2 people share a birthday}

What is $\mathbb{P}(E)$?

Agenda (1)

- Conditional Probability
- Chain Rule
- Independence
- Law of Total Probability
- Bayes Theorem

Conditional Probability (Idea)



What’s the probability that a uniformly random person likes ice cream **given** they like donuts?

Conditional Probability

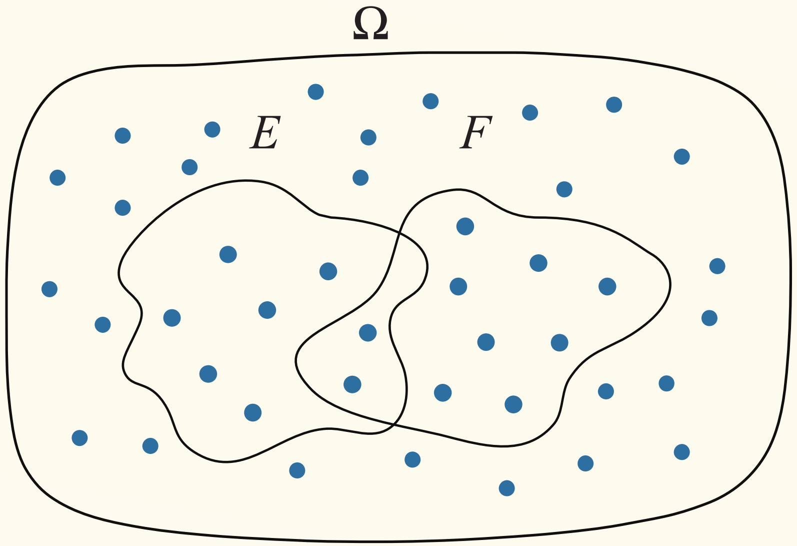
Definition. The **conditional probability** of event E **given** an event F happened (assuming $P(\mathrm{F}) > 0$) is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Conditional Probability

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

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$



Two equivalent views for **equally likely outcomes**:

$$\boldsymbol{P}\{E \mid F\} = \frac{2}{10}$$

(of the 10 outcomes in set F, only 2 of these are in set E)

$$\boldsymbol{P}\{E \mid F\} = \frac{\boldsymbol{P}\{E \cap F\}}{\boldsymbol{P}\{F\}} = \frac{\frac{2}{42}}{\frac{10}{42}} = \frac{2}{10}$$

Conditional Probability on Events

Q: What is $\boldsymbol{P}\{\text{both are colts}\}$?

The offspring of a horse is called a foal. Horse couples have one foal at a time and each foal is either a “colt” or a “filly.”

We’re told that a horse couple had 2 foals and that each outcome is equally likely.

Conditional Probability on Events

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

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Q: What is $\boldsymbol{P}$\{both are colts | $\geq$ 1 colt\}?

The offspring of a horse is called a foal. Horse couples have one foal at a time and each foal is either a "colt" or a "filly."

We're told that a horse couple had 2 foals and that each outcome is equally likely. We are also told that at least one of these is a colt.

Conditional Probability on Events - Answer

The offspring of a horse is called a foal. Horse couples have one foal at a time and each foal is either a "colt" or a "filly."

We're told that a horse couple had 2 foals and that each outcome is equally likely. We are also told that at least one of these is a colt.

$$P\{\text{both are colts} \mid \geq 1 \text{ colt}\}$$

$$= \frac{P\{\text{both are colts} \ \& \geq 1 \text{ colt}\}}{P\{\geq 1 \text{ colt}\}}$$

$$= \frac{P\{\text{both are colts}\}}{P\{\geq 1 \text{ colt}\}}$$

$$= \frac{\frac{1}{4}}{\frac{3}{4}}$$

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

Conditional Probability on Events - picture

$$P\{\text{both are colts} \mid \geq 1 \text{ colt}\} = \frac{1}{3}$$

The offspring of a horse is called a foal.
Horse couples have one foal at a time and each foal is either a "colt" or a "filly."

We're told that a horse couple had 2 foals and that each outcome is equally likely. We are also told that at least one of these is a colt.

	Colt	Filly
Colt	✓	
Filly		

Conditional Probability in a uniform probability space

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

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

If the probability space is uniform, then $P(E|F) = \frac{|E \cap F|}{|F|}$

Agenda (2)

- Conditional Probability
- Chain Rule
- Independence
- Law of Total Probability
- Bayes Theorem

Chain Rule

If $\boldsymbol{P}\{E_1 \cap E_2\} > 0$, then:

$$\boldsymbol{P}\{E_2|E_1\} = \frac{\boldsymbol{P}\{E_1 \cap E_2\}}{\boldsymbol{P}\{E_1\}}$$

Equivalently, we can write:

If $\boldsymbol{P}\{E_1 \cap E_2\} > 0$, then:

$$\boldsymbol{P}\{E_1 \cap E_2\} = \boldsymbol{P}\{E_1\} \cdot \boldsymbol{P}\{E_2|E_1\}$$

Likewise: $\boldsymbol{P}\{E_1 \cap E_2\} = \boldsymbol{P}\{E_2\} \cdot \boldsymbol{P}\{E_1|E_2\}$

Probability outcome is in both E_1 and E_2

Probability outcome is in E_1

Probability outcome is in E_2 given that it's in E_1

Chain Rule for Conditioning – general case.

If $\boldsymbol{P}\{E_1 \cap E_2\} > 0$, then:

$$\boldsymbol{P}\{E_1 \cap E_2\} = \boldsymbol{P}\{E_1\} \cdot \boldsymbol{P}\{E_2 | E_1\}$$

This can be generalized!

Theorem: **[Chain Rule for Conditioning]**

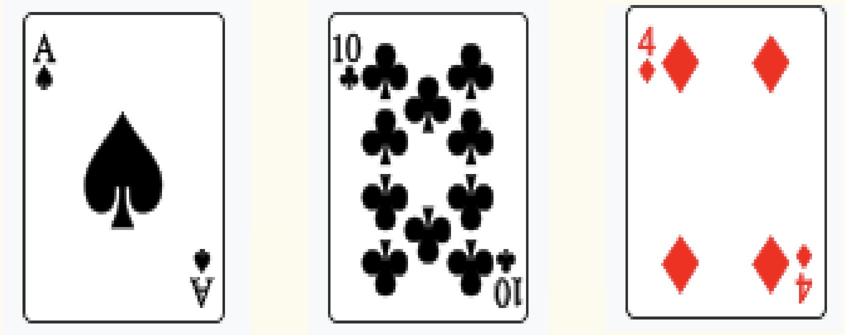
If $\boldsymbol{P}\{E_1 \cap E_2 \cap E_3 \cap \cdots \cap E_n\} > 0$, then

$$\boldsymbol{P}\{E_1 \cap E_2 \cap E_3 \cap \cdots \cap E_n\}$$

$$= \boldsymbol{P}\{E_1\} \cdot \boldsymbol{P}\{E_2 \mid E_1\} \cdot \boldsymbol{P}\{E_3 \mid E_1 \cap E_2\} \cdots \boldsymbol{P}\{E_n \mid E_1 \cap E_2 \cap E_3 \cap \cdots \cap E_{n-1}\}$$

Chain Rule Example

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

What is P (

) = P(**A** ∩ **B** ∩ **C**)?

A: Ace of Spades First

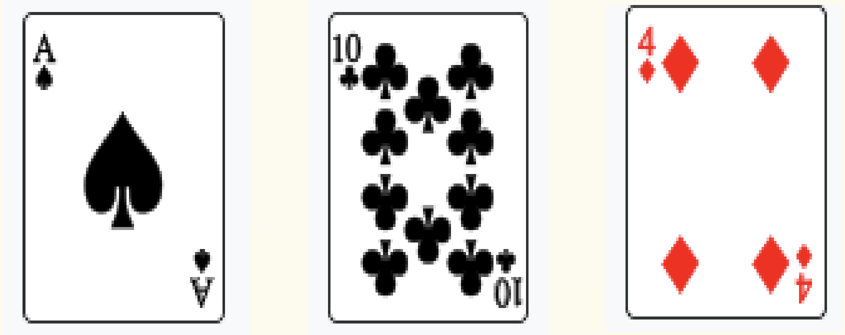
B: 10 of Clubs Second

C: 4 of Diamonds Third

Chain Rule Example

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

What is P (



) = P(**A** ∩ **B** ∩ **C**)?

A: Ace of Spades First

B: 10 of Clubs Second

C: 4 of Diamonds Third

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

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

Agenda (3)

- Conditional Probability
- Chain Rule
- Independence
- Law of Total Probability
- Bayes Theorem

Independent Events

Defn: Events E and F are **independent**, written $E \perp F$, if:

$$\boldsymbol{P}\{E \cap F\} = \boldsymbol{P}\{E\} \cdot \boldsymbol{P}\{F\}$$

Here's an equivalent and perhaps more intuitive definition:

Defn: Assuming $\boldsymbol{P}\{F\} > 0$, Events E and F are **independent,** if:

$$\boldsymbol{P}\{E \mid F\} = \boldsymbol{P}\{E\}$$

If E and F are **independent,** so are $\overline{E}$ and F, E and $\overline{F}$, and $\overline{E}$ and $\overline{F}$

Independent Events – example 1

Defn: Events E and F are **independent**, written $E \perp F$, if:

$$\boldsymbol{P}\{E \cap F\} = \boldsymbol{P}\{E\} \cdot \boldsymbol{P}\{F\}$$

Here's an equivalent and perhaps more intuitive definition:

Defn: Assuming $\boldsymbol{P}\{F\} > 0$, Events E and F are **independent**, if:

$$\boldsymbol{P}\{E \mid F\} = \boldsymbol{P}\{E\}$$

Q1: Can two mutually exclusive, non-null events be independent?

Independent Events – example 2

<u>Defn</u>: Events E and F are **independent**, written $E \perp F$, if:

$$\boldsymbol{P}\{E \cap F\} = \boldsymbol{P}\{E\} \cdot \boldsymbol{P}\{F\}$$

Here's an equivalent and perhaps more intuitive definition:

<u>Defn</u>: Assuming $\boldsymbol{P}\{F\} > 0$, Events E and F are **independent**, if:

$$\boldsymbol{P}\{E \mid F\} = \boldsymbol{P}\{E\}$$

Q2: Suppose we roll a die twice. Each outcome equally likely. Which of these pairs of events are independent:

a. Let E=“1st roll is 6.” Let F =“2nd roll is 6”
b. Let E=“Sum of rolls is 7.” Let F =“2nd roll is 4”

Independent Events - answers

Defn: Events E and F are **independent**, written $E \perp F$, if:

$$\boldsymbol{P}\{E \cap F\} = \boldsymbol{P}\{E\} \cdot \boldsymbol{P}\{F\}$$

Here's an equivalent and perhaps more intuitive definition:

Defn: Assuming $\boldsymbol{P}\{F\} > 0$, Events E and F are **independent**, if:

$$\boldsymbol{P}\{E \mid F\} = \boldsymbol{P}\{E\}$$

Q1: Can two mutually exclusive, non-null events be independent?

No!

Q2: Suppose we roll a die twice. Which of these pairs of events are independent:

a. Let E="1st roll is 6." Let F ="2nd roll is 6"
b. Let E="Sum of rolls is 7." Let F ="2nd roll is 4"

Both!

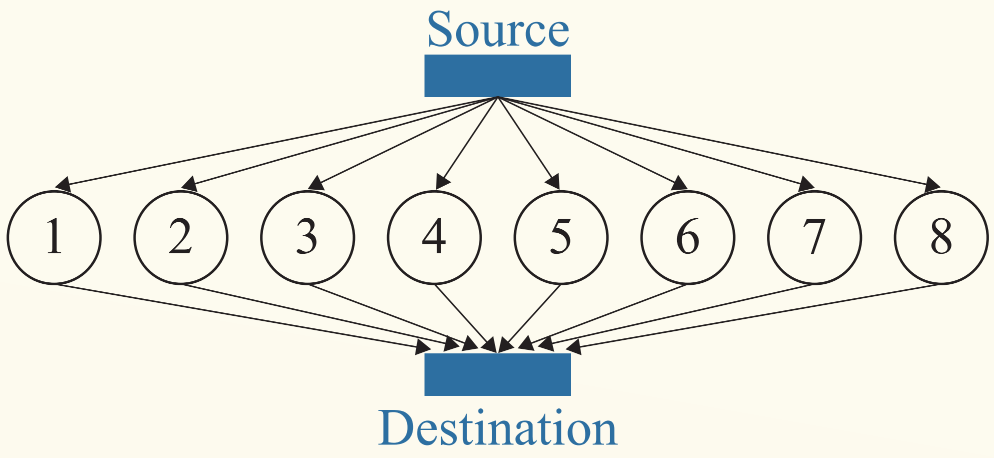
More Independence Definitions

Defn: Events $A_1, A_2, \ldots, A_n$ are **independent** if, for every subset S of $\{1, 2, \ldots, n\}$:

$$\mathbf{P}\left\{\bigcap_{i \in S} A_i\right\} = \prod_{i \in S} \boldsymbol{P}\{A_i\}$$

Another example

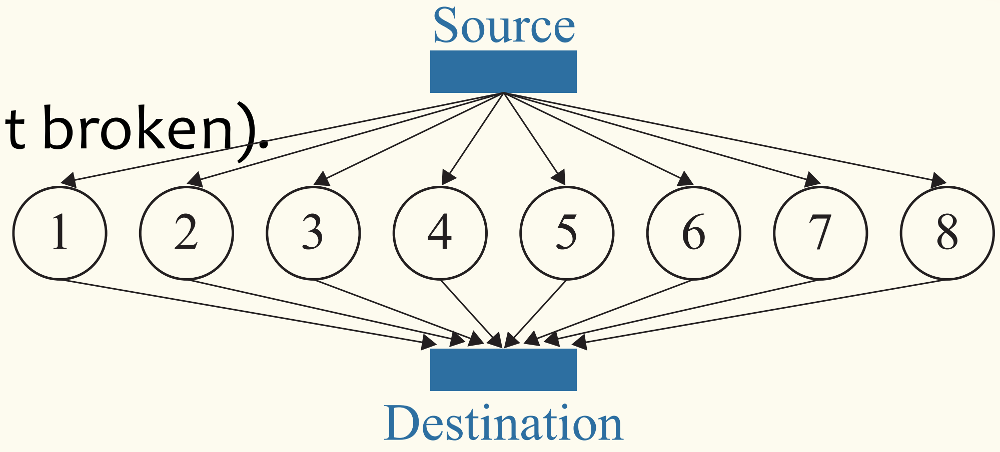
You are routing a packet from the source to the destination.
But each of the 16 edges in the network works with probability p, independently.



Q: What is the probability that you can get the packet from the source to the destination?

Let's break it down

Each edge works with probability p, independently of other edges. There are 8 paths.
Let E_i denote the event that the i^{th} path is usable (not broken).



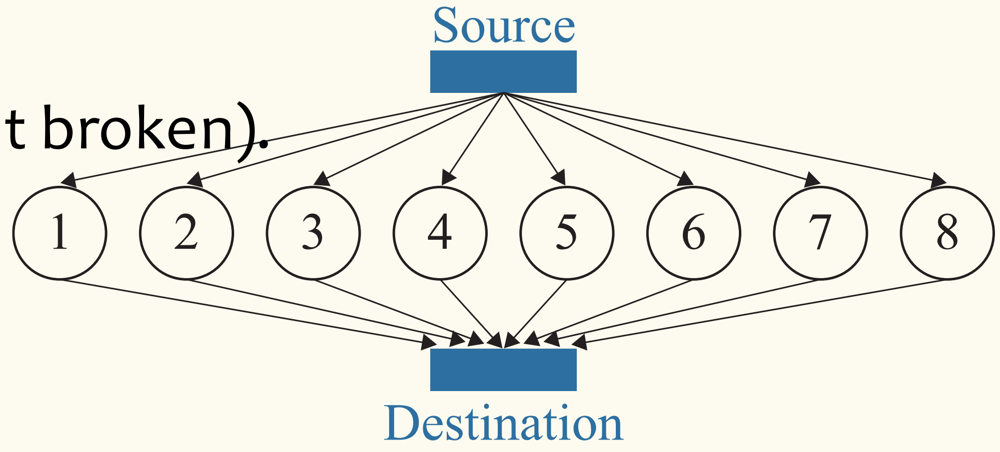
Q1: What is $\boldsymbol{P}\{E_i\}$?

Q2: What is $\boldsymbol{P}\{\overline{E_i}\}$?

More calculations

Each edge works with probability p, independently of other edges. There are 8 paths.

Let E_i denote the event that the i^{th} path is usable (not broken).



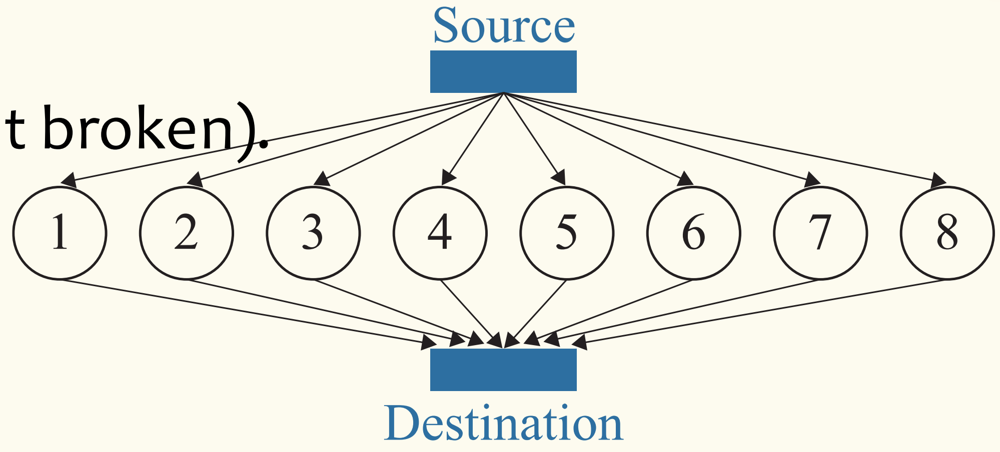
$\boldsymbol{P}\{E_i\} = p^2$ $\qquad$ $\boldsymbol{P}\{\overline{E_i}\} = 1 - p^2$

What is $\boldsymbol{P}$\{Can get from source to destination\}?

$\boldsymbol{P}$\{Can get from source to destination\} $=$

Answer

Each edge works with probability p, independently of other edges. There are 8 paths.
Let E_i denote the event that the i^{th} path is usable (not broken).



$\boldsymbol{P}\{E_i\} = p^2$ $\qquad\qquad$ $\boldsymbol{P}\{\overline{E_i}\} = 1 - p^2$

What is $\boldsymbol{P}\{\text{Can get from source to destination}\}$?

$$\boldsymbol{P}\{\text{Can get from source to destination}\} = \boldsymbol{P}\{\text{At least one path works}\}$$

$$= \boldsymbol{P}\{E_1 \cup E_2 \cup \cdots \cup E_8\}$$

$$= 1 - \boldsymbol{P}\{\text{All paths are broken}\}$$

$$= 1 - \boldsymbol{P}\{\overline{E_1}\} \cdot \boldsymbol{P}\{\overline{E_2}\} \cdots \boldsymbol{P}\{\overline{E_8}\} = 1 - (1 - p^2)^8$$

Agenda (4)

- Conditional Probability
- Chain Rule
- Independence
- Law of Total Probability
- Bayes Theorem

Law of Total Probability

For any sets E and F:

$$F = (F \cap E) \cup (F \cap \overline{E})$$

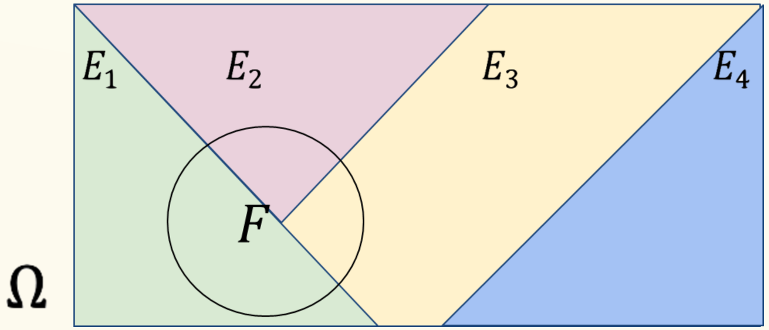
$$\boldsymbol{P}\{F\} = \boldsymbol{P}\{F \cap E\} + \boldsymbol{P}\{F \cap \overline{E}\}$$

$$= \boldsymbol{P}\{F|E\} \cdot \boldsymbol{P}\{E\} + \boldsymbol{P}\{F|\overline{E}\} \cdot \boldsymbol{P}\{\overline{E}\}$$

Generalizing, we have:

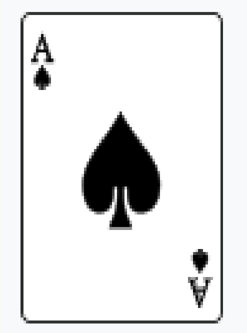
<u>Theorem</u>**: [Law of Total Probability]**
Let $E_1, E_2, \ldots, E_n$ partition the state space Ω. Then:

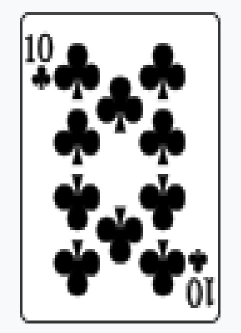
$$\boldsymbol{P}\{F\} = \sum_{i=1}^{n} \boldsymbol{P}\{F \cap E_i\} = \sum_{i=1}^{n} \boldsymbol{P}\{F|E_i\} \cdot \boldsymbol{P}\{E_i\}$$



Are A and B independent?

Have a Standard 52-Card Deck. Shuffle It, and draw the top 2 cards **in order**. (uniform probability space).





A: Ace of Spades First

B: 10 of Clubs Second

Are the events A and B independent?

$$\mathbb{P}(B|A) = \frac{1}{51}$$

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

Agenda (5)

- Conditional Probability
- Chain Rule
- Independence
- Law of Total Probability
- Bayes Theorem

Our First Machine Learning Task: Spam Filtering

`Subject: “FREE $$$ CLICK HERE”`

Suppose you know that 80% of emails you receive are spam.

So a priori, our belief is that any email has an 80% chance of being spam.

How do you update that belief when you see that the subject line contains the phrase “FREE $$$”?

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 follows from the definition of conditional probability

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

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

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

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

Back to Spam Filtering

```
Subject: “FREE $$$ CLICK HERE”
```

What is the probability this email is spam, given the subject contains “FREE”?
Some useful stats:

- 10% of ham (i.e., not spam) emails contain the word “FREE” in the subject.
- 70% of spam emails contain the word “FREE” in the subject.
- 80% of emails you receive are spam.

You receive a random email:

- Let S be event that the email is spam
- Let F be the event that the email contains the word “FREE”.

Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, E_2, \ldots, E_n$ be a partition of the sample space, and F and 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 non-zero probability, then

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