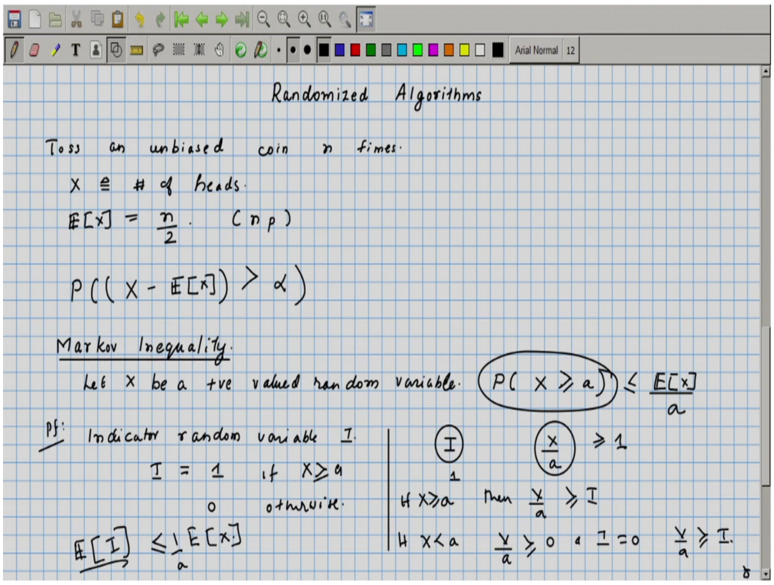
Randomized Algorithms
Prof. Benny George Kenkireth
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture – 08
Markov and Chebychev's Inequalities

So the next thing So, that we have an agenda this plan is we will look at the following problem. Let us say we toss a coin n times ok. So, toss an unbiased coin n times.

(Refer Slide Time: 00:45)



The random variable that we are interested in is the number of heads. We know that the expectation of X is going to be n by 2. If it was a coin of bias p it is going to be n times p. What we are interested in is how much does this random variable deviate from its moment, I mean from its first moment or in other words we will look at various moments of this random variable and look at what is the expected value of those moments, why is this important. Well we are in when we in many of our random randomized algorithms we will have this as a repeated theme. Some experiment is being conducted many times and we expect a certain value that will be our good scenario and the bad scenario will be that the expected value is far the outcome of the experiment is far from the expected value.

So in particular we want to look at probability of such events. The random variable X minus its expectation being greater than some particular value alpha, ok. So, if we want to say that this probability is small ok. There guarantee that we can give our give to our algorithm depends on how well we are able to calculate this probability. So, this probability in case of unbiased coin we will calculate today for different values of alpha. So, our first inequality that we will prove was something called as Markov inequality ok. So, let X be positive valued random variable, we are interested in the probability that X is just greater than some particular value a.

So we know absolutely nothing about the random variable other than that it is a positive value at random variable and we want to compute the probability that it is greater than a particular a ok. So, we will show that this probability is less than or equal to expected value of X divided by a ok. Proof is very simple; we will set up what is called as a an indicator random variable ok. So, I indicates a following event. How does it indicate?.

The value of this random variable i is 1 if X is greater than or equal to a and it is 0 otherwise ok. So, this is a certain event that is the random variable taking values greater than or equal to a that is a certain event. The if a is let us say 100 the random variable takes the value 105, we can say that the event has occurred ok

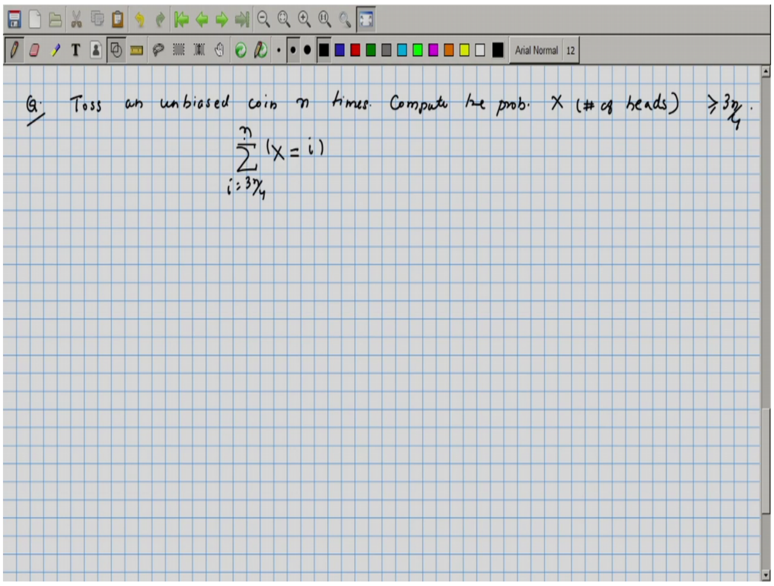
Now, the indicator random variable I that we have defined here defines the same event. It takes the value 1 if X is greater than a and it takes the value 0 otherwise. Now, let us compare 2 different random variables. The first random variable being I and the second random variable being X by a; where a is this particular constant number ok.

If X is a random variable X b is also a random variable and let us look at these 2 random variables point wise. Point wise means on every element in the underlying sample space ok. So, here on every sample point X will take some value some real value, whenever it is greater than a. So, if X is greater than a greater than or equal to a, this random variable will be greater than or equal to 1 ok. Whereas, this random variable in that case will be equal to 1; X is greater than or equal to a, I will take the value 1 and X by a will take the value greater than or equal to 1. So, in the; so, in this case if X is greater than a then X by a is greater than or equal to I. If X is less than a that is the other case remaining if X is less than a by virtue of X being a positive valued random variable X by a is greater than or equal to 0 and I is equal 0.

Therefore even in that case X by a is greater than or equal to I. So, here are 2 random variables X by a and I X by a is always greater than I, therefore, its expectation will also be greater than I ok. So, we can write this as expectation of I is going to be smaller than expectation of X by a; they being a constant we can write it as 1 by a times expectation of X ok. And expectation of I of this indicator random variable is nothing but this probability because I takes only 2 values 1 and 0. So, the expectation will be 1 multiplied by this probability plus 0 multiplied by the probability of the complimentary event. Therefore, the left hand side will be probability that X greater than a expectation of I is going to be this quantity and that is the proof of Markov inequality.

So, now let us ask ourselves this question. Can we apply Markov inequality to compute the probability that if we toss a coin n times, what is the probability that you will get more than three- fourth of the tosses as heads?

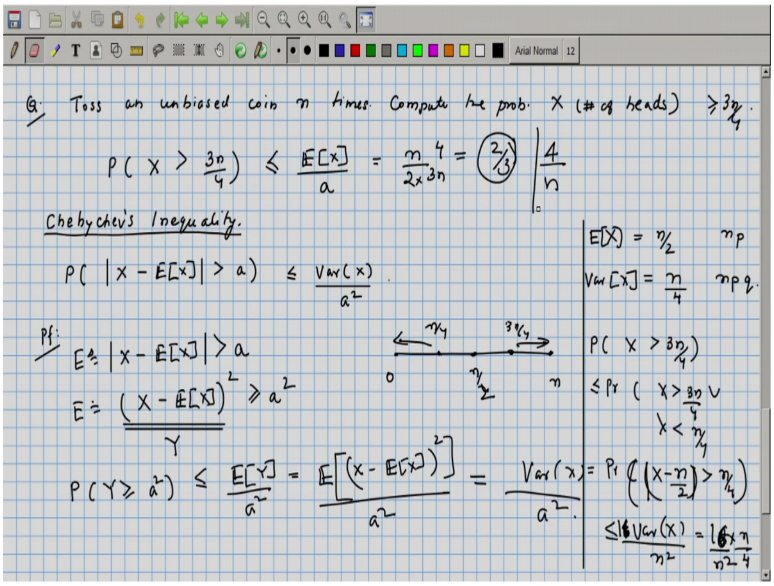
(Refer Slide Time: 08:19)



So, this is the question that we are interested in. Toss an unbiased coin n times, compute the probability that the random variable X which denote the number of heads is probability that X is greater than the random variable X takes a value greater than or equal to 3 n by 4 ok. Of course, we can compute this using our binomial random variable X is the binomial random variable. So, we just need to sum up over the probabilities that X is equal to let us say i waiting from 3 n by 4 to n you will get some value that will be the exact answer.

But you can see that it involves summing up some number of binomial coefficients that is not going to be I mean it can be done, but we want a simpler calculation ok. So, let us see what Markov bound what Markov inequality gives us ok.

(Refer Slide Time: 09:43)



So, we want probability that X is greater than 3 n by 4; this we know is less than or equal to expectation of X divided by a. So, a here is 3 n by 4. So, this and expectation of X is n by 2 and a is 3 n by 4 goes up, so, you will get this as 2 by 3 ok.

So, there is only a less than 60 percent 66 percent chance that this will happen ok, very loose bound. So, we will try and improve this bound ok. You can see that this is very loose bound ok. We will learn something called as Chebychev's inequality. So, here what we are interested in is the probability that X minus expectation of X, the absolute value of that being greater than or equal to some number a ok.

So, here you are interested in the probability that a random variable deviates from its expected value by a specified amount a. So, this we will show is less than variance of X divided by a square the proof was a simple application of Markov inequality, but we will see that Chebyshev inequality gives a better bound than Markov inequality for this particular problem ok.

So, how do we apply Markov inequality to this? So, we are interested in the event that X minus expectation of X the absolute value of that being greater than or equal to a. So, this event is same as the event X minus expectation of X the whole square greater than or equal to a square ok. So, whenever X minus E expectation of X the whole square is greater than a square X is going to be deviating from its mean by an amount more than a. So, these are essentially the same events, but here this is a familiar quantity. This is nothing but the variance of X ok. So, we know that sorry the expectation of this is the variance; this is not the variance, but the expectation of that quantity is the variance. So, the variance comes over here, but this is a random variable which we will denote by let us say the letter Y and we know that it is a positive valued random variable. So, probability that Y is greater than or equal to a square is going to be less than expectation of Y divided by a square that is what Markov inequality says and here expectation of Y is going to be expectation of X minus expectation of X the whole square divided by a square. So, this is going to be variance of X divided by a square ok.

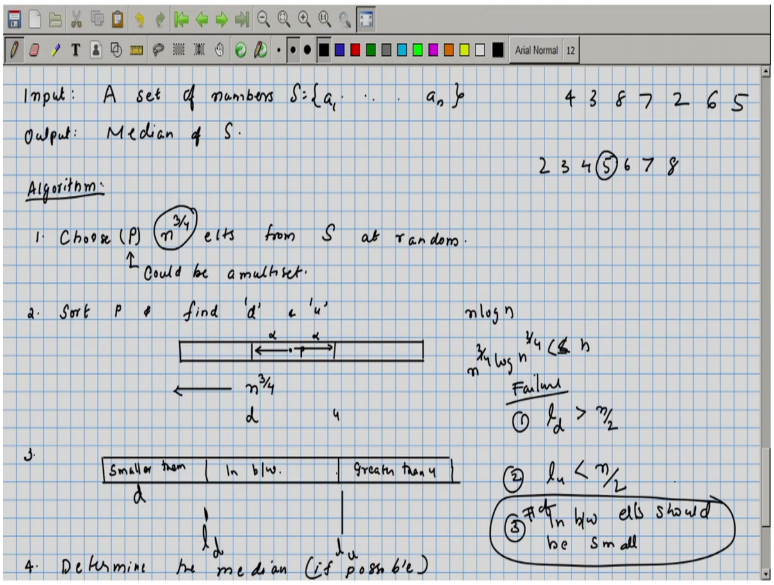
So, now, we will try and apply Chebyshev's inequality to our problem lets toss an unbiased coin n times. So, if you toss a coin unbiased coin n times the random variable X is the Bernoulli random is the binomial random variable. So, the expectation of X expectation of X is going to be n by 2 and variance of X is going to be n by 4 ok. So, if you take a binomial random variable with parameters p and n its expectation is n p and variance is n p into q where q is 1 minus p. So, probability that X is greater than 3 n by 4.

This is certainly less than the probability that X is greater than 3 n by 4 or X greater than X less than n by 4 ok. So, if you look at the values that the random variable X can take can vary from 0 to n and n by 2 being the mean and we are looking at these points 3 n by 4 and n by 4. The probability that it lies in this region is this probability and that is equal to probability that X minus n by 2 greater than, so, the modulus of this is greater than by 4.

So, this is the expression to which we will apply Chebyshev's inequality and that we know that this probability is less than variance of X divided by a square. So, a here is n by 4 square will be n square 4 n square and variance of X is n by 4. So, 4 into n by 4 sorry, this is 16 n by 4 the whole square is 16 n square. So, this will be 16 n by 4 divided by n square. So, the value that we will get is 4 by n square as n becomes large this is a much smaller quantity than this.

So, the probability that if you toss an unbiased coin n times the chances that it will be the number of heads will be greater than 3 n by 4 this significantly smaller than 4 by n. Now, let us think of an algorithm that we will of an algorithm to compute the median of unsorted array. So, let us look at a randomized algorithm the input is the following.

(Refer Slide Time: 17:47)



Input as a set of numbers which we will call 1 which will call a 1 to a n. So, there are n numbers and the output will be the median of this collection ok. So, median would mean that the middlemost element when you sorted it.

So, for example, if your numbers are 4 3 8 7 2 6 5, the sorter order would be 2 3 4 5 6 7 8 and the middle element is going to be 5. We may assume that the total number of elements is odd, if it is even the median is just the sum of the. So, if you have 2 3 4 6 7 8 add these and divide by 2 ok. So, we will just for convenience assume that we have an odd number of elements. There is a deterministic linear time algorithm to compute the median ok. It is a slightly complicated algorithm. Here we will see a very simple algorithm to compute the media and it will work in expected linear time. The analysis is a little more involved, but this is an algorithm that 1 can readily implement ok.

So, let us see the algorithm. So, first step choose and raised to 3 by 4 elements from S at random ok. So, let us call the set as P ok. So, the choices in P are made by picking elements at random from S with replacement. So, P could be a multi set. We may initially assume that the set of elements in S are all distinct, but even then P could have repeated elements. We do this just to I mean it is much easier to analyze this choice than let us say if we do sample with replace without replacement.

The second step is sort P and find d and u ok. So, we have the set P of elements ok, we will sort this and let us say this is the middle point, this is a collection with n raise to 3 by 4 elements, we will go this side a distance of say alpha and we will go a distance alpha is alpha will fix later and this is going to be our d and this is going to be our u. The idea is simple. We will randomly sample our initial set and using that set we will pick 2 elements d and u which will act as pivots ok.

The number of elements that which we chose is n raise to 3 by 4 and the random elements the elements that we pick as d and u are going to be away from the midpoint of P within, but at a distance alpha. The third step is we are going to split the entire set S into 3 parts. So, this is smaller than d this is greater than u and this is going to be the in between ok. So, one can verify that all these 3 steps can be done linear time this is choosing n raise to 3 by 4 elements let takes only n raise 3 by 4 steps and if you have a small sized set small here is n raise to 3 by 4 because n raise to 3 by 4 is less than n and therefore, one can use and use a linear time. So, I mean use an n log in sorting algorithm.

So, if you use an n log in sorting algorithm for n raise to 3 by 4 it is going to take n raise to 3 by 4 times log n raise to 3 by 4 which is certainly less than n, much less than n ok. So, therefore, the sorting and finding d and u can be done efficiently, it can be done in linear time. And once you have this d and u you just need a single scan of the entire set of numbers to identify the numbers which is smaller than d and identify the numbers it is a greater than u and the in between numbers are whatever is remaining. What our algorithm will do is now if this is let us say l u I am sorry l d and if this is l u ok, what we expect is that the middle portion is sufficiently small ok. So, what if this middle portion is sufficiently small and if it contains a median then will determine the median.

So, determine the median if possible. When I write if possible it means if it is possible in linear time to find the median do find it. Otherwise will just fail as we are interested in randomized algorithms failing is perfectly ok, but we should not fail with large probability ok. The probability of failure should be small that is all ok. So, what does it mean to determine the median ok? Suppose this d is greater than n by 2 sorry l d is greater than n by 2; that means, the median essentially lies in this portion ok. If l d is greater than n by 2; that means, median is not lying in the in between region.

Second situation is l u is less than n by 2; that means, this is l u is towards the left of the median in that case also the median is not going to be in between and the third case is if these 2 conditions are satisfied we know that the in between elements contained in the median, but our algorithm is just going to sort those and find the median. So, we want this to be small. So, third condition is so, in between elements should be small, I mean the number of the number of in between elements should be smart.

So, will leave it at this point for today try and figure out what exactly should alpha be, what should be our bound and what should be the small size ok. The only tools that we would require to analyze this algorithm and give reasonable guarantees is say Chebychev's inequality ok. So, you can try various possible values of alpha, but alpha should be chosen appropriately. So, that for that choice of alpha there is a significant probability that the in between elements are small and they will contain the median and so, that will be the final algorithm.