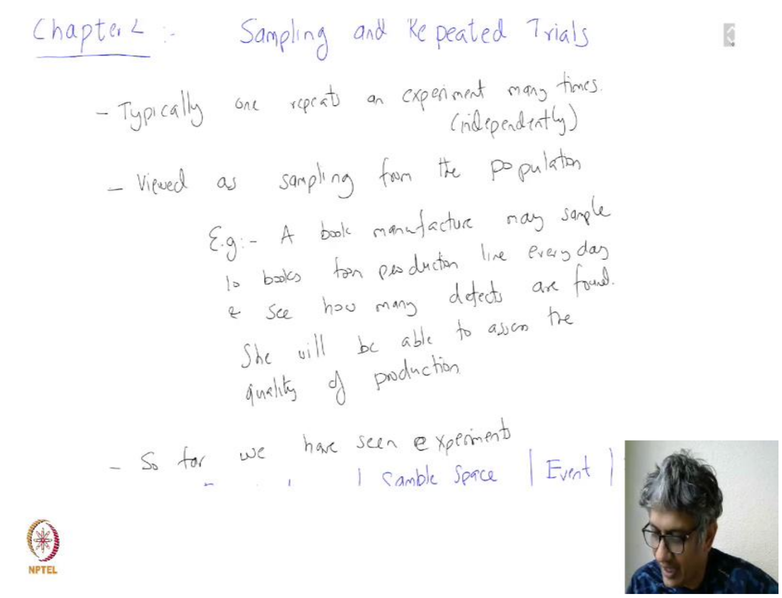
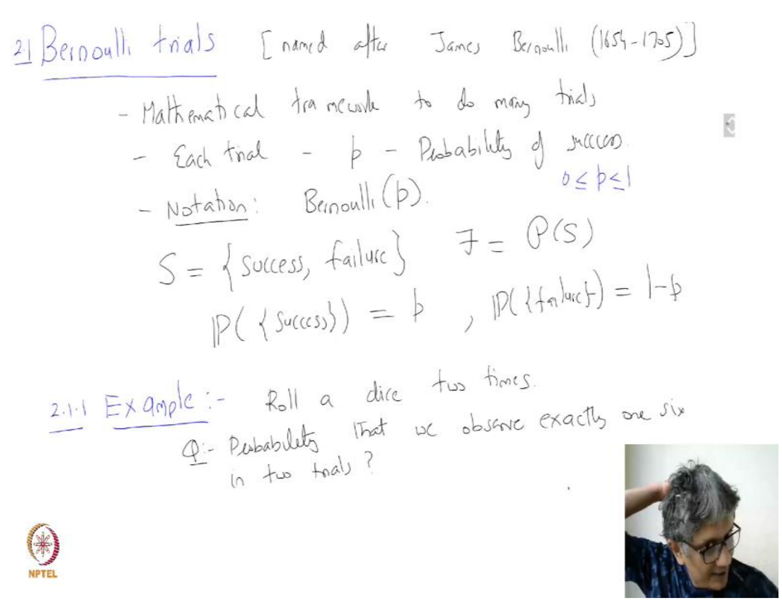
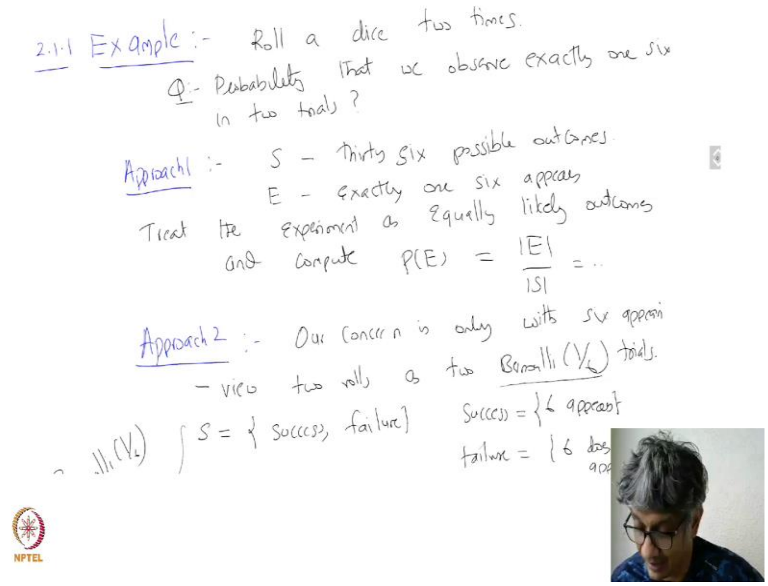
Introduction to Probability – With Examples Using R
Professor Siva Athreya
Theoretical Statistics & Mathematics Division
Indian Statistical Institute, Bangalore
Lecture 10
Sampling and Repeated Trails – Part 02

(Refer Slide Time: 0:23)







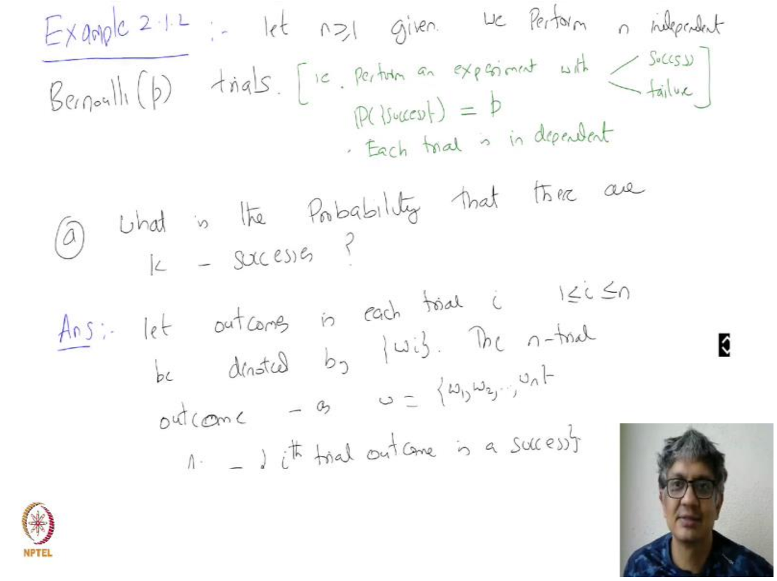
So, we were doing these Bernoulli trials, so let me just quickly recap a bit of what we were doing in the previous class. So, the idea was to understand if you repeat an experiment many many times, how does 1 assess the framework, how does 1 set up the framework and assess the probability of any event that you are interested in.

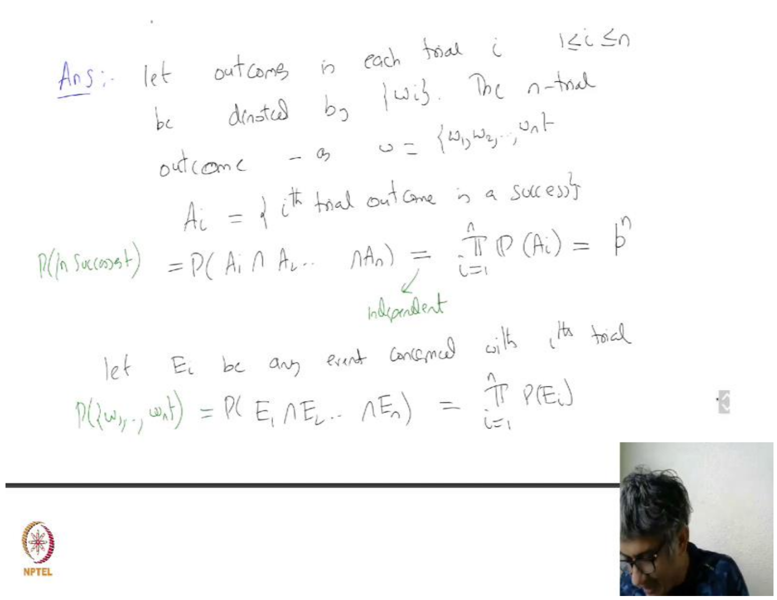
So, idea was to set up as Bernoulli trials, so each trial has the following property that there is a, there is success and failure and the probability of success is going to be p and the probability failure is going to be 1 minus p. So, then we decide an example which we said, we roll a die 2 times and we will not understand the chance that there is a 6 appear only once, so 1 approach we have seen before is that you write down all possible outcomes in 36 (())(1:12) that is how to write down, you write down an event e of interest and write down the probability of interest.

But we know that this also has pitfalls because if we do this n times we cannot possibly write down all the outcomes of e, it is very hard to write and count them. So, then we said that since we are only concerned with 6 we can set this up as a Bernoulli one six experiment and perform the calculations.

So, then we had a we had a sort of structure in which we wanted to find out one six appears in 2 trials that was same as success-failure, failure-success, excuse me, and we did have the disjoint events use independence and Bernoulli to understand the answer.

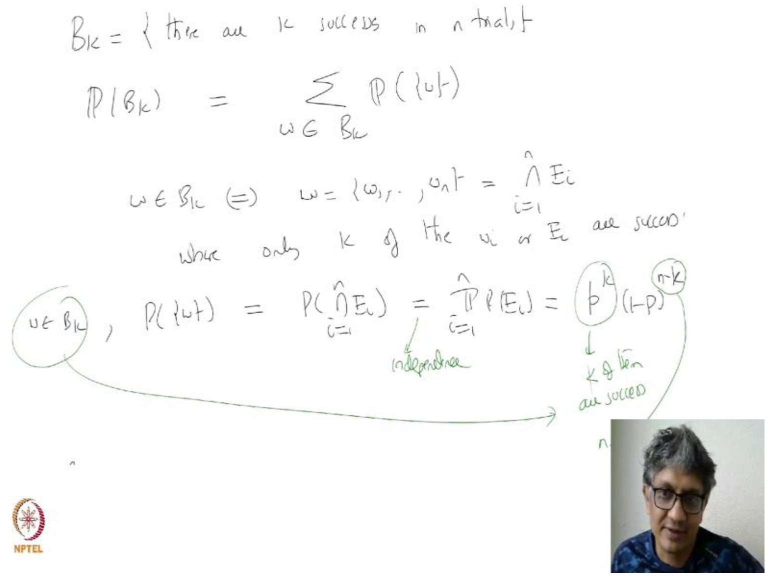
(Refer Slide Time: 2:00)

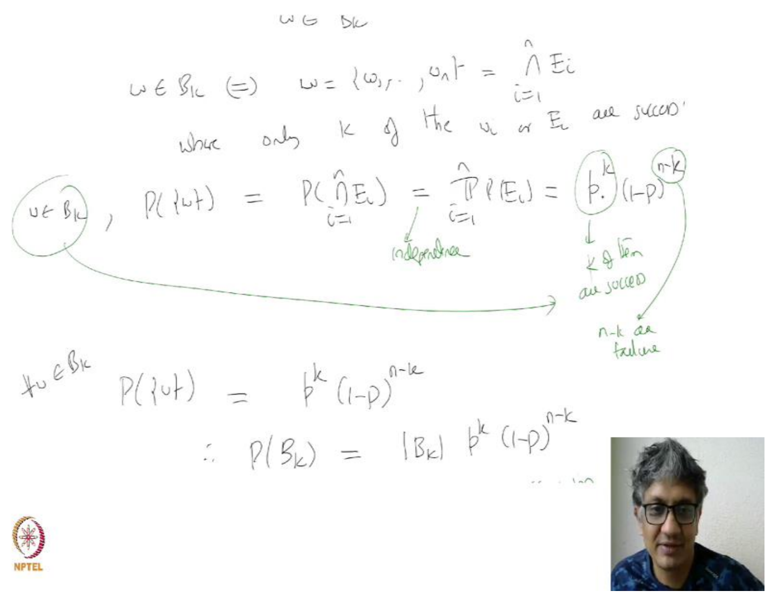


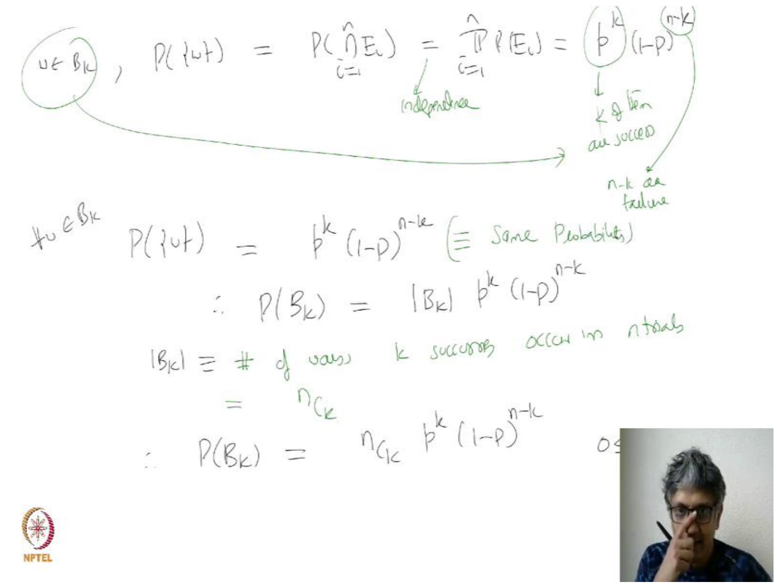


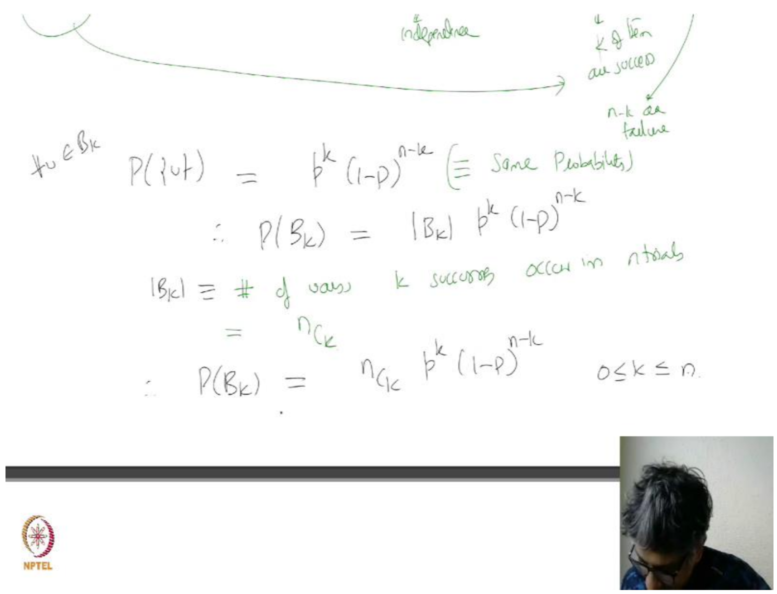
So, then we decide, how does the framework work, so we have n trials, we perform n independent Bernoulli trials, probability in each trial has its success probability is p, each trail independent, how does one do this? So, I want to understand the chance there are k successes. So, the way we did it was we first observed the following idea that since each trial is independent, if you have an event Ei that concerns itself with only the ith trial then the probability of intersection is the broader problem.

(Refer Slide Time: 2:37)





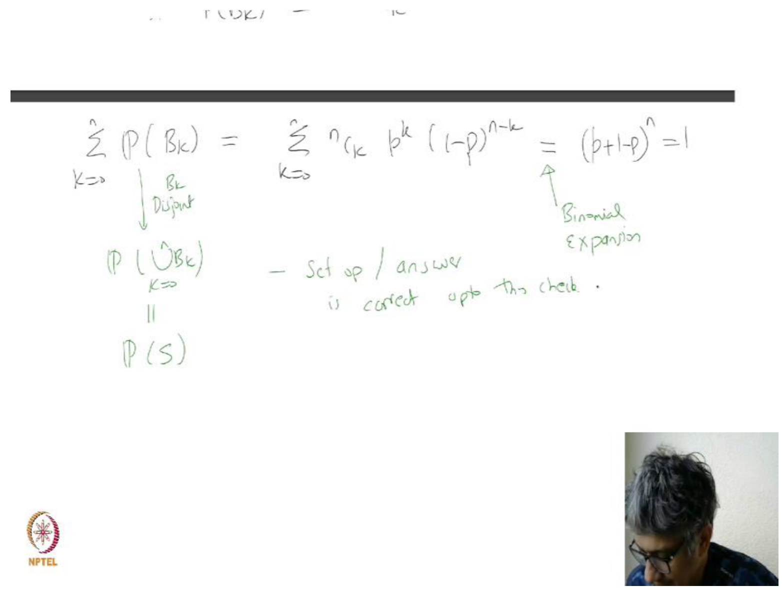




And now we are interested in the chance that Bk that there are k successes in n trails. So, that notation the probability of Bk is the sum of the property of every outcome in the set Bk. But for every outcome in Bk we know that the omega i's that only k of them are successes and the rest of them are all failures.

So, then in the plot of Ei you get p power k for the k successes, 1 minus p power n minus k for the n minus k failures. And then we notice that everybody in Bk has the same probability, so every outcome in Bk has the same probability, the probability being p power k, (())(3:22). So, we luck out, so we know the chance of probability of Bk is just the size of Bk times the probability of every particular outcome and that gives me this nice formula, so that is where we were at last time.

(Refer Slide Time: 3:56)

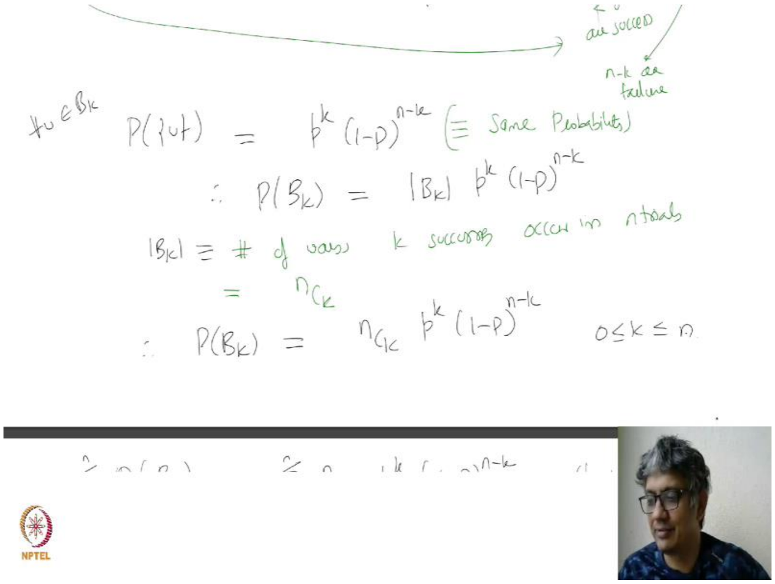


So, now this is, this sort of a thing can generalize for any n and it is sort of a very nice formula that you have. And also one can note the following that you know that the probability of Bk and the sum over k equal to 0 to n that is this the same as so what is this going to be, so what is, this is going to be the same as the sum over k equal to 0 to n, n choose k p power k 1 minus p power n minus k and this is a binomial expansion, we know that that is going to be the same as p plus 1 minus p, p plus 1 minus p to the power n and that is going to be 1.

But then if you look at this side right here, let u go in green, let us see what happens. So, this side is just the I am using the binomial expansion that is a plus b to the power n is just this n choose k a power k b power n minus k and then this guy here since the Bk's are all disjoint events this disjoint Bk's, this is the same as the chance that I have union k equal to 0 to n Bk.

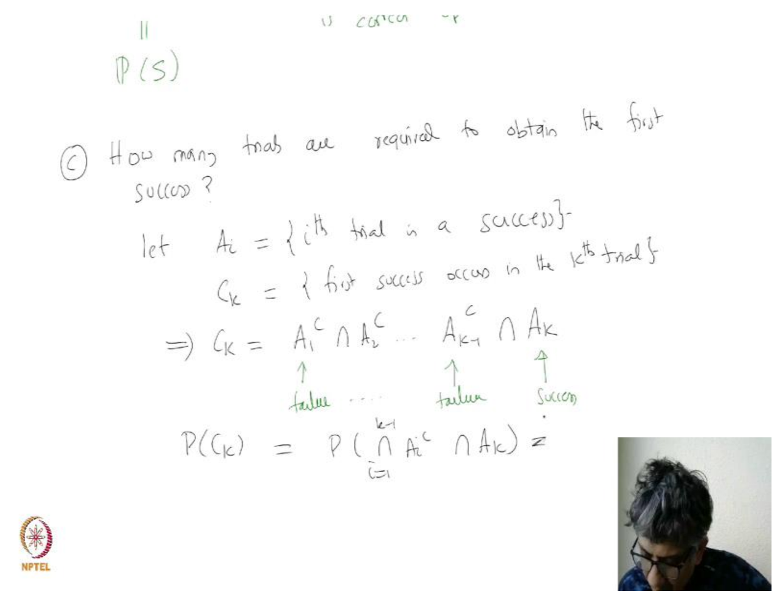
But if you look at this that means in n trials, I can have 0 success, 2 success, blah, blah, all the way to n success that is just the same as the anybody in the sample space character. So, we have, we have just the roundabout way, we have shown that our probability set up correctly and so probability s is equal to 1, so it is in a roundabout way we have sort of set up things correctly, so our setup, set up and answer is correct, do this.

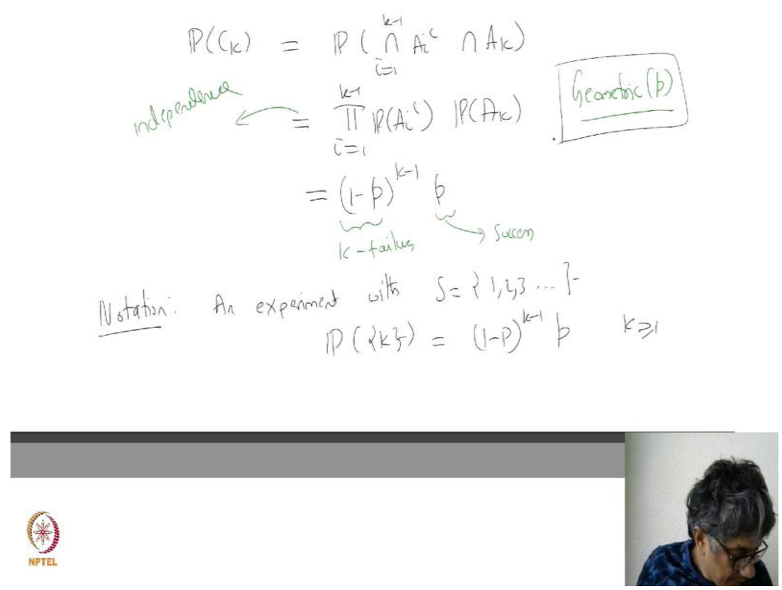
(Refer Slide Time: 6:21)



So, 1 question was asked recently that will the 2 approaches give 2 different answers? So, that is not, that you can verify is not going to happen, the approach will give the same answer to the same event, so probability will be the same so it is not a, because I am using the same structure of independence to get the answer, that is the whole idea of setting it up properly like this is Bernoulli trials.

(Refer Slide Time: 7:02)





So, now this is something we can sort of understand and see and then this 1, this concept one can use quite a bit, so let us do another 1, let us do another thing in terms of independent trials let us see, so let us try and do this, let us try and see, let us try and do another question let us say this is C in the book I will leave it as C, question number C, I will not do B, you can do B.

So, let us say I want to know how many trials are required to obtain the first success, so that means how many trails do you have to make till you observe the first success. So, how does 1 do this? That means, let us say you can have a success in the first trial, so let us say let yeah, let us go Ai, let us do the answer Ai be the trial, be the event, let me write it down properly, Ai is the event that the ith trial is a success.

And let Ck be the event that the first success occurs in the kth trail. Now, so what, I want to understand probability of Ck. So, what is probability, what is Ck going to be? So, this implies what is Ck going to be, Ck is going to be is going to be what, is going to be the intersection, so you have A, A1 complement should happen, A wo complement should happen, big theory and the second trail should happen, all the way and the k minus 1 trial a failure should happen.

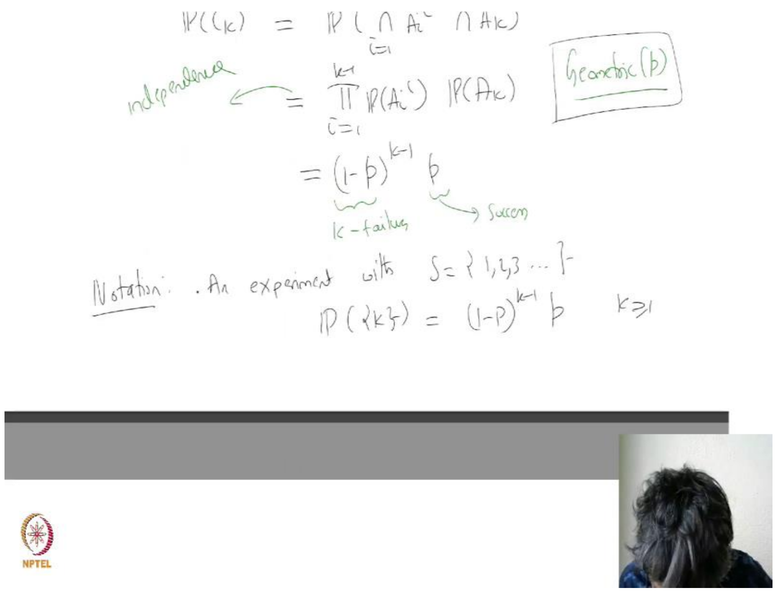
And then in the last trial we should have success that is what Ck, so Ck means the first success occurs in the kth trial that means you have failure, all the way up to the k minus 1 trial and suddenly in the kth trial you have your first success, that is exactly what this problem. So, now this can be calculated quite easily, because now, therefore probability of Ck is going to be the probability of the intersection of i equal to 1 to k minus 1 of Ai complement and Ak and that is just going to be the product of i equal to 1 to k minus 1, probability of Ai complement times probability of Ak.

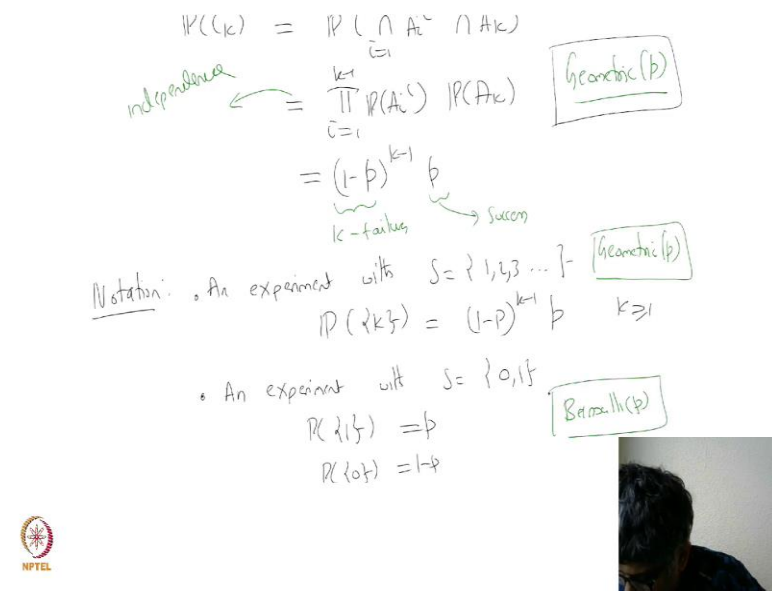
But that we already know is same as 1 minus p to the power, let me write it properly, so as same as the product of i equal to 1 to k minus 1 probability of Ai complement times probability of Ak and that is the same as p power 1 minus p to the power k minus 1 times p power times (())(10:26). So, what have I d1 here, I have, let me use write it in black formula but the probability is always like this.

So, then what have I d1 here, what have I used here, here I use description of Ck from above, here I have used the fact that in our independence, here I have used the fact that these come from k failures and this comes from success, so this is just, this 1 minus p is the probability of failure. So, this has a nice name and this is called, this distribution is called geometric, geometric p, so 1 could, so here is the notation I will rewrite again, so just notation so 1 could view the above question, so what is the question, how many trials are required to obtain the first success is the question.

So, 1 could view this as an experiment with sample space as 1, 2, 3, it takes you so on so many trails with the first success and the probability of every outcome that is k is the chance that the first success comes from the kth trial that is the same as 1 minus p to the k minus 1 times p, so 1 could view it like these experiments. So, this whole thing is called geometric, so it is a distribution on s or you could think of it as just sort of a (())(12:18) Bernoulli trails you put a number of trails to k success.

(Refer Slide Time: 12:52)

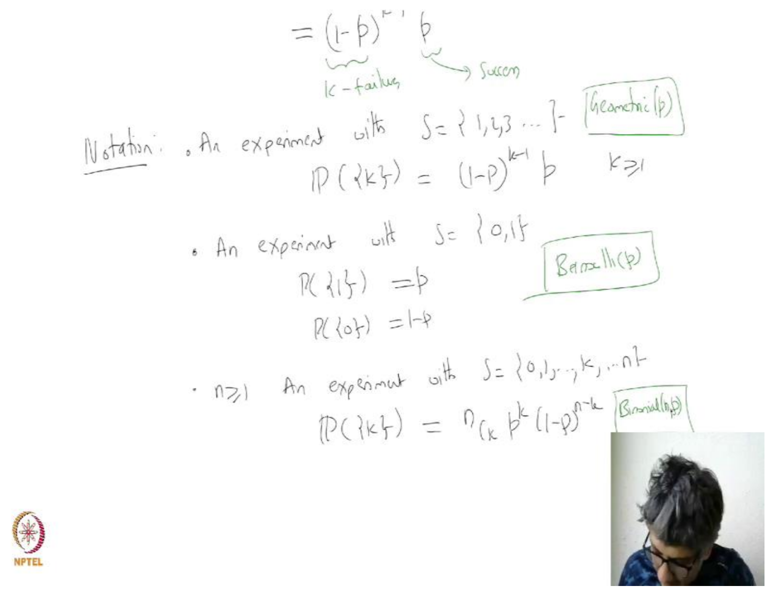


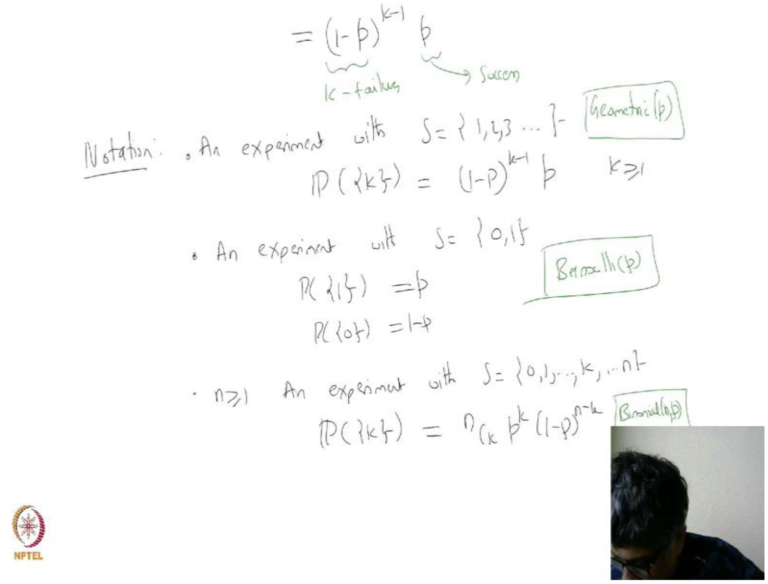


So, we have seen 2 ways of doing it, so 1 is so 1 notation is this, this is 1 notation, the other rotation I would like to introduce is that an experiment you can use, this is how you use geometric, so let us write geometric like this, geometry can be viewed as a coin toss or a rolling die with 2 events success and failure and then you look at the number of trials required to get the kth success, first success let us open the kth trial with this probability.

You could also view this, you can also view the Bernoulli trial as an experiment with 2 outcomes so I can think of the outcomes as let us say 0 or 1 or success-failure whatever you want to do and you can think of the probability of 1 as success as p and probability of 0 as 1 minus p, so this is what you would call as the Bernoulli distribution typically.

(Refer Slide Time: 13:55)



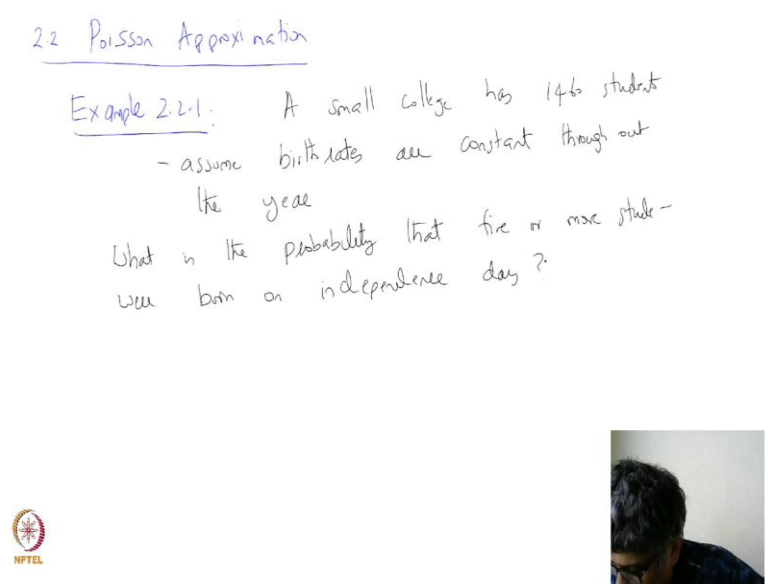


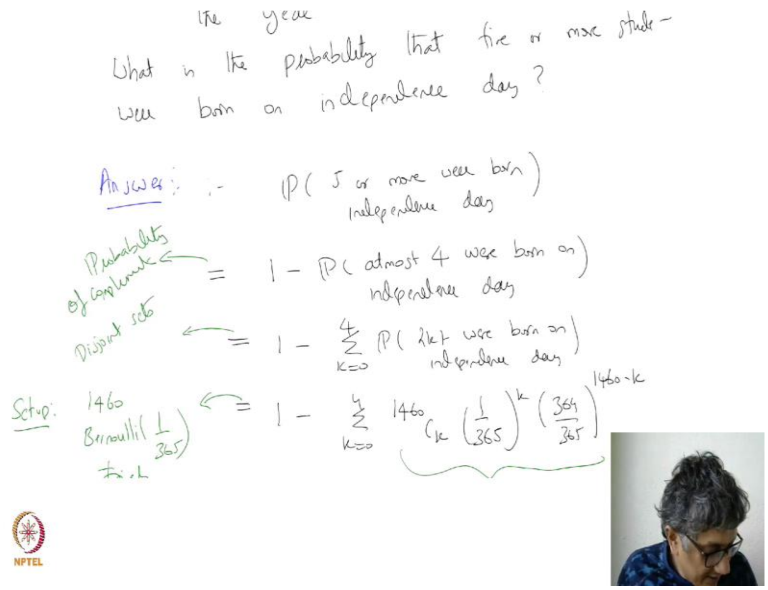
And the experiment where we had this observed k successes we would think of it as n is fixed and you look at the experiment as with s equal to let us say 0, 1 up to k and all the way up to n so this represents the number of successes in n trials and you think of the probability of the of k successes in n trails is probability of k, you think of it as n choose k p power k 1 minus p to the power n minus k, this has a name and this is called the binomial np.

So, we will go up and down these names, so if I say binomial np that means I can think of it as an experiment on with sample space as 0, 1 up to n with these probabilities or I could think of it as performing n independent Bernoulli trials and counting number of successes and I could use any notation I want to use, so that is 1 advantage of this framework.

So, that is that is the broad message in this whole thing. So, if you want to do the same experiment many many times, you just have to make sure that the event interested, you are interested in can be classified as success or a failure in each trial and then you are on your way. And this gives rise to 3 distributions on 3 different sets, 1 set is the set of natural numbers s which is and gives each natural number of rate like 1 minus p to the k minus 1 times p, the other 1 just does 0, 1 and the binomial one does for n from 0 to n, very nice.

(Refer Slide Time: 16:01)





Now, the, this is all very nice and then the thing is that there is something 1 has to do or should be aware of the following is that is computational, so let me just illustrate that and that gives rise to another distribution called Poisson which I will try to explain in the next 10 minutes or 20 minutes.

So, the idea is that let us say you are an ISI, let us do a simple example let me write the examples in blue as before, let us say you are in a small college, so the example let us say 2.2.1 I have put in the book. So, let us say you are in a small college, let us say like a small institution has 1460 students let us say and let us say I ask you, let us assume that everybody, assume that birth rates are constant throughout the year then what is the probability that let us say I ask the question, what is the probability that 5 or more students who were born on Independence Day, suppose I ask this question.

So, this is what, this is, so this one can, one can take a little bit, let me see how do you want to think a little bit about this question, how do you want to do this. So, you can think of it as a Bernoulli trial with, so you have each student has a probability 1 over 365 being born on a particular day and you want only 4 of them to be born and 5 or more to be born on that day.

So, what you could do is you could do a, this is a simple exercise so this is, so you could, I will leave it as an exercise to check the following that this probability that 5 or more were born on independence day is the same as probability, is same as first thing you do is like I will not react, write everything out, so same as 1 minus the probability that at most 4 were born on

Independence Day, that is easy to do, and that is the same as 1 minus the sum from k equal to 0 to 4 the chance that k were born on Independence Day.

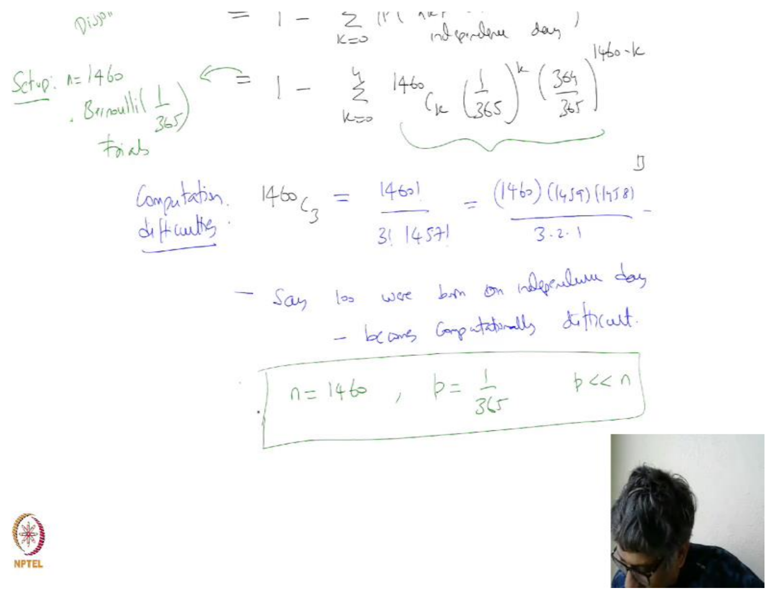
Now, the key is the following idea that this is the same as 1 minus, so you look at every student as a Bernoulli trial and the chance of success is 1 over 365. So, that means this is the same as sum from k equal to 0 to 4, the chance of k successes in 1460 trials that is the same as saying 1460 choose k 1 over 365 to the power of k and 364 over 365 to the power of 1460 minus k and that will be the answer.

So, let me just notation-wise let me rewrite this, let me write this as answer and I will put x as exercise as a sketch, so what you have to justify, let me also write down what I will justify, so the exercises. So, 1 is that this part is clear, this part is just using the fact that (())(21:06) is using the part about probability, there is no problem, this is the probability of compliments that part is easy.

So, now this part also is easy it is just using the fact that you have disjoint sets, so if you have at most k 4 things means you have k equal to 0, 4 and 0, 1, 2, 3, 4 there is a union so here the idea is, here is where the mathematics come in, you set it up, you set up the problem as you set up the problem as 1460 Bernoulli 1 over 365 trails and that was a crucial part actually in this set up here to get this answer, not this on this part, this is just this part, that was a part of it, and that is why I did the answer. So, this is something you have to write out properly but I have sort of given a quick immediate way of getting the answer and the steps are fairly clear.

But now there is another problem now. Now, the thing is that I do not know if you notice it you try to do this in a calculator it will be quite hard, see this is 1 over 365 is small, this is quite large the other 1 and then you have to compute 1460 choose 4 that is quite hard to do, you will do 1460 factorial so 1460 let us say choose 3 so computational difficulties are following.

(Refer Slide Time: 22:53)

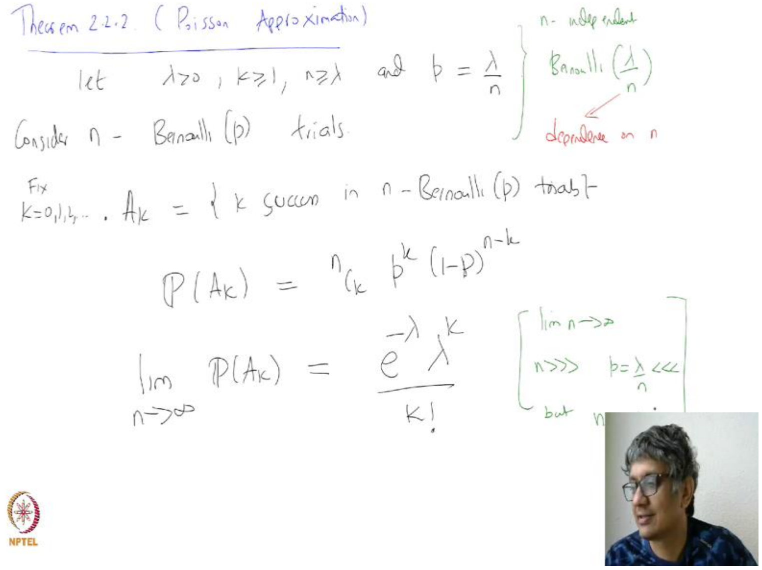


So, let us say computation difficulties like are quite large because you do this, this is same as 1460 factorial by 3 factorial and then the other 1 is 1460 minus 3 which is, this is 1457 factorial so you can do a little bit here you can you can do a little jugglery and get the answer, you can use the 1460 into 1459, in this case you can do this all jugglery (())(23:37) divide by 3 into 2 into 1, that is the answer, that is how you can compute these guys, and then for 4 you can do it.

But the moment I get larger than if I ask the same question let us ask the same question, let us say 100 were born on Independence Day, on Independence Day, the question is quite hard to do, Independence Day becomes computationally difficult. So, we must develop a different way to do this.

So, like what can we use here, so what can we use, what can we use in this problem? So, the idea is that the, what we can use here is the following, use here is that n is 1460 that is total number of trials that is n equal to this and p is 1 over 365 that is a success and p is kind of much much smaller than m. So, can we somehow use this, can we use this, can we somehow use this fact that these this thing is happening, so can we somehow use this idea to try and see to resolve this computationally difficult. Because if 4 was replaced by 100 then this competition becomes much much harder to do.

(Refer Slide Time: 25:28)



So, I will close off with the theorem which will come and show next time and then in the next class, so theorem is the following, so how to resolve the computational difficulties, the theorem is the following, so very very fundamental theorem it is called the Poisson approximation. So, let lambda be positive, k be bigger than equal to 1, n be at least lambda and p be equal to lambda by n, so this is something I am setting up in the beginning.

So, then we consider n Bernoulli p trials, so note all I am doing is I am considering, performing n independent trials with probability p success. So, let me rewrite, erase this little bit, let me write this here. So, let me consider n different trials, so under the 1 of the things to load in mind, 1 is that each trial is independent, n independent trials and then the Bernoulli, the probability depends on n so it is lambda by n, so that is the crucial part you have to be careful about that is the, that is dependence on n.

So, for large n I am having success probabilities getting smaller and smaller because p is lambda by n. So, then I do this, so n is fixed let us say n is fixed then I do this following I want to compute let us say Ak is the event that there are k successes in n trials, so n is sort of hidden in the Ak but n Bernoulli p by trial is a n Bernoulli p trails, n Bernoulli p trails.

Then I want to know this we already know that the chance that A sub k your (())(28:05) k succession n going to be p trails that you already know it is n choose k p power k 1 minus p to the power n minus k this we already know from before there is no problem. But what the theorem says is, limit as n goes to infinity the probability of Ak that is if n is large and p is small and what happens is the limit approaches the fixed number for every k and that is lambda power k by k factorial.

So, this is true for all k so k is equal to 0, 1, 2. So, let k be equal to 1 of these guys, so fix the k, once you fix the k you let Ak be this and this happens so that means what should what, how interpret this if you want to compute, so here the limit as n goes to infinity you interpret it in the following manner, you interpret as n as being very very large and p which is lambda by n is being very very small. But p is not too small, it is just n times p is always a constant lambda. So, if you have a situation where p is very very small but n times p is a constant then you can use this idea to sort of compute the problem through this, we will do this in the next class.