Independence

We have already considered several cases where we have multiple outcomes (e.g., two rolls of a die, several tosses of a coin, a pair of variables (x, y) drawn from a unit-square). In each of these cases we have actually been inherently assuming that one outcome gives us no information about the other (the first roll of the die or toss of the coin has no impact on the outcome for the next roll or toss, and the value of x tells us nothing about the value of y). We can formalize this notion as follows: two events A and B are **independent** if

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B).$$

That is, if two events are independent, we can compute the probability that they both occur simultaneously by computing both of their individual probabilities and combining them (with a product).

Here are other examples of events that are independent:

- The outcomes of consecutive rolls of the dice.
- Whether you are over 6 feet tall, and whether the person sitting next to you is over 6 feet tall.
- Whether two randomly chosen people in the classroom had a dog growing up.
- etc.

Here are examples of events that are not independent:

- The sum of two die rolls and the value of the first roll.
- The first card that you draw out of a well shuffled deck is a spade, and the second card you draw is a spade.
- Whether you are over 6 feet tall, and whether your father is over 6 feet tall.
- Whether Apple stock goes up tomorrow, and whether Google stock goes up tomorrow.

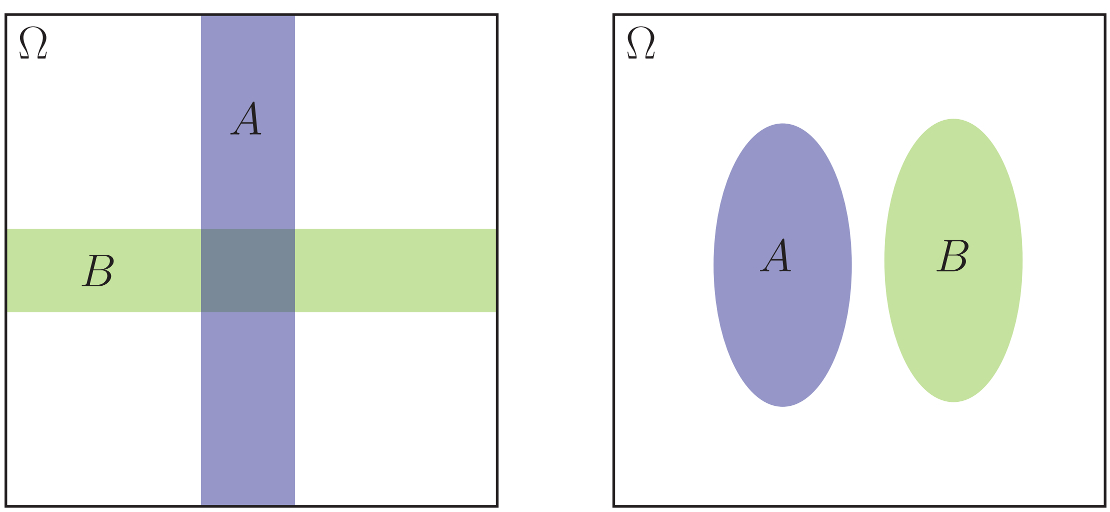
Important note: Disjointness does not imply independence!

In fact, the opposite is true; if A and B are disjoint events (i.e. $A \cap B = \emptyset$)with positive probabilities, then

$$\mathrm{P}(A \cap B) = \mathrm{P}(\emptyset) = 0 \neq \mathrm{P}(A)\,\mathrm{P}(B)\,.$$

In this case, knowing that A occurred tells you something very tangible about B, namely that it could *not* have occurred.

Independent events | Disjoint events



$P(A \cap B) = P(A) \cdot P(B)$ | $P(A \cap B) = P(\emptyset) = 0$

Examples. Consider an experiment involving two successive rolls of a fair (6-sided) die. Which pairs of events below are independent?

1. $A = \{\text{first roll is a } 3\}$,
   $B = \{\text{second roll is a } 6\}$

2. $A = \{\text{first roll is a } 1\}$,
   $B = \{\text{the sum of two rolls is } 2\}$

3. $A = \{\text{first roll is even}\}$,
   $B = \{\text{the sum of the two rolls is even}\}$

4. $A = \{\text{the first roll is even}\}$,
   $B = \{\text{the product of the two rolls is even}\}$

5. $A = \{\text{at least one roll was a } 2\}$,
   $B = \{\text{the sum of the rolls is } 5\}$

6. $A = \{\text{first roll is a } 2\}$,
   $B = \{\text{the sum of the two rolls is } 7\}$
   (tread carefully)

Exercise: We generate two bits B_1 and B_2 independently at random with

$$\mathrm{P}(B_1 = 0) = \mathrm{P}(B_1 = 1) = \frac{1}{2},$$

and similarly for B_2. Set

$$Z = B_1 \oplus B_2, \qquad \text{where } \oplus \text{ is the `xor' operator.}$$

Are the events $\{B_1 = 1\}$ and $\{Z = 1\}$ independent?

Exercise: Suppose that a point (x, y) is chosen from the unit square $[0, 1] \times [0, 1]$ in the plane according to the uniform law. Consider the events

$$\begin{aligned} A &= \{x + y \leq 1\} \\ B &= \{\max(x, y) \leq 0.5\} \end{aligned}$$

Are A and B independent?

Exercise: Suppose that a point (x, y) is chosen from the unit square $[0, 1] \times [0, 1]$ in the plane according to the uniform law. Consider the events

$$A = \{x + y \leq 1\}$$
$$B = \{\{\max(x, y) \leq 0.5\} \ \cup \ \{\min(x, y) \geq 0.5\}\}$$

Are A and B independent?

Exercise: A number D is chosen in $[0, 1]$ according to the uniform probability law. Let

$$D = 0.d_1 d_2 d_3 d_4 \cdots$$

be its decimal expansion. Let A and B be the events

$$A = \{d_1 \text{ is even}\},$$
$$B = \{d_2 \text{ is even}\}.$$

Are A and B independent?

Independence of multiple events

Events $A_1, A_2, \ldots, A_n$ are independent if and only if the probability of an intersection of the A_is is the product of the individual probabilities.

- $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i)\,\mathrm{P}(A_j)$ for all i, j.
  (This is called *pairwise independence*.)
- $\mathrm{P}(A_i \cap A_j \cap A_k) = \mathrm{P}(A_i)\,\mathrm{P}(A_j)\,\mathrm{P}(A_k)$ for all i, j, k.

  $\vdots$

- $\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\,\mathrm{P}(A_2) \cdots \mathrm{P}(A_n)$

Note: Pairwise independence does not imply independence.

To see this, let $s_1 = \pm 1$ and $s_2 = \pm 1$ be independent signs with

$$\mathrm{P}(s_1 = 1) = \mathrm{P}(s_1 = -1) = 1/2,$$

and similarly for s_2. Set $x = s_1 s_2$, and consider the events

$$\begin{aligned} A &= \{s_1 = 1\} \\ B &= \{s_2 = 1\} \\ C &= \{x = 1\}. \end{aligned}$$

Then it easy to see that while A and B are independent, A and C are independent, and B and C are independent, but A, B, C are clearly not independent:

$$\mathrm{P}(A \cap B \cap C) \neq \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C),$$

since knowing two of s_1, s_2, x uniquely determines the third.

It is easy to calculate the probability of the combination of a bunch of independent events.

Exercise: M&Ms come in six colors: red, blue, green, yellow, brown, and orange. What is the probability that the next six M&Ms I pull out of this bag are all green?

Exercise: If I am an 85% free throw shooter, what is the probability that I make my next three free throws?

Exercise: If I open a book to three different pages at random, and place my finger down on the page at a random location, what is the probability that the closest letters spell out 'r','u','n'?

Breaking down problems

One technique we have used to compute the probability of an event is enumerate (i.e., write an explicit list of) all possible outcomes of the experiment. If the outcomes are equally likely, then the probability of the event occurring is the ratio of the number of "successes" (i.e. number of outcomes leading to the event) to the total number of possible outcomes. If the elementary outcomes have different probabilities, then the probability of the event occurring is the sum of the probabilities of the outcomes that correspond to a "success". This is the method we have used the most so far.

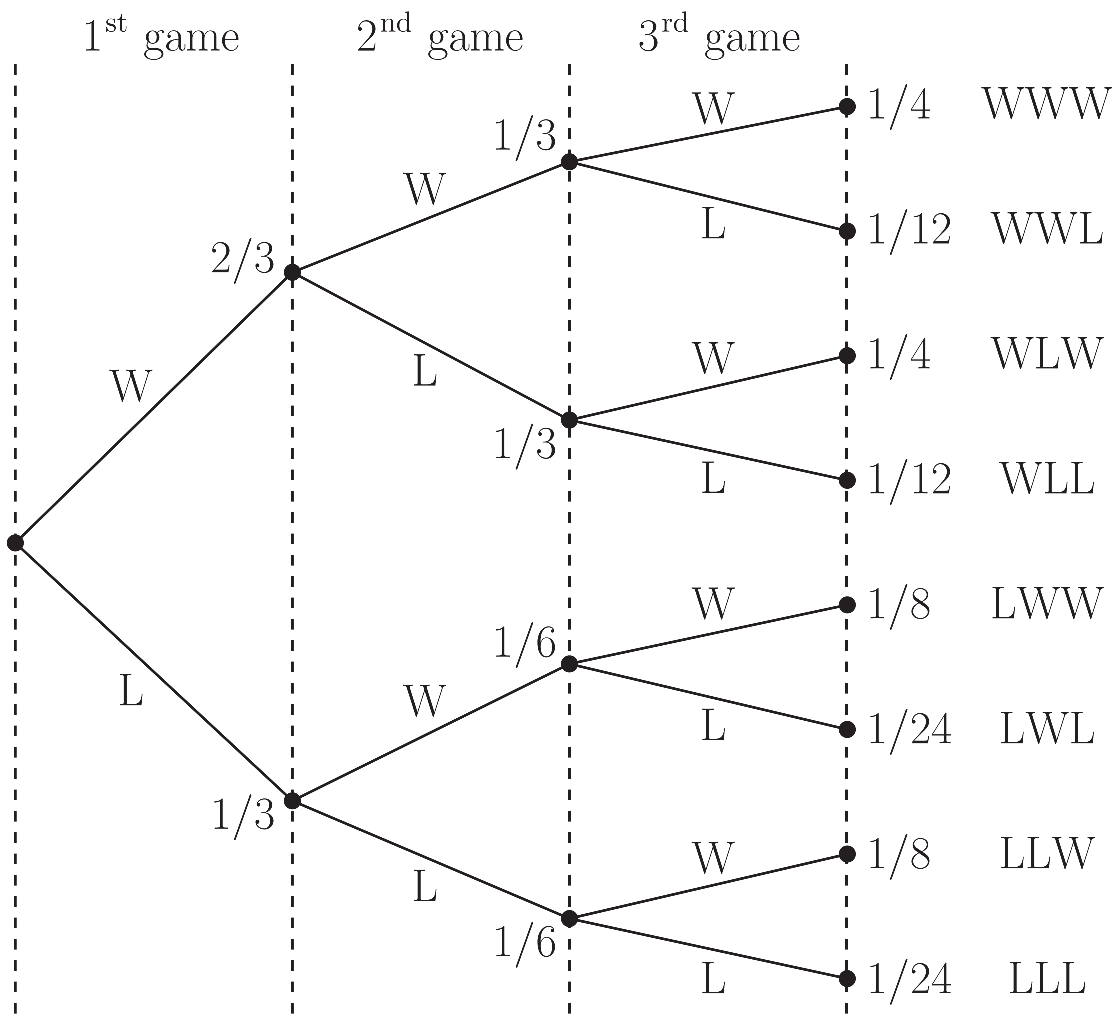
When dealing with a **sequence** of outcomes, it can be useful to break up the calculation of a particular probability into more manageable pieces by following along the history of the sequence of outcomes. We have implicitly done this a few times already, but a couple of examples will hopefully clarify how this technique can be used to help organize and sometimes simplify our calculations.

Example. You are in Las Vegas and decide to bet \$10 on each of 3 football games. For each game, if you win, you will double your money (so that if you predict correctly you make \$10 and if you are wrong you lose \$10). Your knowledge gives you different degrees of confidence. You estimate your chances of correctly predicting each outcome as:

$$\begin{aligned} \mathrm{P}\left(1^{\text{st}} \text{ game correct}\right) &= 2/3 \\ \mathrm{P}\left(2^{\text{nd}} \text{ game correct}\right) &= 1/2 \\ \mathrm{P}\left(3^{\text{rd}} \text{ game correct}\right) &= 3/4 \end{aligned}$$

What is the probability that you come out ahead? (That is, what is the probability that you win at least two of the bets?)

You can think about this by drawing a tree with 3 levels, where the edges depict you winning or losing certain bets, and we label the nodes with the probability that we arrive there.



Thus, our probability of winning at least two of the bets is $\frac{1}{4} + \frac{1}{12} + \frac{1}{4} + \frac{1}{8} = \frac{17}{24} \approx 0.71$.

Now, what happens if we only have \$10 to wager (so that if we lose the first bet, we no longer have enough money to continue)?

In this example, each outcome is independent of the other, but because they occur with different probabilities (as opposed to a sequence of fair coin tosses) and since the order in which they occur can be important, the sequential method can be a useful way to keep track of the relevant probabilities. Ultimately, this is not really dramatically different from just enumerating the possible outcomes and adding up the relevant possibilities, but it often provides a useful way to organize the calculations (and can also be very useful in situations where direct enumeration is not possible).

As we will see in the next section, this type of sequential approach can also be very handy when the sequence of outcomes are not independent.