Discrete - Time Markov Chains and Poisson Processes

Professor Ayon Ganguly
Department of Mathematics
Indian Institute of Technology, Guwahati

Lecture 03

Review of Basic Probability III

In the previous two lectures, we have talked about some concepts of probability and random variables then jointly distributed random variables. Now, in the third lecture of this particular course, we are going to talk about conditional distribution of random vectors or random variables.

So, what is the conditional distribution for discrete random variables, we will first start with the discrete random variable, and then we will proceed to the continuous random variable or continuous random vector also. So, in case of the discrete random vector, how we can talk about the conditional distribution? And the idea in this case is very very simple, suppose that (X, Y) be a discrete random vector with joint probability mass function, $f_{X,Y}(\cdot, \cdot)$ and suppose that the marginal PMF of Y is $f_Y(\cdot)$.

So, what is the marginal PMF of Y? That is nothing but, if (X, Y) is discrete random vector, it can be shown that Y is a discrete random variable, so it has a PMF. So, that PMF is called the marginal PMF in this context, and in this case we can talk about the conditional PMF of X given $Y = y$ is defined by this particular expression

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{P(X = x, Y = y)}{P(Y = y)} = P(X = x|Y = y).$$

And what is this expression, if you see that from the definition of joint probability mass function, the numerator is nothing but $P(X = x, Y = y)$ divided by $P(Y = y)$, the denominator is $P(Y = y)$ that is directly from the definition of joint PMF and the marginal PMF.

Now, this particular quantity is nothing but this conditional probability, that is actually nothing but $P(X = x|Y = y)$. So, that is basically the conditional PMF and this conditional PMF is defined only if $P(Y = y) > 0$, which basically means that $f_Y(y) > 0$. So, this condition has to be there otherwise, I cannot define the conditional PMF of X for given $Y = y$, that is needed.

Now, from the definition it is clear that this function is basically a probability mass function, of course, the conditional probability mass function, but it is the probability mass function. So, I can write the corresponding conditional CDF and which is basically coming from the fact that CDF is nothing but $P(X \leq x|Y = y)$, i.e., probability of X is less than equal to x, but I know that the value of Y is y. So, in this case, I can just sum it up the probability up to the point x. So, that is what basically written here

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{\{u \leq x:(u,y) \in S_{X,Y}\}} f_{X|Y}(u|y),$$

provived $f_Y(y) > 0$.

And again, this probability is only meaningful if the probability of the given event is strictly greater than 0. So, that basically means that $f_Y(y) > 0$. So, in this way, I can talk about the conditional distribution for a discrete random variable. And it is a very very important thing as we see in this course, in most of the cases we are going to talk about the conditional distribution.

So, it is the important thing. Keep in mind that well in case of the discrete random variable, the conditional PMF of X, given $Y = y$ is defined by this particular expression. And this is basically nothing but $\frac{P(X=x,Y=y)}{P(Y=y)}$. And this is only defined if $P(Y = y) > 0$ and note that this is a function from $\mathbb{R} \to \mathbb{R}$ or rather, I can write it a function from $\mathbb{R} \to [0, 1]$ because it is finally a probability, so $\mathbb{R} \to [0, 1]$. Keep this function as a function from $\mathbb{R} \to [0, 1]$.

And now of course, we can talk about the conditional CDF and that is nothing but if you take the sum of the conditional PMF over the appropriate points, we get the conditional CDF. So, again the idea behind this conditional distribution or conditional PDF, just same as that of the conditional probability because it is finally a conditional probability and the idea is that, if I know that $Y = y$ has occurred, what is the probability of $X = x$ has occurred or, if I talk about the conditional CDF that is, if I know that $Y = y$ has occurred, I try to find out what is the probability that $X \leq x$.

Let us proceed now, we can talk about something again very important concept is conditional expectation for discrete random variable and idea is very simple. As I have mentioned that $f_{X|Y}(\cdot|y)$, this one is basically a conditional PMF, so finally it is basically a PMF, it is a conditional PMF fine, but it is basically a PMF.

So, with respect to this PMF I can try to find out what is the expectation of the corresponding random variable. So, that is basically exactly that and, in this case, a little bit generalization I have written, instead of taking only that random variable, maybe I can talk about some function of the random variable.

So, if I try to find out $E(h(X)|Y = y)$, so I try to find out that if I know that $Y = y$ has occurred, what is the expectation or what is the long run average of $h(X)$ given $Y = y$ has already been occurred, I try to find out what is the average of $h(X)$ that is basically this one. And that one basically, I can find, like these way that the simple thing and basically that because I am trying to find out the conditional expectation, here the PMF is basically the conditional PMF of X given $Y = y$.

So, that is basically the thing and, in this case, also, this absolute summability is needed and that basically means that again same as before that it is nothing but $\sum_{\{x:(x,y)\in S_{X,Y}\}} |h(x)| f_{X|Y}(x|y) < \infty$. That absolute summability basically means this in this case.

So, let us now proceed to the continuous random vector. So, in case of the continuous random vector, the joint PDF, the probability density function of X given $Y = y$ is defined by this expression

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

So, just note that the expression is just same as that of the discrete random vector that joint probability density function divided by marginal probability density function and for discrete random vector it is joint probability mass function divided by marginal probability mass function.

So, the definition is exactly same, but this condition again is needed because I am dividing by that probability density function at the point y. So, that condition has to be true if it is 0 then I cannot define this ratio. So, that is why basically this has to be true.

Now, again I can talk about the expectation, the definition of the expectation is same as before just the summation need to be changed with the integration and again this condition is needed otherwise that joint density function is not defined. And then this is the expectation and we say the expectation is finite or expectation exists if the absolute integrability is there if not there, we say that expectation do not exist or expectation is not finite.

So, let us proceed. Now, let us talk about something which is very very useful in computing the expectation. So, we can compute the expectation by conditioning, it is something like resemble with theorem of total probability, I can find out the probability of a set by conditioning on several other sets, the same kind of concept goes through here. So, what happens suppose I try to find out $E(X)$ that if Y is another discrete random variable I can find out $E(X)$ by this expression.

So, what I did, I basically write the conditional expectation multiplied by the marginal probability and I then take the sum. The idea is very simple, suppose in a class there are two columns of benches and the teacher is here. He is teaching to the class. Now, suppose the teacher wants to find out what is the average height of the students in the class, the thing is that X is a random variable which denotes the height, now, the thing is that the teacher try to find out what is the expected height.

Now, there are two ways to do this one, one is that, I measure the height of all the students in the class and then I take their arithmetic average that is standard way to do this one. Another way I can do this one is as follows suppose there are n_1 students in this column and n_2 student in this column. So, what I can do is that I first try to find out what is the average of this column, maybe I call it $\bar{x_1}$ and I call the average of heights of this column as $\bar{x_2}$. So, then to find out the final average I can do this one it is $\frac{n_1\bar{x_1}+n_2\bar{x_2}}{n_1+n_2}$.

All of you know this formula; this is a very preliminary formula that we learn in school levels, so all of us know this formula. Now, note that this one I can write as $\frac{n_1}{n_1+n_2}\bar{x_1} + \frac{n_2}{n_1+n_2}\bar{x_2}$. Now, you see what basically I am doing you can think this one as a probability that a student is in this column. And this one is again the probability that the student in the second column, probability of first column and this is the probability of second column.

So, now what is this $\bar{x_1}$? That is the average of this column; this is the average of this column. So, finally what I have, probability multiplied by the average given that the student in the first column, so that is the conditional expectation. Similarly, plus probability multiplied by the conditional expectation of the height such that the students are in the second column and you see that exactly same thing has written here, the thing is that if you take Y is the random variable taking two values: 1, if the student in the first column, 2, if the student in the second column and then this summation, this expression is exactly the thing that I have written here. So, this one is you can see as a generalization of the concept that you have done in your school level, but of course, that has returned in a general form in terms of probability in terms of the conditional expectations.

So, this is a very very useful expression, useful formulas we can see many times that finding out $E(X)$ is quite difficult, but when you try to find out the expectations by conditioning that become quite simple and in our Markov Chain part and then the subsequent part of Poisson Process we will actually see this kind of examples.

So, this is for Y discrete, similarly I can do it for Y continuous, the only thing is that, keep in mind one thing this probability is basically same as I can write as marginal probability mass function of Y. And then for continuous random variable again the same standard thing is there, that the summation has been replaced by the integration and then probability mass function is replaced by the probability density function. So, these two are quite important, quite useful formulas to compute expectation by conditioning and with that I let us proceed.

And the next slide we are going to see how I can compute probability by conditioning. So, this one again is, you know that generalization of theorem of total probability. So, if Y is district this is given by this if you compare with the last expression in case of the expectation, this was the summation, conditional expectation multiplied by marginal probability mass function.

Now, when I try to find out the probability, it is nothing but the conditional probability multiplied by probability mass function for discrete random variable, and for continuous random variable, conditional probability multiplied by probability density function and then take the integration.

So, that is what the thing and this one actually directly comes from the expectation formula because, just keep in mind if I have a random vector, this one which is basically a function $I_A(x)$ which takes values 1, if $x \in A$ and 0 otherwise, i.e.,

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

then it is very easy to see that the $E(I_A)$, a random variable I_A, is same as the random variable I_A takes two values, one is 1 and other is 0. So, it is nothing but $1 \times P(I_A = 1) + 0 \times P(I_A = 0)$, because I_A is a discrete random variable, it can take only two values.

Now, the second part does not contribute anything because I have a multiplication by 0. So, that basically turns out to be $P(I_A = 1)$ and $I_A = 1$ if and only if $x \in A$. So, that basically turns out to be $P(X \in A)$, so the $P(X \in A)$ can be written as $E(I_A)$ or maybe if I write it correctly, maybe I just put an X here just for understanding that maybe I can write X here, so then everything goes like that.

Now, you see that the probability can be written as expectation and these random variables generally we call indicator random variable. So, probability of X belongs to a set can be written as the expectation of indicator random variable of the set A.

And now using this one, finding out this one is very simple, because see $P(E)$ is basically I can write as $E(I_A(X))$ and then basically this one continues. Then basically we can use the previous slide's material that the expectation can be written in this form, and we can proceed, and finally we have that this expression holds true.

And as I mentioned earlier that computing expectation by conditioning or computing probability by conditioning are very very useful tools to find out the expectation or probability and we will see these, some example, as we go through with our Stochastic Process, with our Markov chain, with Poisson process in this particular course. With that I stop and thank you for listening.