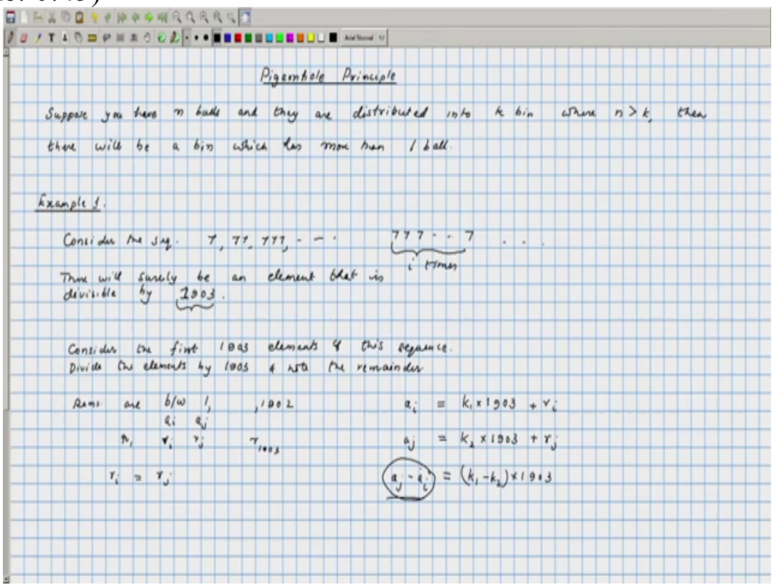
Discrete Mathematics

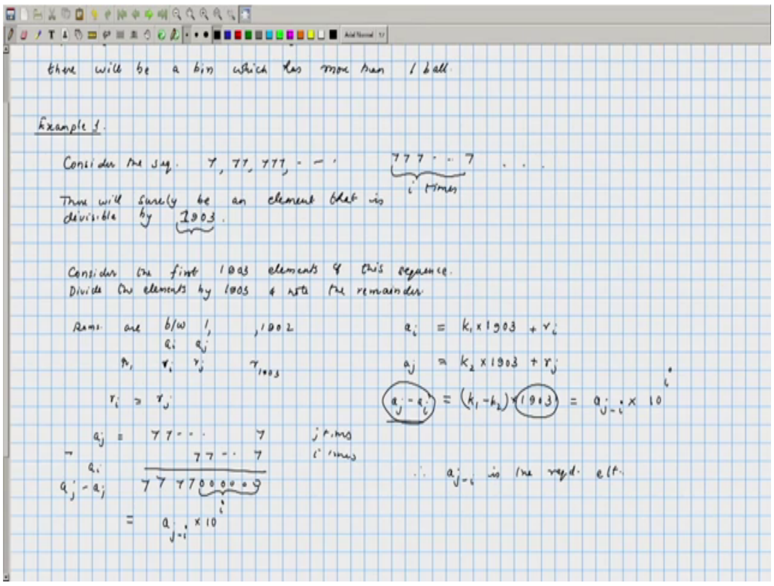
Professor Sajith Gopalan
Professor Benny George
Department of Computer Science & Engineering
Indian Institute of Technology Guwahati

Lecture 27 - Pigeon Hole Principle

This lecture we will see a very useful principle from Combinatorics known as pigeonhole principle. The principle is extremely simple, so simple that it is difficult to imagine that this can be of use. Ok, the principle is very simple to state.

(Refer Slide Time: 0:45)





Suppose you have n objects, say n balls and they are distributed into k bins. If n is greater than k then there will exist a bin with more than 1 ball. So, that is the principle. So if you have n balls and they are distributed into k bins where n is greater than k, then there will be a bin, at least 1 bin which has more than one ball. It is almost self-evident, if the number of bins is small and if all of them had less than or equal to one ball then the total number of balls is going to be only, it is going to be less than or equal to k but the number of balls is greater than k and therefore the statement must be true.

So let us, let us see some applications of this. So, pigeonhole principle of course is very simple to state. The power of this principle comes from the wide variety of ways in which you can set up balls and bins. So, let us look at a problem which can be solved using pigeonhole principle. So, let us look at this sequence. So, consider this following sequence, consisting of numbers 7, 77, 777 and so on. Ok, so the ith element of the sequence would be 777 and so on i times, ok. Now, let us look at the theorem or the results states that there will surely be an element in the sequence which is divisible by, let us say 1903.

Ok, so one element is this. There will be an element in the sequence which is divisible by 1903. Now, how do we prove this? What is the method? This statement is even true. Why should there be an element in the sequence consisting of 7s, the ith element being 7 repeated i times? So one of this is going to be divisible by the number 1903. There is nothing sacrosanct about 1903. There is a wide, you can substitute it with a lot of other numbers. The few numbers that you cannot, you can look at the proof carefully to figure out what all numbers can come in place of 1903, ok.

So, let us try and work out a proof for this. So we will set up the balls and bins in a certain way such that we can argue that there will be a bin which contains more than one ball and that is going to be used to construct the proof of divisibility, ok.

So, here one of these numbers to be divisible by 1903. So, what we can do is try and divide these numbers by 1903. You will get infinitely many numbers. So let us just take let us say 1903 elements of the sequence, ok, so 7, 77 and so on. We will look at the first 1903 elements of this sequence and divide each one of them by 1903, you will get some remainder, ok. If any one of them is 0 then we know that we have some numbers in the sequence which is divisible by 3. What if none of them is 0? Ok, so let us say k is that we are interested, so let us write down the steps.

Consider the first 1903 elements of the sequence. So, divide the elements by 1903 and note the remainder, ok. If any of the remainders is 0 then we have done, we have a proof that look one of the, we have an element which is divisible by 0. So, we may assume that all the remainders we get are numbers between 1 and 1902. So, remainders are between 1, both numbers inclusive 1902, ok.

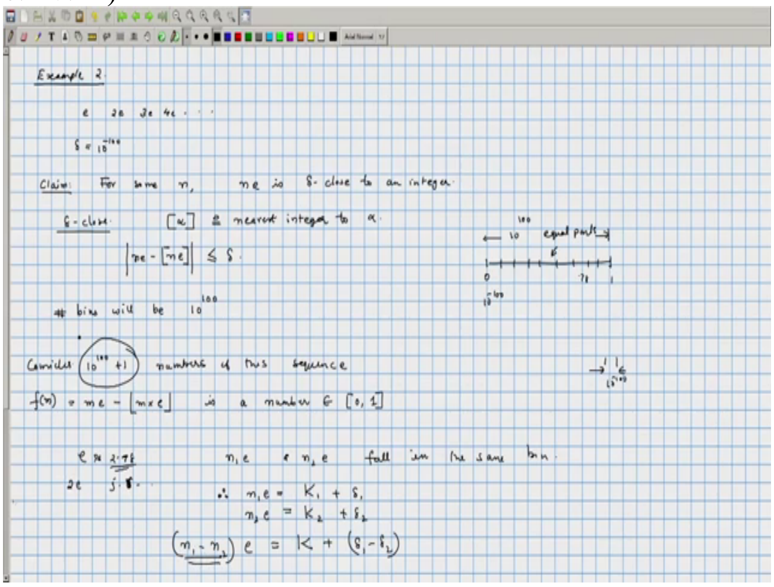
So these are 1902 numbers, the remainders, and we have 1903 numbers in total, ok. So if you denote the remainder by R 1, R 1903 so on, these are 1903 remainders and each of them is one of these numbers between 1 and 1902. So one of these remainders must repeat, so let us say a i leaves remainder r i and a j where j is greater than i leaves remainder r j. So, we have this case that r i is equal to r j, ok. So, what we have is a i is equal to some number k 1 into 1903 plus r i and a j also has a similar equation, this is equal to k 2 into 1903 plus r j.

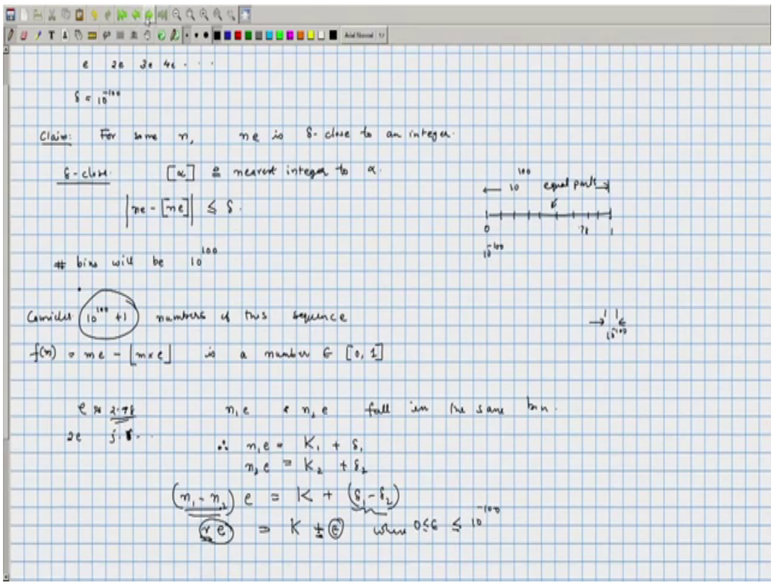
We have r i and r j as the same and therefore if you do a j minus a i what we will get is k 1 minus k 2 into 1903, ok. Now, this 1903 therefore must divide a i minus a j, ok. We are almost done, what we wanted to show is the reason element in the sequence which is divisible by 1903 and not the difference of two elements, ok. But this is good enough for us because what is a i minus a j?

So if you look at a j, a j is equal to 777….j times and a i is equal to 777….i times. If you subtract them, what you get is a j minus a i that is going to be 777 and the last items are going to be 00000, ok. So this is equal to a j minus i, which will consist of j minus i 7's into 10 to the power i, because there are i 0s, this equation is true. So, we can write k 1 minus k 2 into 1903 is equal to a j minus i into 10 power i, ok. So, now if you look at 1903, 1903 does not have any common factor with 10 power i, because the only factors of 10 power i are 2 and the only prime factors of 10 power i are going to be 2 and 5 and they do not appear in 1903.

So 1903 must divide a j minus i and a j minus i is one of the elements in the sequence, ok so we can say therefore a j minus i is the required element, ok. So, that was one application of pigeonhole principle. So, the balls or the pigeons where the remainder that is left by each element of the sequence when divided by 1903 and the pigeonholes are the bins where numbers from 1 to 1902, we took 1903 remainders to fit in 1902 bins. So, now let us see another application of pigeonhole principle. So little more involved application.

(Refer Slide Time: 10:12)





So, here we start with a, with an irrational number, so let us say e, you can replace it with any other irrational number and we consider its multiples, ok so 2 e, 3 e, 4 e and so on. So, this is the sequence that we are considering. Instead of the irrational number if you had started off with a rational number then we know that at some point of time we will get an integer, ok. When we start with an irrational number and if you consider this sequence e to e 3 and so on, the multiples of e. So, we know that none of these elements in the sequence is going to be equal to an integer but what we will show is one of these elements is going to be arbitrarily close to an integer.

There will be some element which is you give a measure of closeness one of those elements is going to be arbitrarily close to some integer, ok. So, we will make the, I mean we will make the statement little more concrete. So, let us say a degree of closeness we will call this 10 to the power minus 100, ok. So we will, what we want to, this is what is delta we will claim the following for some n, n times e is delta close to an integer. So, delta close means, so if I write the square bracket alpha, this will denote the nearest integer to alpha, ok. And if it is exactly midway then we will choose either one of them. In that case let us say we will choose a larger one, ok.

So this is a definition of the nearest integer. So what we want to say is n e so n e minus n e, ok. The number minus its nearest integer its absolute value is going to be less than or equal to delta, this is the claim. I mean this is a definition of delta close and what we are claiming is for some n ne is delta close to an integer, ok. So how do we prove this, what should be the balls and what should be the bins? Ok.

So let us look at each of these elements e to e 3 and so on. Here we have delta is equal to 10 raised to minus 100, ok. So 1 by delta is what we will choose as the number of bins. So, we have not yet told what is a bin but the number of bins will be, we are going to set up balls and bins analogy wherein the number of bins will be 1 by delta or 10 raised to 100, ok. And we are going to look at 10 raised to 100 plus 1 numbers of this sequence, ok. And each of those numbers is going to be put in one of the bins and since there are 10 raised to 100 bins one bin is going to be containing 2 numbers, ok.

Now let us decide how we are going to put these numbers in the bins, ok. So, look at any number. If you consider let us say m times e and if you look at the floor that, this is going to be an integer and m e minus floor of m e is a number which belongs to 0, 1. It cannot be 0 or 1, so we could have said the open interval 0 1, ok. But this is m e minus floor m times e is going to be some number between 0 and 1.

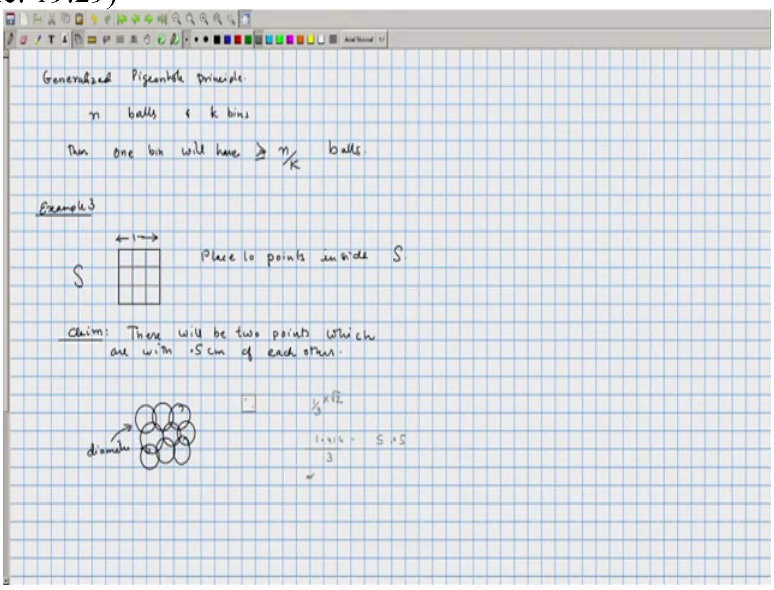
Now let us just look at this interval 0 to 1 and split it into 10 raised to 100 equal parts, ok. So let us call this as f of m, ok. So for the mth element f of m denotes the amount by which m e exceeds the integer part of the number, ok. So f m if it, now we have split this into 10 raised to 100 parts, equal parts, so each part is going to be of size 10 raised to minus 100, ok. So, these parts are going to be our bins, ok. And if f m falls in one of these part then we will say that the number falls in that particular bin, ok. So if you had let us say 2.78, e is approximately 2.78 so look at the, I mean there is some bin which contains 0.78 and that will be where the first element will be there.

Now if you take 2 e that is going to be something like 5.5 something or 5.6 something, so that is going to be fall in a different bin and so on, ok. So, each element we are going to put in one of these bins clearly since we took 10 raised to 100 plus 1 numbers, some 2 bins is going to contain the same, I mean some 2 bins is going to contain 2 numbers, ok. So, let us say n 1 e and n 2 e fall in the same bin. Therefore n 1 e is equal to some integer plus let us say delta 1 and n 2 e is equal to k 2 plus delta 2, ok. And delta 1 and delta 2 are numbers which fall in the same bin.

So if you look at n 1 minus so let us say n 1 is greater than n 2, so n 1 minus n 2 e is going to be k 1 minus k 2 that is going to be an integer let us say k plus delta 1 minus delta 2, ok. Now, delta 1 minus delta 2 both these numbers by virtue of being in the same bin of size at most 10 to the power minus 100 their difference can be at most 10 to the power minus 100. So, that means n 1 minus n 2 times e because this is let us say r e where r is some integer r e is going to be k plus this could be negative as well. So, plus or minus some number epsilon where epsilon is guaranteed to be less than 10 to the power minus 100, ok.

So plus or minus is there let us say epsilon if assume as positive and if delta 1 is greater than delta 2 then we can say it is k plus, the other k's will say it is k minus epsilon, ok. But whatever is the case r e is another element of the sequence and r e is going to be some integer plus a very small number and by very small here we mean that it is going to be less than 10 to the minus 100, ok. So, we have seen 2 cases of pigeonhole principle. Now we will look at little more, slightly more general version of pigeonhole principle, ok.

(Refer Slide Time: 19:29)



So here we will say that this is a generalized pigeonhole principle. So, if you have let us say n balls and k bins then one bin will have at least one bin will have greater than or equal to n by k balls, ok. The same reasoning as to why pigeonhole principle is true will be the reason why the generalized pigeonhole principle is true. So distributed n balls into k bins then there surely should be a bin which has at least n by k balls, because if all bins contain less than n by k balls then the maximum is going to be less than n by k into k which is less than n, ok.

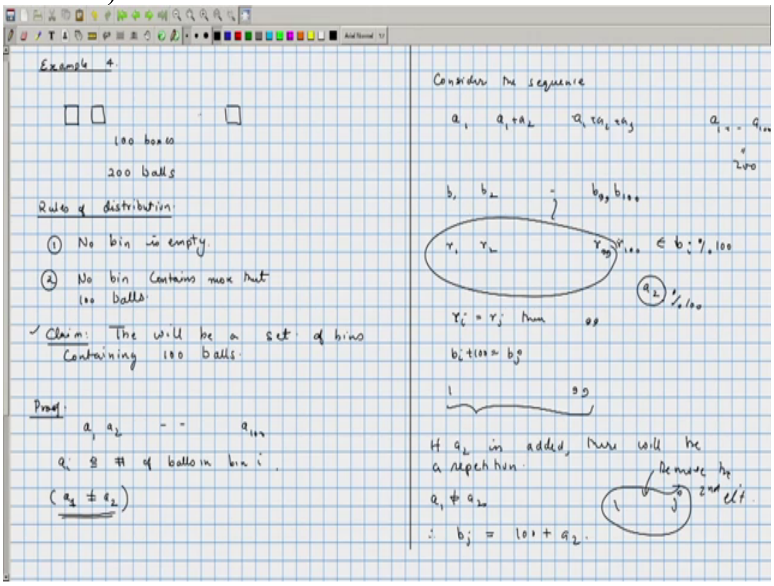
So, now let us see another example involving generalized pigeonhole principle. So, ok now we will look at an example from geometry, a one sided, I mean we have a square whose sides are 1 centimeter long, ok. And we are putting let us say 10 points, place 10 points inside S, so let us call this as a square S and we are placing 10 points inside the square, ok. If we have large number of points then there will be 2 points which are close to each other, ok that is a general statement. Here we want to say that when we have 10 points inside S there will be 2 points, so claim no matter how you place the points there will exist which are, there will 2 points which are within 0.5 cm of each other, ok.

Now how do we prove such a statement? Ok, there will surely be 2 points which are within 5 centimeters of each other, now how do we try to see this as a generalized pigeonhole principle? One way to think about it would be to take, let us say circles. So consider circles of diameter 0.5 and if you can somehow cover the entire sphere using 9 circles whose diameter is less than 5 then we know that there will be 2 points in the same circle, ok. So what we want to try and do is we want to somehow place 10 circles or 9 circles, ok so that they cover the entire region.

Now, here if you want a circle maybe that is little tricky and there is going to be lot of overlap, so it is not very straight forward to see how this can be done but there is a simpler way wherein you can split it into 9 parts, ok. So, you can look at this particular square and let us say we split it into 9 parts, natural split, ok. So it is split into 9 squares, ok. Now if you look at these 9 squares and no matter how you place 10 points there is going to be one particular square which contains 2 points, ok. And the maximum distance between any 2 points in the smaller square is going to be, so smaller square is of side 1 by 3, so this into root 2 will be the maximum distance, ok. And that is 1.414 something divided by 3 that is going to be less than, surely less than 0.5, ok.

So if we have split this in this particular manner, so you can get 2 points which are within 0.5 centimeters of each other.

(Refer Slide Time: 24:51)



We will end our discussion on pigeonhole principle by looking at one more example. So here we are looking at a commentarial problem where we have let us say 100 boxes and we have 200 balls or 100 bins and 200 balls and we are distributing these balls into these boxes following 2 rules. The first rule; no box, no bin is empty. The second rule; no bin contains more than let us say 100 balls, ok.

Now when you distribute it in this way, there are many ways of distributing it in this way. When you distribute, none all of those distributions would have the property that, there will be a set of bins containing, so when I say set of bins containing it take all the balls in those bins together whatever they contain. So the set of bins containing 100 balls, exactly 100 balls, ok. So no matter how you distribute the balls into bins, there will be a subset of bins such that together they will account for exactly 100 balls, ok. Now suppose you had let us say 700 balls and we are done this, ok.

We could let us say give 7 balls in each bin and that is never going to, if you distribute in that particular fashion, there is no way that you can get 100 by combining few bins, ok. So natural question would be to wonder why such a claim is true when you have 200 balls and 100 boxes, ok. So we will give a proof of this via pigeonhole principle, ok. Again the techniques are very similar to the techniques that we had seen. We are going to, now when we want to say that some bin contains 100 balls, ok. Even if we can say that some bins together contains a multiple of 100 balls, ok, the only allowable multiples under these conditions are 100 and 200. But if we skip some boxes then you are going to get something less than 200, ok.

So if you just say that there is a collection of bins which, I mean there is a subset, strict subset which contains a multiple of 100 balls in them then that would mean that it is exactly 100, ok. When we want to look at that multiple of 100 we could try and look at the remainders, the remainder of what? This is where the cleverness comes in. So the proof, so let us look at these boxes they will contain some number, let us say a 1, a 2, a 100 so a i denotes the number of balls in bin i, ok. We have just arranged in some particular way, ok. Now we will insist one thing on this arrangement. We will insist that a 1 is not equal to a 2, ok. Can this always be that? I mean if there are two bins which contain different number of balls in them, we could of course get those 2 bins and do this entire thing.

So, the only case where we cannot get such an arrangement is when every bin contains exactly the same number of balls but that would be the case when every bin contains exactly 2 balls. But in that case our claim is trivially true because we can just take any 50 bins and together they are going to be 100, ok. So we will assume that is not the case and therefore we have 2 bins a 1 and a 2 which has a distinct number of balls in bin 1 which is going to be different for the number of balls in bin 2, ok and those numbers we will call as a 1 and a 2.

Now, let us consider this sequence of 100 numbers. So, the first is a 1, the next is a 1 plus a 2, next is a 1 plus a 2 plus a 3 and last is a 1 plus up to a 100 and this is going to be surely 200. Ok so let us look at these numbers and ask, sorry this, the last sum is going to be 200, then total number of numbers that are there in the sequence is 100. So this is a sequence of 100 numbers, ok numbers varying from let us say a 1 to 200. Now let us divide, so let us call these numbers as let us say b 1, b 2, b 100, ok. Let me consider the remainders r 1, r 2, r 100, so this is b i mod 100, ok.

There are 100 remainders if any of these were zero then we are done, so except r 100 the last one of course is going to be divisible by 100 or the other so let us skip the 100th element, we will just look at the 99 elements, say r 99, ok. So these are 99 elements. If any one of them gives a zero then we have a set of bins which contain 100 balls. So we can assume that none of them is zero and therefore they are numbers within 1 to 99.

If none of them is zero then they have to 1 to 99. Now let us ask this question, is there any repetitions in them? There could be repetitions but that would be the good case for us. So if r i equals r j that means the b i's are increasing in order, ok. So if r i equals r j then b i plus 100 is equal to b j, ok. Therefore if you look at the numbers or look at the bins numbered let us say i plus 1 to j they are going to be of sum 100, ok. So, now that would validate our claim, so we may assume that is also not the case. That means all these numbers are distinct and therefore they are numbers from 1 to 99, ok. Now what do we do? Ok, here is where the assumption that a 1 is not equal to a 2 comes handy, ok.

So let us look at these numbers and we will add a 2 into the mix and if you look at a 2 mod 100 that should be now you have 100 elements, ok, a 2 itself is not 100 because we said that no bin contains greater than or equal to more than 100 balls. If a 2 was exactly 100 we have one bin with exactly 100, ok. There should be some sum here which is exactly equal to a 2, ok. So if a 2 is added there will be a repetition, ok. And that repetition without a 2 there was no repetition, so that sum is going to be, so that remainder is going to be a 2, ok. And by virtue of a 1 not equal to a 2 so this is not equal to b 1, it is going to be some b j. Therefore b j is going to be equal to 100 plus a 2. Since b j is equal to 100 plus a 2 if you look at the set consisting of all elements from 1 to j and remove the second element, we will get a set which contains exactly 100 balls, ok. So that concludes the proof. We will stop here.