PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

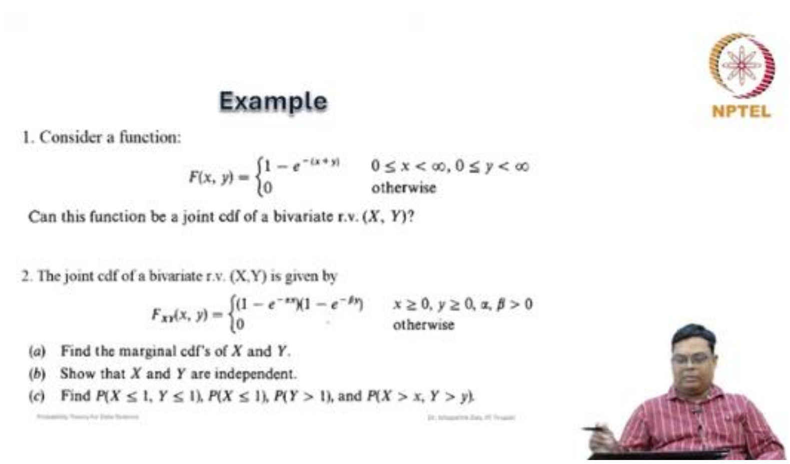
Indian Institute of Technology Tirupati

Week - 07

Lecture - 34

Examples of Joint Cumulative Distribution Functions, Marginals, and Independence

Next, this is another example. Here, it is not explicitly asked, but you can check that it satisfies all the relationships and properties. Here, it is asked that the joint cumulative distribution function (CDF) of a bivariate random variable (X, Y) is given by $1 - e^{-\alpha x} * (1 - e^{-\beta y})$. Let us consider $x \geq 0$ and $y \geq 0$. The joint cumulative distribution function, $F_{XY}(x, y)$, is defined for a bivariate random variable.



It is defined as $1 - e^{-\alpha x} * (1 - e^{-\beta y})$. Whenever $x \geq 0$ and $x < \infty$, $y \geq 0$ and $y < \infty$, with $\alpha > 0$ and $\beta > 0$, and 0 otherwise. This is the cumulative distribution function of a bivariate random variable. You can check that when $x \to \infty$, this approaches 1, and when $y \to \infty$, this also approaches 1. That is why it goes to 1 as x and y approach infinity. It will always be between 0 and 1 because $e^{-\alpha x}$, for positive x and α, will be between 0 and 1. Since α and β are positive, $1 - e^{-\alpha x}$ will also be between 0 and 1. We can also check the other properties of a cumulative distribution function (CDF). Now, the question is that it is not asked, but it is already given that this is the cumulative distribution function of a joint bivariate random variable. The task is to find the marginal cumulative distribution functions of X and Y.

'Marginal' refers to the distribution of a subset of the random variable. So, what is the formula for the marginal distribution? The marginal cumulative distribution function of X is given by $F_X(x)$. This is nothing but the probability that $X \leq x$. So, how can we find this value?

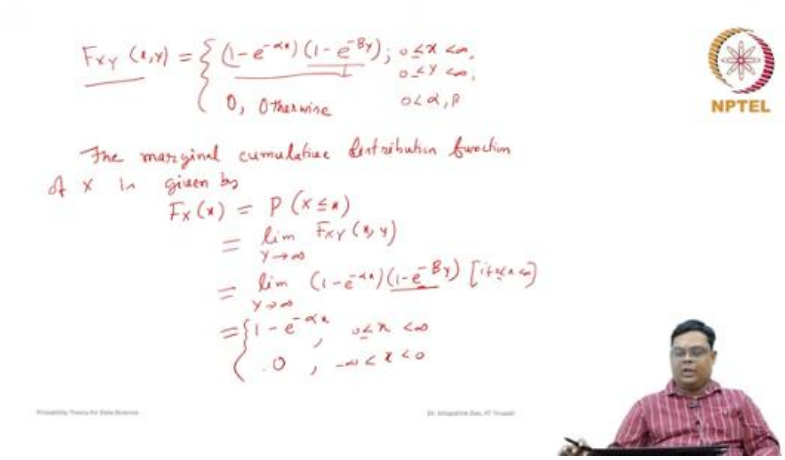
As we have already discussed, it can be found by taking the limit of the other coordinate, where $y \to \infty$. The limit of $F_{XY}(x, y)$ as $y \to \infty$ is nothing but $F_X(x)$. Now, if you take the limit as $y \to \infty$, you can see that this is straightforward. Since $e^{-\beta y}$, with β being a positive number, approaches 0 as $y \to \infty$, this becomes $1 - e^{-\alpha x}$. Another step is to write the limit as $y \to \infty$.

Since this is a positive number, it will be $\alpha x * (1 - e^{-\beta y})$. So, the expression becomes $1 - e^{-\alpha x} * (1 - e^{-\beta y})$. So, as $y \to \infty$, since β is a positive number, this goes to 0, and the expression goes to 1. So, this is nothing but $1 - e^{-\alpha x}$. This holds when $x \geq 0$. If $x < 0$, we cannot take this function, and it will be 0. So, this is valid when $x \geq 0$ and $x < \infty$. So, we can write it like this. Finally, we can say that this is equal to $1 - e^{-\alpha x}$ when $x \geq 0$ and $x < \infty$. When $x < 0$, the value will be 0.

When we write a distribution function, we have to specify the entire range. So, what is the value? $F_X(x)$ is given by the following:

- For $x < 0$, $F_X(x) = 0$
- For $x \geq 0$ and $x < \infty$, $F_X(x) = 1 - e^{-\alpha x}$

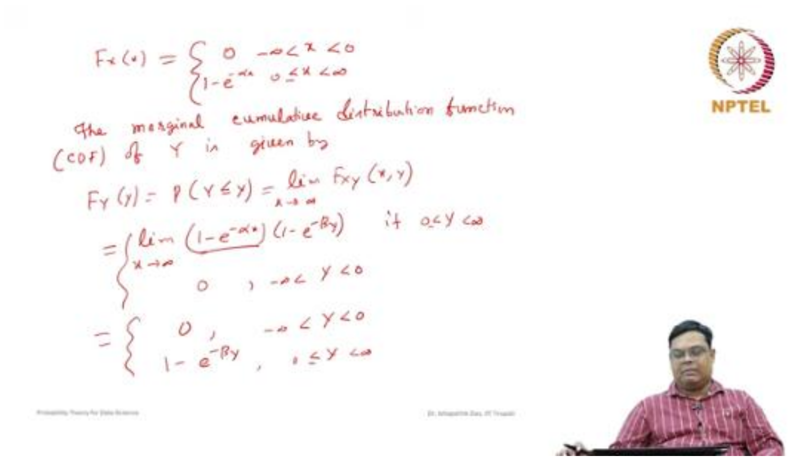
Similarly, you can find the marginal distribution function of Y. Let us compute that.



The marginal cumulative distribution function (CDF) of Y is given by $F_Y(y)$. So, probably that $Y \leq y$. This is equal to the limit as $x \to \infty$ of $F_{XY}(x, y)$. So, this is nothing but the limit as $x \to \infty$. So, basically, $1 - e^{-\alpha x}$ and $1 - e^{-\beta y}$ will be valid if $y \geq 0$ and $y < \infty$.

It will be 0 if $y < 0$ and $y > -\infty$. So, this will be the limit, which can be written as 0 first because this range is from the left-hand side. Whenever it is greater than 0, we will take this function. When $x \to \infty$, this goes to 1 because $e^{-\alpha x}$ is a positive number, and as x increases, it approaches 0. So, this goes to 1, and it will be $1 - e^{-\beta y}$ whenever $y \geq 0$ and $y < \infty$.

So, this is the marginal cumulative distribution function of this joint distribution. Now, for X and Y to be independent random variables, how can we show that? The definition is that two random variables are independent if you can show that $F_{XY}(x, y) = F_X(x) * F_Y(y)$ for all x and $y \in \mathbb{R}$. Since both x and y are positive, $F_{XY}(x, y) = 1 - e^{-\alpha x} * (1 - e^{-\beta y})$, which is nothing but $F_X(x) * F_Y(y)$.



Now, for any other cases, the left-hand side will be 0. If, for example, $x < 0$ and $x > \infty$, or if $y < 0$ and $y > -\infty$, the left-hand side will also be 0. $F_{XY}(x, y)$ will be 0, as we can see. Whenever $x < 0$, $F_X(x)$ will be 0, and if $y < 0$, $F_Y(y)$ will be 0. Therefore, for any other cases, both will be 0.

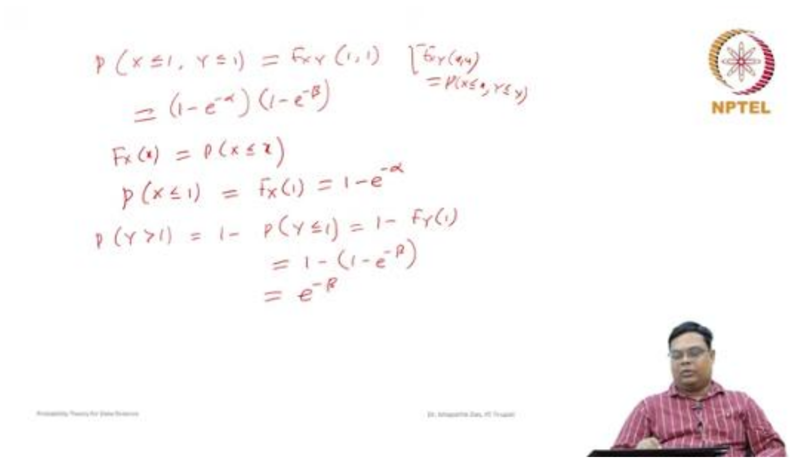
Hence, $F_{XY}(x, y) = F_X(x) * F_Y(y)$ for all $x, y \in \mathbb{R}$. This implies that X and Y are independent random variables. Hence, we can see that X and Y are independent random variables. Now, let's move on to the next question. The next question is to find the probability that $X \leq 1$ and $Y \leq 1$.

So, let's find the probability of $X \leq 1$ and $Y \leq 1$. So, by definition, this is nothing but $F_{XY}(1, 1)$. According to the definition, this represents the probability that $X \leq x$ and $Y \leq y$. When x and y are both 1, it represents the probability that $X \leq 1$ and $Y \leq 1$. We know the function is $e^{(-\alpha * 1)}$ and $e^{(-\beta * 1)}$, which equals 1.

So, this is straightforward. Now, the next question is to find the probability that $X \leq 1$. This is the same because we know that $F_X(x)$ represents the marginal cumulative distribution function. Therefore, the probability that $X \leq 1$ is nothing but $F_X(1)$. So, $F_X(1)$ is the cumulative distribution function of X, which is $1 - e^{(-\alpha * x)}$, for $x > 0$.

So, when $x = 1$, this becomes $1 - e^{(-\alpha)}$. The next problem is also straightforward: the probability that $X \leq 1$ is $F_X(1, 1)$, which equals $1 - e^{(-\alpha)}$. Next, we need to find the probability that $Y > 1$. Similarly, the probability that $Y > 1$ can be found by taking the complement, which is the probability that $Y \leq 1$. We know that this is the cumulative distribution function of Y evaluated at 1.

This is equal to $1 - e^{(-\beta * 1)}$. When you substitute the value for Y, the result simplifies to the exponential term with the negative β. So, that is also straightforward; we just need to take the complement. The next problem is to find the probability that $X > x$ and $Y > y$, where $x, y \in \mathbb{R}$. Note that this is not simply the complement of the probability that $X \leq x$ and $Y \leq y$.



Basically, what we want to say is that "$Y > y$" is not the complement of "$X \leq x, Y \leq y$." We can't just say that the probability is 1 - this. We'll see that this approach may not be true. So, it might seem like a straightforward solution to take the complement, but it may not always be correct. So, if you want to find the complement, how can we find the probability of $X > x$ and $Y > y$?

It is simply $1 - P(X \leq x, Y \leq y)$. First, we write down the event, which is the probability that $X > x$ complement, and $Y > y$. So, for all $x, y \in \mathbb{R}$, $P(X > x, Y > y) = 1 - (P(X \leq x, Y \leq y))$.

Now, if you want to find the probability, we know it in terms of $F_{XY}(x, y)$, which is the probability that $X \leq x$ and $Y \leq y$.

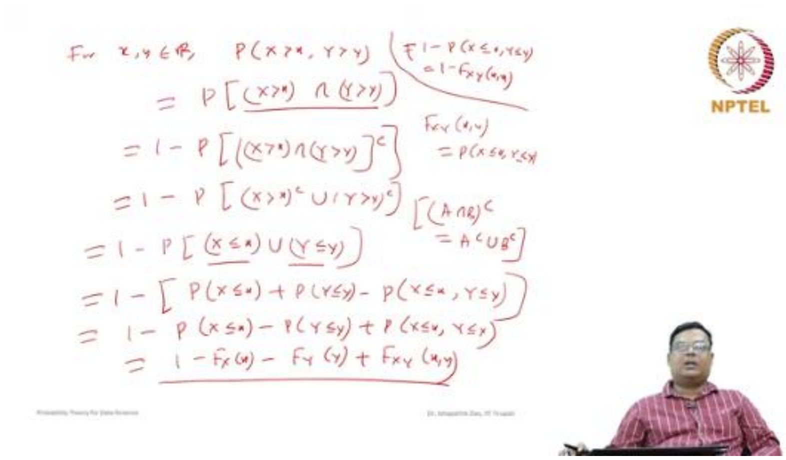
So, it is better to try to find the complement of this, but it's not just a direct complement. We also need to use some other results. So, this is nothing but $1 - P(X > x, Y > y)$, the whole complement of the event. Now, this is A, and this is B. $A \cap B$ complement can be written as A complement $\cup$ B complement.

This is a theorem you know. Sorry, yes, this is the result we are using when we take the whole complement. $A \cap B$ complement is nothing but A complement $\cup$ B complement. We are using this result. This is simply $1 - P(X > x$ complement$) = P(X \leq x \cup Y \leq y)$. This is $A \cup B$. Then, $1 - A \cup B$ formula. We know that this is $P(X \leq x) + P(Y \leq y) - P(A) + P(B) - P(A \cap B)$, $P(X \leq x \cap Y \leq y)$. So, finally, what we get is:

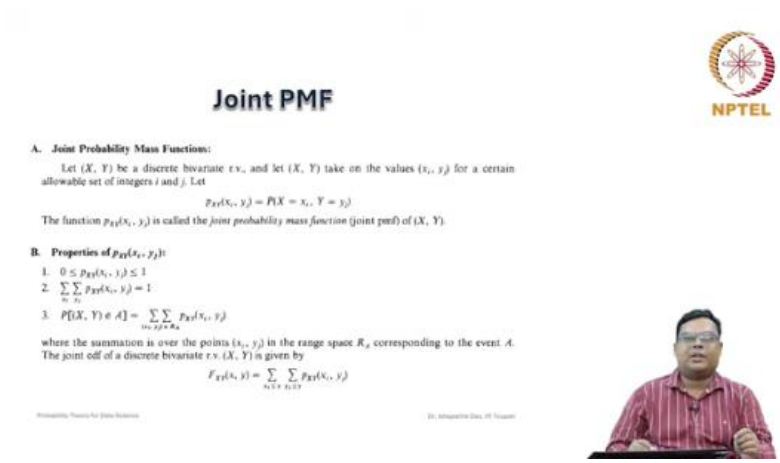
$1 - P(X \leq x) - P(Y \leq y) + P(X \leq x, Y \leq y)$. This is $1 - F_X(x) - 1 - F_Y(y) + F_{XY}(x, y)$. So, this is finally what we got. Basically, initially, we just wanted to say that this is not simply equal to $1 - P(X \leq x, Y \leq y)$, like in the univariate case, where it's straightforward like that, $1 - F_{XY}(x, y)$. But that is not correct. So, it doesn't come out like this. As you can see, it is not like that.

So, we have that $P(X > x, Y > y) = 1 - F_X(x) - F_Y(y) + F_{XY}(x, y)$. So, this is what we found for the cumulative distribution function.

Finally, this is a numerical example we discussed: the cumulative distribution function, the marginal distribution function, and how to find out if the variables are independent, and how to check whether they are independent or not. Finally, we discussed how to compute some probabilities using the joint cumulative distribution function and their marginal cumulative distribution functions. Next, we will discuss discrete bivariate random variables and bivariate continuous random variables, including their probability mass functions and probability density functions.



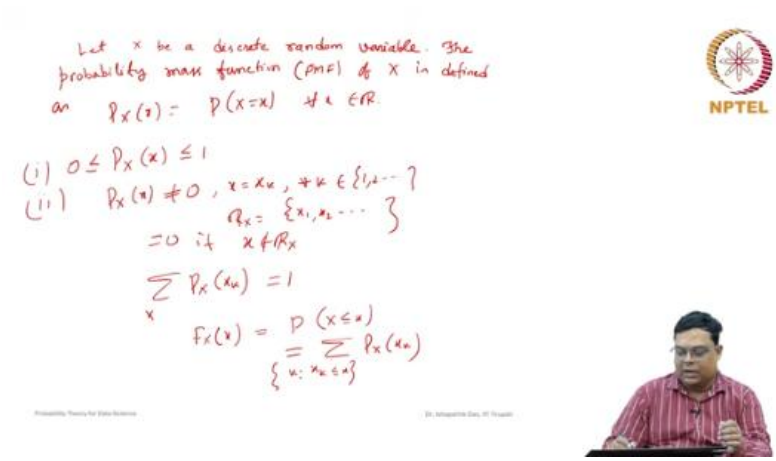
Now, we will discuss the joint probability mass function of two discrete random variables. Whenever X and Y are both discrete random variables, the distribution is called the discrete bivariate distribution function. Similarly, in the univariate case, when X is a discrete random variable, we discuss the probability mass function. The probability mass function gives the probability at a point where X = xi. You can remember that whenever X is a discrete random variable, we define the probability mass function (PMF) of X as follows:



The probability that $X = x$ for all $x \in \mathbb{R}$. Now we can define it for all x, but it will be 0. So, $P(X) \le 1$ and $P(X) \ge 0$. Some of the properties are that $P(X = x) \neq 0$ whenever $x = x_k$ for some $k \in \{1, 2, ...\}$. So, basically, x_k is the range of X, like x_1, x_2, and so on.

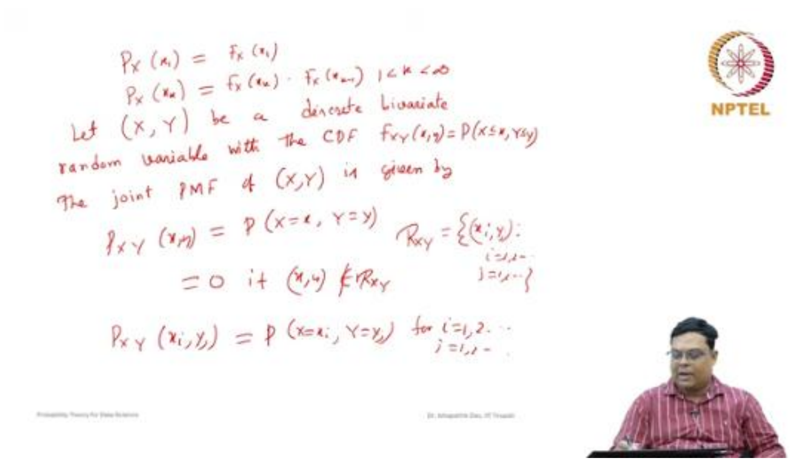
So, for some k, if $x = x_k$, otherwise, the probability is 0 if $x \notin \{x_k\}$ or if x does not belong to the range of X. If $x \notin \text{range}(X)$, the probability is 0. Now, one of the properties to remember is that the sum of the probabilities of $X = x_k$ for all possible values of k is equal to 1. To find the cumulative distribution function of X from the probability mass function, you sum the probabilities for all k such that $x_k \leq x$. This will be 0 when $x < x_1$ (the minimum value in the range of X, for example, x_1, assuming the values are written in ascending order).

So, x_1 will be the minimum. If $x < x_1$, then the cumulative distribution function, F(x), will be 0. So, we have already discussed all these concepts. If you know the probability mass function, you can find the cumulative distribution function. Similarly, if you know the cumulative distribution function, you can also find the probability mass function. So, $P_x(x_1) = F_x(x_1)$, and for any other value P_{xk}, it is $F_x(x_k) - F_x(x_{k-1})$, where $k > 1$ and $k \in \mathbb{N}$.



So, k takes values like 1, 2, 3, 4, and so on. So, that is why if you have the cumulative distribution function, you can also find the probability mass function. Now, for the joint distribution of joint random variables, let X and Y be a bivariate discrete random variable. So, basically, both are discrete bivariate random variables with the cumulative distribution function $F_{x\gamma}(x, y)$. This is equal to $P(X \leq x, Y \leq y)$.

The joint probability mass function of (X, Y) is given by $P_{x\gamma}(x, y)$, which is the probability that $X = x$ and $Y = y$. Now, this will be 0 if, for example, we consider the range of X, where $R_{x\gamma}$ is the range. So, this basically contains all the numbers x, which we can write as x_i, and y, which we can write as y_j, where i = 1, 2, ... and j = 1, 2, ... This is equal to 0 if $(x, y) \notin R_{x\gamma}$, and it is non-zero otherwise. So, that is why you can write $P_{x\gamma}(x_i, y_j)$, which is nothing but the probability that $X = x_i$ and $Y = y_j$, where i = 1, 2 and j = 1, 2.



For any other values of x_i and y_j, this will be 0. So, that is the probability mass function. It is defined as $P_{x\gamma}(x_i, y_j)$, which equals the probability that $X = x_i$ and $Y = y_j$. The function $P_{x\gamma}(x, y)$ is called the joint probability mass function. So, the joint probability mass function has already been mentioned.

Now, let's discuss the properties of this joint probability mass function. First of all, the probability $P_{x\gamma}(x_i, y_j)$ represents the probability that $X = x_i$ and $Y = y_j$. Note that this comma represents the probability of the event $X = x_i$ and $Y = y_j$. We have discussed this several times. This is the event where $X = x_i$. $X = x_i$ is an event, and $Y = y_j$ is also an event.

So, $X = x_i$ corresponds to a set $s \in S$ that belongs to the sample space because X is a random variable, which means it is a measurable function from the sample space S to $\mathbb{R}$. So, all s that satisfy $X(s) = x_i$ and $Y(s) = y_j$ are elements of the sample space S such that $Y(s) = y_j$. These are the events. If you take the intersection of these events, it is the probability that $X = x_i$ and $Y = y_j$. So, this is the joint probability mass function.

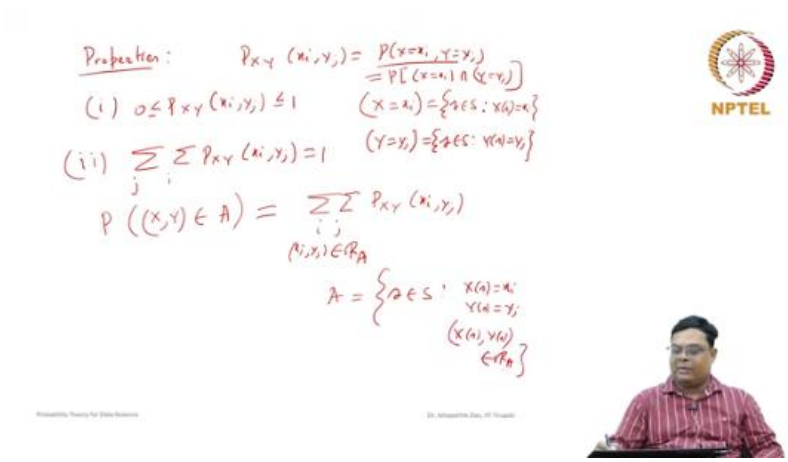
Since this is a probability, by axiom 1, it will always be ≤ 1 and ≥ 0. This is one of the properties mentioned here. Similar to the univariate case, we also discuss the bivariate. The sum of all the possible values will be equal to 1. So, the sum of all i and j, the probability of X and Y at (x_i, y_j), should be equal to 1.

So, this is a certain event because you are considering all possible values that the random variable X and Y can take. The sum of all these probabilities should be equal to 1. This is similar to finding the probability that X and Y, the random variables, belong to A, where

A is a subset of the range (x_i, y_j). So, this is nothing but the summation over all i and j, where x_i and y_j belong to this range. So, this is nothing but all the possible values of x_i and y_j such that they are inside A.

This can be written as $P_{x\gamma}(x_i, y_j)$, where x_i and y_j belong to the range of A. So, it is written that x_i and y_j should belong to the range of A. Basically, here, A is a subset of X. What is A? A is defined as the set of all elements in S such that $X(x_i)$ belongs to S.

In other words, for all $s \in S$, X(s) belongs to the range of A, and Y(s) belongs to the range of Y. In a different way, we can write, sorry. So, this is nothing but when $X(s) = x_i$ and $Y(s) = y_j$, and then the pair (X(s), Y(s)) should belong to the range of A. The range of A is a subset of the sample space. Now, I hope you have understood that.



Next, we will discuss how to find the cumulative distribution function if you know the joint probability mass function.