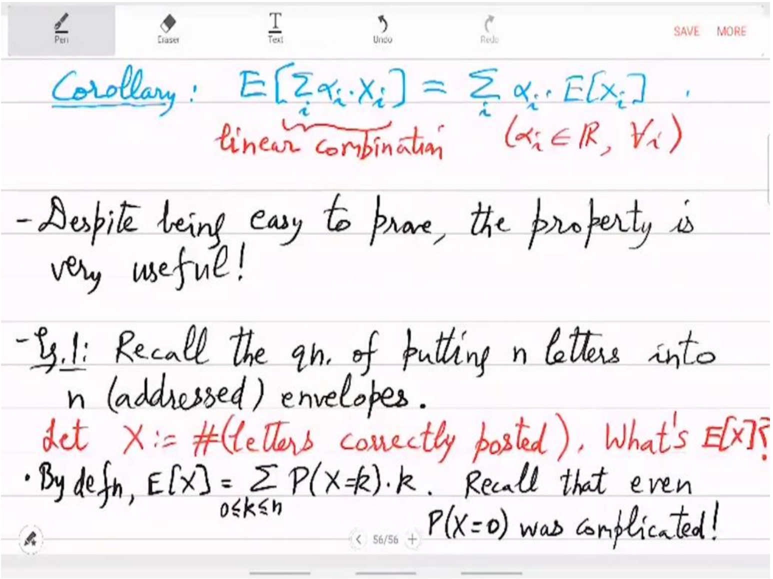
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology - Kanpur

Module - 3
Lecture - 11
Important Random Variables

We proved linearity of expectation, which is that the expectation of X + Y is expectation of X plus expectation of Y, and this was without any assumption. So, you will see in this course again and again the great usefulness of this simple property.

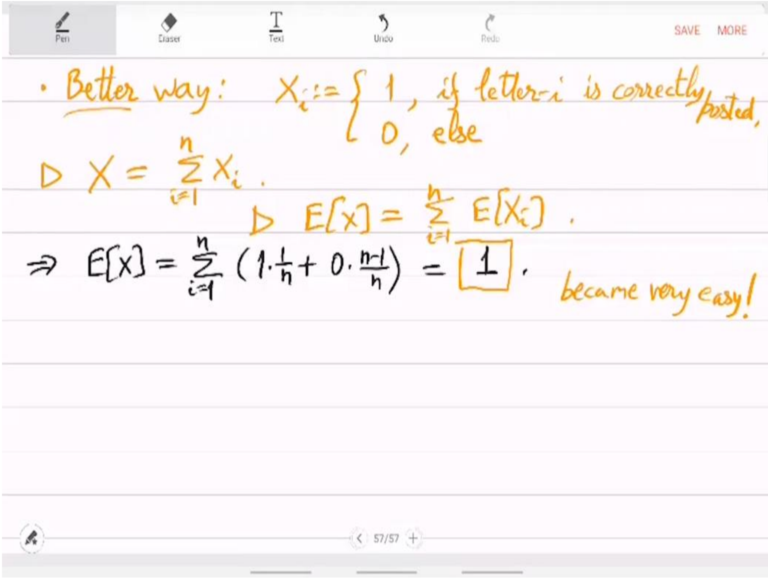
(Refer Slide Time: 00:42)



So, let us start with this example of putting letters into n envelopes which are addressed, in a random way. So, the question is, are the random variable we are interested in is, what is the number of letters which were correctly posted? So, what is expectation of this X? So, if you do it by definition of expectation, so, simply by definition, expectation is probability that X takes value k times the value k, for all k, 0 to n.

So, 0 also because, maybe no letter was posted correctly. And that probability, recall, we have actually evaluated, and it came out to be a complicated expression. It was something like, in the limit it was 1 over e, but otherwise, in general it was quite complicated. So, computing these probabilities is not easy, and hence, this expectation also looks a lengthy calculation that even probability of X equal to 0 was complicated. This, it was an example of inclusion-exclusion principle for probability. So, is there a way to get a sense, better sense of this expectation? So, let us do that.

(Refer Slide Time: 02:21)

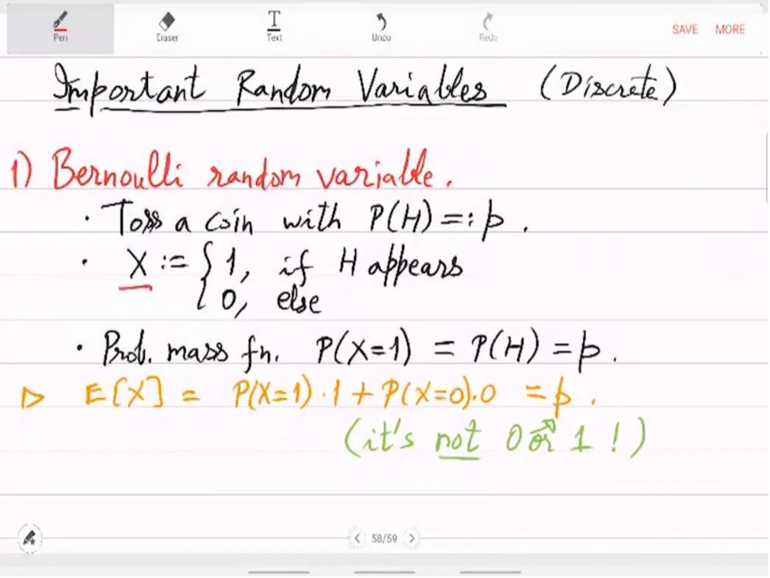


So, define X i to be focused on only the ith letter. So, it is 1 if ith letter is correct, if letter i is correctly placed, and 0 otherwise. So, now, if you go over all the letters, then the sum of these X i's is your variable of interest, which means that expectation of X is expectation of sum which by linearity of expectation is sum of expectation of X i. So, what is that? So, this implies that expectation of X is; so, expectation of X i is essentially the following expression.

So, it will be 1 if the letter i is correctly posted. And what is the probability of that? It is unique, right? So, it is just 1 case, 1 over n; and 0 otherwise, which is n - 1 over n. So, that gives you 1. That is the answer. So, almost immediately, by linearity of expectation, you get that, you expect only 1 letter to be correctly posted. All the n - 1 will be wrongly delivered. That is the expectation, in expectation; it may not happen, but that is what you are getting by probability calculation.

So, this became a really easy calculation. And this is what you will see many times. If you correctly partition, if you correctly use linearity of expectation, then complicated expressions will greatly simplify your life. So, with that example, I hope linearity is clear.

(Refer Slide Time: 04:53)

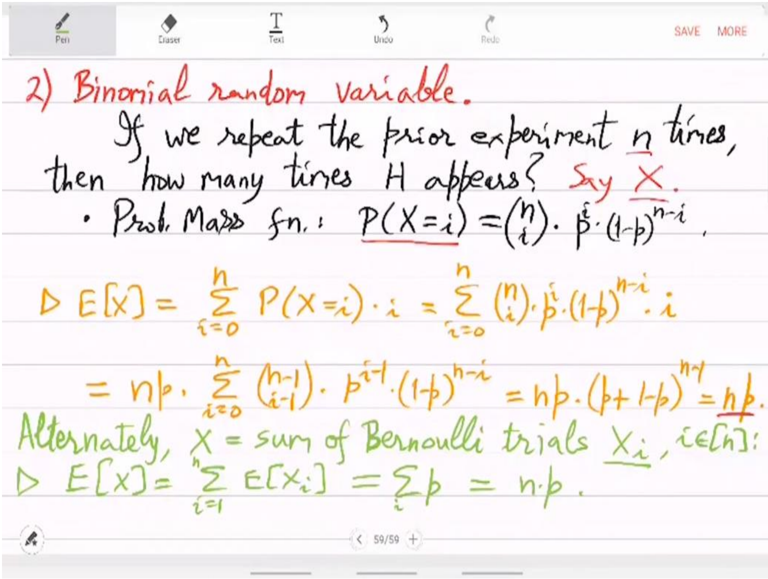


Let us now move to random variables which are famous, which appear in many areas. So, let us start with the discrete first. So, first example is, probably the oldest named example is Bernoulli random variable. This is toss a coin, where the probability of getting a head is not half but p. So, it is a possibly biassed coin. And define the random variable to be 1 if head appears, 0 otherwise.

So, the mass function is, either X takes 0 or 1 value; so, it takes 1 value with probability, same as probability of head, which is p. And the other is X equal to 0; probability is 1 - p. And you have these properties that expectation of X is probability of X equal to 1 times 1 plus probability of X equal to 0 times 0, which is equal to just p. p is the expectation. In particular, you can notice that it is not 0, 1.

So, this is a simple example where the expectation may not actually be a feasible value. In this experiment, the values are only 0 or 1, but expectation is somewhere in the middle, it is p, which can be half or it can be one-third or it can be 1 by 100 or whatever, right? It can be a biassed coin. So, that is the first major example.

(Refer Slide Time: 07:30)



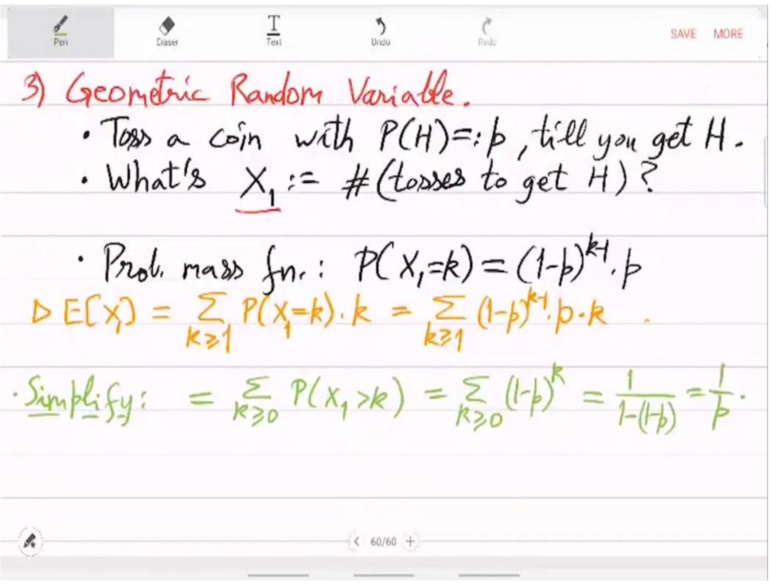
Second is its sum, and it is called binomial random variable. So, if we repeat the prior experiment n times, then how many times head appears? This is what we are interested in. So, this is the random variable X. That is the definition of X; number of heads in n tosses. So, here the mass function is, probability that X = i. So, i heads out of n, exactly. So, that possibility is, which of these tosses were heads? So, that is n choose i many times.

And each time the probability associated is p raised to i times 1 - p raised to n - i. So, that is the mass function. And based on this, you get the expectation. So, expectation is sum of probability X being i times i. It can be either 0 head .. n heads; and which further expands to n choose i times p raised to i times 1 - p raised to n minus i times i; which you can simplify quite easily actually.

So, you just, what you do is, this n choose i times i you write as, bring out np and get n - 1 choose i - 1 p raised to i - 1 n - i. And this thing you see is a binomial expansion; probably that is why we call it also binomial random variable, which is np times p + 1 - p to the n - 1, which is np. So, after this long calculation, what you get is n times p as the expectation. Now, is that a coincidence?

Actually, we can give a simpler proof, which will explain why you should get n times p. So, alternate way of looking at this is, X is the sum of Bernoulli trials. You can write this as a sum of Bernoulli trials or Bernoulli random variables, because, well, each toss is just a Bernoulli random variable. And using that, you can calculate expectation of X to be summation expectation of X i. Expectation of X i, you know is p. So, you get sigma p which is just n times p. This is a more natural way to understand this expectation expression, just n times the Bernoulli expectation.

(Refer Slide Time: 12:11)



Third is geometric random variable. So, this is toss a coin, which is biassed probably, with probability of head being p, till you get heads. You keep tossing it till you get a head. Now, what is the number of tosses? So, what is X 1? I will call it number of tosses to get a head. That is the random variable of interest. So, what is the mass function? Probability that X 1 is k. So, the first k - 1 are, each of those outcomes was tail, and the final one, kth one was head.

So, you get 1 - p to the k - 1 times p. That is the probability mass function. And what is the expectation? Expectation of X 1 is, by definition it is probability that X comes out to be k times k, for all k. That is just the definition, right? And you get then; and k goes from 1 to infinity, p is general; so, you do not immediately see what this value is, how does it depend on p; this is not immediately clear. So, let us do this analysis in a different way, again.

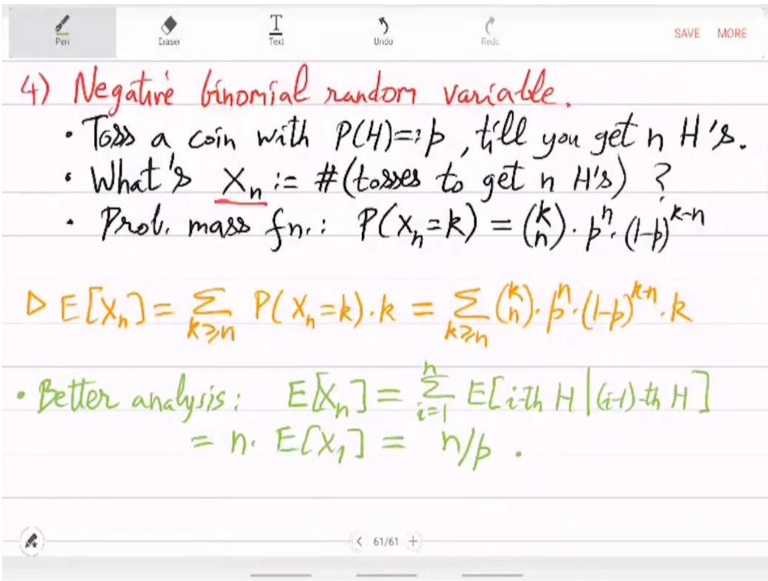
So, let us simplify. Let us rewrite this sum probability X equal to k times k as follows: X 1 greater than k. Let me correct this, this was X 1 and not X. So, I can say this because, in the first X sum, infinite sum, sum value, say k equal to 10, that probability is being multiplied by 10, right? So, you just want to make sure that this probability X 1 equal to 10 in the second sum appears 10 times. So, does that happen?

In the second sum, X 1 equal to 10 probability will be appearing from values 0 to 9; k equal to 0, it contributes; k equal to 1, it contributes; and k equal to 9, it contributes. So, in the second sum, it appears exactly 10 times. At k equal to 10, it stops its contribution, right? So, you can actually rewrite that sum as this. And I did this because, computing this probability becomes easier, it is a simpler expression.

So, probability that X 1 is greater than k basically means that first k were tails. So, that is 1 - p raised to k. And what is this? This is a simple geometric sum up to infinity; that is why the name geometric random variable. And you can see that its value is inverse of 1 minus 1 - p, which is p, 1 over p. So, this is a simple calculation that gives you a 1 over p. And now, thinking about this expression, it is somewhat natural because, the probability of getting a head is p, which is a fraction; so, the expected number of times that you have to flip to get a head should be around 1 over p; also in practice.

If you do this experiment, then you do not expect the head before 1 over p. So, for example, if p was half, then you expect a head to appear when you toss a coin twice; you are not very confident of a head in the first toss, but in the second toss, you get more confident; in 2 tosses, you are more confident. So, that is what this expression is capturing, but you saw the mathematical calculation.

(Refer Slide Time: 17:51)



And finally, we have the negative binomial random variable. What is this? So, toss a coin with probability of head being p, till you get n heads. So, you can compare this with binomial random variable definition. So, there, the experiment was for n times, and you were looking at how many heads. Now, you flip the scenario. You will keep tossing till you get n heads. So, how many tosses? So, what is this random variable X n? Number of tosses to get n heads.

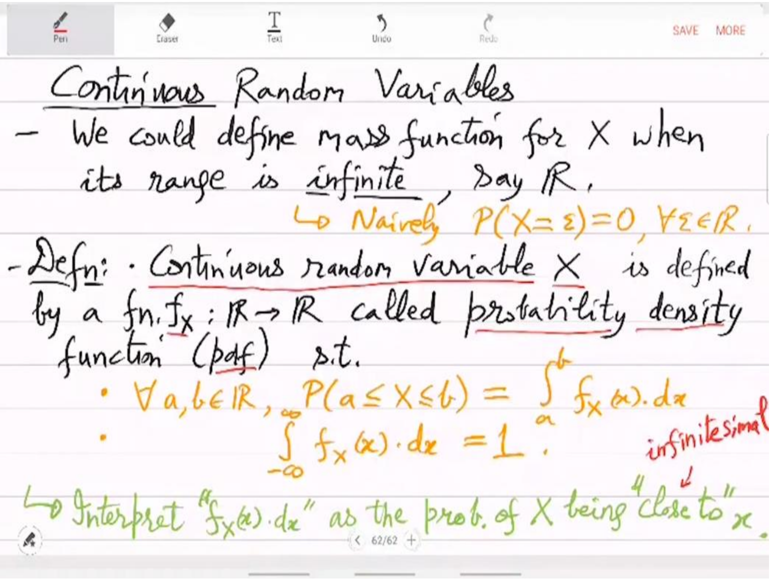
So, I am calling it X n because this is just a generalisation of X 1. Probability mass function, let me first write that. So, probability that X n is k; that is, in k tosses, you get n heads, exactly n heads. That is, k choose n places, you can put the head. So, head is, probability is p raised to n; the rest is, 1 - p raised to k - n. That is the probability mass function, which gives you the expectation. So, to get n heads, you need to toss at least 10 times.

So, you get sum over probability that X n is k times k. And that is just the following expression. So, this is a more complicated expression than before. So, I do not know how to compute this. Let me now give an alternate calculation. So, you can write; see, you want n heads, right? So, what is the expectation to get the first head? So, expectation for the first head is expectation of X 1. Or let me do it like this.

So, ith head, given i minus 1th head. This is what I am looking at. So, first head; then, after the first head, how far away is the second head, after how many tosses? Then the second; after the second head, how many coin toss is required? The third head, and so on. You can actually break up the expectation in this sequence, and that is n times expectation of X 1, which is n over p.

So, believe it or not, this n over p is the same expression as the one in orange, and this is also much better to appreciate. Why do you get n by p? Because, well, 1 head already requires 1 by p, so, n head should require n by p.

(Refer Slide Time: 22:47)



Now, let us look at a different family of random variables, which is continuous random variables. So, what does continuous mean? So, this we will define properly. Obviously, one thing is that you want to work with infinite sample space, that is for sure; when its range is infinite, and it is actually continuous. So, you can say that the range is all the real numbers. And in these real numbers, you are interested now, what is the probability that X takes a value epsilon, epsilon being a real number?

So, naively, it seems to be impossible, because there are so many real numbers that; and then the probability sum, you want to be 1. So, actually, X being some value, some real value, this seems to be 0 probability, in this big space. So, this does not help much or this is kind of inconsistent with our axioms. So, we will actually need some new notions, and let us do that. And that is what we are calling continuous random variable X. So, what is that?

It is defined by a function f from reals to reals, reals to itself, called the probability density function, shortened to PDF, such that; this probability density function will help you to define the mass, but obviously, it will make more sense to talk about a range or interval where X is, instead of focusing on 1 element. 1 element probability will be 0, but what about looking at a bigger set like an interval of reals.

And there, this PDF will help. So, what you want is, first of all; let me put one more thing here. There is a f X, f sub X. Let me clarify that. I will put the big X also. And then, for all a, b, the probability that X is in the interval a, b, that is given by an integral. So, now, integration appears, which might seem shocking at this stage, we did not use integration till now.

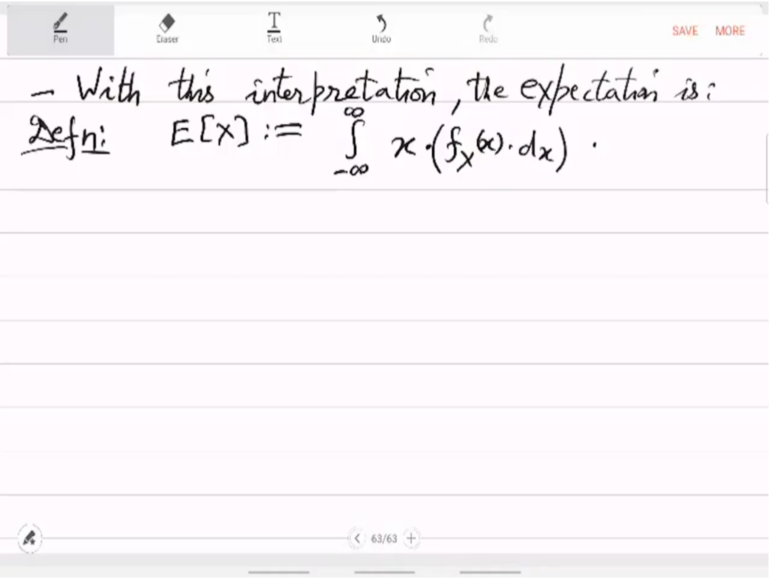
So, I will explain why this is happening, but let us just write down the properties first, which define a PDF, and then through that, a random variable, continuous random variable. Second thing then, what you want is obviously, if you integrate over the whole line, then you should get 1. So, if you integrate in intervals, then you get the probability of the random variable being in that interval. And if you expand it to the whole line, then you get, should get 1.

So, this is what defines a continuous random variable. What is the meaning of this? So, one could interpret this f of X dot dx; can interpret this as a probability; obviously, dx in integration is a formal variable, but you can also think of dx as just an infinity symbol value like epsilon. Integration, you can define as a summing up in these small steps, which then become infinitesimally small.

So, in those terms, this f times dx, it should be interpreted as the probability of X being in the vicinity of x. So, this close to x means infinitesimally close. It is not that you can make sense of big X taking 1 value. But if you look at a small interval around this value small x, and small in the sense of infinitesimal, that it is very small; in the limiting case, it is just x, but formally it is a vicinity of x.

So, X, big X taking a value in that neighbourhood or in that vicinity is given by f of small x times the vicinity size, neighbourhood size. So, once you have this vague interpretation or this, let us say, physical interpretation, then you see that integration is just summing up these probabilities. And the sum is then the mass function for big X. And if you sum up over the whole line, then you get 1.

(Refer Slide Time: 30:27)



So, in that sense, this integration makes sense, is what you need. And with this, the expectation is E of X equals the value small x or kind of in the small vicinity, infinitesimal neighbourhood of small x. What is the probability? That probability is given by f X x times dx. So, this you should think of as the probability. And the first thing is the value. So, that is the expectation.