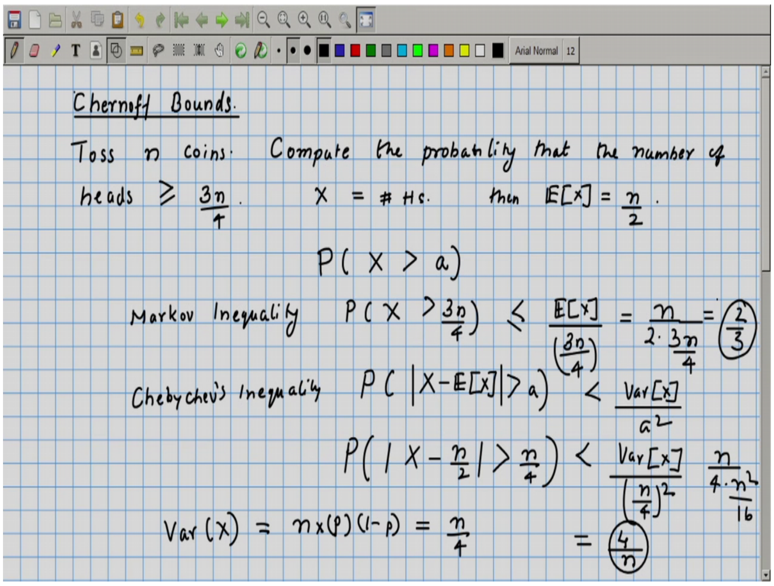
**Randomized Algorithms**
**Prof. Benny George Kenkireth**
**Department of Computer Science & Engineering**
**Indian Institute of Technology, Guwahati**

**Lecture – 10**
**Chernoff Bound**

(Refer Slide Time: 00:31)



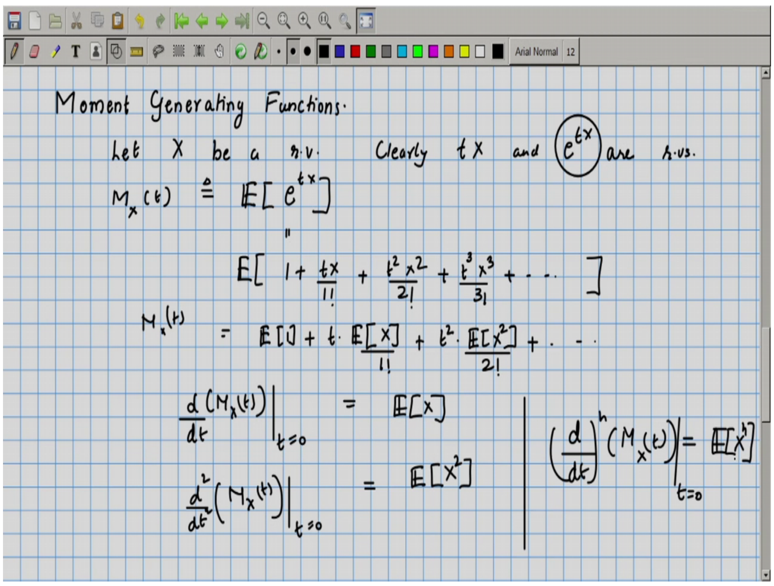

So, in today's lecture we will learn about what is known as Chernoff Bounds ok. So, the question that we will address in this lecture is the following: let us say we toss 1 lakh coins and we will get some number of heads and some number of tails. What is the probability that the number of heads is greater than 75000. So, we look at the following problem toss n coins, compute the probability that the number of heads is greater than say 3 n by 4. Clearly we expect; so, if we denote by X the number of heads then expectation of x is equal to n by 2.

So, now here in this case, the random variable takes a value much larger than its expectation ok. What is the probability of such a thing happening. So, this is a question of the following form; compute the probability that X is greater than a. We already have some tools to address these kind of questions, the first one the Markov inequality. If we use Markov inequality probability that X is greater than 3 n by 4 will be less than or equal to expectation of x divided by 3 n by 4. So, this is equal to n by 2 into 3 n by 4. So, that is equal to 2 by 3 ok. So, we can just conclude that the probability that the number of heads is significantly greater than n is going to be less than 2 by 3.

If you use Chebychev's inequality' so, this is the probability that X minus expectation of x greater than a this is going to be less than variance of x divided by a square. So, here expectation of x is n by 2. So, we have probability that X minus n by 2 to be greater than n by 4 ok. Whenever n is greater than 3 n by 4 X minus n by 2 is greater than the absolute value of that is going to be greater than n by 4. This happens with probability less than variance of x divided by n by 4 the whole square ok. Now variance of this particular random variable is going to be n times p into 1 minus p where, p is the bias of the coin. So, that is going to be equal to n by 4. Therefore, here we will get this as n by 4 into n square by 16.

So, this probability is going to be 4 by n much better than 2 by 3 ok. So, if you toss say 1 million coins the probability that the number of heads is greater than 0.75 million probability that the number of heads is greater than 0.75 million that is going to be something like 4 divided by 4 million ok. So, 1 in a million chance whereas earlier bound would have told us 2 by 3. So, this is a Chebychev's inequality is already a much better bound Markov inequality. We will try to get a significantly better bound in both of these bounds and that is what we will look at in this class.

(Refer Slide Time: 06:19)

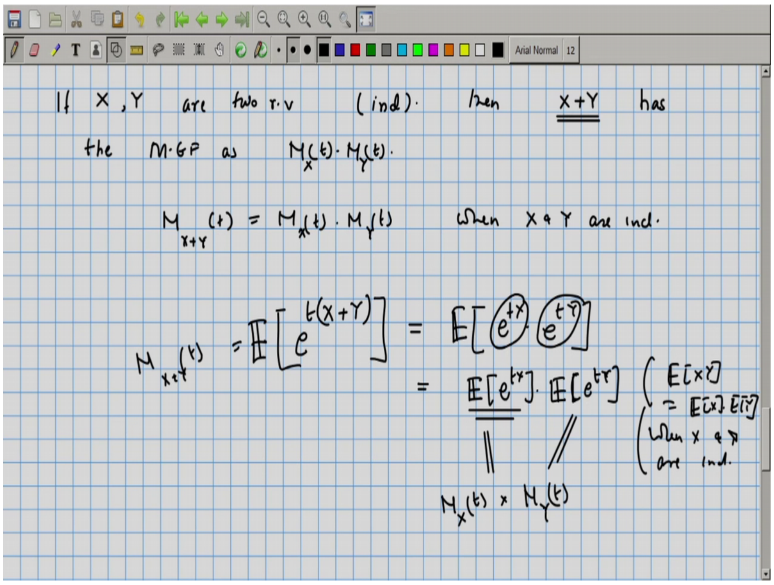


So, first in order to understand what is Chebychev, Chernoff bounds we will look at moment generating functions ok. So, let X be a random variable; so, clearly tx is also a random variable for any value of t and e raise to tx are so, clearly tx and e raise to tx are random variables; now we can compute the expectation of e raised to tx. So, moment generating function of this variable x is defined as the expectation of e raised to tx. So, look at this particular random variable; its expectation is going to be called the moment generating function. Note that the moment generating function is a function of a variable t; for each value of t we will get a different random variable e raised to tx and those random variables can have their own expectations that expectation is called the so, that expectation is the value of M x t.

So, this expression expectation of e raised to tx that is going to be dependent on t that function is called as a moment generating function. So, e raised to tx we can alternatively alternately write as expectation of 1 plus tx by 1 factorial plus t square x square by 2 factorial plus t cube x cube by 3 factorial and so on. And, by linearity of expectation, we can write this as expectation of 1 plus t times; t is a constant expectation of x by 1 factorial plus t square into expectation of x square by 2 factorial plus so on ok; so, this is your M x t. So, you can verify that if you differentiate M x t with respect to t and evaluate this function at t equals 0 you will get expectation of x.

So the moment generating function has in itself the value of the expectation of the random variable embedded. Similarly, if you look at the second derivative of the function M x t and evaluated at t equals 0 you will get this to be equal to expectation ofx square. In general, if you differentiate it n times then you will and evaluate this at t equals 0 what you will get is the expectation of x raised to n. So, these expectations are what are called as the moments and that is the reason why M x t is called as a moment generating function.

(Refer Slide Time: 10:11)

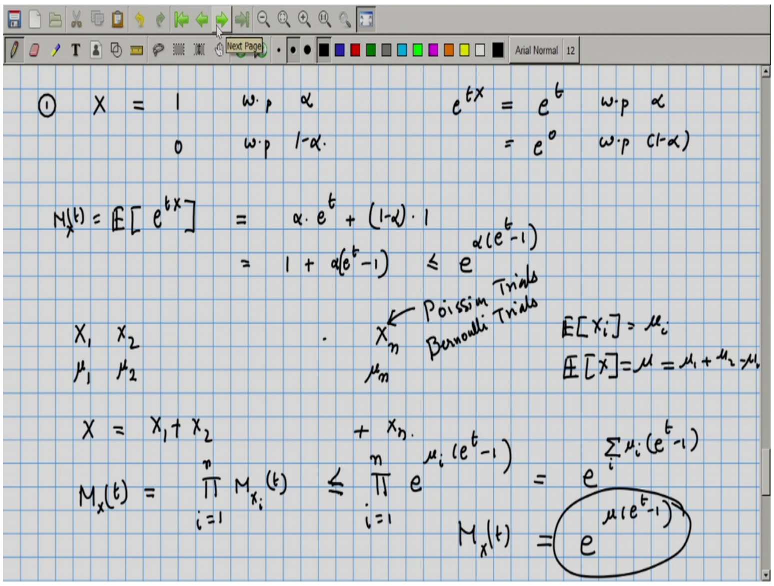



So, now let us just look at the; let us look at this moment generating functions more closely. If you have two random variables X and Y such that they are independent then, when you look at the random variable X plus Y so, this will have moment generating; if whenever it has a moment generating function ok. So, what will be the moment generating function of X plus Y ok; so, this will be the moment generating function of x times the moment generating function of y ok. In other words the moment generating function of x plus y is equal to M x t times M y t when X and Y are independent not too difficult to see. So, if you look at the random variable X plus Y its moment generating function is nothing, but e raised to t times x plus y and the expectation of this.

So, this is the moment generating function of x plus y. This is going to be equal to expectation of e raised to tx times e raised to ty, e raised to tx and e raise to t y are random variables in their own right. And, those random variables are going to be independent random variables because x and y are independent. When you look at independent random variables and you want to compute the product of these random variables those I mean the expectation of the product that will be just the product of the expectation. So, this will be equal to expectation of e raised to tx times expectation of e raised to ty. So, this is because expectation of x y is equal to expectation of x times expectation of y when x and y are independent ok.

So, this quantity is nothing, but moment generating function of x and this quantity is nothing, but moment generating function of y. So, the moment generating function of x plus y is just the product the individual moment generating functions. Now let us look at the moment generating functions of indicator random variables some particular examples.

(Refer Slide Time: 13:01)




So, let us say X takes X is a random variable which takes value 1 with probability say alpha and 0 with probability 1 minus alpha. What will be the expectation of x sorry, what will be the moment generating function of; so, we want to calculate expectation of t e raised to tx this is the moment generating function of x ok. So, this is equal to e raised to tx can take 2 values e raised to tx is equal to either e raise to t, this happens with probability alpha and this is when x is equal to 1 the value of e raised to tx is equal to e raised to t and this is equal to e raised to 0 this with probability 1 minus alpha and that happens when x is equal to 0.

So, the 2 values that it can take are e raised to t and 1. So, the expectation of e raised to tx is alpha into e raised to t plus 1 minus alpha into 1. So, this can be written as 1 plus alpha e raised to t minus 1 ok. So, this is clearly less than e raised to alpha t minus 1 ok. Now, let us look more closely at the problem that we were interested in; you toss a coin many times what is the probability the number of heads is much larger than the expected value ok. So, in that the individual toss we will set up as indicator random variables whether a head has appeared or not and we will look at this sum of these ok.

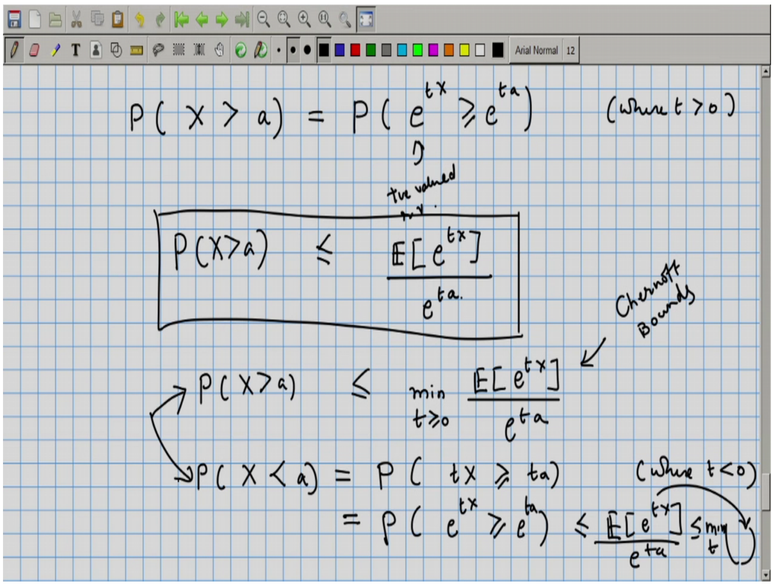
So, that is and we will look at the moment generating function of these independent random variables. So, let us say that we have all these random variables X 1, X 2 up to X n they are all random variables which takes 0 or 1 value ok. So, this takes value let us say X 1 takes the value 1 with probability mu 1, X 2 takes the value 1 with probability mu 2 and X n takes the value n with probability mu n. So, clearly expectation of X i is equal to mu i because this is 0 and random variable their expectation is same as the probability that they take the value 1 and if you look at this random variable X is equal to X 1 plus X 2 plus X n the moment generating function of X is going to be equal to these are independent random variables. So, your moment generating function is just going to be a product of the individual moment generating functions ok.

So, if M x i t is the individual random variable sorry if m if M if x i is the individual random variable M x i t will refer to the moment generating function of the i th random variable and this is equal to product i going from 1 to n this will be less than or equal to e raised to. So, here 1 plus alpha e raise to t minus 1 is less than e raised to alpha e raise to t minus 1 ok. So, here this product is going to be less than e raised to alpha instead of alpha you have mu i e raise to t minus 1 and this is nothing but e raised to the product will translate into summation in the exponent summation over i mu i e raised to t minus 1 and summation of the individual mu i s will be equal to the mean of the entire random variable.

So, if you denote expectation of X to be mu which will be equal to mu 1 plus mu 2 up to mu n then this will be equal to e raised to mu e raise to t minus 1 ok. So, we know that if we look at these random variables X 1 to X n which takes value 0 or 1 with probability mu i each when mu 1, mu 2, mu i then, the moment generating function of the sum of these random variables is going to be less than or equal to e raised to mu e raise to t minus 1. These random variables have a name they are called as Poisson Trails ok. When all these random variables are identically distributed this is also known as Bernoulli trials ok.

So, in our experiment what we have is, we have these coin tosses or them having the same bias. So, we have identical coins being tossed. So, we have a sum of Bernoulli trials. So, our random variable is a sum of Bernoulli trials and we know that in case of these kind of trials the moment generating function has to be less than the quantity that we have derived here ok. So, we will use this fact later on, but right now let us just try and see how we to put these things together and obtain what is known as the Chernoff bound ok.

(Refer Slide Time: 19:46)



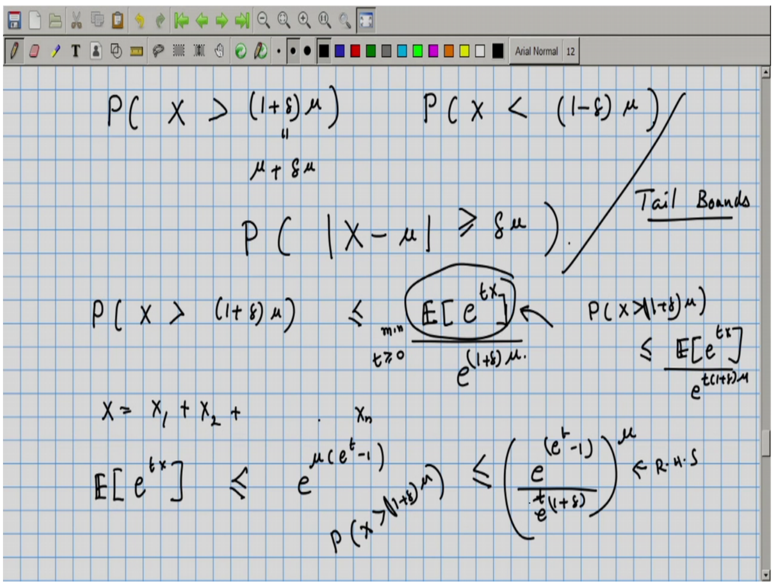

So, we were interested in questions of this kind probability that X is greater than a ok. Now this is same as this probability is same as probability that e raise to tx is greater than e raise to t a where t is greater than 0 ok. So, we had an arbitrary random variable x, we have converted that into a positive valued random variable this applies only when t is greater than 0. And therefore, once we have converted this into a positive valued random variable we can apply Markov inequality. And therefore, we will get this to be less than or equal to expectation of the random variable e raise to tx divided by a, a here being e raised to t a. Now this statement is true for any value of t in particular we can choose that t which minimizes this expression.

So, probability that X is greater than a is going to be less than minimum chosen over all t greater than 0 expectation of e raised to tx divided by e raised to ta. Note that the numerator and denominator are both functions of the variable t ok. So, you can try and optimize depending upon the random variable X and how its distribution is e raised to tx its expectation is going to have some functional form dependent on t and so, numerator and denominator both has t in them you can try and figure out what is the best value of t, but is the smallest value that you can choose, so as to minimize this expression and even for that value this probability will be less than that ok. So, these kind of bounds are in general known as Chernoff bounds.

Now, this is not a very useful form of Chernoff bound, but we will convert this into a nice form when X is when we know something more about the distribution of x. We could also look at inequalities of the form I mean we want to compute the probabilities that X is less than a particular a now this is going to be equal to the probability that minus let us say we choose a negative number t. So, then this probability is going to be equal to the probability that tx is greater than ta here we will choose where t is less than 0 and this probability is going to be equal to the probability that e raise to tx is greater than or equal to t a. Now e raised to tx. So, this is e raise to ta.

Now e raised to tx is going to be always a positive valued random variable and therefore, we can say that this probability is less than or equal to expectation of e raised to tx by e raised to ta. Again we can minimize. So, this is less than minimum over all values of t of this expression ok. So, these bounds are what is called as Chernoff bounds, it can apply both in the case where X is greater than a and X is less than a and for any random variable X ok. So, now, what we will do is we will convert Chernoff bound into a more useful form ok.

(Refer Slide Time: 23:49)

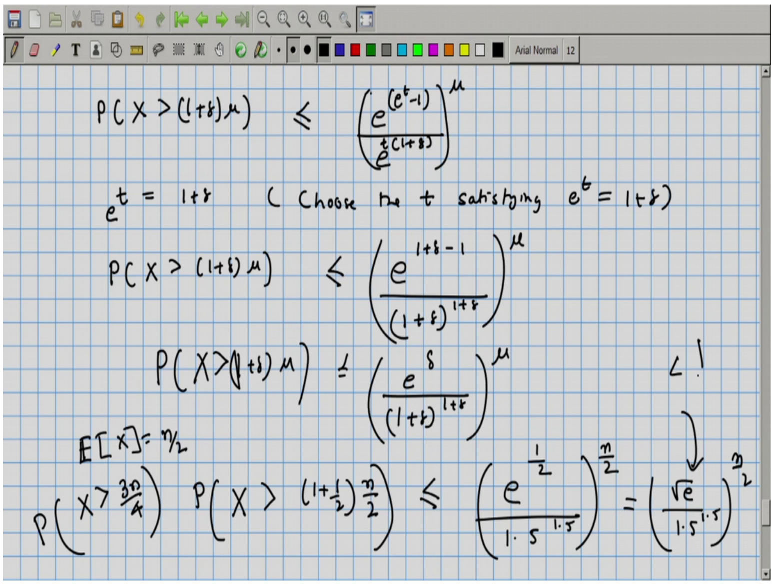


So, we will try and obtain the Chernoff bound for these kind of events X greater than say 1 plus delta times mu. So, delta is a parameter that is given to us and mu let us imagine that it is the expectation of the random variable we want to compute the probability that X is greater than 1 plus. So, you expect it to be around mu but, what is the probability that it is greater than mu by an amount delta mu ok. So, and we will also look at probability that X is less than 1 minus delta mu and together we will also look at we will combine these two things together and compute the probability that X minus mu is greater than delta times mu ok.

So, these will be special cases of what known as tail bounds ok. We will do this for the case where, X is a sum of Bernoulli trials or even Poisson trials. So, we know that probability that X is greater than 1 plus delta mu is going to be less than expectation of e raised to tx e raise to tx divided by e raise to 1 plus delta times mu t minimized over all values greater than or equal to 0. Now if X is the sum of Poisson trials of n Poisson trails then this quantity on the numerator expectation of e raised to tx we know is surely less than some quantity we had derived this. So, this is the moment generating this quantity here is the moment generating function of X we had shown that that is going to be less than e raised to mu e raise to t minus 1.

So, we will use this result. So, we have this to be less than e raised to mu e raised to t minus 1 by e raise to 1 plus delta ok. So, this is going to be this is equal to e raised to e raised to t minus 1 divided by e raised to 1 plus delta the whole raised to mu ok. So, what is the value of t that we will choose in order to make this quantity on the LHS sorry the RHS we need to make as small as possible what is the value of t that we can choose.

(Refer Slide Time: 27:05)




So, probability that X is greater than 1 plus delta times mu shown this is equal this is less than or equal to e raised to e raised to t minus 1 divided by e raised to 1 plus delta the whole raised to sorry there is a t here e raised to the expectation is this quantity and this probability that the probability that X is greater than 1 plus delta times mu is going to be less than this quantity ok

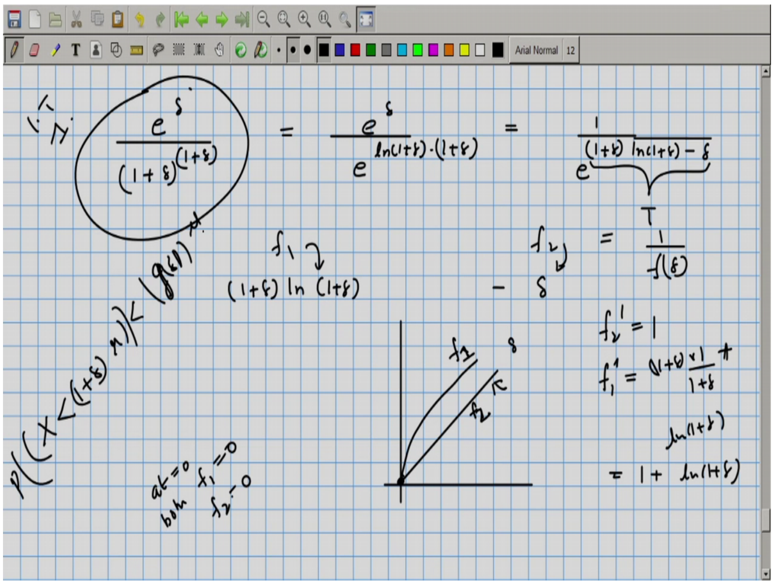
This is because this probability that X is greater than 1 plus delta times mu is going to be less than expectation of e raised to tx divided by e raised to t times 1 plus delta times mu. So, that is how the t comes in the denominator ok. So, this is our expression we need to choose a value of t ok. So, any positive value of t would be fine. So, 1 t which can simplify things is choose t equals. So, let us say we choose e raised to t to be equal to 1 plus delta ok. So, this quantity e raised to t we will choose it to be 1 plus delta clearly. So, choose the t satisfying this expression and clearly that t is going to be positive t is 0 you will get 1. So, little greater than 0 you will get 1 plus delta as delta is delta some positive number ok.

So, now this so, if we had chosen this as our value of t we know that probability that X is greater than or greater than 1 plus delta times mu is going to be less than or equal to e raised to 1 plus delta minus 1 divided by 1 plus delta the whole raised to 1 plus delta the whole raised to mu and this is equal to e raised to delta by 1 plus delta the whole raised to 1 plus delta the whole raised to the power mu. So, this is the probability that X is greater than 1 plus delta times mu.

Now, let us look at this expression closely in our scenario, we had X to be the sum of n i i d random variables and the expectation of X was n by 2. We wanted X to be greater than 3 n by 4 and we wanted to determine the probability of this probability that X is greater than 3 n by 4. Now in this case, delta would be 1 half. So, probability so, we were interested in probability that X is greater than 1 plus 1 half times the expectation n by 2 and this by Chernoff bound is going to be less than or equal to e raise to delta here is half divided by 1.5 the whole raised to 1.5 the whole raised to mu and mu is going to be n by 2 ok. So, this is equal to so, so, this is root e divided by 1 point five whole square sorry 1.5 raised to 1.5 you can verify that this quantity is less than 1.

Computation will tell you that this quantity is less than 1 and therefore, the probability of this happening is some quantity less than 1 the whole raise to n by 2. So, when n was something like 4 million this is some quantity less than 1 the whole to the power 2 million. So, it is a much smaller number than whatever we had computed before 1 by 1 million ok. Chebychev bound gave us the probability to be less than 1 in 1 million, but here were getting a probability some number which is less than 1 the whole raised to 2 million ok. Now this number is it really a small number is it very close to 1 ok. So, in some sense this formula is not really easy to work with does not really tell us how far is this away from 1, if this was let us say half and we readily know if this is less than half we readily know that this is a this gives a very good bound ok. So, let us try and convert this into a more workable form ok.

(Refer Slide Time: 32:29)

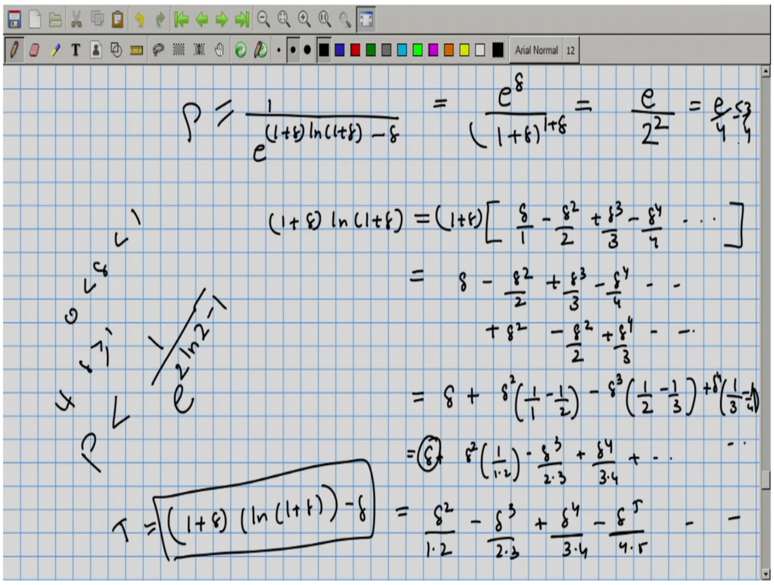


So, we have this probability to be e raised to delta divided by 1 plus delta the whole raise to 1 plus delta this is same as e raised to delta divided by e raised to log 1 plus delta times 1 plus delta and that we can write it as 1 by e raised to 1 plus delta log 1 plus delta minus delta ok. So, if we can show that the exponent here let us call that as T. If T is greater than 0 for any value of delta then, we have some number which is strictly less than 1. So, this quantity this entire quantity will be less than 1 ok. So, this will be less than 1 ok; so, will try to show that. So, if you look at this expression 1 plus delta 2 log 1 plus delta and minus delta you can plot the curves of these.

So, this is a curve delta and 1 plus delta times log delta if that curve lies to strictly above this curve then, we know that 1 plus delta log 1 plus delta minus delta is going to be always positive ok. How do we show that? At 0 the function f 1 and the function f 2 are both 0 because this is 1 plus 0 log 1 and this is 0. So, log 1 is 0. So, the both of them are 0. So, f 1 and f 2 agree at 0 and if you look at the derivative of f 2 that is so f 2 prime is 1 f 1 prime is going to be 1 plus delta into 1 by 1 plus delta plus log 1 plus delta that is going to be 1 plus log 1 plus delta. So, f 1 as a function which grows faster than f 2 and they agree at 1. So, clearly this is going to remove this is going to lie above the so this is a curve of f 1, f 2 and this is the curve of f 1 ok.

So, this difference is always going to be positive ok. Since it is always going to be a positive a positive number we can write this is 1 by f delta where f delta is going to be always greater than 1 ok. So, this entire this tells us that this is going to be some f let us say. So, the probability of success or the probability that X is less than 1 plus delta times mu is going to be always less than some let us say g delta raised to mu and g delta is a number less than 1. Now we will determine the form of g delta.

(Refer Slide Time: 36:37)

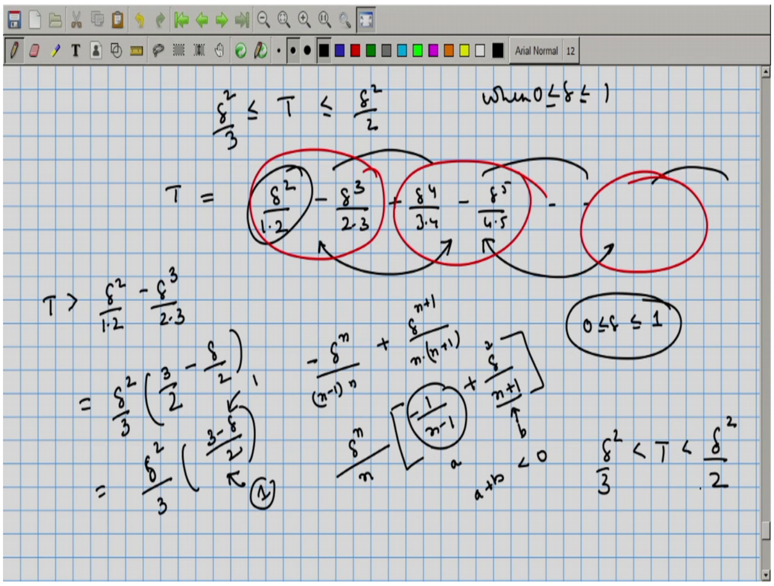


So, we have this expression 1 by e raised to 1 plus delta log 1 plus delta minus delta. So, this we can write it is. So, if you look at 1 plus delta into log 1 plus delta if you just expand this is going to be 1 plus delta into log 1 plus 1 plus delta we can write it as delta by 1 minus delta square by 2 plus delta cube by 3 minus delta raised to 4 by 4 and so on ok. So, this if we just multiply the terms delta minus delta square by 2 plus delta cube by 3 minus delta raised to 4 by 4 and so on, when you multiply the delta you will get plus delta square minus delta cube by 2 plus delta raised to 4 by 3 and so on. So, this is going to be equal to delta plus this is going to be delta square into 1 by 1 minus 1 by 2 minus delta cube into 1 by 2 minus 1 by 3 plus delta raised to 4 into 1 by 3 minus 1 by 4 and so on ok.

Clearly this is going to be equal to delta plus delta square into 1 by 1 into 2 minus delta cube by 2 into 3 plus delta raised to 4 4 and so on and what we were interested in this 1 plus delta into log 1 plus delta minus delta. So, this term goes away. So, this is going to be equal to delta square by 1 into 2 minus delta cube by 2 into 3 plus delta raised to 4 by

3 into 4 minus delta raised to 5 into 4 into 5 and so on ok. How large is this is the exponent of this term ok. So, if we call this as T how large is T ok.

(Refer Slide Time: 38:57)




We will show that T lies between delta square by 2 and delta square by 3 [FL]. So, we will show that T lies between delta square by 2 and delta cube by 3 ok. So, T if we know is equal to delta square by 2 by 1 into 2 minus delta cube by 2 into 3 plus delta raised to 4 by 3 into 4 minus delta raised to 5 by 4 into 5 and so on. Now, if you look at, if you combine any 2 of 2 terms if you look at delta r raised to n the first term is going to be negative term delta raised to n divided by n minus 1 into n, the next term is going to be a positive term delta raised to n plus 1 divided by n into n plus 1.

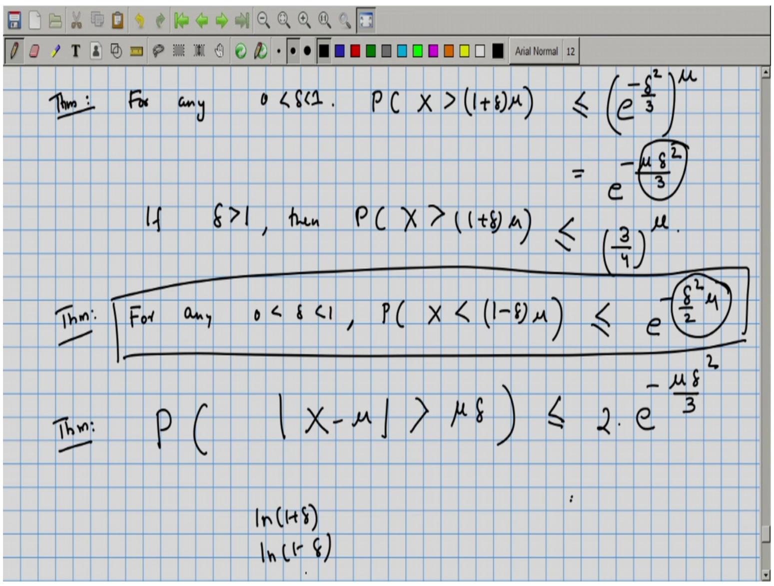
We can take what is common delta raised to n divided by n and then what is remaining is minus 1 by n minus 1 plus delta by n plus 1. The denominator is larger and the numerator is smaller because delta we may assume is between 0 and 1, if delta is between 0 and 1 when delta is between 0 and 1 these sum of terms taken 2 at a time this is going to be a negative number because, this is smaller than this quantity if you call it as b and this as a minus a plus b. So, this if you call this entire thing as a then a plus b is going to be less than 0 because 1 is greater than delta n minus 1 is smaller than n plus 1.

So, the sum of all these terms all these are all these terms are going to be negative terms and their sum is going to be negative. So, T is going to be at most delta square divided by 1 into 2. Now if you take the summation in a slightly different way that is if you take these two together and the next two together and the next two together you can see that each of these terms are going to be positive ok. So, T so, therefore, T is going to be greater than delta square by 1 into 2 minus delta cube by 2 into 3 and this is equal to delta square by 3 into 3 by 2 minus delta by 2. So, this is equal to delta square by delta square by 3 into 3 minus delta by 2 the maximum value that delta can take. So, this delta the maximum value that it can take is 1 and even in that case this expression is going to be 1. So, this is no less than 1. So, the entire value of T will lie between. So, that T should surely lie between delta square by 3 and delta raised to 2 by 2.

Now, let us look at this expression slightly more carefully. We did this analysis for delta lying between 0 and 1. But 1 plus delta log 1 1 plus delta minus delta this is of course, always positive. Since it is always positive, the value of this expression at 1 is going to be I mean whatever is the value of the expression at 1 that is going to be smaller than the values at all higher delta ok. So, if delta is greater than 1 greater than or equal to 1 then this expression p, p is going to be less than the value at 1 which is going to be 1 by e raised to 2 log 2 minus 1 ok.

We could have just simply written this as e raised to delta divided by 1 plus delta the whole raised to 1 plus delta and delta equals 1 this is e divided by 2 square. So, that is e by 4 it is less than 3 by 4 ok. So, when delta is greater than 1 we can readily substitute this by 3 by 4.

(Refer Slide Time: 44:18)



So, 3 by 4 raised to mu and when delta is less than 1 we have so, when delta is less than 1 we have the following theorem. So, for any delta between 0 and 1 probability that X is greater than 1 plus delta times mu is going to be less than e raise to minus delta square by 3 the whole raised to mu ok.

So, this is e raised to minus mu delta square by 3 ok. So, this is a form in which we are going to use Chernoff bound. When delta is greater than 1 of course, so, if delta is greater than 1 then probability that X is greater than 1 plus delta times mu is less than 3 by 4 raised to mu ok. We can also derive a similar bound for any delta between 0 and 1 probability that X is less than 1 minus delta times mu this we can show is less than or equal to e raised to minus delta square by 2 times mu. Now here, if you assume that X is a positive valued random variable delta cannot I mean there is no sense in choosing a delta less than I mean let us say greater than 1 if delta is greater than 1 then 1 minus delta is going to be let us say a negative number ok.

So, we can combine these two theorems and say probability that X minus mu is greater than mu times delta is going to be less than 2 times e raised to minus mu delta square by 3 ok, we take the smaller of these quantities or smaller or larger. So, this probability is going to be the sum of these two probabilities and mu delta square by 3 is going to be a smaller quantity than this. So, we can say that this probability is less than twice; this twice e raise to minus mu delta square by 3 ok. So, when we try to prove this statement, we just have to retrace the steps of the previous theorem. But, instead of log 1 plus delta has expansion we will have to work with log 1 minus deltas expansion and then everything else is the same and we will get this as the bound ok.

We will stop here for the time being. This concludes our discussion on Chernoff bounds.