PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 07

Lecture - 33

Independence Between Two Random Variables

Now, you can also prove it graphically. Suppose this is x_1 and this is x_2. These points may not all be positive; we are just representing two points. Now, it is asking that only y is considered, not two values. So, let us consider just one value for y.

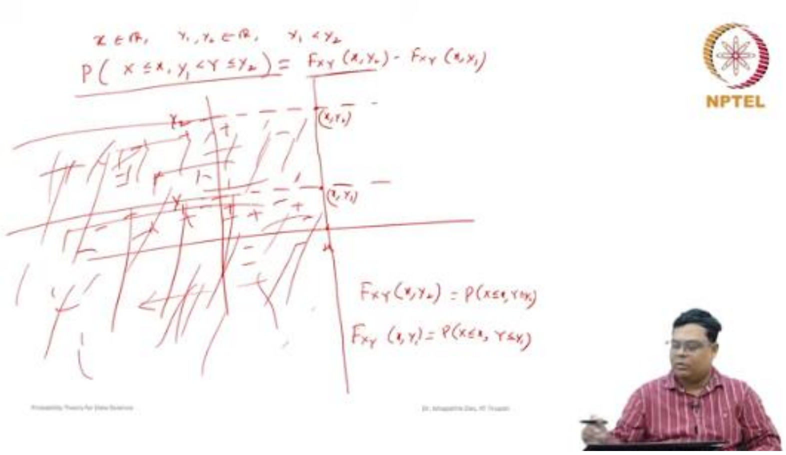
Now, what is $X \leq x$? Sorry, okay, there are two points: y_1 and y_2. We are proving, I actually thought the last result we already proved. So, there are two points, y_1 and y_2, and one value of x. Now, it is asking for the probability that $X \leq x$.

$X \leq x$. That means this area, the probability of this part, and y will be between y_2 and y_1. So, this is y_2 and this is y_1. Now, this value is nothing but (x, y_2), and this is (x, y_1), this point. Now, this probability is the probability of this region, so $X \leq x$, and $Y \in [y_1, y_2]$.

$Y \leq y_2$ in this region. So, how will we get this region? This is represented by $F_{x\gamma}(x, y_2)$. This is nothing but the probability that $X \leq x$, and $Y \leq y_2$. So, $X \leq x$ is this part.

Suppose I'm writing this part: $X \leq x$, and $Y \leq y_2$, so it's the whole part. Now, this is the probability of (x, y_2), or $F_{x\gamma}(x, y_2)$. Now, what is $F_{x\gamma}(x, y_1)$? This is nothing but the probability that $X \leq x$, and $Y \leq y_1$. So, $X \leq x$ is this part, and $Y \leq y_1$ is this part.

So, this is what I am writing, this minus... Now, this is plus, actually all plus. So, basically, if you subtract that part from the whole region, then we are getting this part. That is the probability that $X \leq x$, and $Y \in [y_1, y_2]$. We have proved this by set theory, and we have also presented it graphically.



$x \in \mathbb{R}, \quad y_1, y_2 \in \mathbb{R}, \quad y_1 < y_2$
$P(X \le x, y_1 < Y \le y_2) = F_{XY}(x, y_2) - F_{XY}(x, y_1)$
y_2
y_1
(x, y_2)
(x, y_1)
$F_{XY}(x, y_2) = P(X \le x, Y \le y_2)$
$F_{XY}(x, y_1) = P(X \le x, Y \le y_1)$
NPTEL
Probability Theory for Data Science
Dr. Ishapathik Das, IIT Tirupati

So, this is one of the important properties of the cumulative distribution function. This is very similar to when we discussed the univariate random variable. There are only two points, and the probability that X is between x_2 and x_1 is nothing but $F_x(x_2)$ - $F_x(x_1)$. However, here, we are fixing one coordinate and taking the other coordinate in an interval, and then we are finding the relationship. Until now, it has just been an extension.

The additional property is property 7, which is very important. Any function that satisfies all the properties but does not satisfy this property 7 cannot be considered a cumulative distribution function or a joint cumulative distribution function of a bivariate random variable. So, what is this property? It applies when you consider two points. Here, let us consider two points, x_1 and x_2, which belong to $\mathbb{R}$, such that $x_1 < x_2$, and two other points, y_1 and y_2, which also belong to $\mathbb{R}$, such that $y_1 < y_2$.

Then, what is the probability that $X \le x_2$, $X > x_1$, and $Y \le y_2$, $Y > y_1$? What will this probability be? So, this probability can be represented as $F_{x\gamma}(x_2, y_2) - F_{x\gamma}(x_1, y_2) - F_{x\gamma}(x_2, y_1) + F_{x\gamma}(x_1, y_1)$. So, this is the probability. Now, how can we do this?

Let us first represent it graphically, as it will be easier to understand. This is x_1, x_2, and this is y_1, y_2. So, we are talking about the following: This is $Y = y_1$, $Y = y_2$, and this is $X = x_1$, and this is $X = x_2$. So, this point is (x_2, y_2), this point is (x_2, y_1), this is (x_1, y_1), and this is (x_1, y_2).

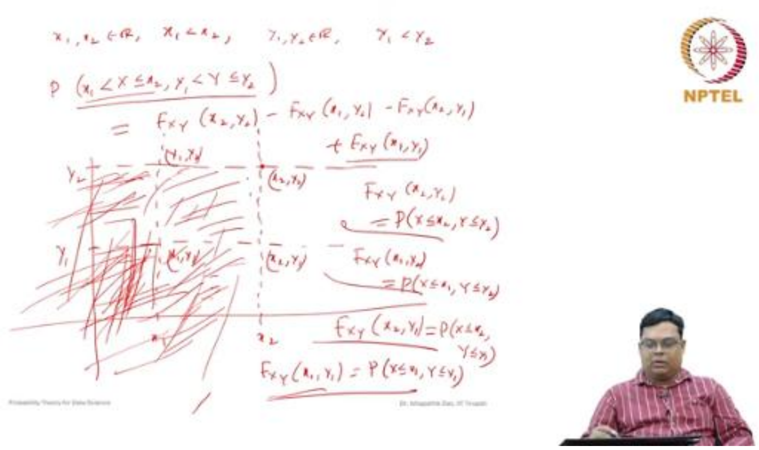
The probability that $x_1 \le X$, and $y_1 \le Y \le y_2$ will be the probability within this region, between x_1 and x_2, and y_1 and y_2. How do we find this region? First, consider $F_{x\gamma}(x_2, y_2)$. By definition, this represents the probability that $X \le x_2$, and $Y \le y_2$. So, the probability that $X \le x_2$ and $Y \le y_2$ represents the probability of the entire region.

However, we don't want the probability of the entire region; we need to subtract some additional values. These values are represented by $F_{x\gamma}(x_1, y_2)$, which is the probability that $X \le x_1$ and $Y \le y_2$. So, $X \le x_1$ represents this region, and $Y \le y_2$ represents this region. We subtract this region because we did not ask to find the probability for it. Additionally, we need to subtract this region as well, which is why we subtract $F_{x\gamma}(x_2, y_1)$.

This represents the probability that $X \le x_2$ and $Y \le y_1$. So, $X \le x_2$ means this part. $X \le x_2$ represents this, and $Y \le y_1$ means it is subtracting this part as well. Now, when subtracting this and this part, we have subtracted this part twice. This part is nothing but $F_{x\gamma}(x_1, y_1)$, which represents the probability that $X \le x_1$ and $Y \le y_1$.

That is why we subtracted this part twice. Therefore, we are adding this term, which is why it is $F_{x\gamma}(x_1, y_1)$. So, this part minus this part, and then again this minus this part, plus this part, because we have subtracted it twice. Using set theory, we can also prove this. You can find the set and represent it accordingly.

Then, suppose you first determine the set by considering and fixing y_1 and y_2. After that, you can apply the previous proof twice. By doing so, you can verify and prove it. Now, the important part is what we want to say. So, what we have found in this property is property 7. Property 7 states that the expression $F_{x\gamma}(x_2, y_2) - F_{x\gamma}(x_2, y_1) - F_{x\gamma}(x_1, y_2) + F_{x\gamma}(x_1, y_1)$ represents the probability that $X \in [x_1, x_2]$, and $Y \in [y_1, y_2]$.



$x_1, x_2 \in \mathbb{R}, \quad x_1 < x_2, \quad y_1, y_2 \in \mathbb{R}, \quad y_1 < y_2$

$P(x_1 < X \le x_2, y_1 < Y \le y_2)$

$= F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$

$F_{XY}(x_2, y_2) = P(X \le x_2, Y \le y_2)$

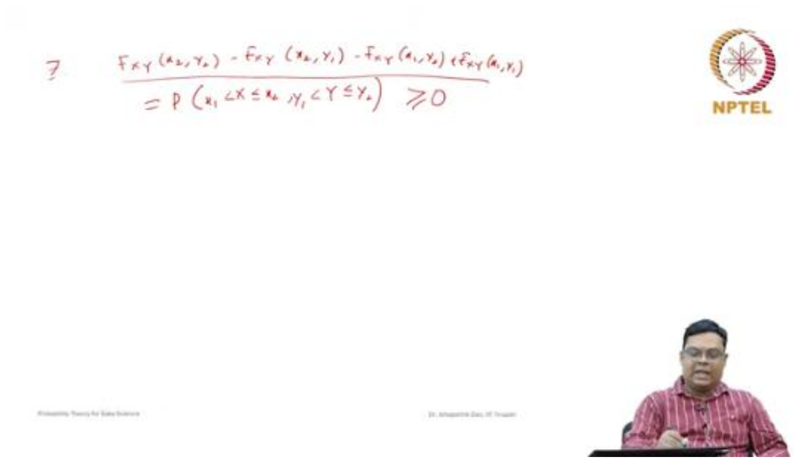
$F_{XY}(x_1, y_2) = P(X \le x_1, Y \le y_2)$

$F_{XY}(x_2, y_1) = P(X \le x_2, Y \le y_1)$

$F_{XY}(x_1, y_1) = P(X \le x_1, Y \le y_1)$

Property 7 states that the given relationship must hold. As per Axiom 1, the probability should always be ≥ 0. Therefore, any cumulative distribution function or bivariate function that satisfies properties up to 6 might still fail if it does not satisfy the relationship described in Property 7. So, we have to check that if you are considering a bivariate function from ℝ² to ℝ, and you are thinking it may be a cumulative distribution function of a bivariate random variable, it may not actually be one. Even if it satisfies the first six properties, it cannot be called a joint cumulative distribution function of a bivariate random variable unless it satisfies this relationship as well.

So, we will discuss some examples where we encounter cases that appear to be cumulative distribution functions but cannot be considered as such because they do not satisfy these properties. Now, we have already talked about this property, which is called independence. So, what is independence by definition? So, whenever we talk about any event, suppose A and B are two events. Let A and B be two events.



7 $\underline{F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1)}$

$= P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \geqslant 0$

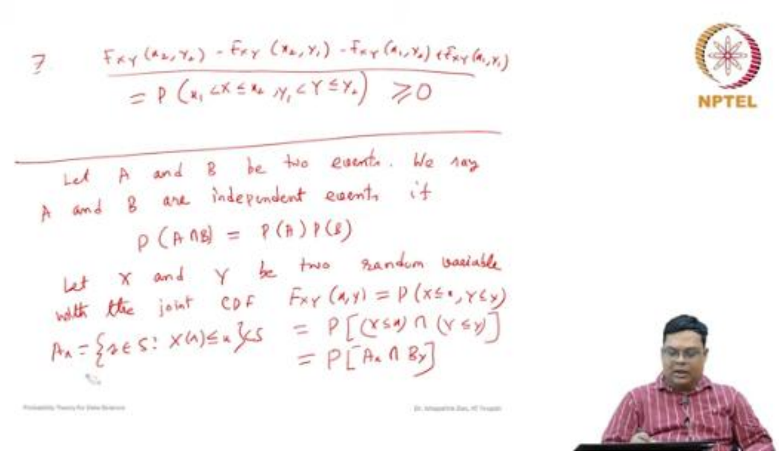
Probability Theory for Data Science

Dr. Ishapathik Das, IIT Tirupati

We say that events A and B are independent if P(A ∩ B) = P(A) * P(B). Now, how do we define two random variables as independent? We have already discussed this, but let us go over it again. Now, how we define two random variables as independent? We already discussed this. Again, we are discussing.

It is because independence for two events is fine. But random variables are functions, measurable functions. We know the joint cumulative distribution function of two random variables. Let X and Y be two random variables with the joint cumulative distribution function Fxy(x, y). This is defined as P(X ≤ x and Y ≤ y). Now, this event X ≤ x and Y ≤ y is nothing but the event Y ≤ y, X ≤ x as an intersection.

So, in our notation, this is nothing but P(Ax ∩ By). Where Ax is the event defined for any real number x as the set of all s ∈ S such that X(s) ≤ x. This is a subset of S. Similarly, By is the event defined as the set of all s ∈ S such that Y(s) ≤ y.



$$? \quad F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1)$$

$$= P(x_1 < X \le x_2, y_1 < Y \le y_2) \geq 0$$

Let A and B be two events. We say A and B are independent events if

$$P(A \cap B) = P(A)P(B)$$

Let X and Y be two random variable with the joint CDF $F_{XY}(x, y) = P(X \le x, Y \le y)$

$$A_x = \{s \in S : X(s) \le x\} \subseteq S \qquad = P[(X \le x) \cap (Y \le y)]$$

$$= P[A_x \cap B_y]$$

Now, Ax ∩ By refers to the event where we will say that the events Ax and By are independent if P(Ax ∩ By) = P(Ax) * P(By). In that case, Ax and By are independent events. Now, Ax ∩ By refers to the event where we will say that the events Ax and By are independent if P(Ax ∩ By) = P(Ax) * P(By). In that case, Ax and By are independent events. That means, what is Ax ∩ By?

Sorry, this is By. Ax ∩ By is nothing but P(Fxy). So, Ax ∩ By is equivalent to X ≤ x and Y ≤ y. So, this implies that this is nothing but P(X ≤ x), and this is nothing but P(Y ≤ y) for all x, y ∈ R, if this happens. So, this implies that we can write this as equivalent to saying that Fxy(x, y) = Fx(x) * Fy(y) for all x, y ∈ R.

If it satisfies this relationship, Fxy(x, y) = Fx(x) * Fy(y) for all x, y ∈ ℝ, then we say that X and Y are independent random variables. Whenever you can express the joint cumulative distribution function as the product of their marginal cumulative distribution functions, then X and Y are said to be independent random variables. This is the definition. Let us discuss some examples. Before that, we need to cover a few concepts because we will be doing some other computations as well.

So, how can we find these marginals? We are talking about the marginal distribution function, Fx(x). Suppose the joint distribution functions are given. Let Fxy(x, y) be P(X ≤

x and Y ≤ y). This is known. Now, we want to find the marginal cumulative distribution function of X and Y.

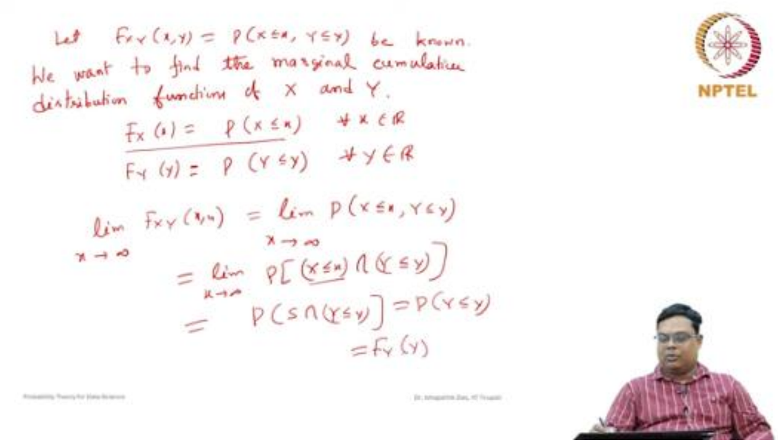
So, that means we want to find Fx(x), which is P(X ≤ x) for all x ∈ ℝ. We also want to find Fy(y), which is P(Y ≤ y) for all y ∈ ℝ. So, suppose the joint cumulative distribution function is known. Now, we want to find how we can compute it. To do this, let's consider the limit as x → ∞ of Fxy(x, y).

What does this represent? This is P(X ≤ x and Y ≤ y) as x → ∞. Now, by definition, this represents P(X ≤ x and Y ≤ y) for this event. Now, whenever x → ∞, as we discussed earlier, notice that y is fixed here. We are only considering x approaching infinity. As x becomes larger and larger, this event is simply P(X ≤ x). This is nothing but all s ∈ S where X(s) ≤ x. As x becomes larger and larger, all s will satisfy this relationship. So, when you take the limit as x → ∞, this set will become S. This is just for intuition and to help understand the concept.

So, we understand that whenever x → ∞, this event actually becomes S. So, this is equal to P(S ∩ Y ≤ y). This is nothing but P(Y ≤ y). So, S ∩ Y ≤ y is nothing but this, and this is equal to Fy(y). So, what is this?

So, this is Fy(y). So, by definition, Fy(y) is the marginal cumulative distribution function of Y. Hence, what we found is this result: whenever you take the limit of one variable, you get this. Similarly, you can show the limit for the other variable.

So, what we found is that as x → ∞, Fxy(x, y) = Fy(y), which we found earlier. Similarly, if you take the limit as y → ∞, fixing the variable x, and let the other variable tend to infinity, you will get the marginal cumulative distribution function of the variable X.



Let $F_{XY}(x,y) = P(X \le x, Y \le y)$ be known.
We want to find the marginal cumulative distribution function of X and Y.

$$F_X(x) = P(X \le x) \quad \forall x \in \mathbb{R}$$

$$F_Y(y) = P(Y \le y) \quad \forall y \in \mathbb{R}$$

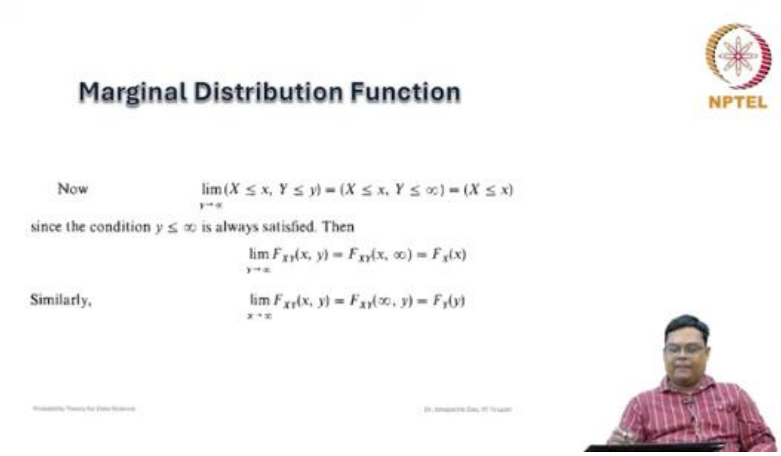
$$\lim_{x \to \infty} F_{XY}(x,y) = \lim_{x \to \infty} P(X \le x, Y \le y)$$

$$= \lim_{x \to \infty} P[\underline{(X \le x)} \cap (Y \le y)]$$

$$= P(S \cap (Y \le y)] = P(Y \le y)$$

$$= F_Y(y)$$

So, that is the method of finding the marginal cumulative distribution function when the joint cumulative distribution function is known to us. So, this is the result. When you take y → ∞, it becomes nothing but X ≤ x, since the condition is always satisfied. Similarly, when y → ∞, Fxy(x, y) becomes Fx(x). If you take x → ∞, you will get Fy(y).



Marginal Distribution Function

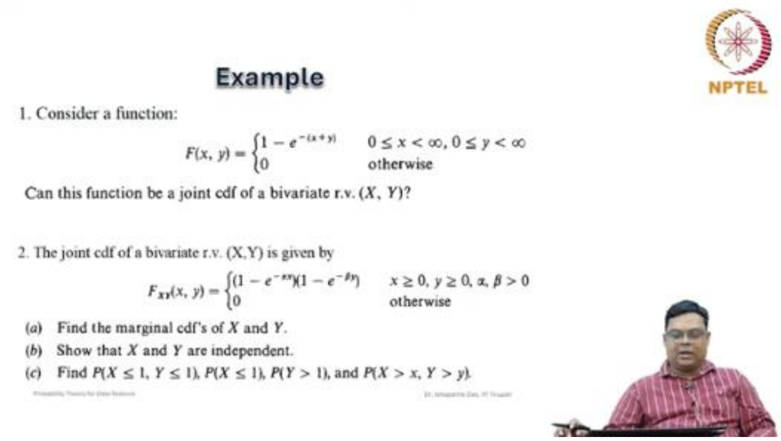
Now $$\lim_{y\to\infty}(X \le x,\ Y \le y) = (X \le x,\ Y \le \infty) = (X \le x)$$

since the condition $y \le \infty$ is always satisfied. Then

$$\lim_{y\to\infty} F_{XY}(x, y) = F_{XY}(x, \infty) = F_X(x)$$

Similarly, $$\lim_{x\to\infty} F_{XY}(x, y) = F_{XY}(\infty, y) = F_Y(y)$$

This is the concept of how to find the marginal distribution function. Let us now discuss some examples. So, consider a function F(x, y), which is equal to 1 - e^(x + y) * e^(xy), whenever x ≥ 0 and x < ∞, and y ≥ 0 and y < ∞, and 0 otherwise. Can this function be the joint cumulative distribution function (CDF) of a bivariate random variable? So, the question is whether this function can be the joint cumulative distribution function of a bivariate random variable.



Example

1. Consider a function:

$$F(x, y) = \begin{cases} 1 - e^{-(x+y)} & 0 \le x < \infty, 0 \le y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Can this function be a joint cdf of a bivariate r.v. (X, Y)?

2. The joint cdf of a bivariate r.v. (X,Y) is given by

$$F_{XY}(x, y) = \begin{cases} (1 - e^{-\alpha x})(1 - e^{-\beta y}) & x \ge 0, y \ge 0, \alpha, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal cdf's of X and Y.
(b) Show that X and Y are independent.
(c) Find $P(X \le 1, Y \le 1)$, $P(X \le 1)$, $P(Y > 1)$, and $P(X > x, Y > y)$.

Let's discuss this example. The function is given, but we don't know if it will be a cumulative distribution function. Let F be defined as F(x, y), which is equal to 1 - e^(x + y), whenever x ≥ 0 and x < ∞, and y ≥ 0 and y < ∞. It is 0 otherwise. So, this is the example. F(x, y) = 1 - e^(- (x + y)), where x ≥ 0, y ≥ 0, and 0 otherwise. So, the question is whether this can be a joint cumulative distribution function of a bivariate random variable. To check this, we need to look at the properties. So, what are the properties? You can see that.

So, it is always between 0 and 1. You have to check all those properties. Let’s look at this. When x ∈ (0, ∞), and y ∈ (0, ∞), this is equal to e^(- (x + y)). Since e^(- (x + y)) is always less than or equal to 1, this will always be between 0 and 1. So, as x → ∞ and y → ∞, e^(- (x + y)) approaches 0. Therefore, F(x, y) approaches 0. Similarly, when x → -∞, this becomes 0 by definition. You can observe that properties 1 to 6 are satisfied. The only issue lies with the 7th property. So, you can remember the first six properties, as these are the key properties to check. It is right-continuous, and those properties will also be satisfied.

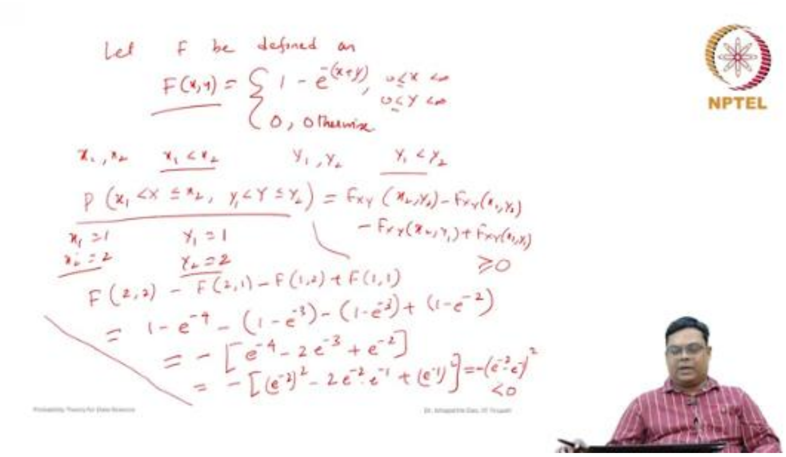
The only property we need to check is the 7th, as we discussed earlier, because the 7th property is very important. The 7th property states that if you consider two real numbers, x1 < x2, and two real numbers, y1 < y2, So, to find the probability, we need to determine the probability of X ≤ x2, X > x1, and Y ≤ y2, Y > y1. This probability is simply given by Fxy, which we can write as Fxy(x2, y2). If it is a cumulative distribution function, then we can say that this property holds: the probability for the region with coordinates (x2, y2) minus the probability for the region with coordinates (x2, y1), minus the probability for the region with coordinates (x1, y2), plus the probability for the region with coordinates (x1, y1). So, we can write that this is always ≥ 0.

We need to check that for any values of x1, x2, y1, y2, this function satisfies this relationship, meaning it must always be ≥ 0. You can check this with any numbers, or generally, you can verify this relationship to determine whether it is ≥ 0. If it cannot be proven that it is always ≥ 0, then we need to find an example where it is not true, because it should hold for any x1 and x2. Let's consider any values for x1 and x2. Suppose y1 = 1. Let x1 = 1 and x2 = 2. We can also take y1 = 1 and y2 = 2. This is just an example we can use.

So, we need to prove that for this case, F(x2, y2) = F(2, 2) - F(2, 1) - F(1, 2) + F(1, 1). Now, what will the values be? So, F(2, 2) = 1 - e^(- (2 + 2)), which simplifies to 1 - e^(- 4). F(2, 1) = 1 - e^(- (2 + 1)) = 1 - e^(- 3). F(1, 2) = 1 - e^(- (1 + 2)) = 1 - e^(- 3). F(1, 1) = 1 - e^(- (1 + 1)) = 1 - e^(- 2).

Now, substituting these values, we get: F(2, 2) - F(2, 1) - F(1, 2) + F(1, 1) = (1 - e^(- 4)) - (1 - e^(- 3)) - (1 - e^(- 3)) + (1 - e^(- 2)) = - e^(- 4) - 2e^(- 3) + e^(- 2). This can be represented as e^(- 2) * e^(- 2) - 2e^(- 2) * e^(- 1) + e^(- 1) * e^(- 1). This is nothing but the negative of e^(- 2) * e^(- 1) squared. So, notice that this is non-zero. Not only that, but it is the square of any real number, whether positive or negative, and a square is always positive. Therefore, this is a positive number, and when we take the negative of this positive number, it will always be less than 0.

For this, x1 = 1, and x2 can be any other value. I just used some values to illustrate the point. When we check, we find that this is < 0. It may also be true for other values of (x1, x2) and (y1, y2), which can be shown through similar algebra. This means that property 7 is not satisfied. Since it is less than 0, the probability would be less than 0, which is not possible. Probability must always be between 0 and 1. Therefore, it must be between 0 and 1. That is why it will not be a cumulative distribution function of a bivariate random variable. So, that is one example we discussed.



Let F be defined on

$$\underline{F(x,y)} = \begin{cases} 1 - e^{-(x+y)}, & 0 \le x < \infty \\ & 0 \le y < \infty \\ 0, & \text{otherwise} \end{cases}$$

x_1, x_2 $\quad \underline{x_1 < x_2}$ $\quad y_1, y_2$ $\quad \underline{y_1 < y_2}$

$$\underline{P(x_1 < X \le x_2,\ y_1 < Y \le y_2)} = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1) \geq 0$$

$x_1 = 1$ $\quad y_1 = 1$

$\underline{x_2 = 2}$ $\quad \underline{y_2 = 2}$

$$F(2,2) - F(2,1) - F(1,2) + F(1,1)$$
$$= 1 - e^{-4} - (1 - e^{-3}) - (1 - e^{-3}) + (1 - e^{-2})$$
$$= -\left[e^{-4} - 2e^{-3} + e^{-2}\right]$$
$$= -\left[(e^{-2})^2 - 2e^{-2} \cdot e^{-1} + (e^{-1})^2\right] = -(e^{-2} - e^{-1})^2 < 0$$

Probability Theory for Data Science

Dr. Ishapathik Das, IIT Tirupati

Let us discuss another. I hope you are following and you understood it. The question was whether this function can be a joint cumulative distribution function (CDF) of a bivariate random variable. What we found is that it cannot be, because property 7 is not satisfied.