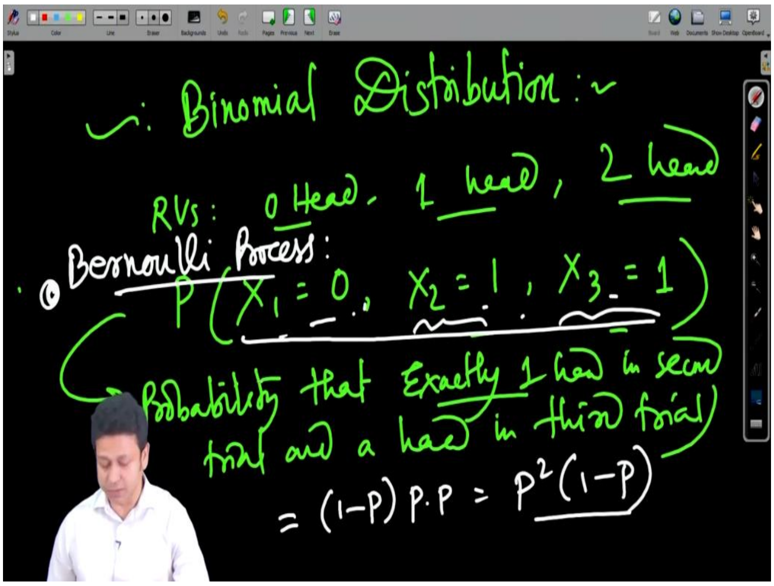
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 2
Lecture - 21
Bernoulli Distribution (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. So, we are talking about probability distribution. And in the previous lecture, we discussed about Bernoulli distribution. Now, going forward, stay forward, we will see the application of Bernoulli distribution. And the next application that we see is the binomial distribution.

(Refer Slide Time: 00:44)



We will see how we add up all the Bernoulli random variables, and we get the binomial distribution. So, essentially, this is, we have seen the random variable in a Bernoulli trial is independent and identically distributed. So, similarly, in a binomial distribution also, you add this, all of this independent and identically distributed random variables and you get the binomial distribution.

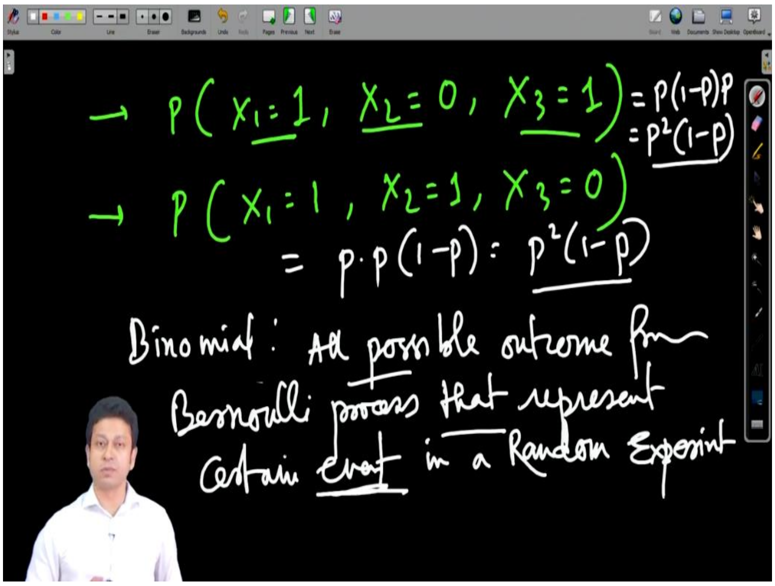
I will give you an example to illustrate what I am trying to say. So, let us say random variables are like 0 head, 1 head, 2 head and so forth. So, these are all possible outcomes that you can get. Now, all of these are independent, and they are all identically distributed, the coin tossing thing is identically distributed. Now, so, these are all IID random variables. And

I need to get the probability of, let us say X 1 being 0 head, X 2 being 1 head and X 3 being another head.

So, depends, like if you want to have 1 head, 2 head or whatever. So, if I want to; let us say this is the probability that I will have exactly 1 head in second trial; or you do not need to write exactly 1, because, in any case, it is going to be either head or tail; and a head in third trial. So, I will write that in this form. So, if I am told that I am tossing the coin 3 times and I want the second and third trial to give me a head each, and the first trial to give me a tail; so, then, I will represent that as this.

So, basically, with different values of X. Now, this is where I talk about; this is essentially representing a Bernoulli process. So, here, I am basically just talking about the independent trials, each of these independent trials, and I am talking about a specific outcome. So, I am talking about specific outcome where I need a 0 at the first trials, 1 at the second trial, third at the second trial. Now, I can actually ask for different types of this process, and they will be different.

(Refer Slide Time: 03:38)



For example, I can ask probability X 1 = 1, X 2 = 0, X 3 = 1. So, essentially, what I am asking is that, to have the same 2 heads but the order has changed. So, it is a different Bernoulli process, but still the outcome is; if I think in terms of the success, so, I have like 2 success here; similarly, I had 2 success here; so, the outcome remains same, but the process is different, because I have; the sequence is different here; I have 0 1 1, and here I have 1 0 1.

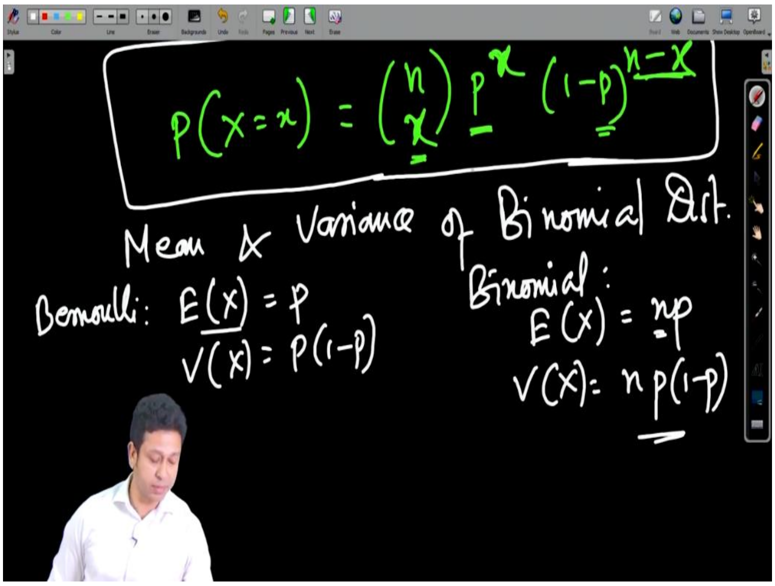
Similarly, I can also think of like this outcome; probability X 1 = 1, X 2 = 1 and X 3 = 0. So, these are all different Bernoulli process. But when I talk about binomial, so, I am actually talking about all these outcomes together. And what do I mean by that? So, when I actually have these outcomes, what is the probability? Let us actually try to calculate that. So, in this case, the probability is going to be, if I have the first one is 0, is 1 - p.

And since this is an independent, events are independent, I am going to have p into p. So, essentially, it is going to be p square into 1 - p. Here, I am going to have similarly, that is p 1 - p into p. So, essentially, it means p square 1 - p. The third one, essentially I am going to have p, a p and a 1 - p. So, it means p square 1 - p. So, all the outcomes, all the different trials that we had are different Bernoulli processes that we have.

So, the probability is going to be same, because I am only talking about 2 successes out of 3 trials. Now, when I talk about binomial, so, it actually represents all the possible outcomes. So, instead of Bernoulli, what binomial is doing is, all possible outcome from Bernoulli process which represent certain event; here, the event is 2 head 1 tail; that represent certain event in all that we were doing in a random experiment.

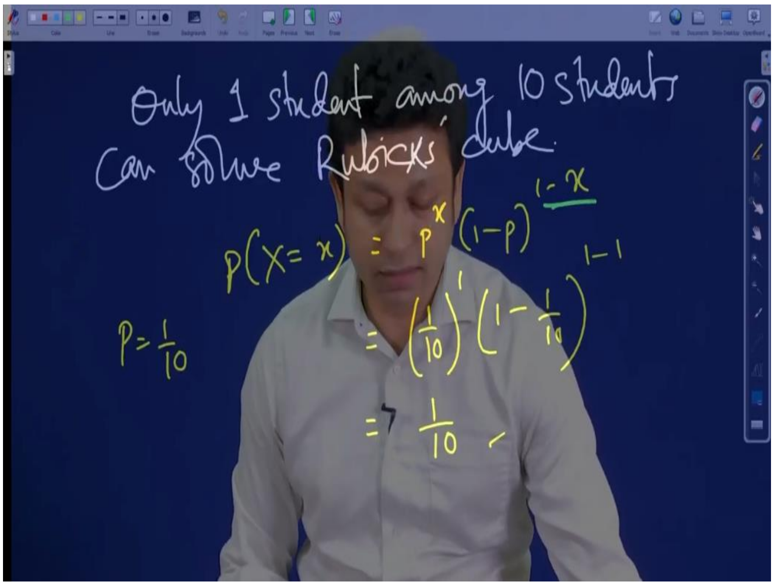
So, we are basically running a random experiment. So, instead of 1 particular outcome, we are talking about all the possible outcomes that represent certain event. The certain event here is 2 head 1 tail. So, the moment I am talking about certain event, it is becoming a binomial distribution. And the moment it is becoming a binomial distribution, you need to add this up.

(Refer Slide Time: 07:02)



And how do I write that? Probability X = x, essentially, out of n trials, you have, let us say r successes. And then you write, probability of r; or if I write x, so, x successes; probability of x, (()) (07:18) of x and 1 - p n - x. So, that is essentially what we are doing for 1 trial in Bernoulli.

(Refer Slide Time: 07:30)

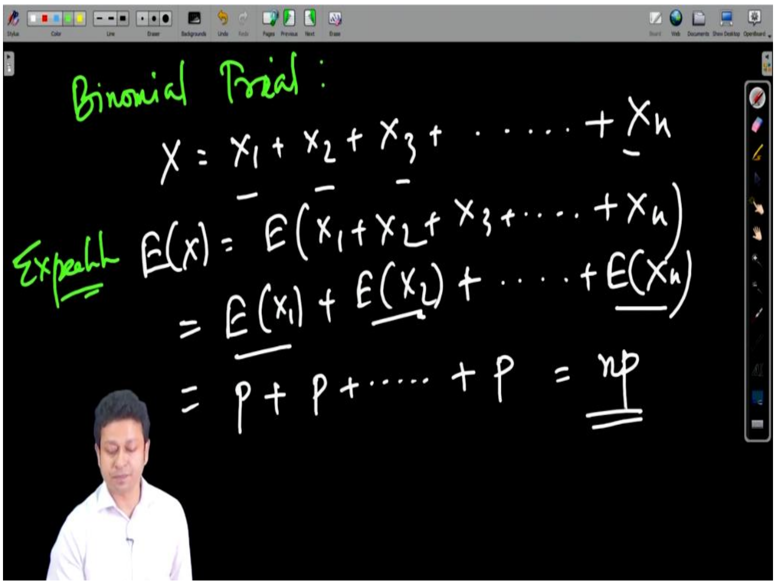


In Bernoulli, we are doing for 1 trial, but here we are doing for n trials. And so, that is why we use n - x, and the number of successes is x. So, that is basically the idea. So, you can also see that, in both the cases, the success; the probability p is going to be fixed, so, it is not changing, just like before. And there, we always talk in terms of success and failures, only thing is that we sum up all the trials in binomial vis-a-vis what you are doing in Burnoulli's.

So, now we will talk about the mean and variance of binomial distribution. So, let us say we know that you basically add up all the trials. So, previously, the expectation of X for a Bernoulli, in a Bernoulli trial is equal to p. And the variance of X in a Bernoulli trial was p into 1 - p; we have seen that. Now, I say that I actually add up all the different trials here; I multiple; Bernoulli trial is actually constituting a binomial distribution.

So, then I can simply write for binomial, expectation of X is going to be np, because I am constituting from n trials. And variance of X is going to be n into p into 1 - p. Now, I have done it simplistically, but actually, I can actually prove that how it is happening from Bernoulli to binomial.

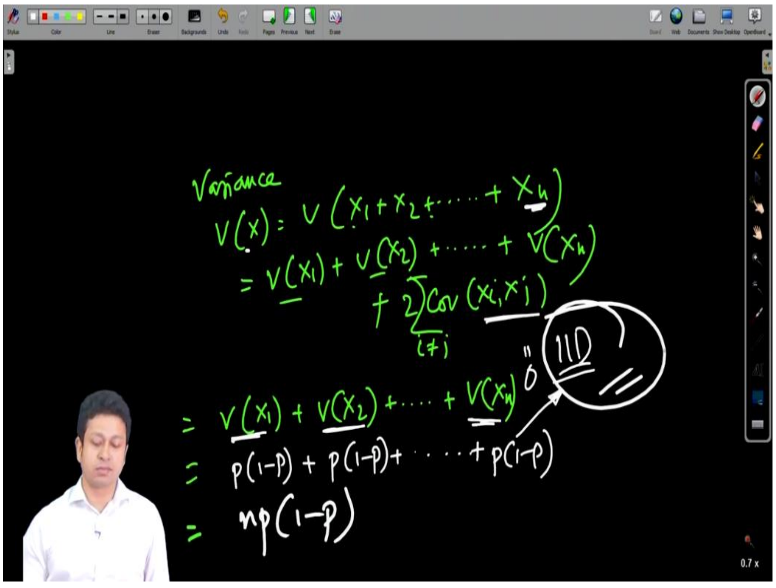
(Refer Slide Time: 09:38)



And I think that is, that would be a good idea to actually do the proof. So, here we can actually write; we should mention what I am; for a binomial trial, let us say X = X 1 + X 2 + X 3 + X n. So, these are all independent Bernoulli trials. The moment I take expectation of X, I take expectation both the sides; X n; and we know from the expectation rule that I can actually write it as X 1 X 2 and expectation of X n.

Now, since these all are constituting independent individual Bernoulli trials, they are all going to have a value of p. So, for n trials, it is going to be np. So, that is how I get the expectation in a binomial distribution. What about the variance? Let us say expectation; this is the expectation we got. What about the variance?

(Refer Slide Time: 11:02)



The variance is going to be, just we will use the same formula, where variance of X is going to be variance of X 1 + X 2 X n. Now, we will use the variance rules. We have already seen the variance rules, and it is going to be V of X 1, V of X 2, V of X n. Now, when we expand these kind of terms, we will also have a covariance term, right? So, we will have 2 of, like, if we multiply, 2 Cov some X i X j, over all i j; when i not equal to j, of course, because this cases are captured here, when i = j.

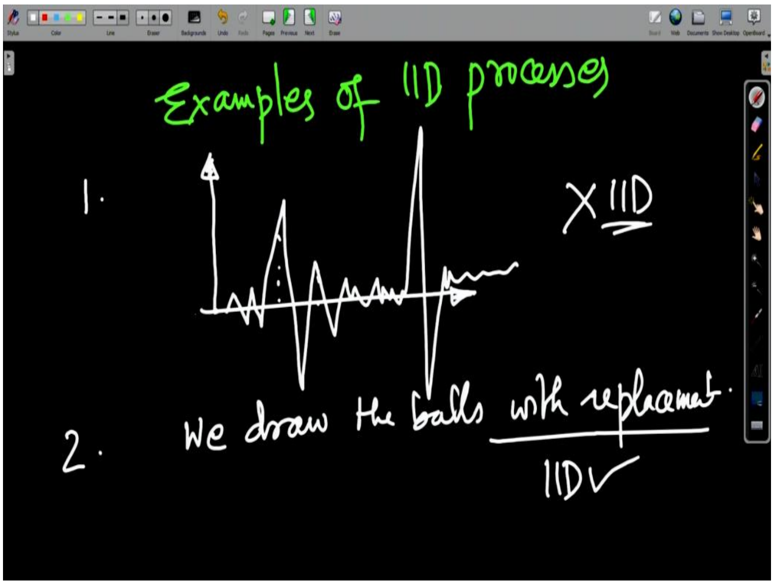
So, when I do that, we have to be mindful that, you remember, we are talking about IID again; coming back to the IID random variable. When I have IID, so, they are independent. So, when 2 events are independent, we have already proved that this term is going to be 0, because they are independent. So, if that is the case, I am going to have; what I am going to have is V X is equal to V X 1 V X 2 upto V X n.

Now, what is V X 1? Now, we again talk about the IID random variable. So, the moment I have IID random variable, all the variances are going to be constant. Now, what is the variance for a binomial distribution? These are all representing each of these Bernoulli trial, right? So, when I have these Bernoulli trials, we are going to have the variance. Variance is going to be p into 1 - p; we have already proved that; p into 1 - p, and so forth, p into 1 - p.

Now, we have all together n cases here, because we have represented X with n different Bernoulli trials. So, if I have this, so, I am going to add this up. And what I am going to get is np into 1 - p. So, that is basically the variance of the binomial distribution. You can also try to expand the probability mass function for binomial distribution, and you can actually try to prove; there is another way of proving it, but I am not getting into that; but you can do that.

So, now, since we have learned the term IID and we have seen the importance of IID when we derive all these different characteristics of these distributions. So, let us delve a little deep into the term IID, and let us see what we can call IID and what may not be an IID, IID process.

(Refer Slide Time: 14:08)



So, let us give some examples. Examples of IID processes: We will talk about different kinds of examples and we will see which one actually is an IID and which one is not an IID. So, let us say, the first case, I draw something here. And I will ask you whether you think that this is IID or not. So, we will think together. So, let us say this is some whatever, let us say stock exchange or whatever.

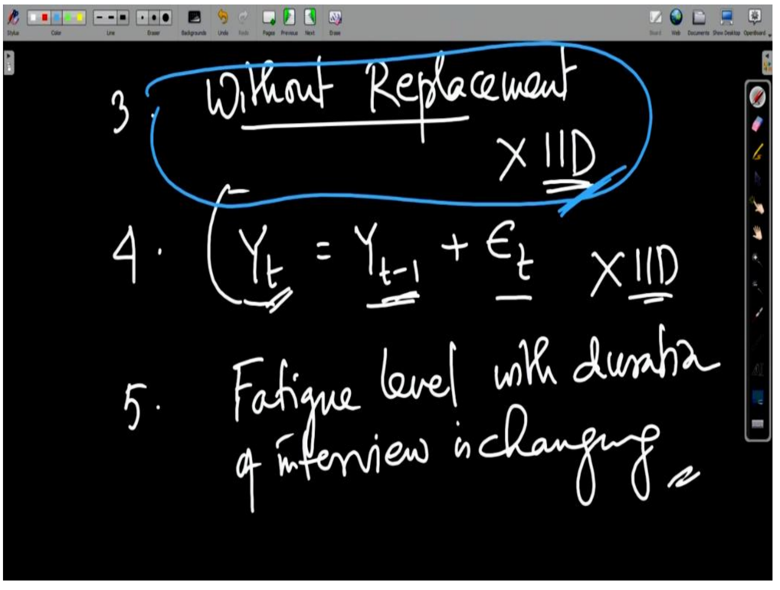
So, I see the variance is like changing like this; something like that. Now, I ask you, just looking at this distribution or the way the numbers are plotted, do you see it is going to be a IID process or a non-IID process? Now, think for a moment, pause for a moment and think, and then I will explain that. Hope you got a chance to think about it. Now, let us try to see what is happening here.

So, perhaps the mean, because it is sort of oscillating almost like equally around the X axis; so, the mean is going to be constant. So, mean is going to be like; yeah, it is going to be constant. Whereas the variance, you see the variance for different points, they are actually varying quite a lot; so, we cannot be sure about whether it is a equal variance process. So, the condition for a process to satisfy the IID process is that, the mean and variance has to be identical for all the different random variables.

So, if the mean and variance has to be constant and the variance is not constant, so, then it is not an IID process, because the variance is not constant. The second problem: let us say there is one urn, and it has like plenty of balls. And we draw the balls with replacement. Now, if you draw balls with replacement, so, what is happening there? What is happening there is, your distribution is remaining constant.

So, of course, since you are drawing from the same population, your mean and variance are going to be constant. And at the same time, since you are drawing from the same population, the number of balls are not changing in the urn, so, they are going to be independent. So, it really does not matter what happened in the previous trial, right? So, that is why, when you are drawing the samples with replacement, it is going to be a IID process. Now, I give you a third example.

(Refer Slide Time: 17:11)



Let us say without replacement, now I talk about without replacement. Now, what is happening here? If I have a same balls in an urn and if I draw some balls, what is happening? Whatever you are drawing in the first trial, some other balls are remaining there. And every time the outcome of the next trial is actually being dependent on the previous trial, because you actually drew some ball out of the urn and you are not returning it back, right?

So, the moment you are doing that, it is no longer a constant probability process, and it is no longer a process where you can say they are actually independent. So, that is the reason it is not an IID. The fourth one: Let us say I write Y t = Y t - 1 plus E t. So, is it an example of IID process? Now, look at it carefully. Y t is actually dependent on the previous; so, here, Y t is basically the outcome of this time period, and Y t - 1 is the outcome of the previous time period.

Now, when your this time period outcome is dependent on the previous time period outcome, so, you cannot call it an IID, because it is dependent on the previous, whatever happened in the previous trial. Let me give some other practical examples. Let us say, it is a pretty common example in fact, that let us say, this is a placement session is going on, and the company people who have come to interview you; actually, the interview is continuing; and let us say fatigue level with duration of interview is changing.

Because, the more they interview a candidate or more number of candidates they interview, what is happening is, they are getting more and more fatigue. And when they are getting more and more fatigued, the decision they make probably is influenced by fatigue. So, when that is happening, we cannot say the process, the selection process is an IID process, because the fatigue level which is crucial in the decision making is actually changing.

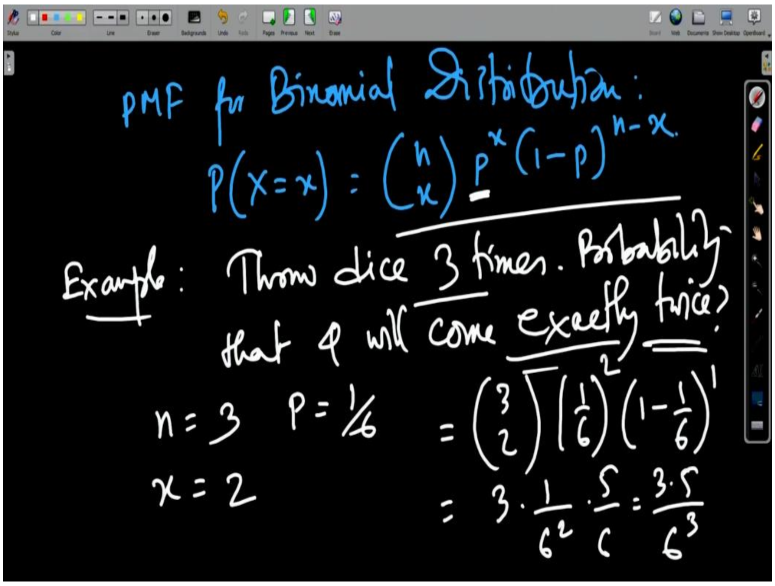
So, it is not a memoryless process, it is a process that is actually influenced by what happened previously. In the same way, and you can say that, depending on how many candidates are selected in the first few trials, whether it can actually influence the next trials, because the company people, they might have a perception about the school or the quality of the rest of the entire batch, or they might actually satisfy all their needs; I mean, they might just not need any more students, even if the student is actually better.

So, whatever happened in the previous trial is actually influencing whatever is going to happen or is happening in the current trial. So, that is what we need to keep in mind that how we have to conceptual, how to think about the IID process. Now, one more point I need to say here is that, when I am talking about this one, without replacement, now, if I have a very large pool of sample, sufficiently large pool of sample; now, what do I mean by very? or what do I mean by sufficient? That is a different question.

But if, let us say it is enough, large enough so that drawing one sample does not really change the constitution of that remaining sample, so, then we can call it an IID process with some error correction. But it is still going to be an IID process, assuming that the drawing of samples or balls is not really going to change the constitution of the remaining balls in that urn.

So, these are the examples I have given just get an idea of the IID process. Now, we spoke about the probability mass function for a binomial distribution. Now, we will see some application of the probability mass function of the binomial distribution.

(Refer Slide Time: 21:37)

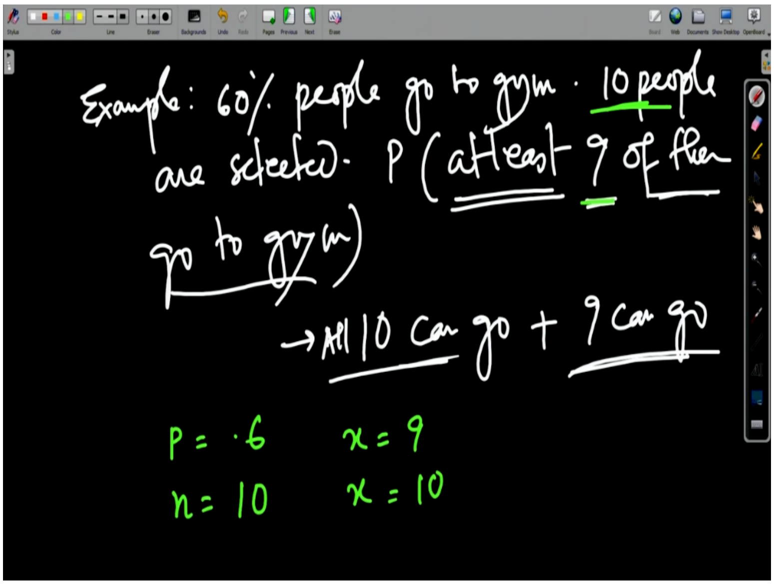


So, we know the PMF for binomial distribution. PMF for binomial distribution is going to be; we know probably X = x is n c x with n trials; n - x, right? Now, let us use this formula for an example. Let us say I draw or I throw a dice 3 times. What is the probability that 4 will come exactly twice? That is the question. To do that, let us first what do we need here to actually create a binomial distribution here?

If we just want to use a PMF, so, what do we need here? We need the probability, the value of p, the constant value of p. The n is 3; your total number of trial is 3. And your success, the success x = 2. You want to have like 2 success. And your probability p is; since you are throwing a dice, there are 6 possible outcomes; so, probability is 1 by 6. And since you want exactly 2 dice, so, what you get is basically n is 3 2 probability 1 by 6 to the power x is 2 and 1 minus 1 by 6 to the power 1.

So, that means, here 3 c 2 is 3 into 1 by 6 square into 5 by 6. So, that means, 3 into 5 by 6 cube. So, that is your answer. So, that is how you use your binomial distribution to get the probabilities. Now, let us say we do another example. Here, we said that exactly twice. Now, it can also say that less than twice or more than twice. So, then, what do you have to do? We have to actually add up the probabilities for each of these individual values. So, let me illustrate that. Let us do another example.

(Refer Slide Time: 24:37)

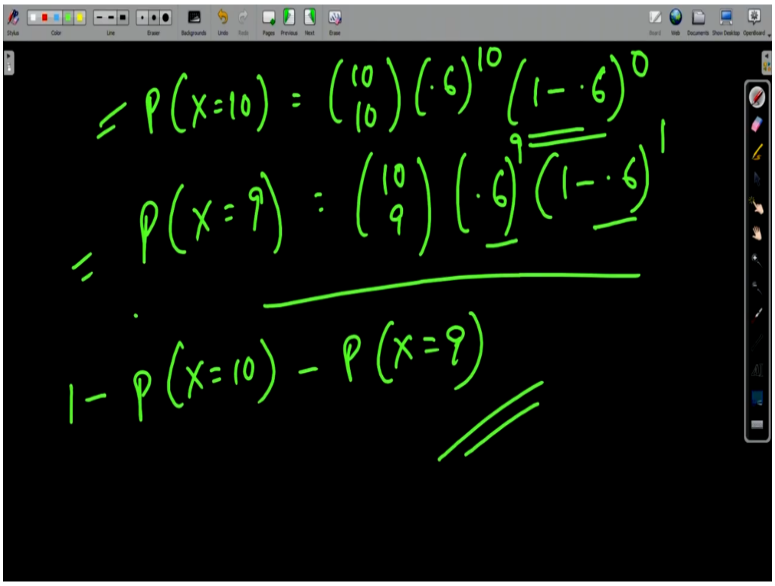


Let us say 60% people in a population go to gym. 10 people are selected; just they are randomly selected. And let us say you ask that, at least; let us say probability that at least 9 of them go to gym. So, here, instead of exactly I am using the word at least. So, if that is the case, all I have to do is, it means, I have to see what are the events that is constituting the random variable here.

So, the events are that the person can go, all 10 can go or plus 9 can go. So, it could be 10 people, it could be 9 people, but if it is less than 9 people, then it is not constituting the random variable that we are looking for. So, if all 10 can go and all 9 can go, so, the probabilities of these 2, you have to add them up. So, for this, you use the, again; so, basically, we will be using the PMF for the binomial distribution.

And so, to use a PMF, you need to know few things. So, here, the p is constant, which is 60% people go to gym; so, 0.6. And 10 people, so, there are n trials, n is 10. And your success x is equal to once it is 9, and x is equal to 10. So, there are 2 possible cases. So, all you have to do, you have to add the probabilities for these 2 possible cases. So, how do we do that?

(Refer Slide Time: 26:44)



It is very simple. You do probability X = 10, is equal to, so, 10 c 10; probability of success 0.6 to the power 10, because you want all the time to go to the gym; so, it is 0.6 to the power 10. And 1 - 0.6 to the power of 0, because nobody goes to gym, this part. And probability X = 9 is going to be, 10 c 9 0.6 to the power 9 1 - 0.6 to the power 1. So, these are the success, these are the failure.

So, what you do for your event to come true, you basically, for your random variable that at least 9 people go to gym, you add these things up. Now, if I want that less than 9 people go to gym, what is the; if you change the random variable that what is the probability that actually less than 9 people are going to gym, so, what do you do? You basically subtract this from 1. So, this is okay, you can do that, when you have to calculate less number of these kind of events.

But if you have to calculate more number of events. So, let us say you have 50 individuals and you want to see what happens if you want to calculate the probability for 35 people, that at least 35 people should go to gym; so, then you have to subtract all of; you have to calculate the probability for all the 15 events, like which are above 35; 36, 37, 38. So, you have to get individual probabilities; you have to add them up; you have to subtract from 1.

So, that is a bit difficult task. And that is where we want something different. And that is where we will be seeing in the next lecture, how we approximate to normal distribution; from a binomial distribution, how we actually end up approximating to normal distribution. And that is really handy, useful, because it reduces the expense of calculation; you really do not have to calculate so much.

So, in the next lecture, we are going to see a normal approximation of the binomial distribution. And with this, we will end the introductory lecture on binomial distribution. Thank you.