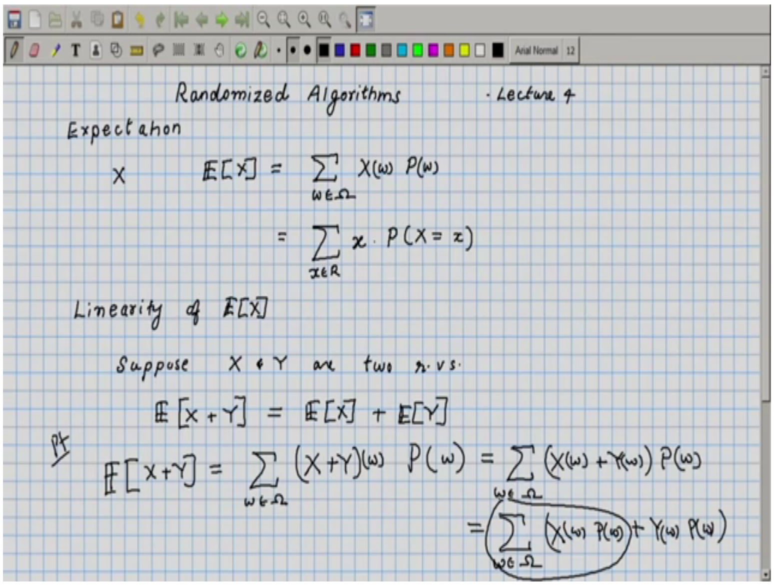
**Randomized Algorithms**
**Prof. Benny George Kenkireth**
**Department of Computer Science & Engineering**
**Indian Institute of Technology, Guwahati**

**Lecture - 05**
**Expectation of Random Variables**

(Refer Slide Time: 00:28)



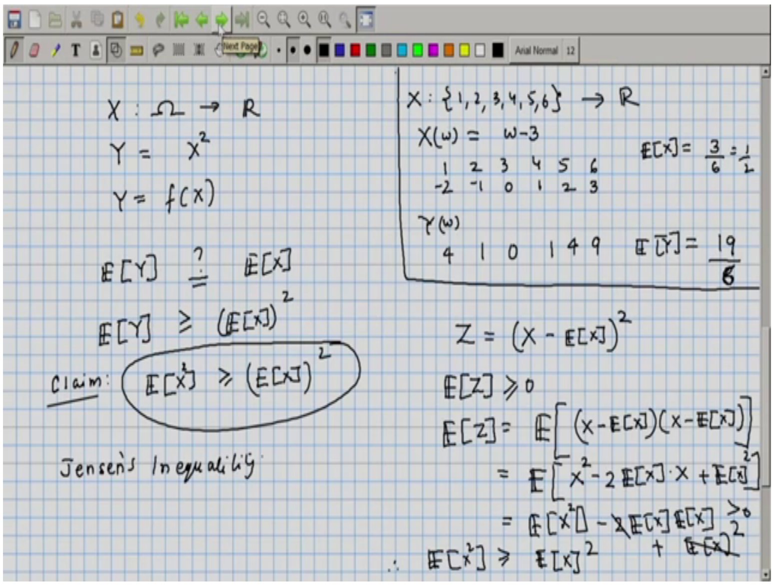

In today's lecture, we will learn more about expectation of random variable. So, if we have a random variable X, the expectation of X we had defined it in earlier class as sum over x belonging to omega; omega belonging to omega X omega probability of w. In other words, this is equal to the values X w denotes the value that the random variable takes, when the sample point is omega.

So, it is this sum over all x belonging to real numbers such that the value of the random variable is x times probability that X is equal to x. This is what we defined as the expectation of a random variable. We will see many properties of random of expectations of random variables today.

The first property that we will learn is known as linearity of expectation ok. So, suppose X and Y are two random variables. We could look at the expectation of X plus Y ok. So, expectation of X plus Y is equal to expectation of X plus expectation of Y, let us see why this is the case. By definition expectation of X plus Y is equal to sum over all omega belonging to omega, X plus Y is a random variable.

So, the value that it takes at the point omega is the value that random variable at the point omega times probability of omega, this is also equal to sum over omega belonging to omega X omega plus Y omega times probability of omega. And this is equal to summation ok, and this clearly is. So, this quantity if you, look at this quantity that is going to be expectation of X and the summation when pushed to the other term is going to be the expectation of Y.

(Refer Slide Time: 04:05)




Now, let us look at an interesting property of expectations. Suppose, X is a random variable ok, so this takes value from omega to reals. And let us look at this random variable Y is equal to X square ok. So, suppose X was the random variable, let us say suppose X is a random variable from 1, 2, 3, 4, 5, 6 to reals, such that X of omega is just let us say omega minus 3 ok. So, on 1 it takes the value minus 2, and 2 it takes the value minus 1, and 3 it takes the value 0, and 4 the value is 1, 5 is 2, and 6 is 3.

The random variable Y omega will be taking the value 4, 1, 0, 1, 4, and 9. We are interested in knowing the relationship between the expectation of X and the expectation of Y, where Y is a function of the random variable X ok. So, Y is equal to f of X. So, here the function is the square. So, how does expectation of Y, and expectation of X relate, what is the relationship between them? We can show that expectation of Y is greater than expectation of X the whole square ok, why is this so well expectation of X here is so let us assume that all these 1, 2, 3, 4, 5, 6 are equally likely. Expectation of X will be 3 by 6, expectation of Y will be 19 by 6.

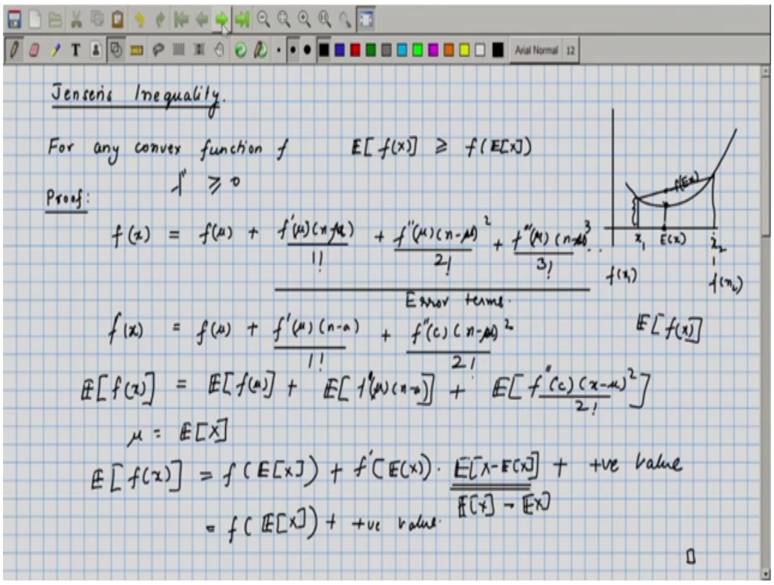
So, we need to compare Y and expectation of X square, this is equal to half. So, squared is going to be less than that this is greater than 1 the square is going to be greater than 1 ok. So, not just for this particular example. This is the case that expectation of Y of X square is always greater than of expectation of X the whole square ok, so this is the case. Expectation of X square is greater than expectation of X the whole square ok. So, let us see why this is the case.

So, let us look at this particular random variable expectation of let us look at this random variable X minus expectation of X the whole square clearly. Z is a random variable, which is always positive ok. As Z is always positive, we know that expectation of Z is also going to be greater than or equal to 0 ok. But, what is the expectation of X expectation of Z, well expectation of Z is equal to the expectation of this whole expression X minus expectation of X times X minus expectation of X, we could expand this out.

And we will get expectation of X square minus 2 times expectation of X times X plus expectation of X the whole square. By linearity of expectation this is going to be equal to expectation of X square minus 2 times expectation of X being a constant that comes out, so we will get 2 times expectation of X, and expectation of X remains plus expectation of X whole square. So, this is expectation of X the whole square one of these cancels. So, what we get is expectation of X the whole square minus expectation of X whole square ok.

So, since this quantity we know is greater than 0, we can conclude this is greater than 0, therefore expectation of X square is greater than expectation of X the whole square ok, so that basically is a proof of this particular statement, then that is something much more general is true ok. So, and the general statement is called as Jensen's inequality ok, so we will state and prove Jensen's inequality that is going to be used many times during our course.

(Refer Slide Time: 10:20)




So, Jensen's inequality states that for any convex function f we have the following property. If you compute the expectation of f X, so let X be a random variable. So, if you compute the expectation of f X that is going to be greater than or equal to the value f of expectation of X ok. How do we view this, how do we prove this see the proof, but intuitively you can think of it in the following way. If you think of a convex function, it is the shape of a convex function. And if you let us say look at a point x 1, and if you look at another point x 2, the average value of x 1 and x 2 will be x 1 plus x 2 by 2, if you take a weighted average, it will be somewhere in between these two points.

So, let us say, so that is going to be the expectation of x. So, let us say if x is a random variable which can take only two possible values namely x 1 and x 2, this is the x this point is the expectation of x. And this is f of expectation of x. If you compute expectation of x, this is going to be f of expectation of x. Whereas this is going to be f of x 1, and this is going to be f of x 2. So, if you look at the function expectation of f X as x varies that is going to be, that is going to be lying on this line. And this is the point which is the expectation of f of x ok. And that is going to be greater than f of expectation of x ok.

Mathematically, what will be a that is a geometric intuitive proof, but mathematically how do we prove this. So, we will assume that f is a twice differentiable function. And for any function, we can write its Taylor series expansion. So, f x can be written as if you expand it about a point let us say mu, you can write it as f of mu plus f prime mu into x minus a by 1 factorial plus f double prime mu into x minus a the whole square by 2 factorial plus the third derivative x minus a whole cube by 3 factorial and so on ok.

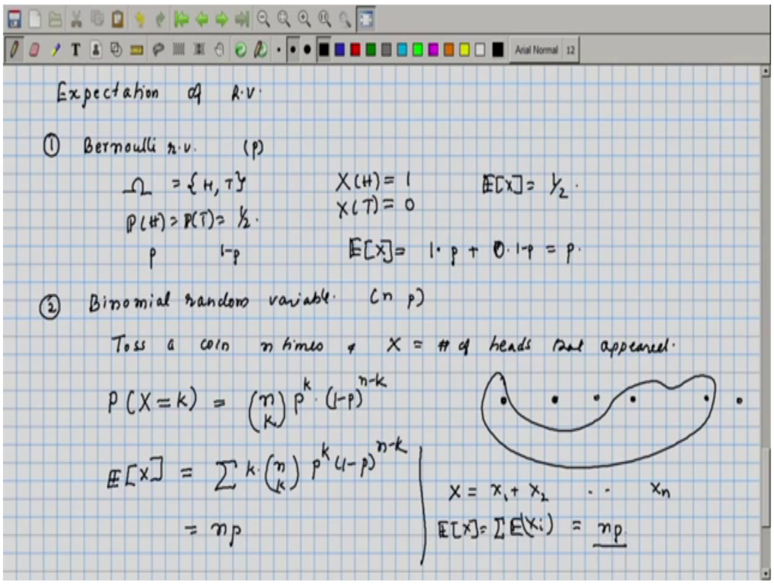
So, these terms can be viewed as the error terms ok. So, if you just approx so this you can think of as a first approximation, combine the first and the second term you get a better approximation and so on. Now, the mean value theorem form of Taylors theorem says that this can be written as f of mu plus f prime mu into x minus a by 1 factorial. So, these higher order terms, we will keep the first derivative, but the higher order terms we can make it equal to f double prime c times x minus a the whole square by 2 factorial, where c is some value between x and a sorry a here being mu ok.

So, now we can just so this is true for any function. Now, if you take x to be a random variable, we can compute the expectation of f x. So, expectation of f x is going to be equal to expectation of f mu plus expectation of f prime mu times x minus a by 1 factorial plus expectation of f double prime mu sorry f double prime c by x minus mu whole square by 2 factorial.

Now, if you choose mu to be the expectation of X ok, so this is the Taylor series expansion, we expand it about the point expectation of X. So, then we will get the left hand side is the quantity that we want expectation of f of x this is going to be equal to expectation of f mu. Mu is the expectation of x which is a constant, so this is going to be just f of expectation of X plus f prime of mu that is going to be f prime of expectation of X that is a constant that comes out times expectation of X minus expectation of X plus. Now since it is a convex function f prime c here is going to be a positive value, convex would essentially mean f double prime is going to be greater than 0, so this quantity is going to be positive and x minus mu whole square by 2 factorial that is also going to be positive. So, we will call this as a plus a positive value ok.

Whereas, this expression it is going to be expectation of X minus expectation of X minus expectation of X. So, this quantity is going to be just expectation of X minus so here we are applying linearity, and expectation of X being a constant will get it as expectation of x. So, this is 0, so we will get this to be f of expectation of X plus some positive number ok. So, if you remove that positive that means, expectation of f x is going to be greater than or equal to this quantity, so that is the Jensen’s inequality proof.

(Refer Slide Time: 18:25)




So, the next thing that we will see is, we will compute the expectation of various random variables commonly occurring random variables ok. The first example that we will see is what is called as a Bernoulli random variable ok. So, the sample space here is the coin toss sample space. And we will assume that probability of H is equal to probability of T is equal to half ok. And the random variable is the following ok. If the coin toss results in a head, then the random variable takes the value 1. And if the random variable takes the value T if the coin toss results in a tail, the random variable takes the value 0 ok. So, this is an example of a Bernoulli random variable. And the expectation clearly is half.

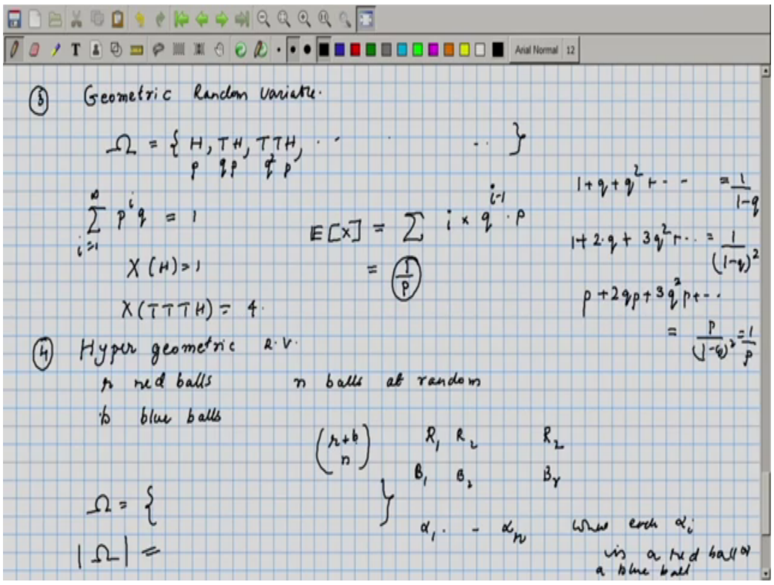
We could also consider the coins being biased. So, let us look at the; so that is also a Bernoulli random variable with parameter p. So, when the parameter is p, the probability of a heads we will assume as p, and the probability of tail will be 1 minus p. And in that case the expectation of this random variable will be 1 times p plus 0 times 1 minus p, which is equal to p. So, this is the Bernoulli random variable.

The second random variable that we will look at is the binomial random variable ok. So, let us consider this experiment of tossing a coin n times, so toss a coin n times and look at the or count the number of heads that appear ok. So, X is equal to number of heads, we will assume that the coins biases p.

So, in that case the probability that the random variable X takes the value K is going to be n choose k into p raise to k times 1 minus p raise to n minus k. So, this is the probability that the number of heads appeared is k. Out of the n tosses, we could pick k of them, and all of them had to be heads that will happen with probability p raise to k. And the remaining had to be tails that will happen with 1 minus p raise to n minus k. So, random variable which has this distribution is called as a binomial random variable. We could also think of binomial random variable with parameter n and p as the sum of n independent coin tosses or in n independent binomial random variable Bernoulli random variables ok.

So, now let us try and compute the expectation of the binomial random variable ok. So, we can compute this as summation k times n choose k times p raise to k into 1 minus p raise to n minus k. This expression will simplify as n times p, but there is another quicker way to compute the expectation of X, we can use the linearity principle. So, if we think of X as the sum of n Bernoulli random variables, then the expectation of X is going to be the sum of expectations of the X i's. And each X i, it is expectation is going to be p, therefore this is going to be n times p.

(Refer Slide Time: 23:27)




The third random variable that we will see today is what is called as a geometric random variable ok. So, let us imagine the following experiment. We keep on tossing coins till we get a head ok. So, our sample space is going to be different from the sample spaces that we have seen so far ok. So, we do not know how many times, we are going to toss the coin ok, but we know that the outcomes can be thought of as H, TH, TTH, and so on ok. So, this is a countable sample space, it is a discrete sample space ok, but it is not the sample space of all possible coin tosses ok.

For example, the outcome T is not present inside the set omega. And the probability is that we can assign to the individual sample points are if we assume that the coins bias is p, so this happens with probability p, this happens with q times p or q is 1 minus p, this happens with q square times p and so on. And you can verify that summation p to the i q, i varying from 1 to infinity, this is going to be equal to 1 ok.

So, what does the geometric random variable say, geometric random variable counts the number of tosses that you have to do before you get the first head ok. So, X of H will be equal to 1, X of TTTH will be equal to 4. And we need to compute the expectation of X clearly, by definition this is going to be equal to summation over the values that the random variable can take, random variable can take the value i with probability q raise to i minus 1 times p that is going to be equal to if you sum this up, we will get 1 by p ok. You can think of this as an arithmetic or geometric series, can look at the following series 1 plus q plus q square so on. Since, q is less than 1, this summation is going to be equal to 1 by 1 minus q.

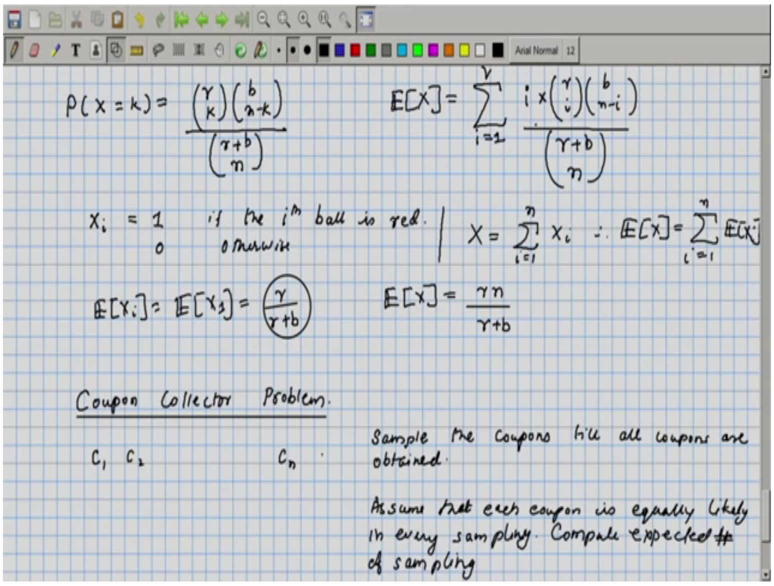
If you differentiate both sides, we will get 1 times so 1 plus 2 times q plus 3 times q square so on. And the differential of this is 1 by 1 minus q the whole square multiply the entire left hand side and right hand side by q, you get q plus 2 q square plus 3 q q so on. This is going to be equal to q by 1 minus q the whole square. And this is the expression that we have here it is if you bring the p outside, so that is going to be equal to so summation q raise to so we just took we did not have multiplied. So, this summation is going to be equal to 1 by 1 minus q the whole square.

So, p into this expression will give you p plus 2 times q p plus 3 times q square p so on ok. So, this is going to be equal to p by 1 minus q is the whole square that is going to be equal to 1 by so 1 minus q is p this is 1 by p ok. So, the expectation of the geometric random variable is 1 by minus p.

The fourth example that we will see is what is called as a hyper geometric distribution or the hyper geometric random variable ok. So, let us imagine the following experiment. We have r red balls, and b blue balls ok, and then we will pick n balls at random. So, this will be uniformly at random. So, the total number of ways in which you could do this is r plus b chose n. And all the possible choices are equally likely ok.

So, the sample space if you write it explicitly, that is going to contain r plus b choose n elements. So, let us say if our balls were numbered, if we thought of them as R 1 we wrote them as R 1, R 2, R r, and B 1, B 2, B r our picks could be a sequence of n balls ok. So, let us say alpha 1 alpha n, where each alpha i is a red ball or a blue ball ok. So, every such choice is going to be included into the set omega. And the size of omega is going to be equal to r plus b choose n.

(Refer Slide Time: 29:59)




Now, we could define a random variable on this probability space, which is so the random variable that we are interested in is a number of red balls, when we choose n balls at random without replacement ok. So, probability that this random variable X is equal to k is going to be equal to r choose k times b choose n minus k divided by r plus b choose k.

And therefore, expectation of this random variable is going to be summation over all possible values that X can take, so values varies from 1 to r, so i varying from 1 to r i that is the value of the random variable times the probability that the random variable takes the value i that is going to be r choose i into b choose n minus i times r plus b I mean divided by r plus b choose k sorry this is not r plus b choose k this is r plus b choose n. Now, how do we calculate the expectation of X? Again we can simplify this expression apply it is a binomial identities and get the correct answer, but these computations can again be done using linearity of expectation ok.

So, let X i denote the random variable the indicate random variable. So, X i equals 1, if the ith ball is red. So, let us say we had picked n balls, and numbered them 1 to n the 1 to i. Whatever is the ith ball, if that ball is red Then we will say that x i equals 1, because 0 otherwise; clearly. X is equal to summation over i X i, where i varies from 1 to n.

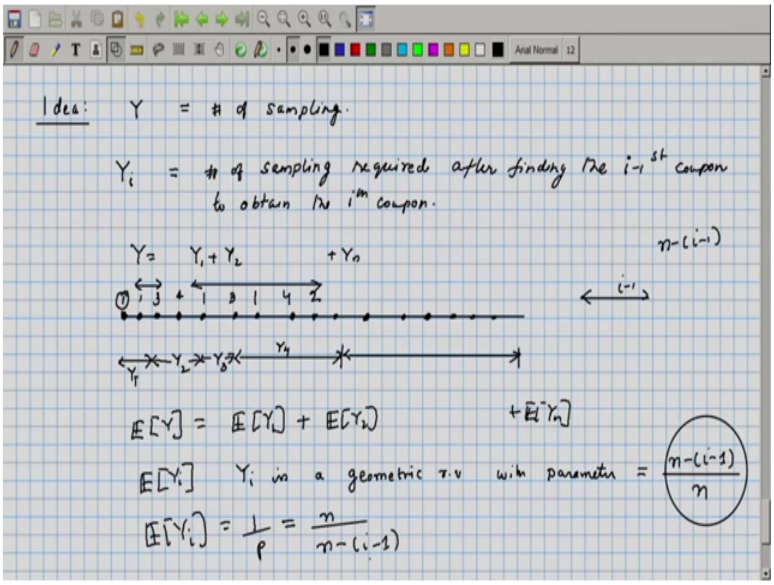
Therefore, expectation of X, there is going to be summation i going from 1 to n expectation of X i. And each of these X i's their expectations has to be equal ok. Because, when we are picking n balls at random uniformly at random, we do not really differentiate between any of the positions, any ball is equally likely to be in any of these positions ok.

And expectation of X i is same as the expectation of any other i so let us say X 1. The first ball can be red or I mean can be red with probability r by r plus b ok. So, this is the probability in the expectation of any X i. So, expectation of X would be r times n divided by r plus b. So, now we have seen many different random variables, and we have computed their expectations.

The next problem that we will address today is something called as the coupon collector problem. So, we will introduce the problem today, and we will work out the details in the next class. So, let us say we have n coupons, which we will call it as c 1, c 2, and c n ok. So, the experiment that we do is the following, we will randomly sample a coupon ok. We will think of the sampling as with replacement.

So, each time we are equally likely to get any one of these coupons c 1 to c n. We will continue doing this, till we get all the coupons ok. What we are interested in is the expected amount of time that we will have to sample in order to get all possible coupons ok. So, we will just write down the requirements. Sample the coupons till all coupons are obtained. Assume that each coupon is equally likely in every sampling. Compute expected number of sampling in order to get all the possible coupons ok, how do we solve this problem. We are going to use the random variables that we have studied so far in order to compute the expectation of this particular random variable ok.

(Refer Slide Time: 35:39)




So, the idea is simple. So, let us say Y is the random variable. So, Y is the number of sampling. We will see Y as a sum of geometric random variables ok. And for any geometric random variables, we know what is expectation is we will apply linearity of expectation and compute the expectation of Y. So, how can we view Y as a sum of geometric random variables. So, what we can imagine this particular experiment, we will keep on taking coupons. And each time we have a collection of coupons that we have already obtained ok. And we can define success as finding a new coupon that we have not discovered so far ok.

So, at some stage let us say we have 5 different coupons out of a total of let us say 20 different coupons. There are 15 other coupons that we can get out of a total of 20 different coupons. So, there is a three-forth chance that our sampling will result in a new coupon. If we do the same analysis, we will see that at any stage we already have certain coupons that we have already picked. The remaining from the remaining set of coupons there is a non-zero probability. There is a fixed probability that well find a new coupon, so that we can think of as a geometric random variable. The some of these geometric random variables one can easily compute.

So, let us introduce some new random variables. So, Y i let us define to be the random variable, which denotes the number of samplings required after finding the i minus 1st coupon to obtain the ith coupon ok. So, at some stage we have already collected i minus

1 different coupon. The additional sampling that we need to do, before we get the ith coupon that is what we will call as Y i clearly.

Y is just Y 1 plus Y 2 plus Y n, we can think of this as the following. So, let us say these are the various coupons that we collected represented by dots on the line ok. So, here let us say we obtained coupon 1, and again the sample we got 1, and then we got a fresh coupon, let us call that a coupon 3, then we got another coupon ok. So, the time that we were in one coupon that is, so that is the time before we got the first coupon. So, this is Y 1, and then we have to do two samplings in order to get the next coupon, so that is going to be Y 2.

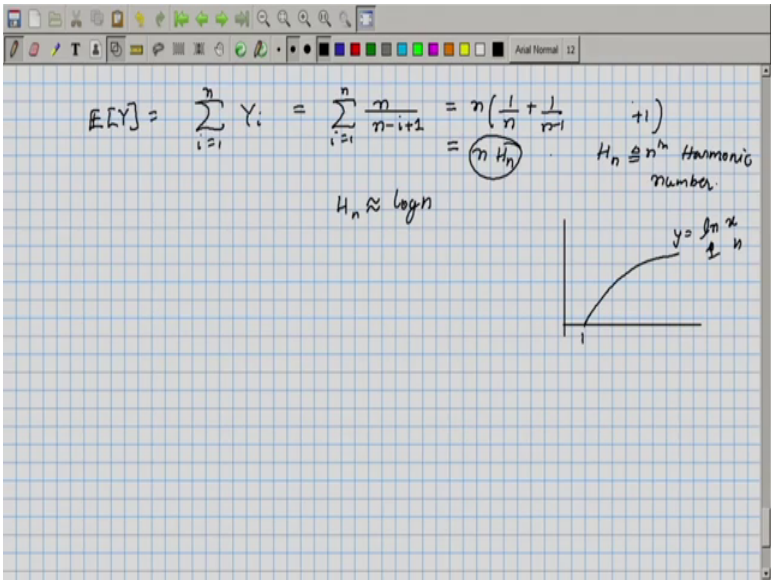
And then the immediate next sampling we got Y the third coupon, so that is Y 3. And maybe next we sample we were getting 1, 3, 1, 4, and after that we got 2. So, we did 5 samplings before we got the fourth token, so that is Y 4 and so on ok. So, this is the time spent after obtaining the 4th token in order to get the 5th token. So, all the samples here essentially must be from one of the samples that we have already obtained ok. So, clearly Y is equal to Y 1 plus Y 2 up to Y n.

And linearity of expectation says expectation of Y is equal to expectation of Y 1 plus expectation of Y 2 plus expectation of Y n ok. So, let us look at what is expectation of Y i, Y i clearly is a geometric random variable, but its parameter is something that we need to determine ok. So, when we are at the when we have already obtained let us say i minus 1 coupons, there are a remaining of n minus i minus 1 coupons.

If we pick any one from there we have a success, otherwise we will need to repeat the sampling process ok. So, Y i we can think of as a geometric random variable. So, Y i is a Y i is a geometric random variable with parameter or the success probability equal to n minus i minus 1 divided by n. Out of n possible coupons, only if we pick, these n minus i minus 1 two coupons, we will consider it a success ok.

So, the expected value of this geometry expected expectation of this geometric random variable is going to be equal to so this is the parameter. So, this is expectation of Y i is equal to 1 by p, where p is this particular quantity. So, this is going to be equal to n divided by n minus i minus 1 ok. Now, we can just sum this up, and that will give us the expectation.

(Refer Slide Time: 42:49)




So, expectation of Y is equal to summation i going from 1 to n Y i, and this is going to be equal to summation, i going from 1 to n, n divided by n minus i plus 1. So, this is going to be n times and i equals 1 this is 1 by n, and the next it will be 1 by n minus 1, and the last term will be plus 1 ok. So, this is equal to n times H n, where H n is the harmonic nth harmonic number ok. H n is approximately log n ok.

You can see this by, if you just integrate the function y equals log x from 0 from 1 to n, so at 1 it is 0, and then it will be some set function ok. So, the area under this curve was going to be till stage n is going to be less than this, and you can bound it. And you can show that it is very much, it is very close to log n. So, the number of times we will have to sample as approximately n times log n ok. So, we will stop today's lecture here, and continue in the next class.