**Conditional Distribution and Independence**

Table of Contents

[Independence 3](#_Toc64381110)

[Properties of Independence 6](#_Toc64381111)

[Independence Among Three Events 7](#_Toc64381112)

[Independence of Events 9](#_Toc64381113)

[Conditional Independence 9](#_Toc64381114)

[Independence of Random Variables 11](#_Toc64381115)

[Expected Values of Independent Random Variables 13](#_Toc64381116)

[Continuous Random Variables 14](#_Toc64381117)

[Independence of Multiple Random Variables 18](#_Toc64381118)

[Independence of Random Vectors 20](#_Toc64381119)

[Conditional Law of Multiplication 21](#_Toc64381120)

[Conditional Law of Total Probability 22](#_Toc64381121)

[Conditional Bayes’ Theorem 23](#_Toc64381122)

[Layered Conditional Probabilities 24](#_Toc64381123)

[Bayes Theorem from a Statistical Perspective 25](#_Toc64381124)

[Hypothesis 26](#_Toc64381125)

[Prior Probabilities 26](#_Toc64381126)

[Data and Likelihood 27](#_Toc64381127)

[Posterior Probability 28](#_Toc64381128)

[Bayes’ Theorem with Random Variables 30](#_Toc64381129)

## Independence

By the term independence we are referring to whether or not some event depends on another, i.e. whether or not knowing that one event has occurred has an effect on the probability of another event occurring.

Consider the procedure of tossing a fair coin twice. The set of possible outcomes is

We know that each of the possible outcomes is equally like, meaning they each have a probability of .

Let be the event that the 2nd toss produces a head. Thus, the possible events are .

Let be the event that the 1st toss produces a tail. Thus, the possible events are .

If we want to find the probability that the event occurs given that the event has already occurred,

The only event common to both and is . Thus, .

From the above example, we can make a conclusion. . This means that knowing that has occurred has no effect on the probability of . Thus, and must be independent. This is denoted as .

If , , given that .

Note the last condition added.

We have previously seen another formula, which is actually the formal definition of independence.

If , then and are independent.

This formula comes directly from the product rule of probability.

If , .

Example

Say we have a bucket with balls, green and red. The procedure is to pick two balls one after another with replacement.

Let be the event that the first ball is red and be the event that the second ball is red.

Since for both picks, we are just performing the same action,

Thus, since , and must be independent.

Now let’s say the procedure is to pick two balls without replacement.

remains . However, can now be found to be , since a red ball has been removed from the bucket.

Given that is the event that the first ball picked was green, if we calculate separately,

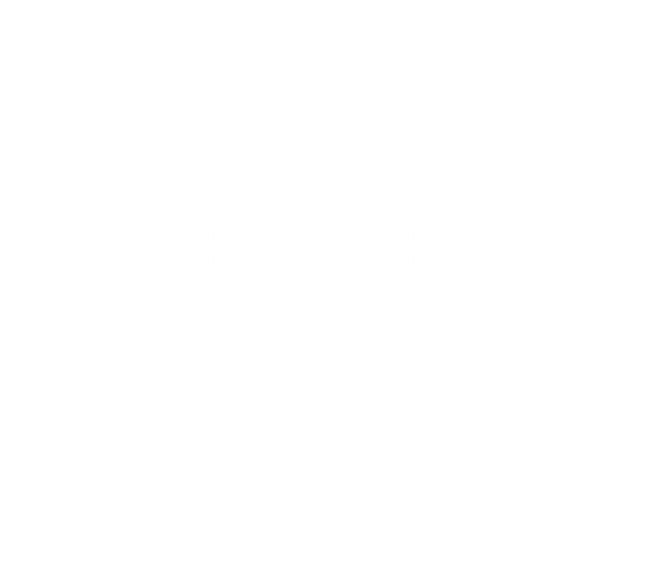
This is interesting, since if we do not know what the first ball picked was, simply ignores that. This is important since it means that whenever we are not given information about an event that the current event could be dependent on, we can safely ignore it. The law of total probability supports this.

Since , must be dependent on .

### Properties of Independence

1. If , then

We can actually prove this quite easily. Consider that and are mutually exclusive and cover the entire sample space, as shown below:



Applying the law of total probability,

since

Thus, .

1. If , then . This property can be proven in a similar manner.
2. If and are disjoint, they cannot be independent. This is why sample spaces are not helpful when we want to prove independence.

This property might be a little confusing since our normal assumption would be to think that two events with no common elements cannot affect each other. However, if and are mutually exclusive, if occurs, we know for a fact that cannot occur. As such, the outcome of one event is affecting the outcome of another, making the events dependant on each other.

### Independence Among Three Events

For three events , and to be independent (), all four of the following conditions need to be met:

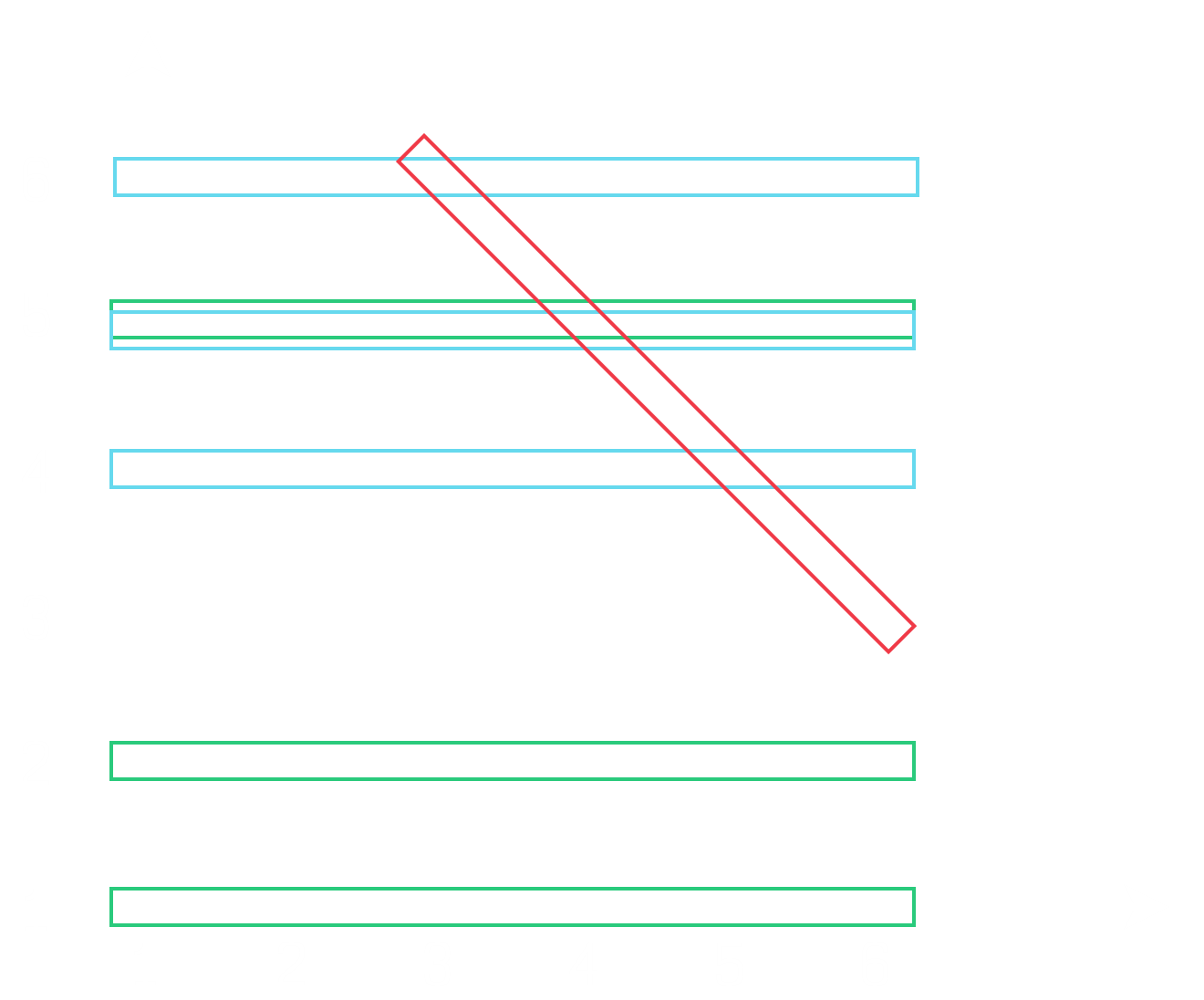
The last condition might not seem necessary, but it is possible that is somehow dependant on . This makes the last condition necessary.

Example

Say we roll 2 dies. Let

* be the event that the second roll is , or
* be the event that the second roll is , or
* be the event that the sum of the two rolls is

We can use a diagram to help us with this:



Since three of the six horizontal lines are covered by the event , .

Since three of the six horizontal lines are covered by the event , .

There are four outcomes that fall in the event , so .

None of the pairwise independences can be proven.

However, it just so happens that the combined independence can be proven.

From this, we can conclude that the fourth condition is a necessary condition, but not a sufficient condition. We need this condition to proven independence, but this condition alone is not enough to prove independence. If it were a sufficient condition, even though the other conditions also exist, they would not need to be proven for the statement to be proven true. This is not the case here.

### Independence of Events

For events , …, , we need to find all possible events or more elements and prove that elements in each subset are independent. Thus, for all subsets where , we need to show that

### Conditional Independence

Consider the following equation:

We have seen this before. It simply tells us that the fact that has occurred does not change the probability of occurring. This means is not dependant on .

Similarly, we can have this equation:

This tells us that an event has occurred that has changed the sample space, and in this changed sample space, the fact that event has occurred does not change the probability of occurring. Essentially, is dependent on , but not on .

Formally put, and are conditionally independent, conditioned on the event .

## Independence of Random Variables

If two random variables and are independent,

Let and be two random variables defined on the same sample space. If and are two real numbers, then we can have two events. The first event is that and the second event is .

and are said to be independent if

We are still dealing with events mind, so this is the same thing as .

Since this is true, we can say

Extending further upon this, say we now have two events and . Again, if and are independent,

, i.e.

In terms of conditional probabilities, the fact that should not affect the probability of having the value if and are independent and vice versa. Thus,

There is a simple proof for this.

There is an important condition to remember here. It is not enough for a specific pair of values of and to satisfy the above equations for them to be true. To prove that and are independent, the equations must be satisfied for all possible pairs is .

Example

Say a computer station is sending data packets to another computer station, of which are dropped. We need to find the probability that packets are dropped from the first packets sent and packets are dropped in total.

The fact that of the packets are dropped means that the probability of any individual packet being dropped is .

Let and .

We know that . This implies that . This is because and are independent. The value of does restrict the value of , but it does not change the probability of that value occurring. Put another way, the fact that packets are dropped in the first packets does not change the probability of packet being dropped in the other . Thus,

If we define a success as a packet being dropped, both and are binomial random variables. Thus,

### Expected Values of Independent Random Variables

If and are independent,

Furthermore, this equation is also true for any random variables derived from and , i.e.

Keep in mind though that the converse may not be true. The fact that these equations hold does not prove that and are independent.

Example

Say is a random variable for which , with each value of occurring with an equal probability of .

Say is a derived random variable such that , meaning . Here, and , since there are two events for which could have the value .

Thus,

However, we know must be dependant on since was derived from . Thus, this is proof that this equation being satisfy does not guarantee that and are independent.

### Continuous Random Variables

A similar formula exists for continuous random variables.

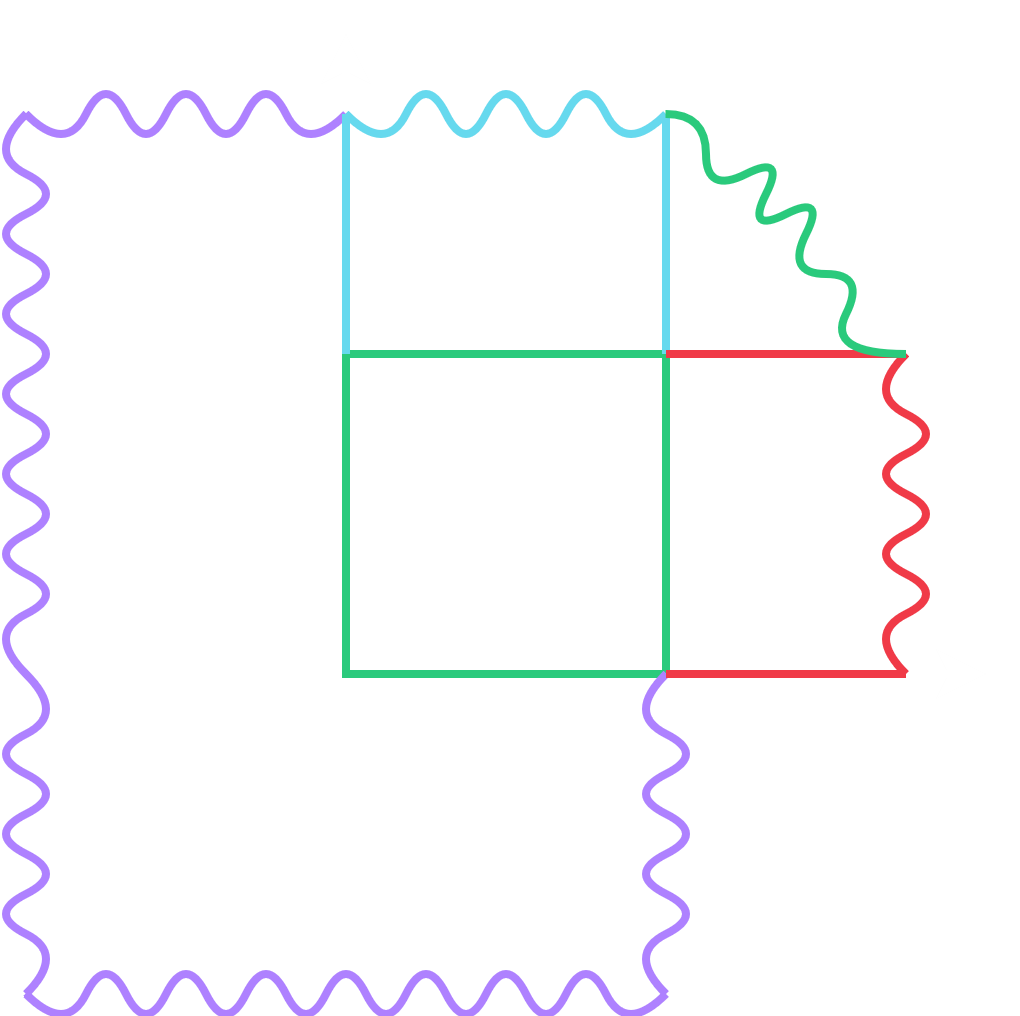
Example

and are independent and have identical distributions. This is the concept of IID (independently and identically distributed). We will be using this more and more with time. Essentially, it means that random variables are independent and have the same distribution functions.

For and ,

We want to find .

and are defined for a square region. We need to find the CDF for five different areas, as shown below:



For or , .

For and , .

For and , .

For and , .

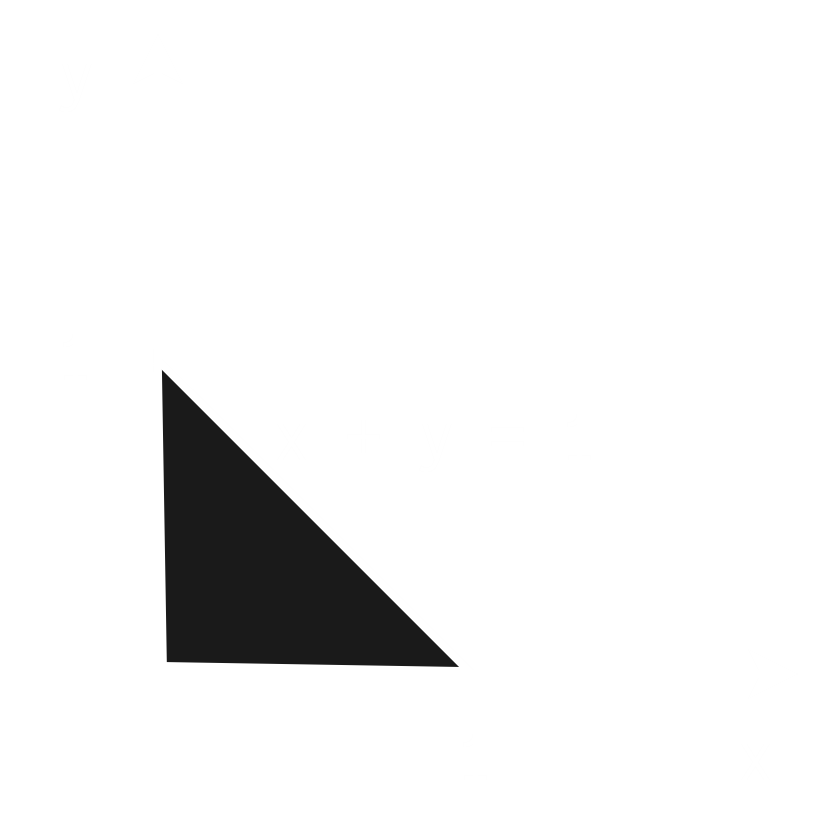
For and , .

Example

and are two random variables which depict the rainfall on May 1 in two successive years. and are independent and have identical distributions.

We want to find .

The area covered by is shown below:



### Independence of Multiple Random Variables

Similar ideas can be used for multiple random variables.

Of course, these equations need to hold for all possible values in .

Example

For four random variables , the joint PMF is given by

We need to find and and figure if the four random variables are independent on not.

Thus, the random variables are not independent.

## Independence of Random Vectors

We can do the exact same thing using random vectors. In reference to the previous example,

Let and .

Thus,

Thus, and are not independent.

## Conditional Law of Multiplication

We have previously seen that the joint probability of events can be found by the product of the conditional probabilities of each of the events conditioned on all of the previous events, i.e.

Extending upon this idea, we can also say that the joint probability of events conditioned on a single event is given by the product of the probabilities of each event conditioned on all of the previous events as well as the original conditioned event, i.e.

The above equations have been shown for three events, but of course, we can extend this for events.

A simple proof for this can be given.

We know that , and so on. Thus,

## Conditional Law of Total Probability

We have previously seen that the probability of an event can be found by summing the probability of the event conditioned on each event in an event space, i.e.

Similarly, we find the probability of an event conditioned on another event by finding the sum of the probabilities of the event conditioned on each event in the event space, which are themselves conditioned on the new event, i.e.

## Conditional Bayes’ Theorem

We have previous seen Bayes’ theorem as

If the event itself is also conditioned on another event,

## Layered Conditional Probabilities

By this time, it should be obvious that literally everything can have a conditional version. Even things that are already conditioned can have another layer of condition added. If we keep adding layer after layer of conditions, the question becomes how do we keep track of where to add variables?

The rule is, all unconditional probabilities in the equation should be replaced by conditional probabilities and all existing conditional probabilities should have the new condition added to the existing conditions.

Thus, if we have a conditional probability like this:

and we add another condition , will be replaced by and will be replaced by . Thus,

## Bayes Theorem from a Statistical Perspective

We have previously seen Bayes theorem as

Another way of looking at this same equation is from the perspective of Hypotheses () and data (). We can have many different hypotheses that might lead to a certain piece of data being found. Thus, is a particular hypothesis, and is the data.

Example

Say we have three types coins.

* The first type of coin is fair, so the probability of getting a head, . There are 2 coins of this type.
* The second type of coin is slightly biased, with . There are two coins of this type.
* The third type of coin is extremely biased, with . There is one coin of this type.

The procedure is to pick a coin at random, flip it, and record whether it is a head as data, .

Before we get into the answer, we need to know a few terms.

### Hypothesis

Firstly, the hypothesis, denoted by . The hypothesis in this case tells us which type of coin we picked. So, we could say that is that we picked a type coin, is that we picked a type coin and is that we picked a type coin.

### Prior Probabilities

Given no other information, just from the numbers of the different types of coins available, we can say that the probability of picked a coin of type , , because of the coins available are of type . Similarly, we can find the probabilities of picking coins of each type:

These probabilities are called the prior probabilities of the hypothesis, since these are the probabilities that the hypotheses are true before we actually perform the experiment and get our data.

### Data and Likelihood

Data simply refers to the outcome of the experiment. Data is denoted by , and for this experiment, we are considered to mean that a head has occurred.

For each of the hypotheses, we can use their corresponding prior probabilities to calculate the probability of getting some data, given that that hypothesis is true. Thus, this is simple conditional probability.

Notice that we did not consider the probability of actually getting the hypothesis for now, since we are not trying to find the overall probability of getting a head. Instead, we are considering the probability of getting some data given that we know a particular hypothesis is true, which is why we do not need to consider the probability of the hypothesis being true.

These probabilities of getting a particular data when we know a hypothesis to be true is called the likelihood. The higher the likelihood, the higher the probability that the outcome we get has occurred due to that hypothesis. In this case, this simply means that if we get a head, it is most likely that we picked the type coin.

### Posterior Probability

The Bayes’ theorem will help us find the probability that a particular hypothesis is true, given some data. It is called the posterior probability, since once we know the outcome of the experiment, the probability of the hypotheses will change.

Thus, , and will each tell us the probability that the hypotheses , and respectively were true, given a particular value of .

We now know the posterior probabilities of the different hypotheses.

Now we want to do a little analysis. For this, we will need a table of all the values we have found so far.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hypothesis | Prior Probability | Likelihood | Joint Probability | Posterior Probability |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Here, we can see that even though a type coin was most likely to give us a head, the fact that we actually got a head tells us that it was probably a type coin that caused it. This can be a little confusing at times. The likelihood only tells us the probability of getting a head once we had already picked a coin. It does not take into account the probability of picking the coin in the first place. Once we take that into account, as we did here, we see that the number of coins begin to matter as well.

Another important pattern to notice here is that, a hypothesis with a low prior probability but a high likelihood, like , will cause an increased posterior probability. However, if the likelihood is low, like in and , the posterior probability will be lower than the prior probability.

In the equation for Bayes’ theorem, the denominator, , is a constant. Thus,

Posterior Probability Likelihood Prior Probability

## Bayes’ Theorem with Random Variables

Let be the data. For the previous example, if we got a head, and if we got a tail .

In Bayesian statistics, is also used as a random variable. Here, let us define as the hypothesis. Thus, , where can have the values , or .

Using this, the original equation becomes

That was for discrete random variables. If we had continuous random variables instead,

If we create a table like we did for the previous example using random variables, it would look like this:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Hypothesis | Prior  Probability | Likelihood | | Joint  Probability | | Posterior  Probability | |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Example

Let us consider a different example now. Say we have the same set of coins as we did in the previous example, but now we toss it two times and both times we get a head, i.e. and .

One way to solve this to find the probability of getting different values of given the joint probability of and .

Another way is to solve the problem step by step.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hypothesis |  |  |  |  |
| Prior Probability |  |  |  |  |
| Likelihood |  |  |  |  |
| Joint Probability |  |  |  |  |
| Posterior Probability |  |  |  |  |
| Likelihood |  |  |  |  |
| Joint Probability |  |  |  |  |
| Posterior Probability |  |  |  |  |

Here, we found the posterior probability of getting the first head, and used that as the prior probability to find the posterior probability of getting the second head. The benefit of this second method is that we can easily extend it for a variable number of tosses.