**Introduction to Probability**

Table of Contents

[Experiments 4](#_Toc63618566)

[Random Experiments 5](#_Toc63618567)

[Probability Models 6](#_Toc63618568)

[Probability Model Examples 8](#_Toc63618569)

[Sample Spaces 11](#_Toc63618570)

[Events and Event Spaces 13](#_Toc63618571)

[Probabilities 14](#_Toc63618572)

[Probability Axioms 14](#_Toc63618573)

[Joint Probabilities 15](#_Toc63618574)

[Probability Assignment 16](#_Toc63618575)

[Relative Frequency Approach 18](#_Toc63618576)

[Subjective Approach 18](#_Toc63618577)

[Conditional Probability 19](#_Toc63618578)

[Law of Multiplication (Product Rule) 22](#_Toc63618579)

[Law of Total Probability (Sum Rule) 23](#_Toc63618580)

[Bayes Formula 25](#_Toc63618581)

[Independence 26](#_Toc63618582)

Probability and statistics is the study of situations and phenomena that produces uncertain outcomes. Essentially, there is a defined range of outcomes, but we do not know for certain what the outcome is until we actually see the outcome.

The course will teach us that, for a problem description or report that can produce some uncertainty, how we can:

* Identify the uncertainty
* Select a probability model to mathematically define the uncertainty
* Predict the most likely outcome
* Make appropriate conclusions and forecasts

## Experiments

An experiment is anything that can produce a result. Experiments can be classified into two categories, deterministic and random.

A deterministic experiment is one which has only one possible outcome, i.e. the outcome is fixed. For example, if we perform a multiplication of two numbers, we are guaranteed to always get the same result.

A random experiment has more than one possible outcome. The number of outcomes can be anything from 2 to infinity. The outcome of any specific instance of the experiment is random, and we do not know which one it is until we actually perform the experiment. A random experiment can be repeated any number of times with identical conditions. Examples of random experiments include tossing a coin, rolling a dice, etc.

For our purposes, since the results are uncertain, we use the term ‘random experiment’.

An experiment has two components, a procedure and an observation.

A procedure is a real-world phenomenon that produces an outcome. For example, the physical process of tossing a coin is a procedure, since we cannot get the outcome of the toss until we actually perform the toss.

The observation is the results we are looking for. For example, if we send a packet of data from one computer to another, we would be observing the fate of the packet, i.e. if it was sent successfully or if it was unsuccessful. In this case, sending the packet is the procedure and whether or not the delivery was successful is the observation.

### Random Experiments

This course will deal mostly with random experiments, and it will do so in two parts: probability and statistics. The first part deals with probability theory, or classical probability, and the second part deals with statistical inference, which will lead up to machine learning later on.

Random experiments can be of two types. The first type is one in which we know everything about the experiment, its internal properties and the set of outcomes. Here, the job of the probability theory is to mathematically define the uncertainty. Once this is done, it is called a probability model. For example, if we are tossing a coin, we need to know about several factors. We need to know whether or not the coin is biased, whether external factors are affecting the toss, etc. Once all of this is known, the uncertainty can be captured by the mathematical description provided by the probability model. This model can help us find the likelihood of getting a certain outcome and what the expected outcome of the experiment will be, which we call the expectation.

The other type is where statistical inference comes in. We are given a random experiment and a set of outcomes for the experiment as data, but we do not have any information about the experiment. Statistical inference can help us figure out the unknown details of the experiment. We can estimate details related to the probability model of the experiment, we can identify the probability model without knowing the experiment and we can forecast future data of the random experiment.

## Probability Models

A probability model is a mathematical description of an uncertain situation (a random experiment). For example, if we are sending a packet of data from one computer to another, the probability model will tell us that the chances of sending the packet successfully is times that of failing to do so.

The probability model can also describe other things. For example, each instance of a packet being sent is independent from the rest. Essentially, the outcome of one instance of the experiment will not affect the other instances. In fact, the concept of independence is extremely important in probability theory. There are situations where we will be forced to assume that successive instances of an experiment are independent, since if we do not, the situation will become very difficult to deal with.

Now let’s look at a few examples. For the first example, consider that the procedure is sending three packets of data from one computer to another. The observation here would be the sequence of successful and unsuccessful deliveries. Possible outcomes include , , , etc. where indicates a successful delivery and indicates a failed one.

In the second example, consider that the procedure is again that three packets of data are being sent from one computer to another, but the observation is the number of successful deliveries. Possible outcomes include , , and .

In the third example, consider that the procedure is to send a packet of data repeatedly from one computer to another until the packet is delivered successfully and the observation is number of attempts that were required. Here, possible outcomes could be anywhere between and .

In the fourth example, consider that the procedure is to send packets of data one after another until three packets of data have been sent and the observation is the number of attempts that we needed. Here, the possible outcomes are between and .

Notice that the procedures are similar for all the examples, but these will not be considered the same experiments. Two experiments are only considered to be identical if they have the same procedure and the same observation.

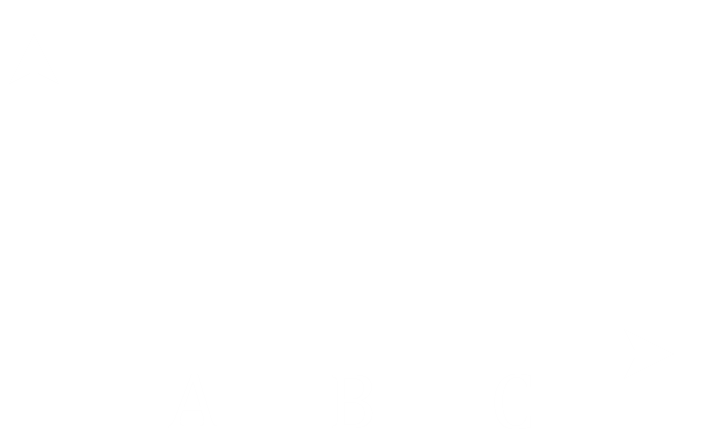
In any probability model, we work with a certain random experiment with a certain set of outcomes. This should include all the possible uniquely distinguishable outcomes. For this set of outcomes, also called the sample space (), we do not know for certain which of the outcomes will appear in one instance of an experiment. However, based on some data, we can make a prediction about the likelihood of an outcome appearing. The chances of each outcome occurring is called the probability of that outcome.

For some experiments, the probability of all the outcomes are equal, such as when we roll a fair die. For others, some outcomes have a higher probability than others. By this, we mean that for a number of runs of the experiment, those outcomes will appear more times.

The probability of an outcome is denoted by a non-negative number between and . This is done by convention. A higher value indicates that an outcome is more likely to occur and a lower value indicates that an outcome is less likely to occur. It can also indicate the relative frequency of a particular outcome occurring. Usually, the probabilities are between and exclusive, but there are cases where the probability may be or . This value is also called the probability measure.

For an outcome , the probability of the outcome is denoted by .

The outcomes and their corresponding probabilities can be represented using a graph. For example, consider an experiment with 3 possible outcomes, , and , which are all equally likely to occur. For this experiment, the graph would look a little like this:



We still need to talk about where we are actually getting the values for the probabilities. These values are either set or predicted. The assignment of these values needs to follow a set of rules called the probability laws, which we will study later.

### Probability Model Examples

Example 1.

Procedure: Send a packet of data from one computer to another.

Observation: Whether the packet was delivered () or the delivery failed ().

Sample Space:

It is very unlikely that the probabilities of both of these outcomes is equal. For now, let us assume that and . These two equations, along with the equation for the sample space, represent the probability model of this experiment. The outcomes here are independent. Note that this is an extremely simple model and is incomplete. We still need to give the actual values of the probabilities for the model to be complete.

Example 2.

Procedure: Sending 3 packets of data.

Observation: The sequence of outcomes.

Sample Space:

The probabilities here are a little more complicated and will be discussed later.

Example 3.

Procedure: Sending 3 packets of data.

Observation: The number of successes.

The outcomes here are, for successful deliveries, , for successful delivery, or or and so on. However, just by looking at these values, we are unable to tell specifically which outcome occurred. For example, the value could mean any of the three possibilities. Thus, the values we are searching for do not form the sample space.

Example 4:

Procedure: Send a packet of data repeatedly, until it is successful.

Observation: The number of attempts required.

Sample Space:

In this case, even though the values we are searching for are not the outcomes themselves, we can tell what the exact outcome is using the values. For example, the value can only mean that the outcome was . As such, this is the sample space.

Example 5.

Procedure: Send packets repeatedly until 3 packets have been sent.

Observation: The number of attempts required.

Here, the actual sample space is . Notice however that is not one of the possibilities. This is because the experiment stops as soon as we have 3 successful deliveries. Thus, that is a condition we must abide by in this experiment.

## Sample Spaces

A sample space is a set of outcomes that are

* Collectively exhaustive
* Mutually exclusive
* Have the finest grain property

A set of elements are considered collectively exhaustive if they contain all of the outcomes of the experiment. If we get some outcome that is not in the sample space , then the outcome is considered unknown. Our experiment should not have any unknown outcomes. Also, we should not have anything in the sample space that can never be an outcome for that experiment.

Mathematically, two elements are considered mutually exclusive if their intersection produces a null set. In other words, we should not have any scenario where more than one element of satisfies the outcome of the experiment.

The finest grain property is related to the granularity of the elements of the sample space. Essentially, this means that if we were to break down any of the elements of the sample space, none of the smaller elements would qualify as an outcome for the experiment. All the elements should be uniquely distinguishable, and a subset of an element of cannot be an outcome.

All of this also means one more thing, that the sample space of an experiment is unique. There can never be more than one sample space for a single experiment.

The finest grain property is likely to be vague and difficult to understand. This can be made easier with the help of an example.

Procedure: Roll a die

Observation: Number of dots on the side facing upwards

Sample Space:

Consider the sample space . Notice that there are two elements such that a subset of those elements, , can be an outcome of the experiment. As such, this sample space does not have the finest grain property and cannot be a sample space.

Whenever we start working on an experiment, the first thing we should do is define our sample space. This will help remove many of the confusions related to the development of the probability model of a random experiment. Incorrect sample spaces can lead to situations where the probability of a result is found to be more than .

## Events and Event Spaces

Formally, an event is a set of outcomes that consists of one or more elements of the sample space.

For the last example we say, . Here, a few possible events are:

From these, we can calculate the probability of a particular event occurring.

Let’s consider another example.

Procedure: Send 3 packets

Observation: Sequence of successes and failures

Sample Space:

Now consider the set . All the elements are mutually exclusive and collectively exhaustive. A set of events that have these two properties is called an event space. However, the elements are not finest grain, since the subset of an element could be an outcome. We can tell that a success occurs if an event occurs, but we cannot specify which outcome occurred. As such, we cannot call such a set a sample space.

## Probabilities

Now we shall discuss how to actually assign a value to a probability. All assignments are of course, not valid. For a probability to be assigned, it needs to satisfy a few conditions. These conditions are called probability axioms.

### Probability Axioms

By definition, axioms do not require any proof or validation. They are accepted as they are. The probability axioms are:

* **Non-negativity** – A probability cannot be negative. Theoretically, there is nothing preventing this. However, this is simply accepted as fact and always followed.

If is an event, , (for all values of ).

* **Additivity** – If and are two events and they are disjoint (mutually exclusive), . Collectively, is the probability that either or occurs.

This same formula applies for any number of events. Thus, generally, if , , are disjoint events, .

* **Normalization** – The probability of the entire sample space is , i.e. . Thus, the result of any instance of the experiment is always in the sample space.

There are a few corollaries we can derive from these axioms.

* A corollary extended from the additivity axiom is the case where the events and are not disjoint. Here, . This last term, , is the probability of the intersection of and , and is called the joint probability. This will be discussed further in a moment.
* Another term we need to be aware of is the complementary probability of an event. The complimentary probability of the event , , is the probability that the event does not occur. Thus, .
* If is a subset of , , then .

There are more corollaries we can derive, but these are the ones we will end up using frequently.

### Joint Probabilities

Joint probabilities are represented as , or . This simple means the probability of both events and occurring simultaneously. One approach to finding the join probability is to simply find the sum of the probabilities of the events that are common to and . Of course, the joint probability can be calculated for multiple events.

### Probability Assignment

There are three ways in which probabilities can be assigned.

* Classical Approach
* Relative Frequency Approach
* Subjective Approach

#### Classical Approach

In the classical approach, it is assumed that the outcomes are all equally likely. Say there are outcomes, , , , . Thus, the probability of each outcome will be . Intuitively, we know that this is correct under this approach, but we still need to check this against the three axioms. We will of course, find that all three check out.

On the other hand, we could assign the probabilities by following the axioms first. In this way, we know for certain that the axioms were followed.

* For this scenario, .
* According to the third axiom, we know that . Thus, .
* We are assuming that each of the outcomes are mutually exclusive. Thus, following the second axiom, we know that .
* We also assumed that each of the outcomes are equally likely, and thus have the same probability. Thus, , and .
* Thus, , meaning , where .

Example:

Procedure: Roll 2 fair 4-sided dice.

Observation: Number of dots on the top face of both dies

Sample Space:

There is a total of 16 possible outcomes. The question here is, are the outcomes equally likely to occur? In this case, they are. Thus, the probability of each of the outcomes is .

Now, let’s change the observation.

Observation: Sum of the dots on both dice

In this scenario, the outcomes are not all equally likely. This is because, some of the sums can occur due to more combinations than others. For example, the pairs , , and will all give us a sum of , but the only way to get the sum is from the pair . Thus, obviously, the outcomes are not equally likely.

However, we can still assign probabilities to the outcomes. This is done using the relative frequency approach.

### Relative Frequency Approach

For the previous example,

However, this is not always easily noticeable. Instead, in the relative frequency approach, we run the experiment a very large number of times, times approaching infinity. From here, is the number of times occurs, or . Thus,

### Subjective Approach

The subjective approach is used in very special scenarios and is not always included in text books. However, this will come in handy when we get to the Statistics part of this course, so we shall discuss it a bit here.

Essentially, the subjective approach is based off of someone’s experience. As such, it varies from person to person.

## Conditional Probability

In conditional probability, some information about the outcomes is given in advance. It is best to see how this works using an experiment.

Procedure: Roll a 6-sided fair die

Observation: Number of dots

Sample Space:

Obviously, the probability of any of the outcomes occurring is .

Now consider these cases.

Now, say we cannot actually see the dice after rolling it, because we have a screen between us and the dice. Someone else on the other side of the screen is informing us about the results. This person is not giving us the exact result, but instead telling is if event occurs or not.

Now, we need to find the probability of event , subject to the condition that event has already occurred.

The reason the probability that event occurs has changed is because we have a changed sample space, called the conditional sample space.

The probability model we have in this situation is called a conditional probability model.

More formally,

,

The conditional probability model satisfies the axioms of probability.

* If and are disjoint events,

and are not necessarily equal, but sometimes they might be. Both of these events deal with the probability of event occurring. The only difference is that in the first case, we know nothing about the experiment, and in the second case, we have some partial information. This difference causes the sample space to be different.

## Law of Multiplication (Product Rule)

This rule gives us the joint probability of two or more events occurring, e.g. . These probabilities are found using conditional probabilities.

We know that,

Thus, ,

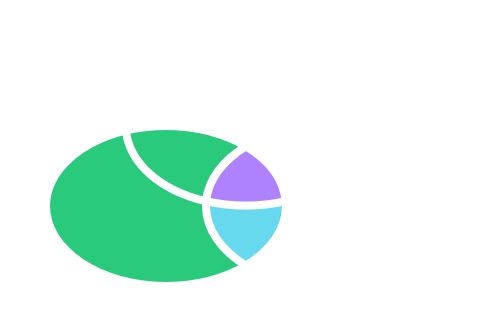
Similarly, , .

Of course, .

This is known as the law of multiplication.

Extending upon this idea,

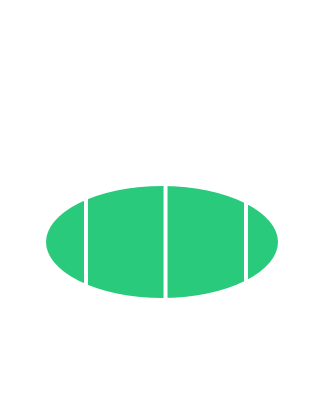
A Venn diagram should make visualizing why this works easier.



In general, . Thus, we are following a chain rule.

## Law of Total Probability (Sum Rule)

Consider that we have a sample space for an experiment, where is an event, and we have an event space consisting of , , and . Using Venn diagrams, these look like this:



Here , , and . Thus, , and so on.

We also have . Thus, , since we know , , and must be mutually exclusive since , , and are mutually exclusive.

Thus, .

Notice how we did not need to find the probability of occurring separately. Instead, we found the conditional probabilities of occurring given that each of the events in the event space occur, and then found the sum. This is the law of total probability.

Example

Consider that we are sending 3 data packets. The possible events are:

, , and .

Thus, .

Say the event .

We know that , , and .

Thus, , since , , and .

Thus, .

## Bayes Formula

Bayes formula combines conditional probability and the law of total probability together. Here, we will not be dealing with conditional probability, but the reverse. Say an event is picking a biased coin from some set of coins, and an event is getting three heads in a row. We want to know the probability that the event had occurred, given has occurred. Notice what this means. The event is not dependant on the event , but it is one of the possible ways in which can occur.

The way to solve this is to recognize that can occur in more cases than just after has occurred. In this problem, can occur for and (not picking a biased coin), with different probabilities in each case. Thus, the result we are looking for is the probability that has occurred after divided by the total probability of occurring.

If there were multiple ways for to occur, for the events from to , the probability that had occurred given has occurred is given by

Say we have 4 events, , , and , each signifying that a machine produces some products. The event signifies that some product is defective. Using Bayes formula, we can find, given that we have found a defective product, the probability that that product was produced by the th machine.

## Independence

Two events, and , are independent if the joint probability of the two events occurring is the same as the product of their individual probabilities, i.e. .

If and are independent, . Thus, .

Example

Procedure: Send 3 data packets

Observations: Sequence of successes and failures

Sample Space:

1.

2.

3.

4.

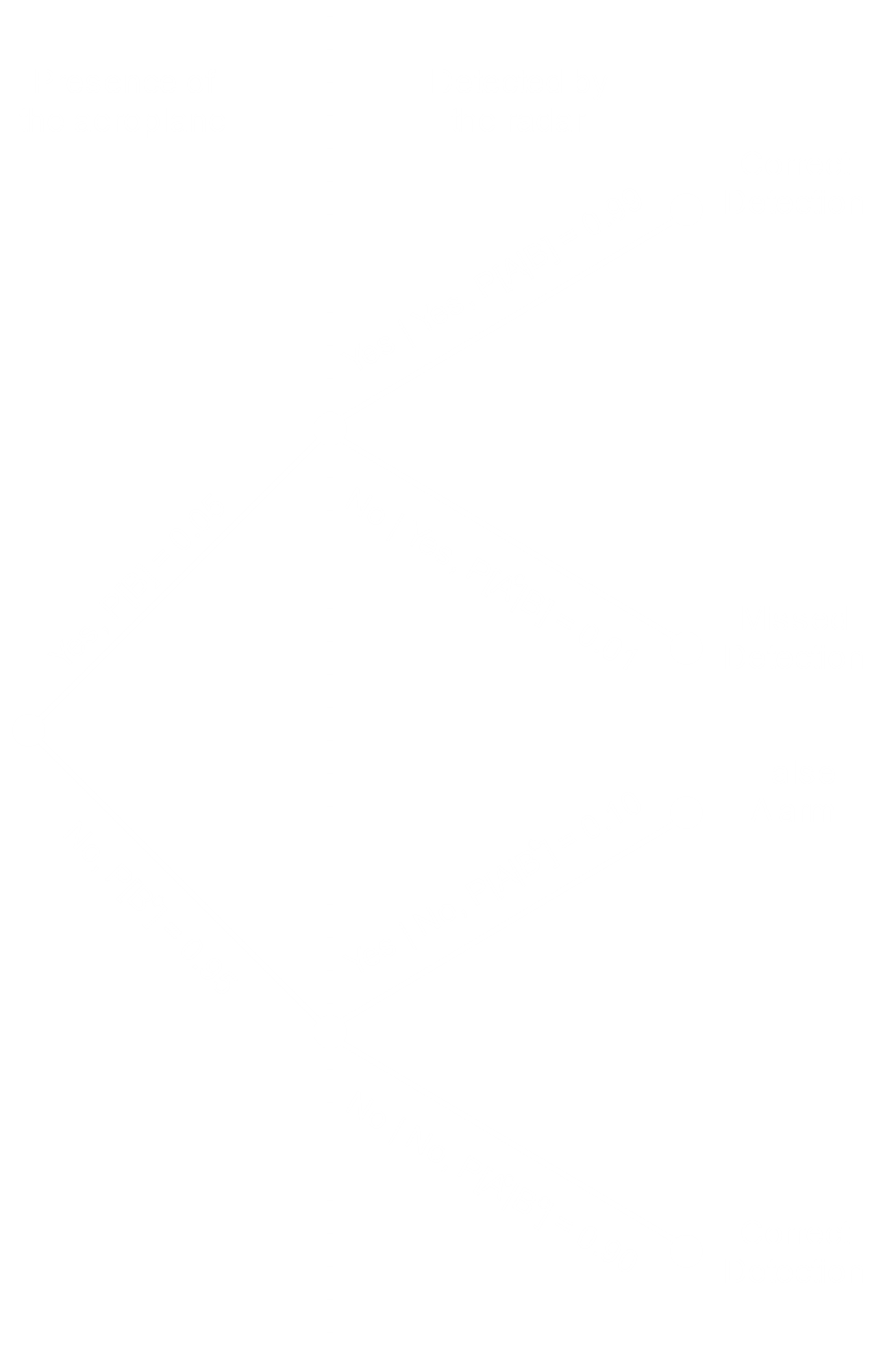
Example

The law of multiplication states that the probability of two or more events occurring simultaneously is given by .

Consider that there is a radar covering a certain area, and a plane enters that area. The radar sends a signal that bounces off of the plane and is detected by the radar. However, no system is perfect, and there is always the possibility the signal is reflecting off of some other body, and the radar is identifying that body as a plane.

Assume that the probability of an aeroplane being present with the range of the radar is . From the cases in which the aeroplane is present, the radar is able to detect it in of cases. The of cases where it is unable to detect the aeroplane are called missed detections. When the aeroplane is not present, the radar gives a false alarm in of cases.

We need to find the probability of a missed detection and the probability of a false alarm. To solve this, let’s make use of a tree diagram. It should make the problem easier for us.



Here, , , , , and .

A missed detection occurs when the aeroplane is present and there is no alarm, thus . Notice that this is the product of the two edges from the tree graph that lead to this result.

A false alarm occurs when an aeroplane is not present and there is an alarm, thus . Again, this is the product of the two edges from the tree graph that lead to this result.

Note that here, the presence of an aircraft is an unconditional event, and the radar detecting the plane is a conditional event, since it depends on the presence of the aircraft.

Example

Let’s look at an example for the law of total probability. Say we are playing in a chess tournament. There are three types of players here, of type 1 and each of type 2 and 3, with each type being an indication of how good the type of player is. Say the probabilities of us beating a player of each of those types is , and resepctively.

For a game with a randomly selected player, what is the probability that we win?

Let be an event where we play against a type player, where , and be an event that we win against a random player.

Example

For the same problem as the previous example, consider the problem using Bayes formula. Say we played against a random player and won. What is the probability that we played against a type 1 player?

Thus, .

How are we supposed to differentiate between the case where we used conditional probability and the case where we used Bayes formula?

|  |  |  |
| --- | --- | --- |
| Events | Play versus type player | Result |
| Order | First | Second |

|  |  |  |
| --- | --- | --- |
|  | Given | Target |
| Conditional Probability | Play | Win |
| Bayes Formula | Win | Type of opponent |

In conditional probability, the event that occurs first is given and we have to find the other event. In Bayes formula, the reverse is true.