Introduction to Stochastic Processes

Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay

Lecture 10

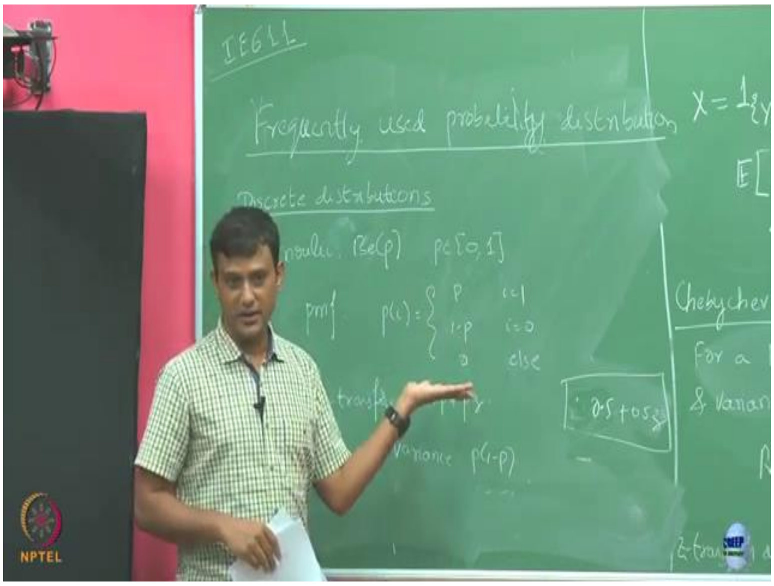
Discrete Probability Distributions

So next, what we will do is, so far we have been just defining random variables what possible characteristic one should be interested and we defined all its expectation variance and then characteristic function. And now we are going to see, we also just defined the notion of a probability distributions right, like before we started defining random variables.

Now, what we want to now look into so good, I can come with any probability distributions I want for which I can define all these, whatever we want do like starting from expectation, I can go all the way of to compute characteristic function and all the moments I can compute. Fine, now we are going to look into certain set of the probability distribution, that are useful for modeling I do not want to use any arbitrary probability distribution, we know that pro, a probability distributions is something as from there is satisfies this axioms this is a probability distribution for it.

But, I am interested in modeling, modeling and analyzing a system when I want to model and analyze I want it to be kind of also mathematically tractable right, so that I can write the expression and try to derive some intuition of what is happening. So people often use a set of distributions which are like seems to reasonably capture the things we want model but also more importantly they are kind of tractable. We can explain to each other well, what we mean by this probability distribution. So somehow the standard distribution that we are going to use, that we are going to discuss next, some of them we have already used but, we did not explicitly named them.

(Refer Slide Time: 02:36)



Now, about this characteristic functions and all, you will see exercise questions related to them in that assignment, like you will see that, we have they are going to be useful, and you are also going to see that, where they are going to come when I am going to discuss various probability distributions here.

So, there are subtract distributions people use, when you are trying to model something discrete where the outcomes are discrete and set of random variables when you try to use, you try to use when you are trying to model and experiment which is going to take continuous valued outcomes.

Student: (())(03:36) variance as the deviation, means as the average value of.

Professor: It is just like a mathematical convenience for us, which will help us to relate different moments with the characteristic function and as I said, it unitly determines the probability distribution function. So, this is mostly like a mathematical convenience for us.

So, now we are going to talk about first, discrete distributions, so first thing is the one which are used already many times, this is going to be, and how this probability distribution looks like, so I am now talking about probability distributions of the discrete random variables. So, I have only I will only give the probability mass functions.

So, now pmf of this in it is going to be defined as p of i equals to p, i equals to 1, 1 minus 2, i equals to 0, and 0 else, and it is the z transform, it is going to be 1 minus p plus pz so, somebody quickly calculate if I have pmf like this what is the mean value of this, p and what is the variance, so goes where is this distribution useful, it we used this distribution already.

So I am saying so I am going to call something a Bernoulli distribution with parameter p where it going to take value 1, with probability p, I am going to take value 0, the probability 1 minus p and all other values are 0. This is already completes like p minus 1 minus p means everything else is 0.

I am going to call such a, such a random, such a distribution as Bernoulli distribution, so where this is useful, coin toss, so if I am going to coin toss my outcomes are heads and tail but, on this I am going to put a random variable and map, head to may be 1, tail to 0, then head is going to happen a probability p, tail then 1 minus p, and this is exactly this, so to characterize this distribution or like it is the single parameter p.

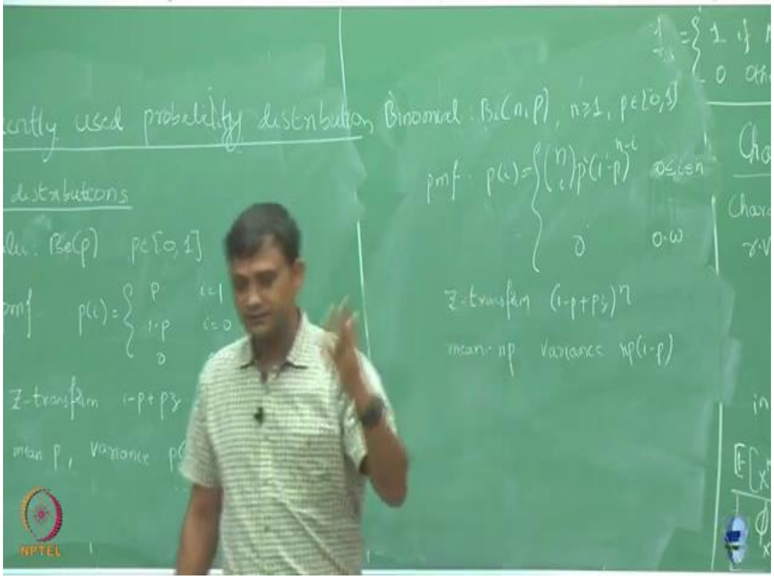
That was what we have been using, when you are doing a coin toss. So p is the bias of the coin when we did a coin toss and that value could be anywhere in the interval 0, 1 right, if it is a fair coin, p is his half otherwise it could be any value between 0 to 1.

And now, you should can for this like, we can compute z transform and whenever you see a somebody gave you z transform if it is of this form so for example let say, I have something like 0.5 plus 0.5z if I say that a random variable has a z transform of this form then what is the understanding you will have, that it is a it corresponds to Bernoulli distribution with what parameter.

Student: 0.5

Professor: It is going to be 0.5 or may be like somebody can just write just give it in this form, then may be you can just manipulate and see that you can relate it in this form, and then realize that, just like a Bernoulli random variable with this parameter.

(Refer Slide Time: 08:21)



The second is you call it as Binomial and it has two parameters n, p then n is going to be greater than or equals to 1 and p into belongs 0, 1 and its pmf is of this following format. So, suppose I have a pmf in this format, where probability that it takes value i is, n choose i p to the power i 1 minus p, n minus i and any other value is 0. So, you can clear you can verify that this will have some to 1. So, this is a valid probability mass function. So, now can somebody compute quickly mean value for this?

Student: (())(10:15).

Professor: np why, what, why it is np, what is variance?

Student: np 1 minus p (())(10:28).

Professor: np 1 minus p, so why it was np what?

Student: (())(10:37).

Professor: It is n time Bernoulli distribution what is that means.

Student: (())(10:46).

Professor: So, what is this capturing here, if you want to model something, if you have made n successive process of your coin, what is the probability that you are going to see, n heads from this sorry, i heads from this in out of this n trials given that probability that had coming in each of this trials is going to be p.

So, and you see that like if I am going to have a probability mass function this, I have a many situations that I can use this 1, 2 model and this is given a like binomial distribution. The another example is suppose, so let say there is an target and you want to hit it down, and if you are you have let say missiles and each missile has a probability that it is going to hit the target is some p.

Now you can identify, each one has a probability of hitting the target is p, what is the probability that if I am going to hit 10 missiles, let 3 are going to hit the target, or let 5 are going to hit the target, or alternately you can pose a design question as following suppose, there is a target, I want to make sure that I need to hit the target, with probability 0.99 you call like, let say I am guarding a very secure area which need to be given a very very high standard of security and I need to deploy all my security missiles, whatever.

Now, the question is, if a target comes, into this territory, how many missiles I should deploy so that I hit the target, with probability 0.99. So, in that how you are going to pose this question, may be you want to say, I want this probability, to be greater than or equals to 0.99, and you know p, you know p the success probability of your missile.

Now, then you want to say like what is that n, so that I should use so that this probability becomes larger than 0.99. So, you can come up with many such cases where you can easier Bernoulli distribution to model things.

Then, geometric and it has a pmf which is, so I have a pmf which says that for any i greater than or equals to 1, the probability is 1 minus p to the power i minus 1p. So suppose, if I want to ask the question what is the probability that i equals to 3 this will say that, that is going to be 1 minus p square and then p, and any other values are going to be 0.

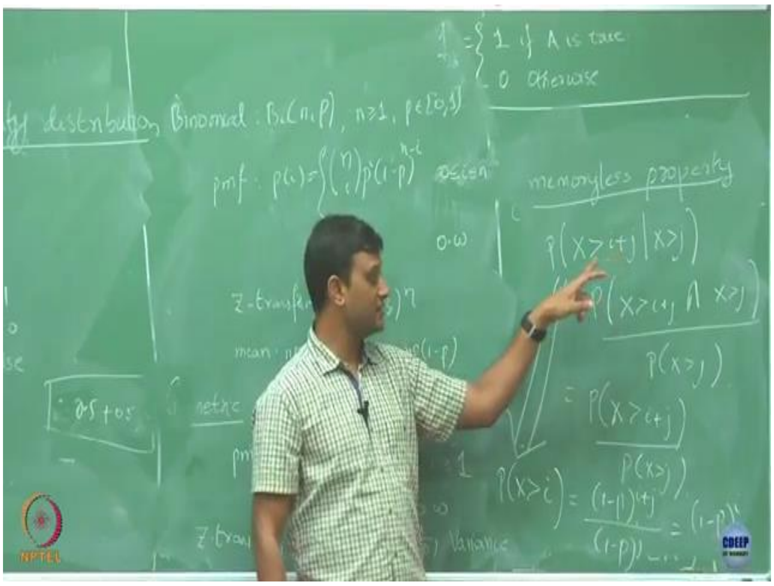
And here, I was written i to be greater than or equals to 1, so this can go on, it can take 123 all the way up to any value if like so like these are only restricted to 1,0 and here it is only up to n. Now, suppose you have a case where you want to model something like, it is not necessary that every time I am going to try, I am going to succeed but I keep on repeating, repeat the same thing again and again till I succeed.

Now, your question could be how many trials I need or what is the probability that I will succeed in some i th attempt and that probability could be given by this priorate that the probability that your success in each attempt is, p here.

So for example let say, again let us go back to other coin toss problem, let say if you keep on tossing your coin if you are going to see a head you win, otherwise you lose, and every time you toss a coin the probability that head shows up is p. Now what is the probability that you need i travels to see a head, it will exactly this, so you fail in the first i minus 1 trials and then you succeed.

So this is basically gives the probability on succeeding for the first time, in the i th trial, and the like it depend, like you may if you have a scenario like this we want to ask the question, what is the probability you succeed in 10 trial, 100 trial like that whatever it is. So, geometric distribution here, the way I have defined here has one nice property called, memory less property.

(Refer Slide Time: 17:29)



So, let discuss what is that, so suppose I want to ask the question so let say x is why, a random variable which is going to capture the number of trials before i succeed. Now, you have been told that the number of trials it took to succeed is already more than j. And now you are asking the question what is the probability that this is going to be more than i plus j, you already have this information let say, somebody told that guy took 4 attempts to clear the exam.

Now, we want to ask the question what is the probability that guys going to take 7 attempt to clear the exam, given that he was already taken 4 attempts to do this, how does it is going to be what?

Student: (())(18:41).

Professor: This memory loss now, you are calling doctors memory loss property.

Student: Because it has forgotten that it has already taken there. it starts from a new point.

Professor: What, start from new point?

Student: (())(19:00) after j.

Professor: And first once lets do not jump into, so I have this explanation, so when you have this all I know is, before I do anything this is the conditional probability and then to apply the definition of my conditional probability and see what this look like, so if I am going to apply my conditional probability, definitely this will x greater than i j into section x greater than j given, x greater than j. I have just applied the definition of my conditional probability.

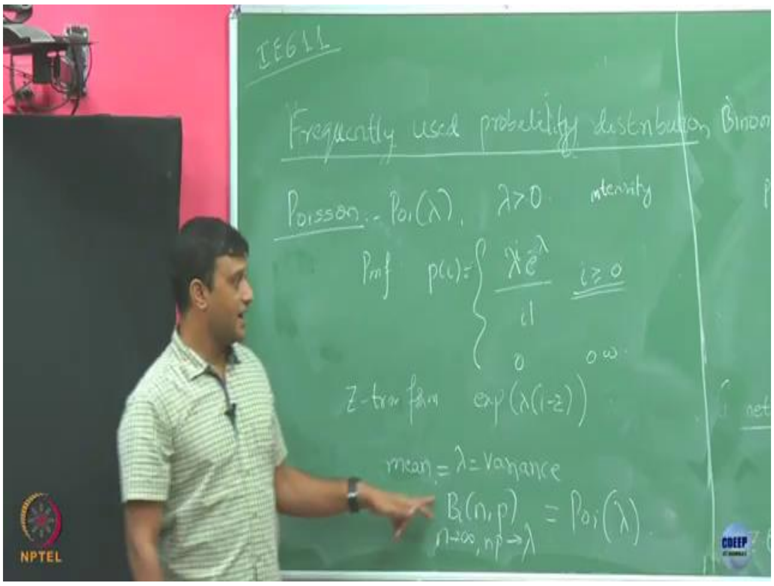
Now, what is this probability in the numerator, x greater than i plus j, now for this case can we compute what is the probability that if x is my geometrically distributed can you compute what is the probability that x is greater than j. So, probability that x is greater than j, you need to add all these terms for values greater than j, so if you add that you will see that it is simply going to be 1 minus p to the power j. So, in this case it is going to be and this is going to be, 1 minus p to the power j, and if you simplify this, 1 minus p to the power i.

So, given that you are already taken j attempts and requiring further i more attempts that you already taken more than j attempts is going not helping you anything about in understanding whether you need at least j more attempts, i more attempts. So in that sense this is the memory less property.

You see this like why, there is a given that you already took more than j attempts and now if you want to ask i more attempts in addition to this, the fact that you already taken j attempts is not going to help you so that you are going to, you are understanding about you are going to need i more attempts.

So, in that way this is what you are going to that is why we are going to call this as property and this is nothing but simply, probability that x is into i. And this probability 1 minus p to the power is exactly x is greater than i. So, we are saying that this probability is equals to this, it is x is going to greater than given that x is going to greater j, this is going to happen is simply, x is going to greater than i.

(Refer Slide Time: 22:03)



Then, we have Poisson. So, this geometric as I said it, takes all possible integer values starting from 123 like goes on like that. Another such thing is Poisson 1 and its pmf is given as, and it defined as, Poisson of lambda when lambda is greater than 0. And it has p of i is equals to lambda i p to the power lambda divided by i factorial for i greater than 0 and 0 other wise for any other non-integer.

So, when I say i greater than or equals to 0 these are like integer values i0 1234 like that and if you are take any value this is going to be 0. That means it has a mass only on the integer values, and then its z transform is going to be, and its mean is equals to lambda and that also happens to be variance. So, if I have a distribution a probability mass function like this I am going to call it as Poisson distribution.

Now, so where you thing such a probability mass function is going be useful, it is basically telling what, it is basically assigning some number to each integer value, so this Poisson and here this lambda is called intensity. So, this Poisson distribution comes very handy in situations where you want to basically do counting.

Suppose we want to count let say you are, you want to count let say, you are in a some traffic signal just imagine like IIT main gate, you are there, you want to count how many vehicles are passing when signal opens like, so may be in the late night like around 2 3 or 4 am may be 1 or 2 vehicles will pass but as (())(24:58) like actually nothing passes right, everything is traffic jam.

But, still like you can count it as like lot of vehicles moving and you can basically want to model something like, how many vehicles passed, so here like number of vehicles is an integer number right, so it could be and it could be taking any value, 123 and it could be taking large number of value. So here, may be like to model such things Poisson distribution come very handy. We have the geometric distribution but, that was more suitable for what, number of trials till success. But, this Poisson distribution has somewhat properties which we will discuss later.

So, but this comes very much when we actually want to count how many things happened, like for example, may be in a traffic signal you want to count how many number of vehicles passed, or whatever you want to count. So, I think this is a popularly used in a case where, you want to count in a given duration how many calls arrived, like so, you are in a telephone exchange, you want to manage your resources right, like you have certain number of switches or bandwidth whatever, you want to see like a people are calling from different location, at any given time, how many calls come, and that may you may want to model as a Poisson distributed.

So, it depend on all applications I am just listed you, some set of distributions here, and it is not necessary that whatever the thing you want to model is not like I it is not like something, geometric is the best it depends on the application for something may be geometric use and not perfect but, the best thing I can go for, I mean the nearest realistic approximation is geometric. Or for something may be Poisson is the nearest realistic approximation, I would like to go for, and these are the one which we have better understanding of this distributions.

We know how they look, what is their mean and also we are able to I mean, analytically play with them well. As somebody said, you can verify that Poisson, so maybe I will just write it as when you have this binomial n, p then you let n go to infinity but, you will let go it in a control fashion such that the product n, p goes to infinity.

So, this will be equals to Poisson of lambda, you understand this notation, what I am saying, if you have a binomial distribution in this like this you will at, let n go to infinity but, p go to 0, that is you have large population but success in each round is very very small, and you make sure that this n, p product convert just to lambda and then in that case your binomial distribution is nothing but the Poisson distribution, this is an exercise please complete this, check that you are able show this formulae.

And all these z transform mean and variance, I have just written please verify them they are correct, and there is something called hyper geometric distribution, I do not have complete information about, just go and look for it what is hyper geometric distribution, what are the parameters in that, what are the mean, variance of that.