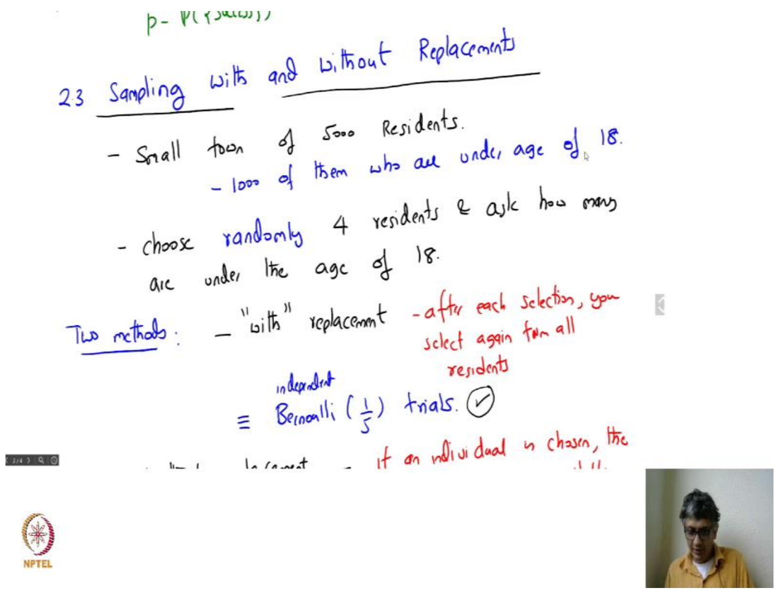
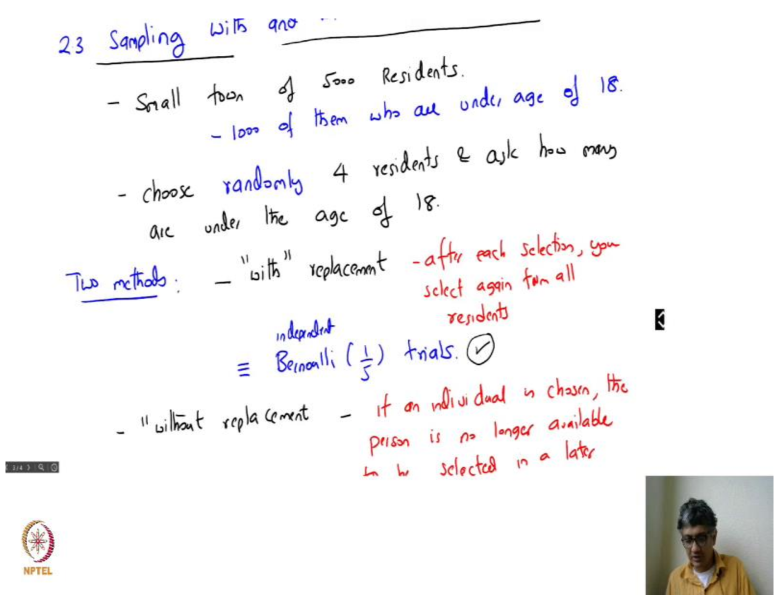
**Introduction to Probability – With Examples Using R**
**Professor Siva Athreya**
**Theoretical Statistics & Mathematics Division**
**Indian Statistical Institute, Bangalore**
**Lecture 12**
**Sampling Without Replacement**

(Refer Slide Time: 00:15)

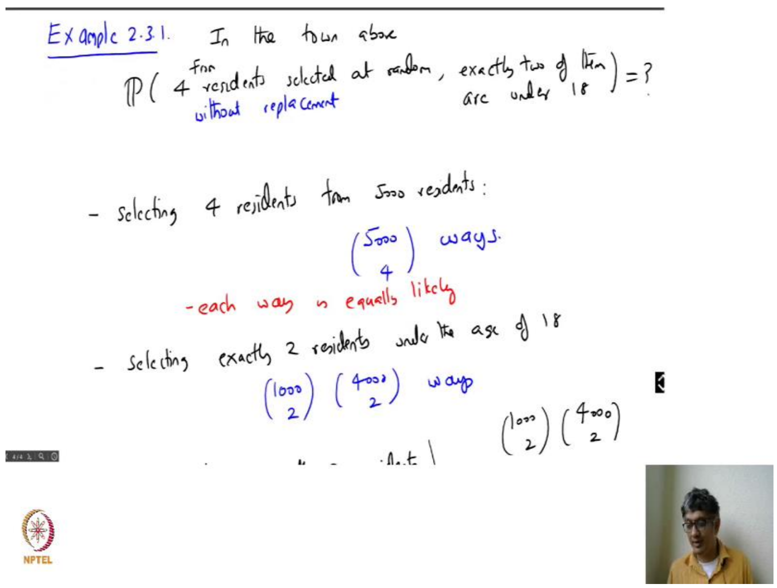



So, we were doing this sampling without, with, or with replacement for the last class. So, the idea is that we have a small town of 5000 residents, 1000 of whom are under 18. So, we choose 4 of them at random, and ask how many people are under the age of 18. Now, there are two ways of doing it; one way is that, you can say is that after picking everybody, I put that person back in the population, and pick again.

So, that is with replacement. So, everybody is available for selection in the start. So, if you think a little bit, that is just like choosing, so you have a total of 1000 whom you are interested in, and there are 5000 of them, so 1000 by 5000 is 1 by 5, and that is why you get independent Bernoulli 1 over 5. Then, the other option is that once you pick a person, you do not put the person back. So, it is without replacement, without replacement. So here, if an individual is chosen, that person is no longer available for selection. So, there are two methods of doing this.

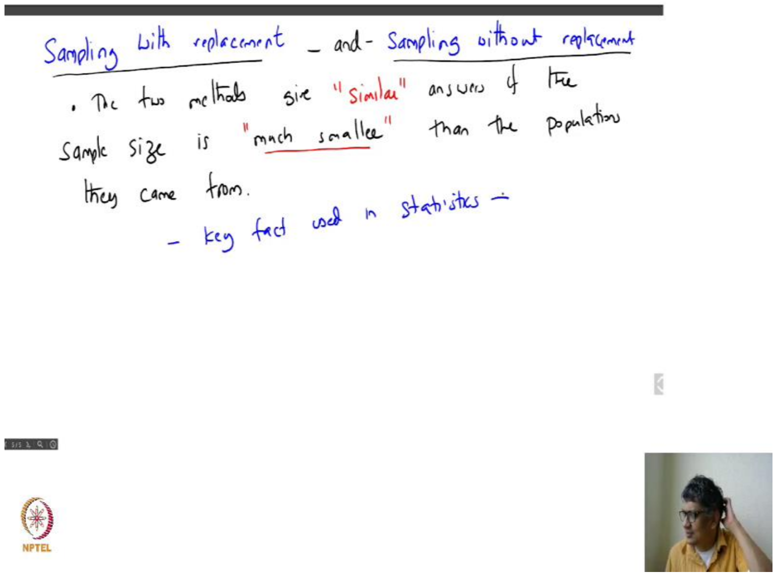
(Refer Slide Time: 01:40)

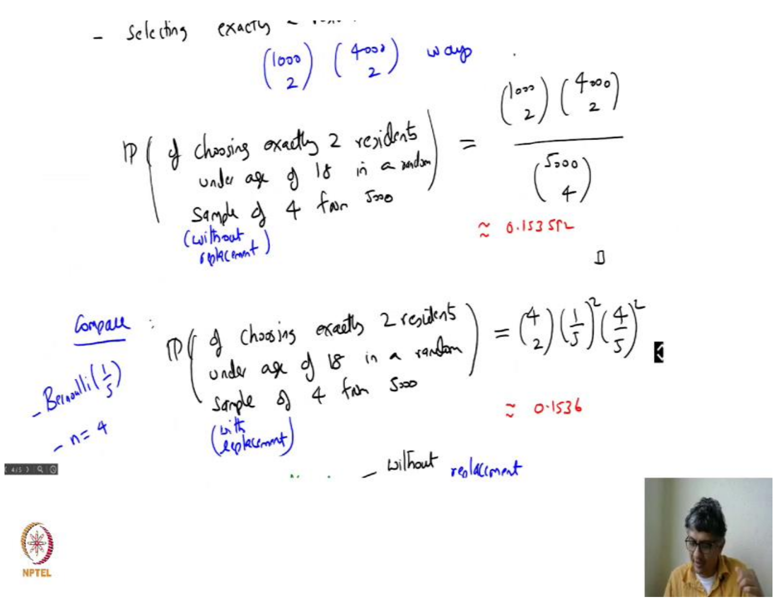



So, the first method we have seen, so let us say we want to understand a question that exactly 2 of them are under 18 from the 4 that we have chosen. So, way one does it is, you think of it in the following way, you have 5000 choose 4 ways of choosing 4 as in 5000; each way is equally likely. Selecting 2 of them under the age of 18 is 1000 choose 2, 4000 choose 2 ways. So, the net probability is going to be 1000 choose 2, 4000 choose 2.

So, that is the answer you get there. And then what we found out was that if you did independent Bernoulli trials, that is, if you chose them with replacement, then the answer would be 4 choose 2, 1 over 5 choose 2, whole power 2, 4 5th whole power. And then if you do the computation, the answer in red turns out to be quite a close by. So, the key thing is that the two methods are very different. So, one is without replacement and one is with replacement. And in one case, we have independence, in the other case, we do not have, but the answers are not so far. So, this can have a very useful idea that is used in statistics a lot.

(Refer Slide Time: 3:12)



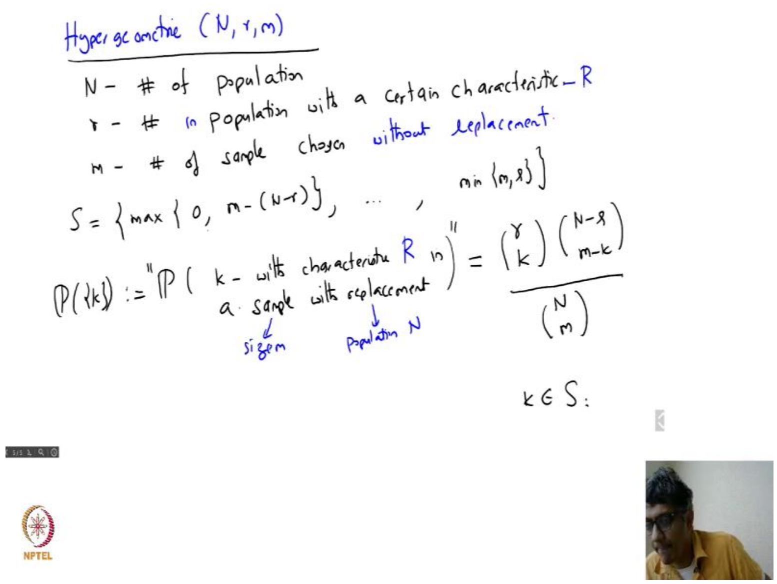




So, the idea is the following that, so sampling with replacement, with replacement, with replacement, and sampling without replacement, and sampling without replacement. So, there are two different, are very, they are, they are sort of, so let me write down this way, sampling with replacement and sampling without replacement, without replacement. The, two methods so let us have two methods, there are two methods. So, I should say give, (())(4:28) they are not shown a proof for this, give similar answers if the sample size is much smaller.

What is much smaller, we do not know, much smaller we do not know, is much smaller than the total population size, so we say the population size. Then the populations (())(05:11). So, of course, this is something that I have not shown or proved for you but that is the formula; this, with replacement and without replacement. They are two different methods, they give two different outputs, but the methods give similar answers if the sample size is much smaller.

So, this is a sort of a, if you do statistics, this is a key fact that is used in statistics a lot. So, again like, before when we did this n go to infinity and we computed the probability of AK and I said that it goes towards Poisson, similarly the above is also can be, can be thought of a distribution, this quantity right here, this 1000 choose 2, 4000, 5000 choose 4.

(Refer Slide Time: 6:26)




So, the distribution is called hypergeometric, it is called hyper geometric distribution, and it has several parameters. So, one is capital N, is the, is like the 5000 people. Then it has little r which is in our case, the none of people who are below 18, and m is the sample size of the population. Let me write it, let me write this in blue so that it is clear so N, r, and m.

So, the idea is that, let me write the whole thing in proper form. So, n is the, is the, is the number of people, or the size of population, the population. This could be people, it could be balls. And then r is the number of people, number of, number in population, number in population with a certain characteristic, certain characteristic.

And then m, little m is the number of the sample chosen. So, here is the key without replacement. So, you choose m from the population without replacement, from capital N people, and all of whom are the certain characteristic. Of course, the I mean what you are interested in is the number of people with characteristic r.

Of course, your m has to be less than r, less. If it is more than r, then you cannot even pick more people outside the characteristic. So, then let us, let us write this down. Let us say S. So, S is, the sample space S is going to be, you can have 0 successes, 1 successes, so on and so forth, and up till, so at the most r.

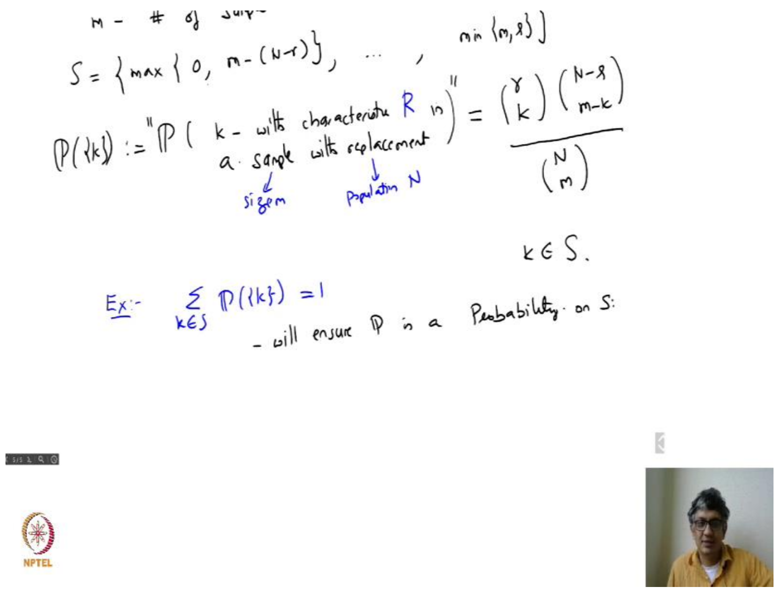
So, what we will do is, so, we will, let me just backtrack a little bit. So, before I set up sample space, so, so what can you do? You can, you can, you can pick at the… at least 0, maximum of 0, and just to be mathematically precise, I divide it as m minus N minus r. That way, I do not have to fix what r and m is.

So, S is from here. We start with this number and then it all, it goes all the way till the minimum of m and r. That is the number of people I can pick in my sample, and the chance of picking k people with characteristic r, with replacement, in a sample with replacement, (())(10:00) this, in a sample with replacement. And the sample is of size m, sample is of size m. And the population is of size capital N, and there are r people with characteristics.

So, let us say 7 characteristics. I will not say characteristic r, let us call it characteristic, let us call it characteristic, let us say capital R. But here I will have to say, capital R. So, what do I get? This is going to be equal to, so it is going to be, I have to pick the k people from little r, or choose K.

In the rest of them, I have to choose them from N minus r, and I have chosen them from m minus K. And the number we are choosing K people is just N choose. And m, m people, is just m choose. So, that, this is the, called the hypergeometric distribution. And you can think of this as, this probability of K or probability of K people (())(11:29). So, this probability of K is defined as, this, this is how we think about this probability, and the answer is just, is finding K.

(Refer Slide Time: 12:20)




So, of course, one is, one way is to check. This is some, so that you can think about is, is it a clever way of check. Up till now, I the way I set up the experiment, it is kind of clear that this is exhaustive, but one exercise to check is that is it really a probability? Does P define a probability? So, you are doing, so, does, the question is that, then it says is that does P define a probability?

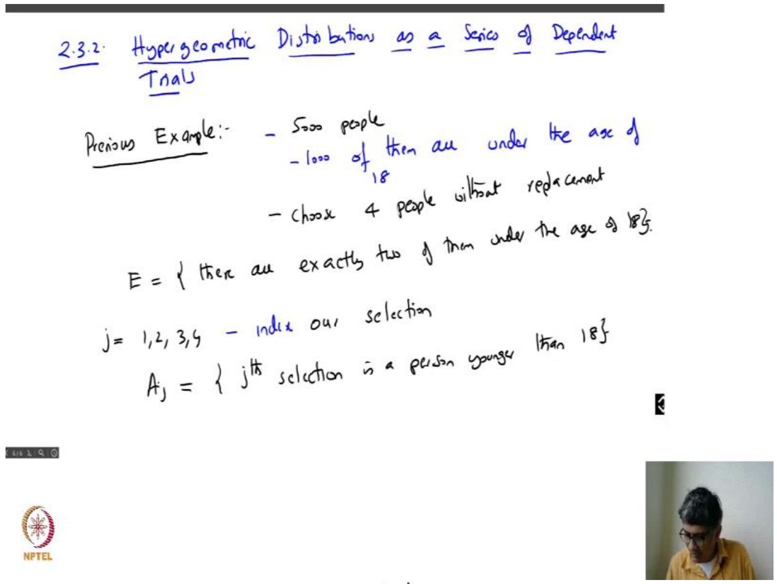
The idea is that, can you show that summation KS, (())(12:27) with K, is actually equal to 1? Let me think about it. So, it will ensure that P, P is equal to 1. So, this is a very important distribution, and it is a, it is a very very powerful one, and used widely in application. Hypergeometric means with without replacement. And you have to be careful that dependency is involved, and that is how you compute the probability.

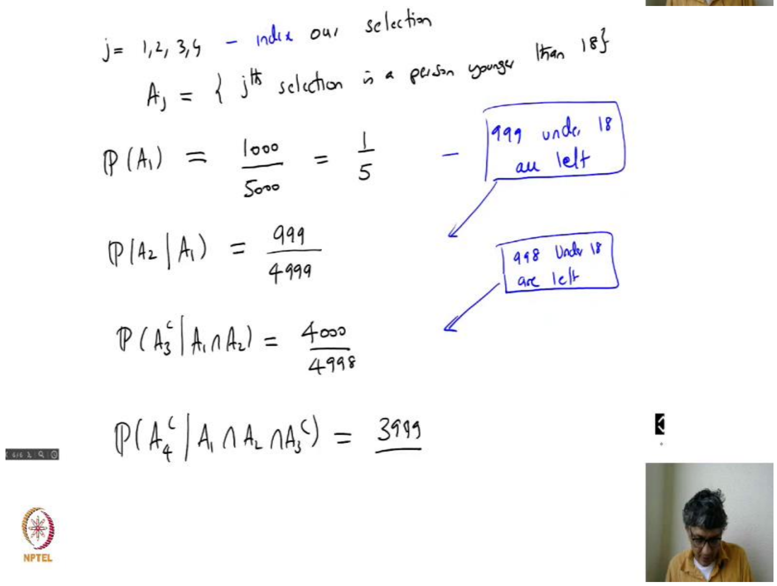
(Refer Slide Time: 13:37)

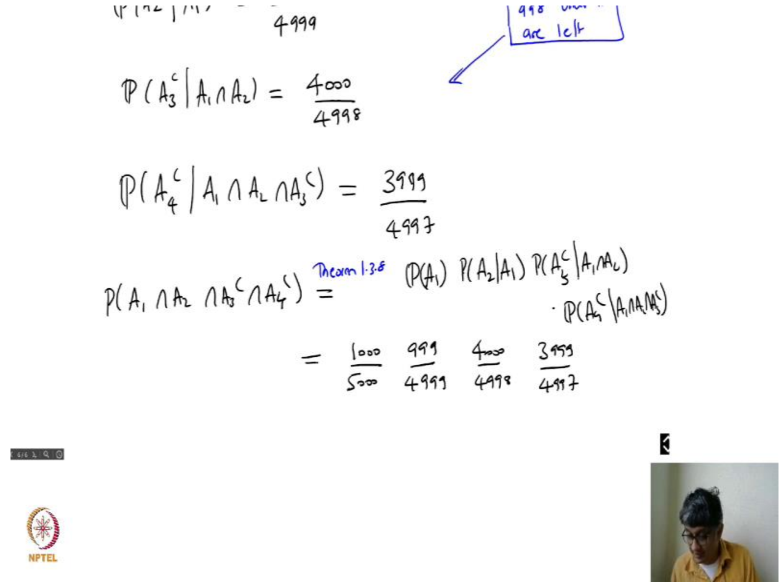

So, now let us, let us, I sort of, I sort of was a little below. So, I said that, so let us just go back to this idea and try and make it, make something, make some sense of it. Let us try and go back to the operations. So, I had said that if I do with replacement, with replacement, then I get a series of independent Bernoulli p trials.

And then I had said here without replacement, I had said, without replacement, and I had said that this can be thought of as dependent trials. Of course, but I did not tell you why they were dependent, and so dependency is clear because the moment I pick somebody up, the sample is getting smaller, so the dependency is obvious. But can I, can I make this idea rigorous? Let us try and let us try and see if I can do this.

(Refer Slide Time: 15:09)






So, I want to make the idea rigorous that I want to think of hypergeometric as a case of dependent Bernoulli trials, or dependent trial, not Bernoulli trial. So, let us try and understand that. So, this is 2.3.2. This is hypergeometric distributions as a series of dependent trials. So, now, I want to just try and see how (())(15:41). So, this... So, now here let us try and see how to, how to resolve this.

So, let us go to our previous problem, our previous example. So, previous example, I do the same thing. I say that I have, I have 5000 people here, I have 5000 people; and I mean that is the population size. And I knew that 1000 of them are under 18, are under 18, of 18, is what we had before.

And then, I said that fine, now I am going to go and pick 4 people, so choose 4 people without replacement. And suppose you want to set it up as so dependent trials, the way one would do it is, you would, you would first do the formula. Let us try and do it, so let us see. So, again, we wanted to (())(16:55), event interest of event of interest E is that there are exactly 2 of them, 2 of the people, 2 of them under 18.

That is again our (())(17:18). So, how did I do this again? Now, what I will do is, if I wanted to set up the, before, I know the answer to this, it is just 5000, if 1000 choose 2, 4000 choose to divided by 5000 choose 4, but I want to set it up a set of independent trials. So, what I do is, I say I take j equal to 1, 2, 3 and 4 to index, index my trials, index my samples, or index, index (())(17:53).

Then what I do is I will define Aj as the event is that the jth selection, for the means of sample, I will just say index my selection, index our selection, sorry. Then Aj is the fact that the date selection is a person under, younger than 18. That is the, that is the selection I have. So, I have, I choose 4 people, but I just order it in this.

So, now, let us do, let us do a little bit systematically. So, 1000 people are there. I want to look at probability Aj. So, first step is clear; probability of A1. A1 is clear, A1 means set of first selection under 18. So, that is clear, so, that is just I have 1000 people to choose, who are, who are favorites and I have 5000, so, it is 1 over 5.

This is exactly the way I would have computed the Bernoulli people. So, the first one is clear. Now, the second one gets a little tricky. We have done this before, let us do it again. So, now clearly let us go back here. So, after the first selection, 999 are left. So, then probability of A2, if you want to calculate, then you have to do it in two ways, you have to do it in two ways.

So, you have to do it as probability of A2. So, let us not do, let us come back to this in a second. So, let us just do, let us do the simpler one. So, now let us say I, I want to do something simpler. I want to do probability of A2, given A1, so given A1. So, what is that going to be? That is going to be 999 left, so, that is going to be just at the bottom. So, 999 people under 18 are left. And 999 people under 18. Let us write this here 999 under 18 are left.

So, now I have already chosen one person here. So, when I come to this competition, when I come to this competition right here, so, I will get, here I get is, is 999 people I have to choose from, and everybody left is for 4999. So, now let us say, that means that first one I have chosen under 18, the second one I have chosen under 18. And let us say I chose the, I, I look at the third guy whom I do not want under 18, because my event team will need exactly 2 people.

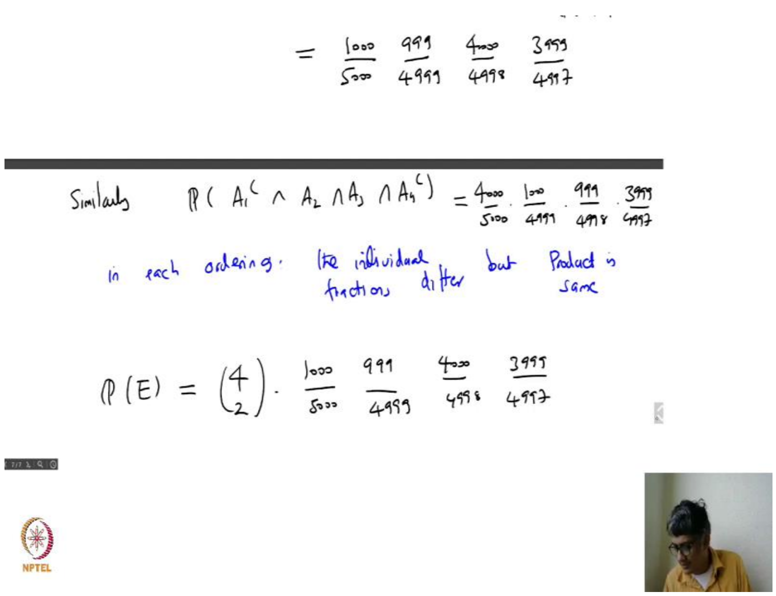
So, let us, let us do one particular computation, so A complement, given A1 and A2 is going to be what now? After this, there are only, at this step, there are 998 left, under 18, at this step. So, but here I want A to be a complement, that means I do not want the person to be under 18. So, then I have how many people I have left? Everybody has left, almost 4000 of them left, divided by 4998, and I come down here, and I do the same thing again.

If I will do the 4th person, I do not want to be under 18. And I do the event under which I have A1, A2, and A3 complement. First trial, I took under 18, second, under 18, third, I do not pick under 18. This is going to be again equal to 3999 by 4997. So, each time the population goes down by 1 and the appropriate category also goes down by 1 depending on (())(22:18).

Then we know this idea, we know that probability of A1 and A2 and A3 complement, and A4 complement. That is one way of choosing 2 people under 18 in a size of 4, this we did this in theorem, if we go back to this Conditional Probability Theorem 1.3.8, this, we know is the same as probably of A1, probability of A1, A2 given A1, probability of A2, A3 complement given A1 and A2 times the probability of A4 complement given A1 and A2 and A3 complement. So, that is how you figure it out.

And that you multiply it out, and what do you get? You get 1000 by 5000, and you get 999 by 4999. And then you get, the next one, you get is 4000 by 4998. And you get 3999 divided by 4997. That is how you (())(23:35). Of course, there are many, many ways of doing this. There is one way of getting 18 people, in getting 2 under 18. So, you have to go through all possible ways. Let us do that.

(Refer Slide Time: 24:00)



5000 4999 4998 4997

in each ordering. the individual fractions differ but Product is same

$$\mathbb{P}(E) = \binom{4}{2} \cdot \frac{1000}{5000} \frac{999}{4999} \frac{4000}{4998} \frac{3999}{4997}$$

$$\underset{(Ex.)}{=} \frac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}}$$

Example 2.3.1. In the town above

$\mathbb{P}$( 4 residents selected at random, from without replacement, exactly two of them are under 18 ) = ?

- Selecting 4 residents from 5000 residents:

  $\binom{5000}{4}$ ways.

  - each way is equally likely

- Selecting exactly 2 residents under the age of 18

  $\binom{1000}{2}\binom{4000}{2}$ ways

- Selecting 4 residents from 5000 residents:

  $\binom{5000}{4}$ ways.

  - each way is equally likely

- Selecting exactly 2 residents under the age of 18

  $\binom{1000}{2}\binom{4000}{2}$ ways

$\mathbb{P}$( of choosing exactly 2 residents under age of 18 in a random sample of 4 from 5000 (without replacement) ) $= \frac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}}$

$\approx 0.153592$ □

So, similarly, one can check a probability of A1 complement, that is an A2, and let us say A3, and A4 compliment is also that same product, of the same, in the next slide you can check, the same product is 1000 by 5000. That is the first one. So, maybe you should do a different, let us do, it is the same thing as, let us write the same product but to do the correct order, give A1 complement first, so 4000 by 5000. Then it is A1 complement given A2, that is going to be 1000 by 4999.

The third one is going to be A3 given the previous two, A1 compliment at A2, that is 999 by 4998. And the last one is going to be A4 complement given the other three things, A1 complement, A2 and A3. So, you are going to get 3999 divided by 4999. So, the products are the same. So, you notice that the, that the, that the net product is the same, even though the order is different, the net product is the same.

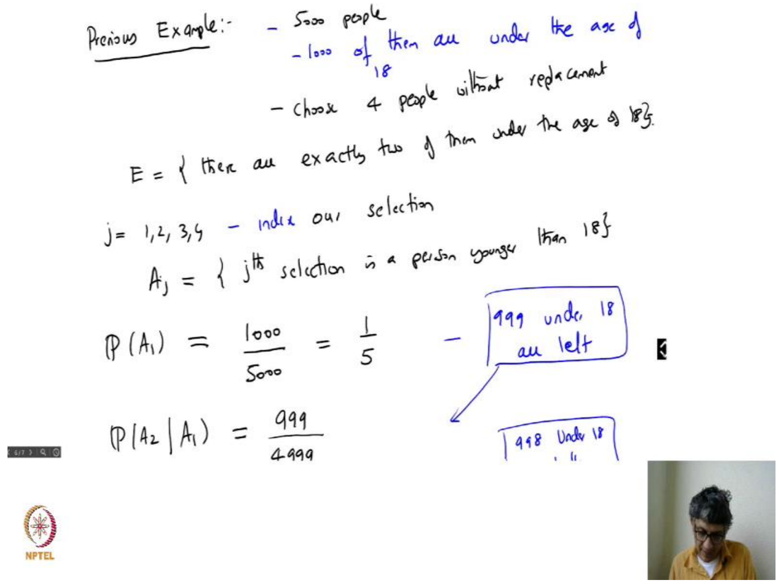
So, we notice that in each ordering, the individual fractions are different, individual fractions differ, but product is the same. So, then it is a, then you, then you set it up as a nice dependent Bernoulli trial, where in each selection you pick under 18, but you want exactly two of them, so you do one of these things, one of these things that have a complement, two of them will have complements, and two of them will have no complement.

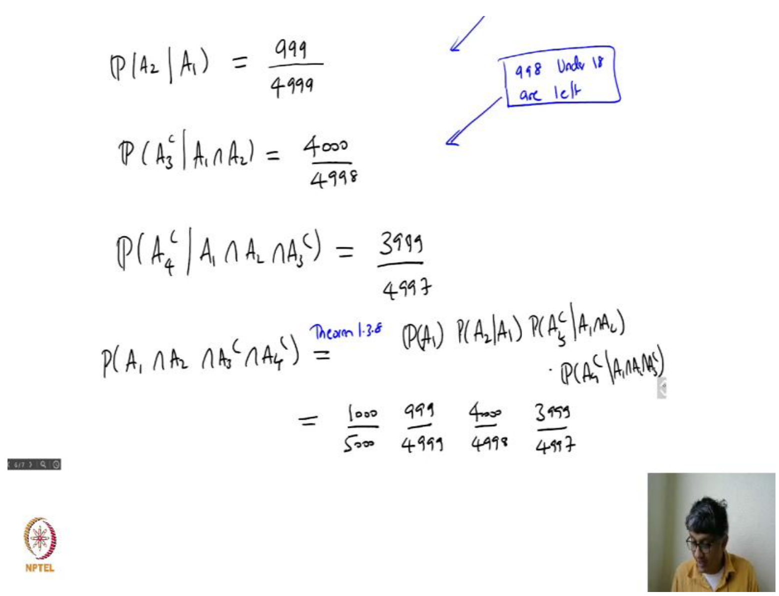
But any ordering has the same product. So, then we can think a little bit, just the probability of E is just the number of ways of choosing 2 people from 4, or the number of ways of putting 2 complements in 4 sets, you get 4 choose 2, and we put the same product, 1000 by 5000, or 999 by 4, it can be anything you want, 4999, 4000 by 4998, and 3999 by 4997.

So, that is the (())(26:41). But this, if you, if you just do some algebra, so, this is an exercise, I will leave as an exercise, algebra. This will exactly be the same as 1000 choose 2 into 4000 choose 2 into 4000 choose 2 divided by 5000 choose 4. That is how you will get it. So, that is the kind of the, the way of phrasing the whole thing as a dependent trial, and then you can get the same answer. So, it is kind of a nice idea. You can phrase a, hypergeometric in the following manner, like I did before in this example first. So, what I will do was I said, fine.

If I want to think of it as choosing 4 people from 5000, I look at the total number of ways of choosing them, that is 5000 choose 4. They are all equally likely, the number of ways I can choose my favorite sample is 1000 choose 2 times 4000 choose 2, and the ratio is my answer. That is one way of doing it.

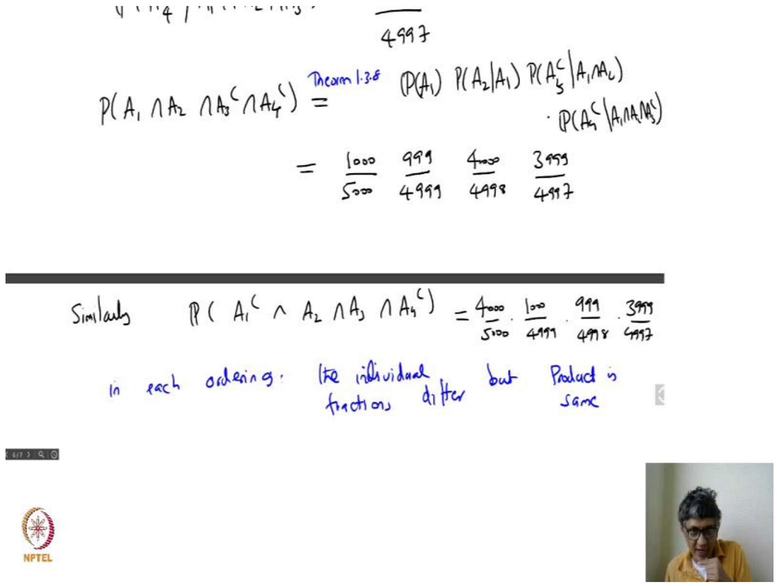
(Refer Slide Time: 28:10)

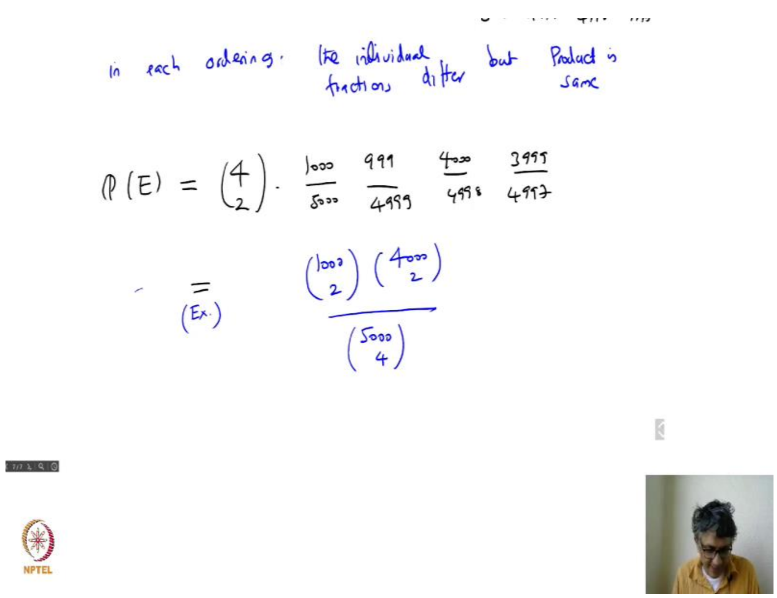



But I can also phrase the whole thing like a dependent trial. And the way I did it was I said, I will look at each selection, and I check if the person is under 18 or not. Then, ultimately, I want only 2 under 18. So, I want A1, A2, A3, A complement, and A4 complement. That is one way of getting under 18. Then I, I compute that expression, I get something using the (())(28:30) probabilities that we had before.

(Refer Slide Time: 28:36)

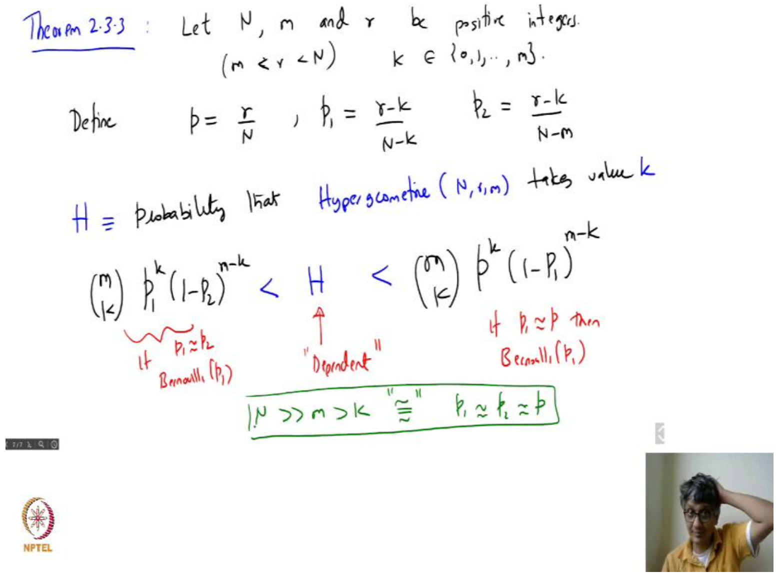



Now, if I reorder the selections, if I do, if I do not pick the first person under 18, and the second person under 18, third, and then I do not pick the fourth one under 18, that will give me a certain answer. But I find that the, even though the individual fractions are different, the products are the same. Then all I have to do is just find the number of ways I can put two complements in 4 sets. So, that is the number of ways is 4 choose 2. And the answer is the same for every selection, every, so, every (())(29:03) of, of complements. So, that means I get the answer to be 1000 choose 2, 4000 choose 2, and 5000 choose 2.

So, that is how you can order independent trials and dependent trials. So, let me write that theorem down, that is sort of, that quantifies this in a certain nice way. So, the idea that I said that sampling with replacement, without replacement, can be thought of as dependent, independent trials.

(Refer Slide Time: 29:55)




Let us try and write the theorem down, that is, that has to be exactly how, how the dependency can be captured. Here is the theorem. So, let N, m and r be positive integers. And let us just make everything simple. Let us assume that little m is much smaller than r, it is much smaller with the previous jugglery of, of N minus m minus r (())(30:38).

Let K belong to R, 0 1 up to m. So, now when we define p as r by N, we define p1 as r minus K by N minus K. I will define p2 to be equal to r minus K by N minus K. So, now let us do the following, you think of capital H as the probability of choosing let me write it down in black, choosing K people, (())(31:41) I will come back at K people, choosing K people from N people, so actually let us not say this.

We will just, let me not, let me not go into writing the experiment down. Let K be the probability, H be the probability, H be the probability that the probability that the hypergeometric distribution of N r and m takes the value of K. So, by this I mean this H is going to compute the probability that I choose K people with characteristic R with replacement, the sample of size m.

Then, I can then the following inequality. So, on one side I have H I also (())(32:58). So, if I want to compare them with independent trials, on this side, I will write down, it is m choose K, p1 power K into 1 minus p2 power m minus K. On the other side, I will write down, m choose K, p power K, and 1 minus p1 to the power m minus K. This is a (())(33:33) on both sides.

This is something that the mathematical computation one can do and one can show this, this I will, I will write down an answer. I will try and see if I can prove it today. If not, maybe today is too late, may be next class I will prove it. But that is the idea that this is a computation one can use to understand two things.

So, this is, this is kind of the comparison to independent trials. So, your, if you want to know your sampling with replacement, that is, that is your dependent trial, your dependent trial, these are not quite independent, because the p, if the p was equal to p1, if p was p1 and equal to p, then this is Bernoulli, this is like Bernoulli here.

Bernoulli p, p1, here if p1 is equal to p2, this is like Bernoulli here p2. I will say Bernoulli p1 (())(34:43). So, then the question is when, when is this true? So, when is this, when is this true, this? So, now the key thing is that when are your to, your dependent trials are going to be like independent trials.

So, when will p1 be like p2? So, the idea is that little m is small compared to capital N, then p1, p2, and p are all the same. So, the key is, the key factor is that if, let me go up a little bit, we can see the screen, very good. So, the key idea is that if N is much larger than little m, and if, if K is, of course, much larger, is much less than K, then this will imply, this, this will (())(35:45) that p1 is the same as p2, is the same as p.

So, in some sense your dependent trials can be thought of independent trials if, N is very, very large compared to little m. And that is actually what they do in statistics, they do not worry about with replacement, without replacement. They just think about independent trial. So, you would not pick the same person twice, even though you are actually doing that, but independents and dependents are all not the same.

This proof I might leave as an exercise and assignment, and I will ask you to prove it, but the proof is not so hard. It, it just follows from what I did before. Then I tried to understand how hypergeometric was a was dependent trials with different problems. Maybe I will stop, I will stop here.