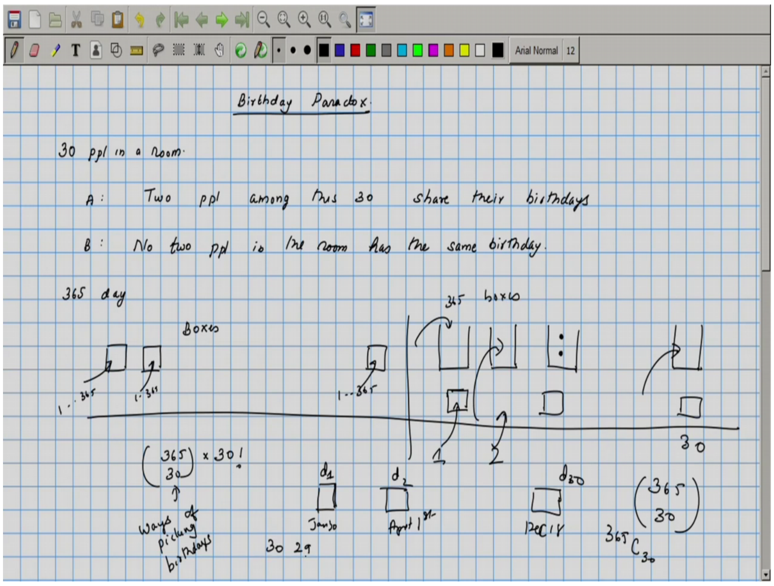
Randomized Algorithms
Benny George Kenkireth
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture – 07
Birthday Paradox

(Refer Slide Time: 00:25)



In this lecture, we will learn about what is known as Birthday Paradox. So, here is the question. Suppose you have let us say 30 people in a room, what are the chances that two of them share their birthday? Now a words, we can ask the following question we describe two events and we have to guess which of these two events are more likely. So, event A is that two people among this 30 share their birthdays and the other event is the complement of this event that is no two people in the room has the same birthday. Which of these two events is more likely?

Since they are the complement events, one of them is going to have probability greater than half; the other one is going to have probability less than or equal to half ok. We want to know which one is more likely think about it for a minute. So, we need to model this problem carefully. So, let us first model the problem carefully. So, there are 365 days in a year, we will ignore leap year for the moment and each person is being randomly assigned a birthday; that means, any day is equally likely to the birthday of any person that is there in the room. So, under this assumption, how do we compute the probability that two people share their birthday? Two people could share their birthday in many different ways like three people could have the same birthday 30 people all of them could have the same birthday, 15 of them could have one birthday and let say the remaining 15 could have another day as a birthday and so on.

It is a very complicated thing to count, but the complement event is rather easy to count. So, we will basically count the complement event and based on that determine the probabilities. So, you could think of this way. There are 30 boxes corresponding to people and each box you put a number and this number could be any number from 1 to 365 and each number is equally likely to be chosen to go into the box.

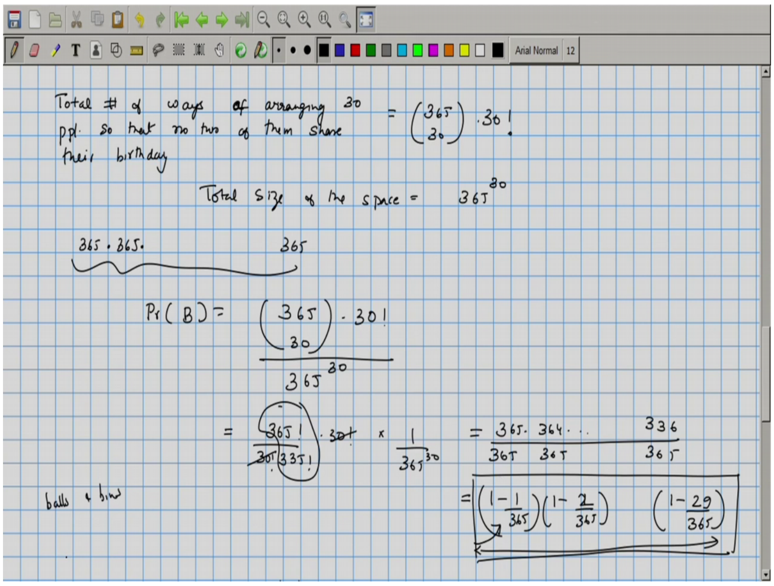
Another way of thinking about the same thing is you have 365 boxes and each person has been put into one of these with equal probability and we want to compute the probability that, there is a particular box with two people ok. They are essentially recasting the same problem in different ways. So, we are interested in no two people going into the same birthday ok. So, how many ways that can be done?

So, there are 30 people. So, if we want all of them to have distinct birthdays, we should first choose 30 days and for these distinct choices of 30 days, we can assign any of the 30 people to any of the 30 rooms chosen. So, 365 choose, 30 is a number of ways of picking birthdays and once you have picked the birthdays. So, let us say these are the days we have picked; we could either assign the first person here and second person, thirtieth person here or any permutation of this 1 to 30 for these rooms chosen or for these boxes chosen ok.

So, 365 chose 30 into 30 factorial is the total number of ways of assigning the 30 people in the room to 365 days. So, that no two people are assigned the same day ok. Once again we first choose 30 days. So, let us call this is d 1 d 2 d 30; let us say these days are numbered I mean. They are the increasing order. This is this is d 1 is the first d that might be let us say Jan 30th, this could be April 1st and this could be December 18 then so on ok, but they are in the increasing order.

So, these 30 days were chosen and now we have to assign people to these days. These 30 days could be chosen in 365 chose 30 different ways. 365 C 30, those many different ways. And once these days have been picked, who gets assigned Jan 30th has birthday that can be decided in 30 ways; who gets assigned April 1st is the birthday that can be decided in 29 ways. There are 30 factorial ways of assigning birthdays to people ok.

(Refer Slide Time: 06:29)



So, that the total number of arranging people so 30 people; so, that no two of them share their birthday. This is going to be equal to 365 chose 30 into 30 factorial. The entire sample space that is the total ways of picking us, the first person could have the any of the 365 days, the second person could have any ofthe 365 days into so there are 30 such people. So, total size the sample space is going to be 365 days to 30.

So, probability of the event B is going to be 365 chose 30 into 30 factorial divided by 365 days to 30. We can try and simply this,this is just 365 to 30 will be 365 factorial divided by 30 factorial into 3 35 factorial into 30 factorial into 1 by 365 days to 30. So, these cancels and whatever we have here as is following factorial.

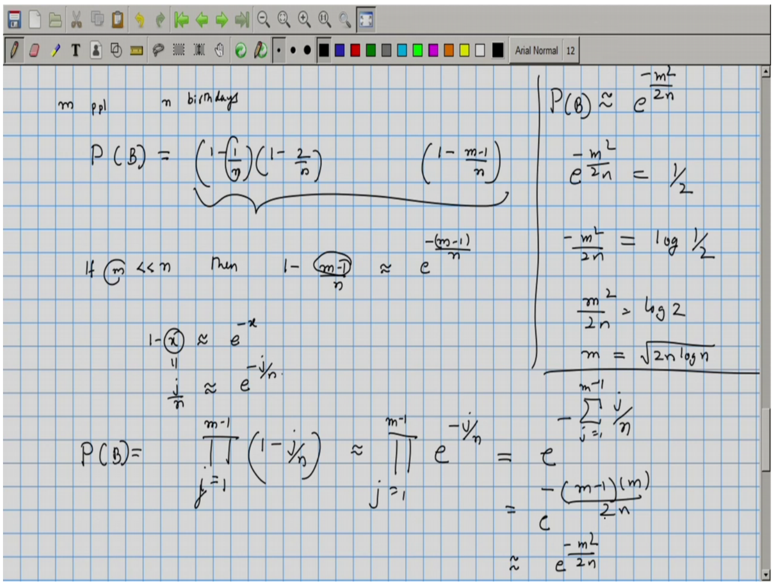
So, that is going to be so 365 into 364 all the way we have to 336 would be there above because all the other terms will cancel with the denominator 335 factorial and then there are. So, these are 30 terms and there is another 30 terms which is which are all 365. So, the first term is 1 by 1 that goes away. The remaining is 1 minus 1 by 365 into 1 minus 2 by 365 into 1 minus 29 by 365 ok. There are 29 terms; the first term been 1 by 1 that is vanished.

so this is a probability. Now that is a difficult thing to compute, but we can see this answer in yet another way. We want no two people to share their birthday. Well the first person whoever it is, he can be assigned some birthday. There is no clash and the second person if his birthday has to be a on a different day that could be done in 364 by 365 days or that one out of this 365 that birthday cannot be chosen that is his 1 minus 1 by 365.

Now once the first two people have been assigned birthdays, there is a 2 by 365 chance that the third person might share the birthday the probability that does not happen is going to be one minus 2 by 365. So, that is the telescoping product. So, 1 minus 1 by 365 and 1 minus 2 by 365, all we have 29 terms in this and that product will be the probability that no two people share their birthday.

Now we need to estimate this and think of this as a balls and bins problem. We had 365 bins and the balls were people we have to throw them into these 365 bins so that no two people share the same birthday.

(Refer Slide Time: 10:33)



So, let us say we have m people and n birthdays and we want no two of them to share the birthday. So, those could be done in 1 minus 1 by n into 1 minus 2 by n all the way up to 1 minus m minus 1 by n. This will be the total probability that no two people share their birthday ok.

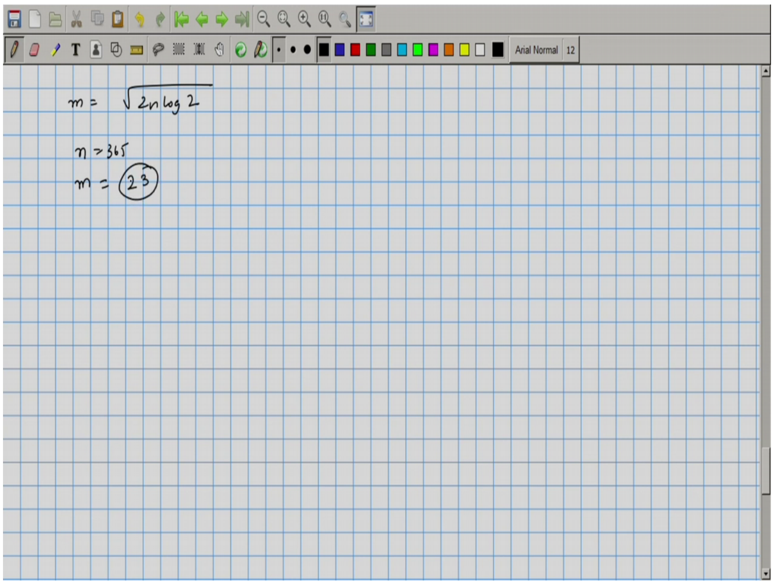
So, we can call that as probability of B, we want to estimate this. So, if m is smaller than n, then 1 minus m minus one by n or this can be written as this is approximately for small values of the numerator this can be written as e rise to minus m minus 1 by n ok. It can be done for all the not just for this for any of these terms, they can be approximated as e raise to minus m minus 1 by n. 1 minus x this approximately e raise to minus x for small values of x.

So, when x is equal to let say j by n this is approximately equal to e raise to minus j by n ok. So, the probability of B is going to be the product i going from 1 to m minus 1; let say j going from 1 to m minus 1 1 minus j by n and this is approximately product j going from 1 to m minus 1 e raise to minus j by n. And that is going to be equal to e raise to minus summation j going from 1 to m minus 1 j by n that is equal to e raise to minus m minus 1 into m by 2 n which we can approximate as e raise to minus m square by 2 n ok.

So, in the general situation what we have is probability of B is equal to is approximately equal to e raise to minus m square by 2 n. We want to know at what threshold does, what is a number of people we should have in the room so that the probability that two people have their birthdays on this same day is going to be at least half ok. So, in other words that threshold will happen at the value where P B is equal to half. So, if we equate this to half so, we want e raise to minus m square by 2 n to be equal to half; that means, minus m square by 2 n is equal to log half or m square by 2 n is equal to log 2. So, m is equal to 2 n log 2 under root ok.

So, if m is under root 2 n by log n and for this you can see that m is significantly less than n because its only root of n log n 2 times root of n log n. So, for m equal to this value the probability will be approximately half ok.

(Refer Slide Time: 14:39)



And when so, m is equal to 2 n log 2 and when you plug in n equals 365 can see that the value of m that you will get something like 23 the rounded off to the integer, it is 23. So, what it means is that if you have 23 people in a room, there is a 50 percent chance that 2 of them share the birthday. Although the total number of birthdays are much larger than 23; it is something like 365 by having this two 23 people in the room, we have a significant probability that is the 50 percent chance that two of them will share their birthday

Will stop here and we will learn about tail bounds in the next week.