Introduction to Stochastic Processes
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 12
Jointly Distributed Random Variables and Conditional Distributions

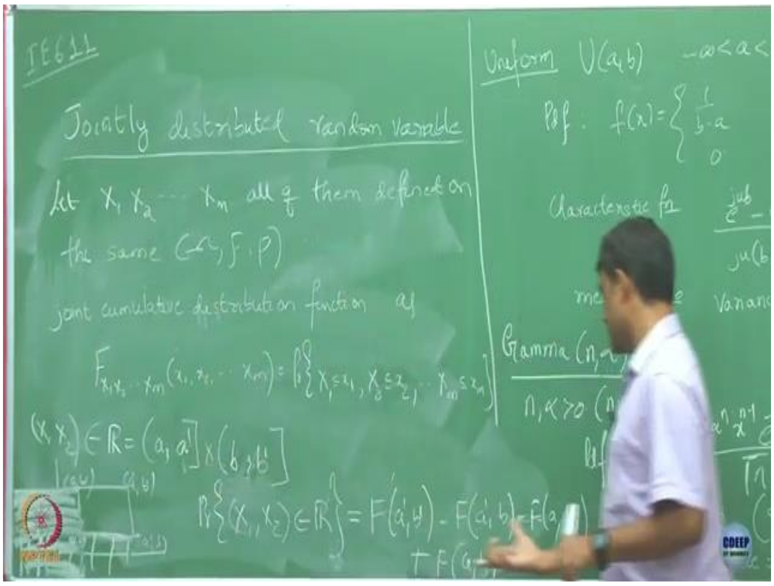
Next we are going to talk about how to handle a set of random variable. So far we have told random variable x. What we just defined it CDFs pdf, if it is continuous and then talk to how to compute its expectation, variance and all this thing, but often it is not that one variable you have to deal with.

You have to deal with multiple variables, multiple random variables. A simple thing we already discussed this, if you throw a 2 dice? First one outcome you want to represent by x 1 and the second one you want to represent by x 2, you have already 2 random variables here and it may so happen that one of the dice is biased, it is not necessary that it is going to take all the outcome uniformly likely.

So, maybe with one it with one it with probability 0.9 it may take 1 and in rest of the outcomes it will take only with 0.1 your other dice could be uniform, it can take all the possible outcome equally likely. So, then in this case, if you want to see the outcome, the joint outcome what came when you throw both of them simultaneously? You have to jointly characterize what is the outcome or if you think that it is enough like I characterize the distribution of x 1 separately and I will characterize the outcome x 2 separately.

Then when I have to look at the outcome jointly, if I just separately characterize them that will tell me how the joint one look like. It is not necessary like the joint one need not look necessary if I just know the individual one. So, let us, we will make this more clear. So, for that, in any experiments, where we have multiple random variable, we need to define the joint distributions.

(Refer Slide Time: 02:36)



And now we are going to study that jointly distributed. So let I have 10 random variable which I am going to denote like this. All of them defined on the same probability space. Then we are going to define joint cumulate distribution function as, so now I am not looking at 1 random variable, but I am looking at now m random variable and the outcome of them, maybe like joint, they are dependent on each other. So that is why I need to queue, a joint distribution here and how I am defining it.

The CDF the Cumulative Density Function of that Cumulative Distribution Function of that at point x 1, x 2, x m, I am going to take it as probability that x 1 is less than x 2, x 2 is less than or x 2, and all the we have to xm less than or equal to xm. So this is what I call my joint, CDF.

And just to visualize this, suppose I have a simple case of 2 random variables, let us say x 1 and x 2. And I want to ask the question, what is the probability that these values take in some region R and that region R could be let us say, I am in a 2 dimensional space, that region R could be simply a rectangle. For example, let us say this region R could be some rectangle here. In which case, this I can write it as, let us say, I am not including this boundary here.

Let us say a, let us call this a b. Let us call this a prime b. And let us call this a b prime. And this is going to be what? a prime b prime. So let us say this is going to be a into b then 2 a prime into coma b prime. So this interval and then this interval, so you understand this. I have x 1 and x 2 both, let us say both of them taking real numbers. Now I want to ask the question, this is entire 2D dimensional, now I may be asking the question, what is the probability that they are going to take value in some particular region R and that region is here rectangle.

Student: (())(07:09).

Professor: How do I represent this, this is going to be a a, a b is what? The way I have written it, this is going to be what? So this is this interval is going to be a a prime and this vertical is going to be b b prime. So now you want to ask the question, what is the probability that this kind of variables x 1, so let us say my x axis represents x 1 and my y axis represents x 2. What is the probability that they will fall in this region.

Student: (())(08:04).

Professor: So how you are going to find that? So, one possibility I can think of, so if I want to now find, let us say, probability that x 1, x 2 belong to R, what I can do, I can look for, I can look for a probability that I takes value below this and then I can look for value that is taken below this region.

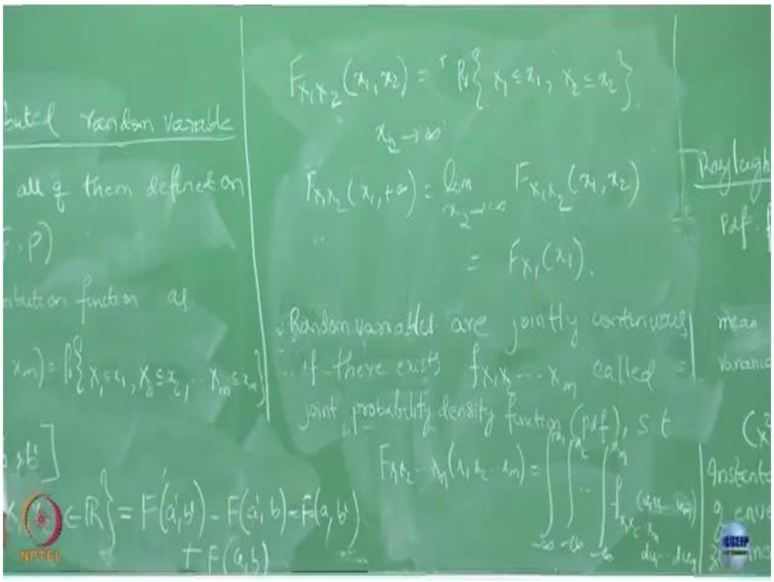
So when I do this, I am subtracting this region, price? Maybe I have to add it back. So let us write this. So, if I do this, is this correct if I do this, this is going to be x, x 1, x 2 and what is this? This is going to be the region. I want to go all the way up to a prime and this is going to be b. And so first I will take the entire thing? Everything below this. So this is going to be a prime, b prime minus F. I am going to all this below all the way is going to a and b and then F all the way up to that first (())(10:09) is going to be a and the height is going to be b prime.

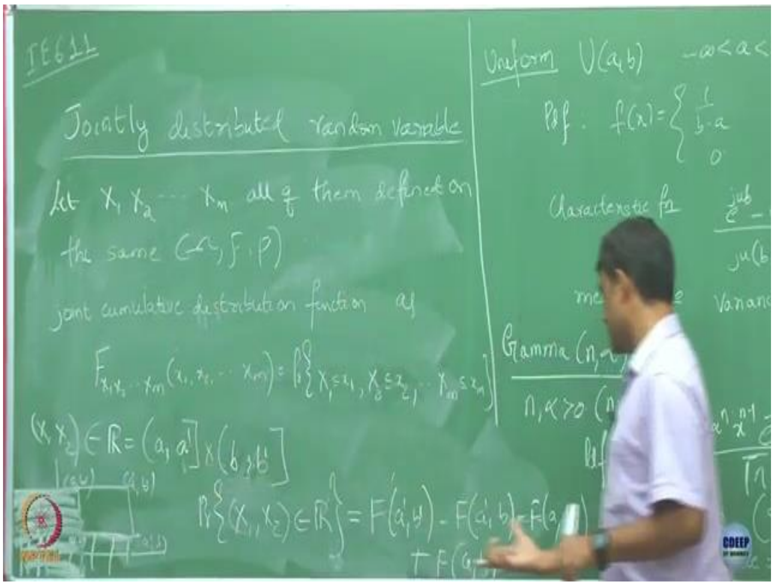
Student: Plus F of ab.

Professor: F of ab, so if I do this, I am only going to retain this area and now this is going to characterize that probability. So, the point here is when I have this set of random variable, I want to ask the question jointly, what is the probability that this belongs to this area? I have to further, so to express that I need to know the joint CDF also?

I mean this is like I have taken a special case just for visualization. R is like a rectangle, if you can take any shape and in that case to find this you need to have this CDF and that is why to understand how this set of random variables are characterized the set of random variable I need to define this joint CDFs.

(Refer Slide Time: 11.32)





Now another thing is to, so if I just go and write it as F of x 1 just, let us say for 2 random variables I have this and I am going to write it as F of x 1. This is going to be what, probability the way I have defined it, this is going to be x 1, x 2. Suppose I let x 2 go to infinity. So in this case, this second part let x 2 less than or equals to infinity that is always holds, like x 1 x 2 has to. So, in that means this probability that x 2 spanning all its value is already covered, then this simply turns out be what is the probability that x 1 is less than or equals to x.

So, that is why we say that F of x 1, x 2, x 1 plus infinity that is written as x 2 goes to infinity (())(12:44) x 1 x 2 we are simply going to denote it as F of x 1 of x 1. So, by letting one of the variables go to infinity we have recovered what?

Student: F of x 1.

Professor: F of x 1 in this case the marginal CDF from the joint this was a simple case with 2 random variables from the joint we have recovered the marginal CDF of x 1. So, similarly you can plug in for any given x 2, they can let x 1 go to infinity. In that case what you will recover?

Student: (())(13:22)

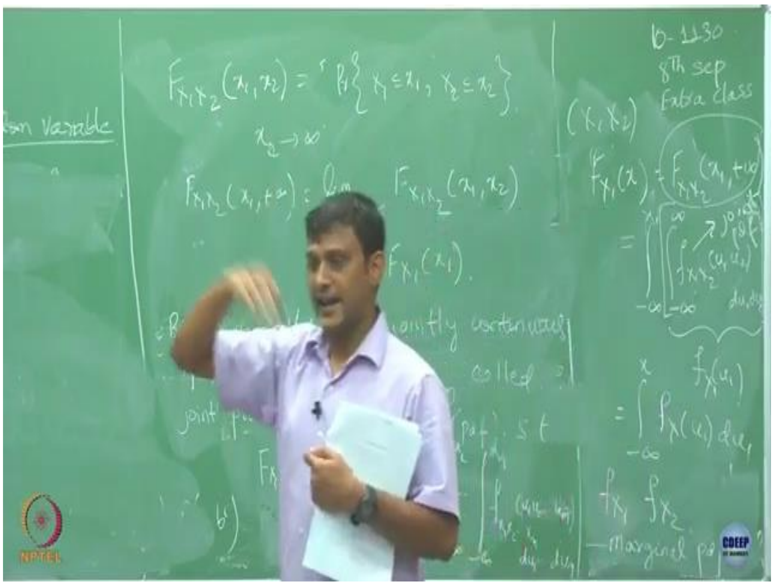
Professor: You are going to recover F of x 2 by letting x 1 go to infinity. We are going to say random variables are so this definition I have said are all on the same space and the joint cumulative distribution function is this.

This I do not, did not specify anything like these are discrete or continuous random variable. Now, I am going to say that here I had only say that they are jointly distributed. Now I am going to say that random variables are jointly continuous if there exist function which I am going to call (())(14:32) called and now I am going to call this as called jointly, joint probability and the function here I do not need to say this is such that if you can express this F of x 1, F of x 2, all the way up to x of n, x 1, x 2, x of n as an integral of and this is like a not like integration over one variable but integration over multiple variables here.

This is going to be, so this is just like a generalization of what we mean by continuous random variable to jointly continuous random variables for a continuous random variable how we have defined, if we said that x is a continuous random variable, if its associated CDF can be expressed as integral of some function F.

So that is our so, we said that if there exists a function F such that F of x of x is equal to integration of minus infinity to x, F of x u, d u then that x is continuous, but we are just now extending that notion to m random variables, if I have such a multivariate function here. I am calling it multivariate because it is now taking m variable instead of one variable. And if I can express the CDF, which is the joint F, in this fashion, then I am going to call the set of random variables are jointly continuous and from the joint continuous PDF.

(Refer Slide Time: 17:20)



So, from this joint PDF, we can again go and recover the marginal PDF as we did earlier. So what we can do is suppose now, we know that F of x 1 so let us say for the simple case of 2 random variables x 1 and x 2 how did we represent this in terms of the joint distributions? We had said of x 1 x 2 x plus infinity, now plugging support let us say x 1 and x 2 are jointly continuous, suppose they are then according to my definition there exists F of x 1 x 2 which will satisfy this condition. So, then I have minus infinity to x 1 minus infinity to plus infinity and what is this F of x 1, x 2 of u 1, u 2 which is and d 1.

Now if you look at to this integral here. Now, I am integrating this function over all u 2, is this whatever this is in the square bracket, is this a function of u 2 anymore? No, because I integrate it over all possible. So let us call this as, now it is only a function of what u 1 now, inside this and so then this integration simply becomes minus infinity to f of x, u 1 of d u. So, we have now recovered basically if it is a jointly continuous we have also said that for x 1 also from that F x 1, x 1 jointly PDF now we have recovered a marginal PDF.

So, now suppose x 1 x 2 are jointly continuous by that I know that by this definition there exists such a function F of x 1 x 2 which satisfies the solution and now from that what I have derived for F of x 1, x. I have this function, what is this implies that, is this implies that my random variable x n is continuous random variable.

So, jointly continuous I have we are now recovering the our earlier definition of Continuous Random Variable for a single random variable and in this case this F of x, u 1 this is going to be called as marginal distribution of x 1. So, this marginal PDF of x 1 and this one was what this was joint PDF and this is like a marginal PDF. And so similarly, so this is now f of x 1 so, similarly, you can recover f of x 2 also. So both you can recover from f of x 1, x 2 and these are called marginal PDFs and this one is called joint PDF, so this is for what?

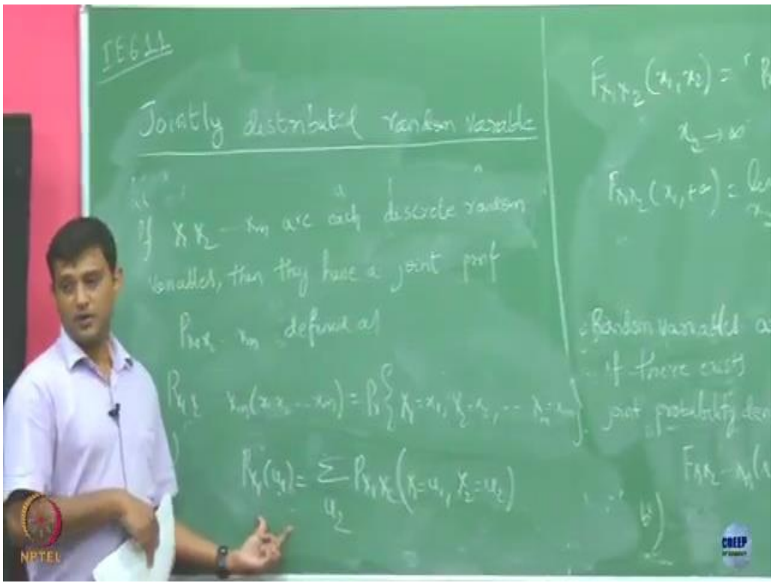
Student: (())(21:01).

Professor: I will say we wanted to cover what is the meaning of this, the meaning of this is x 1 is less than our probability that x 1 is less than or equal to x 2, and probability that x 2 is less than or equals to infinity. So we wanted to let x 2 span all the values x 2 if you make x 2 less than or equals to minus infinity, nothing is left, that is going to be 0.

Now let us turning our focus on the discrete random variables. So we already note discuss that for the discrete random variables our interest is in what, probability mass functions because it is a discrete random variability is going to take only maybe at most countably finite, countable infinite number of points.

And if I know the probability at each of this point that is enough, so probability mass function is enough. So, maybe in this case also if some random variable is jointly discrete then maybe I can I should be have a similar probability mass function definition here.

(Refer Slide Time: 22:31)



So, we are going to say that if x 1, x 2 are each discrete random variable then they have a joint pmf, P of x 1, x 2 equal to x m which is defined as P of x 1, x 2 (())(23:30) let say x 1, x 2, x m this is nothing but probability that x 1 equals to x 1, x 2 equals to x 2 into x m equals to x m, so you are asking basically the question if x 1, x 2, x m are all discreet random variable.

You want to know what is the probability that x 1 takes exactly value x 1 and x 2 takes exactly value m and x m takes exactly value small x m, then this is going to be the, if we have such function this is going to define the joint Probability Mass Function of my set of discrete random variables x 1, x 2 up to x m and for this case also the way we recovered my marginal PDF here we can also recover your marginal PMF, I will just do for the simple case of 2 random variable suppose this x 2 and x y discrete and jointly distributed.

Then I am going to get P of x 1 of u 1 simply going to be just submit over the second random variable. So I have and this is whatever x 2 whatever its range is whatever the possible outcomes of x 2? And that is what I am submitting over all possible of value of u 2. If I do this I have already covered all possibilities of x 2 then what remains is only this part so, and that is what the probability of u 1 for random variable x 1, so it is clear what I mean by joint PDF and joint a Probability Mass Functions. And is that clear what I mean by marginal PDF and marginal Probability Mass Functions.

So, when I said joint you have m random variables that joint is going to say how it is kind of characterizing that all of their behavior together. And when I say marginal, I have recovered the distribution of the individual random variable in that from that joint distributions. So let us say let us give an example for joint, let us say PMF. So you might have already looked at all these urn kind of examples? Let us say you have in one urn there are 10 balls. In another urn, there are only, let us say 5 balls but of different colors.

Now you want to understand, you are going to pick some balls from this and you are going to pick some balls from this and now you want to come out with the joint distribution what is the probability that. So, you are going to pick some numbers from this some numbers from this you know what if the first ball has only first urn has only 10 balls and the second one has only 15.

So x 2 value can take what values all the way up to 1 to 5 and x 1 take all the values all the way from 1 to 10 and I do not know maybe like in some cases it may happen that when you have pick 5 from the first urn you will be allowed to pick only 2 from the second urn suppose let us say there is such a condition then it is forcing you a joint distribution there like some and when you can pick let us say 3 from the first urn, you will be allowed to pick only one from the second urn.

This has kind of enforced like joint thing. So, there it may be useful to come up with useful to use this joint distributions and from there you can further once you have a joint distributions, we have already said how to recover my marginals just integrate over all the possibilities. So, joint distributions basically come into run, when there is a constraint that the first occurrence puts on the second or like the second occurrence put some constraint on the second on the first.

So, for example, if x 1 is about the temperature and x 2 is about some other properties like it is very cloudy and it is very windy or such things. I mean you can or correspondingly map this outcomes to some numbers so, that you have a associated real random variable so, the first random variable temperature, the second random variable is the values associated with like cloudy windy all these things.

So, you naturally expect that if the second observation is cloudy too cloudy you may not expect the temperature to be more than like 30 or something of this sort and also the second random variable could also take one of the output could be like rainy, if you see that x 2 is already rainy.

Maybe the first temperature could be maybe just about 25-26 degrees. So, one event already constraints the possibility that the other event, the other experiments or other the random outcome of the second experiment. So, in such case we have to come maybe like in such case, the joint distribution is what completely determines, completely characterize how this how the outcomes look together when I am interested in both the values not about individual values.

(Refer Slide Time: 30:16)

Now we will define what we mean by formerly independence. So that is our next, so what do you mean by independence of random variables? So, we know what we mean by independence of events? What we mean by independent of 2 events?

Student: (())(30:34).

Professor: If I have to events a in b the probability of the joint happening together is nothing but the product of each of them. But random variable defines so many possible events it is not just one event. So, we have to appropriately define what we mean by independence for the random variables. So you are going to say that let us say the how m random variables to be independent you should take any set of events m events.

So, if you have any sets A 1, A 2 all the way up to A m and if you ask this like if you consider this events x 1 belonging to A 1, x 2 belonging to A 2 all the way up to x m belonging to A 1 if this events are independent I know what is independence of events these are independence.

If these events turn out to be independent. Then I am going to say that my random variable x 1, x 2 all the way to x m are independent. What is the crucial thing is, here the crucial thing is for any subsets it is not that when he said, so if it is raining, it is going to constraint my temperature to be within some range.

I just do not want, that is one even possibility I what I am asking for independence definition here is for every possible events if that set of events happened to be independent. Then I am going to call this x 1 x 2 all the way up to x m to be independent random variable. Is that clear?

Student: (())(33:38).

Professor: All possible subsets of any subsets m sets you just give me m subsets, like sets like this. And now I am going to construct this event x 1 belong to A 1 and all the way up to x m. And now I have m set of events. I know how to verify this m sets of events are independent or not, so how many conditions I need to check to verify this?

Student: (())(34:04).

Professor: Yes we already discuss, so if I if that happens, then I am going to call this x 1, x 2 all the way up to x m to be independent. So, independence is not for a specific event, this is for all possible. So if I am going to look for different A m maybe I will end up with different events. All possible events that are arising through these sets A m all the way this A 1, A 2, A m, if they are independent then I have to call this random variables independent. Fine so, as you see that if you want to do this, this is going to be a horrendous task, never ending task.

So, what is the consequence of this definition? It so happens that here is the result? Maybe we can take it as a theorem, so let like this point let us point ponder on some of the technicalities involved in this definitions. So, as of now, we are going to treat we are going to set any subset.

I mean A 1, A 2 any subset means it could be a closed subset or open subset, whatever subset it is. But there is, but any set you are going to take did not be measurable. So there are some, that theory that establishes this any set, any subset of R need not be measurable, but one of the things from the very definition of x 1, x m I know that if x 1 is the random variable it has to measurable.

So that this question that x 1 belongs to A 1 is a valid question. If this sets are not measurable, maybe I will not be able to assign probabilities to this there will be some issue, but that is going to take us to some measure theoretic things and we have to define this sets to be appropriately and there is some notion of Borel subsets that I need to look worry about. So, when I have Borel subset I need not worry about whether I can measure it and apply probability on that, but for our class and for our purposes we will just take this any subset is fine.

But if you make it we want to make it very very regressed you have to actually be calling it as Borel subsets, which has certain meaning, but that meaning is certain, like we do not need to diverge into that. But for as for all our practical purposes, any subset should be fine. And that is just an I said like I do not want to delve into that aspect which is I mean, that is not essential for this class.

So suppose I have such set of random variables x 1 x 2 all the way up to m, then we are going to say that are independent if and only if the joint CDF factors. So, what I mean for how x 1 x 2 all the way up to x m. So if you are going to take this joint CDF then, F of x 1 is a marginal CDF, F of x 2 is a marginal CDF for random variable x 2.

They are saying that this joint CDF splits as a product of individual CDFs. So instead of checking for all this, if you can check this condition, then it guarantee then it means that your set of random variables are independent. I am not going to prove this but I think this is not hard to show this if you at least one direction like if this holds you can right away write this expression for this and show that this quantity holds that it splits. And similarly if this holds you can show that the events are going to be independent for any subset of any set collection of sets A 1, A 2 all the way up to here fine.

So, if I say and notice that this is only if and only if condition so, if I say, my random variables x 1, x 2 all the way to x m are independent. Then you just need to check that their joint distribution splits. And now suppose I say my x 1 is through of 1 dice, and x 2 the through of another dice they are independent and then if I ask the question, what is the probability what is the joint probability that x 1 equals to 5 and x 2 equals to 6? How you are going to compute?

Student: (())(40:31).

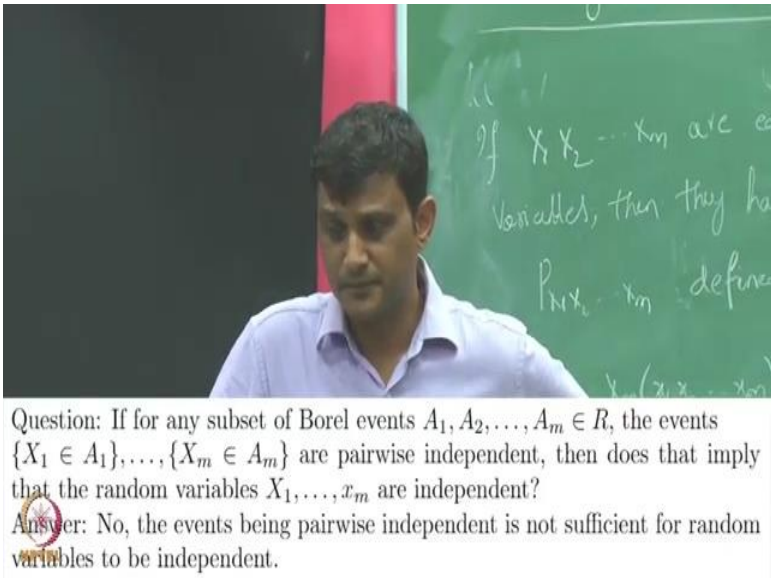
Professor: Multiply, why that multiplication is coming?

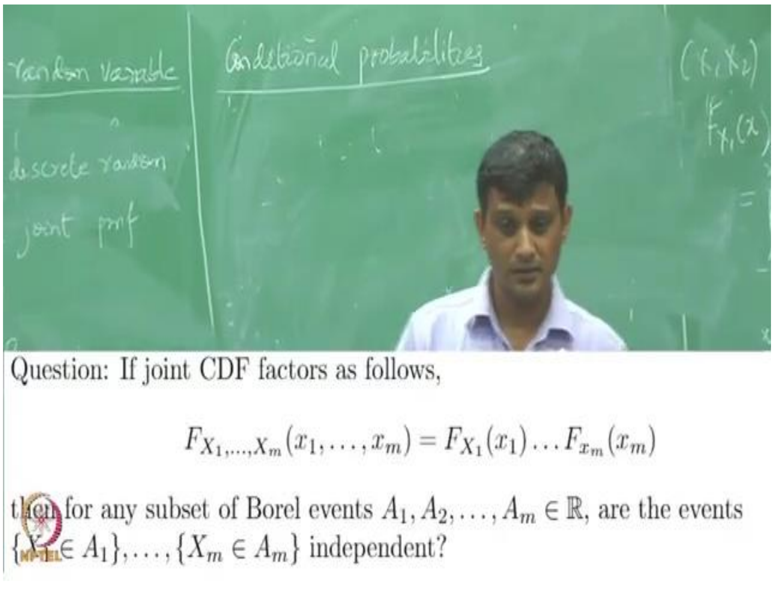
Student: Independent.

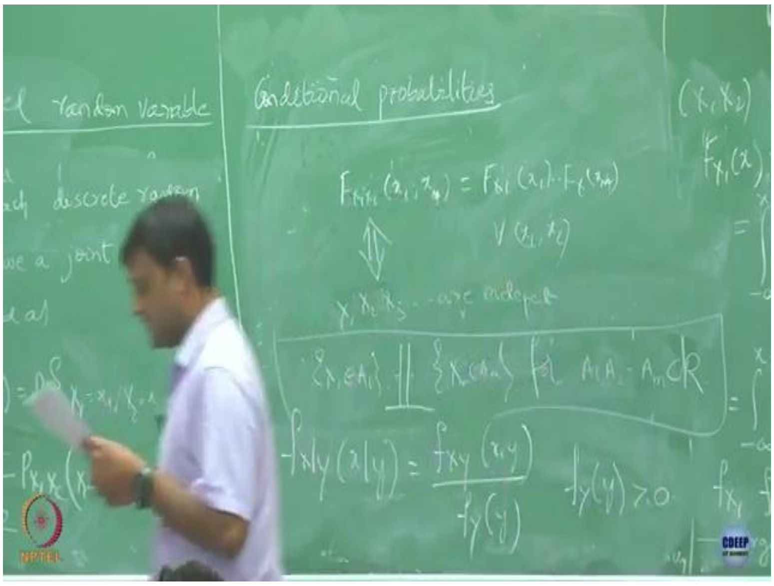
Professor: Because of independent so if you are going to just went to apply this for that particular case. And because this has independent I just asked first value is x 1 equals to 5 and x 2 equals to 6. Suppose if I asked the question x 1 equals to 3 and x 2 equals to 2, what would I have done?

I can do that multiplied? Even though I change the values. This is because independence is not for a particular event. This should happen for all possible events. So that is why when I say independence. You just do not worry which event is that you tell any possible event we are just going to look at the probability of individual events multiply and sell this is the joint probability of these things.

(Refer Slide Time: 41:46)







Professor: Conditional probabilities.

Student: (())(42:00).

Professor: No, we have to so we have distinguished what is independence and pair wise independence? So we have clearly said the set of events has to be independent. If they are pair wise independent, no I do not know. This is not does not implies my random variables are independent, they have to be independent, not just pair wise independent.

Student: (())(42:29).

Professor: If your joint CDF is going to split into each of them marginals one then because of the if and only if case it is true, that if you take any set A 1, A 2, A m, then the events that x 1 belongs to A 1 all the way up to x m belongs to A m. They need to be independent, they will be it is not like they are, that means they will be pair wise independent anyway and that independence means it is much more than that. Let us take the case of 2 ones if it is like this and it so happens that this is x 1 and F of x 2 for all x 1 x 2 when I said the splits which means all possible values of x 1 x 2, if it so happens then it by our definition, sorry from the previous result x 1 and x 2 are already independent.

Student: If there are m variables (())(43:44) it does not implies anything (())(43:50).

Professor: It depends.

Student: (())(43:53).

Professor: No, we are just saying, see like even you are going to take m random variable whatever, if this happens, what is our definition? So what the amplification of our definition? We are just saying x 1, x 2, x m are independent that is the only inference we are saying.

Student: (())(44:25).

Professor: That naturally happens like because we are saying this implies if this happens we already know x 1, x 3 are independent, this already implies that if you are going to take x 1 belongs to A 1 all the way to x m belongs to A 1 for any A 1, A 2 all the way up to A m, which are a subset of R that independent wherever this holds.

Student: (())(45:09).

Professor: So what is bothering here?

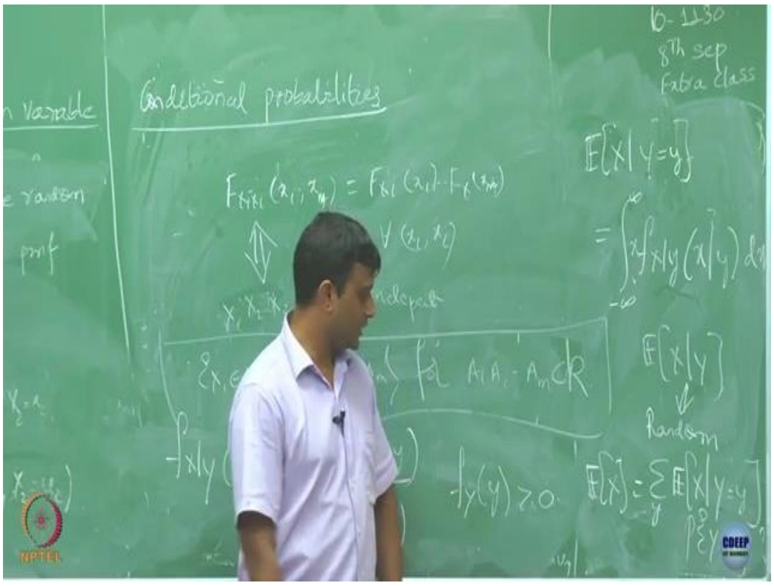
Student: (())(45:13).

Professor: If you satisfy that for all possible x 1 and x 2, we are saying that this is also guaranteed for all subsets of R. So x 1 independent means these are like these sets are independent. That is the definition, so if this satisfied of course this will automatically satisfied and this happens for all (())(45:47) we are also saying this automatically holds.

So for the conditional probabilities we are going to say that f of x given y, now what we are interested is suppose we have multiple outcomes, we want to ask that, given that this has happened what is the other events? What are the possible outcomes of the other experiments? We have already defined conditional probabilities.

Now we want to talk about conditional distributions extensions of that. So we are going to say that this is going to be f of x y, x y divided by f of x of y provided when naturally we want this to be such that the f of f y should be greater than 0 otherwise this is not useful. And last thing, so this is for the continuous random variable. For the probability mass functions when we have only 2 values, we already know how to define conditional probabilities. So I am only defining it for the continuous random variable.

(Refer Slide Time: 47:20)



So, one last thing now how to define them. What we mean by conditional distributions. So, one event has happened, something is told now I want to see what is the expected outcome from the second event. Let us say I have 2 outcome, 2 experiments, like 1 denoted x 1 another is going to be denoted x 2, if I see the first outcome happen to be something, what is the expected outcome from the second.

And this is what the meaning of this and here when I write it like this, what it means is, you have told me what is the value of y and then conditioned on this what is expected value of x. I have already defined given y what is the PDF? So, now you just go and apply the same definition of your expectation on this. Suppose, let us say I have like this, how we are going to define the CDF expected value of this, this is going to be minus infinity, infinity f of x given y of x given y to dx, this value I already know from this.

Student: (())(48:37).

Professor: In x and one more thing. Now, this offer will be written as we just write it as expectation of X given Y, if I just write this, it means that for given value of Y this is going to be once you specify me what is going to be Y this is going to be the expected value. If I do this there is no randomness in this. But suppose if I just write it X given Y I have not specified you, what is the value and conditioning on?

Y is that under variable, it could be taking many different possibilities here I have specified what is the value it is going to take, and then computed it expectation, but if I do not specify this, this means for different different values of Y this is going to take a different values. So, in that case, this itself is a random variable because it depends on what is the value of Y that is going to take. So, this is going to be random. Now, another thing you can see that, if you want to compute expectation of Y, what you could do is expectation of X and Y equals to y and probability Y equals to y and this is all possible values of y. So just like go and convince yourself how I am getting this. So this is like from conditional expectations, you are getting the expectation of X. So fine I had some.

Student: So that is into probability or...

Professor: Yes this is into you are multiplying this probability with respect to this expectation. So just to give an example for this suppose, let us say there are 2 things. One is you want to take, let us say you are going to take 5 exams and you (())(50:51) do not know which exams were going to be given so X 1 so Y, let us say Y denotes your which is that exam that you will be given and X denotes given that exam, what is your score? If I am going to tell you so different exams, you may be scoring different values.

So but before I give an exam, maybe like I am going to choose that exam with some probability, so first, we are going, so am going to choose that exam with some probability and after choosing that exam, this is the expected value from that exam we are going to get but now, if I will tell you all before, I will not tell you what exam I am going to give you just tell me the expected value you are going to score out if you have to take any exam. So then what we will do is we will basically average your score from all the possible exams you could take right?

So let us say there are 5 exam first one you are going to take with 0.1 seconds with 0.2, 0.3 like that, if you are going to take the first exam that would have taken the probability 0.1 and had you taken that what is your score? And like that, you do it over all exams and then this is your expected score in that case fine. So, this is what I want to talk about. There are a bit more about this conditional probabilities but related to that we will give you assignments. And from that, you will learn the things I have skipped in conditional probabilities, so let us stop here.