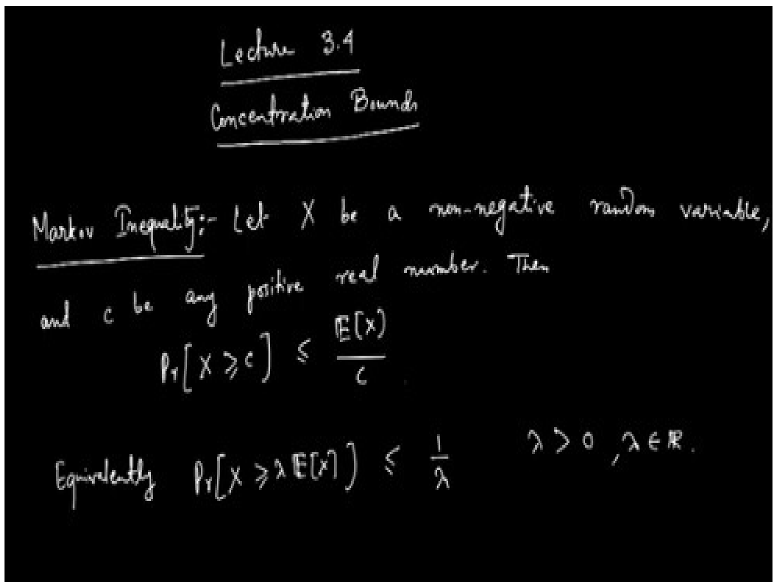
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module No # 03
Lecture No # 14
Concentration Inequalities: Markov, Chebyshev, Chernoff

Welcome so in this lecture we will study some important but fundamental concentration inequalities for probability distribution.

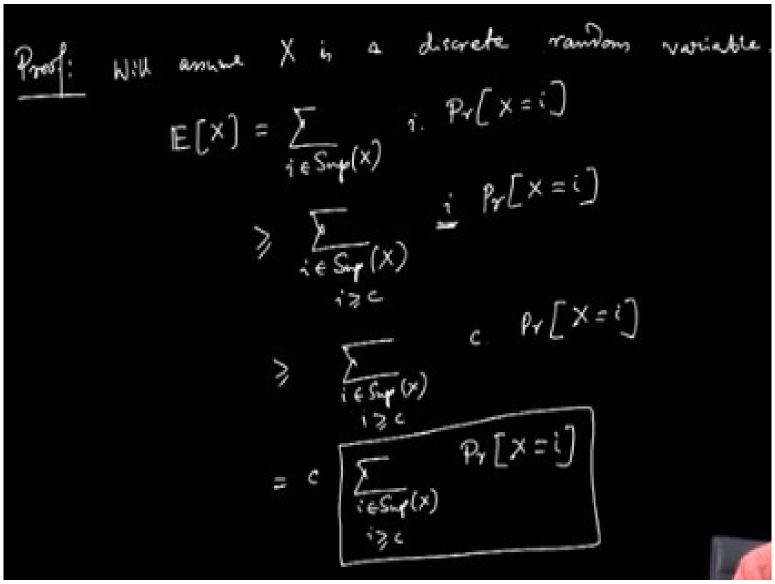
(Refer Slide Time: 00:39)



So this lecture 3.4 concentration bounds our first concentration bound concentration inequality is the Markov inequality. So let X be a non-negative random variable and c be any positive real number. Then probability that X takes value at least see is less than equal to c by expectation of X. Equivalently probability that X takes more than λ times expectation of X this is less than equal to this is expectation of X by c is less than equal to $\frac{1}{\lambda}$.

This holds for all λ greater than 0 and this is sort of the best bound possible if you only use the expectation of X to bound it is to get this concentration bound.

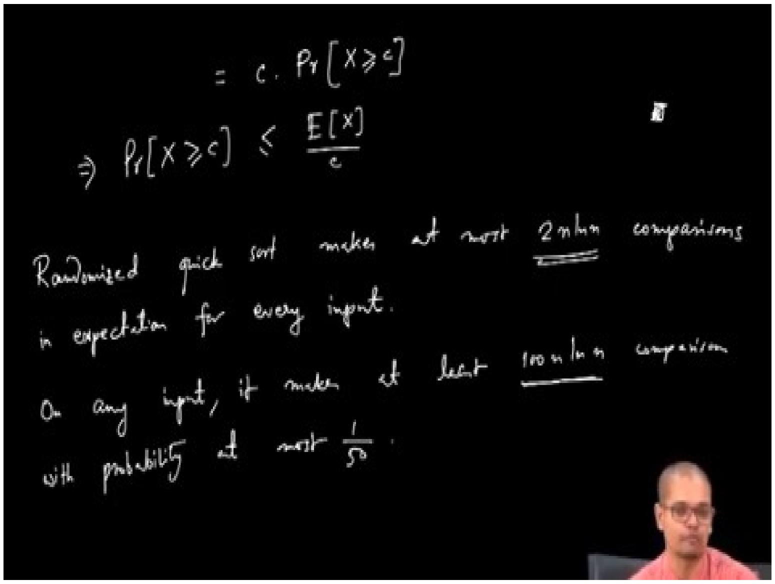
(Refer Slide Time: 03:37)



So proof so let will prove this result for discrete random variables only so because you know for randomized algorithms this is the case when you what we need mostly. So we will assume is not without loss of generality that X is discrete random variable. For discrete random variable expectation of X is $\sum i\, Pr[X=i]$ this is greater than equal to i in support of X and i greater than equal to c.

And this is greater than equal to you know in this region i is greater than equal to c. So this i we can lower bound it using c that is what we do sum i in support of X, i greater than equal to c; c times probability X = i it may. Let us bring c out i in support of X, i greater than equal to c probability X = i.

(Refer Slide Time: 06:16)

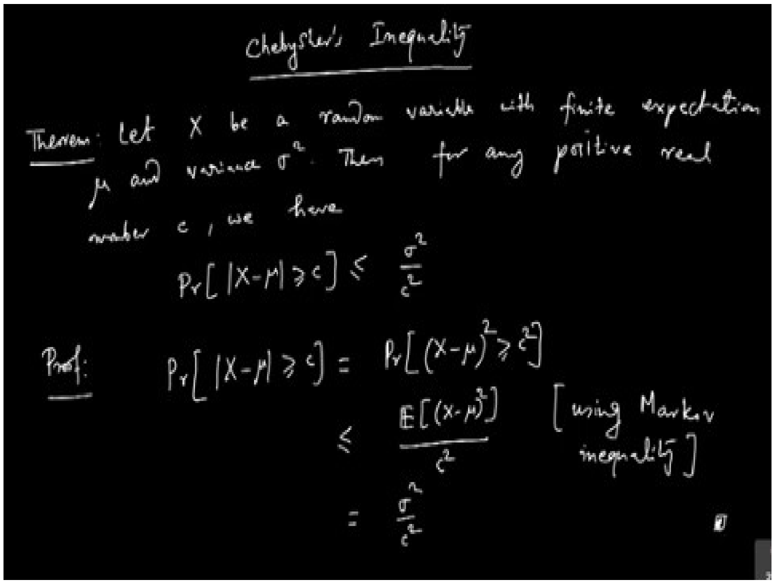


Now this is greater than equal to not greater than equal to the last one this also should be equal. Now this probability is the probability that X takes value at least c this is c times probability that X takes value greater than equal to c. So probability that X takes value greater than equal to c is less than equal to $\frac{E[X]}{c}$ which concludes the proof. Now this shows that you know how I can use Markov inequality?

We had observed that you know that the randomized quick set in the last lecture we have seen randomized quick sort makes at most twice in $\ln n$ comparison in expectation for every input. Hence on any input it makes say at least say $100 n \ln n$ comparisons with probability at least or at most using Markov. This probability is upper bound by at most expectation which is at most twice in $\ln n$ and X value should be greater than equal to $100 n \ln n$ so this at most $\frac{1}{50}$.

And so whenever we have an on the performance of in algorithm we have a whenever we have a bound on the expected cost we can use Markov to get this straightforward you know probabilities. So from expectations we can go to probabilities using Markov in this way direct application of Markov. Now we can improve Markov significantly when we have when we use the second moment that means variation and the inequality that you get is called Chebyshev's inequality.

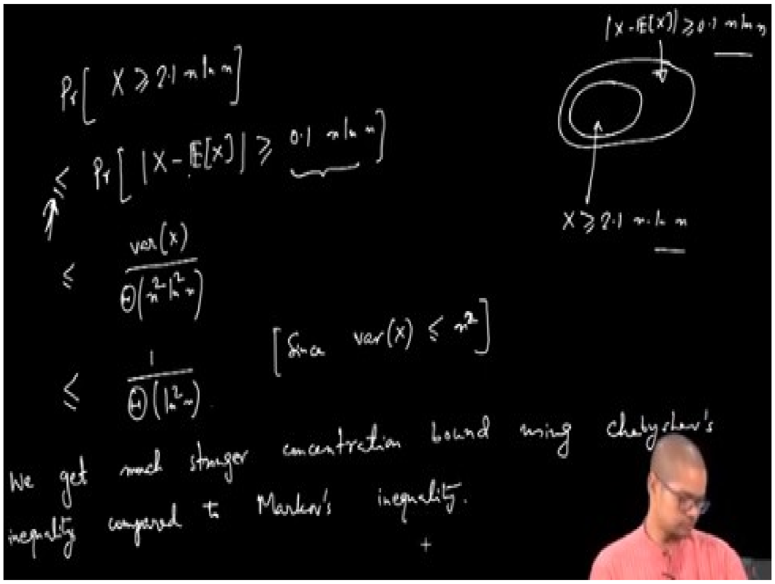
(Refer Slide Time: 10:00)



So what is the theorem? Let X be a random variable with finite expectation mu and finite variance σ^2. Then for any positive real number c we have probability that $|X-\mu|$ deviates from mu by at least c this is bounded by $\frac{\sigma^2}{c^2}$ poof. Probability $|X-\mu|$ is greater than equal to c. This is same as probability that $(X-\mu)^2$ is greater than equal to c^2.

Now $(X-\mu)^2$ this is a positive random variable so I can use Markov and get $\frac{E\left[(X-\mu)^2\right]}{c^2}$ using Markov inequality. And $E\left[(X-\mu)^2\right]$ is nothing but variance this is $\frac{\sigma^2}{c^2}$ which concludes the proof. So let us apply now Chebyshev to get how much the expected number of comparisons can vary from its mean for randomized quick sort.

(Refer Slide Time: 13:34)



So probability that X is an expected number of comparisons that the randomized quick sort algorithm makes on any particular input fix a particular input. And let X be the number of comparisons it is a random variable so it is slightly more than the expectation in $\ln n$ expectation is at most twice in $\ln n$. And what is the probability that the number of comparisons is more than

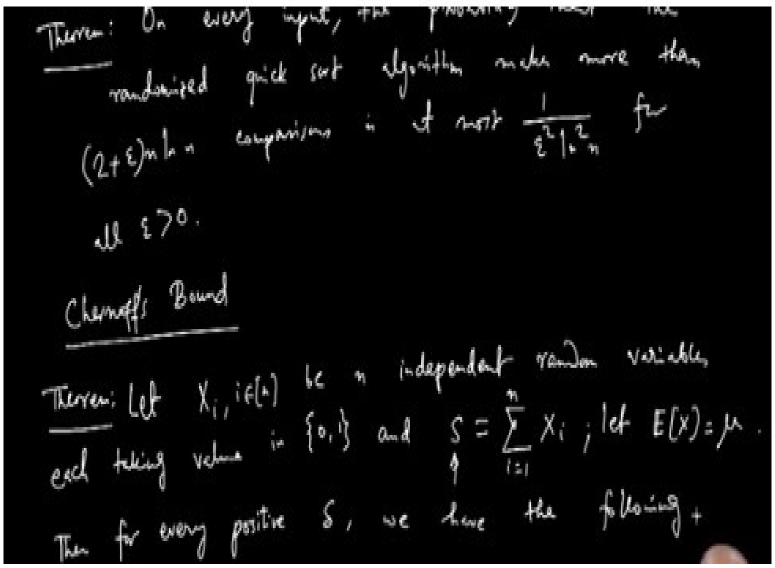
$2.1 n \ln n$.

So this is I can write it this way this is less than equal to probability that X deviates from its expectation this deviation is less than $0.1 \ln n$. Again why I have used this less than because if the deviation is less than $0.1 \ln n$ then X is this deviation is more. So if X deviates from its expectation by more than 0.1 in learning that is a larger event because this is the event where X is more than $2.1 n \ln n$. And in this case and what is this event that X deviated this is a larger event.

Because if X is more than 2.1 in learning then X has divided from its expectation by at least 0.1 in running but it could it could deviate in the other way also, it can be less than its expectation and that less; X takes value or 0.1 in $\ln n$ less than its expectation so that way also it can deviate so this is a larger event. That is why the probability increases this that is why this less than equal to this is mod of X minus expectation of X is greater than equal to $0.1 n \ln n$.

Now this is less than equal to now let us apply Chebyshev variance of X by this square this $\Theta\left(n^2 \ln^2 n\right)$. Now variance is at most, n^2 so this is $\frac{1}{\Theta}\left(\ln^2 n\right)$ in since variance of X is less than n^2. I would let you check as a homework now you see that this bound is this is a much more we get much stronger bound much stronger concentration bound using Chebyshev's inequality compared to Markov's inequality.

(Refer Slide Time: 19:01)

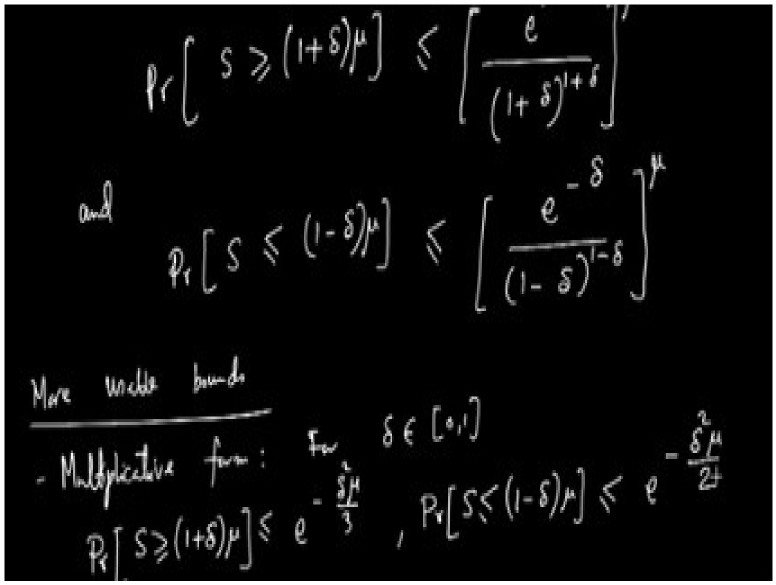


This could be written as a theorem more concretely on every input the probability that the randomized quick sort. Algorithm makes more than 2 plus Epsilon in learning comparisons is at most 1 over Epsilon square Lon square in for all Epsilon greater than 0. Now no wonder that you know using higher moments we can get even stronger concentration inequality and that is the idea of Chernoff bound.

So what is Chernoff bound? Let me write where you know in turn of bound the variables are on or the setup is such that the it means all the moments exist expectation of x to the power i exist for all integers i. So in a specialized setup we write like this let $X_1, \ldots, X_n;\ X_i, i \in [n]$ be an independent random variables each taking value in 0, 1. And $S = \sum_{i=1}^{n} X_i$ you know this so it is a concentration on S.

Now you see that S has is already has much more structure than Markov inequality or Chebyshev's inequality. It is a sum of 0-1 random variables let expectation of X be μ then for every positive delta we have the following.

(Refer Slide Time: 23:14)

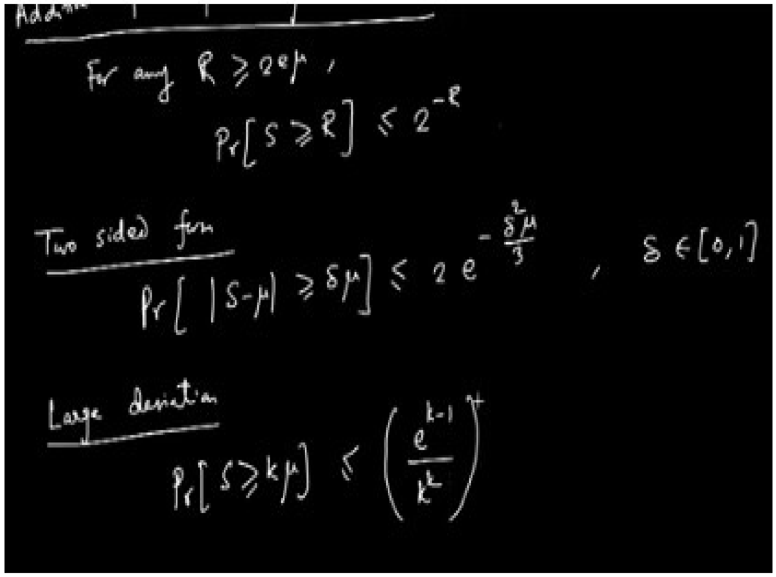


Probability that is deviates from μ by a multiplicative factor of $1+\delta$ is less than equal to $\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$. And probability that S deviates by $1-\delta$ of μ X is less than equal to $1-\delta$ of μ this is $\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$. We will prove this in next class but let me make some useful I mean this is the most tight Chernoff Bound in most tight form.

But we can relax this slightly and make it make it make this bounce more useful or more usable. So more usable bounds these bounds look complicated and difficult to work with. So the first one is multiplicative form what is the multiplicative form? For any $\delta \in [0,1]$ the previous 2 inequalities δ could be any positive number greater than 0 any positive number. But now if $\delta \in [0,1]$ we can simplify this right hand side further and get this.

Probability is greater than equal to $1+\delta$ times μ is less than equal to 2 to the less than equal to $e^{-\delta^2 \mu / 3}$. And probability S is less than equal to $1-\delta$ to the μ $e^{-\delta^2 \mu / 2}$.

(Refer Slide Time: 26:46)



This among the most useful forms of Chernoff bound is the multiplicative form and there is another additive form which is also very useful additive form for large deviation is not applicable for small diffusion. So for any mod greater than equal to twice μ probability that S is sum is greater than equal to R is less than equal to 2^{-R}. There is a 2 sided form which is also very useful probability that $|S-\mu|$ is greater than equal to $\delta\,\mu$.

That S deviates from μ by at least $\delta\,\mu$ this is less than equal to $2e^{-\delta^2 \mu/3}$. This holds for all $\delta \in [0,1]$. This can be obtained from the multiplicative form and there is another one for large deviation probability that S is greater than equal to $k\,\mu$ is less than equal to $\left(\frac{e^{k-1}}{k^k}\right)^{\mu}$, so we will stop here today.