

Refresher - Tree Diagrams

Department of Statistics, University of Warwick

September 2025

1 Multi-Step Probability Problems

1.1 What is a Multi-Step Problem?

In basketball, players attempt two free throws in a row when fouled and receive one point for each score. In Dungeons and Dragons, a player attacks by rolling a 20-sided die to determine success or failure, and then a 6-sided die to calculate the damage. In an exam, you answer a succession of questions with some probability of getting each one correct.



A dungeon master's cache. Source: freepik

Each of these examples is a combination of two or more separate, but possibly related, experiments. These are **multi-step problems**, where multiple experiments take place in a row.

An outcome for a multi-step problem is a sequence of outcomes from the individual experiments. For example, if a free throw is either scored (S) or missed (M), the sample space for one shot is $\{S, M\}$. Then, the sequence SM would correspond to scoring the first and missing the second. The sample space for both shots, or the **product space** is $\Omega = \{SS, SM, MS, MM\}$, all possible combinations.

Note: the order here is important; SM is a different outcome to MS as it shows the order of occurrence. In ST118 you will define these rigorously.

Think of the sequence of outcomes as a series of **AND** statements. If we flip a fair coin twice with options of heads (H) or tails (T) each time, then flipping two heads in a row is the same as flipping heads **AND** flipping heads. To get the probability of both happening, we **multiply** the individual probabilities:

$$P(\{HH\}) = P(\{H\}) \times P(\{H\}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Hence, we can find $P(\{HH\}) = \frac{1}{4}$ in two ways; finding the product space explicitly and assuming all 4 outcomes are equally likely, or by multiplying $\frac{1}{2}$ by itself.

Outcome (ω)	HH	HT	TH	TT
Probability (p)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

However, if we look at the probability of scoring two free throws in a row, perhaps the result of the first affects the probability of the second. If so, we can't just multiply the same probability by itself.

1.2 Independence and Conditional Probabilities

Multiplication of probabilities for outcomes carries over to events. Considering outcomes are also events and events are generally more interesting, it is common to refer to events in general.

When we find the probability of a sequence of events, we are finding the **joint probability** of those events happening. So, the probability of scoring two free throws is the joint probability of scoring the first **and** scoring the second.

Definition 1.1 In two experiments, the probability of event A occurring in the first experiment and event B occurring in the second experiment is the **joint probability** $P(A \cap B)$ or $P(A, B)$.

The intersection symbol $\cap$ implicitly represents AND; the event that both A **AND** B occur.

Multiplying $P(H)$ by itself to get the joint probability $P(H, H)$ works because flipping two fair coins are **independent trials**; the first experiment has no effect on the second. We commonly refer to these as **independent events**, where one event occurring has no effect on another.

If missing the first free throw makes a player less confident and decreases the chance of scoring the second, then these are **dependent events**. We want some way of expressing the probability of scoring the second given what happened first.



A New York Knick stepping up to shoot. Source: Michael Barera / Wikipedia

Definition 1.2 In two experiments, the probability of event B occurring given event A has occurred is the **conditional probability** $P(B|A)$.

The key difference between joint and conditional probabilities is that in the joint, there is uncertainty on both A and B . In the conditional, A is known to have happened, and now we are interested in how this affects B .

Example 1.1 Let A be the event it rains today, and B be the event it rains tomorrow. Then, the probability it rains on both days is $P(A \cap B)$. If I know it rained today, the probability it rains tomorrow is $P(B|A)$.

There is a relationship between the joint and conditional probabilities:

$$P(A \cap B) = P(B|A)P(A)$$

In words, we find the probability of B given we know A , and then multiply by the probability A happened in the first place. This formula is a rearranged version of the general formula for conditional probabilities:

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

The definition of joint probabilities allows a rigorous definition of independent events:

Definition 1.3 Two events A and B are **independent** if and only if $P(A \cap B) = P(A)P(B)$.

Thus, if A and B are independent then $P(B|A) = P(B)$; A happening has no effect on our belief for B . If A and B are dependent events and $P(B|A) \neq P(B)$, we need $P(B|A)$ to get the joint probability.

The joint probability formula can be expanded to any finite length of sequence. Suppose we perform $n < \infty$ experiments and are interested in the sequence of events $A_1, A_2, \dots, A_n$:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|(A_1 \cap A_2 \cap \dots \cap A_{n-1})) \quad (1)$$

Thinking of $P(A_1) = P(A_1|\emptyset)$, the conditional probability of A_1 given no information, then the joint probability is the sequential product of conditional probabilities. Think of the conditional probabilities as the building blocks, and the joint probability as the final product.

Definition 1.4 For an event A , the **complementary event** $\Omega \setminus A$, written A^c , A' , $\bar{A}$ or $NOT A$, is the event that A does not occur.

A^c contains all outcomes that are not in A such that $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$ and $P(A^c) = 1 - P(A)$. So if A is rolling a 6 on a 6-sided die, then $A^c = \{1, 2, 3, 4, 5\}$.

Together, A and A^c form a **Bernoulli trial** (see refresher on probability distributions), and so any experiment can be transformed into a Bernoulli trial by focusing on whether or not one event occurs. For example, if rolling a 6 wins a game and any other score loses, then rather than consider all possible outcomes it is easier to focus on the events 6 and NOT 6.

Note: Conditional probabilities are not restricted to multi-step problems. For example, when rolling a 6-sided die, let A be an even number and B be a prime number. Then, $P(B|A)$ would be the probability of rolling a prime number given an even number was rolled.

The following question was Q18 on the 2024 A level Maths Paper 3.

Click [here](#) for video.

Question 1.1 The Human Resources director in a company is investigating the graduate status and salaries of its employees. Event G is defined as the employee is a graduate. Event H is defined as the employee earns at least £40 000 a year. The director summarised the findings in the table of probabilities below.

	G	G^c
H	0.21	0.18
H^c	0.07	0.54

An employee is selected at random.

- Find $P(G)$
- Find $P((G \cap H)^c)$
- Find $P(H|G^c)$
- Determine whether G and H are independent

Solution 1.1 a) Each cell of the table is a particular joint probability; e.g. $P(G \cap H) = 0.21$, 21% of employees are graduates earning at least £40,000. To find $P(G)$, we sum both cells in the G column, or:

$$\begin{aligned} P(G) &= P(G \cap H) + P(G \cap H^c) \\ &= 0.21 + 0.07 \\ &= 0.28 \end{aligned}$$

So 28% of the company are graduates.

b) We know $P((G \cap H)^c) = 1 - P(G \cap H)$ and so $P((G \cap H)^c) = 1 - 0.21 = 0.79$.

c) We use the formula:

$$P(H|G^c) = \frac{P(H \cap G^c)}{P(G^c)}$$

We know $P(H \cap G^c) = 0.18$ and we find $P(G^c) = 0.18 + 0.54 = 0.72$, so:

$$P(H|G^c) = \frac{0.18}{0.72} = 0.25$$

So given the employee is not a graduate, there is a 25% chance they are earning £40,000 or more.

d) G and H are independent if $P(G \cap H) = P(G)P(H)$. We know $P(G) = 0.28$ and $P(G \cap H) = 0.21$. Summing the row for H gives $P(H) = 0.21 + 0.18 = 0.39$. So:

$$\begin{aligned} P(G)P(H) &= 0.28 \times 0.39 \\ &= 0.11 \neq P(G \cap H) \end{aligned}$$

and so they are not independent. This makes sense, as we would expect graduates to be more likely to be higher earners...

Key Takeaways:

- Independent events have no effect on each other
- A joint probability is the probability of two events both happening
- A conditional probability is the probability of one event happening given another has happened
- To get the joint probability of a sequence of events, multiply the sequence of conditional probabilities (AND / $\cap$).
- Every event defines a complementary event and so we can transform any type of experiment into a Bernoulli trial centered on that event

2 Tree Diagrams

2.1 What is a Tree Diagram?

Often, multi-step problems are given in the form of a word problem. Consider the following example.

Example 2.1 A game consists of flipping a fair coin, rolling a fair die and then picking a piece of paper from a hat. If the coin is heads, then the die rolled is a 4-sided die. If the coin is tails, the die is a 6-sided die. There are 10 pieces in the hat and the score on the die tells how many of those pieces are winning. What is the probability of winning the game?

This can be difficult to parse in words but it is helpful to then visualise the problem, like drawing a graph when working with functions.

Definition 2.1 A **tree diagram** is a graphical representation of all possible sequences of outcomes from a series of experiments.

There are two scenarios where trees are particularly useful:

- When the experiments in the sequence are not the same experiment repeated (flipping a coin and then rolling a die)
- The experiments are dependent (the coin flip result dictates what die is chosen).

Tree diagrams are a useful way of representing the multiplication rule in (1), where each branch in the tree corresponds to a conditional probability.

2.2 Drawing the Tree

An example and video walkthrough on drawing tree diagrams follows, but there is a text explanation below.

Example 2.2 Suppose we toss a fair coin three times. Draw a tree diagram to represent this process.

[Click Here for Video](#)

To draw the tree, we begin with the first experiment and a starting **node**, the **root**. We draw a branch or **edge** outgoing from the root for each possible outcome of the first experiment, and label the edges with the outcomes they correspond to. Each edge leads to a new node. This node represents two things:

- The second experiment
- The outcome of the first experiment (based on the incoming edge)

Note: it is not always necessary to represent a node as a literal point. Any junction where one edge splits into multiple is implicitly assumed to be a node.

We then draw and label outgoing edges from each of these nodes to represent the possible outcomes of the second experiment. Repeat this process until each experiment has been recorded. Each node in the tree uniquely determines both which experiment it corresponds to and what the prior sequence of outcomes was.

After drawing the tree, we equip each edge with the associated conditional probability. For example, in the tree for flipping a fair coin 3 times, let x be the node after flipping H twice. Then, the two outgoing edges from x will have labels H and T , and the probabilities will be $P(H|HH)$ and $P(T|HH)$.

Note: rigorous definitions of joint probabilities should use the notation $P(A \cap B)$ or $P(A, B)$. However, when dealing with sequences of outcomes that are easily representable (such as HHH, H6W, etc.), we can short-hand this (to $P(HHH)$, $P(H6W)$, etc), with or without set notation. This is useful when outcomes are repeated; writing $P(H \cap H)$ or $P(\{H\} \cap \{H\})$ is confusing as to which experiment each H corresponds to. In ST118 you will learn how to make these definitions rigorous.

2.3 Using Tree Diagrams

The visual representation of the tree diagram breaks the process down into more manageable chunks, making it easier to tackle the problem.

Definition 2.2 *A complete sequence of edges from beginning to end of a tree diagram is called a **path**. Each path uniquely determines a sequence of outcomes from the individual experiments, and thus a single outcome in the product space.*

A path can be labeled with the sequence of outcomes when convenient (for example, HHH). To obtain the probability of this sequence, multiply the conditional probabilities of the edges in the path, just as in (1).

The tree diagram demonstrates how each path and thus sequence is mutually exclusive from the others. Thus, once the probability for each path is known, the tree extracts probabilities of events by summing the probabilities of paths that constitute the event. For example, $P(\text{At least one H}) = P(HH) + P(HT) + P(TH)$.

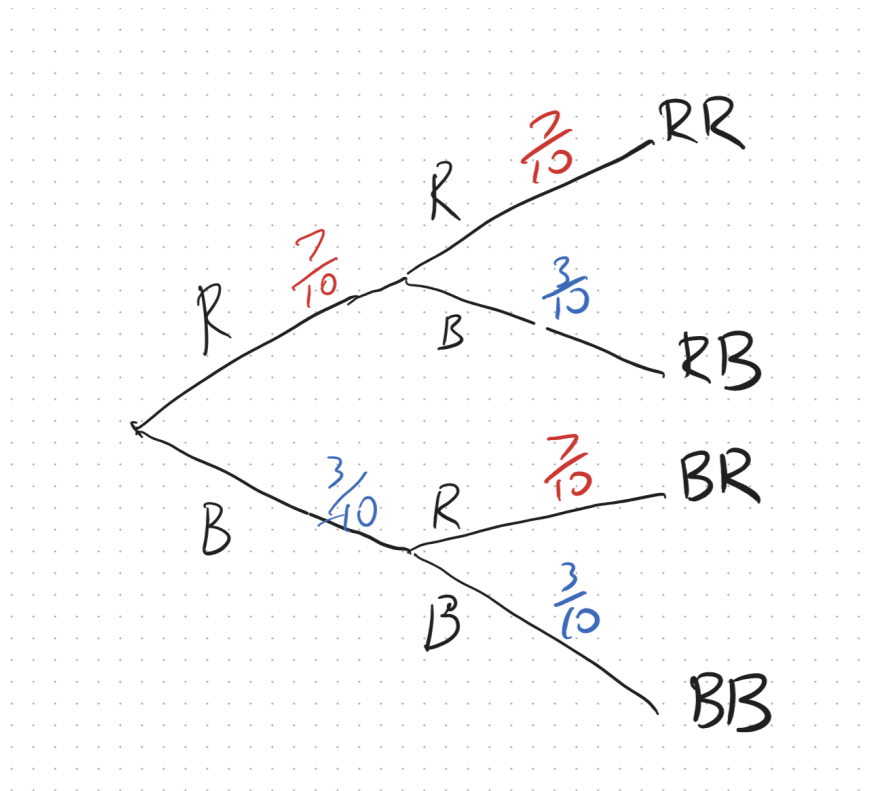
Recall the table in Question 1.1. Think of each cell in the table as matching a path; a unique combination of the events. A 2x2 table is very easy to work with but once the process grows the tree becomes more helpful.

It is generally more natural to think of a joint probability in terms of a series of conditional probabilities. For example, rather than thinking of the probability of it raining today and tomorrow, it is more natural to think of the probability it rains today, and then the probability it rains tomorrow given it has rained today. A tree diagram displays this well; the path gives the joint probability but we build it piecewise using conditional probabilities.

Example 2.3 *Suppose we have a bag of 10 marbles, of which 7 are red and 3 are blue. We pick a marble from the bag, record its colour, and then place it back. We then draw another marble and record its colour. What is the probability both balls are different colours?*

We know that $P(R) = \frac{7}{10}$ and $P(B) = \frac{3}{10}$. The sequence of choosing a red ball followed by a blue ball is written RB.

As the marbles are replaced, the two experiments are independent and the conditional probabilities are the same as the initial ones. A tree diagram for this series of experiments would look like:



The event of interest is that both balls are different colours, consisting of the sequences RB and BR . To find $P(RB)$, multiply the edge probabilities;

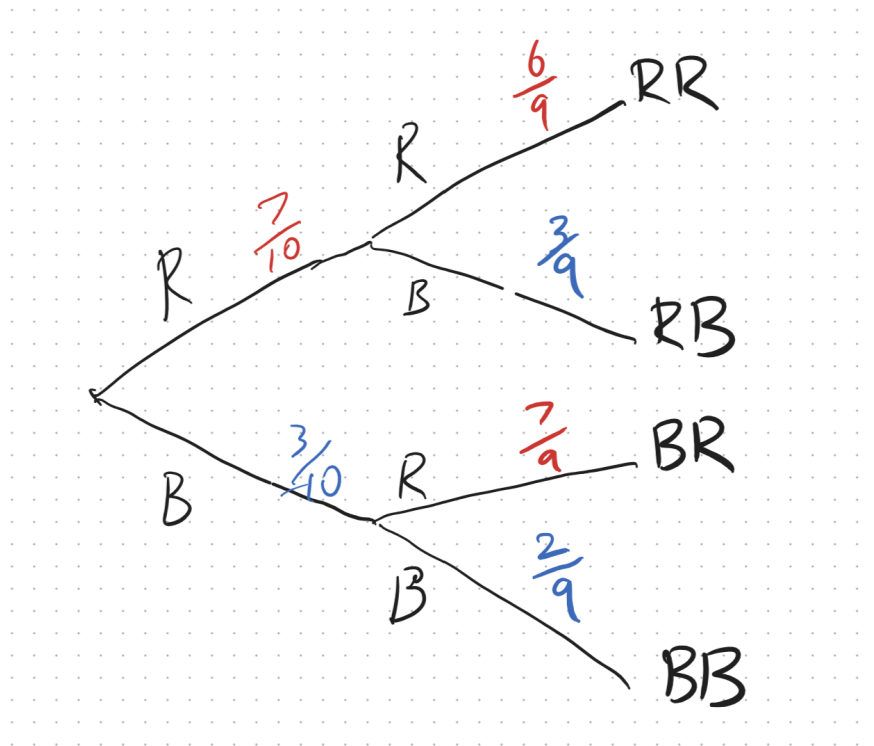
$$P(RB) = P(R) \times P(B|R) = P(R) \times P(B) = \frac{7}{10} \times \frac{3}{10} = \frac{21}{100}$$

Similarly, $P(BR) = 0.21$ as the order doesn't matter, by independence. Then, the probability of drawing two balls of different colours is:

$$P(\text{Different Colours}) = P(RB \cup BR) = P(RB) + P(BR) = \frac{21}{100} + \frac{21}{100} = \frac{21}{50}$$

Next, suppose the marble is not replaced. Then, drawing a red marble would make it less likely to draw a red marble in the future; the events are now dependent. If we draw a red marble, then there are now 9 marbles left of which 3 are blue, and so $P(B|R) = \frac{3}{9} = \frac{1}{3} \neq P(B)$.

The tree diagram would look like this:



Now, $P(RB) = \frac{7}{10} \times \frac{1}{3} = \frac{7}{30}$ and $P(BR) = \frac{3}{10} \times \frac{7}{9} = \frac{7}{30}$. Hence, the probabilities of both paths are still equal to each other but different to before. Thus:

$$P(\text{Different Colours}) = P(RB \cup BR) = P(RB) + P(BR) = \frac{7}{30} + \frac{7}{30} = \frac{7}{15}$$

which is bigger than the case where the marble was replaced. This makes sense as we would expect the probability of the two marbles being the same to decrease.

Finally, suppose we want to find the probability a blue ball was picked last. This event constitutes the sequences RB and BB . We know $P(RB) = \frac{7}{30}$ while $P(BB) = \frac{3}{10} \times \frac{2}{9} = \frac{1}{15}$. Then,

$$P(\text{Blue Last}) = P(RB \cup BB) = P(RB) + P(BB) = \frac{7}{30} + \frac{1}{15} = \frac{9}{30} = \frac{3}{10}$$

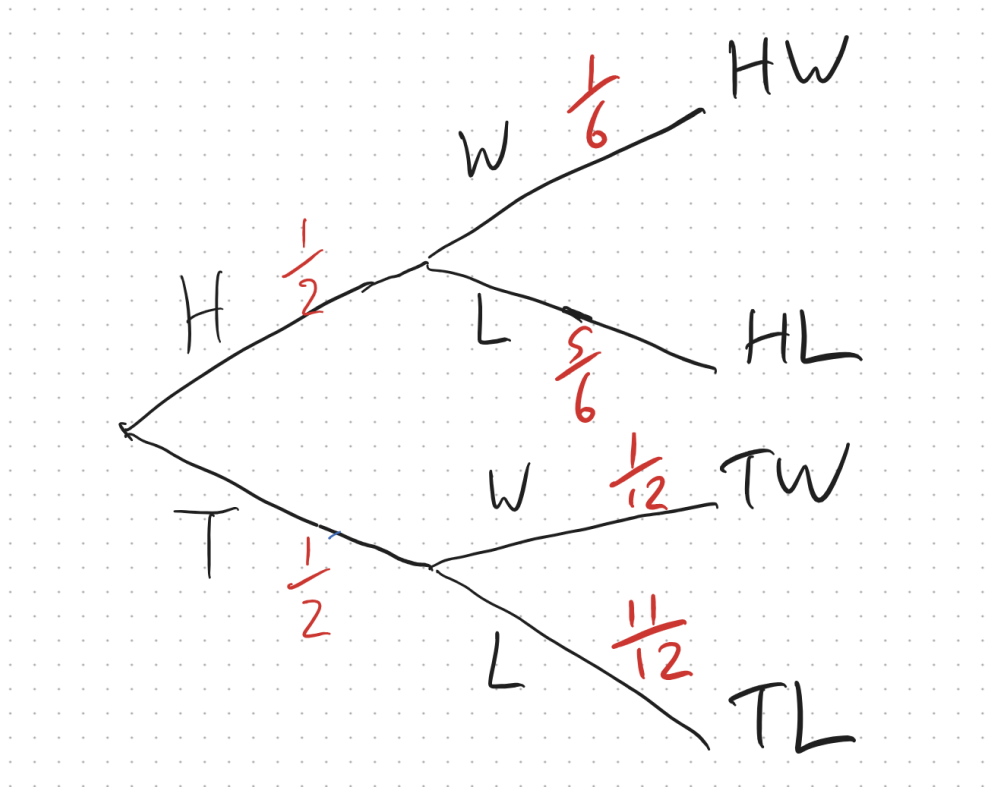
Interestingly, $P(\text{Blue Last}) = P(\text{Blue First})$! Think about why that is...

Now, try the following question.

Question 2.1 First, you flip a fair coin. If the result is heads, you receive a fair 6-sided die. If the result is tails, you receive a fair 12-sided die. You then roll your die and win the game if you roll a 6. The coin flip has sample space $\{H, T\}$ while for the die roll only consider whether you win or lose, $\{W, L\}$.

- Draw a tree diagram representing this game. Label all edges with their appropriate probabilities. Label each path with the sequence of outcomes it represents.
- Find the joint probability of each sequence of outcomes.
- Find the probability you win the game.

Solution 2.1 (a) The tree diagram can be seen below:



(b) The probabilities for each of the four sequences are:

$$\begin{aligned}
 P(HW) &= P(W|H)P(H) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} \\
 P(HL) &= P(L|H)P(H) = \frac{1}{2} \times \frac{5}{6} = \frac{5}{12} \\
 P(TW) &= P(W|T)P(T) = \frac{1}{2} \times \frac{1}{12} = \frac{1}{24} \\
 P(TL) &= P(L|T)P(T) = \frac{1}{2} \times \frac{11}{12} = \frac{11}{24}
 \end{aligned}$$

(c) The event you win the game is $\{HW, TW\}$, and thus:

$$P(\text{Win}) = P(HW) + P(TW) = \frac{1}{12} + \frac{1}{24} = \frac{1}{8} \quad (2)$$

Key Takeaways:

- Each edge in a tree corresponds to a conditional probability
- Each path corresponds to a unique outcome in the product space
- Probability for a path can be calculated by multiplying the edge probabilities along the path
- Probability for an event can be found by finding which paths make up the event and summing their probabilities

3 Law of Total Probability and Bayes Theorem

3.1 Partitions

In the marble example above, we found $P(\text{Blue last})$ by breaking one event into two mutually exclusive events and summing probabilities. We can similarly divide the whole sample space into a group of mutually exclusive events.

Definition 3.1 A series of events $A_1, A_2, \dots, A_n$ forms a **partition** for Ω if:

- i) $A_i \subseteq \Omega$
- ii) $\bigcup_{1 \leq i \leq n} A_i = \Omega$
- iii) $A_i \cap A_j = \emptyset \forall i \neq j$
- iv) $A_i \neq \emptyset, \forall i$

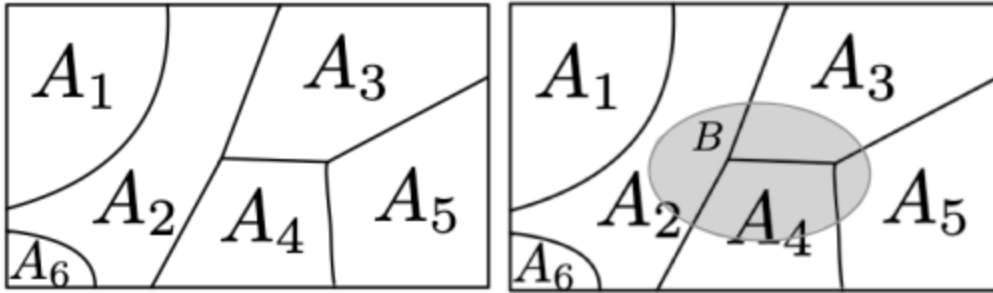
So, we divide the whole sample space into separate, non-intersecting pieces.

An example of a partition would be dividing the sample space of a 6-sided die into even and odd numbers. When rolling a 20-sided die in Dungeons and Dragons, we could partition Ω into success or failure.

Let $B \subseteq \Omega$. Then, we write B as the disjoint union of B intersected with each partition event:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$

The image below demonstrates both the partition and taking an intersection of the partition.



The easiest and finest partition of the sample space consists of dividing it into its outcomes (when finite), represented by the paths in a tree diagram. Then, as we move backwards through the tree we join these outcomes together to make a coarser partition.

For example, let $\Omega = \{HH, HT, TH, TT\}$, the results of two coin flips. The finest partition is $A_1 = \{HH\}$, $A_2 = \{HT\}$, $A_3 = \{TH\}$, $A_4 = \{TT\}$. We could create a new two set partition by going one step back in the tree, with $B_1 = A_1 \cup A_2$ and $B_2 = A_3 \cup A_4$. B_1 is Heads first and B_2 is Tails first.

3.2 Law of Total Probability

So we've seen how to make a partition, but the next question is why? Well, the key is mutual exclusivity. For an event B and a partition, $(B \cap A_i) \cap (B \cap A_j) = \emptyset, \forall i \neq j$. So, we break B up into mutually exclusive pieces too and use the law of union to find $P(B)$.

Definition 3.2 If $A_1, A_2, \dots, A_n$ forms a partition for Ω and $B \subseteq \Omega$, then the **Law of Total Probability** states:

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i) \quad (3)$$

Note: The formula splits $P(A_i \cap B)$ into $P(B|A_i)$ and $P(A_i)$ because we often use those to find $P(A_i \cap B)$

Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ and $A_1, A_2, \dots, A_n$ be the finest possible partition such that $A_i = \{\omega_i\}, \forall i$. Then, for an event $B \subseteq \Omega$, because each A_i is a singleton set, we have either:

$$\begin{aligned} B \cap A_i &= A_i = \{\omega_i\}, \text{ if } \omega_i \in B, \text{ or} \\ B \cap A_i &= \emptyset, \text{ if } \omega_i \notin B \end{aligned}$$

Thus, $P(B) = \sum_{i: \omega_i \in B} P(\{\omega_i\})$. This is what we do when we sum the path probabilities from a tree diagram to get the probability of an event. A tree is the finest possible partition but you don't always have to use that partition.

Another way to think of the Law of Total Probability is as a **weighted average**. We are finding the average of the conditional probabilities weighted by the probability of the event being conditioned on. If one event in the partition has a much greater probability than the others, its corresponding conditional probability will have a greater effect on the total probability.

We saw the Law of Total Probability in Question 2.1 when calculating the probability of winning: we looked at the probability of winning given a heads first, followed by the probability of winning given a tails first, and then summed. The following is another example.

Example 3.1 Suppose a genetic defect has a 1% prevalence in the population and there is a test which can be positive or negative. The test is positive for someone with the defect 95% of the time. However, in someone without the defect, the test will still be positive 10% of the time (a false positive). What is the probability the test will be positive?

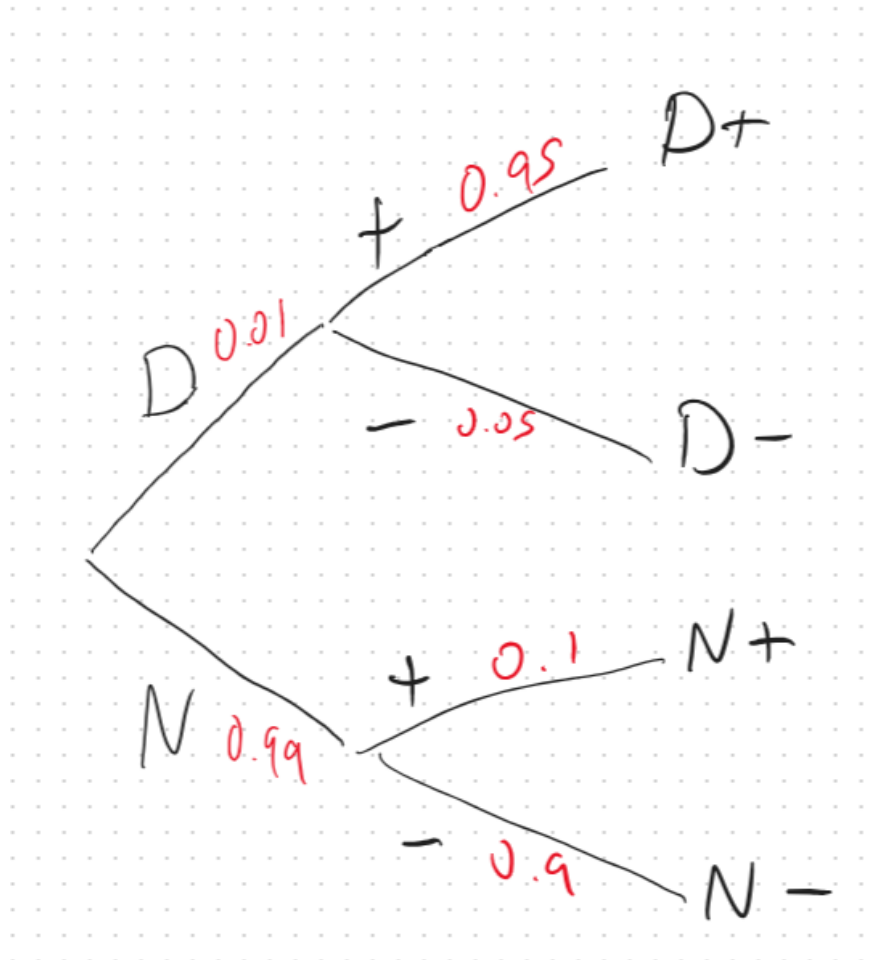
To answer, we first need notation. Let $+$ represent a positive test and $-$ represent a negative. Then, let D be someone with the defect and N be someone without.

Next, using this notation let us write down what probabilities we are given. We are told that 1% of the population have the defect so $P(D) = 0.01$. As such, $P(N) = 0.99$.

We are also told that given someone has the defect, the test is positive 95% of the time, or $P(+|D) = 0.95$. This means $P(-|D) = 0.05$, the probability of a false negative.

For someone without the defect, the test gives a false positive 10% of the time, or $P(+|N) = 0.10$. Thus, $P(-|N) = 0.90$.

We draw a tree diagram below to represent the problem.



Whether or not someone has the defect represents a partition: either someone has it or not. Thus, we write $P(+)$ as:

$$\begin{aligned} P(+) &= P(+|D)P(D) + P(+|N)P(N) \\ &= (0.95)(0.01) + (0.10)(0.99) \\ &= 0.0095 + 0.099 \\ &= 0.1085 \end{aligned}$$

So, there is a 10.85% chance that the test is positive despite only 1% of the population having the defect.



A test being performed. Source: freepik

3.3 Bayes Theorem

So far, we have looked at tree diagrams solely as a way to propagate information forward; we conduct an experiment, note the result, then move to the next. However, tree diagrams also allow us to look *backward*. For example, if we know it is raining today, what is the probability it was raining yesterday?



Assume for two events A and B we know $P(A)$, $P(B)$ and $P(B|A)$. Then, if B is recorded, how does this change our belief in A , $P(A|B)$?

Definition 3.3 *Bayes Theorem states*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

Often, the problem requires solving for $P(B)$ first using the Law of Total Probability before looking backwards.

Example 3.2 *Revisit the game from Question 2.1 that involved winning when a 6 is rolled. Suppose we know that someone has won. Can we find the probability they flipped a heads?*

Let W be the event they won and H be the event they got a heads. We want $P(H|W)$. We know $P(H) = \frac{1}{2}$, $P(W|H) = \frac{1}{6}$ and $P(W) = \frac{1}{8}$ from before. Then:

$$P(H|W) = \frac{\frac{1}{6} \cdot \frac{1}{2}}{\frac{1}{8}} = \frac{2}{3}$$

Interpret this as two-thirds of the time, a winner will have rolled heads. We find the corresponding probability for tails as we know $P(H|W) + P(T|W) = 1$ and so $P(T|W) = \frac{1}{3}$. Notably, this is half as likely as rolling heads, not surprising given we switch from a 6-sided die to a 12-sided die.

Bayes Theorem will become particularly important once you learn about the philosophy of Bayesian statistics. We now revisit Example 3.1.

Example 3.3 We saw that the test for the genetic defect had $P(+)=0.1085$. Given someone tests positive, what is the probability they actually have the defect? The probability we are interested in here is $P(D|+)$, the probability of the defect given the test result was positive. By Bayes Theorem, this is:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)}$$

We are given $P(+|D)$ and $P(D)$ in the question and we have solved for $P(+)$, so:

$$P(D|+) = \frac{(0.95)(0.01)}{0.1085} = 0.0876$$

This means that only 8.76% of the time, a positive test result actually indicates the genetic defect is present. This is a bad test because the false positive rate is too high, and 99% of the population not having the defect means it has a high weight in the calculation.

3.4 Common Mistakes

Below, see two common mistakes when using the Law of Total Probability, Bayes Theorem or conditional probabilities in general.

3.4.1 Common Mistake 1: Inconsistent Conditioning

In Example 3.2, we used

$$P(A|B) + P(A^c|B) = 1$$

This only works when the event being conditioned on is the same for both. It is not true if we condition on two different events, say B and C:

$$P(A|B) + P(A^c|C) \neq 1$$

It is also very common to get the order of the complement mixed up, and it is important to know the difference between the below statement and the one that does sum to 1.

$$P(A|B) + P(A|B^c) \neq 1$$

In general, we should only sum probabilities that are conditioned on the same event or have no conditioning.

3.4.2 Common Mistake 2: Wrong Conditioning Order

Similarly, it is common to be confused about which order to write a conditional probability when given a problem in words. This can lead to a common logical fallacy called **confusion of the inverse**, where $P(A|B)$ and $P(B|A)$ are falsely equated. For example, consider the following statement:

“75% of accidents occur within 5km of your home. Thus, it is safer to drive further from your home.”

We divide this statement into the **evidence** (75% of accidents occur close to home) and the **conclusion** (it is safer to drive further from home). While the conclusion *seems* wrong, the evidence appears to support it. We will display the conclusion is nonsensical using events (and some formal logic).

Let A be the event you suffer an accident driving. Let H be the event you are driving within 5km of your home. The conclusion is commenting on $P(A|H)$, the probability of suffering an accident given you are near your home. Comparing this probability with $P(A|H^c)$ tells whether you are more or less likely to suffer accidents close to home or not.

The evidence may appear to say $P(A|H) = 0.75$, but let us think about what is **given**:

“Given you suffer an accident, 75% of time you are within 5km of your home”.

It is clear that we should write this probability as $P(H|A) = 0.75$. Thus, while the conclusion is on $P(A|H)$, the evidence is for $P(H|A)$.

So why are so many accidents occurring near home if it's not more dangerous? The answer: so much of *driving* itself occurs close to the home. If 90% of your driving occurs within 5km of your home and 75% of accidents, then this would suggest it is actually safer to drive closer to home than not.

We show the fallacy mathematically using Bayes Theorem:

$$P(A|H) = \frac{P(H|A)P(A)}{P(H)} \quad (5)$$

Based on our available evidence, to comment on $P(A|H)$ we would also need to know the accident rate ($P(A)$) and how much driving occurs within 5km of the home ($P(H)$).

3.5 Example Question

Now, attempt the following question; a worked solution follows.

Question 3.1 *A factory has three types of machine that make the same part. Machine A makes 50% of parts, machine B makes 30% and machine C makes 20%. Each machine has a certain probability of making a defective part, D:*

- 6% of parts A makes are defective
- 10% of parts B makes are defective
- 8% of parts C makes are defective

An auditor choose a part at random from the factory and inspects it. Complete the following tasks:

- a) *Express the above information as 6 probability statements*
- b) *Draw a tree diagram to represent this information, complete with labeled edges and paths.*
- c) *Find the probability that the inspected part was both defective and came from machine A.*
- d) *Find the probability that the inspected part was defective.*
- e) *Given the defective part was found to be defective, find the probability it was made by machine C.*

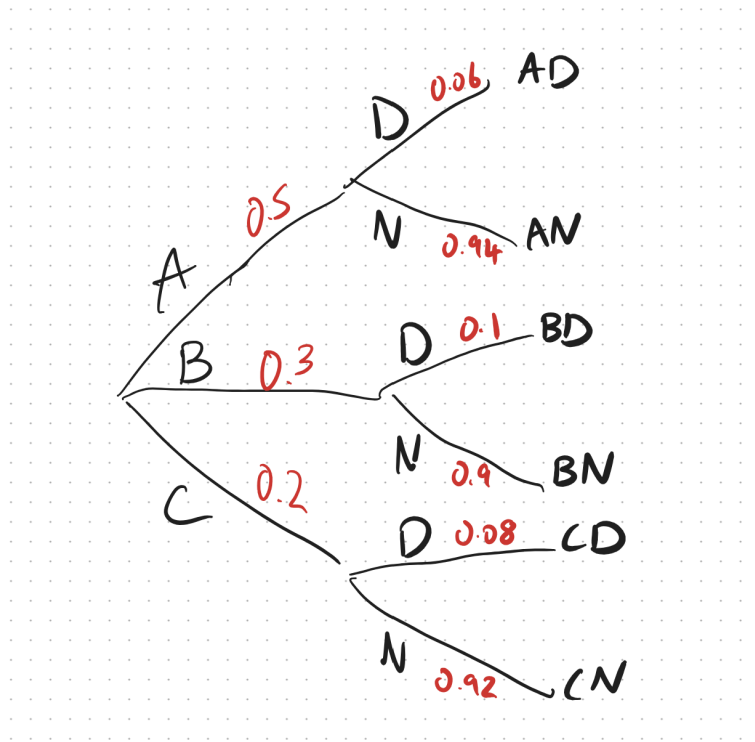
Solution 3.1 (a) *We are first given the proportions of parts each machine makes. That is:*

$$\begin{aligned} P(A) &= 0.5 \\ P(B) &= 0.3 \\ P(C) &= 0.2 \end{aligned}$$

We are then given the conditional probabilities of a part being defective given it came from a particular machine:

$$\begin{aligned} P(D|A) &= 0.06 \\ P(D|B) &= 0.10 \\ P(D|C) &= 0.08 \end{aligned}$$

(b) *Let N be the outcome that the part was not defective. A tree diagram expressing the manufacturing process is below:*



(c) To find the probability a part was both defective and from machine A, this is the sequence AD and we want $P(A \cap D)$. This is:

$$P(A \cap D) = P(D|A)P(A) = 0.06 \times 0.5 = 0.03$$

We can think of 3% of all parts as being both defective and from machine A.

(d) To find $P(D)$, we will use the Law of Total Probability. We have $P(A \cap D)$ and so:

$$P(B \cap D) = P(D|B)P(B) = 0.10 \times 0.3 = 0.03$$

$$P(C \cap D) = P(D|C)P(C) = 0.08 \times 0.2 = 0.016$$

Thus:

$$\begin{aligned} P(D) &= P(A \cap D) + P(B \cap D) + P(C \cap D) \\ &= 0.03 + 0.03 + 0.016 = 0.076 \end{aligned}$$

Hence, 7.6% of all parts are defective.

(e) Finally, we are given that the part is defective and asked to find the probability it came from machine C. That is, we want $P(C|D)$. We find this using Bayes Theorem:

$$\begin{aligned} P(C|D) &= \frac{P(D|C)P(C)}{P(D)} = \frac{P(C \cap D)}{P(D)} \\ &= \frac{0.016}{0.076} = 0.21 \end{aligned}$$

21% of all defective parts come from machine C. If we think of $P(D)$ as the weighted average of defective probabilities weighted by how much the machines are used, then $P(C|D)$ is the percentage of this probability C accounts for. As it makes the fewest parts and does not have an overly high defective rate, C accounts for less than the other two machines.

Key Takeaways

- The Law of Total Probability calculates the probability of an event by separating it across a partition
- Bayes Theorem allows us to look backward in a tree diagram
- Conditional probabilities can only be summed if they are conditioning on the same events
- Be careful which order you should be conditioning on; always think of what is given.

4 Worked Questions

The following questions will be accompanied by video walkthroughs.

We begin with an example based on basketball.

Question 4.1 *A game consists of 3 rounds of basketball shots. The first shot is from a short distance and is worth 1 point. The second shot is from a medium distance and worth 2 points. The third shot is from a long distance and is worth 3 points. The probability of scoring each shot, events represented by the number of points, is:*

- $P(1) = 0.8$
- $P(2) = 0.6$
- $P(3) = 0.2$

Let S represent a score and M a miss. Thus, a sequence of shots could be represented by SMS and would be worth 4 points ($1 + 0 + 3$).

- a) *Draw a tree diagram to represent the game. Label each path with the sequence of shots and the number of points scored.*
- b) *List the sequences that correspond to exactly 3 points.*
- A player wins the game only if they score 3 or more points in total.*
- c) *What sequences lead to the event W , winning the game?*
- d) *What is the probability of winning?*
- e) *Given an individual wins, what is the probability they scored the long distance shot?*

Now, suppose a player still has only 3 total shots but only attempts the next furthest shot if they score. For example, if they miss the first 1-point shot, they must try the 1-point shot again. If they score the second time, they take the 2-point shot as their third shot, and never try the 3-point shot. If they score the first and miss the second, they try the 2-point shot again.

- f) *Draw a tree diagram to represent this updated game, making sure to clearly label all edges and paths.*
- g) *If you still need 3 or more points to win, what is the updated probability of winning?*
- h) *Given an individual won the game, what is the probability they scored their first shot?*

Solution 4.1 [Click Here for Video](#)

The following is Question 12 from STEP II, 2021.

Question 4.2 *i) A game for two players, A and B , can be won by player A , with probability p_A , won by player B , with probability p_B , where $0 < p_A + p_B < 1$, or drawn. A match consists of a series of games and is won by the first player to win a game. Show that the probability that A wins the match is*

$$\frac{p_A}{p_A + p_B}$$

- ii) *A second game for two players, A and B , can be won by player A , with probability p , or won by player B , with probability $q = 1 - p$. A match consists of a series of games and is won by the first player to have won two more games than the other. Show that the match is won after an even number of games, and that the probability that A wins the match is*

$$\frac{p^2}{p^2 + q^2}$$

- iii) *A third game, for only one player, consists of a series of rounds. The player starts the game with one token, wins the game if they have four tokens at the end of a round and loses the game if they have no tokens at the end of a round. There are two versions of the game. In the cautious version, in each round where the player has any tokens, the player wins one token with probability p and loses one token with probability $q = 1 - p$. In the bold version, in each round where the player has any tokens, the player's tokens are doubled in number with probability p and all lost with probability $q = 1 - p$.*

In each of the two versions of the game, find the probability that the player wins.

Hence show that the player is more likely to win in the cautious version if $1 > p > \frac{1}{2}$ and more likely to win in the bold version if $0 < p < \frac{1}{2}$.

Solution 4.2 [Click Here for Video](#)