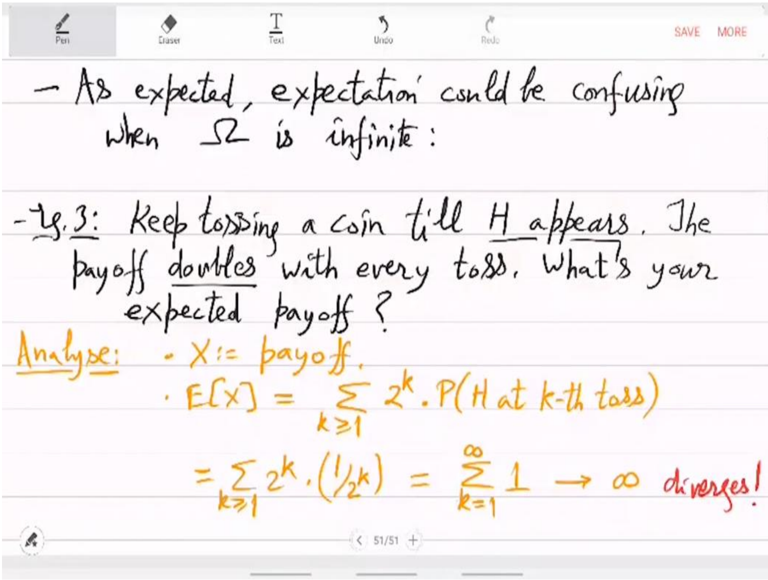
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology - Kanpur

Module - 3
Lecture - 10
Conditional Expectation

(Refer Slide Time: 00:13)



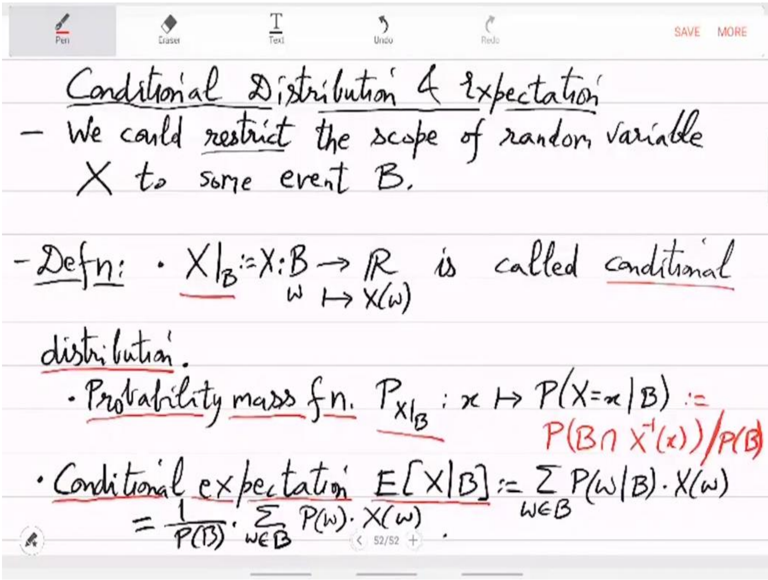
So, this example will show you the problem with infinite sample space. So, suppose you do an experiment where you keep tossing a coin till head appears, and when a head appears, the game ends and you get a payoff. So, the payoff doubles with every toss. So, the later head appears or the longer the game continues, more payoff. So, what is your expected payoff in this game?

So, now, this is a weird game because, if you define X to be your payoff, then expectation of X is what? So, this is, if the game continues for k tosses, so, the kth place you get a head, first time. So, the payoff will be 2 raised to k, because with every toss, the payoff is growing. And what is the probability of this happening? Head appears at kth toss. So, that probability is 1 over 2 raised to k, because everything is fixed.

First k - 1 tail, kth is head; so, that probability is 1 over 2 raised to k. So, you are summing just 1. So, this actually is divergent. So, you can design these games with infinite sample space, where expectation does not make sense; I mean, you will not get a number. So, this is just a diverging to infinity, which is saying that this game is very lucky for you, you will get extremely rich if you play this game, but you do not really get a number.

Finite version of this will be, you play it only, toss it only n times; and then, the expectation comes out to be n, but infinite version is divergent. So, you have to make sure that expectation that the sum actually is, has a convergent value.

(Refer Slide Time: 03:52)



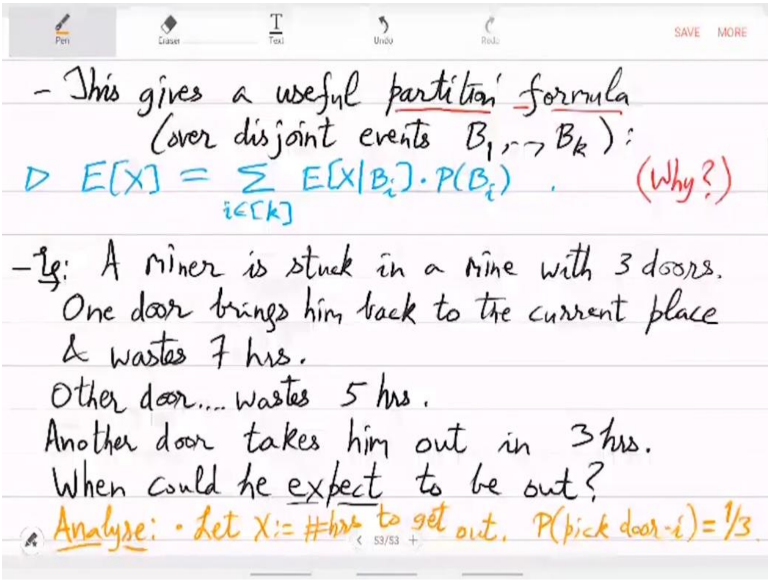
Next is conditional expectation. So, we could restrict the scope of random variable X to an event. There is this natural thing to do with a random variable, you can restrict it to some event B, just like you restrict event A to event B, A given B; similarly, X given B. So, how do you define this formally? So, X given B is just the restriction of X, as it says, it is restriction of X, map X to the domain B. So, you just send omega to X omega.

It is just a restriction, is called conditional distribution. And I can also write, it is just restriction of X to the domain B, from omega. And probability map is, it is P, probability map on this random variable X restricted to B. This is, you map a value x to the probability that big X is this value, given B; which is simply probability of B intersection. So, all those outcomes which lead to little x value, intersect them with B, outcomes in B divided by the probability of B.

So, that is the, you can say probability mass function. So, this is quite intuitive, as you would expect the conditional distribution and conditional probability mass function, and finally the expectation. So, X given B, what is the expectation of that? So, here you will do probability over all the outcomes in B. And what is the numerical value? So, which is also equal to, probability of omega given B is probability omega divided by probability B, conditional probability.

So, you get this. So, conditional expectation is just a, like expectation definition, but you restrict to only outcomes of B and then you divide by probability of B. So, note that, when B is omega, then you recover the original expectation definition. So, this is good to remember, but not really much different from what you have already seen before. Now, let us see an application of this.

(Refer Slide Time: 09:44)



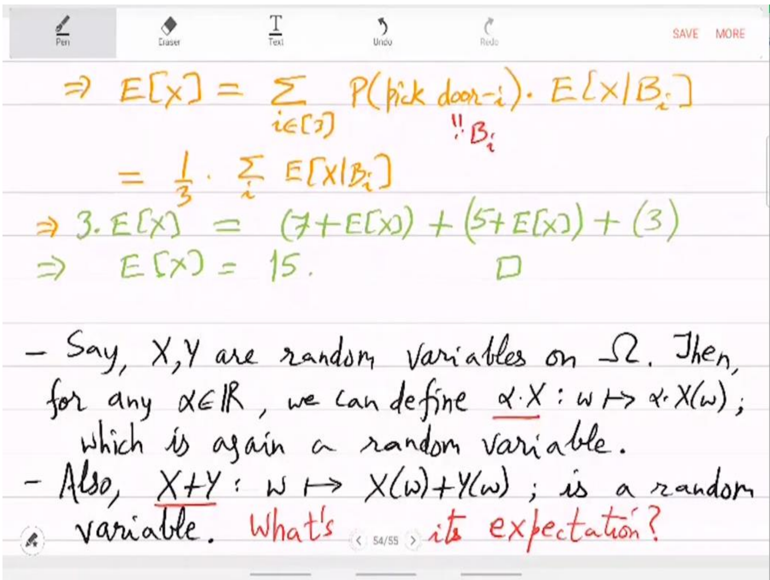
So, this gives a useful partition formula over disjoint events B 1 to B k. So, expectation of X, this can be broken up into conditional distributions or conditional expectations over these disjoint events B 1 to B k, and can be written as expectation of X given B i. But this is not quite correct, right? You should have some weight also. So, expectation of X given B i, but also how likely is B i itself? So, it is times probability of B i; that is the correct formula.

So, from the definition of conditional distribution, conditional expectation, you can actually write this term; and in fact conditional probability, from that you can prove it. So, this I leave as an exercise, formally showing this. So, this partition formula can help you simplify problems by reducing the expectation computation to kind of smaller events which you understand. Let us see an example immediately.

So, say a miner is stuck in a mine with 3 doors. 1 door, if he opens a door and enters, this 1 door will take him, will bring him back to the current place, one without much progress; brings him back to the current place and wastes 7 hours. Other door wastes 5 hours. So, does the same thing. Other door, if the miner takes, then he will come back to the same location wasting 5 hours. And there is a third door which is good, takes him out in 3 hours.

And obviously, miner does not know which door is expected to do what, which door is for what. So, he will just randomly pick 1 and do this. So, if he takes the first two, then he will be wasting time, it will be a cycle, infinite cycle. If he takes this third one, then he will be out in 3 hours. So, the question is, in how many hours is he expected to come out of the mine? So, when could he expect to be out. Let us analyse this. Let X be number of hours to get out. Probability of picking ith door, this is 1 by 3. Let us start with this and find the expectation of X.

(Refer Slide Time: 14:48)



So, this you can write as 3 doors. So, probability of picking ith, times what? So, this is number of hours to come out, to get out, conditioned on picking door i. In fact, use this formula. So, probability; let me simplify this. I wanted to show you the application of that formula. So, you call this event P i. And then, that times; so, expectation of X given B i. B i's are clearly disjoined, because miner will pick only 1 door, cannot pick 2 doors.

So, what do you get? So, one-third for the first one. One-third is the probability, and expectation of X given B 1 and so on. So, this means that 3 times E X is equal to, what is expectation of X given B 1? So, first door, remember, wastes 7 hours, right? And it will first waste 7 hours, and then it will; from that, you will come back to the same place you started, the miner started. So, after that, the expectation is E of X plus X given B 2.

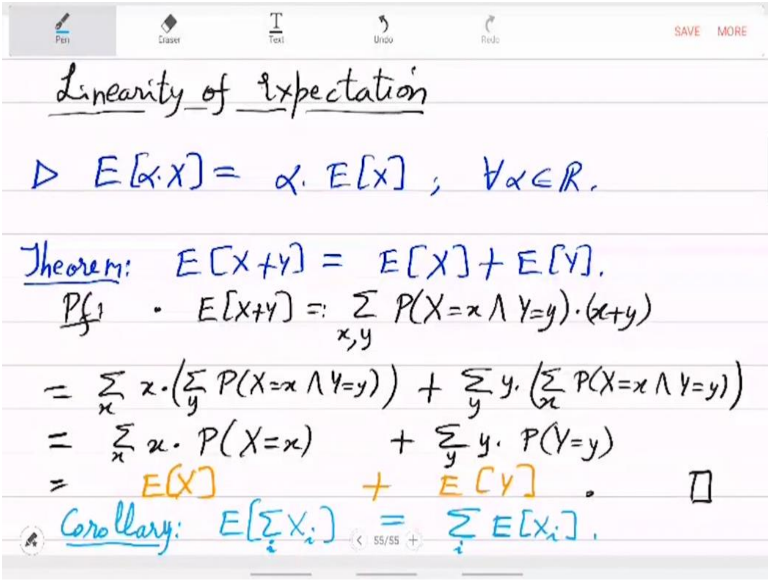
That door will waste 5 hours, come back to the place he has started. And again, from that point onwards, the expectation is E of X plus. The third one only requires 3 hours, and then he is out. So, this means that expectation of X is; so, the beauty of this example, this calculation is that, expectation is unknown, but it gives you an equation, a linear equation, and you get 15 as the answer.

So, that is how this formula of a partition helped you simplify the problem. So, in 15 hours, the miner is expected to come out. Obviously, there are some cases where he will be stuck for a very long time in a cycle, but those things will happen with low probability. So, it is a very good probabilistic algorithm which will ultimately bring the miner out with high probability. So, already, this suggests some kind of linearity of the expectation, and that is what we will do next.

And before that, say X, Y are random variables on a sample space omega. Then, for any constant alpha, we can define another random variable alpha X, which does what? Which just sends outcome omega to the numerical value alpha times X omega, which is again a random variable. And in a similar spirit, we can actually also add up the values of X and Y. So, X + Y, we call it.

What it does on an outcome omega is, adds up the value; is a random variable. So, the question that you should ask here is, what is its expectation? So, what is the expectation of alpha times X? And what is the expectation of X + Y?

(Refer Slide Time: 21:22)



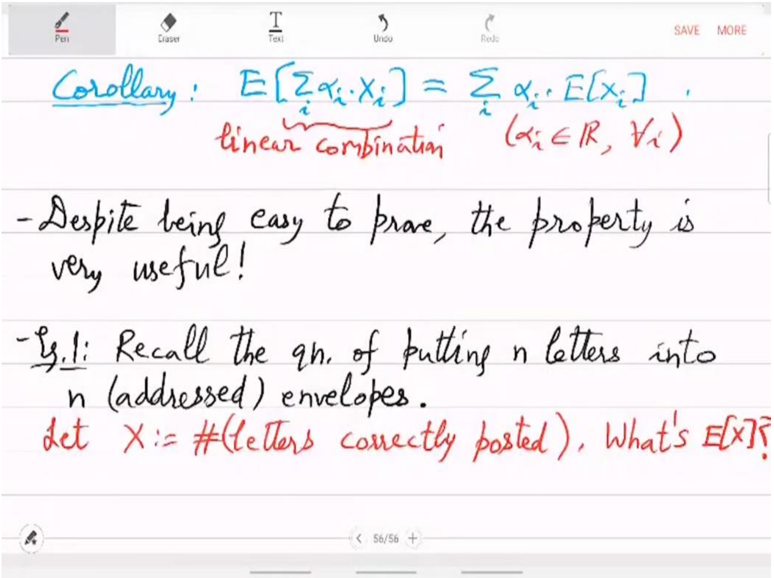
So, first is quite clear, expectation of alpha times X. Since you are only scaling up the numerical value, the whole expectation scales up for all real numbers. This follows from the definition actually. But the second thing is more interesting. So, expectation of X + Y is actually sum of expectations. This is what is called linearity of expectation. So, let us prove this.

So, this left-hand side is equal to, by definition in fact, it is equal to probability that big X is little x and big Y is little y times little x + y being the value obtained; go over all x, y; little x, small x and small y. That is by definition the expectation of the sum, which we can write as; let us first focus on small x; so, x times sum over y probability of X equal to x and Y equal to y; then, similar thing for small y.

And now, what is this sum over y of probability X equal to x and Y equal to y? Since you are going over all the small y's, this is nothing but the probability of first event. So, you get x times the probability of X, and symmetrically obviously, y times the probability of Y, which is expectation of X and expectation of Y. So, this are the very simple proof. You get this linearity property.

And you can also show as a corollary that expectation of sum of random variables is the same as sum over the expectation. So, that is the most general form, summing up over all these random variables, indexed by i. So, the proof is quite simple. And you can also actually apply both these and get the following:

(Refer Slide Time: 24:54)



So, alpha i times X i is alpha i times expectation of X i, where these are constants. So, truly this operator E, it is a linear operator, because it is a very well-behaved on linear combinations. And this is a very useful property. So, despite being easy to prove, this property is usually useful. So, we will do this example next time; the question of putting n letters in n envelopes, n letters into n addressed envelopes.

So, you can ask the question that, if you do this assignment randomly without looking at the addresses, how many letters are correctly addressed, correctly posted? So, let random variable be number of letters correctly posted. So, what is the expectation? As a function of n, how many letters do you think will be correctly delivered or correctly posted, if the assignment was done in a random way?

So, remember that this calculation was quite involved when we did it last time, the number of derangements that all the letters are wrongly posted. So, we will see how linearity of expectation helps you.