Explainer: Tree Diagrams and Conditional Probability

In this explainer, we will learn how to use tree diagrams to calculate conditional probabilities.

When working out probabilities for more than one event, it can be helpful to illustrate the problem using a tree diagram. These can be particularly useful when we are looking at conditional probabilities. (Recall that the conditional probability of event B is the probability of event B given that event A has occurred and is written as $P(B \mid A)$.)

In our first example, we will use a tree diagram to calculate probabilities for selection without replacement.

Example 1: Compound Probability, Taking Two Balls from a Bag without Replacement

A bag contains 22 red balls and 15 black balls. One red ball is removed from the bag and then a second ball is drawn at random. Find the probability that the second ball is black. Give your answer to three significant figures.

Answer

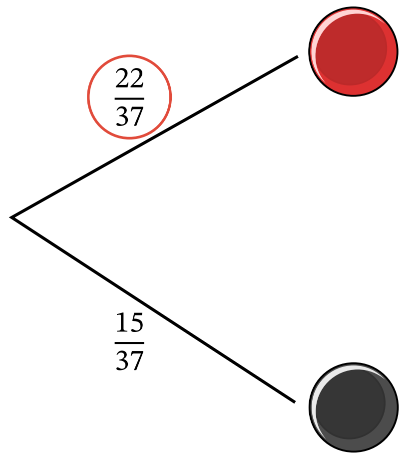
To find the probability that the second ball drawn is black, we note that since the second ball is drawn without replacing the first, the probability of the second being black is dependent on what happens in the first selection. And to calculate our second probability, we first need to work out the probability that the first ball selected is red.

To begin, there are 22 red balls and 15 black ones, so there are $22 + 15 = 37$ balls in total. The probability of drawing a red ball from the bag on our first pick is, therefore,

$$P(\text{Red}) = \frac{\text{number of red balls}}{\text{total number of balls}} = \frac{22}{37}.$$

We can draw the first branches in our tree diagram using this information.

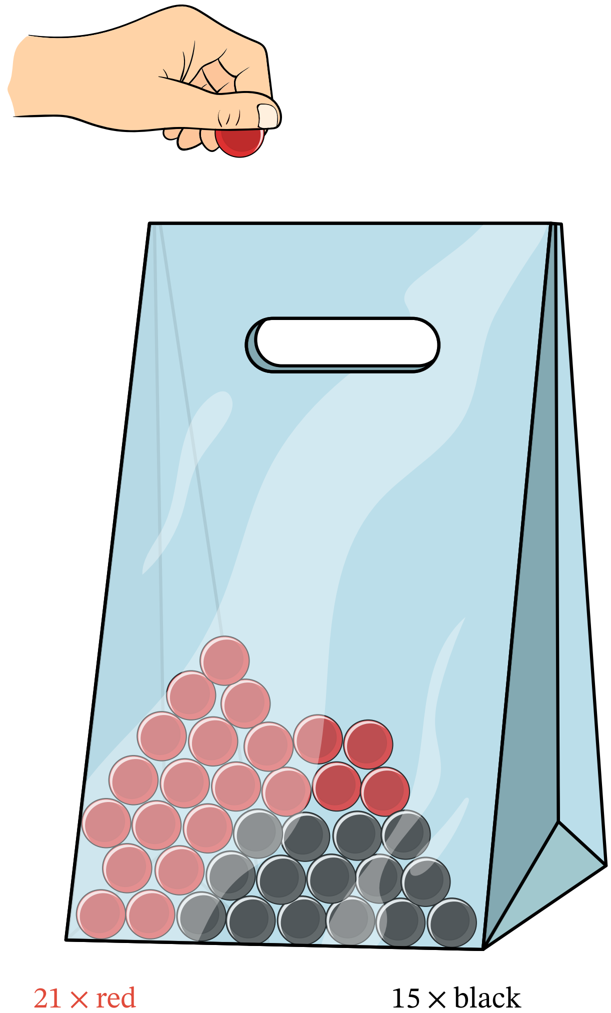
First ball



$$\frac{22}{37} + \frac{15}{37} = 1$$

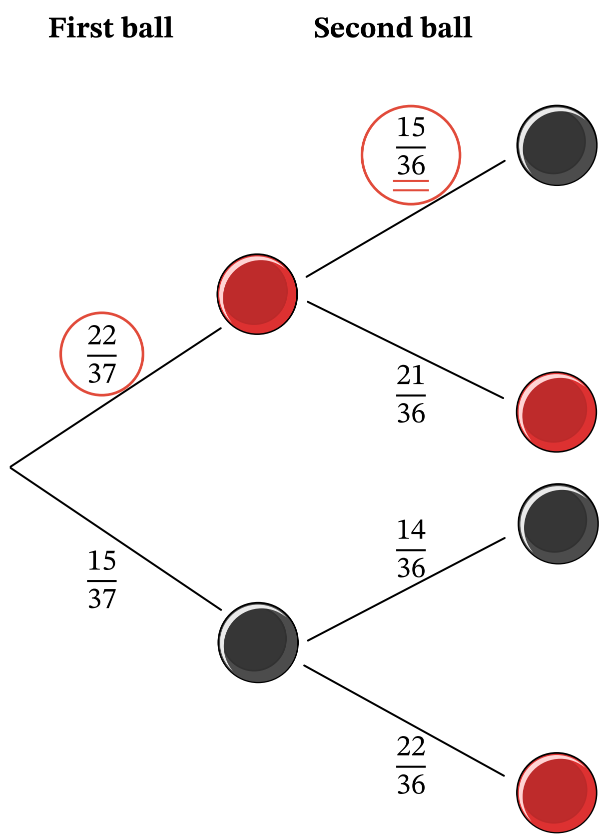
Notice that the probabilities, which are on the two branches, add up to 1. This must be true for each set of branches in the tree diagram since, at every stage, all possible outcomes are covered by each set of branches.

Since we are keeping the first ball out of the bag, there is one less ball in the bag in total, so there are now 36 balls in the bag. Of those 36, there is one less red, so 21 red balls remain in the bag.



$$21 + 15 = 36$$

From each of the first branches, we now have two possible outcomes for our second selection. The second ball could be either red or black. Let us complete our tree diagram with all possible outcomes for the second selection.



$$P(\bullet \cap \bullet) = \frac{22}{37} \times \frac{15}{36} = \frac{330}{1{,}332} \approx 0.248$$

The probability that the second ball is black is found by multiplying the probabilities on the branches corresponding to "first ball red" and "second ball black":

$$\begin{aligned} P = (\text{Red} \cap \text{Black}) &= \frac{\text{red balls initially}}{\text{total balls initially}} \times \frac{\text{black balls}}{\text{total balls left}} \\ &= \frac{22}{37} \times \frac{15}{36} \\ &= \frac{22 \times 15}{37 \times 36} \\ &= \frac{330}{1{,}332} = 0.248 \text{ to 3 significant figures.} \end{aligned}$$

The probability that the second ball drawn is black, having already drawn a red ball, is therefore 0.248. We can say that there is approximately a 25% chance of selecting a red ball and then a black ball (since $0.248 \times 100\% = 24.8\%$).

Note

The probability of drawing a black ball in our second selection is dependent on what we selected first, and in our calculation we have actually used the formula for dependent events:

$$P(B \cap R) = P(B \mid R) \times P(R).$$

The probability $P(B \mid R) = \dfrac{15}{36}$ is the conditional probability of selecting a black ball given that a red ball has already been taken from the bag. $P(R) = \dfrac{22}{37}$ the probability that the first ball selected is red. Using the formula, we therefore have

$$\begin{aligned} P(R \cap B) &= P(B \mid R) \times P(R) \\ &= \frac{15}{36} \times \frac{22}{37} \approx 0.248. \end{aligned}$$

Before looking at some other examples, let us remind ourselves of some of the rules of probability that we will need.

Some Probability Rules and Definitions

For any event A, if $P(A)$ is the probability of event A occurring, we have the following:

Rule 1

$$0 \leq P(A) \leq 1$$

Rule 2

Total probability: the sum of the probabilities of all possible outcomes is equal to 1 (or 100%).

The complement of event A, written as $\overline{A}$, refers to everything that is **not** A.

Rule 3

$$P\left(\overline{A}\right) = 1 - P(A)$$

Note: The complement of event A is sometimes also written as A'.

Two events, A and B, are

- **independent** if the fact that A occurs does not affect the probability of B occurring,
- **dependent** if the fact that A occurs does affect the probability of B occurring.

Rule 4

For **independent** events, $P(A \cap B) = P(A) \times P(B)$.

Rule 5

If two events are not independent, however,

$$P(A \cap B) = P(A \mid B) \times P(B),$$

where $P(A \mid B)$ is the conditional probability for event A, given that event B has occurred. If the events A and B are independent,

$$P(A \mid B) = P(A)$$

and

$$P(B \mid A) = P(B).$$

In our next example, we will look at conditional probability applied to the weather, using a tree diagram.

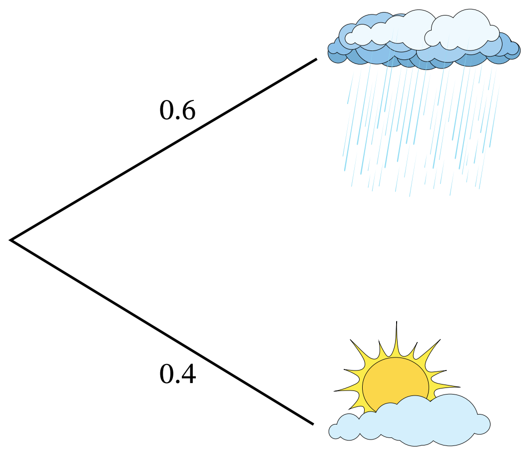
Example 2: Conditional Probability Applied to the Weather Using a Tree Diagram

The probability that it rains on a given day is 0.6. If it rains, the probability that a group of friends play football is 0.2. If it does not rain, the probability that they play football rises to 0.8.

1. Work out the probability that it rains on a given day and the friends play football.
2. Work out the probability that it does **not** rain on a given day and the friends play football.
3. What is the probability that the friends will play football on a given day?

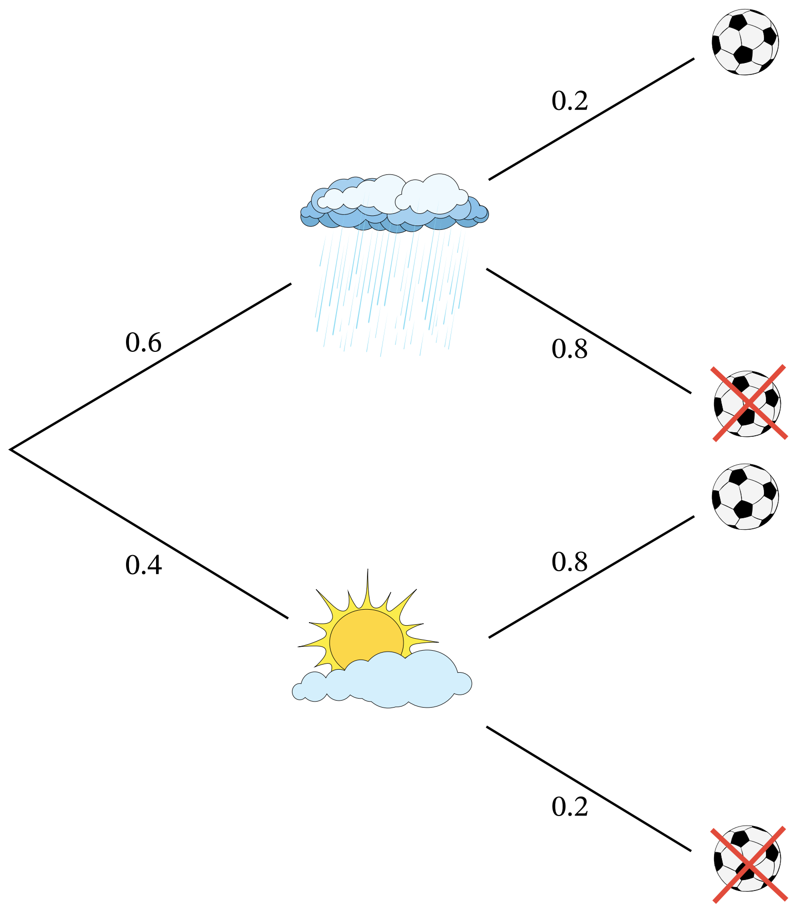
Answer

Since we have a small number of outcomes, Rain/No Rain and Football/No Football, we can easily represent the whole set of outcomes on a tree diagram. The outcomes related to football are dependent on whether or not it rains, so our first set of branches should cover whether or not it rains. We know that the probability of rain is 0.6, so the probability that it does not rain, $P\left(\overline{\text{Rain}}\right)$, is $1 - P(\text{Rain}) = 1 - 0.6 = 0.4$, and we can attach these probabilities to the relevant branches.



The second sets of branches covering football-related outcomes will lead off from these two branches and, as we did for Rain/No Rain, we can work out the probabilities for Football/No Football from the information we have.

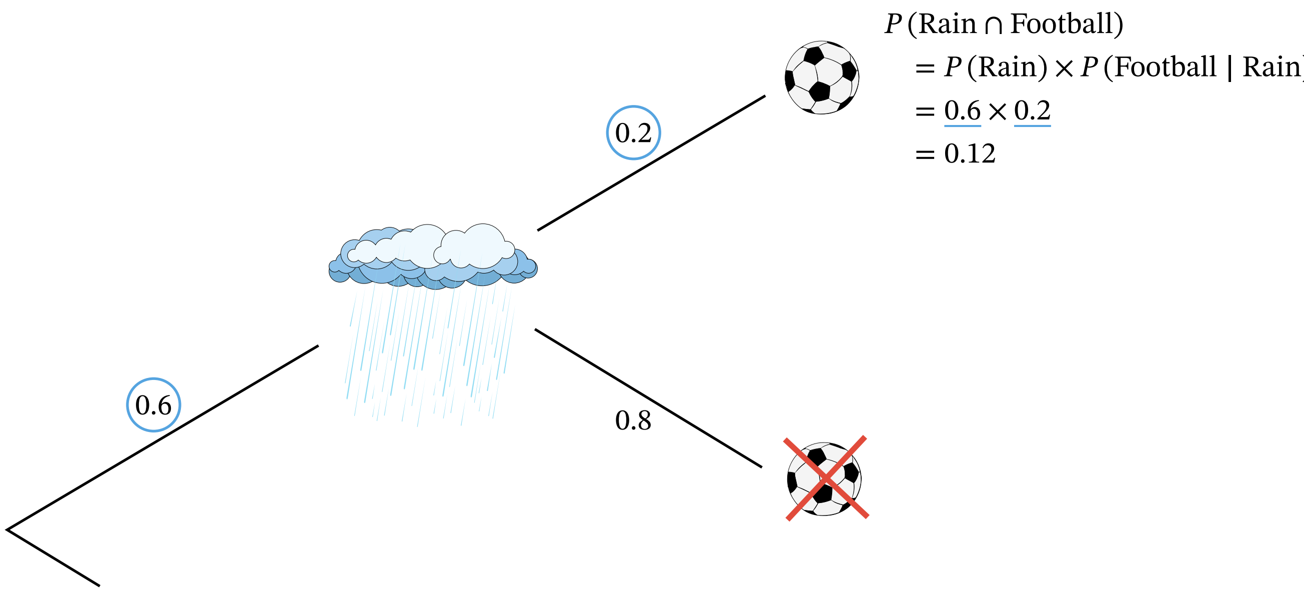
We know if it rains the probability the friends play football is 0.2, so the probability of not playing football if it rains is $1 - 0.2 = 0.8$. Similarly, if it does not rain, the probability that they play football is 0.8, so the probability that they do not play football if it does not rain is $1 - 0.8 = 0.2$. With this information, we can fill in the new branches and probabilities on our tree diagram.



Now that we have our tree diagram, we can calculate the required probabilities.

Part 1

To work out the probability that it rains on a given day and the friends play football, we multiply the probabilities on the branches along the top of the diagram, corresponding to Rain and Football, highlighted below.

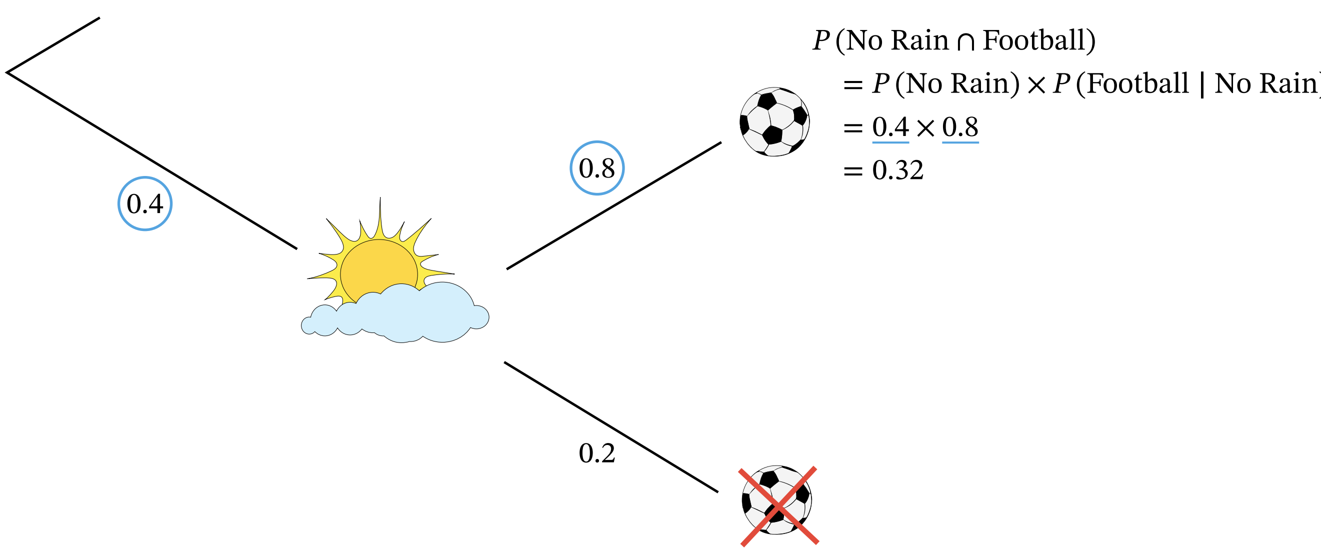


The probability that it rains and the friends play football is, therefore,

$$\begin{aligned} P(\text{Rain} \ \cap \ \text{Football}) &= P(\text{Rain}) \times P(\text{Football} \ | \ \text{Rain}) \\ &= 0.6 \times 0.2 \\ &= 0.12. \end{aligned}$$

Part 2

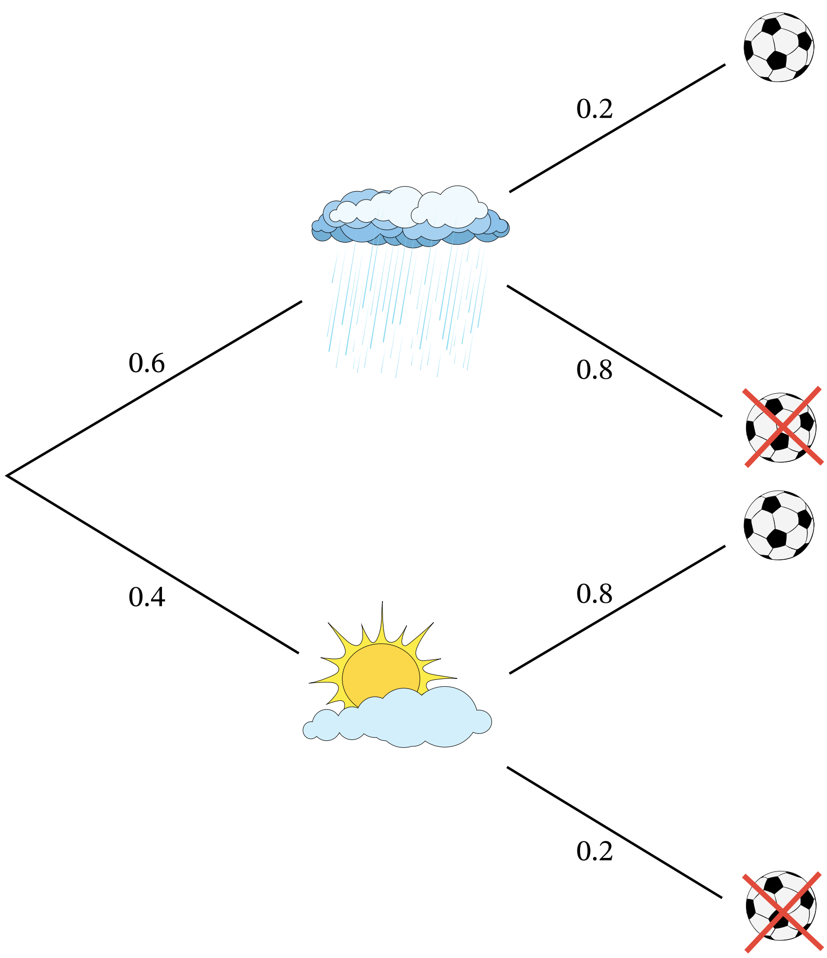
To work out the probability that it does not rain on a given day and the friends play football, we multiply the probability on the branch for No Rain, with the probability of Football on the branch leading from No Rain.



The probability that it does not rain on a given day and the friends play football is therefore 0.32. That is, there is a 32% chance that it will not rain and the friends will play football.

Part 3

To find the probability that the friends will play football on a given day, we must consider every combined outcome where the result is "playing football." This means adding the probability that it rains and they play football to the probability that it does not rain and they play football.



$$\begin{aligned} P(\text{Rain} \cap \underline{\text{Football}}) &= P(\text{Rain}) \times P(\text{Football} \mid \text{Rain}) \\ &= 0.6 \times 0.2 \\ &= \underline{0.12} \end{aligned}$$

$$\begin{aligned} P(\text{No Rain} \cap \underline{\text{Football}}) &= P(\text{No Rain}) \times P(\text{Football} \mid \text{No Rain}) \\ &= 0.4 \times 0.8 \\ &= \underline{0.32} \end{aligned}$$

$$\begin{aligned} P(\text{Football}) &= P(\text{Rain} \cap \text{Football}) + P(\text{No Rain} \cap \text{Football}) \\ &= 0.12 + 0.32 \\ &= 0.44 \end{aligned}$$

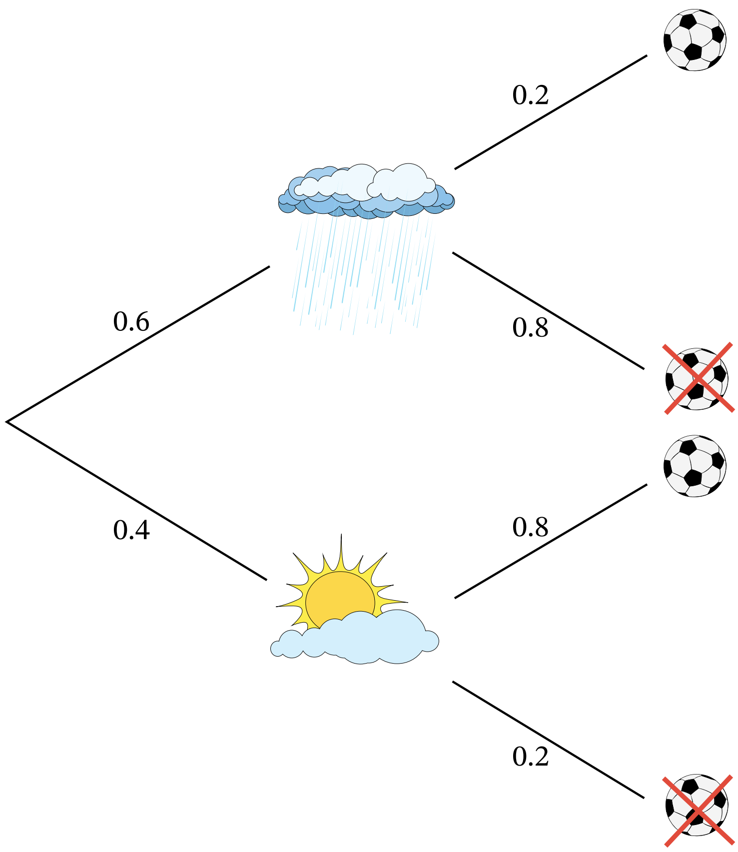
The probability that the friends play football on a given day is, therefore,

$$\begin{aligned} P\,(\text{Football}) &= P\,(\text{Rain} \cap \text{Football}) + P\,(\text{No Rain} \cap \text{Football}) \\ &= 0.12 + 0.32 \\ &= 0.44. \end{aligned}$$

We can say that there is a 44% chance that the friends will play football on any given day.

Note

The sum of the probabilities for every possible outcome, combined, is equal to 1. This should always be the case.



$$
\begin{aligned}
&P(\text{Rain} \cap \text{Football})\\
&\quad = P(\text{Rain}) \times P(\text{Football} \mid \text{Rain})\\
&\quad = 0.6 \times 0.2\\
&\quad = 0.12
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{Rain} \cap \text{No Football})\\
&\quad = P(\text{Rain}) \times P(\text{No Football} \mid \text{Rain})\\
&\quad = 0.6 \times 0.8\\
&\quad = 0.48
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{No Rain} \cap \text{Football})\\
&\quad = P(\text{No Rain}) \times P(\text{Football} \mid \text{No Rain})\\
&\quad = 0.4 \times 0.8\\
&\quad = 0.32
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{No Rain} \cap \text{No Football})\\
&\quad = P(\text{No Rain}) \times P(\text{No Football} \mid \text{No Rain})\\
&\quad = 0.4 \times 0.2\\
&\quad = 0.08
\end{aligned}
$$

Sum of probabilities

$$
\begin{aligned}
&= P(\text{Rain} \cap \text{Football}) + P(\text{Rain} \cap \text{No Football}) + P(\text{No Rain} \cap \text{Football}) + P(\text{No Rain} \cap \text{No Football})\\
&= 0.12 + 0.48 + 0.32 + 0.08\\
&= 1
\end{aligned}
$$

When using a tree diagram to illustrate probabilities, it should always be the case that

- the sum of the probabilities for each set of branches should equal 1,
- the sum of the probabilities of all the final outcomes should also equal 1.

Our next example shows how conditional probabilities and tree diagrams can be used in situations where a “false positive” might occur.

Example 3: Conditional Probabilities, Tree Diagrams, and False Positives

It is a little known fact that drugs have been used to enhance performance in sports since the original Olympic Games (776 to 393 BC). In fact, the origin of the word "doping" is thought to come from the Dutch word "doop," which is a type of opium juice used by the ancient Greeks.

Currently, in all major sporting events, drug testing has become standard practice. However, it is known that not all those who test positive for drugs have actually taken drugs. This is called the false positive effect. Similarly, there is a false negative effect, that is, that someone who tests negative for drugs has actually taken drugs.

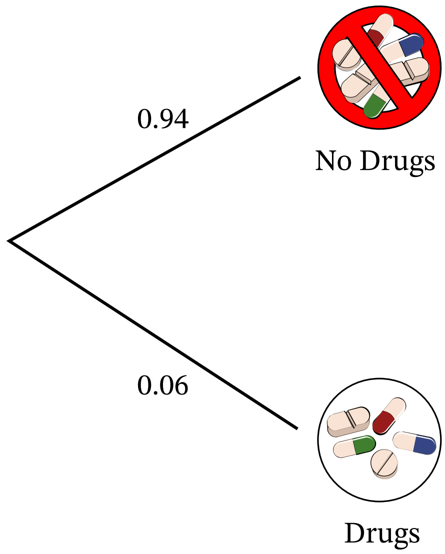
In 2003, after anonymous testing of almost 1,500 players, the MLB (Major League Baseball) announced that approximately 6% of MLB players used performance enhancing drugs. There was, however, a 5% chance that those who tested positive had not taken drugs and a 10% chance that those who had taken drugs tested negative.

1. Find the probability that an MLB player chosen at random had **not** taken drugs and tested positive.
2. Find the probability that an MLB player chosen at random had taken drugs and tested positive.
3. Find the probability that an MLB player chosen at random had positive test results.

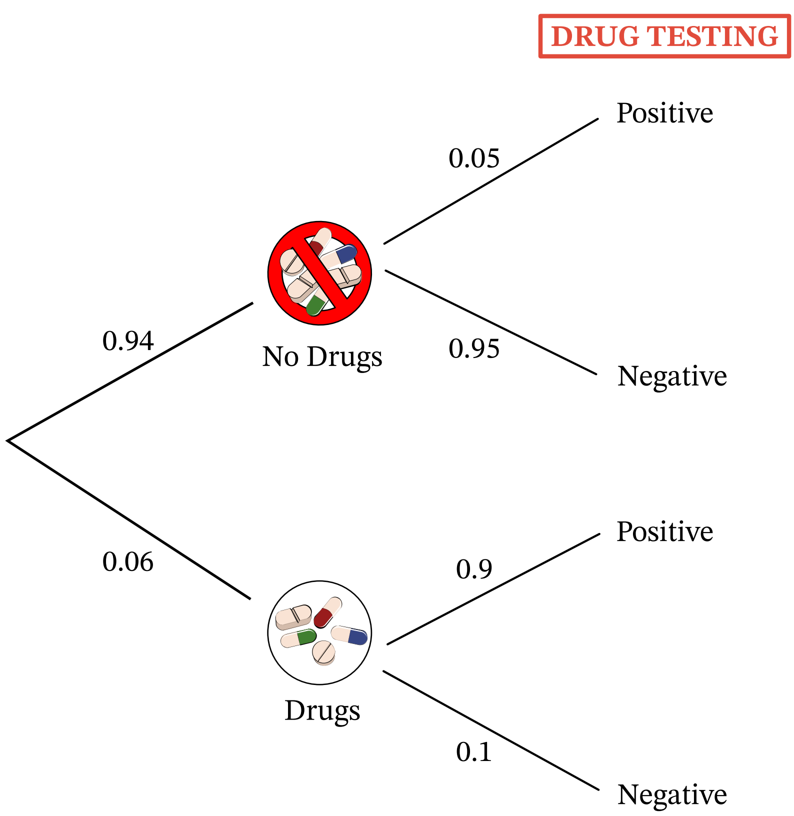
Answer

Our situation is that an MLB player chosen at random had either taken drugs or had not taken drugs and that, either way, they were given a drug test with either positive or negative results. Our first step is to draw a tree diagram using the information we have. We can then use this to work out the required probabilities.

The first branches illustrate the outcomes "Drugs" and "No Drugs," that is, whether an MLB player took drugs or not. We know that approximately 6% of MLB players took drugs, so the probability that an MLB player did not take drugs is 0.94 (since $6\% = 0.06$ as a probability and $1 - 0.06 = 0.94$).



Our next step is to add branches to each of the outcomes, No Drugs and Drugs, for the test results, which were either positive or negative.

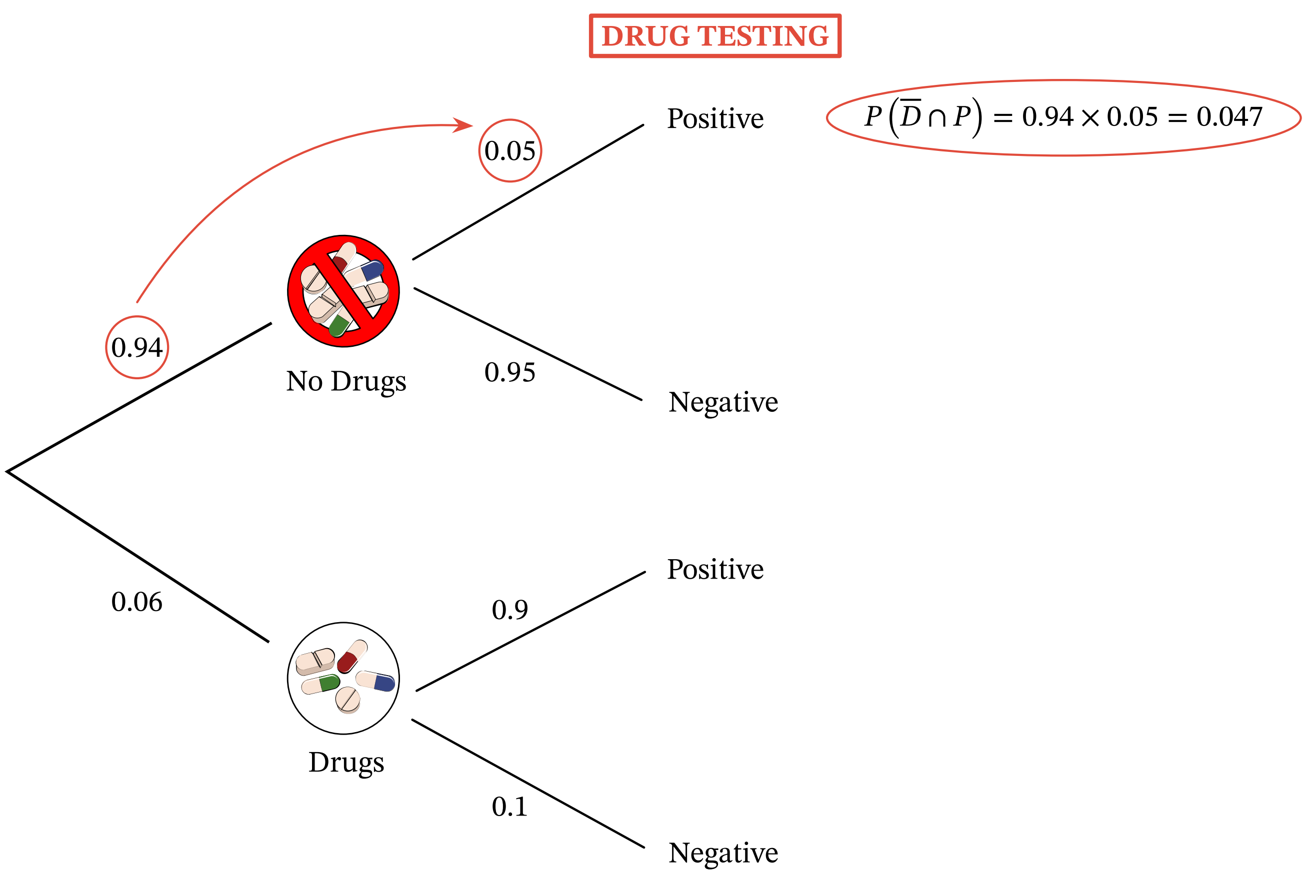


Since the probability of a false positive was 0.05 (5%), that is, it was known that no drugs had been taken but the test was positive anyway, the probability of testing negative when no drugs had been taken was $1 - 0.05 = 0.95$. These two probabilities (0.05 and 0.95) have been placed on the tree diagram next to the relevant branches (Positive or Negative), from the No Drugs branch.

Similarly, when it was known that drugs had been taken, the probability of testing negative was 0.1 (i.e., 10%) and, hence, the probability of testing positive was $1 - 0.1 = 0.9$. These probabilities (0.1 and 0.9) have been placed next to the Positive and Negative branches from the Drugs branch.

Part 1

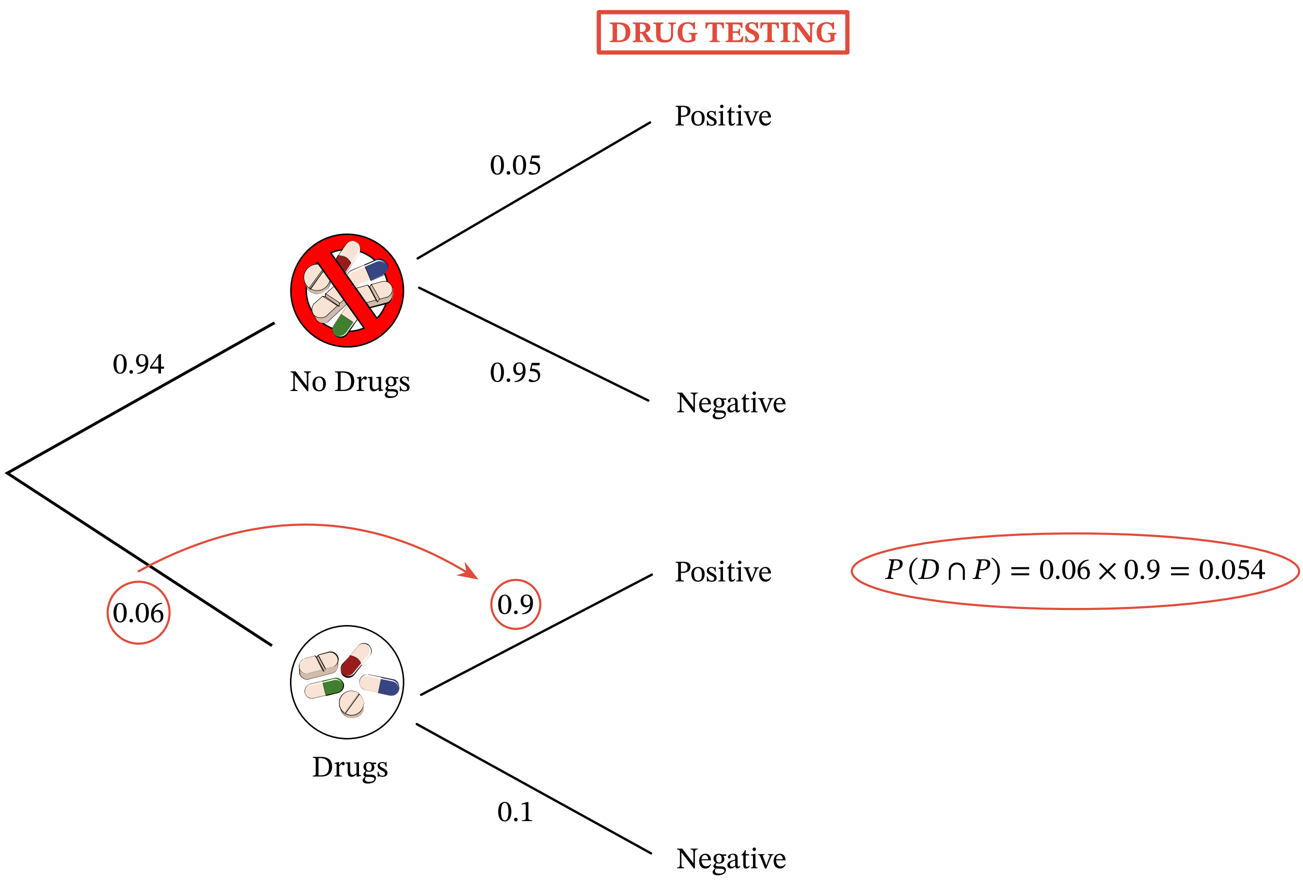
To find the probability that an MLB player had not taken drugs *and* tested positive, we multiply the probability on the No Drugs branch by that on the associated Positive branch.



Hence, the probability that an MLB player chosen at random had not taken drugs and tested positive was 0.047 or $0.047 \times 100\% = 4.7\%$.

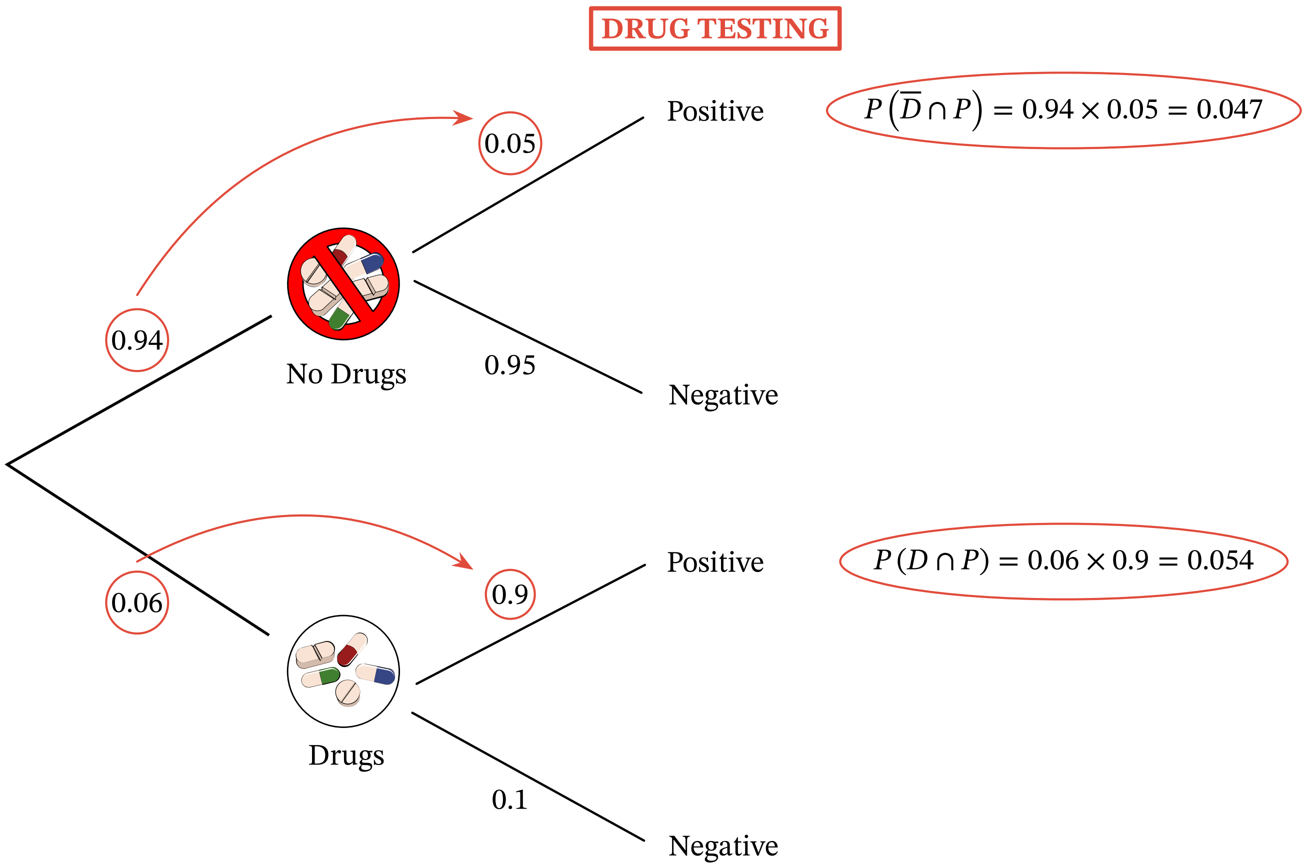
Part 2

To find the probability that an MLB player chosen at random is on drugs and tests positive, we multiply the probabilities for Drugs on the first set of branches and Positive on the second set.



Part 3

To find the probability that an MLB player chosen at random had positive test results, we must add together the probabilities for every possible situation where an MLB player may test positive. Fortunately for us, there are two ways a player might test positive and we have already worked out the probabilities for them both! These are the probability that a player was not on drugs and tested positive ($P\left(\overline{D} \cap P\right) = 0.047$) and the probability that a player was on drugs and tested positive ($P(D \cap P) = 0.054$).



$$P\left(\text{Positive}\right) = P\left(\overline{D} \cap P\right) + P\left(D \cap P\right) = 0.047 + 0.054 = 0.101$$

Hence, the probability that an MLB player chosen at random tested positive is $0.047 + 0.054 = 0.101$. That is, approximately 10% of the players tested positive.

Note

The probability that a player selected at random took drugs *and* tested positive ($P\left(\text{Drugs} \cap \text{Positive}\right) = 0.054$) is not the same as the probability that a player tested positive given that they took drugs ($P\left(\text{Positive} \mid \text{Drugs}\right) = 0.9$). In the second case, we have the prior knowledge that the player had taken drugs; in the first case, we do not know their drug status.

Let us recap the main points of using a tree diagram to work out conditional probabilities.

Key Points

When there is a relatively small number of outcomes, for compound (more than one) events, a tree diagram is a useful way of illustrating a probability problem.

On each branch of the tree, we write the probability of that outcome and the following should be true in any tree diagram:

- The sum of the probabilities for each set of branches should equal 1.
- The sum of the probabilities of all the final outcomes should also equal 1.

Recall that two events, A and B, are

i. **independent** if the fact that A occurs does not affect the probability of B occurring,

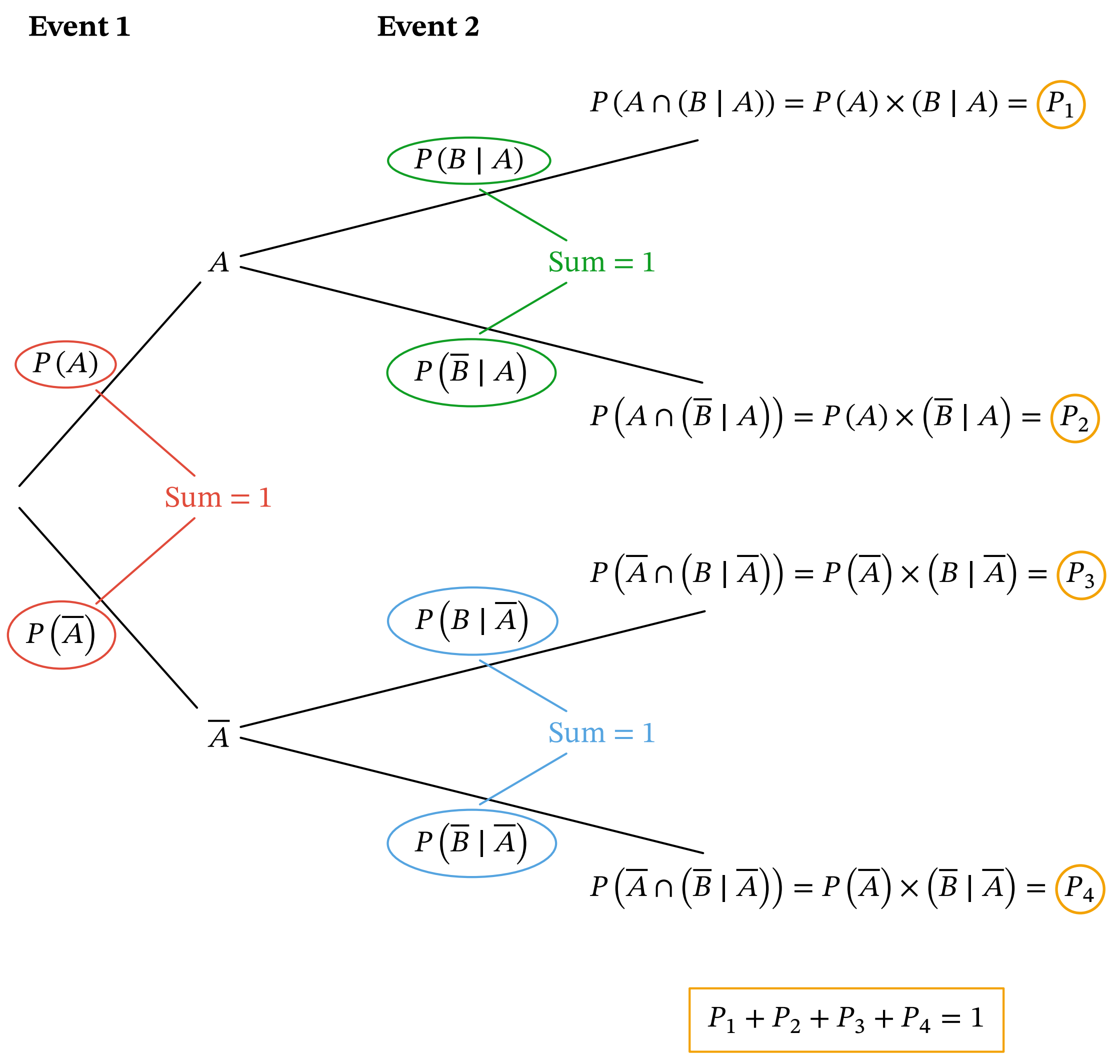
ii. **dependent** if the fact that A occurs does affect the probability of B occurring.

- For **independent** events, $P(A \cap B) = P(A) \times P(B)$.
- If two events are not independent, however,

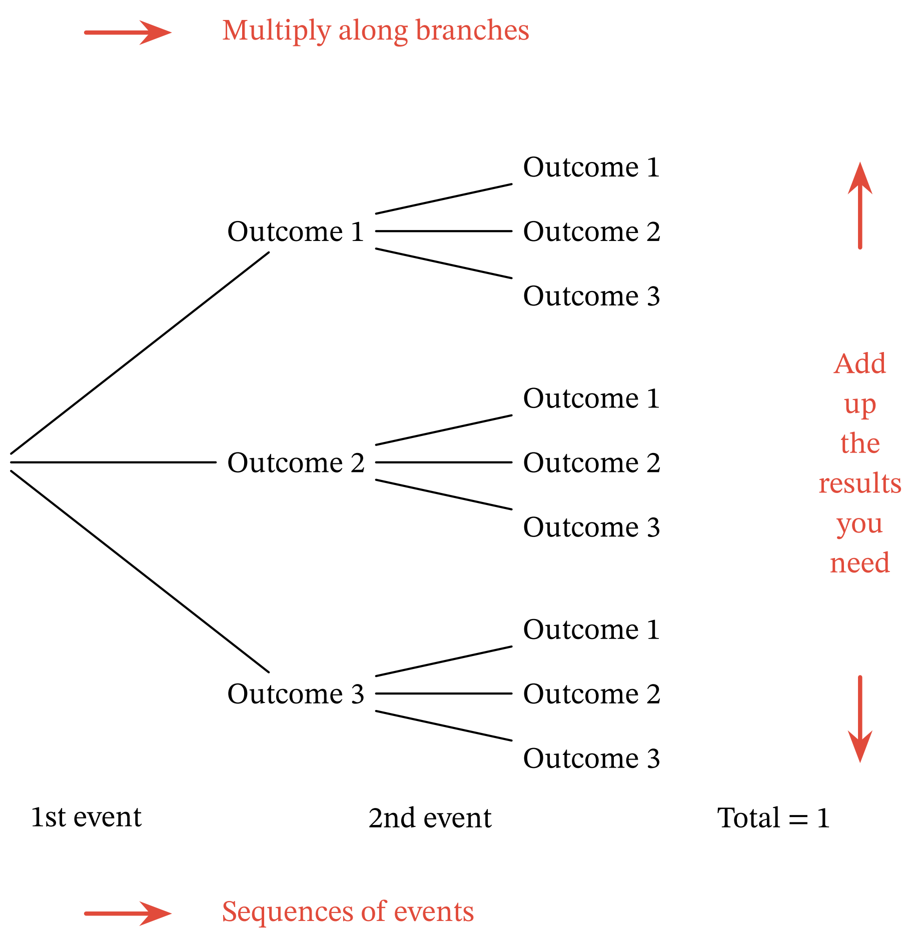
$$P(A \cap B) = P(A \mid B) \times P(B),$$

  where $P(A \mid B)$ is the conditional probability for event A, given that event B has occurred. Note, though, that if the events A and B are independent, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

If we have two events and each event has two possible outcomes, our tree diagram will look like this:



We can use tree diagrams where there are more than two events and also where each event has more than two outcomes. However, the sum of the probabilities for each set of branches and the sum of the probabilities of all the final outcomes must still equal 1. The example below has two events and 3 possible outcomes for each event. With computers and statistical software, we can of course create much more complex tree diagrams.


Multiply along branches
Outcome 1
Outcome 2
Outcome 3
Outcome 1
Outcome 2
Outcome 3
Outcome 1
Outcome 2
Outcome 3
Outcome 1
Outcome 2
Outcome 3
Add up the results you need
1st event
2nd event
Total = 1
Sequences of events