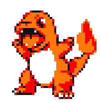
**Learning objectives:**

- ☐ identify the outcomes, events and sample space in a probability problem,
- ☐ use counting techniques to enumerate outcomes and build the sample space,
- ☐ compute the probability of an event by building a probability tree.

Our last major topic in this course is probability, which can be very tricky if you let your intuition take over. Solving probability problems can be much simpler if you follow a basic set of steps.

**Why do I need to know probability?**



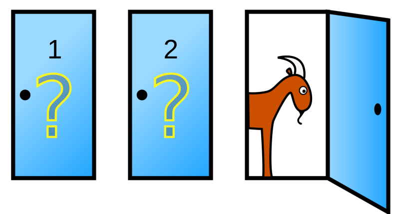
Probability is a very important topic in computer science! It has applications in algorithm design in which random choices in an algorithm can actually improve the performance over deterministic algorithms. Randomness is also useful for filtering noise in signal processing and for achieving security in cryptography applications.

**Example 1:**
The Monty Hall Problem is a famous problem that is based on the game show *Let's Make a Deal*. The problem is stated as follows:

*Suppose you're on a game show and you get to chose one of three doors. Whatever is behind the door you pick is yours to keep. Behind one door is a car and behind the other two doors are goats. You select a door. The host then reveals one of the other two doors, and always opts to reveal a door that hides a goat. You are now given the option to either stick with your original door, or switch doors.*

Do you stick with your original pick or switch doors? Why?

Your intuition might tell you that it doesn't matter what door you pick. You had a $1/3$ chance of picking the door with a car in the first place, so that doesn't change. However, the fact that the host *always* opts to open a door with a goat actually changes the probability.




The host always reveals a door with a goat. (image source)

## 1 Terminology: outcome, event and sample space

Before giving a solution to the Monty Hall problem, we need to develop some terminology.

**Definition 1.** *An **outcome** (also known as a **sample point**) consists of all the information about an experiment, including all the values of random choices made during the experiment.*

For example, an outcome of the Monty Hall problem could be any triple (1) pick door A, door B is revealed, stick with door A, (2) pick door B, door A is revealed, switch to door C, (3) pick door B, door C is revealed, switch to door A, etc.

**Definition 2.** *A **sample space** is the set of <u>all</u> possible outcomes.*

The sample space for the Monty Hall consists of all possible triples of (a) door selection, (b) door revelation and (c) decision to switch or stick with original door.

**Definition 3.** *An **event** is a subset of the sample space.*

There are two possible events for the Monty Hall problem: (1) you win the car or (2) you adopt a goat (this isn't exactly losing because goats are pretty cool).

## 2 A winning strategy for solving probability problems

It would be nice if we could solve probability problems with some kind of algorithm, right? Well, leave your intuition behind and just follow these steps!

1. Find the sample space.
2. Identify the events of interest.
3. Determine outcome probabilities: assign edge weights as the probabilities that a decision is made. Multiply edge weights as you traverse the probability tree to compute the probability of an outcome.
4. Compute event probabilities by adding the probabilities of outcomes that fall into the desired event.

**Definition 4.** *If all outcomes in sample space $S$ are equally likely, then the probability of event $E$, $E \subseteq S$ is*

$$P(E) = \frac{|E|}{|S|}.$$

*If the outcomes occur with unequal probabilities, then the probability of event $E$, $E \subseteq S$ is*

$$P(E) = \sum_{e \in E} p(e),$$

*where $p(e)$ is the probability of event $e \in E$, which can be calculated by multiplying the edge weight of the outcome $e$ as you traverse the probability tree.*

**What about the probability that something *does not* happen?**

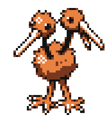
It is sometimes useful to compute the probability that an event $E$ *does not* occur. Since all probabilities must add up to 1, we can calculate the probability that $E$ does not occur is

$$1 - P(E).$$

## 3 Always switch doors!!!

Whether to switch doors in the Monty Hall problem has created a lot of debate. Luckily, we can approach the problem from a mathematical standpoint. Consider the experiment in which (1) a door is chosen for the car, (2) you pick a door, (3) the host reveals a door containing a goat and (4) you switch doors. The full probability tree that lists all possible outcomes is drawn in Figure 1.

**Wait! State your assumptions!**



As in any math problem, be sure you state your assumptions. In the Monty Hall problem, we are assuming

- The car is behind each door with probability $1/3$.
- No matter where the car is, the player picks each box with probability $1/3$.
- No matter where the car is, if the host has a choice, they pick each door with probability $1/2$.

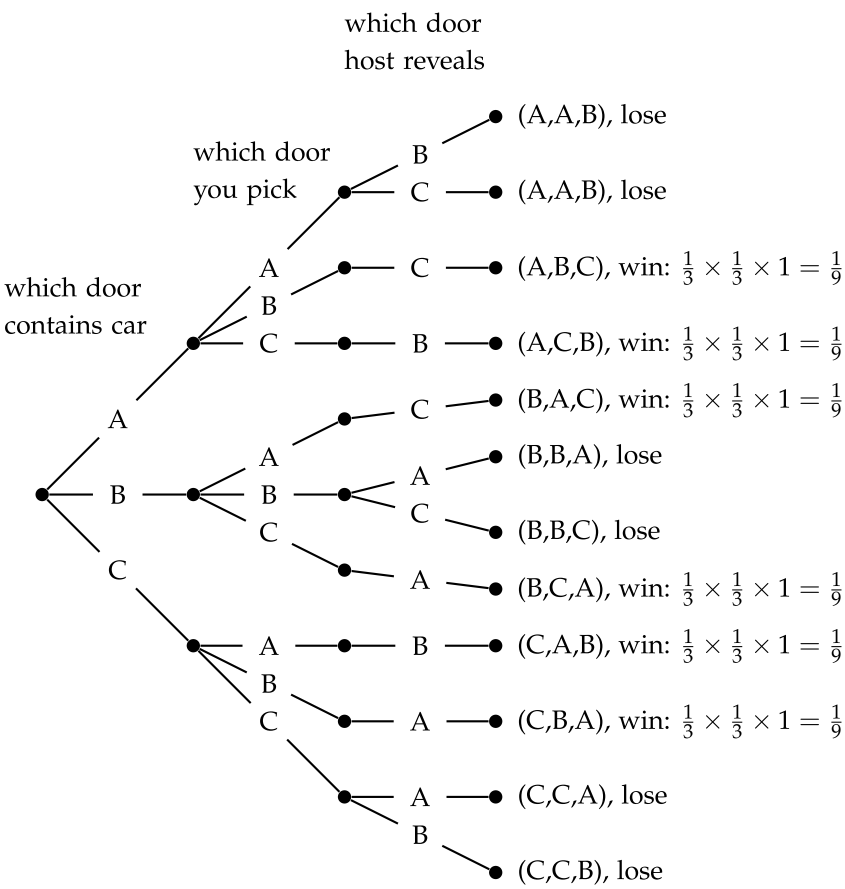



**Figure 1:** Probability tree for the Monty Hall problem.

Under our assumptions, the probability of the car being behind a particular door is $1/3$. There is also a $1/3$ probability you pick a particular door. The door revealed by the host depends on how many possibilities there are. Note that if the car is behind door A and you pick door A, then the host can choose either doors B or C to reveal. However, if the car is behind door A and you pick door B, then the host will only pick door C to reveal a goat. The same applies if the car is behind door A and you pick door C: only door C can reveal a goat.

To calculate the probability of winning by switching doors, we just need to find all the sample points that give a winning event. Under the assumption that you *always* switch doors, a winning strategy is one in which you switch to the door that the car is hidden behind. For example, if the car is behind door A, you pick door A, door B is revealed, meaning you switch to door C means you lose. However, if the car is behind door A, you pick door B, door C is revealed, meaning you switch to door A means you win! These winning and losing events are listed alongside all possible outcomes.

To calculate the probability that you win when switching doors, we need to add up the probabilities of all winning outcomes. There are 6 possible outcomes that create a winning scenario, each with probability $\frac{1}{9}$. Therefore, the probability of winning when you *always* switch doors is $6 \times \frac{1}{9} = \frac{2}{3}$. If you switch doors, then you will win two-thirds of the time! This is much better than one third, which is what your intuition might tell you!

**Probability of losing?**

It's a good idea to check your understanding and compute the probability of a losing event (one in which you switch doors and lose). Check that you get $P(\mathsf{L}) = 1 - P(\mathsf{W}) = \frac{1}{3}$.

## 4 Examples

That was a pretty complicated example but a good way of illustrating all the steps you should take when solving a probability problem. Draw the tree! In some cases you might not be able to draw the entire tree, but you should draw enough of it so that you observe a pattern and then can compute the probability of a winning event using this pattern and some of the counting techniques we previously used in the course.

Let's do a few simpler examples now.

**Example 2:**
A standard deck of 52 cards has 4 suits (hearts, diamonds, clubs, spades) and 13 values (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K). A hand is a set of cards one is dealt. In poker, a flush is a hand with all cards of the same suit. A straight flush is a hand in which all cards are in a row, all cards have the same suit. For example, a flush could be the (A, 2, 3, 4, 5) of hearts or the (10, J, Q, K, A) of clubs, but not the (Q, K, A, 2, 3) of spades (the ace A can come before a 2 or after the king K, but cannot be both). What is the probability of a 5-card straight flush?

**Solution:**
First let's calculate the size of the sample space $S$. We need to count the number of ways we can pick 5 cards from a deck of 52 cards, which is $\binom{52}{5}$. Now we need to calculate the size of the event space. Let's just focus on a single suit: in how many ways can we create a straight flush for a single suit? We could create a flush starting at the A, 2, 3, 4, 5, 6, 7, 8, 9 or 10, hence 10 possible options. Thus there are 10 possible straight flushes per suit and a total of $|E| = 4 \times 10$ straight flushes for a deck (for all 4 suits). The probability of a straight flush is then

$$P(\text{straight flush}) = \frac{|E|}{|S|} = \frac{40}{\binom{52}{5}} \approx 0.00154\%.$$

**Do I need a calculator?**

You can leave your answer in terms of $C(n,k)$ or $P(n,k)$ instead of calculating all those factorials (and possibly making a mistake) :)

**Example 3:**
On a lottery card, you pick 6 numbers out of a set of 55. You win if these 6 numbers match the 6 numbers drawn by the lottery. What is the probability you win if (a) the order of the numbers you pick does not matter and (b) the order of the numbers matters?

**Solution:**
The first thing we need to do is count how many possible lottery tickets there are (the size of the sample space). If order does not matter, then there are $C(55,6) = \binom{55}{6}$ since this is the number of ways to pick 6 numbers from a set of 55. The event we are interested in is the one in which you pick the same 6 numbers drawn by the lottery. The cardinality of the event space is then 1. Therefore the probability of picking a winning card is $1/C(55,6) \approx 3.4 \cdot 10^{-8} = 3.4 \cdot 10^{-6}\%$ (about 34 in a trillion!). If the order does matter, then the size of the sample space is bigger, and becomes the number of ways we can pick a certain set of numbers. First, pick one number (55 choices), which leaves 54 to choose from for the second number. Therefore, there are a total of $55 \cdot 54 \cdot 53 \cdot 52 \cdot 51 \cdot 50$ possible choices - note this is $P(55,6)$. The size of the event space is still 1, so the probability of picking a winning number when order matters is $1/P(55,6) \approx 4.79 \cdot 10^{-11} = 4.79 \cdot 10^{-9}\%$.