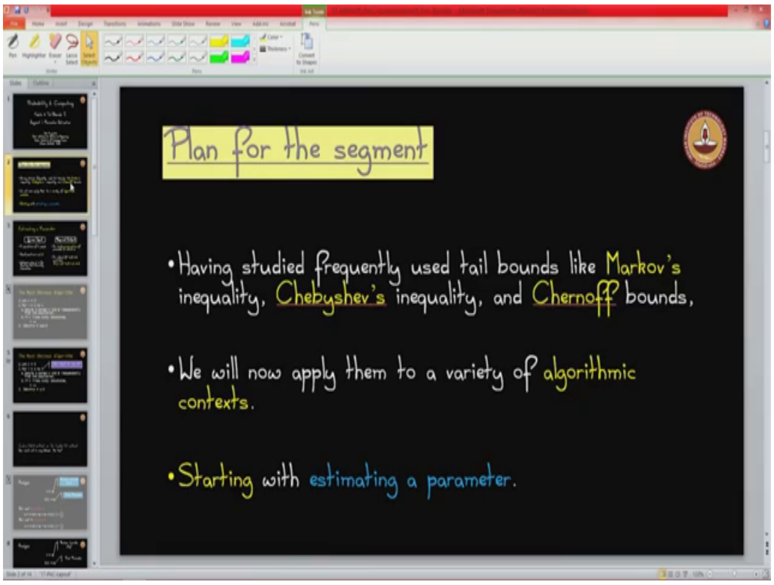
Probability and Computing

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

Module – 04
Tail Bounds I
Lecture - 20
Segment 1: Parameter Estimation

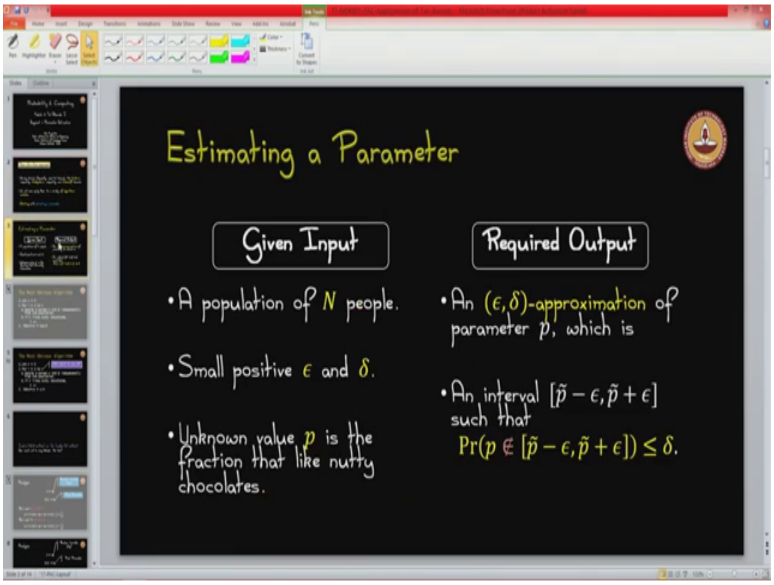
So, we are going to start a new module today, in this module assisted basically an extension of module 3 we are just going to look at some applications of a tail bounds; particularly from the from an algorithmic and computer science point of view, the first 1 is estimating a parameter. So, let me give us some.

(Refer Slide Time: 00:38)



Give ourselves some context, we have looked at all of these inequalities we want to look at the following problem.

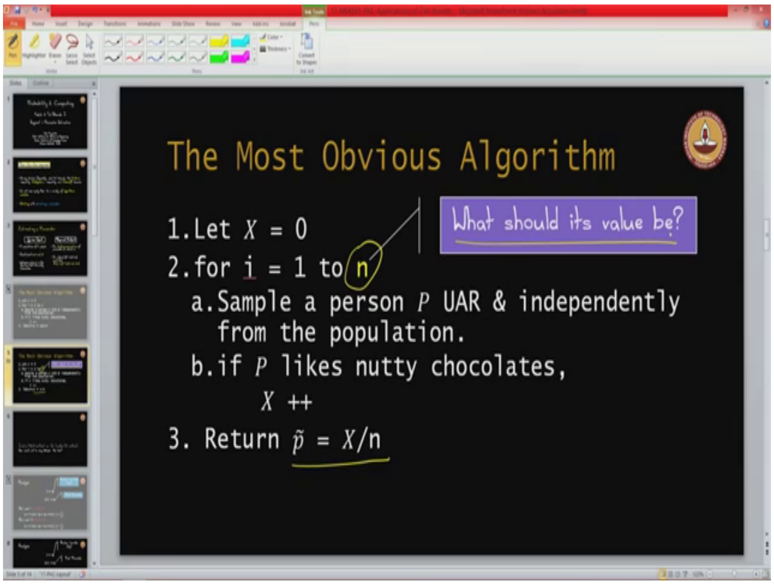
(Refer Slide Time: 00:51)



So, what is given you are given a population of N people. So, think of the country like India or something like that see and you are given 2 small constants epsilon and delta and you want to find the fraction of the people who like something say nutty chocolates ok. So, more practically be used to estimate for example, people who want to vote for or against a certain policy or something and what is this epsilon and delta they are specifying the accuracy with which you want to estimate this parameter p. So, in particularly you want an epsilon delta approximation of the parameter p ah.

What is an Epsilon delta approximation? It is basically an interval given by some p tilde minus epsilon to p tilde plus epsilon, so epsilon is the error term and what is the what role does delta play it 1 minus delta is often called the confidence interval your estimation; should be correct with probability at least 1 minus delta another way of stating that is that your estimation can be incorrect with probability at most delta ok. That is the way I have stated it in this particular case ah.

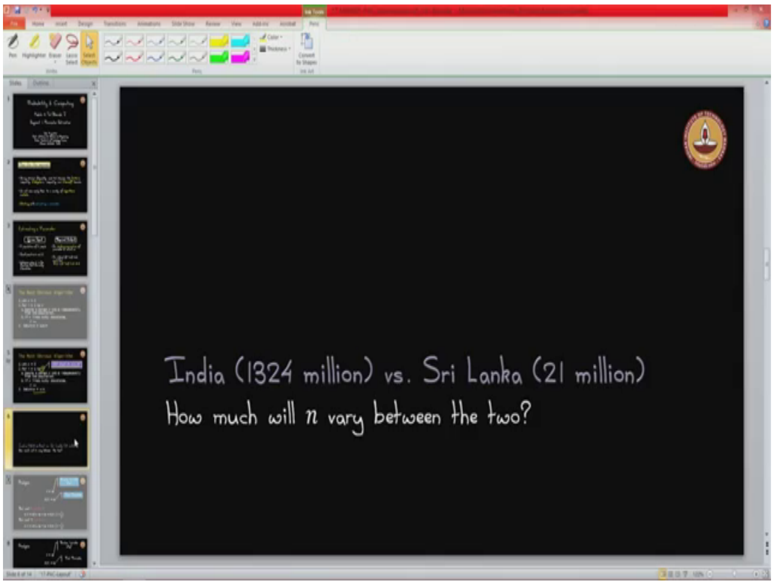
(Refer Slide Time: 02:16)



So, everybody clear about this problem statement because, very simple problem you just need to estimate the fraction of the people who like something and the nice thing is the algorithm we are going to consider for this case is extremely simple, we are just going to run a for loop N times and each time we are going to sample someone uniformly at random and independently and ask them do you like nutty chocolate or not; if they like nutty chocolates we are going to increment the counter otherwise we do not.

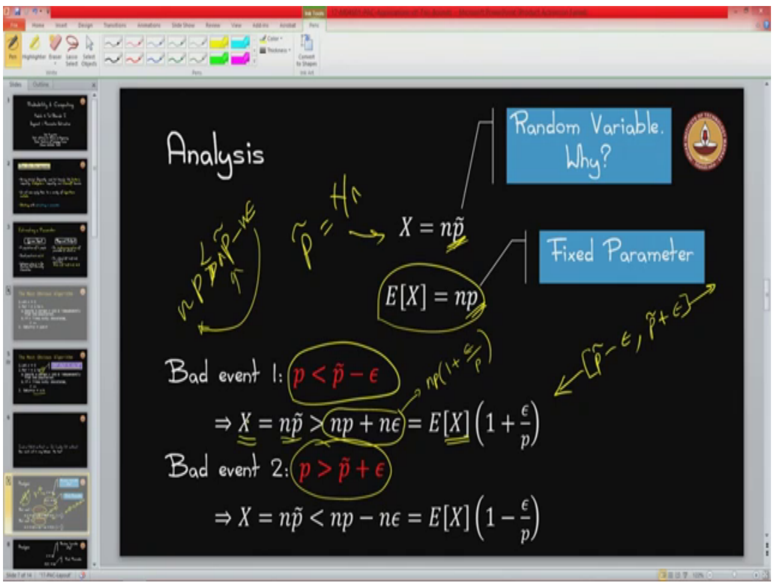
So, then what we do we basically estimate p tilde, we estimate p and denote that estimate by p tilde as just simply X over N, after we have iterated through some small N number of samples. So, the algorithm is very straightforward the only question that we need to worry about is what should the value of N be because, obviously if you just sample 1 person you are not going to get an accurate estimation and if you are going to sample all the upper case N number of people, that is not very useful either that is too difficult. So, the population of India is what some thousand 1.6 billion, no 1.3 some billion I think ok.

(Refer Slide Time: 03:44)



So, the question is we have let us let us play with our intuition a little bit, let us consider 2 countries India and Srilanka. I mean we have been having cricket matches recently and we have been doing pretty well. So, feels good we have 1.324 billion people and Srilanka has 21 million people, how does this N change between the 2 countries that will be an interesting question to ask right. So, out of curiosity how much what proportion how much more samples do you need for India, but a common intuition is that ok. So, country like India needs a lot more effort to get the estimates right, so let us let us see what happens.

(Refer Slide Time: 04:32)



So, here going back to the algorithm your X is the total number of people who said they like nutty chocolates right. So, X is a random variable here because that depends on p which is a random variable, why because you the population is fixed there are upper case n number of people, but what we do the randomness comes from the fact that we are sampling from this population and this p tilde represents the fraction of the people in the sample who liked nutty chocolates ok, so and X so basically we remember we calculated p tilde to be nothing but X over n and that we are just rewriting it this way here.

However, what is the expectation of X this is a well defined quantity at the even though we do not know p, p is a fixed parameter it is an exact fraction of the number of people in say India who like nutty chocolates. So, this p is a fixed quantity so the expectation of X is also a fixed quantity we do not know it, but it is fixed in this algorithm what are the ways in which things can go around again, once again we need to figure out what the bad events are and make sure that the bad events happen with low.

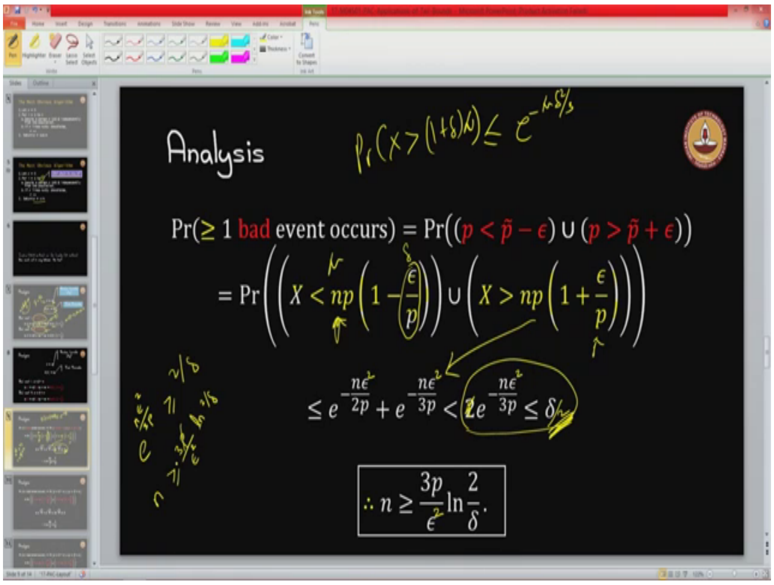
Probability 1 remembers our estimation is in the we are going to give the range p tilde minus epsilon 2 p tilde plus epsilon. So, anything below or above that range is bad and this is actual p value being less than the left end of the bound and this is the p value being to the right of the bound ok, these are the 2 ways in which things can go over ok. So, let us try to formally state these 2 bad events again if the pattern, I hope you are seeing is repeating itself here you are clearly specifying what the bad event and you are trying to capture that bad event in a way that can be fit into a known tail bound ah.

So, here what is how do we define this bad event well let us start with X ok, X we know is in p tilde that is that we know and another way of getting this now you just take this equation p greater than or inequality rather p minus epsilon and you multiply throughout by N and here you have then p tilde term. So, you isolate the np tilde term it is be less than here ok, you isolate then np tilde term and so on the right hand side you will get np and this term will go here plus N epsilon ok, so that is what you get over there alright.

So, now you can in this you can take out np, so you will get 1 plus epsilon by p, but np of course is nothing but E of X expectation of X times 1 plus epsilon by p ok. So, this ultimately what is your what is your bad event written in the form that can be captured by chernoff bounds X greater than expectation of X times 1 plus epsilon by p, remember that is the way in which you want chernoff bounds X greater than some 1 plus delta times mu.

So, you notice that we have gotten exactly the form that we, the same thing can be done with the other bad event as well ok; you will again get X to be less than some 1 minus delta times the expectation of X ok. All of this is basically taking the bad event and rephrasing it in a manner that it can where such that it can be plugged into the chernoff bounds ok. So, then we can do that ah.

(Refer Slide Time: 09:32)



So, our analysis basically is the following what is the probability that either 1 the bad events occur ok, there are 2 bad events that we listed it at least 1 of them should occur; what is the probability well that is equal to the union of the 2 probabilities there are 2 probabilities are shown in red over here and in the previous slide we worked out a formal way in which we can express those bad events in a manner a mean able to chernoff bounds.

So, that is what is written over here these 2 things and now this is the union of 2 bad events and if you re call the union bound if you have the union of 2 bad events, that is at most the sum of the individual probabilities of those bad events and so now you can simply apply a chernoff bounds in the first 1 it is.

Let us see here you have mu and this plays the role of delta just to be clear, this delta is different from the delta we are using in this, in this segment this is the delta coming from the way we stated chernoff bounds now. So, apologies for reusing the same delta, so then if you recall the chernoff bounds so what does it say probability X greater than 1 plus delta times mu is less than or equal to E to the minus mu delta square by 3, you recall this was 1 turn off bounds and that is that is what is showing up over here for instance. So, here it is 1 plus delta you have the 1 plus epsilon by p, so instead of it.

Student: (Refer Time: 11:40).

So, here Oh yeah trying to see why that and yeah you are right. So, let us make sure that we are not missing something, yeah I think there is some title here yeah. So, epsilon square by yeah that is right, so this epsilon square here also there will be an epsilon square the p term there will be a p square at the bottom, but there will be a p at the top as well, so one of the p is will get cancelled so as I miss the square over here.

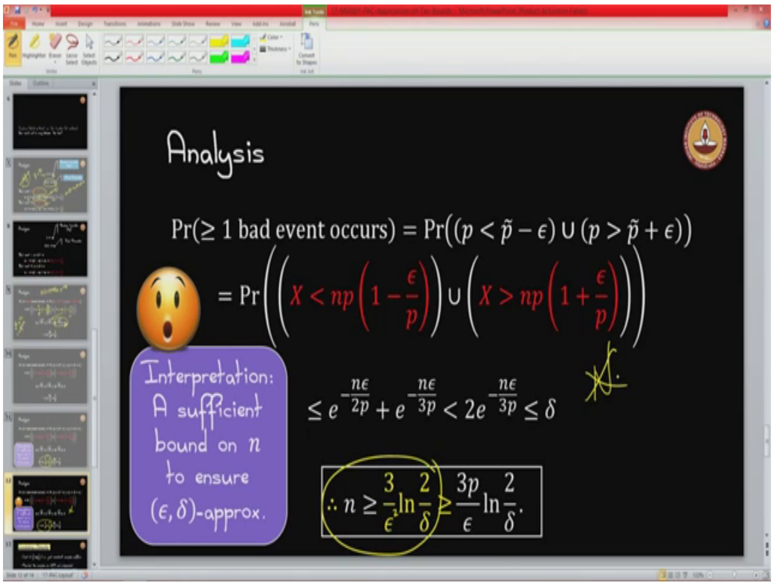
So, the square will continue to play role, so this is square over here ok. So, this is just I am approximating instead of dividing by 2 p in the exponent, I am just if you divide by 3 p you only get a larger bound. So, since we just want an upper bound I am just calling these 2 in individual terms as 2 times E to the minus N epsilon square by 3 p and recall that we want to make sure that this probability is at most delta remember this is one of the input parameters the probability that we will get into a bad event should be at most a delta ok, so that is where this delta shows up.

So, now, let us let us try to work with this inequality, so here if you take the log on both sides. So, first let us take the 2 to the other side it is delta by 2 and then let us also do 1 more thing to get rid of the negation let us make this let me write it over here. So, what can we do about this I can say E to the n epsilon squared by 3 p should be greater than or equal to 2 by delta ok. I am just taking the reciprocal on both sides, now I can take the log on ln on both sides rather. So, then what I will get is an I will get an epsilon square by 3 p, so what I am going to do is this I am going to take the epsilon square by 3 p to the other side. So, I am going to make that 3 p by epsilon squared and there will be ln over here ok.

This is what we get over here and that gives us how do we interpret this. What does this even mean? it just means that as long as we our N value is at least this much we are fine, but there is 1 pesky issue there is this p showing up, p is really what we want to estimate, but there is a p showing up there what do we do that yeah exactly. So, you just want an up a large enough value of n, so simply get rid of the p and you are still going to you your.

Rest your value of N is continue it is going to continue to be sufficiently large right ah.

(Refer Slide Time: 15:27)



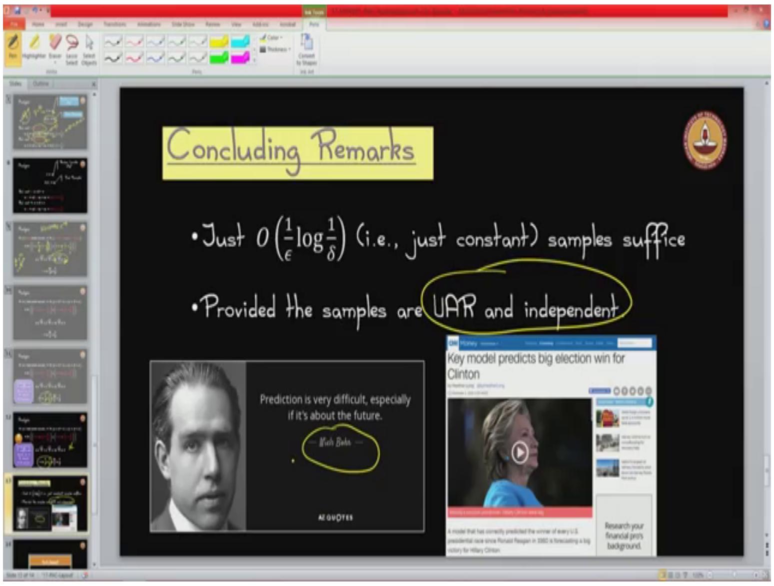
So, basically at the end of the day if you want to look at it what you have is you have gotten rid of the p here, what you have here is a sufficient bound on N to ensure that you are going to get an epsilon delta approximation this of course this type of. Which is which brings us to the issue that we talked about, if you look at this Bharath intuition was correct, what is the surprising fact over here.

There is no occurrence of capital N which means that whether you are you are working in Srilanka or in India, the capital N does not occur in this bound. So, it is only going to your estimation is only going to depend on epsilon and delta not on n. So, that is a fairly often a surprising fact because, you are where you know what you may fail to realize is that regardless of the size of the population your algorithm. Now focus is just based on just sampling and each time you are going to sample with certain probability p right sorry, you are going to get someone who likes nutty chocolates through some probability p; that does not change based on whether it is sri lanka or India that is the that is the intuition that is going over here.

Student: (Refer Time: 17:10).

That is correct, so you are absolutely right. In fact, that does bring me to the.

(Refer Slide Time: 17:16)

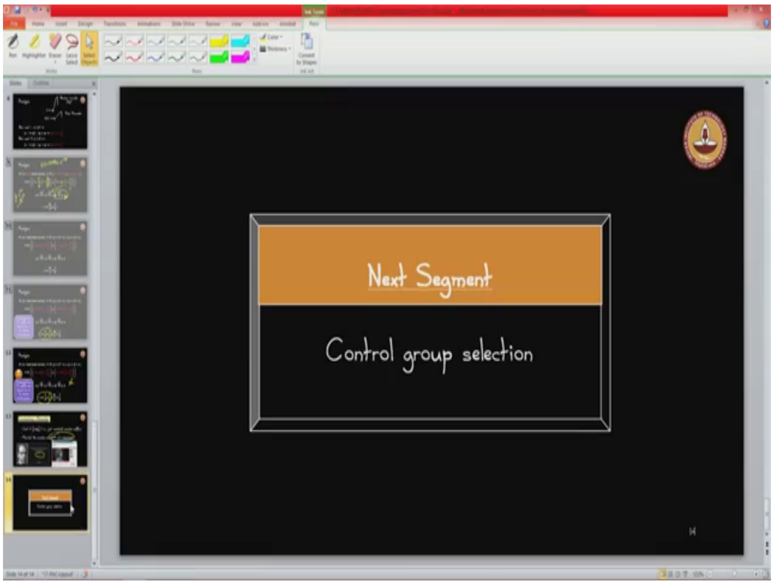


Conclusions like where I am you know wont emphasize that it really heavily depends on the fact that we are doing uniformly at random and independent samples and as you can tell from recent past, predictions go wrong all the time right and we can also take a leaf from Niels Bohrs book and say look prediction is very difficult especially if it is about the future.

Student: (Refer Time: 17:50).

So yeah sometimes you know when your memory as you start getting older, your memory becomes bad and even the prediction about the past is coming difficult [laughter]. So, with that we conclude the this segment.

(Refer Slide Time: 18:09)



So, next again we are going to see 1 more simple application of chernoffs bound and we conclude with that.

Thank you.