Discrete-Time Markov Chains and Poisson Processes Professor Ayon Ganguly Department of Mathematics Indian Institute of Technology, Guwahati Lecture 2 Review of Basic Probability II

Welcome to this lecture. In this lecture, we are going to talk about jointly distributed random variables. (Refer Slide Time: 00:37)

As you recall that in the last lecture, we have talked about probability and random variables. We will recap the concepts of jointly distributed random variables in this particular lecture. Now what is jointly distributed random variable?. It is nothing but a function $X : \Omega \to \mathbb{R}^n$. And what is the useful of a jointly distributed random variable?. Just recall that the use of the random variable is as follows that most of the times we have told that the outcomes of some experiment are numerical. Moreover, if it is not numerical, I can convert it into a numerical thing. If I map it into a real line, then I can able to use different sophisticated tools of real analysis to analyze the data or to analyze the characteristics of the probability distribution of a random variable. In this case that transformation is basically from Ω to $\mathbb{R}^n$ and Ω is basically our sample space. And here the idea is that, in some scenarios I need to talk about more than one random variables at a time. For example, maybe I am interested to see the relationship between height and weight. The intuition is that if height is more weight is more. So, I want to check that whether probabilistically that is true or not or with a high probability that statement is true or not that maybe I can try to see. Now, in such scenarios, I have two numerical characteristics, one is height and another is weight. So, naturally, we want to have two random variables together maybe X_1 and X_2, and if I clump them together that $\underline{X} = (X_1, X_2)$ that I can think of a function from Ω to $\mathbb{R}^2$. And when you just generalize this one to $\Omega \to \mathbb{R}^n$, I have basically n random variables there and I am talking about something about these n random variables. So, that is basically the intuition behind jointly distributed random variables. And in this case, I can again talk about joint cumulative distribution function. And what is joint cumulative distribution function?. Just recall that in case of one random variable, the cumulative distribution function was that it is nothing but $P(X \leq x)$. Similarly, the same definition is directly extended for a joint cumulative distribution function and the definition is given by this. Clearly, this is a function from $\mathbb{R}^n$ to $[0, 1]$. And the definition is strict direct generalization that the joint cumulative distribution function at the point $(x_1, x_2, \cdots, x_n)$ is same as $P(X_1 \leq x_1, \cdots, X_n \leq x_n)$

i.e.,

$$F_X(x_1, x_2, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n) \text{ for all } (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n.$$

So, that definition is direct extension of the definition from one dimensional case. And keep in mind that here I am basically talking about joint occurrence of each of the individual events. And these individual events are basically $\{X_1 \leq x_1\}$, $\{X_2 \leq x_2\}$, and similarly, the n^{th} event is $\{X_n \leq x_n\}$. So, I have n events first one is $\{X_1 \leq x_1\}$, second one is $\{X_2 \leq x_2\}$ and so on and so forth, final one is $\{X_n \leq x_n\}$. So, these are the events we want to consider and we will try to find out what is the probability of joint occurrence of these events and that probability is basically nothing but my joint cumulative distribution function of $(x_1, x_2, \cdots, x_n)$. And here this extra term joint is incorporated just to understand that, this is the cumulative distribution function corresponding to a random vector, not a random variable. (Refer Slide Time: 05:46)

Let us proceed. And the next thing we are going to see is that some property of joint cumulative distribution function. Just recall that we had for a distribution function F.
Now, the distribution function of a random variable, we have some property to be true. For example, we have, of course, that F is non-decreasing, F is right continuous and then we had that $\lim_{x\to\infty} F(x) = 1$, and finally, we had $\lim_{x\to-\infty} F(x) = 0$.
These are the properties are there. Now, so similar kind of properties we will try to see in case of a two dimensional joint cumulative distribution function. And keep in mind that in this case, of course, I have defined the cumulative distribution function for n variable in this slide, but when we are talking about these properties, I am basically writing this thing for only for two-dimensional thing. Basically my random vector $\underline{X} : \Omega \to \mathbb{R}^2$ in this case. That basically means that it is a two-dimensional random vector. And I am just talking about the property of JCDF, Joint Cumulative Distribution Function for a random vector which has two components in it. So, the reason behind this is nothing but this is much easier to write it down, otherwise the writing it is cumbersome, but that these properties can be easily extended to any general value of n. There is no problem about that.
With that let us start what are the properties. See the first property is that if I take the limit $\lim_{x\to\infty} \lim_{y\to\infty} F_{X,Y}(x, y) = 1$. And the idea is very simple. So, this is basically corresponds to the property, $\lim_{x\to\infty} F(x) = 1$, in case of one dimension and that becomes this one. The idea is very simple. The idea is basically nothing but, suppose I have the point (x, y) here, and joint cumulative distribution function is basically a probability. So, this is the probability of which region. This is the probability of all the points here I calculate the probability of this region and that is basically nothing but our CDF. So, the probability of this region is basically joint CDF at the at the point (x, y). Now, if I take $x \to \infty$ and $y \to \infty$, so x goes this side, y goes this side. So, finally, what will happen, I will cover whole $\mathbb{R}^2$. So, when I cover whole $\mathbb{R}^2$, the probability has to be 1. That is why this particular condition actually some correspondence to thie condition, $\lim_{x\to\infty} F(x) = 1$, in case of one-dimensional random variable.
Let us move to the second condition which is also corresponds to one of the condition in case

of one dimensional. $\lim_{x\to-\infty} F_{X,Y}(x,y) = 0$ and $\lim_{y\to-\infty} F_{X,Y}(x,y) = 0$ both correspond to this condition that $\lim_{x\to-\infty} F(x) = 0$.

So, from that we got it. And the idea again the same. If you see that again the cumulative distribution function at the point (x, y) is the probability of this region. Now, if I take $x \to -\infty$, this straight line actually is moving to this wards, and finally, when it moves to the side, then I have no points on which I try to find out probabilities. So, basically, it is something look like probability of $\emptyset$ and that is why the probability, that limit actually equals to 0 and note that this is true for all $y \in \mathbb{R}$.

And similarly, the same thing happens with the third property when this straight line goes to minus infinity, I basically have to find out the probability of $\emptyset$ so that is why basically this property also comes into the picture. Then I have this right continuity thing, but in this case note that this is the function of two arguments. So, the right continuity thing look like that it is the right continuous in each argument keeping the other fixed. So, if I fixed x then it is the right continuous function with respect to y. If I fixed y then it is a continuous function with respect to x. And where I am fixing x or y?. I am fixing anywhere I can fix, but it will remain right continuous with respect to the other one. That is basically next condition and it is clear that next condition corresponding to this condition here. And the final condition, the fifth condition corresponding to the first condition in case of the one dimensional random variable. And notice that the first condition actually comes from the fact that if you have, $a \subseteq b$ then $P(a) \leq P(b)$. So, if I have $a \subseteq b$, then I have that $P(a) \leq P(b)$. This condition I have. And this particular one actually coming from that.

Let us see how we get this one. Idea behind getting this one is nothing but, suppose I have a rectangle here and I try to find out what is the probability of the rectangle. Note that I can easily write the probability of this rectangle in terms of the distribution function. How?. That I can do in this way that suppose I just take this one. This is the distribution function at this point. Call this point as (b_1, b_2). Now, from that, everything here is now incorporated into the probability. So, what I have to do?. I have to basically subtract all the points outside this particular rectangle, but already incorporated. So, if I try to do that, you see that I can just subtract this part which is basically nothing but the distribution function at this point and this point maybe I call it (b_1, a_2), and then I am subtracting this part, which is basically cumulative distribution function at this point which maybe I call (a_1, b_2). And finally, what I have done this part I have actually excluded two times. So, I have to add the probability of this part and this point is basically nothing but (a_1, a_2). And the thing is that now you see that exactly what I have written here that is what I got that, I have taken the cumulative distribution function at this point. I have subtracted the cumulative distribution function of these two points. And finally, because these areas are subtracted twice, I have just finally added to that and this one basically nothing but the point (a_1, a_2), so I added this one here. Now, the point is that, the expression

$$F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2),$$

is nothing but the probability of this rectangle, and we know that probability is always greater than or equals to 0 for all $-\infty < a_1 < b_1 < \infty$ and $-\infty < a_2 < b_2 < \infty$. And that is why we have that this particular expression has to be greater than or equals to 0 for all possible values of (a_1, b_1) like this and (a_2, b_2) like this. So, that is why that fifth condition has to be true. So, these are some property of joint cumulative distribution function at two dimension. Of course, I can extend it to any higher dimension, but writing this condition will be little complicated in that case. So, that is why I am not discussing this one here. But a very important point I should mention here that if any two-dimensional function from $\mathbb{R}^2 \to [0, 1]$ satisfies all these properties, then that function has to be a joint cumulative distribution function at two dimension. So, these five properties are kind of a if and only if condition. If I have a two-dimensional joint CDF then they has to satisfy this point. On the other hand, if any function from $\mathbb{R}^2 \to [0, 1]$ satisfy these five conditions, then it has to be a two-dimensional joint CDF. With that, let us proceed. (Refer Slide Time: 17:10)

Discrete random vector that is what basically we are going to talk now. In case of that random variable we talked about is discrete random variable. Same definition we have extended here to discrete random vector. How the definition goes like?. The definition goes like I have to find out a countable set $S_{x,y}$ and this is a set in $\mathbb{R}^2$. So, I have to find out a subset of $\mathbb{R}^2$ such that it is at most countable. And if $P((X, Y) = (x, y)) > 0$ for all $(x, y) \in S_{X,Y}$ and $P((X, Y) \in S_{X,Y}) = 1$, then we can say that the random vector is having a discrete distribution. So, a random vector (x, y) is said to have a discrete distribution if there exists an atmost countable set $S_{x,y} \subseteq \mathbb{R}^2$ such that I have $P((X, Y) \in S_{X,Y}) = 1$. In this case, generally the set $S_{X,Y}$ is called support. And in case of a discrete random vector we have the notion of joint probability mass function and the joint probability mass function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ at the points (x, y) is nothing but

$$f_{X,Y}(x, Y) = \begin{cases} P(X = x, Y = y) & \text{if } \ (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

And of course see that if (x, y) does not belongs to the set $S_{X,Y}$, then of course that probability has to be 0. There is no other option. And this function is defined from $\mathbb{R}^2$ to $\mathbb{R}^2$ and this function is called joint probability mass function. So, it is again the direct generalization of that of probability mass function PMF in case of one-dimensional random variable and that has extended to two dimensional or three dimensional, even you can write it for any general n dimensional random vector. (Refer Slide Time: 19:23)

Next one is basically how I can find out the expectation of a function of discrete random vector. And the definition is exactly the same before. So, I try to find out the expectation of the function $h(X, Y)$ which is nothing but the summation

$$E(h(X,Y)) = \sum_{(x,y)\in S_{X,Y}} h(x,y) f_{X,Y}(x,y)$$

provided

$$\sum_{(x,y)\in S_{X,Y}} |h(x,y)| f_{X,Y}(x,y) < \infty.$$

And that sum I have to need to take over the support just like in case of the discrete random variable and in this case I have to check this property and this property again needed so that this summation is a meaningful summation. (Refer Slide Time: 20:20)

Now, let us move to continuous random vector. And again, the definition of the continuous random vector goes exactly the same way that of the continuous random variable. In this case what do we have?. We have basically if I can find out a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ that is non-negative integrable function and if I can write the joint cumulative distribution function as the integeral of this non-negative integrable function i.e.,

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$$

for all $(x, y) \in \mathbb{R}^2$. Note that this x is here, this y is here. So, if I can write in this particular form then I can say that the corresponding random vector has a continuous distribution. So, the definition is exactly same. And in this case the function f is called joint probability density function. That function $f(X, Y)$ is called the joint probability density function. (Refer Slide Time: 21:35)

Let us proceed with that definition and let us see how I can find out the expectation. The idea of finding out the expectation, again same as that of continuous random vector. And in this case, if I try to find out the expectation of $h(X, Y)$, I take the double integration of $h(X, Y)$ multiplied by the density function. And this condition again need to be checked so that this integration is a meaningful integration. So, we have to check whether this integration is finite or not. Again, note that the same thing we have mod here and the rest of the things are same as before. (Refer Slide Time: 22:20)

With that let us proceed. Now, let us talk about something called independent random variables. Independent random variables plays a major role in probability and statistics. What is the independent random variables?. The intuitive idea is that, if suppose I have two random variables x and y, or x_1 and x_2. Now, if the value of one random variable does not affect the value of another random variable anyway, we say that random variables are independent. So, just like suppose I am tossing two coins separately. The outcome of one coin does not have any effect on the outcome of the other coin. Now, in this case, these two things are completely independent. One has no effect on other. The same kind of thing we will going to talk about now, but we will going to give the mathematical definition of that and the mathematical definition goes like this. That if $X_1, X_2, \cdots, X_n$ is a collection of random variables, we say that these are independent if the particular expression

$$F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \Pi_{i=1}^{n} F_{X_i}(x_i),$$

is hold true for all $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$. This for all is important. So, what does this mean?. This basically mean, notice that the left-hand side here is nothing but the joint cumulative distribution function of $(x_1, x_2, \cdots, x_n)$, and on the right-hand side, this quantity is nothing but the CDF, the Cumulative Distribution Function of xi and then I am taking the product over all. So, this quantity I can write as $F(x_1)$ at the point x_1, $F(x_2)$ at the point x_2 so on so forth, finally, $F(x_n)$ at the point x_n. And in this case, you see that basically this means that if the joint cumulative distribution function can be written as the product of marginal cumulative distribution functions, and if it is true for all values of $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, we say that the random variables $X_1, X_2, \cdots, X_n$ are independent random variables. So, by the way in the context of jointly distributed random variables, this CDF is called the marginal CDF just to signify in the context of jointly distributed random variables. Now, this is of course a very general definition, because the cumulative distribution function and the jointly cumulative distribution function exists for any random variables and any random vectors, respectively. So, this is a very, very general definition and this definition I can use for any kind of random vectors. In case of the discrete and continuous random vectors, we can give the equivalent definition. Notice that the main definition is basically given in terms of the cumulative distribution function as I mentioned. Now, in case of the discrete random vector, suppose my (X, Y) is a discrete random vector. In this case, the equivalent definition can be given in terms of joint PMF and marginal PMF. Suppose (X, Y) is discrete random vector, then if I can write the joint PMF of (X, Y) same as product of the marginal PMFs of X and Y for all (x, y) in $\mathbb{R}^2$, we can also say that X and Y are independent random variables. The similar thing is also true for continuous random variable. Suppose x and y are continuous random variables and also, I am assuming that (X, Y) as a vector are continuous random vectors, then if I can write the joint PDF is same as product of the marginal PDFs for all values of (x, y) in $\mathbb{R}^2$, I can tell that X and Y are independent random variables. So, of course, in this case, I have written these for two-dimensional random variable, but it can be easily generalized for a multi-dimensional one. For example, if I have n random variable $X_1, X_2, \cdots, X_n$, in case of $(X_1, X_2, \cdots, X_n)$ is discrete random vector, this is the joint PMF, and in case of the continuous random vector it is a joint PDF. If I can write this one is same as product of $f(x_i)$ at the point $x_i, i = 1, 2, \cdots, n$ for all $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, then we can tell that $X_1, X_2, \cdots, X_n$ are independent random variables. So, again, notice that in case of the discrete random variable this is the joint PMF and in case of the continuous random vector this is the joint PDF. And similarly, in this side also, in case of the discrete random variable it is the marginal PMF and in case of the continuous random variable it is the marginal PDF. So, this is the equivalent definition. And of course, this definition is only valid for either discrete random vector or continuous random vector. And in general, of course, I cannot use this definition, because for general kind of random variable the PDF and PMF may not have any meaning. So, that is why basically this is a particular case of discrete or continuous random variable and sometimes this definition we can use very easily compared to the original definition which is given in terms of the cumulative distribution function. Now, you see that in this case actually what I have is that, I can again find out the expectation very easily. So, suppose X and Y are independent here, if I try to find out the expectation of $g(X)$ into $h(Y)$, then I can write it as the expectation of $g(X)$ multiplied by expectation of $h(Y)$ i.e.,

$$E(g(X)h(Y)) = E(g(X))E(g(Y))$$

provided all the expectation to exist. And keep in mind here the functions are product and g is the function of X only, h is the function of Y only. Then I can find out the expectation of $g(X)$, I can find out the expectation of $h(Y)$, I product them, I get the expectation of the product. So, easier way to remember this one if X, Y are independent, the expectation of product is same as product of expectations. That is the easier way to remember.

With that, I stop for this particular lecture. In the next lecture we will see some more concepts of probability random variable, random vectors, especially we are going to see the conditional distribution in case of random vectors and which has a lot of use in our Markov chain part. Thank you for listening.