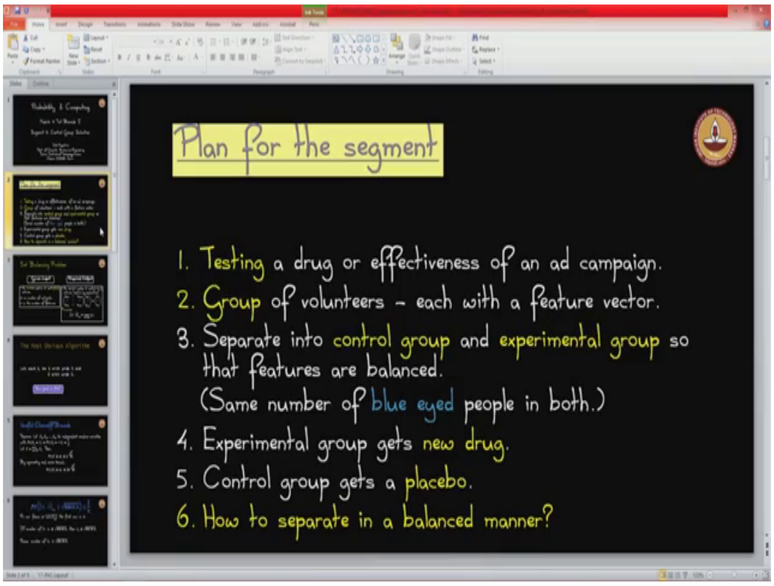
Probability and Computing
Prof. John Augustine
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Module – 04
Tail Bounds I
Lecture – 21
Segment 2: Control Group Selection

So let us start with the second module I mean a second segment of module 4, and it is about how experiments are done in the real world, especially in the context of drug design or trying to understand the effectiveness of a campaign or something like that.

(Refer Slide Time: 00:21)



So, basically what you do is, you collect a group of volunteers they are also called subjects in the experiment and this is typically people who consent to this experiment.

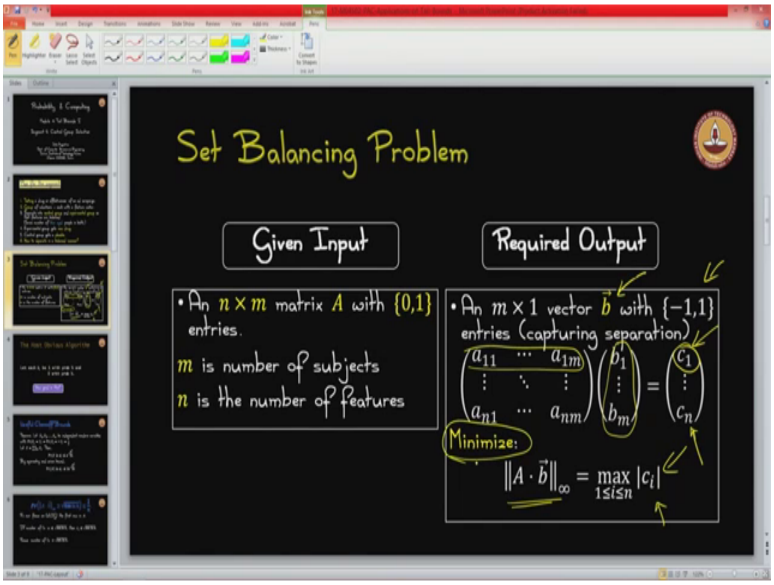
And you can what you want to do is, somehow separate this group of volunteers into the control group and the experimental group ok. And you want to administer the drug to the experimental group, you want to administer a placebo basically what looks like the drug, but is just you know an inert substance to the control group, and you want to be able to say well the there has been significant improvement in the experimental group compared to the control group.

Even before we go into interpreting the outcome of the experiment and all that, the important prerequisite for getting things right is that we need to ensure that the separation between the control group and experimental group is done appropriately. Because suppose for example, you are testing the drug and somehow your control group people have certain feature. Some drug let us say acts better with people with blue eyes versus those who do not have blue eyes and suppose all the people who blue eyes get put in the control group, and the people with non blue eyes get put in the experimental group, then you are not your experiment is going to give the wrong result it is going to say that the drug has no effect whereas, in fact, it you might have been able to figure out later on that you know the does have some effect.

So, basically what you need to do is to try to characterize each individual, and what we do is we characterize them by a feature vector is a. So, let us say you have some n features height you know let us think about binary features. So, it is just blue eyes versus not blue eyes curly hair versus non curly hair tall versus short what not ok.

So, you characterize each subject by this sort of a feature vector, and you want to make sure that for any feature there are sufficient number of people in the control group and a sufficient number of people in the experimental group. So, you do not have you want the separation to happen. So, the to avoid situations where a particular feature is only found in the control group or only found in the experimenting ok. So, this is the context in which this problem is study.

(Refer Slide Time: 03:26)



So, let us try to formalize this. So, you are given an input matrix A it is an n cross m matrix, m is the number of subjects. So, each column corresponds to some person let us say, and n is the number of features ok. So, each feature could be you know blue eyes or tall or whatnot right and if the feature is present you put a 1. So, far if the ith if the ith feature of the jth person is present, then you put a one there otherwise it is 0. So, that is how you interpret this matrix n cross m matrix and what is your required output? it is basically this vector b whose values are either minus 1 or 1. Minus 1 means that that person is going into the control group and plus 1 means that the person is going to go into the experimental group. So, and how do you ensure that the separation is good let us let us take one feature for example, let us take the first feature.

Basically when you multiply this matrix a with the vector b you are going to do the dot product of the first let us say the first row with the with vector b ok. And by doing so, you will get c 1 and so, whenever there is all those people who do not have the feature, they are not going to participate I mean they are not going to contribute to this dot product.

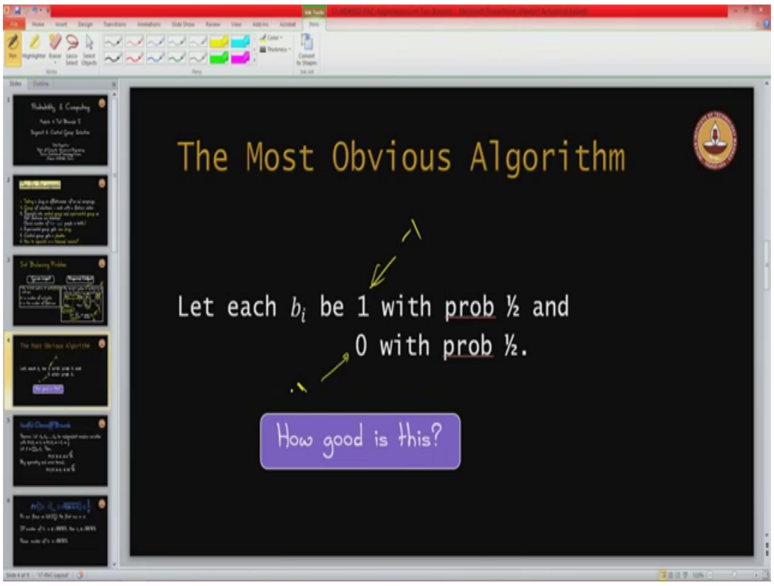
The people with the feature are the ones who are going to contribute to this dot product. And in that case the if look let us look at the perfect case, where those who can do those who contribute in the control group and those who contribute in the experimental group are both equal then what will the c 1 value be? it will be.

Student: 0

It will be 0 because it will be minus 1 times the number of people who contribute to the control group, one times the number of people who contribute to the experimental group. So, the best value you can hope for c 1 is 0, but that is too much to hope for we want; however, to minimize this c 1, but we do not just want to minimize c 1 we want to minimize c 1 this is just one feature you want to minimize across all the features. So, in other words you want to minimize.

The max over all the c is their absolute values because this remember it is 0 is the perfect value. So, it can air on the left side or on the right side. So, you have to take the absolute value and minimize that absolute value across all the c is and how do you how do you get this max over all and the absolute value of c is? It is basically taking the matrix a times the vector b and taking the l infinity norm of that ok. So, you are going to get them by doing this multiplication you are going to get our matrix and you are going to pick the that is essentially this matrix this vector c, and picking the maximum element is the l infinity norm and this quantity has to be minimized. So, that is the goal.

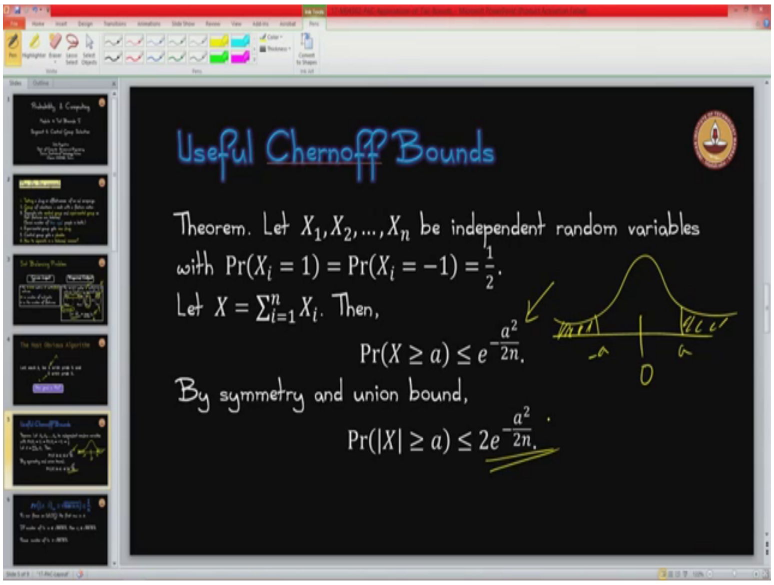
(Refer Slide Time: 07:36)



So, again we are going to apply the most obvious algorithm ok, we are going to do the silliest thing possible, each bi remember bi is either plus or minus 1 is going to be 1 with sorry.

Before someone points it out minus 1 with probability half and 1 with probability half ok. So, this is the most obvious algorithm that I do not think of. So, we need to figure out a way to analyze this. So, how good is this such a silly looking algorithm. So, we have to figure out whether it is any good.

(Refer Slide Time: 08:30)



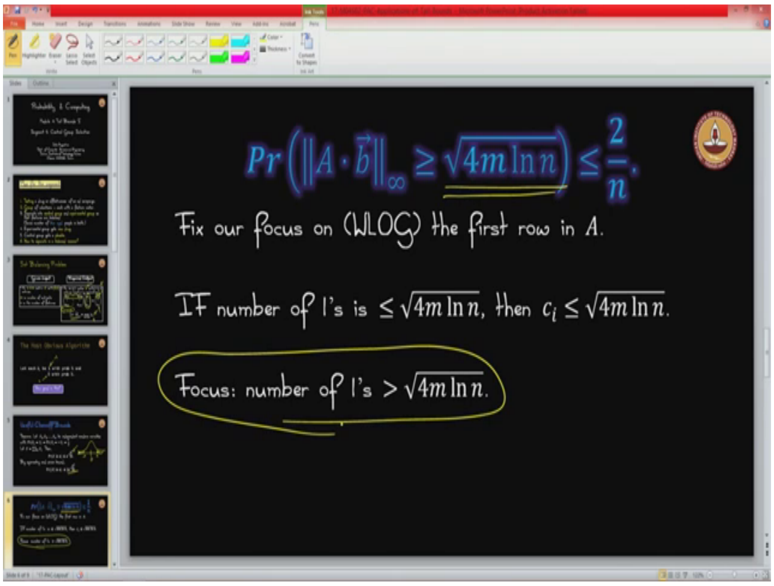
So, again we are going to apply Chernoff Bounds, and we are going to apply another variant that we have not seen, but it is actually not hard to prove this the book has the proof of this, this variation as well. So, here X 1 through X n are either 1 or minus 1 with probability half each, again uniformly and independently at random and upper case X is the summation of all the individual X i s and so in this case what you have is the mean is 0 ok.

So, it is either plus or minus 1. So, the mean is going to be 0 and you are asking what is the probability that x is going to exceed some a that is what is being asked over here, and what is been proved is that it is at most e to the minus a square over 2 n ok. And if you want to do if you want both want to bound both sides of the tail both tail ends, then it is at most two times e to the minus a square over 2 1. So, it is fairly straight forward.

Student: (Refer Time: 09:54) them is divided by 3 right side is given 2.

Correct. So, here the proof is a little bit different. So, here this is remember this is symmetric. If you just change if you notice that if here you are you are there, you are looking at 0 one variables here you are looking at minus 1 1 variable. So, all the variables is symmetric about the 0 point. So, when you prove a tail bond on the right side the same tail bound will hold on the left side as well, exact same tail bound will hold on the other side. So, you can simply use that symmetry and the union bound to get this 2 times e to the minus a square over 2 so.

(Refer Slide Time: 10:35)

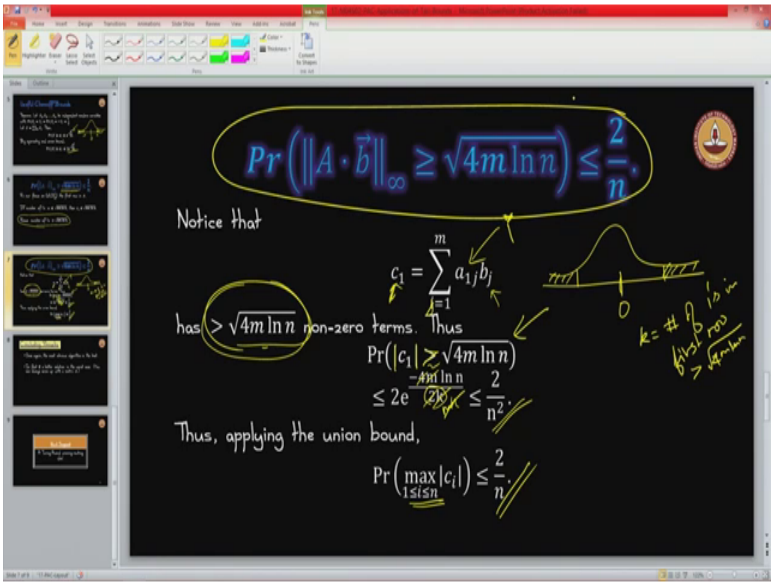


So, now just keep this bound, in mind and we are just going to apply that. So, re call what do we need to prove? We need to prove that this the l infinity norm of eight times being greater and here this is what we want to prove in particular, greater than square root of 4 n lon n ok. So, basically lon n is relatively small essentially what we are saying is that most of all the people all the for every feature the number of people who might be more in the control group with a particular feature that might be more in the control group versus the experimental group of vice versus at most roughly square root of m ok. So, that is the way to interpret it.

So, we want say that exceeding that 4 n lon n is at most 2 n 2 over n ok. So, this is across remember this is across all the all the n features right. So, what we are going to do is focus on one feature first and. So, without loss of generality let us focus on the first feature. So, this is just row 1. If the number of ones in that feature is small, it is itself within 4 m square root of 4 m or lon n then we are immediately done because; however, if you split it if the number of people who are tall are at most square root of 4 m over lon n.

And; however, you split it is going to be equally split and I mean it is not equally split I am sorry it is the number in the control group versus the experimental group can will not exceed that. And in a sense this is fine because it is a feature that affects very small number of people. So, then you immediately get what you want that appropriate c 1 in this case, is at most square root of 4 m over lon n 4 m lon n. So, it is you are done. So,. So, we are going to focus therefore, on the case where the number of ones is greater than this quantity ok. Only then do we have to care about making sure that both the control I mean the control group and the experimental group are roughly equal ok.

(Refer Slide Time: 13:23)



In this case let us look at c 1, c 1 is nothing, but this dot product right summation over i equal to 1 to m a 1 sorry. So, this is just make that it j a 1 j bj ok. This is and we know that this at least the number of positive terms. So, the number of terms here that there are once is at least 4 m over lon for 4 m lon n square root of 4 m lon n ok. So, now, what is the probability that this c 1 is greater this greater than this quantity.

Remember c 1 now is these bjs. So, let us let us be a little bit careful about what is going on over here, when the term a one j is a 0 this is the random variable bjs are the random variables this is the ones that your algorithm chose randomly a 1 j was given to you, it was either a 0 or a 1. If it was a 0 it is not going to play any role you can ignore that particular jth term ok. The only terms that matter are the terms where the a 1 j is equal to 1 and that is at least these many terms and you want to still make sure that the c 1 remember now the c 1 is the summation of all of them some bjs are paused plus 1 some bjs are minus 1.

So, the expectation is still going is going to be 0 and you still have the same thing happening over here, you will have a some distribution with mean 0 and each individual term that matters is either plus 1 or a minus 1 and you have set yourself some bounds to say, what is the and you are asking what is the probability that you are going to lie on either side of the bounds that is that is what is being asked over here.

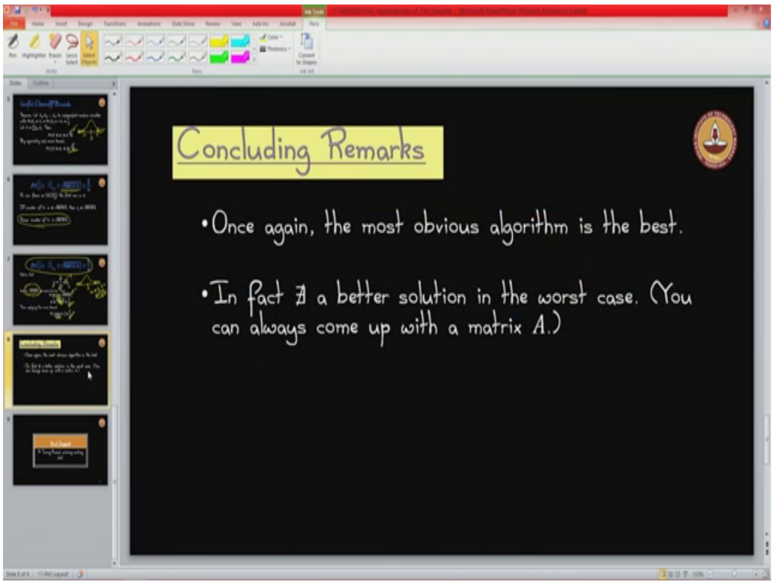
So, we can simply apply the formula, it is 2 e to the and this there is a square root on the right hand side of this event you just any the formula has a square in the exponent you have minus 4 m lon n, divided by white 2 k here what is k? K equal to the number of ones in first row which is greater than square root of 4 m lon n ok. So, that is the k over here because those are the only ones that matter, but again now if you just want an upper bound, you can just replace the K by n and still fine and what is that going to work out to. So, if you if you make that an n.

Student: (Refer Time: 17:00).

Oh ok. So, oh yeah you are right. So, you can make it an I am sorry not an n. So, ta was the one who pointed that out. So, you get a day off. So, the m cancels out we the m cancels out, you are left with and there is a 2 here that can. So, you get 2. So, you basically get two times n to the minus 2, which is what we have over here ok. So, it is 2 over n squared, but now we want to we what we have done is we have only bound this probability for one feature there are a total of n features.

So, again apply the union bound over all ends. So, then you multiply this bound by n. So, this is this is why you carefully doing this reverse engineering knowing that you want it to be of the form 1 by n, you want to bound each individual probability to be 2 over n squared. So, that the n and the n squared cancel out to be. So, you are left with 2 over n all right. So, that is that essentially proves this theorem for us ok.

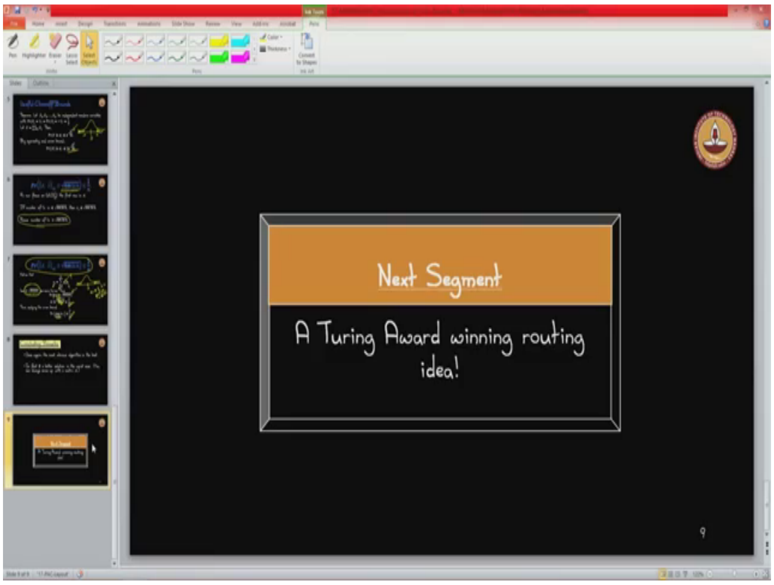
(Refer Slide Time: 18:33)



So, that we can conclude again we have seen a very fundamental problem shows up all the time, and there is an obvious algorithm in fact, practitioners are using this all the time, but what we have done is given a formal analysis and given a proof that that obvious algorithms are. In fact, actually a good algorithm in and if time permits we will actually address some techniques where we can actually show that does not exist a better solution in the worst case.

In other words you can always come up with input matrices, where there is there does not exist any significantly better b vector. So, you cannot come up with a better output. So, this is what this is the best that you can essentially hope for ok.

(Refer Slide Time: 19:29)



So, with that we will conclude this segment, the next segment it will be little bit more non trivial, but at the same time a lot more fun as well in some sense, it is going to be a topic on routing and it is one it is a contribution by Leslie valiant who recently won the Turing award for this and a few other similar such contributions. So, that will be next segment ok.

Thank you.