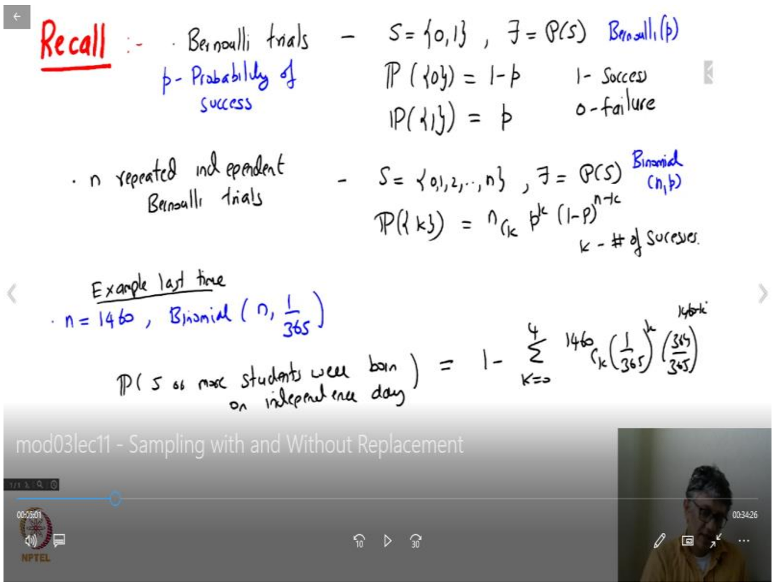
Introduction to Probability – With Examples Using R
Professor Siva Athreya
Theoretical Statistics & Mathematics Division
Indian Statistical Institute, Bangalore
Lecture 11
Sampling with and Without Replacement

(Refer Slide Time: 0:26)



For today's class, we will just recall little bit what we did last time. Recall. From last time. So, what we did last time was the following; we discussed Bernoulli trials. So, these are repeated experiments. We could think of them as experiments with sample space 0 or 1, and, and the events is (())(00:57) power set of S, there will be all possible subsets.

And the probability we put was the probability of every outcome was, was 1 minus p for a failure, and we thought of success as P. So, 1 was thought of as success in the experiment and 0 was thought as a failure. This is what we discussed as I said earlier. And when we performed this experiment n times, n repeated Bernoulli trials, so independent Bernoulli trials.

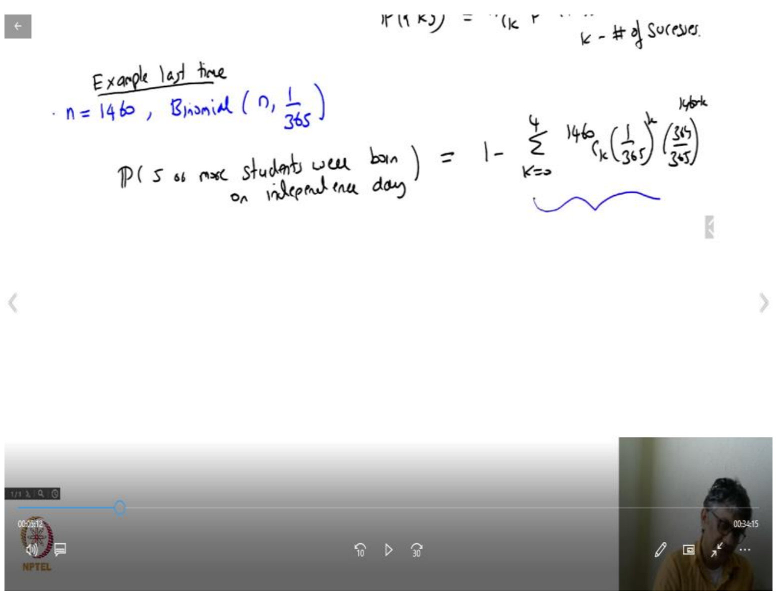
Then we did this. Then what happened was the sample space became 0, 1 up to n. And F again was the event space is again the power set of S with all possible subsets. And then, we noticed that the chance of K we computed was going to be n choose K, p power K, 1 minus p power n minus K. And, here again K is the number of successes in n trials.

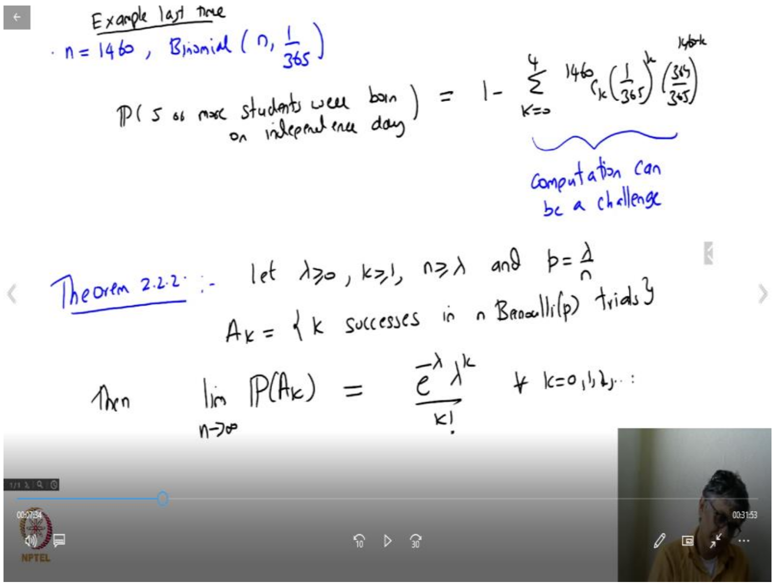
And, here, in this whole experiment, the idea was that P was the probability of success. And these were called Bernoulli p distribution. This was called the Binomial n, p distribution. So, that is where we were at last time. And then, I had done an example, let me show you an example, one second.

The example I had done was at this small college, maybe, size only (thou), n was 1,000 and I wanted to compute the probability that there were at least 4 successes in that. So, this is where we were at. So, I had an example of n equal to, I think it is (4) 1460 with the calculation, and I had to, I had a binomial n, 1 over 365, so this P is 1 over 365. So, n is, n, 1 over 365 experiment. And this was your example last time. Let me just recall the example, so, example.

And the example was that we had to compute the chance that, that 5 or more students were born on Independence Day. And this came out to be 1 minus the chance from K equal to 0 to 4, 1460 choose k, 1 over 365 to the power K times 364 by 365 to the power 1460 minus K. And then we add here, I had explained to you that computing these probabilities can be quite hard, can be quite challenging.

(Refer Slide Time: 05:11)



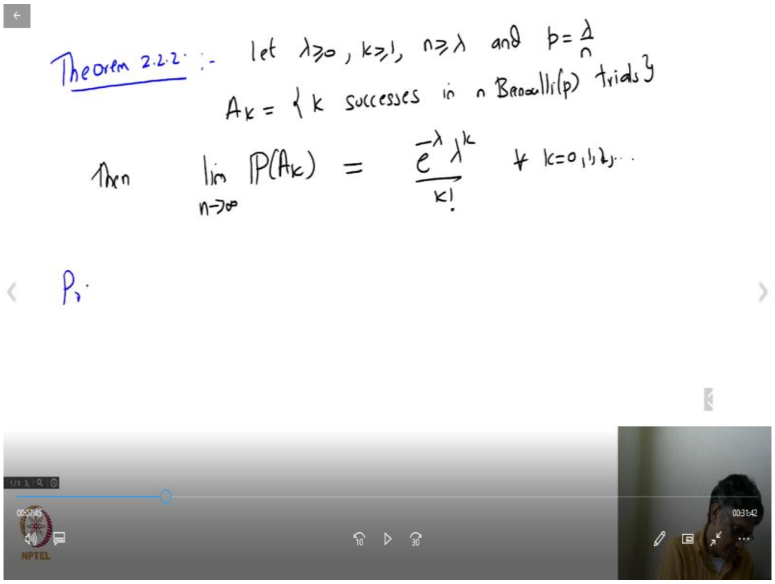


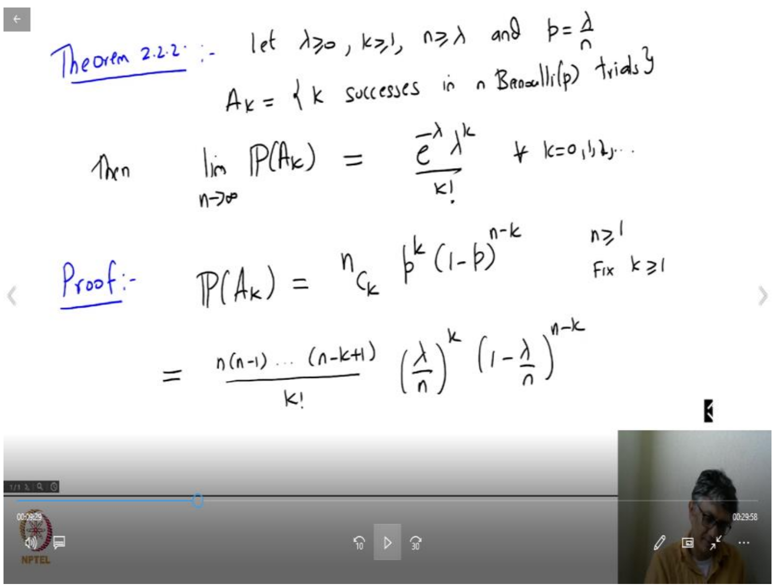
So, in this case, it is easy. So, sometimes computation can be a challenge. That is the, that is one thing which (())(05:30). That is where we were. So, today, I am going to like start off from this theorem that I wanted to show last time. So, let me just write the theorem down. So, theorem that I wanted to show last time was, theorem was the following; this theorem was the Theorem of 2.2.2. This is what I wanted to show today. I will try and show today.

The following, we let lambda be is on negative, K be bigger than equal to 1, n be bigger than equal to lambda, and p be equal to lambda by n. So, they have defined A sub K as K successes n Bernoulli, and n Bernoulli p trials. So, now once you do this, then what you can show is the following; you try to compute the probability of AK.

So, you want to compute a property, there are K successes in n trials. We know a formula for that. And then the key thing is the result is that as n goes to infinity, so as of now, it is an artificial concept because this AK depends on (())(07:16). So, n to infinity, this converges to a number so e to the minus lambda, lambda power K by K factorial. This is true for all K equal to 0, 1, 2, etcetera. So, take a fixed K (())(07:33). So, that is the theorem.

(Refer Slide Time: 07:45)

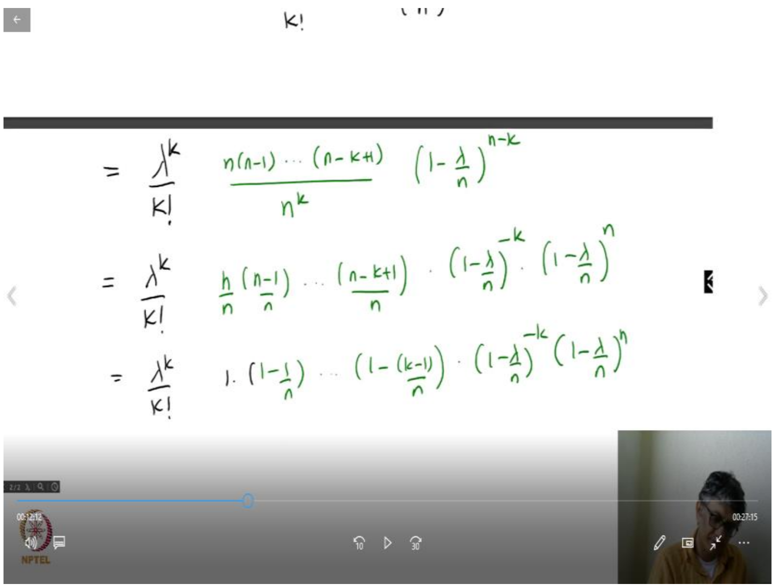




So, we will try and prove that today. The proof is not so hard. So, the proof is the following that you have to just write the formula down and try and see if the, if the mathematics works out. So, what do you get? You, you first write down probability of AK. So, probability of AK, we know what it is going to be. It is just n choose K, p power K, and 1 minus p to the power n minus K, for a fixed K. So, take n bigger than equal to 1, n bigger than equal to, sorry, 1. And, so you fix the K. So, if k is some number, you choose an n larger than K, so let us just fix, so if n is going to infinity if we can choose n large enough.

Once you do this, then all is just, this is just, this is just a probability, this is a computation in which you have to just keep track of the terms and it will all get (())(08:56) problem. So, n choose K, you can write it as n into n minus 1, all the way up to n minus K plus 1, the whole thing divided by K factorial. That is where we were at. And then, I have P is what, lambda by n. So, I have a lambda power by n, the whole power K, and I have 1 minus lambda by n, the whole to the power n minus K is (())(09:28). And that is where we are at. So, now I have to sort of understand how to play with these terms.

(Refer Slide Time: 09:47)



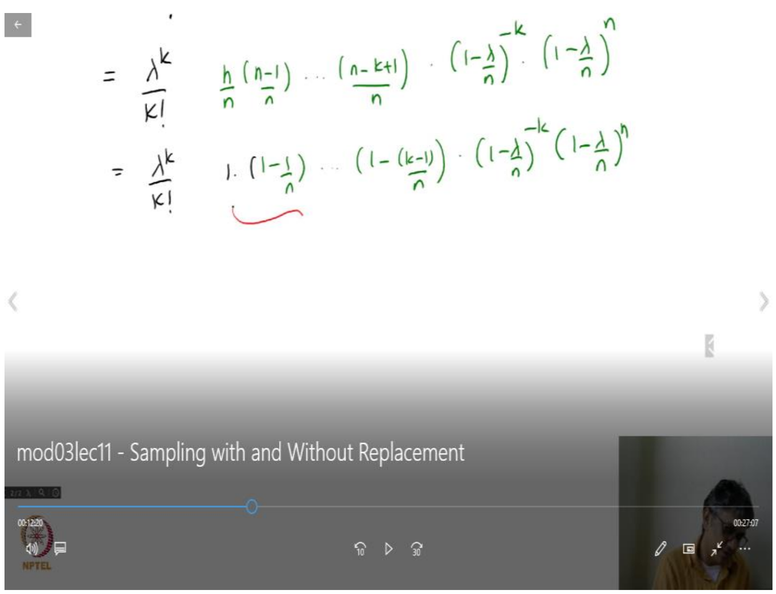
So, what I do is I remove all the terms that do not depend on n. So, from the top 1, I have, I have, so, let me write it in blue, green, let us say. So, the first term I have is lambda power K, it does not depend on n. And K factorial does not depend on n. So, these two terms do not depend on n. Then what depends on n, so, maybe I will write it in reverse, maybe I will write the term that do not depend on n in black is lambda power K by K factorial, it does not depend on n.

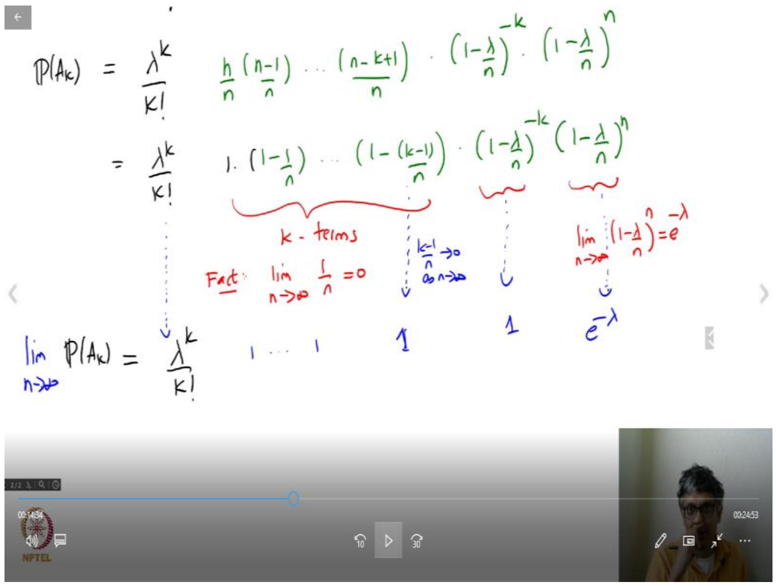
Then everybody else has a term n in it, so here this will be n into n minus 1 all the way till n minus K plus 1. These terms how (())(10:24). I have an n power K at the bottom from this first, second term. And this one depends on, and again, I will write it as 1 minus lambda by n to the power n minus K. So, if n goes to infinity, I have to keep track of what is happening to these terms. So, now what I do is I do another small trick. So, I do this lambda power K by K factorial on the outside.

And here, I will push, there are K n here, so I will push each of them to each term here, so what I will do is I will just get 1, I will get 1, and n minus 1 by n, I will think of it as n minus 1 by n all the way till n minus K plus 1 by n. I get a 1 here because it is n by n, maybe I will write it down, n by n, I will do that. And here, I will split this up into 1 minus lambda by n to the power minus K. So, that is just a fixed number, fixed number. And then, I will have a 1 minus lambda by n to the power n. So, that is how I am just splitting it up. This you can further split up.

So, I will write it as lambda bar K by K factorial. And these terms are fixed k terms, so it is 1 into 1 minus, there is no n in that. So, it is 1 minus 1 by n all the way till 1 minus of K plus 1 by n, this is n minus K. So, I shall write it properly, 1 minus, like this, and 1 minus K minus 1 by n. And then here, I have 1 minus lambda by n to the power minus K, and I have 1 minus lambda by n to the power n. So, that is straight forward.

(Refer Slide Time: 12:20)





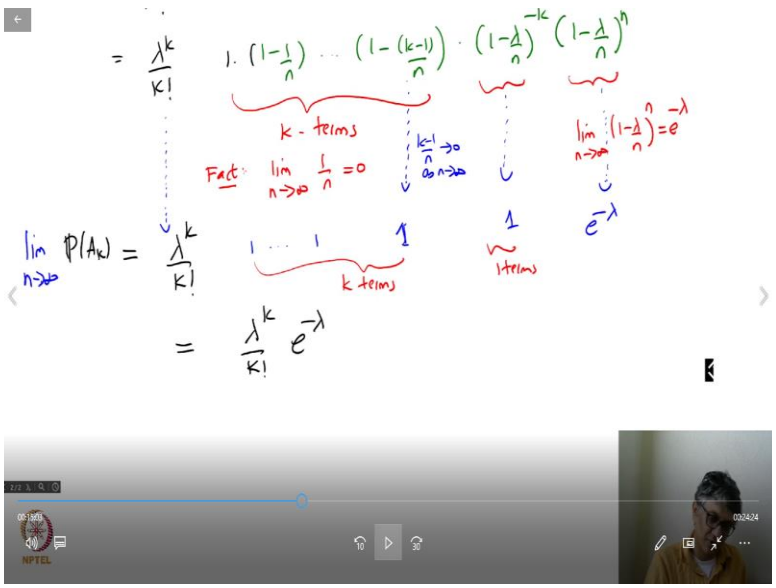
Now, what happens is you notice, this is just a fixed number of K terms. This is just K terms that are fixed. And here is this one term so, we will come back to this, and we will treat this term separately. So, as n goes to infinity, each of these terms, we will have to work it out, it is not that hard. So, we know the fact that the limit as n goes to infinity, 1 over n is equal to 0. This is a fact figure. We also know this fact, here it is again, it is a fixed number of K minus K products.

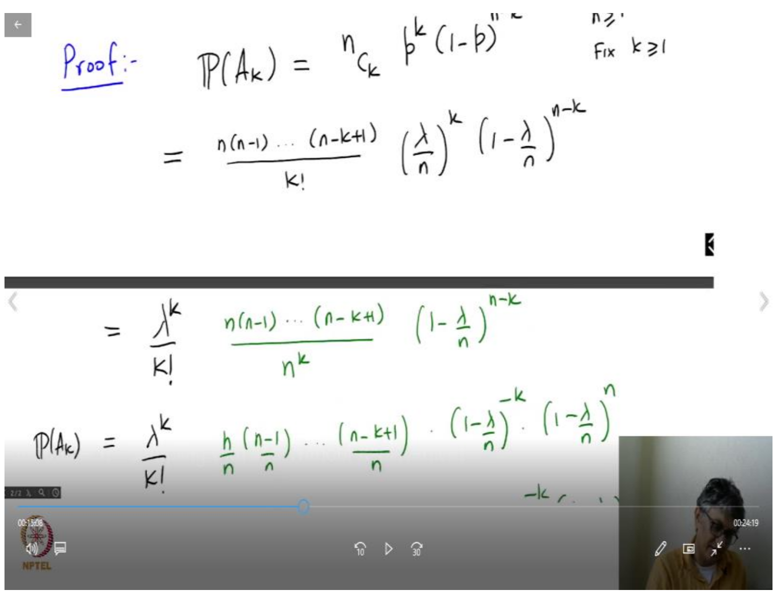
Here also we know the fact that limit as n tends to infinity, 1 minus lambda by n to the power n is e to the minus lambda. So, these two facts we will use. So, once we have these two facts, it should be fairly easy. These are K terms. So, this person stays the same as n goes to infinity. Therefore, let me write down, on this side, let me write down, limit, and so all this is probability AK.

So, limit as n tend s to infinity, probability of AK, is going to be equal to what? So, let me write the limit in blue so it is clear what the limit is about, so limit in blue. So, limit is, limit as n tends to infinity, probability AK. This person right here stays constant. So, this person is going to come down at lambda power K by K factorial, that person? So, here, there are K terms.

But each of them, each of them is going to go to 1 because 1 over n is going to go to 0 and K is fixed, so K minus 1 by n goes to 0, goes to 0 as well. So, K minus 1 by n will also go to 0 as n goes to infinity for all these terms. So, this will all go to 1, so it will all go to 1, so, it will go to 1, and this is multiplying 1 K times. And here again, I have lambda as fixed and K is fixed, so this will also go to 1. And here, by this fact, I know that this person is going to go e to the minus lambda.

(Refer Slide Time: 14:41)

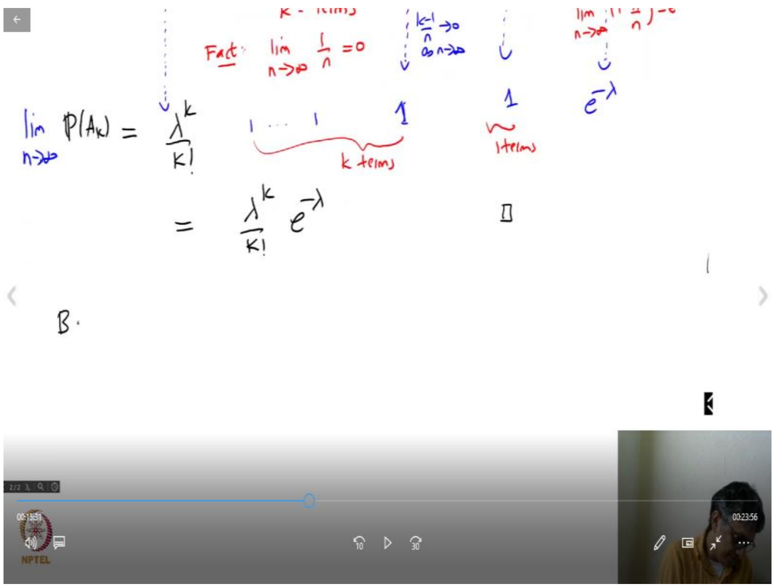


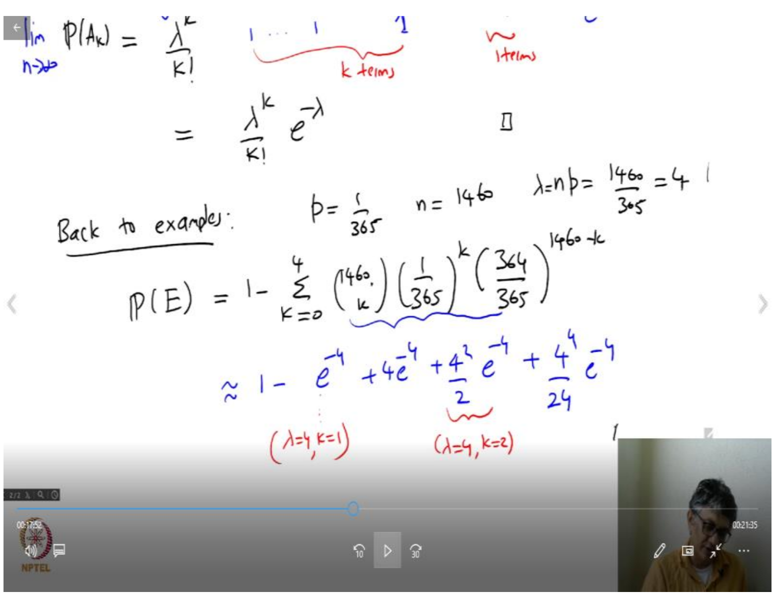


So, now if you, if you just rework this whole thing out, what do you get? You get limit as n tends to infinity is going to be lambda power K by K factorial. I am multiplying only K 1s here so this is K terms, this is K terms. This is rather one term, this is one term.

So, the whole thing becomes e to the minus lambda, (lam) e to the minus lambda. So, that is the answer. So, so we have shown that probability of AK as n goes to infinity is going to be equal to e to the minus lambda times lambda power K, by K factorial. So, it turns out that this is sort of a max (())(15:18)distribution, so.

(Refer Slide Time: 15:25)





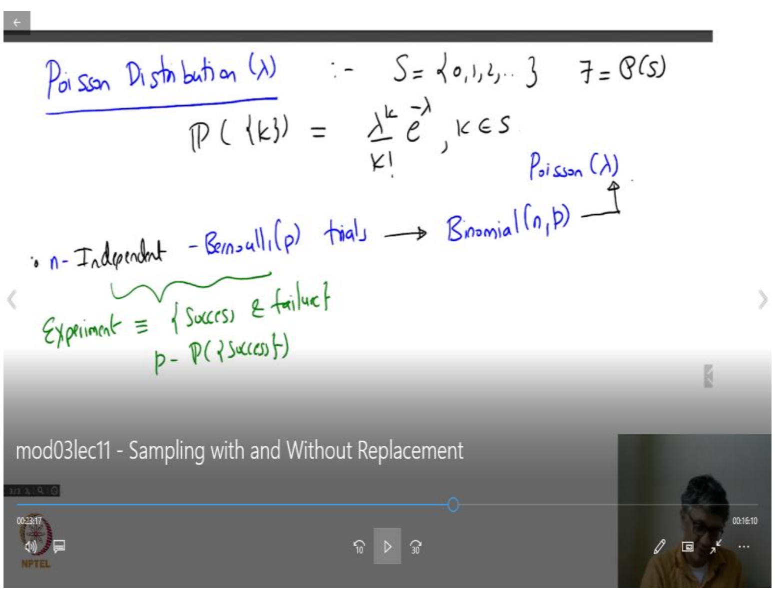
So, how could you use it in an example? So, we will come back to this. This is our answer. We have proved the result. So, if you go back to the example, so back to example. So, example, we have, we had p equal to 1 over 365. We had n equal to 1 over 460. So, n times p is lambda, we will just go to the, something like 1460 by 365, which is like 4 I think if you multiply it out.

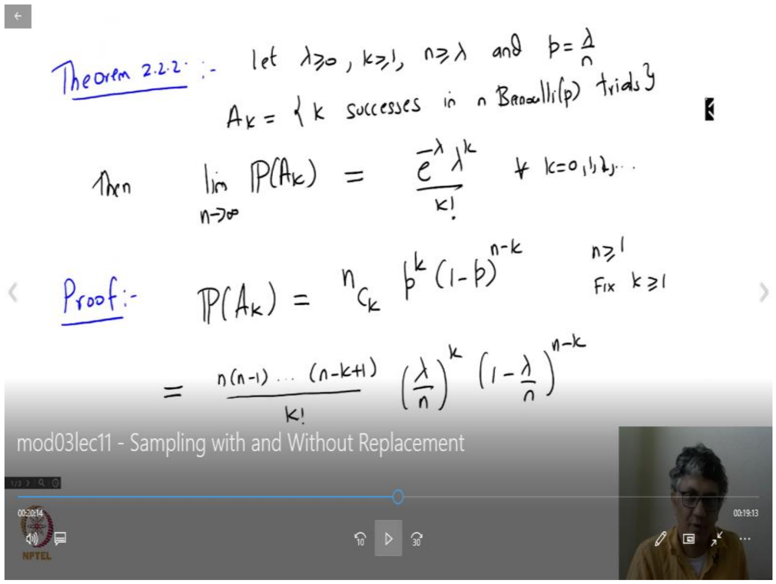
Then, you know that the chance of the probability the, of the event that we are discussing, let us call that event as E, as is going to be 1 minus the chance from K equal to 0 to 4, 4, 14 choose 460 choose K. And then we add 1 over 365 to the power K, and then we add 364 by 365 to the power 1460 minus K. That is the expression we had.

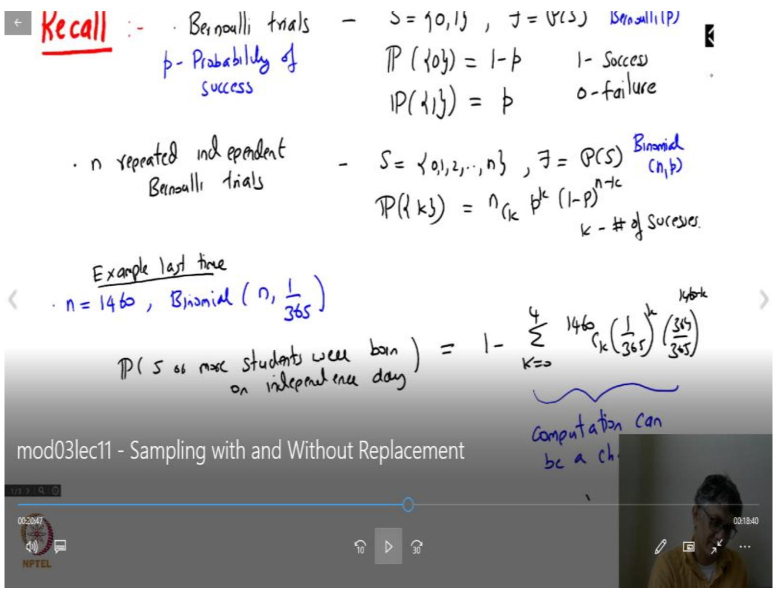
And, so now what we do is, we use the theorem, we know that as n goes to infinity, all these terms are converging nicely, so, think of this as 1 minus this each of the terms was like an AK. So, it is going to be e to the minus 4 for the first guy second guy is 4 e to the minus 4 plus 4 square by 2 e to the minus 4 plus 4 to the 4 by 24 e to the minus 4.

So, I have used each of these terms, each of these terms converts to lambda power K by K factorial times e to the minus lambda. So, this term would correspond to, for example, lambda is equal to 4 and K is equal to 1, and this term, and this term again is corresponds to lambda equal to 4 and K equal to 4, and so on forth, and so on and so forth (())(17:44). And this term, this, this can be computed. You can compute this whole thing. This is one way you can use the theorem to compute things. But it turns out that this is a nice distribution. Let me write this down.

(Refer Slide Time: 18:11)







So, this comes out to be, there is no, there is a way to model the distribution, but as of now we could think of it as the, the Poisson distribution. Let me write that in blue so that, all headings I will remove. So, it is a Poisson distribution, parameter lambda. So, one could think of this as an experiment, there is no experiment in our head. It says a limit of n goes to infinity of n Bernoulli trials. But you could think of this as S as 0, 1, 2 up to n, up till infinity, (())(18:48), sorry about that. Then you take F to be the power set of S. And then you take the probability of an outcome K to be equal to lambda power K by K factorial or KS. And, that defines the problem, e to the minus lambda.

So, this is this is what is called the Poisson distribution and it is a very powerful tool, and sometimes people just call it as a Poisson lambda. So, there are, this is sort of a, a, this, this can be, this also arises naturally in an experiment but as of now we will think of it as a, as a, comes out from a limit of n tends to infinity of Bernoulli Trials.

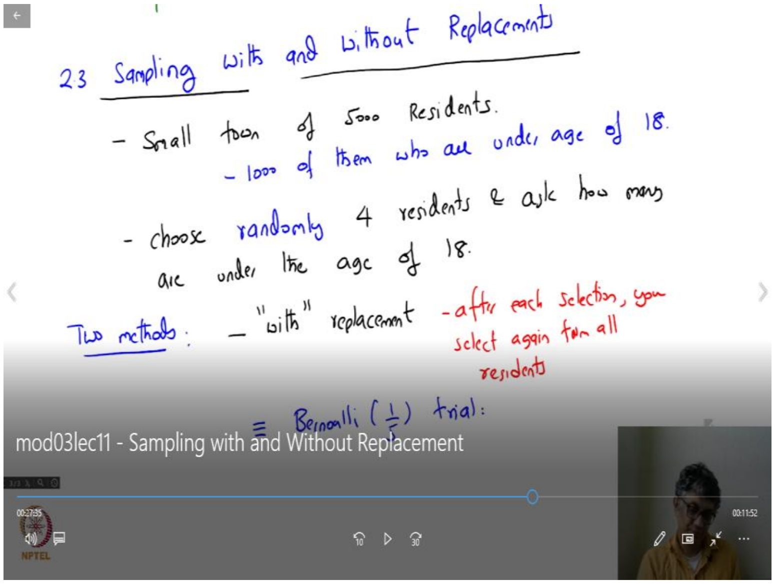
And that is primarily because of the theorem we have showed, we have showed this theorem that, that if you add k independent trials, then the number of successes in k trials is going to be converging to e to the minus lambda by K factorial times lambda (())(20:28). It is, so this is popularly referred to as the approximation of binomial Poisson.

And the main way I am motivated is what I said it is a computational challenge. That is how I, I am motivated, so, (())(20:48). Very good. So now this is, this whole, this whole idea is, is for independent trials. So, now I would like to shift little bit and understand what happens if the trials are dependent in some way. Suppose I have a, I, so now if I, if I keep doing n independent trials, then structure is kind of clear.

So, let us just wrap up a little bit so, yeah, so what all have we done so far? We have sort of developed the structure in the following way; so, we have n independent, maybe independent I will write it in different color so it is all clear. So, we have n independent Bernoulli P trials. We know what they are. So, they are, they are experiments with, they are experiments with, where, each experiment has only two outcomes, each experiment has only two outcomes: success and failure. And probability of success is P, and that is how we think of this experiment.

Of course, we thought of success as 0s or 1s but it is the same thing. It is just notational convenience. So, experiment. And then we said we proceeded to, once we had this, we went to, we went to binomial n, p. We learnt about Bernoulli distribution and then from binomial, we, we took n going to infinity and relied up to Poisson. So, then we, I am just, I will not, (())(23:08) e lambda to the Poisson. So, that is where we were. That is the structure we have been following so far.

(Refer Slide Time: 23:24)



So, now I want to discuss this next topic. So, let us say 2.3, which is sampling with and without replacement. So, the motivation is from I want to get to, dependent Bernoulli trials (())(23:41). Let me try and motivate this. So, here the, key stumbling block is the following, so, so let us say you are a, let us imagine a situation where, and you are in a small town of let us say 5,000 residents.

So, there are 5,000 people. And, let us say there are 1,000 of them who are under 18, under the age of 18. And the rest of them are people who are above 18. And then, what I want to do is I want to choose, I want to choose four, I want to choose the right, let me write this down, in a, in a sort of a way, I want to choose randomly. I have not told you what randomly means.

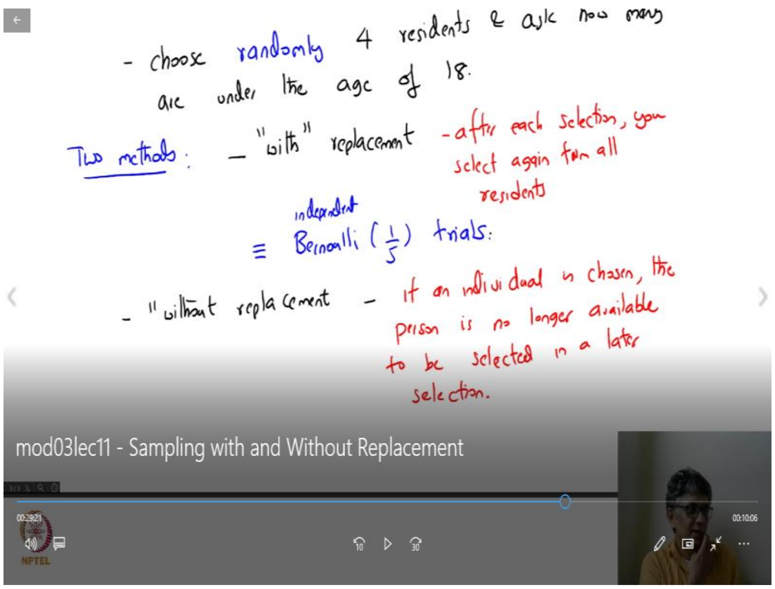
4 residents, 4 visitors, and ask how many are under the age of 18? Is the question clear? The question is that I, I pick 4 people and I want to understand how many of them are under 18. That is the question. So, now of course we have to understand what randomly means. I have not specifically said that.

So, one thing is that you could do two ways. You could pick four people in two different ways: one is you could pick one person, put the person back in, and again another person, put the person back in, and so on and so forth. So, that is called with replacement. So, there are two ways of doing it. So, there are two ways of doing it, there are two methods. There are two methods of doing this.

So, method 1 is, is with replacement. So, in this, in this what do we do? We, we choose everybody and then put them back, so, after each selection, after each selection, that is, after each selection, you select again from the whole population, from all residents. So, this is like, that means each time, every resident has a chance of being picked.

So, you could have multiple people, you have same people again and again. So, this is like a, if you think a little bit, this is like a Bernoulli. (Ou) Our chance of picking is 1,000 by 5,000, so it is 1 by 5 experiment, just think a little bit. This is like a Bernoulli 1, 1 5 th trial. This is one method.

(Refer Slide Time: 27:42)

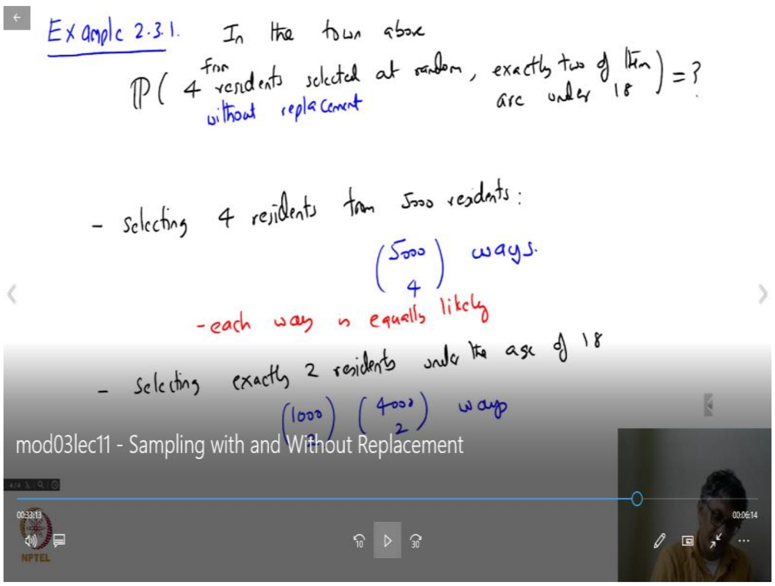


Another method is, is to pick without replacement. That means once you pick a person, you leave the person out. You do not, you do not, you choose the next 3 outside this person. So, it is here, if an individual is chosen, is chosen and, is chosen, the person is no longer available to be chosen, person is no longer available to be picked in the next selection, only this lot, available to be, to be selected, let me write here, selected, in a later selection. And that is not, they are two different ones. So, here, here, you have to be a little careful also, here, there are, there are two different schemes now with and with the (repla) without replacement.

And with replacement, somehow, you, you are lucky mathematically you are, you get trapped (())(29:06) because your independent Bernoulli trials, Bernoulli trials, independent, you have independent Bernoulli 5 trials, 4 or 5 trials. And then, in the second one, you cannot view them independently because the moment you choose one individual, that individual is no longer available for selection.

So, the, so the idea is that you have to sort of understand that they are dependent trials, they are not independent anymore. So, let us just try and see how to sort of solve these problems, so now, in both these situations here. So, let us ask (a), in with the first case we know how to solve it. It is, excuse me, let us try and do it in second case.

(Refer Slide Time: 30:00)



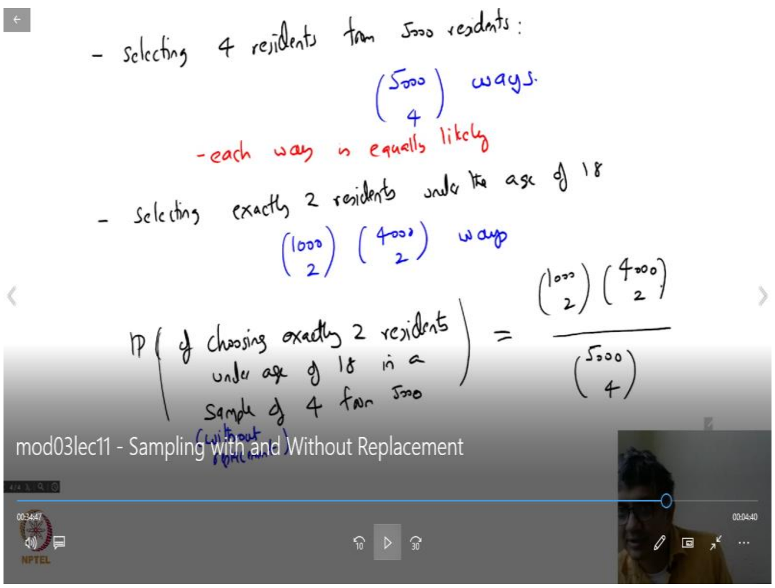
Let us try and do it, let us try and, let us try and understand, let us try and understand the examples. So, let us say example 2.3.1. Let us try and understand this. Let us try and see. So, in the town, in the town above, let us say or what is the probability that four residents selected at random, let us use the, let us use the without replacement.

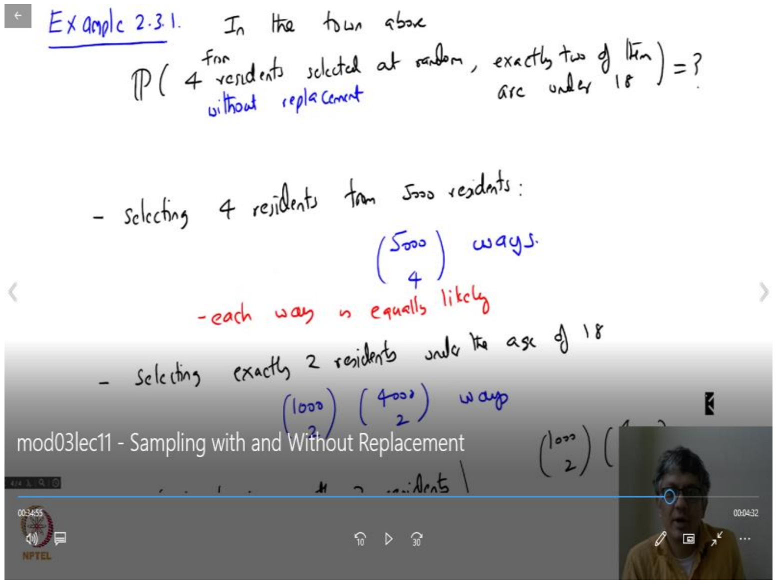
So, let us write down, from so from, four residents without replacement, exactly, let us say two of them are under 18. How do I compute this? So, here the, experiment has, cannot be thought of independent kind of experiments, but you can do it in, in another way. So, let us see. So, let us do the following. So, how do I choose 4 residents? Shall I pick 4 residents from 5,000, from 5,000 residents?

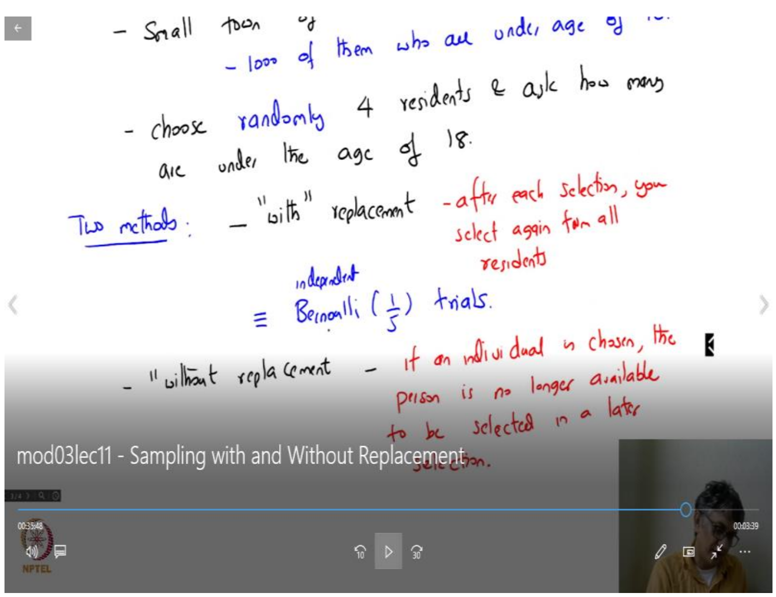
So, that is like choosing four people from 5,000. So, can be done in, the first can be done in 5,000 choose 4 residents. We have so many ways of choosing 4 residents from 5,000 people. So, now each, so when we know, since randomly, each, each, each, is each way is equally likely. Each way is equally likely. So, what we get is the probability of choosing just two under 18 so we just say we look at the, the selecting exactly two residents under the age of 18.

What, how do you do that? That means you have to pick two from 1,000 people. So, what, we had, we had 5,000 and 1,000, right, 5,000, so we had to pick 2,000 people. So, so, can we know how many ways you pick 2 of them from 1,000, and then you pick the rest from the rest 4,000. So, these are number of ways in which you can pick, two people, exactly two people under age of 18.

(Refer Slide Time: 33:26)







Again, since each way is equally likely, you can go back and say the probability of, of choosing, I am going to write down properly, is this, probability of, of choosing exactly two residents under the age of 18 from a sample of 4 in a town of, in a sample of 4, I am sorry, in a sample of 4 from 5,000, and the sample is without replacement, that is, once I pick a person, I cannot put the person back in.

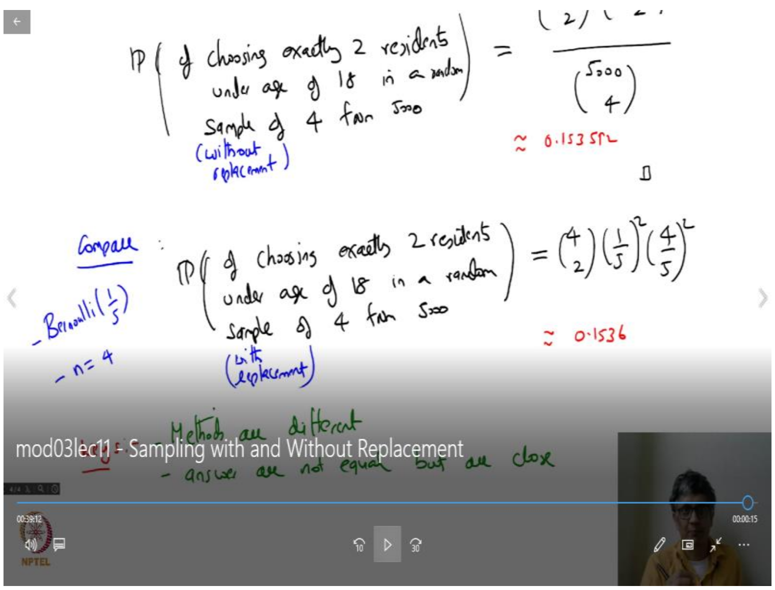
This probability is going to be equal to like we did, it is just divided. So, bottom is, since there are so many ways of choosing them, it is 5,000 choose 4, and the top we put as 1,000 choose 2, and 4,000 choose 2. That is one way of calculating this problem. Let me go up to the, is that clear, the idea is clear?

So, you, you sort over the idea that you wanted to pick 4 people, and you want exactly 2 of them to be under 18, and (())(35:01), you look at it, you first say, fine, I can pick four people from 5,000 and 5,000 choose 4 ways. I can pick two residents under the age of 18 in 1,000 choose 2 at 4,000 choose 2 ways.

The probability is going to be the total number of ways in choosing them, choosing only two under the age of 18 divided by total number of ways of choosing four people because each way is equally likely and that is how I got it. So, I mean I will just write the word randomly, random somewhere. That is one way of completing it.

So, this, this contrast to the fact that you should think of this, this, just think a little bit how do, if you do this calculation right here, it is, it is again, it is a, in this calculation, if we do this calculation with independent, there are different answers like a binomial A4, A2, and here you get exactly something else. So, be careful how you write this answer.

(Refer Slide Time: 36:12)



So, let me just compare, so just compare with the following idea that the probability of choosing exactly two residents under the age of 18 in a random sample of four from 5,000, and here if I say with replacement, that means I have independent Bernoulli trials. So, independent Bernoulli trials means, that means the experiment is just a, a Bernoulli, so, let me write (())(37:03), the experiment here is just Bernoulli 1 over 5 trials, and there are, that is Bernoulli, and then the number of trials is 4.

So, what I will get is, I will get 4 choose 2, and I will have P is 1 over 5, whole square and 4 5th whole square. That is a, so that is a little different answer than the one I got above. So, with and without replacement, has two different answers. So, the 1 K is the round of successes. In both cases, the round of successes is 2 but if you do independent trials, then you get one answer, dependent trial, you get one answer.

But if you just do the calculations, now if you do calculations, then this comes out to be, I think this comes out to be 0.15392, 3592 and this comes out to be 0.1536. So, they are not too far apart, but the answers are too, are not equal but they are very close. So, the, so what, so the key findings or the keys are, one is the models are different; that is one thing.

The answers are not equal, but are close. 3 things are sort of, sort of, so some models are just same and methods are different also. So, one is with replacement, without replacement. The answers are not the same, but somehow, for large n, it seems like the answers are close. We will try and sort of understand this concept.