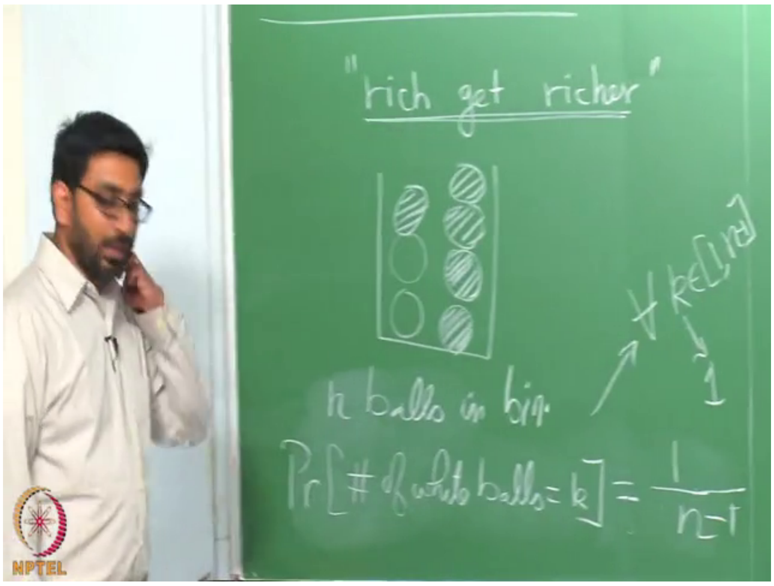
Algorithms for Big Data
Prof. John Ebenezer Augustine
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 08
A Simple "Rich get Richer" Game

In this lecture segment I am going to get you started on a problem.

(Refer Slide Time: 00:30)



Problem number 1.6 in Marker and. At this problem, models a phenomenon that we often see in real life, called the 'rich get richer'. And here is how this problem works.

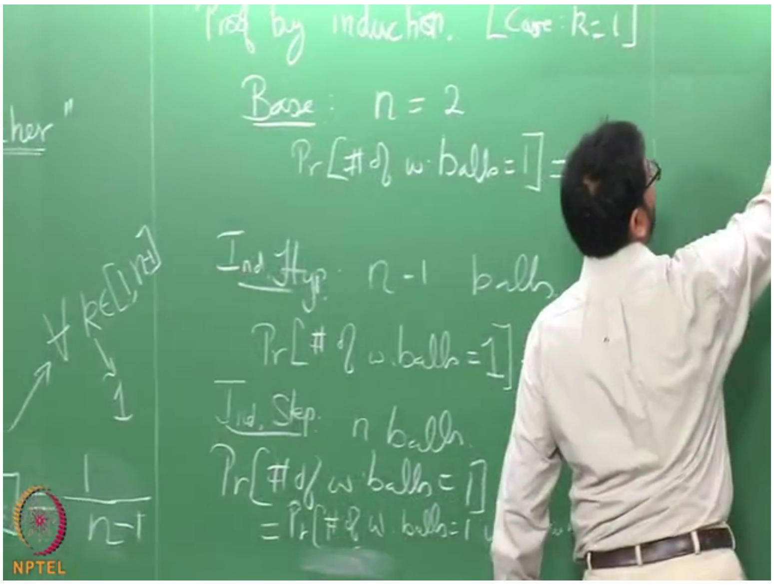
We have a bin and at the start of time have 2 balls in this bin; one is a black ball and one is a white ball. And each time step we pick one of the balls in this bin uniformly at random, and then if it happens to be a black ball, we will replace the black ball and add a black ball. If we have pick the black ball, then we would have replaced this black ball and added a black ball. Now let us say the next time step, we pick the ball uniformly at random and we pick the white ball, then we will replace this white ball and add another white ball.

So, this is the way this random process precedes and as you can imagine overtime due to the randomness in this process one color might exceed the other color. And once that happens, now let us say over the course of time, there are several more white balls than there are black balls, then, you can see how the rich get richer phenomenon comes into play here. Because now, you are more likely to pick a white ball and then once you picked a white ball you going to replace that white ball and add a new white ball, so the number of white balls is going to increase. So, this is the, in that sense it matches this phenomenon that I just talked about.

This question, problem number 1.6 asks us to prove something about this process. This random process now proceeds until we have a total of n balls in bin. And we need to prove something about this process at the end, when we have n balls in the bin. In particular we want to prove the following statement. This statement must hold for all k through range 1 through n minus 1. In particular, we want to be able to prove that it is equally likely that the number of white balls be - 1, then the number of white ball be - 2 all of these events are equally likely events. And this is what we need to prove according to this problem.

So, let us see how this will work. But I am only going to get you started on this problem. What I am going to prove is the case where k equals 1. Then using the idea that I used to prove this k equal to 1, you have to, your job is to extend this to the case where k can be larger than 1. For this we can use proof by induction.

(Refer Slide Time: 04:47)



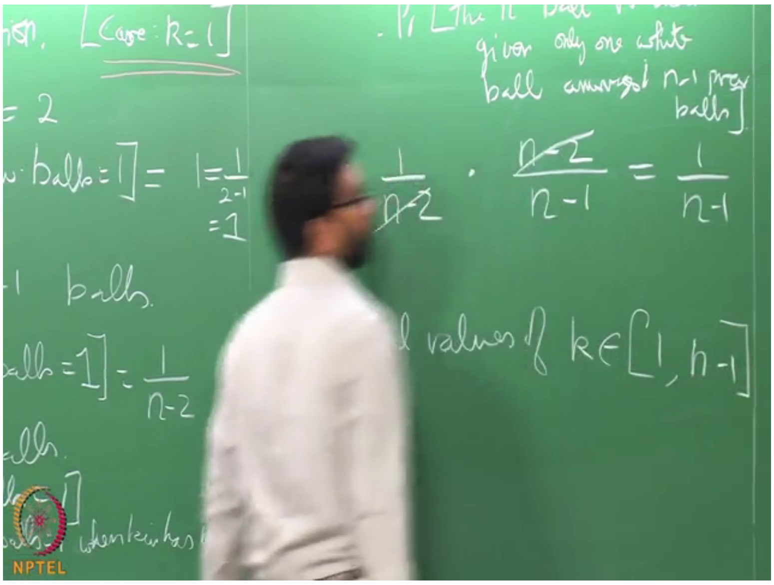
So, our base case is the following that is at the end when n equal to 2. Here of course, probability that the number of white balls equal to 1, and number of k we are gone assume the case where k equal to 1. The number of white balls is equal to 1 is equal to 1, because we start out with exactly 1 white ball. And this matches our 1 over 2 minus 1 equal to 1 and so the base case holds.

Now what we are going to do are inductive hypothesis is the following; we are going to assume, let us say that there are, n minus 1 balls and when there are n minus 1 balls our assumption is going to be, the probability and the number of white balls equal to 1, that is equal to 1 over n minus 2. This n minus 2 is coming from the fact that we want to prove this probability to be able to form 1 over n minus 1 and here it is n minus 1 ball, so we have this take n minus 2 here.

So, this is our Inductive Hypothesis. And now we are ready for Index Step. For this of course, we have to work with n balls. And let us look at the probability, that the number of white balls is equal to 1. Now think about it, how you can have 1 white ball after the total number of balls in the bin has reached n.

Well, two things must happen think about the very last step, prior to the last step, you should, you will have to have at the least 1 white ball because you cannot start with 0 white balls and then you should remain at that 1 white ball. In other words, in the last when you had only n minus 1 balls, there must have been exactly 1 white ball, and in the last step when we reached n balls in the bin, you should not have added a white ball, you should not have added a black ball. When these two conditions hold, we will be ending up with exactly 1 white ball.

(Refer Slide Time: 09:01)



And probabilistically speaking here is how we can write that, so this probability equals the probability that the number of white balls equal to 1 when bin has n minus 1 balls times the probability that the nth ball is black, given only one white ball amongst n minus 1 previous balls. This is the event that you have exactly 1 white ball after you have reached n minus balls in the bin, and then the last ball that is get added, given that you only have one white ball amongst the n minus 1 previous balls, we are interested in the event that the nth ball remains black.

So, let us work this out. What is this probability? Well, here we can take advantage of the inductive hypothesis. This probability is 1 over n minus 2, times this probability here, what is the probability with which the last ball will be a black ball, given that there are a total of n minus 1 balls, out of which n minus 2 are black. So here, the moment you pick a black ball you going to replace that ball and add a black ball, that is what going to happen, if this probability is really the probability that you pick a black ball. So, times there are n minus 2 black balls remember, this is exactly 1 white ball among a total of n minus 1 balls, because these two cancel out and you come out at 1 over n minus 1 which is precisely what we need to prove.

But of course, we have only solved the case where k equals to 1. What we really need to be solving is, general values of k. How do we solve that? Well, the same idea can be extended, I am going to leave this part as an assignment for you to solve. The same technique can be applied except now we need to be a bit careful, in order to achieve a final goal of just 1 white ball we needed exactly 1 white ball in the previous step as well. But let say k equal to let's say 5, talking about the case where we have interested in the probability of achieving 5 white balls at the end of this process.

But keep in mind you can achieve 5 white balls in two ways; one is you could have achieved the 5 white balls at step n minus 1, when the number of balls was n minus 1 and then added a black ball which we regain the 5 white balls. Alternatively, you could have reached the case where there are 4 white balls after the bin has n minus 1 balls and then when the nth ball is added a white ball is picked and therefore the new ball also becomes white. And therefore, the 4 white balls become 5 white balls. There is two ways in which you can achieve the number 5. So, that is something to think about and keep in mind while you work on the general values of k which will be your assignment.