PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 06

Lecture - 31

Example of Bivariate Random Variable

So, we can define different random variables on the same sample space. Let us discuss an example. Consider the random experiment of tossing a coin twice. Basically, you have the same coin, and here we didn't mention the probability of getting a head. So, we assume P(Head) = 1/2.

Essentially, we assume it is an unbiased or fair coin. Now, what will the sample space be? In this case, the sample space S = {(Head, Head), (Head, Tail), (Tail, Head), (Tail, Tail)}. We are tossing a coin two times. So, that means we will get two observations for every run of this experiment.

It could be both heads, one head and one tail, or one tail and one head, or both tails. So, the possible outcomes are (Head, Head), (Head, Tail), (Tail, Head), or (Tail, Tail). This is the sample space, which includes all possible outcomes. In each run, we do not know which outcome will appear. So, now, let us define a random variable X.

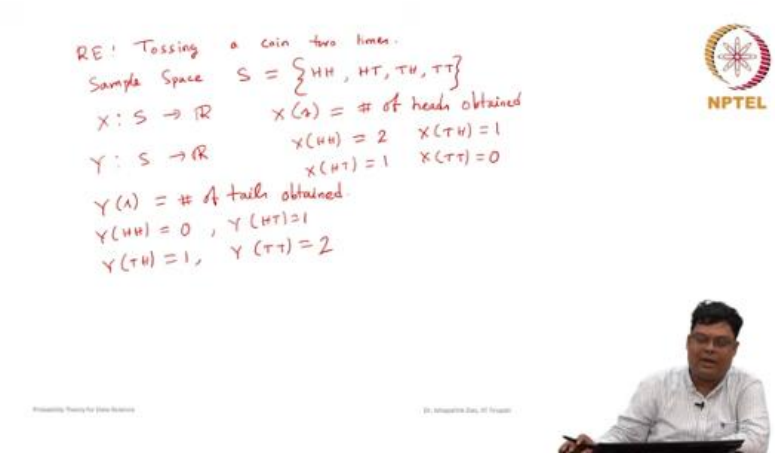
It will be a measurable function from S to $\mathbb{R}$, defined by $X(\omega)$ = number of heads observed or obtained, where $\omega \in S$. So, basically, whenever tossing a coin, we obtain X as the number of heads. For example, if we get two heads, then X(Head, Head) = 2 because the number of heads is 2. X(Head, Tail) = 1 because there is only one head. X(Tail, Head) = 1, and X(Tail, Tail) = 0, as there are no heads.

Another random variable we can define is Y, from S to $\mathbb{R}$. This is defined by $Y(\omega)$ = number

of tails observed, where $\omega \in S$. So, in that case, Y(Head, Head) = 0 because we are talking about the number of tails. What is the number of tails? Since two heads are observed, the number of tails will be 0.

Y(Head, Tail) = 1. Y(Tail, Head) = 1. And Y(Tail, Tail) = 2. Sorry, I mean for (Head, Tail), the number of tails is 1, that's why it's 1. For (Tail, Head), the number of tails is 1, that's why it's 1. For (Tail, Tail), the number of tails is 2.

So, this is actually the opposite of X; here, we have to consider the number of tails for this random variable. So, we have considered two random variables. Now, we define a bivariate random variable. So, let's look at the definition of a bivariate random variable. Let S be the sample space of a random experiment.



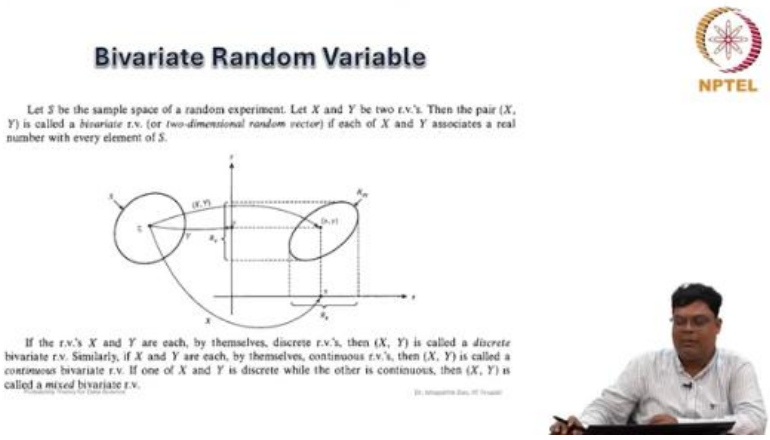
For example, here, S is the sample space. Let's consider this example. So, S = {(Head, Head), (Head, Tail), (Tail, Head), (Tail, Tail)}. Now, by the random variable X, suppose this is 0, this is 1, this is 2, this is 3, this is -1, -2, -3. Here, this is 1, this is 2, this is 3, and so on. Now, X(Head, Head) = 2, so this is okay. I'm just mapping X on the x-axis and this on the y-axis.

So, X(Head, Head) = 2. This is 2, which is X(Head, Head). Now, X(Head, Tail) = 1, so this will be 1. This is also 1. So, this is X(Head, Tail), Head. Now, this is X(Head, Tail). X(Head, Tail) = 1. X(Tail, Head) = 1, and for (Tail, Tail), it is 0. So, X(Tail, Tail) = 0. Now, what about the random variable Y? For the random variable

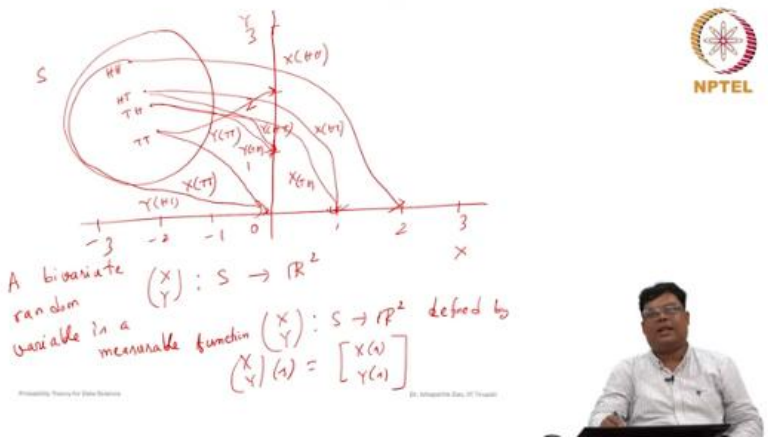
Y, X(Tail, Tail) = 2, so Y(Tail, Tail) = 2. So, this is going here. This is Y(Tail, Tail), and both are going to 1 here. This is the Y-axis, and this is 1. This is also 1. So, we will represent Y(Tail, Head). This is Y(Tail, Head), and this is Y(Head, Tail).

Then, this value will go to 0. Y(Head, Head) will also go to 0. So, this is nothing but Y(Head, Head). This is a little complicated in the graph, but we have understood it. Now, why are you drawing this graph? Because in this definition, let S be a sample space, and I'm going to explain.

Let X and Y be two random variables. Then, the pair (X, Y) is called a bivariate random variable, a two-dimensional random variable. If each of X and Y associates a real number, then every element of S will be associated with X and Y. So, now we define this random variable. Usually, X, like any random variable, is a function from S to $\mathbb{R}$.



Now, we consider this as a vector. This is denoted by (X, Y), which is a function from S to $\mathbb{R}^2$. This defines a bivariate random variable. A random variable is a measurable function. (X, Y) is defined from S to $\mathbb{R}^2$ by (X, Y)(s).



This is a function from S to $\mathbb{R}^2$, so for every element of S, it will map to $\mathbb{R}^2$. The value will be X(s), Y(s). So, X(s) is a real number and Y(s) is a real number, so this is a vector in $\mathbb{R}^2$. This is called a bivariate random variable. Now, if both random variables are discrete, then the random variable (X, Y) is called a discrete bivariate random variable.

If both X and Y are continuous random variables, then (X, Y) is called a bivariate continuous random variable. If one of them is discrete and the other is continuous, then it is called a mixed bivariate random variable. Now, if we graphically represent this, what is the range of X? The range of X is the range of the random variable X, which is {0, 1, 2}. What is the range of Y?

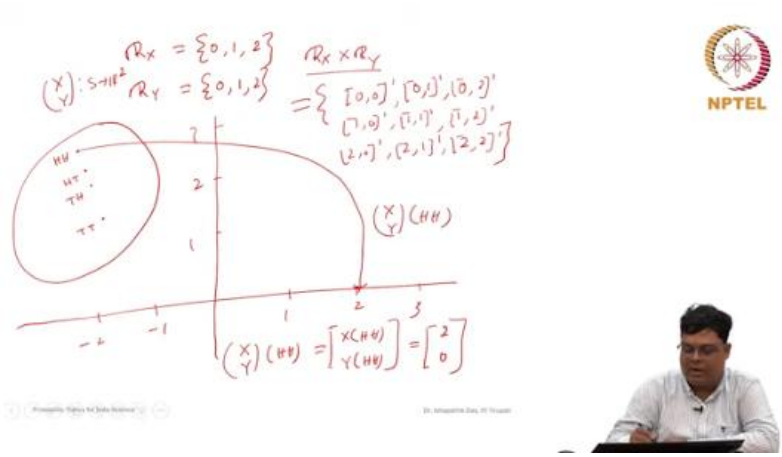
The range of Y is {0, 1, 2}. So, if you consider the cross product, that is, if you consider $\mathbb{R}_x \times \mathbb{R}_y$, this is nothing but it contains 3^2 elements, which means it will consider 9 elements. So, basically, this is (0, 0), one vector. For simplicity, let us denote this as a vector in the first bracket, to indicate that this is a vector. But actually, if we properly represent that, we usually use the third bracket and this transpose dash.

This is nothing but the row-column vector represented as a column. So, now this is (0, 0), $(0, 1)^T$, then $(0, 2)^T$, and then (1, 0), (1, 1), (1, 2), and then (2, 0), (2, 1), (2, 2). So, this is the range of this random variable. Sorry, the range is not—it is nothing but $\mathbb{R}_x \times \mathbb{R}_y$. Now, let us again find out what this X, Y is.

So, this is the sample space: head-head, tail-head, head-tail, tail-head, and tail-tail. So now, this is 1, this is 2, this is 3, -1, -2, and so on, 1, 2, 3. Now, under this random variable X, Y, this function is from S to $\mathbb{R}^2$. So, how is it? What is the (X, Y) of head-head? This is defined by X(head-head), Y(head-head).

So, X(head-head) is nothing but 2, and Y(head-head) is equal to 0 because X(head-head) represents the number of heads. Sorry, let me just write this down again. Let us find out what (X, Y) of head-head is, one of the elements of S. By definition, this is nothing but X(head-head) and Y(head-head). So, X(head-head) is the number of heads obtained, which is 2, and Y(head-head) is 0.

Therefore, this is (2, 0). So, basically, this is coming to (2, 0). This point is mapped to this point, which is nothing but X, Y of head-head. Now, what is the value of (X, Y) of head-tail? (X, Y) of head-tail is nothing but X(head-tail) and Y(head-tail)



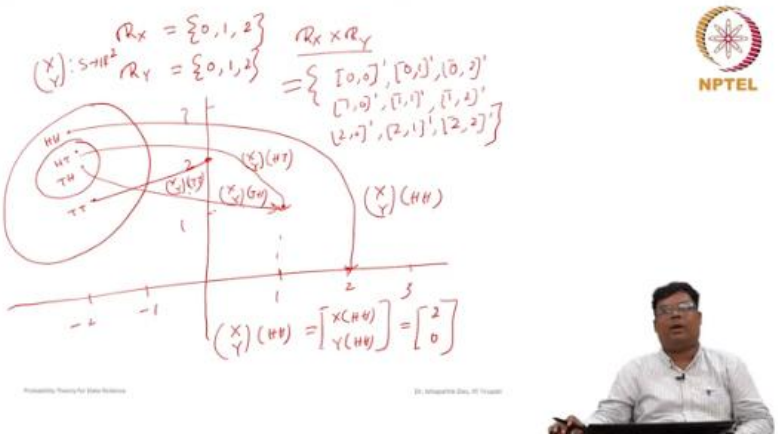
So, X(head-tail) = 1, and Y(head-tail) = 1. Similarly, what will be (X, Y) of tail-head? This point is X(tail-head), Y(tail-head), which is again (1, 1). Now, if you consider the point tail-tail, X(tail-tail) and Y(tail-tail) are 0 and 2, respectively. So, the result is (0, 2).

Now, if you represent this graphically, (X, Y) of head-head and (X, Y) of tail-tail both go to the same point, (1, 1). So, here is the point (1, 1). Both head and tail go to the same

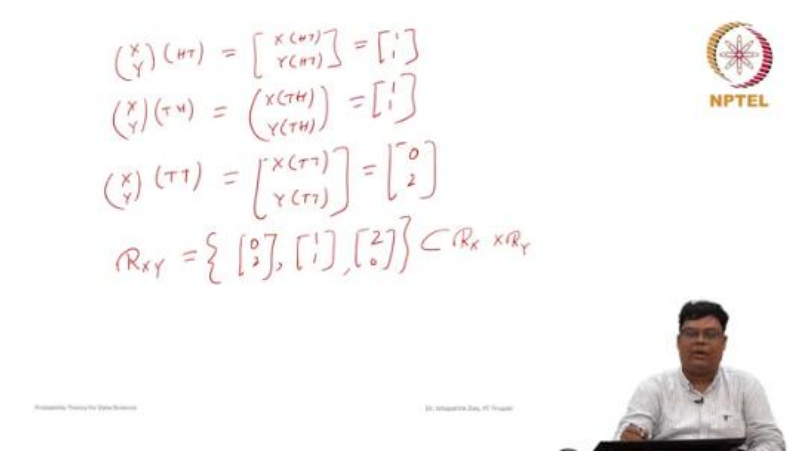
location. This is also true for (X, Y) of head-tail and (X, Y) of tail-head. Now, what will be the (X, Y) of tail-tail?

This function, this value, is nothing but (0, 2). So, (0, 2) is here. This is going to this value. This is (X, Y) of tail-tail. So, what did we find?

What is the range of this bivariate random variable? The range of this bivariate random variable is denoted by $\mathbb{R}_{x\gamma}$. $\mathbb{R}_{x\gamma}$ will contain the points: (0, 2), which is one value; (1, 1), and another point is (2, 0). However, you can see that $\mathbb{R}_x \times \mathbb{R}_\gamma$ contains many more points, not just the three points like (0, 2), (2, 0), and (2, 2). These three points are part of the range of the bivariate random variable.

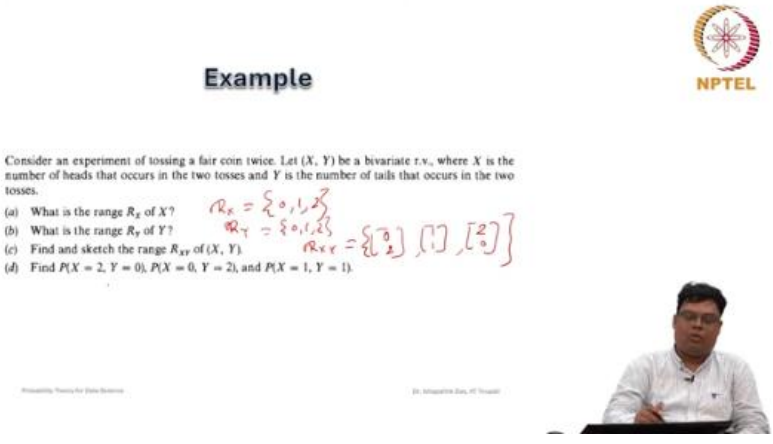


So, $\mathbb{R}_{x\gamma}$ may not be the same as $\mathbb{R}_x \times \mathbb{R}_\gamma$, but it will always be a subset of the range of $\mathbb{R}_x \times \mathbb{R}_\gamma$. So, it may sometimes be exactly equal, but most of the time, as we can see from this example, it will be a subset of $\mathbb{R}_x$. The range of this bivariate random variable, $\mathbb{R}_{x\gamma}$, will be a subset of $\mathbb{R}_x \times \mathbb{R}_\gamma$. I hope you have understood the definition of the bivariate random variable. Now, let's consider the example we have already discussed: the experiment of tossing a fair coin twice.



Let X and Y be a bivariate random variable, where X is the number of heads that occur in two tosses and Y is the number of tails that occur in the two tosses. What is the range, $\mathbb{R}_x$, of X? We have already discussed that the range of X is {0, 1, 2}. Similarly, the range of Y, $\mathbb{R}_y$, is also {0, 1, 2}. We have already discussed and found the range of $\mathbb{R}_{xy}$, which consists of the points (0, 2), (1, 1), and (2, 0).

All these points are in $\mathbb{R}^2$. Now, the question is: What is the probability that X = 2 and Y = 0? So, this notation, which we have already discussed, will be explained again. Let us first find the probability that X = 2 and Y = 0. As we have already discussed, the notation X = 2 means all s in the sample space S such that X(s) = 2.



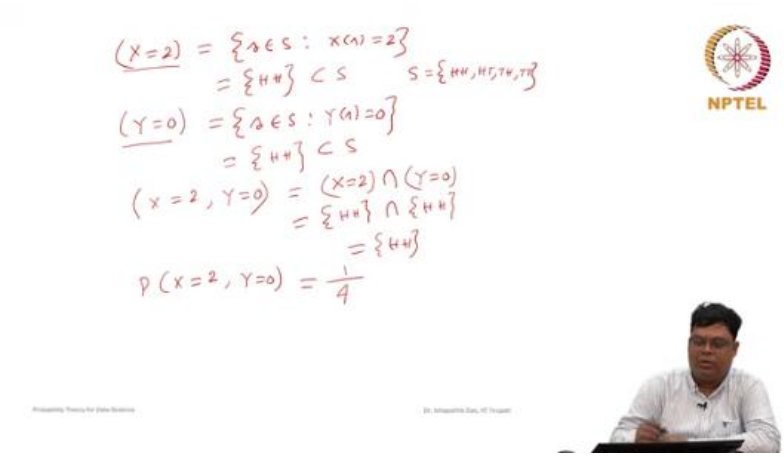
Another question is: What does it mean for Y = 0? So, Y = 0 is also known. Now, let's find out what those values are. Remember that the sample space contains these four points: S =

{head-head, head-tail, tail-head, tail-tail}. We are looking for all s ∈ S such that X(s) = 2. X(s) = 2 means the number of heads will be 2, which corresponds to the point head-head.

Now, the next question is what Y = 0. We also know this. This refers to all s ∈ S such that Y(s) = 0. So, Y(s) = 0 means the number of tails is 0. Therefore, we are looking for the point where the number of tails obtained is 0. This means the point is head-head. So, this corresponds to head and head. Now, if you want to define what this means—X = 2, Y = 0—this is the event corresponding to this, because we already know what the event is. This is a subset of S, and it is also a subset of S.

So, we already know that it will be a subset of S. But what does this notation mean? This notation corresponds to X = 2, which is an event in S, and Y = 0, which is also an event in S. The comma means the intersection of these two events. So, we have found that this is nothing but head-head, and this is the intersection with head-head again.

So, this is simply head-head. So now we can easily find the probability that X = 2 and Y = 0. This probability is simply 1/4, because there are four equally likely points, and we assume the probability of heads is 1/2, as it's an unbiased coin. So, this is 1/4. So in general, we can define this by stating that X = x and Y = y.



X is the random variable, and Y is another variable. We can also talk about cases like X ≤ x and Y = y, or we can discuss situations like X = x and Y ≤ y. How is it defined? If you want to define it, we can define it as an event Ax such that this is equivalent to X = x. How is it defined? All s ∈ S such that X(s) = x, and now, Y is another event.

It is defined by $Y = y$, which means all $s \in S$ such that $Y(s) = y$. Small y is a real number. So, Bx, Ax, and By are all subsets of S. Therefore, they form an event. This is equivalent to the event $Ax \cap By$.

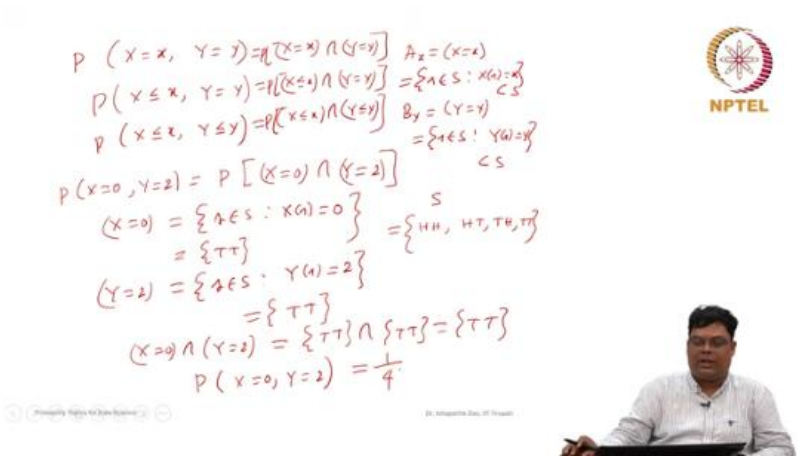
For simplicity and clarity, we use this notation, or you can directly write it like this. So, this corresponds to the event where $X = x$. For an enumerated random variable, we have already discussed that this is an event. Similarly, the intersection where $Y = y$ is another event. In the same way, the event $X \leq x \cap Y = y$. Likewise, the event $X \leq x \cap Y \leq y$.

So, we can now talk about this probability, which refers to the probability of this event. The probability of this event is what we are focusing on. It is important to understand the notation being used here. Once you follow this notation, we can proceed with other examples, such as finding the probability that $X = 0$ and $Y = 2$.

So, the probability that $X = 0$ and $Y = 2$ is nothing but the probability of the event where $X = 0 \cap Y = 2$. Now, what is the event when $X = 0$? $X = 0$ refers to all $s \in S$ such that $X(s) = 0$. This means $X(s) = 0$ for all s in the sample space. Here, X represents the number of heads. So, what is the sample space?

The sample space S is already known and consists of: head-head, head-tail, tail-head, and tail-tail. So, what are the points in S such that $X(s) = 0$? This corresponds to tail-tail because the number of heads is 0. Tail-tail is the only outcome where the number of heads is 0. Now, how do we find $Y = 2$? It is all $s \in S$ such that $Y(s) = 2$. Here, Y is defined as the number of tails. When the number of tails equals 2, the outcome is again tail-tail (T, T).

So, this is tail-tail (T, T). Then, $X = 0 \cap Y = 2$. In this case, both events are the same. The same was true in the previous example as well. Hence, the probability is 1/4. Here too, $X = 0$ and $Y = 2$. This represents the probability of this event. This event is nothing but, by the classical approach, there are four equally likely points, and out of that, one point satisfies the condition. So, the probability is 1/4. So, the next example is to find the probability of $X = 1$ and $Y = 1$. How do we find that?



The probability that X = 1 and Y = 1 is nothing but the probability that X = 1 ∩ Y = 1. We have to find the corresponding events, then take the intersection to find the probability. So, let's write down what S is. S = {head-head, head-tail, tail-head, tail-tail}. Now, what is X = 1?

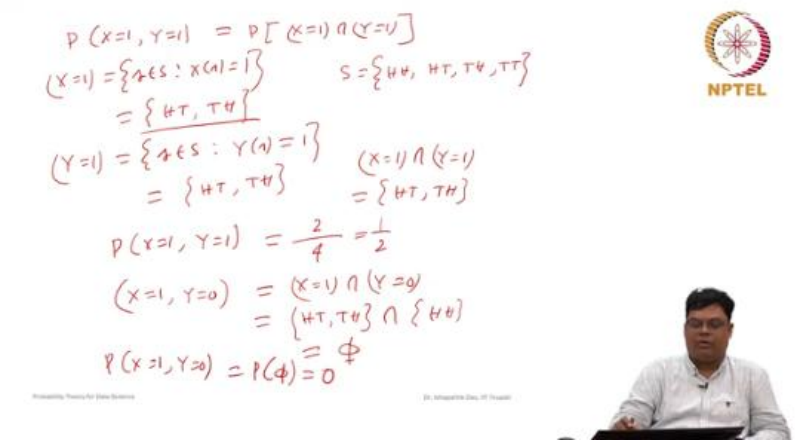
X = 1 refers to all s ∈ S such that X(s) = 1. This means the number of heads will be 1. So, the outcomes where the number of heads is 1 are head-tail and tail-head. Similarly, if you consider Y = 1, it refers to all s ∈ S such that Y(s) = 1. In this case, the number of tails will be 1, which gives head-tail and tail-head.

In fact, for all the examples, it is the same. So, there is no intersection; it is simply head-tail and tail-head. Now, what will the probability be? Probability of X = 1 ∩ Y = 1 is 2 out of 4 equally likely points, so the probability is 2/4, which simplifies to 1/2. It seems like this is always the case.

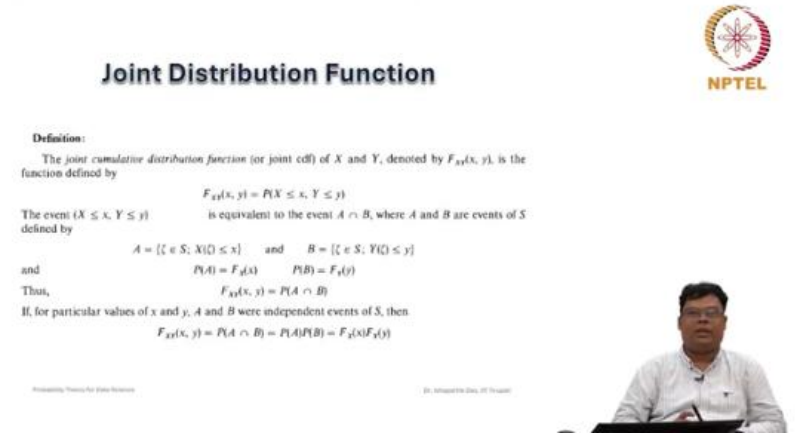
Now, let's consider the case where we need to find the probability of X = 1 and Y = 0. To find this probability, X = 1 corresponds to this event, which is the intersection of X = 1 ∩ Y = 0. So, X = 1 corresponds to head-tail and tail-head. Now, what is Y = 0? Y = 0 means the number of tails is 0, so this corresponds to the intersection of head-head.

So here you see that not all are the same, and that’s why Y = 0. You can find that Y = 0 corresponds to all s ∈ S such that Y(s) = 0. Since Y represents the number of tails, Y = 0 means there are no tails. So, the only outcome is head-head. This is correct.

However, the intersection is empty, meaning it's the null set. Hence, the probability that X = 1 ∩ Y = 0 is the probability of the null set. It is not possible, so the probability is 0. So, you are tossing two coins. You can get one head or zero heads.



Both events cannot happen simultaneously because if one head occurs, then you cannot have both one head and one tail. If one head occurs, the number of tails will be 0. Therefore, this is not possible, and the probability will be 0. This is one example. Now, we will introduce the joint distribution function.



So, the joint distribution function. Now, you can just remember what the cumulative distribution function is for the univariate case. As we have discussed, we have already defined the cumulative distribution function for a univariate random variable. Let X be a random variable. So, it is a function from the cumulative distribution function.

Recall that the cumulative distribution function of a univariate random variable X is defined as F_X(x), which is the probability that $X \leq x$, where $x \in \mathbb{R}$. Now, we have a bivariate random variable. So, how do we define this bivariate random variable? The joint cumulative distribution function, or joint CDF, is defined as the joint CDF. We say the joint CDF of a bivariate random variable.

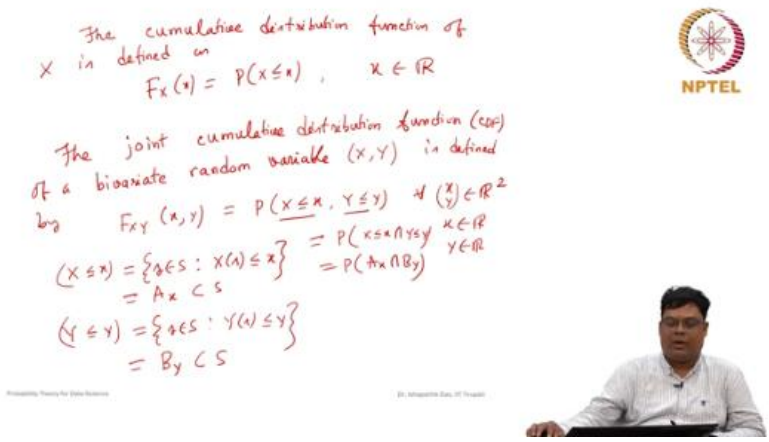
Usually, we write it as (X, Y) or in third brackets, (X, Y). But for simplicity, sometimes we will write it as [X, Y]. A bivariate random variable X, Y is defined as follows: it involves two variables. F_{X,Y}(x, y), where both x and y are real numbers, is the probability that $X \leq x$ and $Y \leq y$. What is a comma? We have already discussed that for all x, y is a vector in $\mathbb{R}^2$, or you can say that $x \in \mathbb{R}$ and $y \in \mathbb{R}$.

Then (x, y) is an ordered pair in $\mathbb{R}^2$. So now, what is $X \leq x$, $Y \leq y$? So, $X \leq x$, we know it's an event. This is nothing but all $s \in S$ such that $X(s) \leq x$, because X and Y are defined on the same sample space, and $X(s) \leq x$. And this is, suppose we denote it by an event A_x, because it depends on the function X.

Similarly, $Y \leq y$ is again a subset of S. It is an event where all $s \in S$ such that $Y(s) \leq y$. So, this is nothing but we denote it as B_y, which is a subset of S. It is an event. We assume that because X is a random variable, it is a measurable function, so it will be in the sigma field.

Similarly, since Y is also a measurable function, B_y will be in the sigma field. Therefore, this is also an event, and it is a subset of S. So, $A_x \cap B_y$ is in the sigma field. Therefore, $A_x \cap B_y$ will be inside the sigma field. The probability is nothing but the probability that $X \leq x$ and $Y \leq y$, which is the same as the probability of $A_x \cap B_y$ for all x and y.

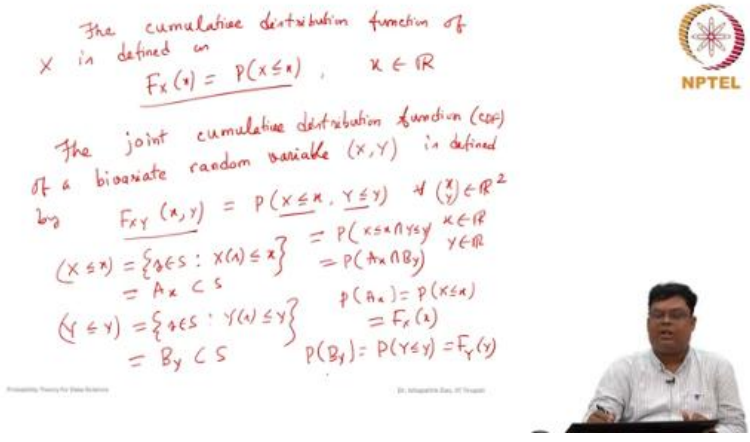
This is the definition of the cumulative distribution function. So, let us consider the joint cumulative distribution function, or joint CDF, of X and Y, denoted by Fxy(x, y). This function is defined as Fxy(x, y), which is the probability that $X \leq x$ and $Y \leq y$. Now, the event $X \leq x$, $Y \leq y$, is equivalent to the event A. I have denoted it this way because it depends on x and y.



So, A_x ∩ B_y, where A and B are events in S, are defined as follows: A is all s ∈ S such that X(s) ≤ x, and B is all s ∈ S such that Y(s) ≤ y. We have already defined this. So, this is A_x, and this is B_y. Now, what is the probability of A_x? So, the probability of A_x is nothing but the probability that X ≤ x.

We have already used this notation. This is nothing but the cumulative distribution function of X. Similarly, the probability of B_y is denoted by the probability that Y ≤ y. This is nothing but the cumulative distribution function of Y. So, whenever we are talking about a bivariate random variable, suppose we denote F_{X,Y}(x, y) as the bivariate cumulative joint distribution function of the bivariate random variable.

Then, the cumulative distribution function of the subset of the random variable, in this case, X and Y, is defined. So, the cumulative distribution function of X and the cumulative distribution function of Y, we call it the marginal distribution. It's just a name, it's usually called the marginal distribution function. So, P(A_x) is F_X(x), P(B_y) is F_Y(y), and F_{X,Y}(x, y) is nothing but A_x ∩ B_y. This depends on Y, and that's why for particular values of (X, Y), if A and B are independent events, then this will happen.

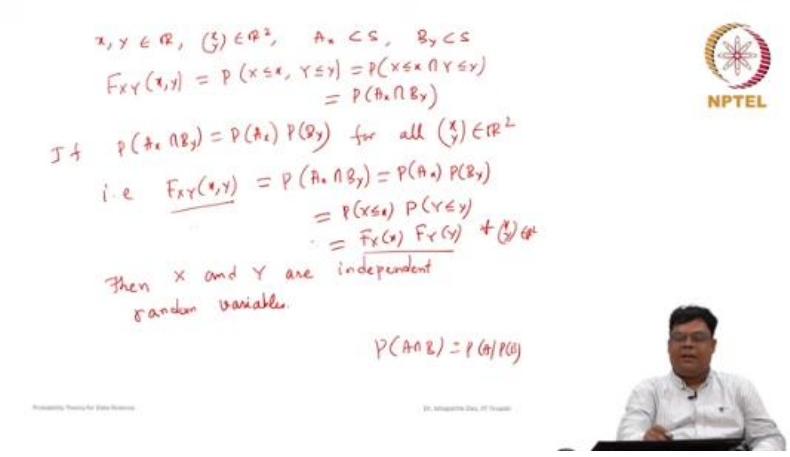


Now, we will discuss some important properties. Let's suppose we say two events are independent. Suppose for any real numbers x and y, where x, y ∈ $\mathbb{R}$. So, basically, we are considering (x, y) ∈ $\mathbb{R}^2$. A_x is a subset of S, and B_y is also a subset of S. Both are in the sigma field. We are discussing that A_x ∩ B_y represents the joint distribution function, which is nothing but the probability that X ≤ x and Y ≤ y. This is denoted as P(X ≤ x, Y ≤ y), which we denote by A_x ∩ B_y.

Now, for fixed x and y, A_x is an event and B_y is an event. A_x and B_y will be independent if P(A_x ∩ B_y) = P(A_x) * P(B_y).

So, now, if this is true for all (x, y) ∈ $\mathbb{R}^2$, then F_{X,Y}(x, y) = P(A_x ∩ B_y) = P(A_x) * P(B_y). The probability of A_x, where X ≤ x, is the cumulative distribution function of X, denoted as F_X(x). The probability of B_y, where Y ≤ y, is the cumulative distribution function of Y, denoted as F_Y(y), for all x, y ∈ $\mathbb{R}^2$.

So, if this happens, then the bivariate random variables X and Y are called independent. So, this is the definition. We know that the definition of two independent events, A and B, is that P(A ∩ B) = P(A) * P(B). But here, X and Y are also involved. For particular values of x and y, it may be true, but we cannot say that X and Y, as random variables, are independent. If this is true for all x, y ∈ $\mathbb{R}^2$, or for any real number x and any real number y, then these two random variables, X and Y, will be called independent random variables. That is why this is the definition. If, for particular values of x and y, we have independent events, then this is true.



However, we say the events are independent when it holds true for all x and y. If it is true for all x and y, then we say that X and Y are independent random variables. We will discuss in the next part some of the concepts of the independence of two random variables. We

also need to discuss the properties of bivariate random variables, as well as the properties of the joint cumulative distribution function of univariate random variables.

Next, we want to discuss the properties of F(x, y). We have already discussed the properties of F_X(x), such as the univariate random variable and the marginal distribution function, and we are familiar with some of those properties. Now, we want to discuss the properties of the cumulative distribution function. Additionally, we will go over some numerical examples of the bivariate cumulative distribution function and the joint cumulative distribution function of the bivariate random variables X and Y.