MATH 7: HANDOUT 11

Conditional Probabilities and Monty-Hall Problem

Conditional Probability and Dependent Events

Why Conditional Probability?

Sometimes we learn *extra information* about an experiment after it happens (or while it is happening). That information can change how likely other events are. Conditional probability tells us how to update probabilities when we know that some event has occurred.

Definition

For events A and B with $P(B) > 0$, the **conditional probability** of A given B is

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}.$$

Read "$P(A \mid B)$" as "the probability of A given B". Intuition: we restrict attention to the cases where B happens, and measure how often A also happens among those.

Die roll with information

You roll a fair die. Let A = "number is even," B = "number is at least 4." Then $P(B) = \frac{3}{6} = \frac{1}{2}$ and $A \cap B = \{4, 6\}$, so $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$. Hence

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Knowing the number is ≥ 4 makes "even" more likely (now only $\{4, 5, 6\}$ are possible).

Multiplication Rule (General Form)

From the definition,

$$P(A \text{ and } B) = P(B)\, P(A \mid B) = P(A)\, P(B \mid A).$$

This is useful for building complicated probabilities step by step.

Independent vs. Dependent Events

Events A and B are **independent** if knowing one does not change the probability of the other:

$$P(A \mid B) = P(A) \quad \text{(equivalently, } P(B \mid A) = P(B)\text{)}.$$

In that case, $P(A \text{ and } B) = P(A) \cdot P(B)$.

If $P(A \mid B) \neq P(A)$, then A and B are **dependent**.

With vs. without replacement

A box has 3 red and 2 blue marbles.

- *With replacement:* draw one marble, replace it, draw again. The two draws are independent.
- *Without replacement:* draw one marble, keep it out, draw again. The second draw depends on the

first (dependent events).

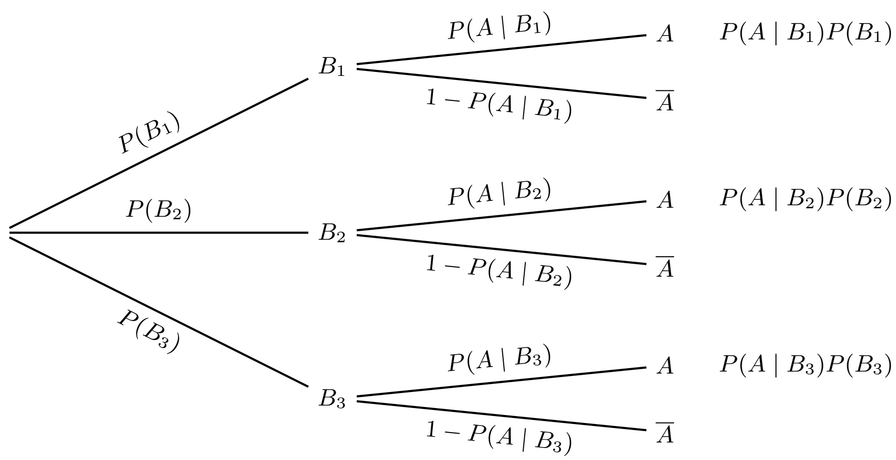
Law of Total Probability

Sometimes an event can happen in several different *ways* or *cases*. If these cases cover all possibilities and do not overlap (that is, exactly one must happen), then the total probability of A can be found by **adding the probabilities from each case**.

Suppose the experiment can happen through cases B_1, B_2, and B_3. Then:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + P(A \mid B_3)P(B_3).$$

In words: *"Find the probability of A in each case, multiply by how likely that case is, and then add them all up."*



Tree diagram for the law of total probability.

Examples

Cards without replacement

A standard deck has 52 cards (13 of each suit). You draw 2 cards without replacement.

1. Probability both are hearts?
2. Probability second card is a heart?

Solution.

1. $P(\text{first heart}) = \frac{13}{52} = \frac{1}{4}$, and $P(\text{second heart} \mid \text{first heart}) = \frac{12}{51}$. Hence $P(\text{both hearts}) = \frac{1}{4} \cdot \frac{12}{51} = \frac{12}{204} = \frac{1}{17}$.
2. Law of total probability:

$$P(\text{2nd heart}) = \left(\frac{12}{51}\right)\left(\frac{13}{52}\right) + \left(\frac{13}{51}\right)\left(\frac{39}{52}\right) = \frac{1}{4}.$$

Marbles without replacement

A box has 3 red (R) and 2 blue (B) marbles. You draw *two* marbles without replacement.

1. What is the probability that both are red?
2. What is the probability that the second is red?

Solution.

1. Let $A_1 =$ "first is red," $A_2 =$ "second is red." Then

$$P(A_1) = \frac{3}{5}, \qquad P(A_2 \mid A_1) = \frac{2}{4} = \frac{1}{2}.$$

By the multiplication rule,

$$P(\text{both red}) = P(A_1 \text{ and } A_2) = P(A_1)\,P(A_2 \mid A_1) = \frac{3}{5} \cdot \frac{1}{2} = \frac{3}{10}.$$

2. Use the *law of total probability* (see below) with the first draw splitting the cases:

$$P(\text{second is red}) = P(A_2 \mid \text{first R})P(\text{first R}) + P(A_2 \mid \text{first B})P(\text{first B}).$$

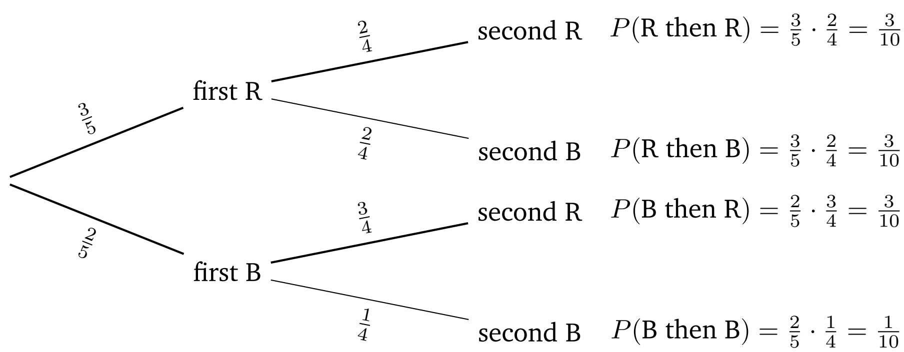
Here $P(\text{first R}) = \frac{3}{5}$, $P(\text{first B}) = \frac{2}{5}$,

$$P(A_2 \mid \text{first R}) = \frac{2}{4} = \frac{1}{2}, \qquad P(A_2 \mid \text{first B}) = \frac{3}{4}.$$

Therefore

$$P(\text{second is red}) = \left(\tfrac{1}{2}\right)\left(\tfrac{3}{5}\right) + \left(\tfrac{3}{4}\right)\left(\tfrac{2}{5}\right) = \tfrac{3}{10} + \tfrac{6}{20} = \tfrac{3}{10} + \tfrac{3}{10} = \tfrac{3}{5}.$$

Notice it equals the initial red fraction $3/5$ (symmetry).



Tree diagram for two draws without replacement.

Testing a Simple Device

A machine can either **work** or **fail**:

- $P(\text{works}) = 0.9$
- $P(\text{fails}) = 0.1$

If it works, a light turns green with probability 0.95. If it fails, the light still turns green with probability 0.2 (a false signal).

We first find the overall probability that the light is green. There are two disjoint cases:

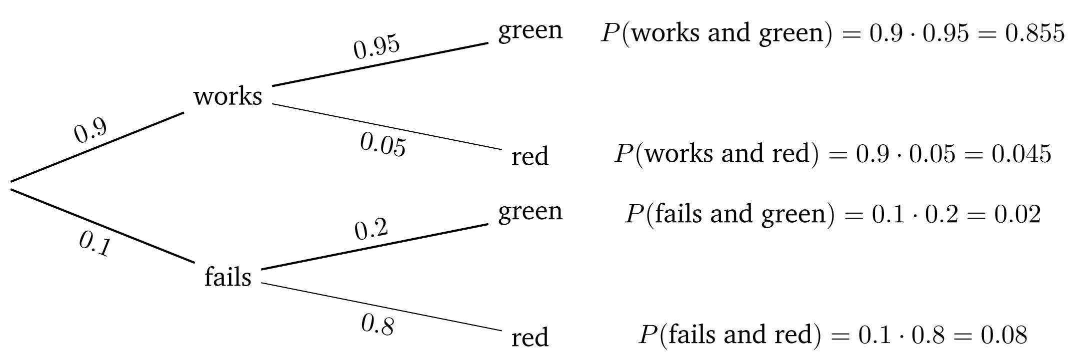
$$\begin{aligned} P(\text{green}) &= P(\text{green} \mid \text{works})P(\text{works}) + P(\text{green} \mid \text{fails})P(\text{fails}) \\ &= 0.95 \times 0.9 + 0.2 \times 0.1 = 0.855 + 0.020 = 0.875. \end{aligned}$$

So the light is green about 87.5% of the time.

Now suppose you see that the light is green. What is the probability the machine is actually working?

$$\begin{aligned} P(\text{works} \mid \text{green}) &= \frac{P(\text{works and green})}{P(\text{green})} \\ &= \frac{P(\text{green} \mid \text{works})P(\text{works})}{P(\text{green})} \\ &= \frac{0.95 \times 0.9}{0.875} \approx 0.977. \end{aligned}$$

Even though the green light appears 87.5% of the time, it correctly indicates a working machine about 97.7% of the time. This small drop shows the effect of "false greens."



Tree diagram for testing a simple device.

Medical Screening Test (P(disease) = 1%)

A person can either **have a certain disease** or be **healthy**:

- $P(\text{disease}) = 0.01$
- $P(\text{healthy}) = 0.99$

A screening test can return either **positive** or **negative**.

- If the person has the disease, the test is positive with probability 0.99 (high sensitivity).
- If the person is healthy, the test is still positive with probability 0.05 (a false positive).

We first compute the overall probability that the test is positive. There are two disjoint cases:

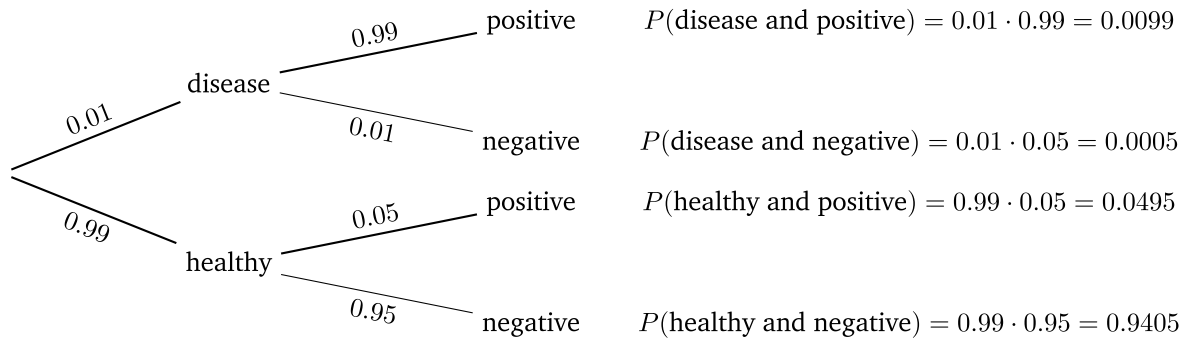
$$\begin{aligned} P(\text{positive}) &= P(\text{positive} \mid \text{disease})P(\text{disease}) + P(\text{positive} \mid \text{healthy})P(\text{healthy}) \\ &= 0.99 \times 0.01 + 0.05 \times 0.99 \\ &= 0.0099 + 0.0495 = 0.0594. \end{aligned}$$

So about 5.94% of people tested receive a positive result.

Now suppose a person tests positive. What is the probability that the person actually has the disease?

$$\begin{aligned} P(\text{disease} \mid \text{positive}) &= \frac{P(\text{disease and positive})}{P(\text{positive})} \\ &= \frac{P(\text{positive} \mid \text{disease})P(\text{disease})}{P(\text{positive})} \\ &= \frac{0.99 \times 0.01}{0.0594} \approx 0.167. \end{aligned}$$

Even though the test is very accurate, a positive result means the person actually has the disease only about 16.7% of the time. This is because the disease is rare, so false positives greatly outnumber true positives.



Tree diagram for medical testing.

Medical Screening Test (P(disease) = 10%)

A person can either **have a certain disease** or be **healthy**:

- $P(\text{disease}) = 0.1$
- $P(\text{healthy}) = 0.9$

A screening test can return either **positive** or **negative**.

- If the person has the disease, the test is positive with probability 0.99.
- If the person is healthy, the test is still positive with probability 0.05 (a *false positive*).

We first compute the overall probability that the test is positive. There are two disjoint cases:

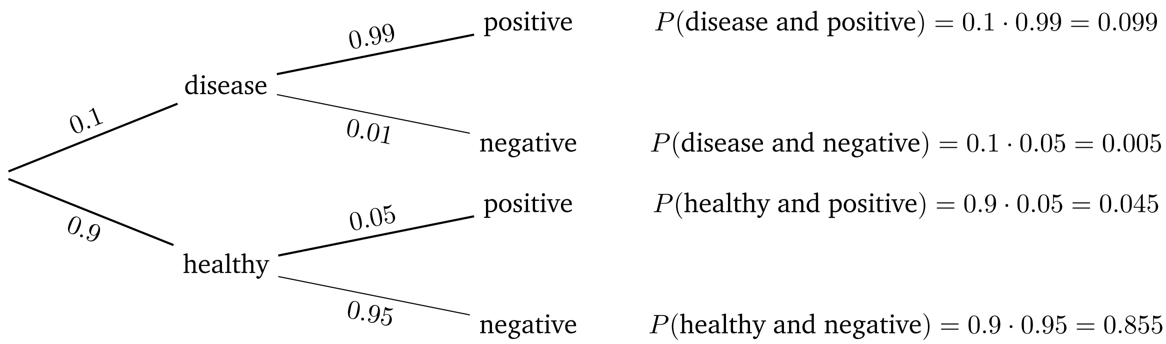
$$\begin{aligned} P(\text{positive}) &= P(\text{positive} \mid \text{disease})P(\text{disease}) + P(\text{positive} \mid \text{healthy})P(\text{healthy}) \\ &= 0.99 \times 0.1 + 0.05 \times 0.9 \\ &= 0.099 + 0.045 = 0.144. \end{aligned}$$

So about 14.4% of people tested receive a positive result.

Now suppose a person tests positive. What is the probability that the person actually has the disease?

$$\begin{aligned} P(\text{disease} \mid \text{positive}) &= \frac{P(\text{disease and positive})}{P(\text{positive})} \\ &= \frac{P(\text{positive} \mid \text{disease})P(\text{disease})}{P(\text{positive})} \\ &= \frac{0.99 \times 0.1}{0.144} \approx 0.6875. \end{aligned}$$

Even though the test is very accurate, a positive result means the person actually has the disease about 68.8% of the time. This shows how false positives still matter, even when the disease is not rare.



Tree diagram for medical testing.

Bayes' Rule in Spam Filtering

Email services try to decide whether a message is **spam** or **not spam** based on certain keywords. Suppose we know:

- $P(\text{Spam}) = 0.20, \quad P(\text{Not Spam}) = 0.80$
- $P(\text{“free”} \mid \text{Spam}) = 0.70$
- $P(\text{“free”} \mid \text{Not Spam}) = 0.05$

Question. An email contains the word *“free”*. What is the probability that the email is spam?

Step 1: Find the total probability that an email contains “free”.

$$\begin{aligned} P(\text{“free”}) &= P(\text{“free”} \mid \text{Spam})P(\text{Spam}) + P(\text{“free”} \mid \text{Not Spam})P(\text{Not Spam}) \\ &= (0.70)(0.20) + (0.05)(0.80) \\ &= 0.18 \end{aligned}$$

Step 2: Apply Bayes' Rule.

$$P(\text{Spam} \mid \text{“free”}) = \frac{P(\text{“free”} \mid \text{Spam})P(\text{Spam})}{P(\text{“free”})} = \frac{0.14}{0.18} \approx 0.78$$

Conclusion. If an email contains the word *“free”*, there is about a **78% chance** that it is spam.

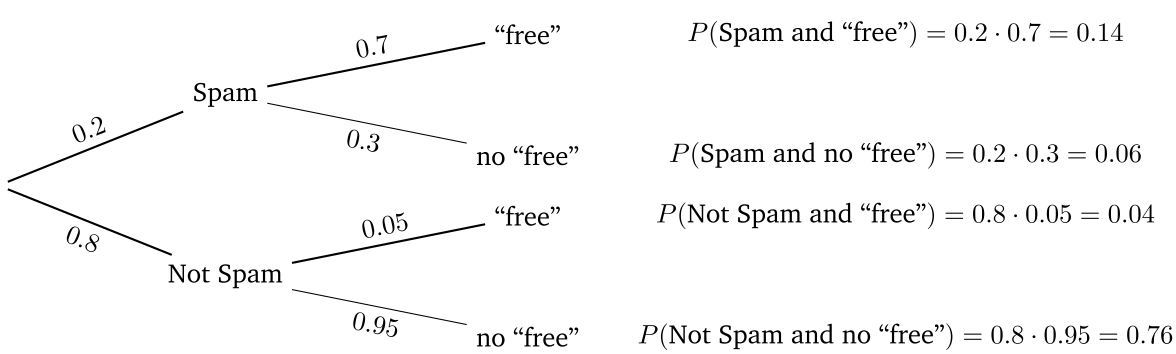
How real spam filters use Bayes' Rule.
Real spam filters do not look at just one word. Instead, they:

- examine **many words and features** (such as “free”, “winner”, links, or sender address),
- estimate how likely each feature is to appear in spam and non-spam emails,
- and then **combine these probabilities** using Bayes' Rule.

Each word slightly increases or decreases the probability that an email is spam. After combining all the evidence, the filter:

- marks the email as **spam** if the probability is high enough,
- or lets it into the inbox if the probability is low.

This means that spam filtering is not guessing — it is using **conditional probability and data** to make a decision.



Tree diagram for spam filtering.

Bayes' Rule (Gentle Preview)

When $P(B) > 0$,

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

It is often used together with the law of total probability to compute $P(B)$ in the denominator. We will apply this idea explicitly in the **Monty Hall** section.

Why Bayes' Rule Is So Important

Bayes' Rule helps us answer the question we *actually care about*:

"Given what I observed, what is really going on?"

In many real situations:

- We observe an **effect** (a positive test, a warning signal, a symptom),
- but we want to know the **cause** (disease, failure, danger).

Bayes' Rule tells us how to reverse conditional probabilities:

$$P(\text{cause} \mid \text{effect}) = \frac{P(\text{effect} \mid \text{cause}) P(\text{cause})}{P(\text{effect})}.$$

Why intuition often fails:

- People focus on how accurate a test is,
- but ignore how **rare** the condition is.

When the condition is rare, even a very accurate test can produce many false positives. Bayes' Rule correctly balances:

- **test accuracy** (sensitivity and false positives), and
- **base rate** (how common the condition is).

Where Bayes' Rule is used:

- medical testing and diagnosis,
- spam filters and facial recognition,
- quality control and safety systems,
- legal evidence and forensic science.

Bayes' Rule does not change the facts — it changes how we *interpret* the facts once new information appears.

Common Pitfalls

1. **Confusing $P(A \mid B)$ with $P(B \mid A)$.** They are generally different.
2. **Forgetting to divide by $P(B)$.** The definition $P(A \mid B) = P(A \cap B)/P(B)$ always needs $P(B) > 0$.
3. **Assuming independence without reason.** Check whether the second step depends on the first (e.g., without replacement).

Conditional Probability and Dependent Events

- Conditional probability measures how likely an event A is, given that another event B has occurred:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}, \qquad P(B) > 0.$$

- **Multiplication Rule:**

$$P(A \text{ and } B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A).$$

- **Independence:** Events A and B are independent if

$$P(A \mid B) = P(A) \quad \text{(or equivalently, } P(B \mid A) = P(B)\text{)},$$

  so $P(A \text{ and } B) = P(A)\,P(B)$.

- **Dependent events** occur when one event changes the probability of the other (e.g., draws *without* replacement).

- **Law of Total Probability:** If $\{B_1, \ldots, B_m\}$ split the sample space,

$$\begin{aligned} P(A) &= \sum_{i=1}^{m} P(A \mid B_i)\,P(B_i) \\ &= P(A \mid B_1)\,P(B_1) + P(A \mid B_2)\,P(B_2) + \cdots + P(A \mid B_m)\,P(B_m). \end{aligned}$$

Practice Problems

1. A jar has 7 red and 5 blue candies. You pick one at random, do *not* replace it, then pick a second.
   (a) Find $P(\text{both red})$.
   (b) Find $P(\text{second is blue})$.
2. A fair die is rolled once. Let $A = \{\text{even}\}$ and $B = \{\text{number} \geq 4\}$.
   (a) Compute $P(A \mid B)$ and $P(B \mid A)$.
   (b) Are A and B independent? Justify.
3. A box contains 4 good bulbs and 1 defective bulb. Two bulbs are chosen without replacement.
   (a) Find $P(\text{both good})$.
   (b) Find $P(\text{second is good})$ using the law of total probability.
4. In a school, 40% of students take Spanish (event S), 30% take Art (event T), and 13% take both.
   (a) Compute $P(S \mid T)$ and $P(T \mid S)$.
   (b) Are S and T independent? Why or why not?
5. A family has two children. Each child is independently a boy or a girl with probability $\frac{1}{2}$.
   (a) Given that the *older* child is a boy, what is $P(\text{both are boys})$?
   (b) Given that *at least one* child is a boy, what is $P(\text{both are boys})$?
   (c) Explain why the answers to (a) and (b) differ.
6. A medical test for a disease has sensitivity 0.98 (true positive rate) and specificity 0.97 (true negative rate). The disease prevalence in the population is 1%.

(a) If a randomly selected person tests positive, compute $P(\text{diseased} \mid \text{positive})$.
(b) If a randomly selected person tests negative, compute $P(\text{healthy} \mid \text{negative})$.
(c) Explain in a sentence why positive results can still be false when prevalence is low.

The Monty Hall Problem (Switch or Stay?)

Setup and Host Rules

There are 3 doors. One hides a **car**, two hide **goats**. You pick one door; then the host, who knows where the car is, opens one *other* door showing a goat and always offers you the chance to switch to the remaining closed door.

Host assumptions:

- The host never opens your door.
- The host never opens the car door.
- If the host has a choice of two goat doors to open, he follows a fixed rule (e.g., picks randomly); this does not change the final $2/3$ vs $1/3$ result.

Key Question

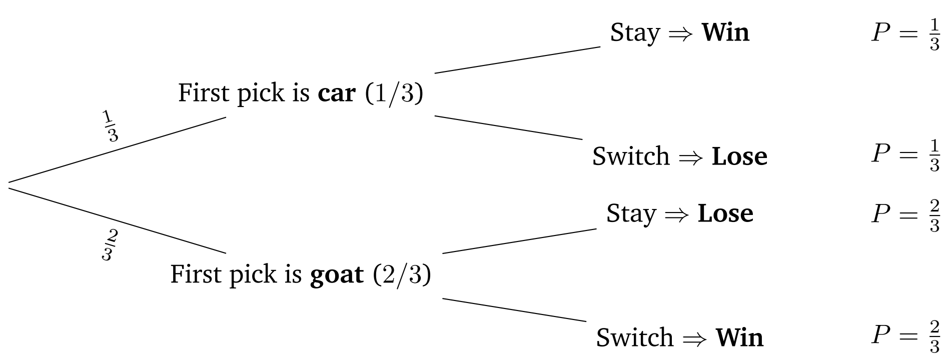
Is it better to **stay** or to **switch**? Which strategy has the higher probability of winning the car?

Quick Intuition

Your first pick has probability $\frac{1}{3}$ of being the car and $\frac{2}{3}$ of being a goat. When the host opens a goat door, he does not change the fact that your original pick was likely wrong. **Staying** wins only if your first pick was right ($\frac{1}{3}$ chance). **Switching** wins whenever your first pick was wrong ($\frac{2}{3}$ chance). Therefore, switching wins with probability $\boxed{\frac{2}{3}}$.

Tree Diagram Argument

There are two cases for your *initial* choice:



If first pick is correct (prob $1/3$), staying wins; if first pick is wrong (prob $2/3$), switching wins.

Therefore:

$$P(\text{win by stay}) = \tfrac{1}{3}, \qquad P(\text{win by switch}) = \tfrac{2}{3}.$$

Tiny Outcome Table

Assume you initially pick Door 1; each car placement has probability $\frac{1}{3}$.

Car behind	Your pick	Host opens (goat)	Switch goes to
1	1	one of $\{2, 3\}$	goat (lose)
2	1	3	2 (car, win)
3	1	2	3 (car, win)

In 2 out of 3 equally likely cases, switching wins.

Events and Notation

Let:

- C = "your initial pick is the car" (so $P(C) = \frac{1}{3}$),
- $\overline{C}$ = "your initial pick is a goat" ($P(\overline{C}) = \frac{2}{3}$),
- W = "you win the car."

Stay vs. Switch using Conditional Probabilities

We compute the win probabilities by splitting into cases based on whether your first pick was right.

Staying.

$$P(W \mid \text{Stay}) \;=\; P(W \mid \text{Stay}, C)\, P(C) \;+\; P(W \mid \text{Stay}, \overline{C})\, P(\overline{C}).$$

If C happens, staying wins for sure: $P(W \mid \text{Stay}, C) = 1$. If $\overline{C}$ happens, staying certainly loses: $P(W \mid \text{Stay}, \overline{C}) = 0$. Hence

$$P(W \mid \text{Stay}) \;=\; 1 \cdot \tfrac{1}{3} + 0 \cdot \tfrac{2}{3} \;=\; \boxed{\tfrac{1}{3}}.$$

Switching.

$$P(W \mid \text{Switch}) \;=\; P(W \mid \text{Switch}, C)\, P(C) \;+\; P(W \mid \text{Switch}, \overline{C})\, P(\overline{C}).$$

If C happens, switching loses for sure: $P(W \mid \text{Switch}, C) = 0$. If $\overline{C}$ happens, the host is forced to reveal the only goat among the two unchosen doors, so the remaining closed door is the car: $P(W \mid \text{Switch}, \overline{C}) = 1$. Therefore

$$P(W \mid \text{Switch}) \;=\; 0 \cdot \tfrac{1}{3} + 1 \cdot \tfrac{2}{3} \;=\; \boxed{\tfrac{2}{3}}.$$

Conclusion: Switching doubles your chance of winning compared to staying.

Why Opening a Goat Door Helps You

The host's action is *not random ignorance*—the host avoids the car on purpose. This "reveals information" that effectively moves the entire $\frac{2}{3}$ chance of your first being wrong onto the other unopened door.

Extension: 4 Doors, Host Opens 2 Goats

Suppose there are 4 doors (1 car, 3 goats). You pick one. The host opens *two other doors*, both goats, then offers you the chance to switch to the single remaining closed door.

$$P(\text{initially correct}) = \tfrac{1}{4} \quad \Rightarrow \quad \text{staying wins with } \tfrac{1}{4}.$$

If your first pick was wrong (probability $\frac{3}{4}$), the car must be among the other doors; after the host opens *all but one* goat door, the last closed door (not yours) carries the entire $\frac{3}{4}$ chance.

$$\boxed{P(\text{win by switching}) = \tfrac{3}{4}.}$$

So the advantage of switching becomes even larger with more doors.

Practice Problems

7. **Four doors, host opens only one goat door.** There are 4 doors. You pick one. The host opens *one* goat door (not two) and offers a switch. Compute $P(W \mid \text{Stay})$ and $P(W \mid \text{Switch})$ by conditioning on C and $\overline{C}$ and carefully justifying the "switch" outcome in each case.

8. **General n-door version.** There are n doors (1 car, $n-1$ goats). You pick one. The host opens $n-2$ goat doors, leaving you and one other closed door. Argue (without advanced formulas) that switching wins with probability $\frac{n-1}{n}$.

The Monty Hall Problem — Switch or Stay?

- 3 doors: 1 car, 2 goats. You pick one; the host (who knows) opens a goat door and offers to switch.
- Your first pick is correct with probability $\frac{1}{3}$, wrong with $\frac{2}{3}$.
- **Stay:** win only if first pick was right $\Rightarrow P = \frac{1}{3}$.
- **Switch:** win if first pick was wrong $\Rightarrow P = \frac{2}{3}$.
- Host's reveal gives information — the unopened door inherits the $\frac{2}{3}$ chance.
- With n doors (1 car, $n-1$ goats, host opens $n-2$ goats):

$$P(\text{win by switch}) = \frac{n-1}{n}.$$

- **Common mistake:** Thinking it's 50–50 after one door opens — it isn't!

Practice Problems: Solutions

1. A jar has 7 red and 5 blue candies (12 total). Draw 2 without replacement.

   (a)
   $$P(\text{both red}) = \frac{7}{12} \cdot \frac{6}{11} = \frac{42}{132} = \frac{7}{22}.$$

   (b) By symmetry (or a quick check with total probability), the probability the second is blue equals the overall fraction of blue candies:
   $$P(\text{second is blue}) = \frac{5}{12}.$$

2. A fair die is rolled once. $A = \{\text{even}\} = \{2, 4, 6\}$, $B = \{\geq 4\} = \{4, 5, 6\}$.

   (a) $A \cap B = \{4, 6\}$, so $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$ and $P(A) = \frac{3}{6} = \frac{1}{2}$, $P(B) = \frac{3}{6} = \frac{1}{2}$.
   $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}, \qquad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

   (b) Check independence:
   $$P(A \cap B) = \frac{1}{3} \quad \text{but} \quad P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

   Since $\frac{1}{3} \neq \frac{1}{4}$, A and B are **not** independent.

3. A box has 4 good (G) bulbs and 1 defective (D). Choose 2 without replacement.

   (a)
   $$P(\text{both good}) = \frac{4}{5} \cdot \frac{3}{4} = \frac{3}{5}.$$

   (b) Use law of total probability by conditioning on the first bulb:
   $$P(\text{second good}) = P(G_2 \mid G_1)P(G_1) + P(G_2 \mid D_1)P(D_1).$$

   Here,
   $$P(G_1) = \frac{4}{5}, \quad P(D_1) = \frac{1}{5}, \quad P(G_2 \mid G_1) = \frac{3}{4}, \quad P(G_2 \mid D_1) = 1.$$

   So
   $$P(\text{second good}) = \frac{3}{4} \cdot \frac{4}{5} + 1 \cdot \frac{1}{5} = \frac{3}{5} + \frac{1}{5} = \frac{4}{5}.$$

4. In a school: $P(S) = 0.40$, $P(T) = 0.30$, $P(S \cap T) = 0.13$.

   (a)
   $$P(S \mid T) = \frac{P(S \cap T)}{P(T)} = \frac{0.13}{0.30} = 0.039, \qquad P(T \mid S) = \frac{P(S \cap T)}{P(S)} = \frac{0.13}{0.40} = 0.052.$$

   (b) Independent means $P(S \cap T) = P(S)P(T)$.
   $$P(S)P(T) = 0.40 \cdot 0.30 \neq 0.13 = P(S \cap T).$$

   Since they don't match, S and T **are not independent**.

5. Two children (independent, boy/girl each $\frac{1}{2}$):

   (a) Given older is boy: outcomes $(B, B), (B, G) \Rightarrow P(\text{both B}) = \frac{1}{2}$.

   (b) Given at least one boy: $(B, B), (B, G), (G, B) \Rightarrow P(\text{both B}) = \frac{1/4}{3/4} = \frac{1}{3}$.

(c) The information differs (conditioning on a specific child vs. on the set "at least one").

6. Medical test: sensitivity 0.98 so $P(+ \mid D) = 0.98$; specificity 0.97 so $P(- \mid H) = 0.97$. Prevalence 1%: $P(D) = 0.01$, $P(H) = 0.99$. Also $P(+ \mid H) = 1 - 0.97 = 0.03$ and $P(- \mid D) = 1 - 0.98 = 0.02$.

   (a) First find $P(+)$:

   $$P(+) = P(+ \mid D)P(D) + P(+ \mid H)P(H) = 0.98(0.01) + 0.03(0.99) = 0.0098 + 0.0297 = 0.0395.$$

   Then Bayes' rule:

   $$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+)} = \frac{0.98 \cdot 0.01}{0.0395} = \frac{0.0098}{0.0395} \approx 0.248.$$

   So the chance the person is diseased given a positive test is about 24.8%.

   (b) First find $P(-)$:

   $$P(-) = P(- \mid H)P(H) + P(- \mid D)P(D) = 0.97(0.99) + 0.02(0.01) = 0.9603 + 0.0002 = 0.9605.$$

   Then

   $$P(H \mid -) = \frac{P(- \mid H)P(H)}{P(-)} = \frac{0.97 \cdot 0.99}{0.9605} = \frac{0.9603}{0.9605} \approx 0.9998.$$

   So the chance the person is healthy given a negative test is about 99.98%.

   (c) Even with a good test, when the disease is rare, there are many more healthy people than sick people, so the small false-positive rate applied to a huge healthy group can produce lots of positives that are actually false.

7. **Four doors, host opens only one goat door.**

   Let C be the event "your first choice is the car." Then

   $$P(C) = \frac{1}{4}, \qquad P(\overline{C}) = \frac{3}{4}.$$

   The host *always* opens exactly one goat door among the doors you did not choose (so he never reveals the car).

   Stay. If you stay, you win exactly when you initially chose the car:

   $$P(W \mid \text{Stay}) = P(C) = \frac{1}{4}.$$

   Switch. Condition on C vs. $\overline{C}$.

   - If C happens (probability $\frac{1}{4}$): your chosen door is the car. The host opens one goat door among the remaining 3 doors. The two unopened doors besides yours are both goats. If you switch, you must switch to a goat, so
     $$P(W \mid \text{Switch}, C) = 0.$$
   - If $\overline{C}$ happens (probability $\frac{3}{4}$): your chosen door is a goat. Among the other 3 doors, there is 1 car and 2 goats. The host must open a goat door (he has *two* goat doors available), leaving *two* unopened doors besides yours: one is the car and the other is a goat. Since the host opened only *one* door, switching does *not* force you onto the car: you switch to one of those two remaining unopened doors, and exactly one of them is the car. With no additional information, symmetry gives
     $$P(W \mid \text{Switch}, \overline{C}) = \frac{1}{2}.$$

Therefore,

$$P(W \mid \text{Switch}) = P(W \mid \text{Switch}, C)P(C) + P(W \mid \text{Switch}, \overline{C})P(\overline{C}) = 0 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}.$$

Answer:

$$P(W \mid \text{Stay}) = \frac{1}{4}, \qquad P(W \mid \text{Switch}) = \frac{3}{8}.$$

(So switching is better, but not as dramatic as the usual Monty Hall.)

8. **General n-door version (host opens $n-2$ goat doors).**

   Initially, your chance of picking the car is $\frac{1}{n}$, and your chance of picking a goat is $\frac{n-1}{n}$.

   - If your first choice is the car (probability $\frac{1}{n}$), then all other doors hide goats. The host opens $n-2$ goats, and the one remaining closed door must also be a goat. Switching loses.
   - If your first choice is a goat (probability $\frac{n-1}{n}$), then the car is among the other $n-1$ doors. The host opens $n-2$ goat doors *from those other doors*, leaving exactly *one* other closed door. It cannot be a goat (all goats among the other doors were opened), so it must be the car. Switching wins.

   Thus switching wins exactly in the case you initially picked a goat:

$$P(\text{win by switching}) = \frac{n-1}{n}.$$

Homework

1. A fair die is rolled once. Let $A = \{\text{even}\}$ and $B = \{\text{number} \geq 4\}$.
   (a) Compute $P(A \mid B)$.
   (b) Compute $P(B \mid A)$.
   (c) Are A and B independent? Justify.
2. A jar contains 5 red and 7 blue candies. You draw two candies *with* replacement.
   (a) Find $P(\text{both red})$.
   (b) Find $P(\text{second is red})$.
   (c) Are the two draws independent in this setup? Explain briefly.
3. The same jar (5 red, 7 blue), but you draw two candies *without* replacement.
   (a) Find $P(\text{both red})$.
   (b) Find $P(\text{second is red})$.
   (c) Are the two draws independent now? Explain briefly.
4. In a school, $P(S) = 0.4$ take Spanish, $P(A) = 0.3$ take Art, and $P(S \cap A) = 0.12$.
   (a) Compute $P(S \mid A)$ and $P(A \mid S)$.
   (b) Are S and A independent? Explain.
   (c) What is $P(S \cup A)$?
5. A medical test is used to detect a certain virus.
   - The probability that a person **has** the virus is 0.05.
   - If a person has the virus, the test shows **positive** with probability 0.9 and **negative** with probability 0.1 (a false negative).
   - If a person does **not** have the virus, the test still shows positive with probability 0.1 (a false positive) and negative with probability 0.9.

   (a) Find $P(\text{positive})$, the probability a random person tests positive.
   (b) Find $P(\text{has virus} \mid \text{positive})$, the probability that a person actually has the virus given that they tested positive.
   (c) Find $P(\text{no virus} \mid \text{positive})$, the probability that a person does not have the virus given a positive result.
   (d) Find $P(\text{has virus} \mid \text{negative})$, the probability that a person has the virus given a negative result.
6. Two urns:
   Urn I: $4R$, $1B$ Urn II: $1R$, $4B$.
   You choose an urn at random, then draw one ball.
   (a) Find $P(\text{red})$.
   (b) Given that the ball is red, what is $P(\text{it came from Urn I})$?
7. A coin is either fair ($P(H) = \frac{1}{2}$) with probability 0.6, or trick ($P(H) = \frac{3}{4}$) with probability 0.4. You pick one coin at random (according to these probabilities) and toss it three times, observing HHT.
   (a) Compute $P(\text{HHT})$.
   (b) Compute $P(\text{trick} \mid \text{HHT})$.

(c) If you toss one more time (a 4th toss), find $P(\text{H on 4th} \mid \text{HHT so far})$.

8. You roll two fair dice.

   (a) Given that the sum is 8, find $P(\text{a die shows 3})$.

   (b) Given that the product is even, find $P(\text{sum is 7})$.

   (c) Given that the product is even, find $P(\text{sum is 8})$.

   (d) Given that the maximum of the two results is 5, find $P(\text{both are at least 3})$.

9. Consider a factory with two machines producing bolts. Machine A makes 60% of all bolts and has a defect rate of 2%; Machine B makes 40% with defect rate 5%. A quality-control test flags defects correctly with probability 0.9 and falsely flags good bolts with probability 0.03.

   (a) Pick a random bolt and test it. It is flagged. Compute $P(\text{defective} \mid \text{flagged})$.

   (b) Compute $P(\text{made by A} \mid \text{flagged})$.

   (c) If the flagged bolt actually came from Machine B, what is $P(\text{that bolt was truly defective} \mid \text{flagged, from B})$?

10. **Biased first pick.** You are told that an over-eager contestant's first pick is correct with probability q (not $\frac{1}{3}$), because they have some (imperfect) clue. The host follows the classic rules otherwise.

    (a) Compute $P(\text{win by stay})$ and $P(\text{win by switch})$ in terms of q.

    (b) For which q (if any) does staying become better than switching?

*11. **Monty with uncertain host behavior (trickier).** There are 3 doors, 1 car. You pick one. With probability p the host behaves like classic Monty (always opens a goat and offers to switch); with probability $1-p$ the host picks one of the two *other* doors uniformly at random and opens it (this could reveal the car; if the car is revealed, the game ends with no switch offer). Suppose the host opens a door and shows a goat, and you *are* offered to switch.

   (a) For a given $p \in (0, 1]$, compute $P(\text{win by switching} \mid \text{goat opened and switch offered})$.

   (b) Determine the values of p for which switching still beats staying.