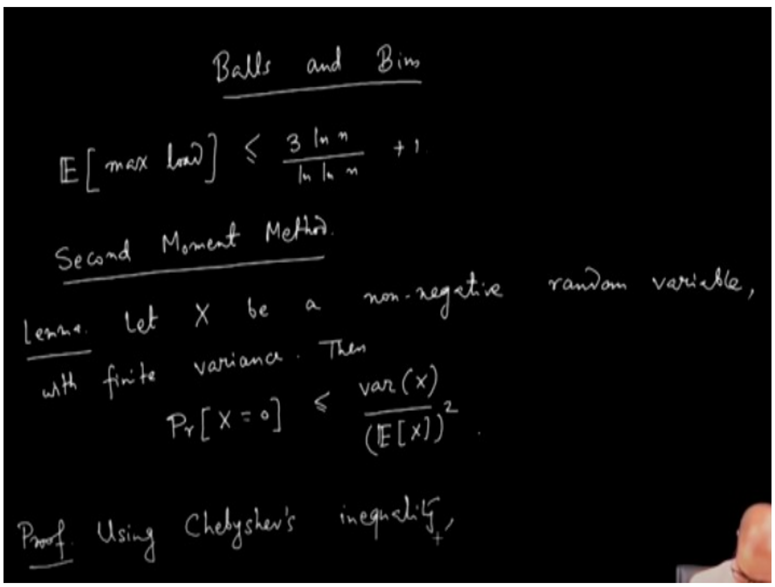
Selected Topics in Algorithm

Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module No # 04
Lecture No # 18
Balls and Bins

Welcome in the last lecture we have studying we were studying the balls and bins experiment and we have bounded the maximum load in today's lecture we will continue that experiment and will show that bound is essentially tight.

(Refer Slide Time: 00:45)

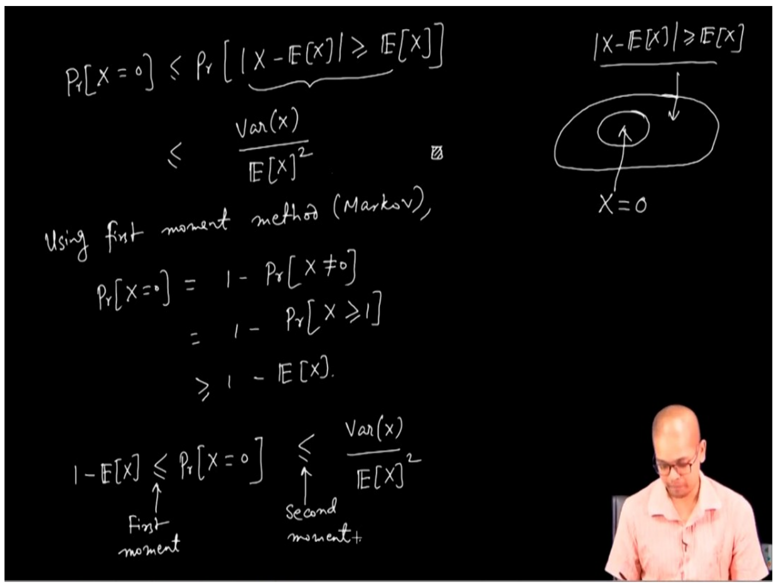


So right is balls and bins and we have shown the last proof last class that expectation of maximum load of any bins is less than equal to $\frac{3\ln n}{\ln\ln n+1}$. So today we will show that you know this is essentially tight and for that what we will use, will show lower bound. And so this is an upper bound and now we should lower bound and for that we need what is called second moment method.

So this second moment method is used to show that you know some random variable takes at least some at least some value it is like you know anti concentration for concentration. We should want to show that you know this random variable ticks what is the probability that random variable takes at most that at most certain value. But this is in this second mode method we can show that it takes at least certain value so the first Lemma the second moment method this Lemma is also often called the second moment mirror.

This is as follows let X be non-negative random variable random variable with finite variance then probability X = 0 is at most $\frac{var(X)}{E[X]^2}$. So let us prove it proof is simple using Chebyshev's inequality.

(Refer Slide Time: 04:11)

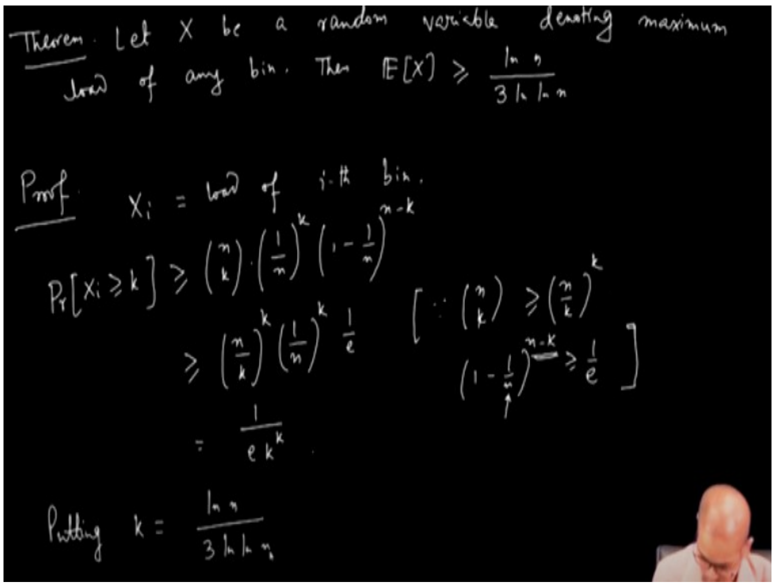


What you get is probability that X = 0 is can be bounded by probability that X minus expectation of X this is greater than equal to expectation of X. Again this inequality follows from if X = 0 then these holes but not the other way that means this is a super event. There is a sample space for mod X minus expectation of X is greater than equal to expectation of X and here is a subset where we have X = 0.

So if X = 0 then this hold this inequality holds but not the other way that is why is less than equal to and now using Chebyshev's inequality this is $\frac{var(X)}{E[X]^2}$ so which concludes the proof so using this. Now let us contrast this with first moment method first movement method is the Markov's inequality using fast moment method probability that X = 0 this is the probability that 1 - probability X not equal to 0.

And this can be bounded by this is 1 - probability that X takes value greater than equal to 1 and this is greater than $1-E[X]$. So what we have is you know probability that X = 0 this is less than equal to $\frac{var(X)}{E[X]^2}$ greater than equal to $1-E[X]$. This is second moment method this is first moment method and this is second moment method. Now using second moment method we will show that our analysis for maximum load of any bin is essentially tight.

(Refer Slide Time: 08:05)



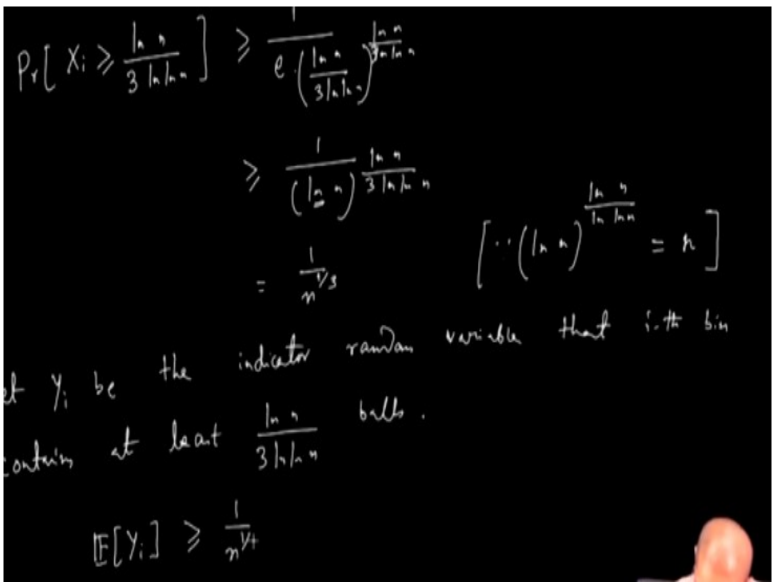
So here is the theorem it is tied up to constant factors again let X be a random variable denoting maximum load of any bin then expectation of X is greater than equal to $\frac{\ln n}{3\ln\ln n}$. So it shows that expectation of X, we earlier we showed that expectation of X is less than equal to $\frac{3\ln n}{\ln\ln n+1}$. Now it shows that it is greater than equal to 3 ln n by 3 ln n run n. So it is essentially tied up to a constant factor of 1. So proof again X_i is the load of ith bin and probability X_i greater than equal to k.

Now before we had upper bounded it now we want to lower bound it and this is ${}^{n}C_k$ I go over all k size sets and all k subset of balls and sum the probabilities that exactly these k balls fall into i th bin and others do not fall so this is really then sub event so this is $\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k}$. So this is greater than equal to now ${}^{n}C_k \geq \left(\frac{n}{k}\right)^k$ is $\left(\frac{1}{n}\right)^k$ and this is most $\frac{1}{e}$.

This can be obscene because you know 1 minus instead of this 1 by n -k. Then it will be close to e asymptotically for large n but this is even bigger 1 by n is bigger and then this we get this inequality. So this is greater than equal to $\frac{1}{e}$ so one way of viewing this n - k if it is n then it is like e.

But this is not in this n - k and each of this is a fraction. So this is greater than 1 by e now by canceling we get 1 by e times k to the power k now putting $k = \frac{\ln n}{3 \ln \ln n}$.

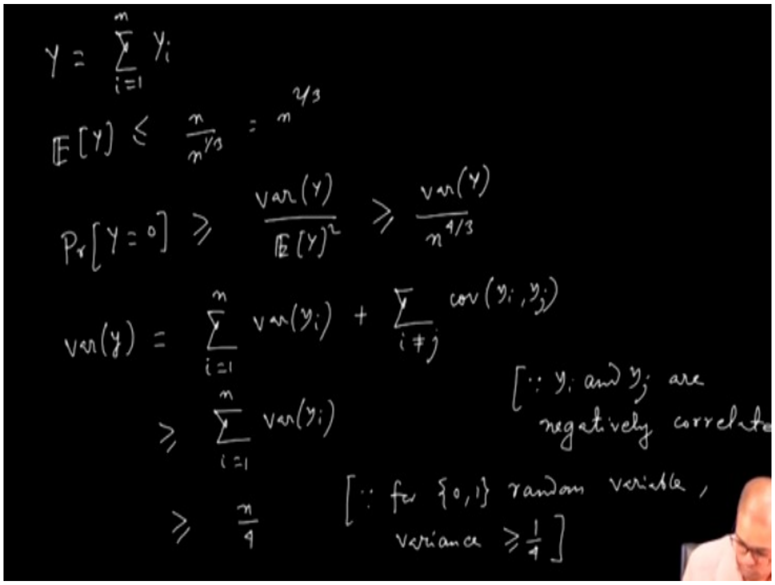
(Refer Slide Time: 12:52)



What we get is probability that X_i is greater than equal to $\frac{\ln n}{3 \ln \ln n}$ is greater than equal to $\frac{1}{ek^k}$ this is $\left(\frac{\ln n}{3 \ln \ln n}\right)^{\frac{\ln n}{\ln \ln n}}$. This is greater than equal to 1 by a ln n to the power here 3 is here $\frac{\ln n}{3 \ln \ln n}$ now exchanging this ln n and n this is $\frac{1}{n^{1/3}}$. This is 1 by n to the 1 third this is because ln n to the power $\frac{\ln n}{\ln \ln n}$ this is 1 is in now what we do is that.

So let Y_i be the indicator random variable that ith bin contains at least $\frac{\ln n}{3\ln\ln n}$ balls. So expectation of Y_i is the probability of this event which is greater than equal to $\frac{n}{n^{1/3}}$.

(Refer Slide Time: 15:59)



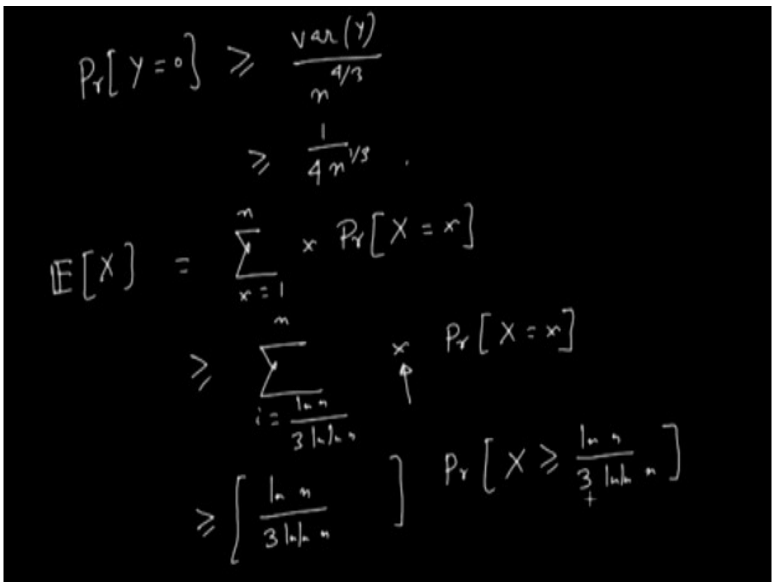
Y is summation Y_i it is the number of bins containing at least $\frac{\ln n}{3\ln\ln n}$ involves. So expectation of Y_i is by n to the 1 third which is n to the 2 third and now recall what is i want to bound the probability that Y = 0. I want to lower bound it i want to show that you know this probability is high if this is the probability is high that means that the expectation you know typically the maximum load is at least $\frac{\ln n}{3\ln\ln n}$ which we want to show.

So again this is by second moment method this is greater than equal to $\frac{var(Y)}{E[Y^2]}$ so this is not equal this is less than equal. So this for expectation of Y we get lower bound which is exactly what we need I want to write this so for that i I did i expect lower bound on expectation of y which I got into the 2/3. So here it is into the 4 third and here is variance and what I need?

I get an upper bound of expectation of Y what we needed but I need a lower bound on variance of Y so what is variance of Y? Variance of Y is $\sum_{i=1}^{n} var(Y_i)$ + covariance terms $\sum_{i\neq j} cov(Y_i, Y_j)$. So covariance terms now if you see that you know if Y_i is more then that increases or that decreases the probability that Y_j is smooth so Y_i and Y_j are negatively correlated and if for negatively correlated random variables covariance's are negative.

So this is greater than equal to $\sum_{i=1}^{n} var(Y_i)$ this is since Y_i and Y_j are negatively correlated and let us see what we get? And variance of Y_i variance of Y_i and see Y_i is an indicator random variable here Y is an integrated random variable in a hence it is a 0. 1 random variable so variance is at so these are 0-1 random variable so variance is less than n. So this for boundary random variable this is at least variances at least 1 fourth this is at least n by 4. Since for 0, 1 random variable variance is greater than equal to 1, 4 and it is of course at most 1.

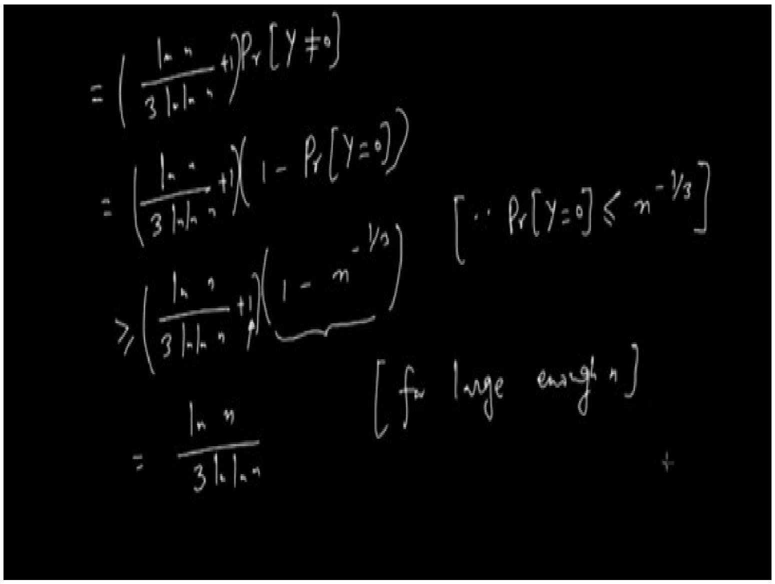
(Refer Slide Time: 20:31)



So now what we get is probability of Y = 0 this is greater than equal to it was variance of Y by n to the 4 by 3 it is from here and variance of y is greater than n by 4 this is greater than 1 by 4 into the 1 third. So now let us come to expectation of X expected maximum load it is again sum of X probability X = x, x = 1 2 this is greater than equal to. If I ignore some term i = $\frac{\ln n}{3\ln\ln n}$ +1 to, n, X probability X = x again in this thing X can be lower bounded with $\frac{\ln n}{3\ln\ln n}$ +1.

This is $\frac{\ln n}{3\ln\ln n}$ +1 and this probability that X is let us sum this from here this is X is greater than equal to $\frac{\ln n}{3\ln\ln n}$. And this is exactly that the probability that Y is not equal to 0 that means there exists at least 1 bin whose load is $\frac{\ln n}{3\ln\ln n}$.

(Refer Slide Time: 23:15)



This is $\frac{\ln n}{3\ln\ln n}$ probability Y not equal to 0 probability that y not equal to 0. So this is $\frac{\ln n}{3\ln\ln n}$ 1 - probability that Y = 0 and probability of Y = 0 is upper bounded by this and also probability of Y less than Y = 0 is lower bounded by. You know let us go back to the first moment method is 1 - expectation of x this is probability that Y = 0 is lower bound by n to the fourth third. So what we have here is this is greater than equal to $\frac{\ln n}{3\ln\ln n}$ 1 - n to the -1 third which is $\frac{\ln n}{3\ln\ln n}$.

So here let us retain +1 keep it +1 and here also +1 this will be needed here +1 and then here we can ignore this is for large enough n so using second moment method. This is probability of X = 0 is less than equal to this will be this is less than not greater than this is less than and expectation of Y this is summation expectation of Y_i and expectation of Y is greater than n to the 1 third. So this will also become greater than $n^{1/3}$ here correct now it is fine.

So now this is from here what we get is this is n by this and here what we need is not the lower bound on upper bound on lower bound on variance what we need is an upper bound. So variance of Y is less than sum of variance of Y_i i = 1 to n and each variance again Y_i is a Bernaulli random variable so each variance is less than equal to 1 this is less than equal to n. So from here we get this which is like 1 by n to the 1 third. So this bound we need variance of y = 0.

So this was again less than equal to this and now it is fine because here since probability of Y = 0 is less than equal to $n^{-1/3}$. This is for large enough n this is close to 1 + term. So this concludes so in the next class what we will see is that how using Chebyshev's method we can use few random coin tosses to boost up the error probability of or the success probability of a randomized algorithm thank you.