**Joint Random Variables**

Table of Contents

[Joint Discrete Random Variables 3](#_Toc64363385)

[Distribution Functions 6](#_Toc64363386)

[Joint Probability Mass Functions 6](#_Toc64363387)

[Expected Values 8](#_Toc64363388)

[Joint Cumulative Distribution Function 8](#_Toc64363389)

[Joint Continuous Random Variables 11](#_Toc64363390)

[Joint Cumulative Distribution Functions 11](#_Toc64363391)

[Joint Probability Density Functions 14](#_Toc64363392)

[Expected Values of a Function of Two Random Variables 22](#_Toc64363393)

[Covariance and Correlation 23](#_Toc64363394)

[Correlation Coefficient 26](#_Toc64363395)

In this section, we will be looking at random experiments that have two random quantities as the outcome. The two random variables could be any combination of discrete and continuous random variables, but to begin with, we will consider joint discrete random variables.

## Joint Discrete Random Variables

Consider that we are sending three data packets, and that each packet is successfully sent, the event , with probability . For simplicity, let’s say . Thus, the probability of failure, the event , is . The probabilities are independent.

For this experiment, say we have two observations:

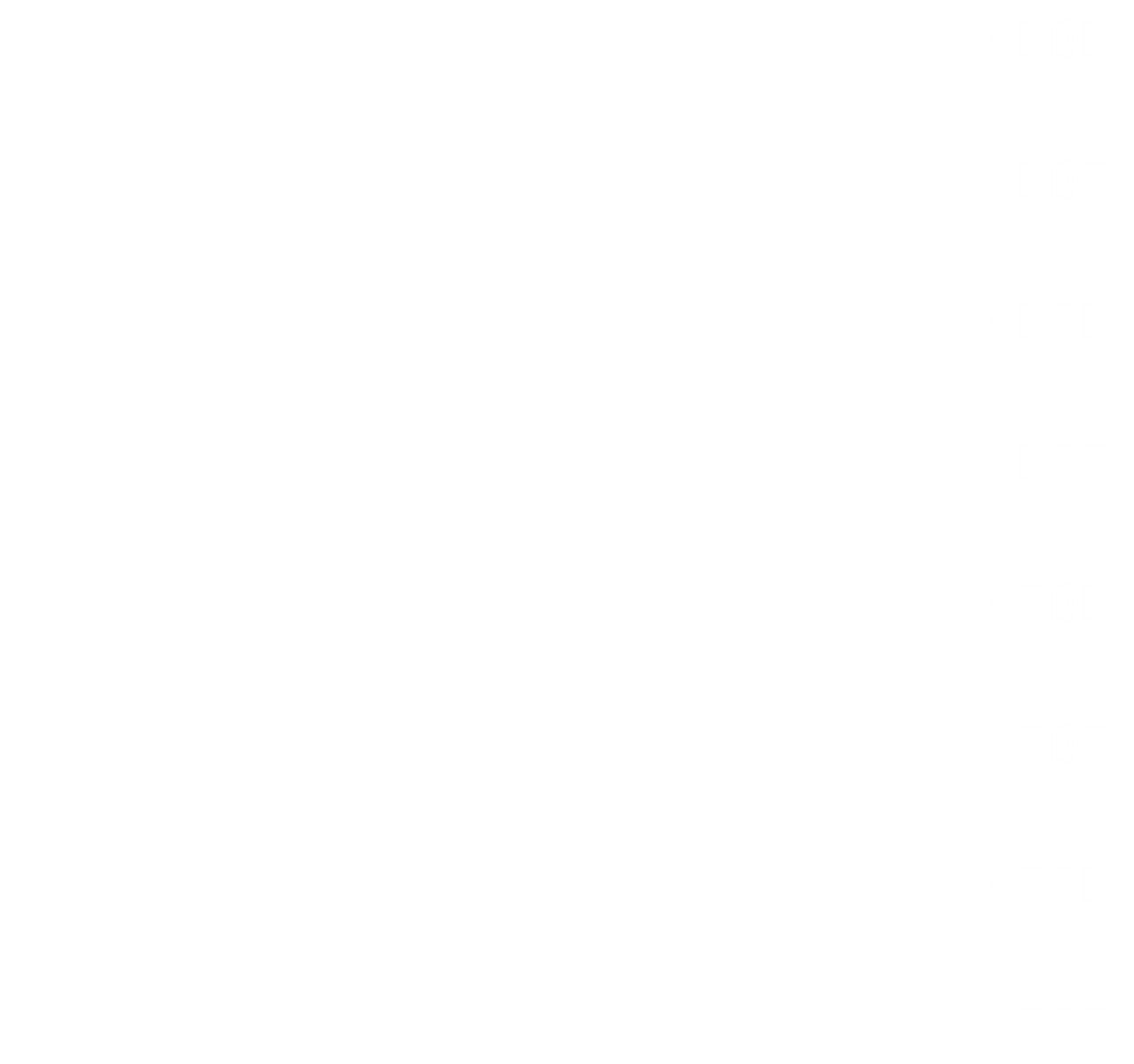
* The total number of success(es)
* The total number of success(es) before the first failure

This second observation is a little tricky, because a scenario exists where all three packets are successfully sent. This would cause an undefined result, so in that scenario, we will assume that the fourth packet would have failed and say the answer is . We could also have assumed this to be if we wanted.

We know .

Let and . and , with the last member of being the exceptional case that we assumed above.

One thing to keep in mind is that for joint random variables, both the random variables have to be defined based on the same sample space. This is why joint random variables are also called bivariant random variables, since if one varies, the other varies too (unless they are independent).



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome | -Value | -Value | Joint Value | Probability |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

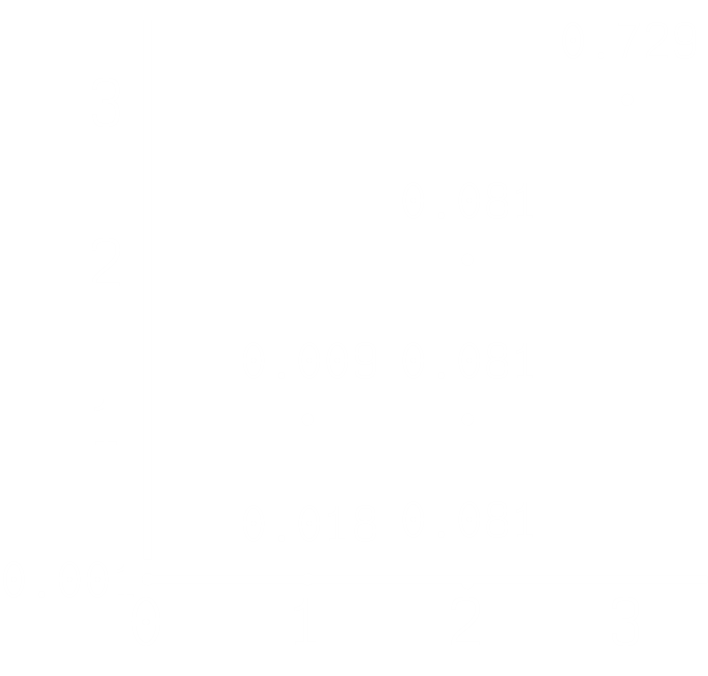
The diagram and table above are based on the scenario we just described and it gives us a huge amount of information. Firstly, we have a tree that shows us all the possible outcomes. We also have the probabilities of each outcome listed, as well as the corresponding values for and . However, the part that is of interest to us right now is the column of pairs. There are ordered pairs of the values , and this is what we will consider to be the outcome of the experiment.

Thus, .

For single random variables, we had functions that essentially converted the values of into points on a number line. Now we have a joint random variable, so we need functions that will convert the values of into points on a 2D plane.

For single random variables, we studied probabilities like , where we essentially attempted to find the probability of landing at a point on the line. Now, we will do the same, the only difference being that we will try to find the probability of landing on the point in the 2D plane, .

For single random variables, we had a 2D graph, with the -axis representing values of and the -axis representing probabilities. For joint random variables, we will need a 3D graph, with the - and -axes representing the values of and respectively, and the -axis being used to represent probabilities.



The graph is fairly simple to understand. The third axis has not been shown, with the actual values of probability being written instead.

The only notable feature of the graph is the point . If we go back to the diagram we saw earlier, we will see that this point occurs twice. Thus, its probability is doubled in the graph. This should be obvious if comparisons with similar scenarios for single random variables are made.

## Distribution Functions

Joint random variables have three distribution functions associated with them, joint probability mass function (JPMF), joint CDF and joint PDF.

### Joint Probability Mass Functions

With single random variables, we defined PMFs for the event . Similarly, we defined CDFs for the event .

For joint random variables, things are quite similar. JPMFs are the probability that the event occurs. It is denoted as

Of course, the axioms of probabilities should hold.

#### Representation of JPMFs

We can represent JPMFs in three ways. The firth method is to use a list.

The second method is the graphical representation that we saw above.

The third method is to use a matrix.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Notice that, in the rightmost column of the matrix, we have . This is given by the summation of each of the rows. Similarly, in the bottommost column, we have , given by the summation of each of the columns.

The two PMFs we have in the matrix, where we are considering just one of the random variables even though two are involved, are called the marginal PMFs, since we are marginalizing one of the variables. Thus, the marginal PMF of is given by

Similarly, the marginal PMF of is given by

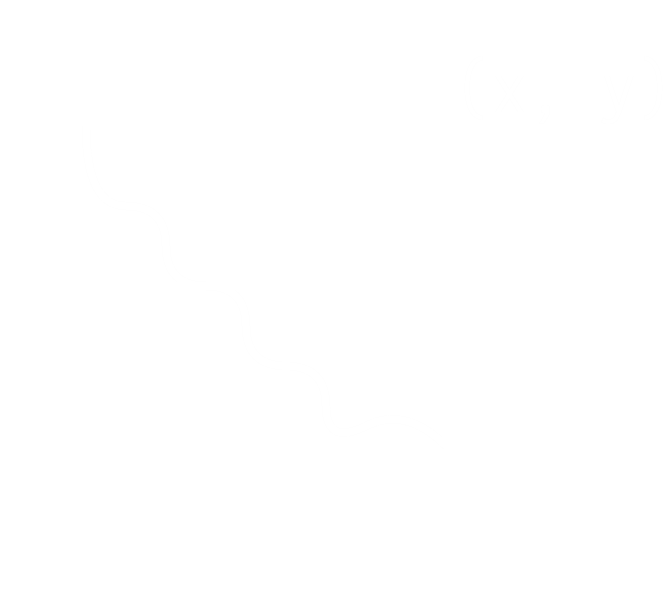
### Expected Values

Once we have the marginal PMFs of and , we can calculate the expected values from the marginal PMFs. However, that is not what we want to do. Instead, we want to calculate the expected values using the JPMF.

Similarly,

### Joint Cumulative Distribution Function

The JCDF is defined by the probability that the event occurs. It is denoted as



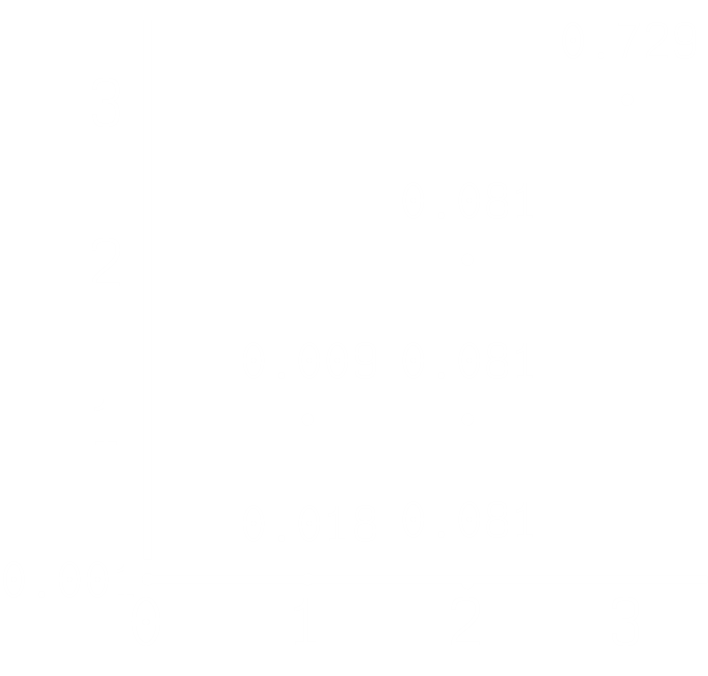
Here, the curly line indicates .

Some properties for JCDFs are:

* , meaning the value will be if either of the random variables has a value .

CDFs for discrete random variables have always been difficult to define. For single random variables, if the values of ranged from to , we could say if and if . Between this range however, we needed to find separate values for for each value of we had in between and . From this, we obtained a stair case model.

For JCDFs, we will have to do similar stuff. And since it will be equally difficult to define, JCDFs are not really used for discrete joint random variables. First, take a look at the JPMF graph from the example we did in the last lecture.



There are valid values here, the smallest of which is . Thus, we can say that if or . Similarly, the largest valid value is , which means we can say if and . This is similar to the easier parts of the CDF definition. Now for the tricky part. Well, no. Well come back to this part. There’s easier stuff we can go over first.

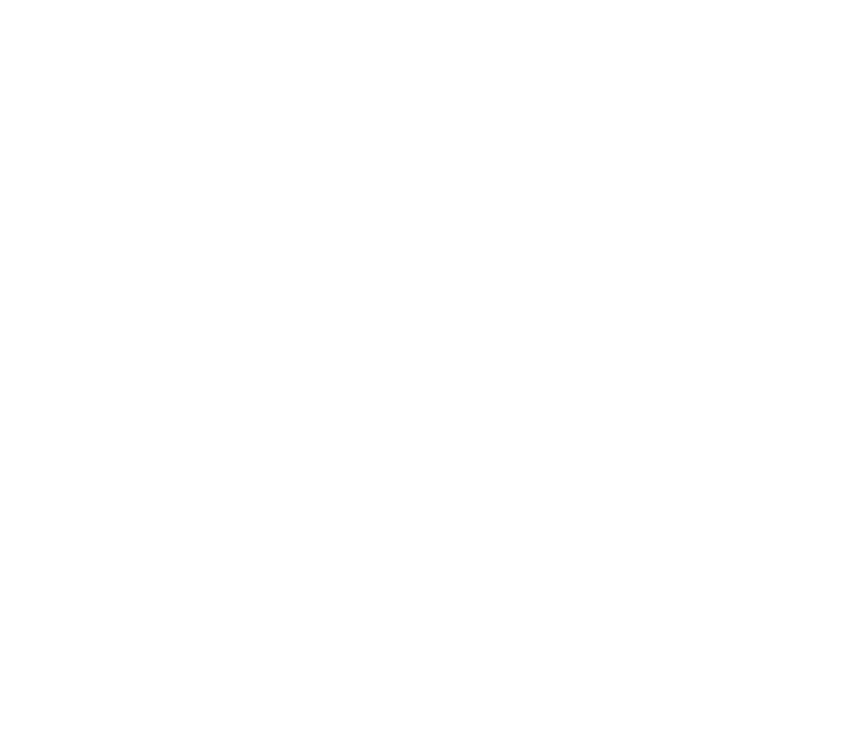
#### Marginal CDF

Marginal CDFs are like marginal PMFs, the CDF of just one of the random variables involved in a joint random variable.

We are considering to be here so that all possible values of can be covered. Similarly,

## Joint Continuous Random Variables

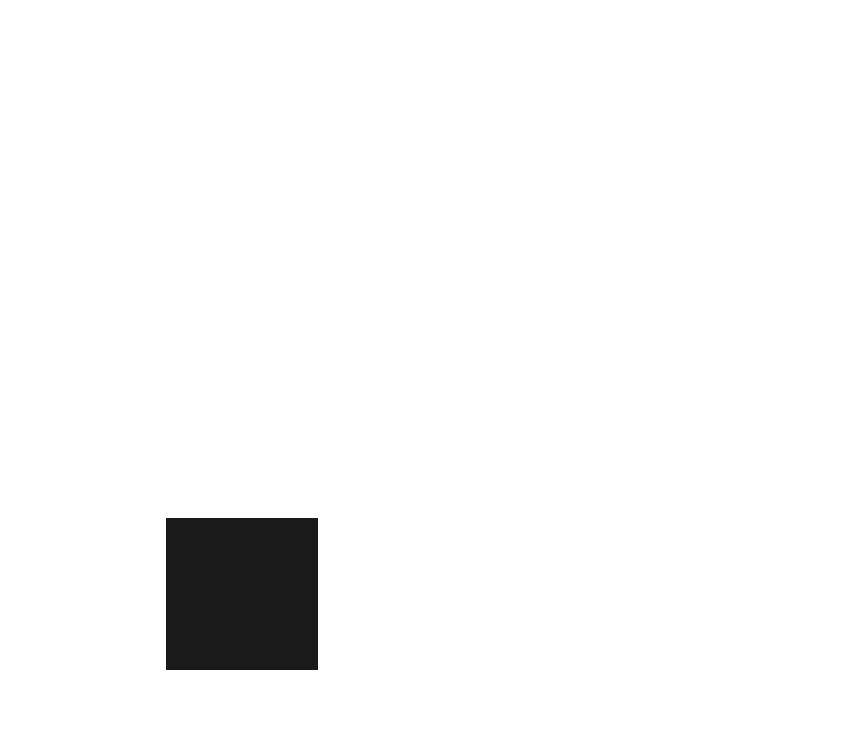
Say we have two random variables, , and , .



There is an infinite number of points within this region, which means like single continuous random variables, the probability of picking a single point from the region will be . Thus, JPMFs cannot be used.

### Joint Cumulative Distribution Functions

We can however, find the probability in a specific region, meaning JCDFs can be used. Say we divide the region into four parts.



Then, . Of course, this is assuming and are uniformly distributed over the region. In general,

Say , , , . Then,

For the region in the graph where either or , no probabilities are defined. Thus,

* If or ,

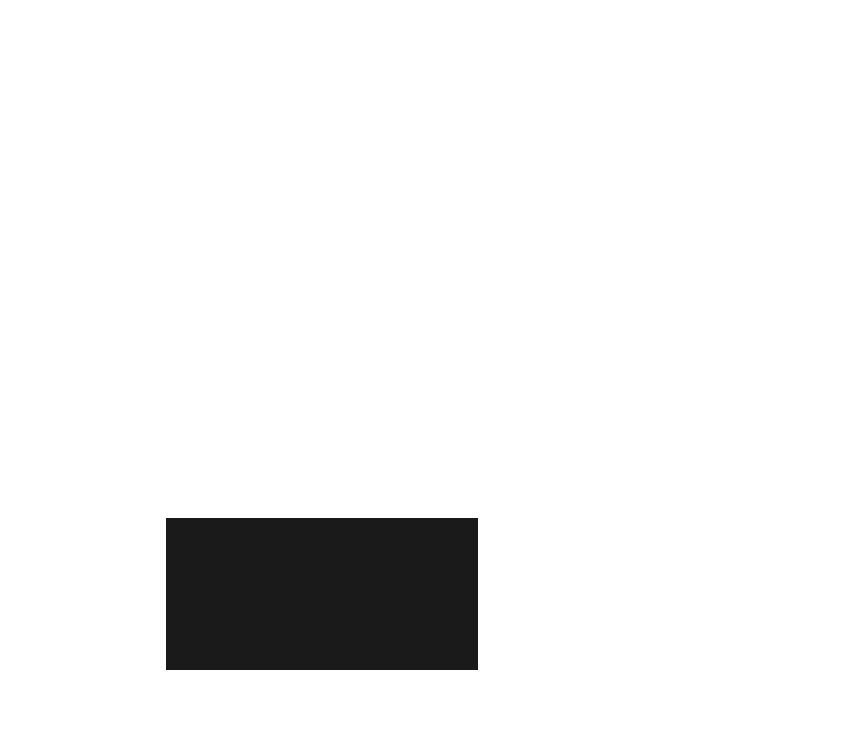
For the region in the graph where both and , the entire area of valid probabilities will be covered by the area of the JCDF. Thus,

* If and ,

Of course, these two properties are for the scenario where and are defined for the region . For other regions, the properties will change their numbers accordingly.

We have the two extreme cases. Now for the cases in between.

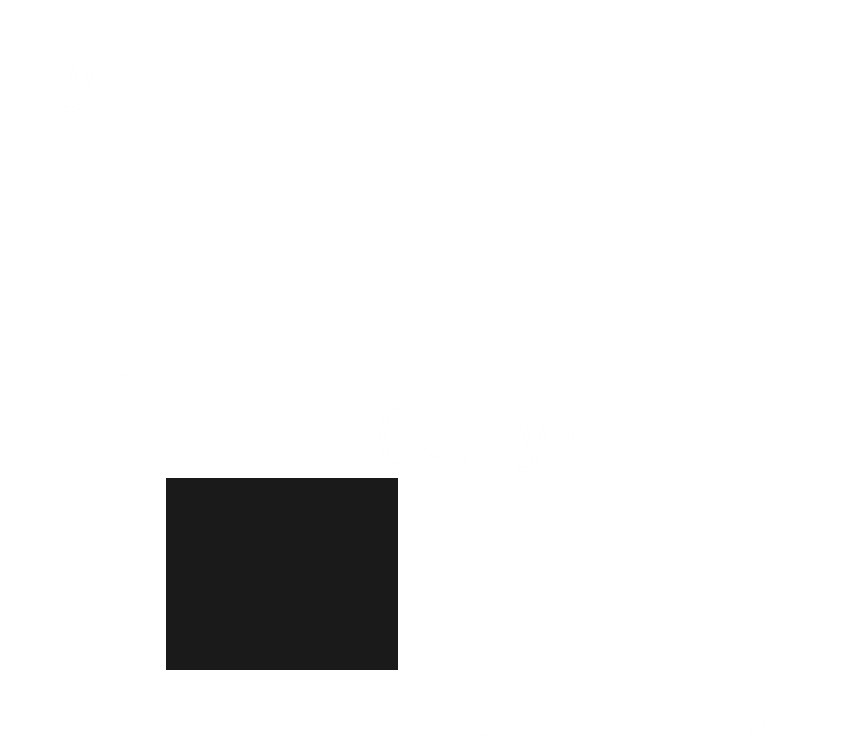
For the area where and , .



Similarly, for the area where and , .

Again, this is for the cases where and are defined between . Changing this will change the numbers, but the basic idea holds.

Finally, for a point such that and , , since the area covered is given by .



That was a lot of stuff, but we essentially just said this:

All that work, and that too for the simplest possible case of joint continuous random variables. Our life will get far easier if we use JPDFs.

### Joint Probability Density Functions

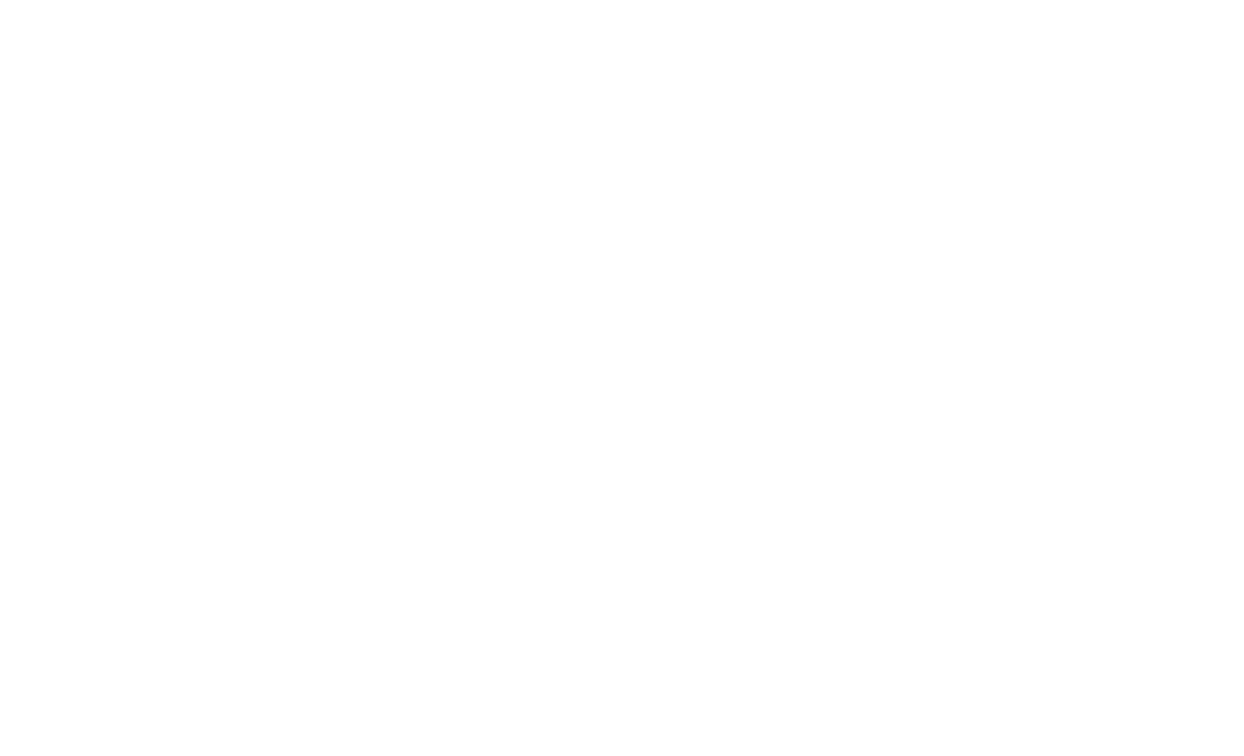
We know that the PDF is defined as .

Similarly, JPDF is defined as .

Thus, the JPDF is the second order derivative of the JCDF.

For the example we saw in the last section, where and were uniformly distributed over each,

If we draw the JPDF curve for this, it will be a cube. We have a 2D graph for and , and each point has a PDF value of , so the entire thing will have a height of on the -axis, thus forming a cube.

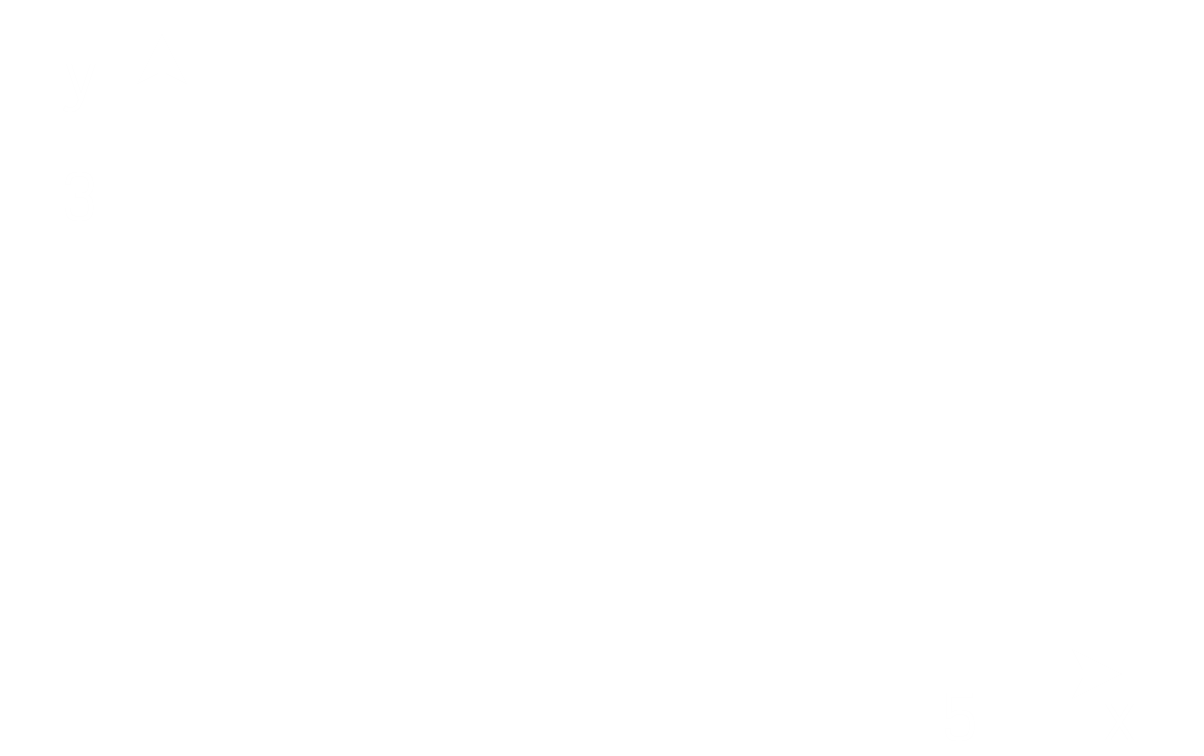


The properties of JPDFs are:

Example

1. Find the value of .
2. Find
3. Find

If we draw the graph for the range of values of and we have been given, it will look like this:

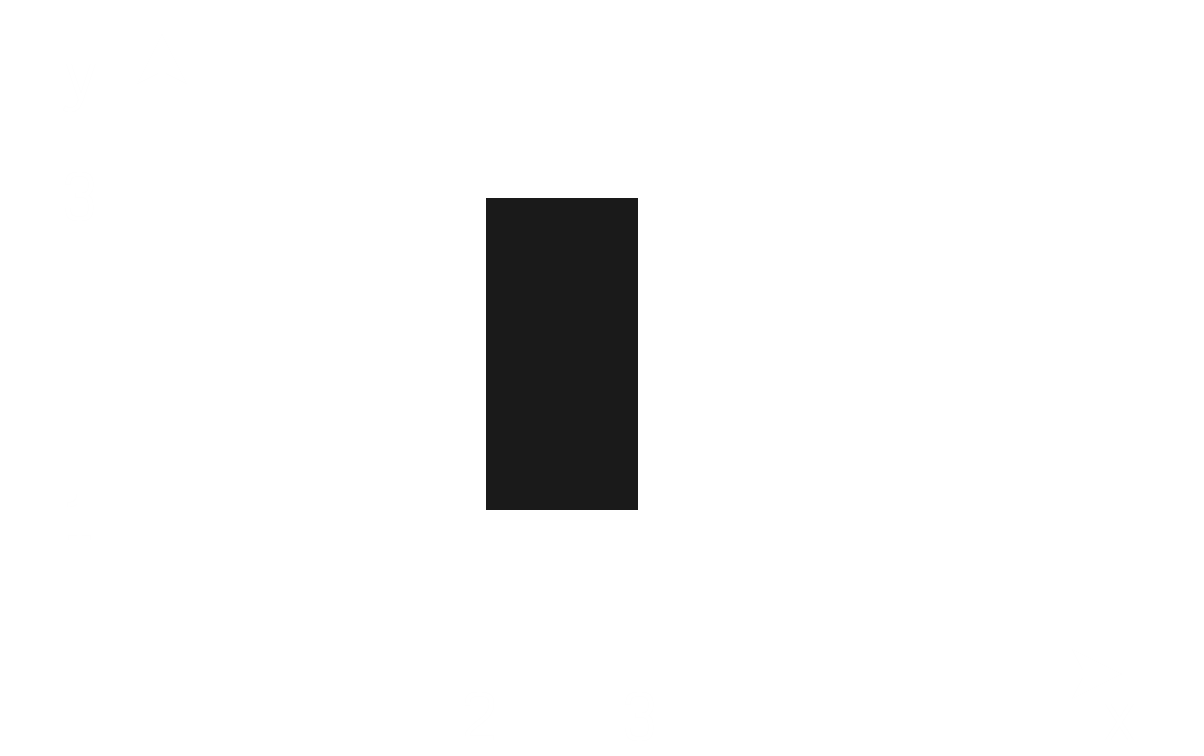


From this, we can tell that the area of the rectangle is .

Since the JPDF is a constant, this must mean that the random variable is uniformly distributed. Thus,

We can find this mathematically as well.

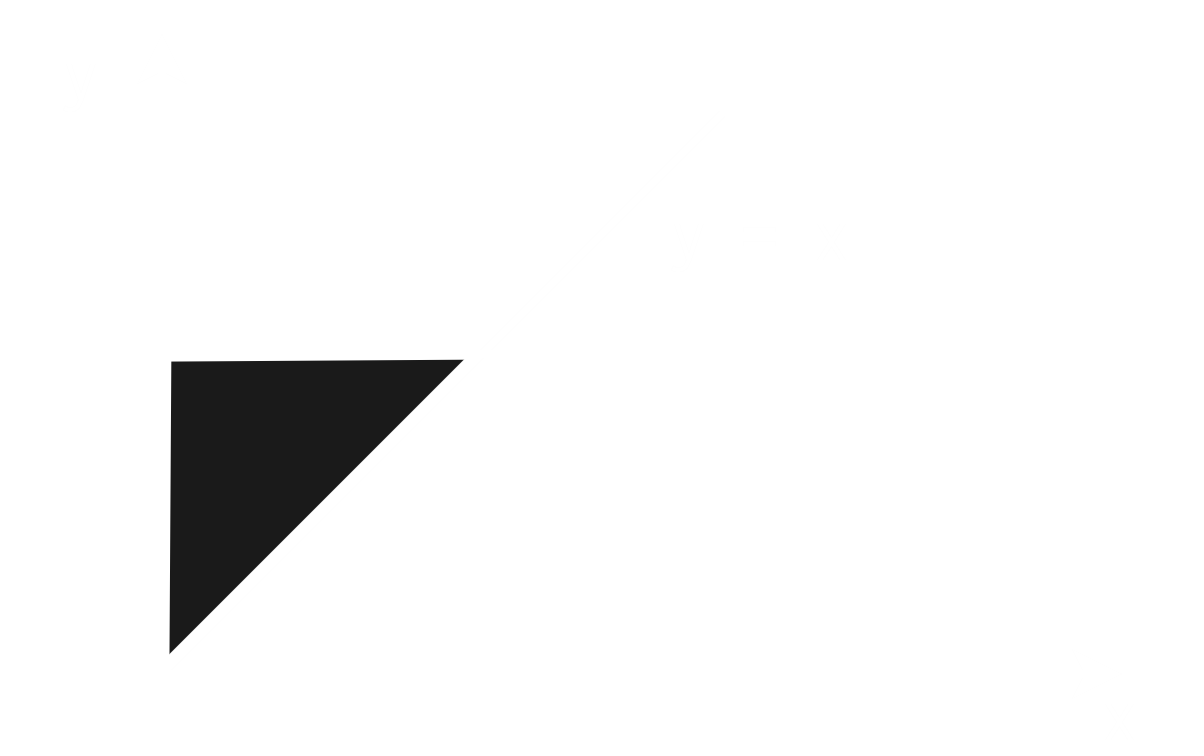
We could actually have reached this conclusion intuitively as well.



Since the area covered is and the random variable is uniformly distributed, . Of course, this simpler approach will not work if the random variable were not uniformly distributed.

This just goes back to proving that . This is because the probability is simply given by . For a uniform distribution, the probability density isn’t even the average, since it’s constant.

For the third question, consider this figure:



We essentially want to find the probability that we have a point in the shaded region.

From the diagram, we can tell that ranges from to . However, we need to pay attention to the range of . The range of is not from to , but rather from to , whatever the value of is.

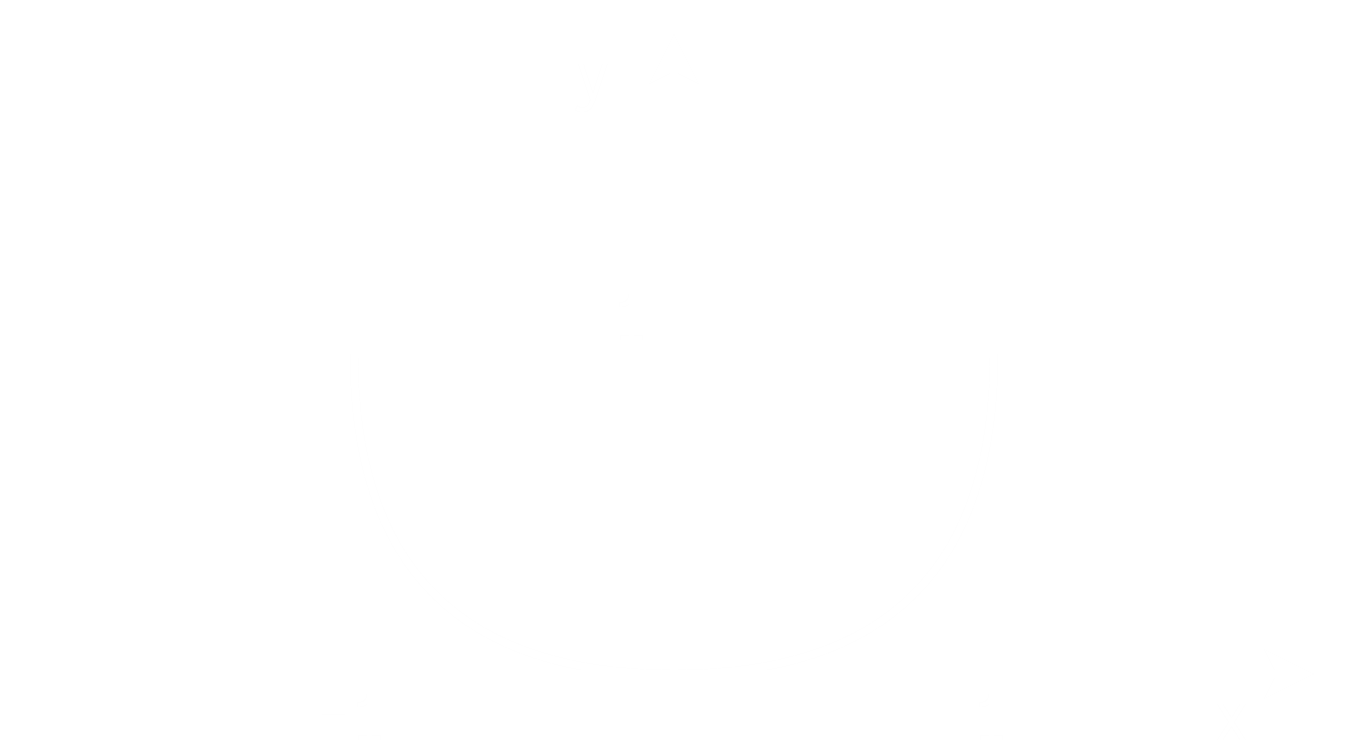
Alternatively, we can take a stupidly simply approach. The area of the shaded region is just , which means .

#### Marginal PDF

A marginal PDF is the individual PDF of one of the random variables involved in a joint random variable.

Example

Consider the figure for this:



For any given value of , ranges from to . Thus,

We know

For any given value of , ranges from to .

## Expected Values of a Function of Two Random Variables

Let be a function of two random variables that are jointly distributed.

For joint discrete random variables,

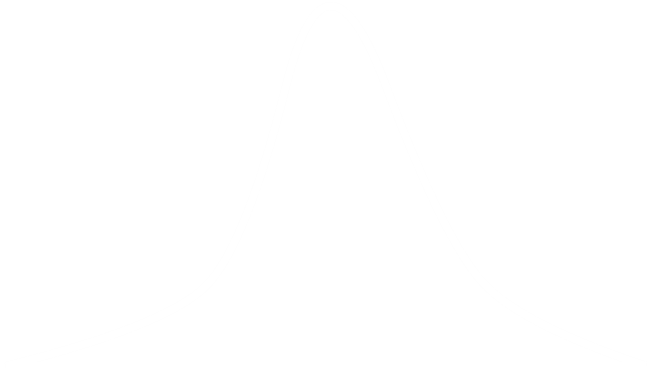
For joint continuous random variables,

This rule applies even if and are not jointly distributed.

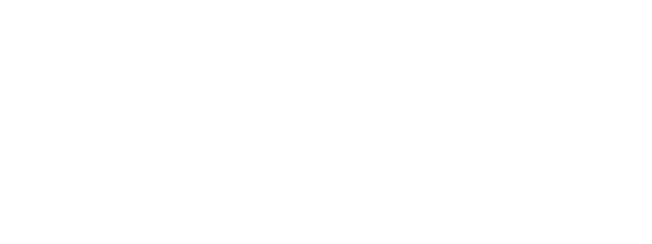
## Covariance and Correlation

We know that the variance of a random variable is given by

The variance represents the dispersion or spread of the distribution of about its expectation. As such,



this figure will have a lower variance, and



this figure will have a higher variance.

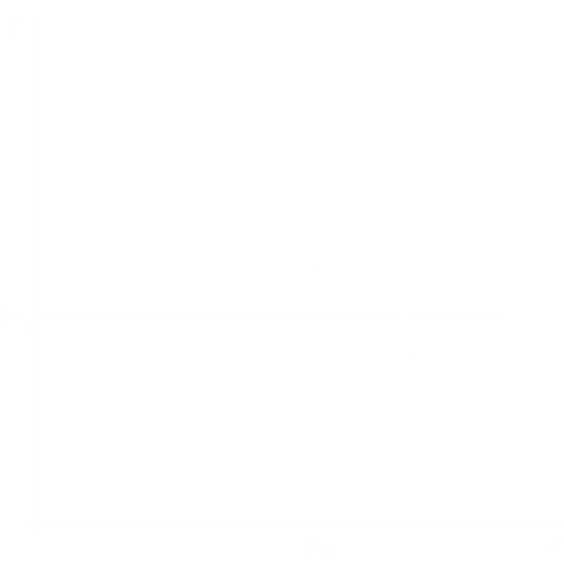
Essentially, we want to know how close the values are to the average value, , at the centre.

The variance of represents the spread of along the -axis and the variance of represents the spread of along the -axis. However, we are interested in the joint spread of the two random variables. This joint dispersion will be calculated in reference to the line .

If we assume and ,

This last term is called the covariance,

Consider a very specific scenario that looks like this:

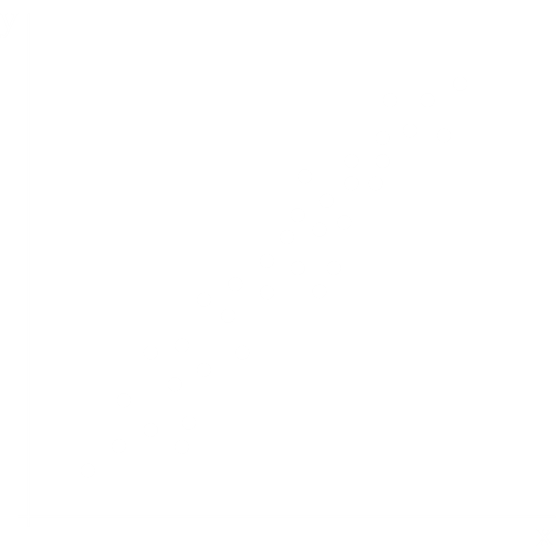


Here, for all the points that are on the right of , is positive but is negative.

For all points on the left of , is negative but is positive.

Thus, for all possible points, the value of is always negative. In this scenario, we say that the values are negatively correlated, meaning if one increases, the other decreases.

Similarly, we can have a scenario like this:



Here, the covariance will be positive and we can say the points are positively correlated.

There is a third scenario where we find the value of the covariance to be . Here, we say that the values are uncorrelated. The values are so scattered, that a correlation does not exist.

### Correlation Coefficient

The covariance is strongly dependant on the unit of measurement we use, so if we change the unit, the covariance my experience a large change. This is why we use the correlation coefficient to normalize the value of the covariance and bring it between to .

where and are the standard deviations of and respectively.

Note that is the symbol for the Greek letter (rho), not the letter .

Example

The value is also called the correlation. When the correlation has a value , we do not say they are uncorrelated, but rather we say they are orthogonal.