CS70 @ UC Berkeley, Spring 2026

Lecture 17
Conditional Probability

March 19, 2026

Last Lecture

- Sample space $\Omega = \{\text{possible outcomes}\}$.
- An event is a subset $E \subseteq \Omega$.

Definition (CS70 version): Discrete probability space $(\Omega, \mathbb{P})$.

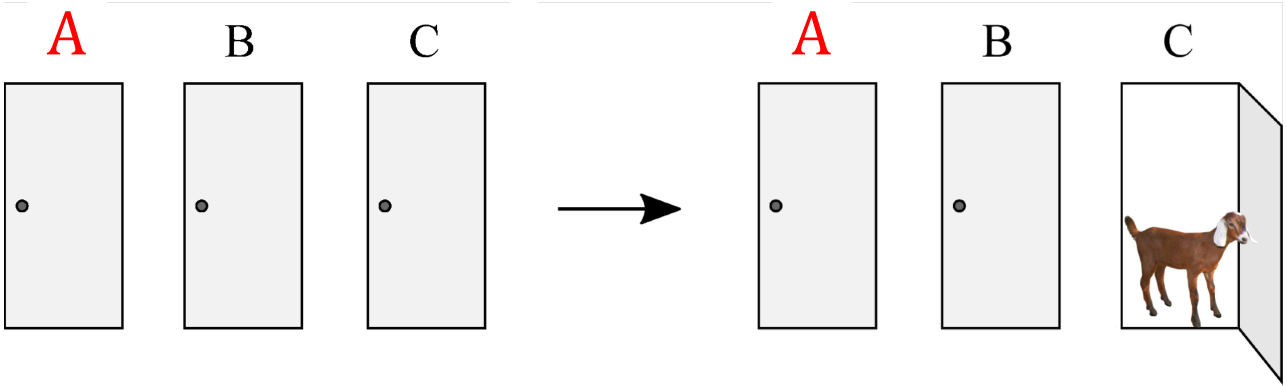
1. Sample space Ω.
2. A probability $\mathbb{P}(\{\omega\})$ for each $\omega \in \Omega$ such that
    i. $\mathbb{P}(\{\omega\}) \in [0,1], \forall \omega \in \Omega$.
    ii. $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$.

Given an event $E \subseteq \Omega$, we have $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\})$.

- For finite Ω, if all $\omega \in \Omega$ are equally likely, then $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$.
- In general, the function $\mathbb{P}$ (called probability measure, which is a systematic way of assigning a "size" to sets) needs to satisfy certain mathematical constraints.

Monty Hall (Review)

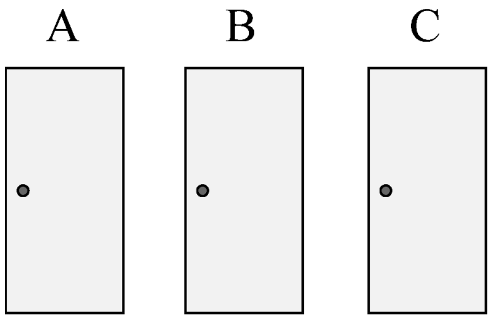
- A famous game show from 1970s, hosted by Monty Hall.
- One door has a **car** behind it. The other two doors have **goats**.



Question: Does the contestant have a better chance of winning if they switch doors?

- The contestant picks a door (A in this example), but doesn't open it.
- Monty Hall's assistant (Carol) opens one of the two remaining doors to reveal a goat (since Carol knows where the prize is, she can always do this).
- The contestant is then given the option of sticking with their current door, or switching to the other unopened one.

Monty Hall (Review)



Notation: (Prize door, Contestant's choice, Carol's choice)

$$\Omega = \{(A,A,B),(A,A,C),(B,B,A),(B,B,C),(C,C,A),(C,C,B),$$
$$(A,B,C),(A,C,B),(B,A,C),(B,C,B),(C,A,B),(C,B,A)\}$$

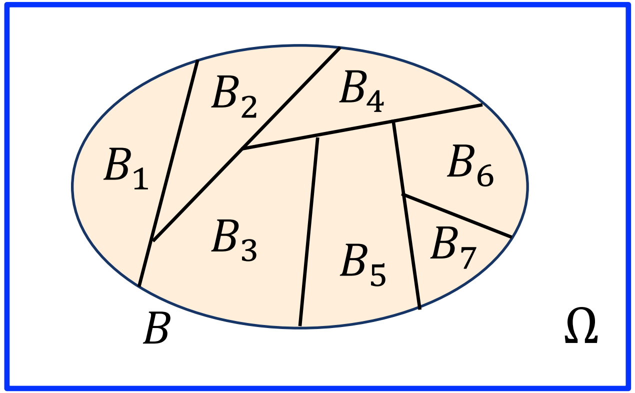
- For each $\omega \in \Omega$ in the first row, $\mathbb{P}(\{\omega\}) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}$
- For each $\omega \in \Omega$ in the second row, $\mathbb{P}(\{\omega\}) = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}$
- So, the contestant would be better off switching, since $\frac{\mathbb{P}[\{(B,A,C)\}]}{\mathbb{P}[\{(A,A,C)\}]} = 2.$

Rules of Probability

Definition (Partition): An event $B \subseteq \Omega$ is said to be partitioned into n events $B_1, \dots, B_n$ if

1. $B = B_1 \cup \cdots \cup B_n$,
2. $B_i \cap B_j = \emptyset$, for all $i \neq j$ (that is, $B_1, \dots, B_n$ are mutually exclusive).

More generally, infinitely many mutually exclusive B_ks may be involved ($B = \bigcup_{k=1}^{\infty} B_k$).



1. **(Non-negativity)** $\mathbb{P}(A) \geq 0$, for all $A \subseteq \Omega$.
2. **(Countable Additivity)** If $B_1, B_2, B_3, \dots$ is a partition of B, then

$$\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(B_k)$$

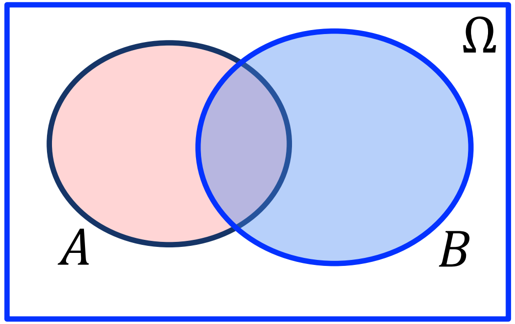
3. **(Normalization)** $\mathbb{P}(\Omega) = 1$.

Multiple Events

- In many applications of probability theory, we will want to think about multiple events defined on the same probability space.
- Examples include
  - the probability of words given their context (e.g., LLM)
  - the probability of states given observations (e.g., speech recognition)
  - the probability of similar items given user history (e.g., recommendation systems)
  - the probability of having a disease given a positive test result
  - the probability of rain given current atmospheric conditions
  - the probability of system failure given some components have failed
- In the next couple of lectures, we will learn key concepts involving multiple events, including conditional probability, joint probability, and independence.

Conditional Probability

- Consider two events $A, B \subseteq \Omega$.
- **Example:** Two-day weather forecast: **R**ain or **N**o rain
  - $\Omega = \{(R,R),(R,N),(N,R),(N,N)\}$
  - $A =$ "it rains on day two" $= \{(R,R),(N,R)\}$
  - $B =$ "it rains on day one" $= \{(R,R),(R,N)\}$
  - $A \cap B =$ "it rains on both days" $= \{(R,R)\}$
  - $B \setminus (A \cap B) = \{(R,N)\}$



- B can be partitioned into $A \cap B$ and $B \setminus (A \cap B)$.

$$1 \geq \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus (A \cap B))$$ By additivity

$$\geq \mathbb{P}(A \cap B) \geq 0$$ By non-negativity

So, if $\mathbb{P}(B) > 0$, we obtain $0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1$

Given that B happens, what is the probability that A also happens?

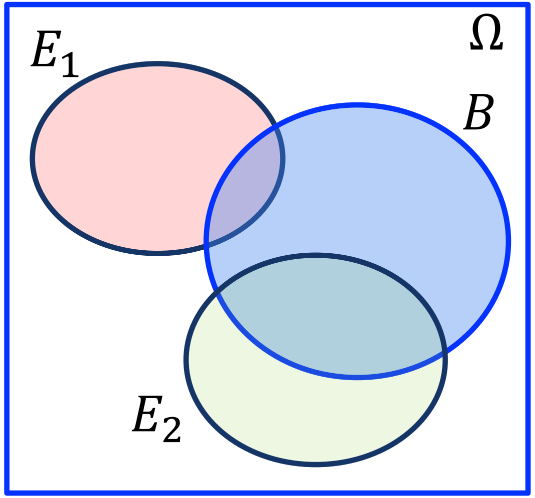
Conditioning on event B changes the probability space from $(\Omega, \mathbb{P})$ to $(B, \mathbb{P}_B)$.

Conditional Probability

- Conditioning on event B changes the probability space from $(\Omega, \mathbb{P})$ to $(B, \mathbb{P}_B)$.
- How should $\mathbb{P}_B$ be defined? It has to be consistent with $\mathbb{P}$.
- For all events $E_1, E_2 \subseteq \Omega$ where $E_1 \cap B \neq \emptyset$ and $E_2 \cap B \neq \emptyset$, we want

$$\frac{\mathbb{P}(E_1 \cap B)}{\mathbb{P}(E_2 \cap B)} = \frac{\mathbb{P}_B(E_1 \cap B)}{\mathbb{P}_B(E_2 \cap B)} \quad \Longrightarrow \mathbb{P}_B = c\mathbb{P}$$

$$1 = \mathbb{P}_B(B) = c\mathbb{P}(B) \quad \Longrightarrow \quad c = \frac{1}{\mathbb{P}(B)}$$

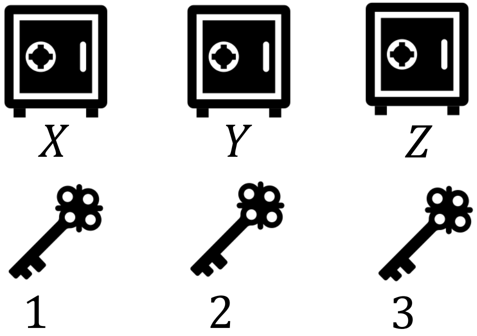


- The conditional probability of A given B is defined as

$$\mathbb{P}(A|B) := \mathbb{P}_B(A \cap B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1.$$ (see previous slide)

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$ (Multiplication Rule)

Three Keys and Three Locks



- We don't know which key opens which lock.
- Each lock can be opened by exactly one of the three keys.
- You try guessing randomly.
- Let event B = "Key 1 opens Lock X"
- Let event A = "Key 2 opens Lock Y"
- Question 1: What is $\mathbb{P}(A|B)$?
- Question 2: What is $\mathbb{P}(A|B^c)$?

How do they compare relative to $\mathbb{P}(A)$?

We expect

$$\mathbb{P}(A) < \mathbb{P}(A|B)$$

$$\mathbb{P}(A) > \mathbb{P}(A|B^c)$$

Why?

Let's calculate these probabilities exactly.

Three Keys and Three Locks

- We don't know which key opens which lock.
- Each lock can be opened by exactly one of the three keys.
- You try guessing randomly.
- Let event B = "Key 1 opens Lock X" $= \{\omega_1, \omega_2\}$
- Let event A = "Key 2 opens Lock Y" $= \{\omega_1, \omega_6\}$
- Question 1: What is $\mathbb{P}(A|B)$?
- Question 2: What is $\mathbb{P}(A|B^c)$?
- $B^c = \{\omega_3, \omega_4, \omega_5, \omega_6\}$
- $A \cap B = \{\omega_1\}$
- $A \cap B^c = \{\omega_6\}$

Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$

$\omega_1 = (1 \text{ opens } X, 2 \text{ opens } Y, 3 \text{ opens } Z)$

$\omega_2 = (1 \text{ opens } X, 3 \text{ opens } Y, 2 \text{ opens } Z)$

$\omega_3 = (2 \text{ opens } X, 1 \text{ opens } Y, 3 \text{ opens } Z)$

$\omega_4 = (2 \text{ opens } X, 3 \text{ opens } Y, 1 \text{ opens } Z)$

$\omega_5 = (3 \text{ opens } X, 1 \text{ opens } Y, 2 \text{ opens } Z)$

$\omega_6 = (3 \text{ opens } X, 2 \text{ opens } Y, 1 \text{ opens } Z)$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}$$

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

$$\mathbb{P}(A|B^c) = \frac{|A \cap B^c|}{|\bar{B}|} = \frac{1}{4}$$

So, the probability event A depends on whether event B happens or not.

Conditional Probability for a Uniform Probability Space

- If Ω is a finite set of equally likely outcomes, and $A, B \subseteq \Omega$ such that $|B| \neq 0$, then

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

Law of Total Probability

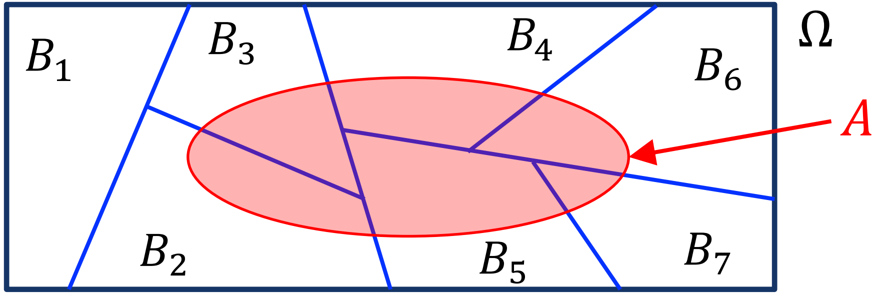
Theorem (Law of Total Probability or Total Probability Rule). Suppose $B_1, \dots, B_n$ is a partition of Ω. Then, for any $A \subseteq \Omega$,

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

Why is this useful?

a.k.a. rule of average conditional probabilities.

Proof:



- $B_1, \dots, B_n$ a partition of $\Omega \Longrightarrow (A \cap B_1), \dots, (A \cap B_n)$ a partition of A .

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) \quad = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

By additivity   By definition of conditional probability

Independence

Consider any two events $A, B \subseteq \Omega$ and suppose the chance of A does not depend on whether or not B occurs, i.e., $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$. (Note: in the Three Keys and Three Locks example, this property did NOT hold.) Then,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$ By the Total Probability Rule

$$= \mathbb{P}(A|B)[\mathbb{P}(B) + \mathbb{P}(B^c)]$$ Since $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$

$$= \mathbb{P}(A|B)\mathbb{P}(\Omega)$$ Since $B \cup B^c = \Omega$

$$= \mathbb{P}(A|B) = \mathbb{P}(A|B^c)$$ Since $\mathbb{P}(\Omega) = 1$

$$\Longrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$$

When this condition holds, we say that A and B are independent events. Next time, we will discuss independence involving multiple events.

Probabilistic Modular Arithmetic

- For $n \in \mathbb{Z}^+$, let $S_n = X_1 + X_2 + \cdots + X_n$, where each X_i is sampled **uniformly at random** (i.e., equal chance for each element) from $I = \{1,3,5,7,9\}$, independently of the other samples.
- What is $\mathbb{P}(S_n \equiv 0 \pmod 5)$?

$$\mathbb{P}(S_n \equiv 0 \pmod 5) = \sum_{i \in I} \underbrace{\mathbb{P}(S_n \equiv 0 \pmod 5 \mid X_n = i)}_{\mathbb{P}(S_{n-1} \equiv 5 - i \pmod 5)} \underbrace{\mathbb{P}(X_n = i)}_{\frac{1}{5}}$$

$$= \frac{1}{5} \sum_{j=0}^{4} \underbrace{\mathbb{P}(S_{n-1} \equiv j \pmod 5)}_{E_j} = \frac{1}{5}\,\mathbb{P}(\Omega) = \frac{1}{5}$$

E_0, E_1, E_2, E_3, E_4 partition Ω

i	$5 - i \pmod 5$
1	4
3	2
5	0
7	$-2 \equiv 3$
9	$-4 \equiv 1$

Which Coin?

	Coin 1	Coin 2	Coin 3
Probability of H	p_1	p_2	p_3
Probability of T	$1 - p_1$	$1 - p_2$	$1 - p_3$

- Pick a coin uniformly at random (u.a.r.) and toss it once.
- **Q:** What is the probability of observing Heads?
- $\Omega = \{(1,H),(1,T),(2,H),(2,T),(3,H),(3,T)\}$
- Event C_i = "coin i is chosen" $= \{(i,H),(i,T)\}$
- Note: $C_1 \cup C_2 \cup C_3 = \Omega$ and $C_i \cap C_j = \emptyset$ for $i \neq j$, so C_1, C_2, C_3 is a partition of Ω.
- $\mathbb{P}(H) = \sum_{i=1}^{3} \mathbb{P}(H|C_i)\mathbb{P}(C_i) = \frac{1}{3}(p_1 + p_2 + p_3)$

↑ Prior (belief before any observation)

Pick a coin u.a.r. and toss it twice.

Is $\mathbb{P}(HH) = \mathbb{P}(H)\mathbb{P}(H)$?

$$\mathbb{P}(HH) = \sum_{i=1}^{3} \mathbb{P}(HH|C_i)\mathbb{P}(C_i)$$

$$= \frac{1}{3}\sum_{i=1}^{3} p_i^2$$

$$\neq [\mathbb{P}(H)]^2$$

$$= \frac{1}{9}(p_1 + p_2 + p_3)^2$$

However,

$$\mathbb{P}(HH|C_i) = \mathbb{P}(H|C_i)\mathbb{P}(H|C_i)$$