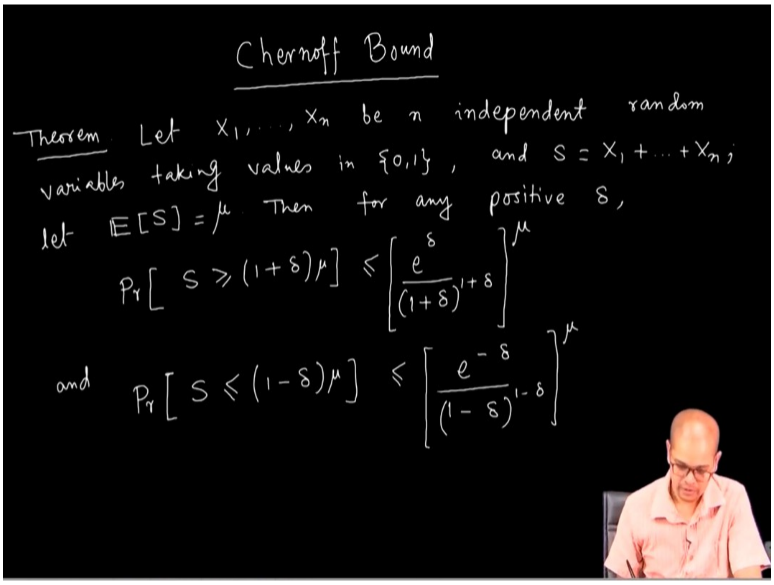
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module No # 03
Lecture No # 15
Proof of Chernoff Bound

Welcome in the last lecture we started looking into concentration inequalities and we have looked at Markov's inequality and Chebyshev's inequality and Chernoff bound. And we have also seen how using Markov's and Chebyshev we can give concentration bounds on the performance of randomized algorithm around its screen. So in today's class we will prove Chernoff bound in the last class we had only stated it. So today we will see the proof.

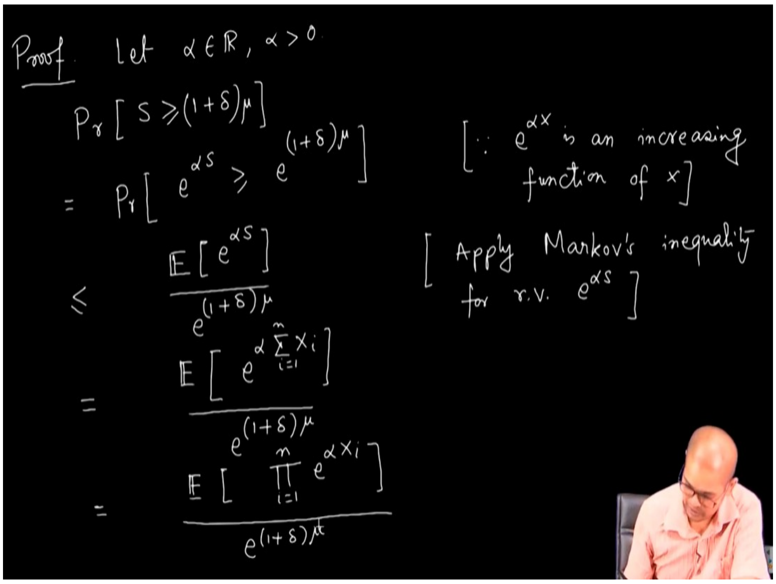
(Refer Slide Time: 01:04)



We will prove the most general version so let me write the theorem again let $X_1, \ldots, X_n$ be n independent random variables taking values in $\{0,1\}$ random variable. And $S = X_1 + \ldots + X_n$ and let expectation of S be μ then for any positive δ we have probability that S is greater than equal to $(1+\delta)\mu$ is at most $\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$.

And probability that S is less than equal to $(1-\delta)\mu$ is less than equal to $\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$.

(Refer Slide Time: 04:18)

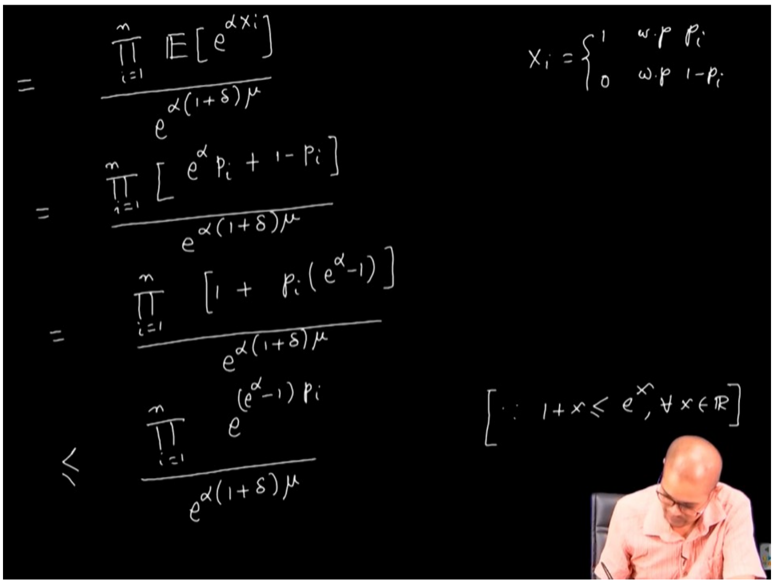


So proof so we take any positive real number alpha and this proof basically uses Markov's inequality this phenomenon. We have already also seen in the proof of Chebyshev's inequality so both Chebyshev's inequality and Chernoff bound in the sense uses Markov's inequality at its core. So let α beginning real number and α greater than 0. So probability that S is greater than equal to $(1+\delta)\mu$ this is probability that e^{α} is greater than equal to $e^{(1+\delta)\mu}$.

This hold because this is since $e^{\alpha x}$ is an increasing function on of x. Now $e^{\alpha S}$ is a positive random variable so now we apply Markov inequality and this is less than equal to $E\frac{[e^{\alpha x}]}{e^{(1+\delta)\mu}}$. This is applying Markov's inequality for random variable $e^{\alpha S}$. Now we write S as $\sum_{i=1}^{n} X_i$ by, $e^{(1+\delta)\mu}$.

Now this is $\frac{E\left[\prod_{i=1}^{n} e^{\alpha X_i}\right]}{e^{\alpha(1+\delta)\mu}}$. Now because X_i's are independent random variable $e^{\alpha X_i}$ are also independent.

(Refer Slide Time: 07:52)

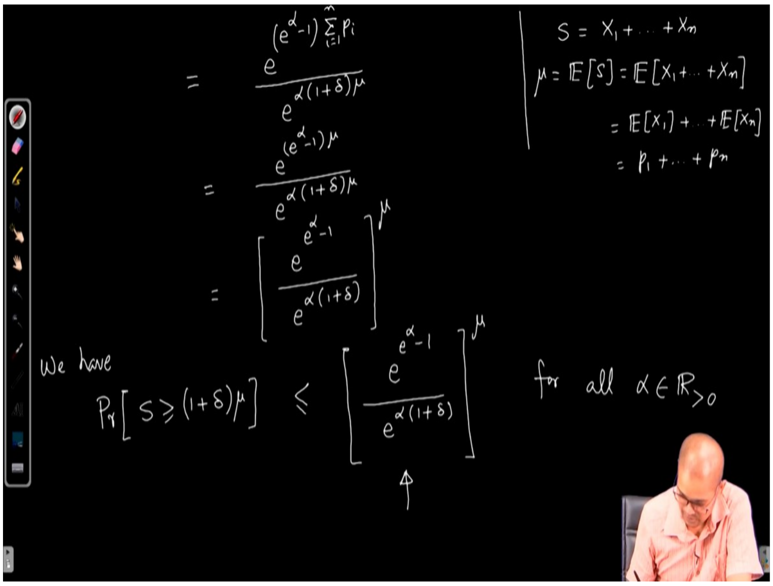


And hence the numerator is $E\left[\prod_{i=1}^{n} e^{\alpha X_i}\right]$ there should be alpha again here. And this α will be here and will be here $e^{\alpha(1+\delta)\mu}$ and now we compute what is the $E\left[e^{\alpha X_i}\right]$. This is product i = 1 to, n so suppose X_i takes value with probability p_i so we have assumed only independence but they need not be i.i.d they did not be identically distributed.

So suppose X_i take values 1 with probability p_i and 0 with probability $1-p_i$ so it is $e^{\alpha} p_i$ + this is the probability with which X_i takes value 1. And if X_i is 0 e need to the power alpha 0 is 1 and that happens with the probability $1-p_i$ by, $e^{\alpha(1+\delta)\mu}$. Now this is $\frac{\prod_{i=1}^{n} 1+p_i(e^{\alpha}-1)}{e^{\alpha(1+\delta)\mu}}$.

Now this can be upper bounded as $\frac{\prod_{i=1}^{n} e^{(e^{\alpha}-1)p_i}}{e^{\alpha(1+\delta)\mu}}$. This is because $1+x \leq e^{x}$ this holds for all real number $x \in \mathbb{R}$ and this particular inequality holds for all α.

(Refer Slide Time: 11:35)

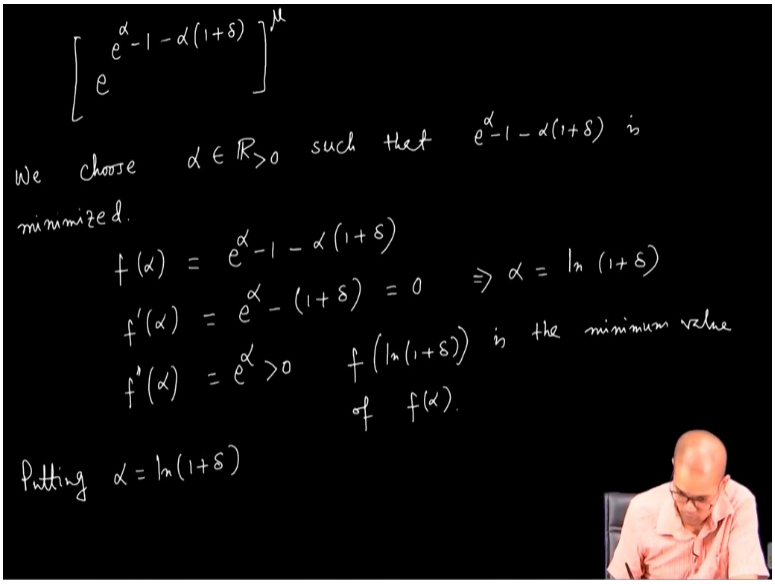


And next we simplify it further and we get this is $\frac{e^{(e^\alpha - 1)\sum_{i=1}^{n} p_i}}{e^{\alpha(1+\delta)\mu}}$. Now $\sum p_i = E[S]$. So let us recall S was $X_1 + \ldots + X_n$ and μ was expectation of S which is $E[X_1 + \ldots + X_n]$. Now we apply linearity of expectation is $E[X_1] + \ldots + E[X_n] = p_1 + .. + p_n = \mu$.

So we put it here $\frac{e^{(e^\alpha - 1)\mu}}{e^{\alpha(1+\delta)\mu}}$. So this can be written as $\left(\frac{e^{e^\alpha - 1}}{e^{\alpha(1+\delta)}}\right)^{\mu}$. Now these holes this inequality so what do we have in somebody we have that probability that S is greater than equal to $(1+\delta)\mu$ this is less than equal to $\left(\frac{e^{e^\alpha - 1}}{e^{\alpha(1+\delta)}}\right)^{\mu}$.

This holds so we have this for all α positive real number now we pick an α which minimizes the right hand side and that will give me the tightest bound. Now how to pick α which minimize this?

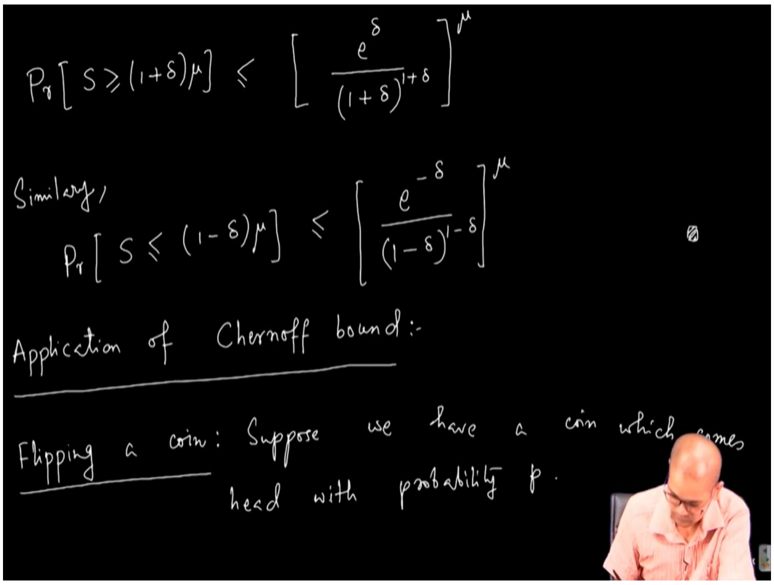
(Refer Slide Time: 14:44)



So this right hand side is nothing but $(e^{\alpha}-1-\alpha(1+\delta))^{\mu}$. So you can ignore μ and try to maximize $e^{\alpha}-1-\alpha(1+\delta)$. And because it is a when we want to minimize and again because e^{x} is increasing function of x. So we pick we choose $\alpha \geq 0$ such that $e^{\alpha}-1-\alpha(1+\delta)$ is minimized.

So let us call it $f(\alpha)=e^{\alpha}-1-\alpha(1+\delta)$ so $f'(\alpha)=e^{\alpha}-(1+\delta)$ set it to 0. So $\alpha=\ln(1+\delta)$ $f''(\alpha)=e^{\alpha}$ so that means that $f''(\ln(1+\delta))$ is the this is greater than 0. Since α is greater than 0 so this is the minimum value of $f(\alpha)$.

(Refer Slide Time: 17:31)



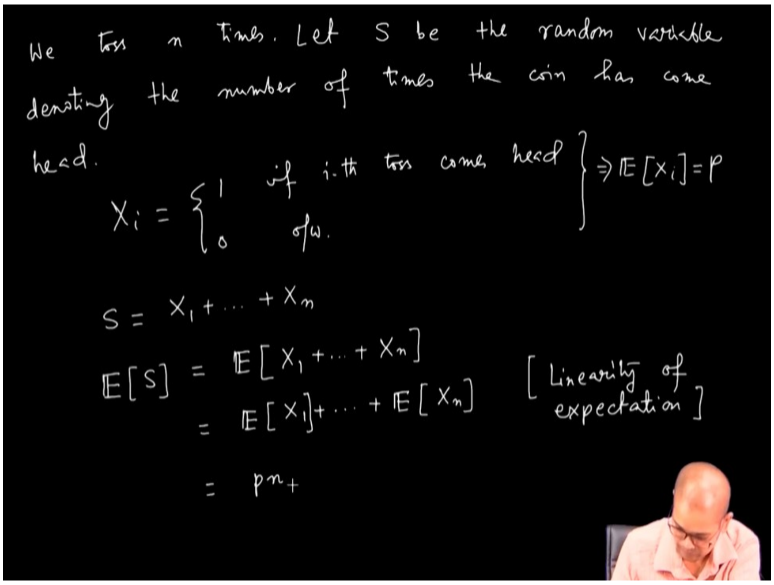
So we put alpha equal to so putting $\alpha=\ln(1+\delta)$ what we have is that probability that S is greater than equal to $(1+\delta)\mu$ is less than equal to. So in this expression here we put

$\alpha = \ln(1+\delta)$ and let us see what we get? So e^{α} for $\alpha = \ln(1+\delta)$ will be $1+\delta$ and that 1 get canceled. So it will be e^{δ} by and in the denominator we have $e^{\alpha(1+\delta)}$.

And if we put $\alpha = \ln(1+\delta)$ it is coming $(1+\delta)^{(1+\delta)\mu}$. So these proves the upper bound so similarly using the take technique we can show the probability that $S \leq (1-\delta)\mu$ is less than equal to $\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$ which concludes the proof of Chernoff bound.

So now let us see some application of Chernoff bound how we can use it to get tight concentration around mean. So the first so application or use of Chernoff bound the classic application is flipping a coin. So suppose we have a coin which comes head with probability p.

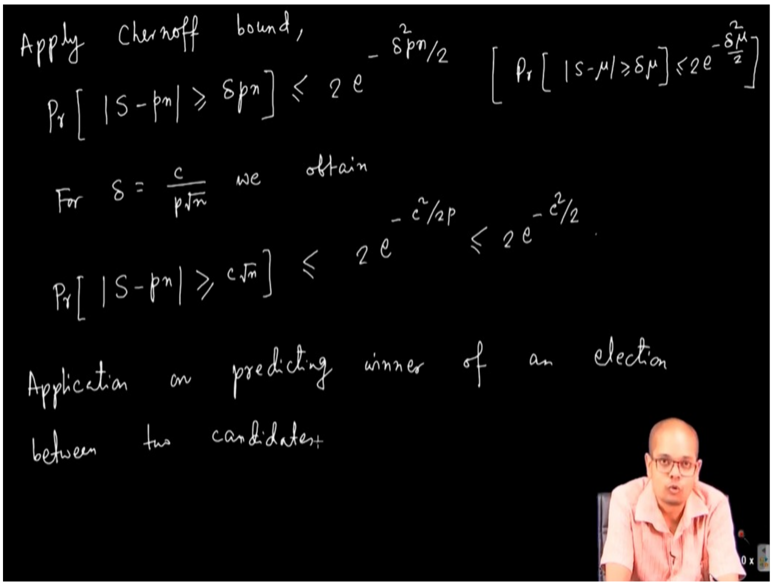
(Refer Slide Time: 21:09)



We toss n times let S be the random variable denoting the number of times the coin has come head. So how can I compute is? So towards that let us define X_i recall to 1 if i'th toss comes head and 0 otherwise. So we see that S is $X_1 + \ldots + X_n$ it is a sum of random variables where each random variable takes a value in between 0 and 1. And expectation of S is expectation of $X_1 + \ldots + X_n$ and now we apply linearity of expectation.

And now you see that each X_i so this is linearity of expectation and each expectation of X_i from here these are indicator random variable it takes value 1 if some particular event happens. And expectation of an indicated random variable is the probability of that event and this is the probability that head comes which is p. So this is p each expectation of X is p this is *pn*.

(Refer Slide Time: 24:35)



So now applying Chernoff bound what we get is? Probability that S deviated from its mean by more than δpn it is like assume that suppose we are using this experiment to toss the coin n times to have an estimate of p. What is the probability that this coin comes up head? And this is the sort of application of finding whether a coin is unbiased or not and what is the empirical probability?

Empirical probability is number of times it has come up head by n and in that we make an error of at most delta this is the probability of error is at least delta this probability is bounded this is less than equal to $2e^{-\delta^2 \mu/2}$. So this is one of the special forms of Chernoff bound so this is applying probability that $S-\mu$ is greater than equal to $\delta \mu$ this is less than equal to $2e^{-\delta^2 \mu/2}$.

So applying this form of Chernoff bound we get this and now the question is what is the delta or what is the error bound? So for $\delta = \frac{c}{p\sqrt{n}}$ we obtain probability that $|S-pn|$ is greater than equal to $c\sqrt{n}$ is less than equal to $2e^{-c^2/2p}$. Now p is at most 1 and this is negative so this is less than equal to $2e^{-c^2/2}$.

So these are applicable in even in sort of voting application so suppose mu as application on predicting say winner of an election say between 2 candidates. Suppose one candidate has got p'th fraction of all votes and we pick a port uniformly at random and if that candidate that random vote is for candidate a then it is like head and if that vote is for other candidate. Then it is like tail and it is boils down to estimating what is the fraction of votes? That one candidate as got and from that information we can predict what will be the outcome of the election. So we will stop here today.