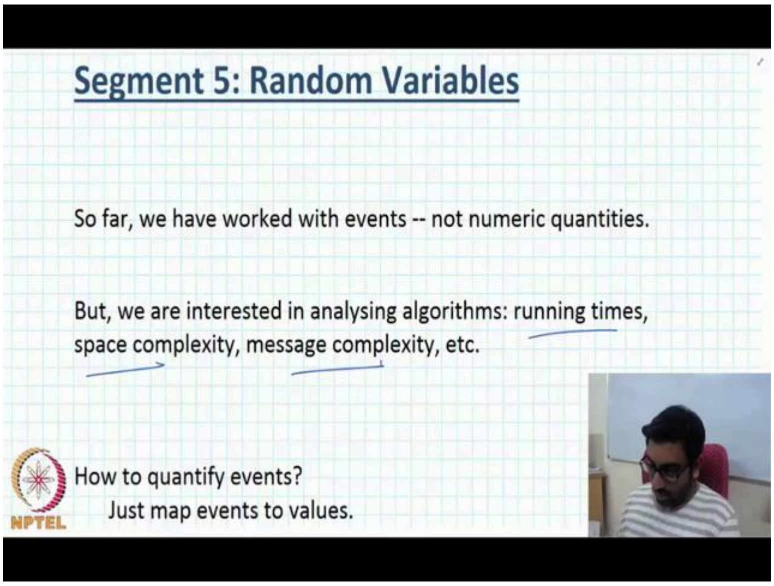
Algorithms for Big Data
Prof. John Ebenezer Augustine
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 06
Random Variables

Hello. So, this is segment 5 and the first lecture, the lecture on probability theory.

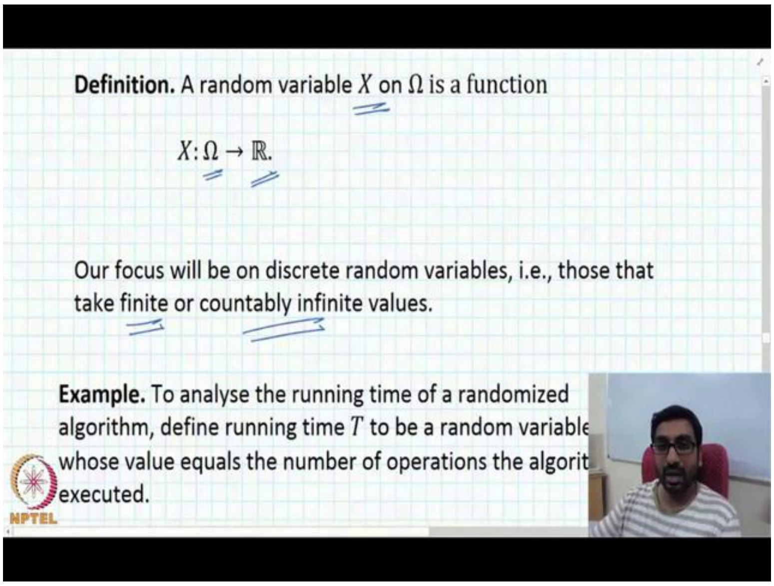
(Refer Slide Time: 00:18)



Quick review, and in this segment we are going to talk about random variables. So, let us motivate notion of Random Variables. We have worked with events. So, far, if the issue with events is that they are not numeric quantities events are (Refer Time: 00:44) with the sample base, often times we are interested in quantities.

For example, in the contexts of analyzing algorithms, randomized algorithms we are interested in understanding running time, space complexity, message complexity, these are all quantities you have ensure that these quantities have small upper bound. So, how do we achieve that using events is the question? To do this all we have to do is to map the events to values, to the quantity values and that is exactly what is achieved by the notion of random variables.

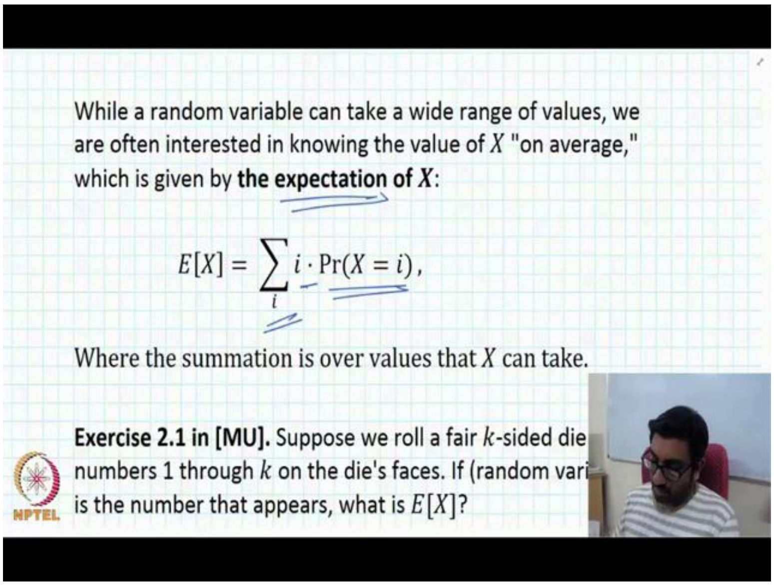
(Refer Slide Time: 01:34)



So, let us look at the definition of a random variable. A random variable x on samples base omega is a function, and this function maps omega to real values. And in particular we will focus on discrete random variables, where the sample space elements will be mapped on to either of finite or accountably infinite, values. So; now, every time an event occurs the random variable x goes track, take on us certain quantitative value and we can use this to quantitative analysis.

As an example consider, the context of analyzing randomized algorithm and we are interested in bounding the running time of this algorithm, and the running time of this algorithm depends on the execution of the algorithm. For the same input because the random chooses made by algorithm, it could execute in different ways, in different attempts. Each possible execution of the algorithm can be thought of an event and the set of all possible execution is the sample space, and for each execution the number of operations performed by their execution is the random variable t. The goal in randomized algorithm analysis is just to gain a good understanding of this particular random variable t.

(Refer Slide Time: 03:42)

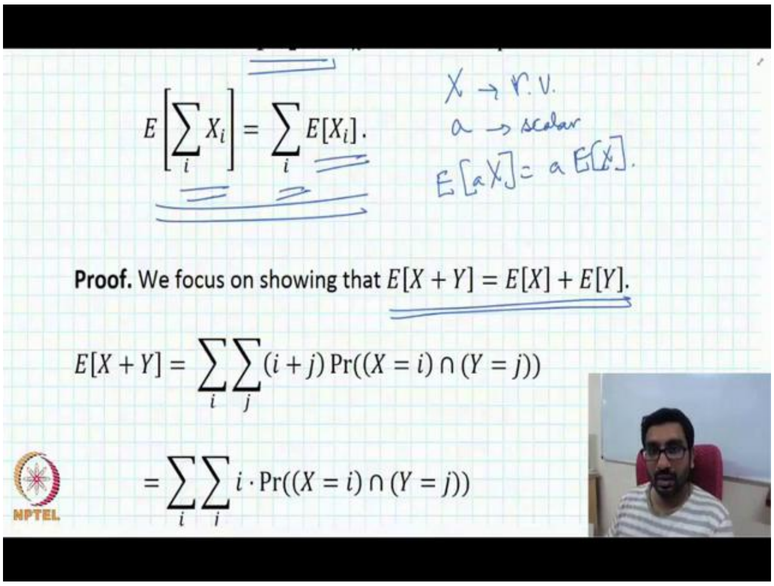


So, random variables can take a wide range of values, going back to our example on analyzing running time of randomized algorithm. A particular algorithm on particular input could possibly take a wide range of running times based on the outcomes of random choices are made by the algorithm.

However, we are after interested in understanding the average running time, what happens on average and to gain and understanding of what happens on. We need to understand the use to notion of expectation. So, let us say excess random variable. The expectation of x is given by the weighted average, its summation overall i where i ranges over the values that x can take i times with the probability or that x actually takes the value i. So, this summation gives us understanding of the average value that x will take.

So, here is a quick exercise for you. Suppose we have a k sided die with numbers one through k on the dies face, now that x being the random variable, that denotes the number that appears when you roll a dice question is what is expectation of x. So, simple application of this formula that we have here but be helpful in that you use to the notion of expectation.

(Refer Slide Time: 05:53)

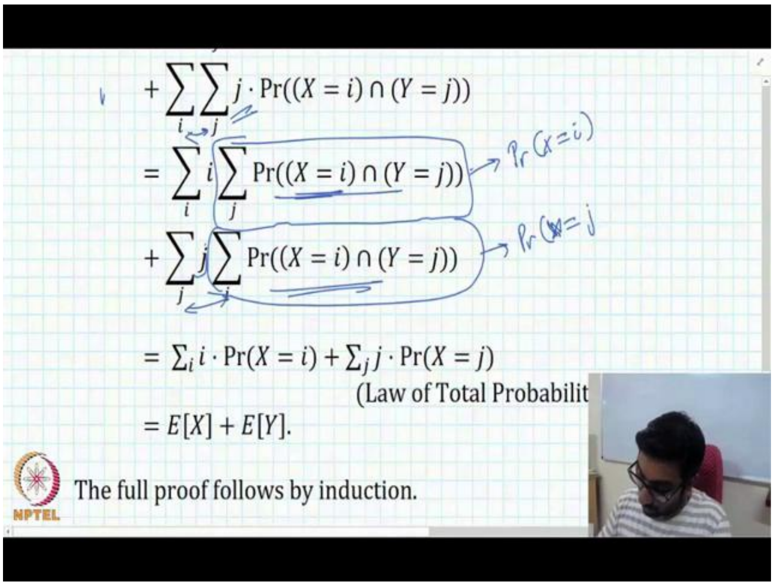


Now, there is very fundamental and their useful property of expectation. Call the linearity of expectation let us suppose we have a finite collection of i have random variables, x1 to x n, and each of them has a finite expectation. Now the expectation of the sum of the random variables, is as general is equal to the sum overall i the expectation of the individual random variables.

This is the very useful property and its call the linearity of expectation, there are some variations to it we also let us say x is a random variable is a random variable, and let us say a is a any scalar value. So, you also know that e of a times x is equal to k times e of x, but let us this is the main form of this linearity of expectation. So, let us try to prove this form, let us focus on proving this particular form where there is 2 random variables x and y. We want to prove that e of x plus y e equals e of x plus e of y.

So, let us start with the left hand side, e of x plus y and x plus y itself is a random variable and therefore, what are the possible values that x plus y can take it, can take or forms i plus j, where i is all the values that x can take, and j is all values y can take. And for each of these i plus j values what is the property that the random variable x plus y will actually take the value i plus j. Well, it is the probability that x individually takes i and y is individually takes j, is a 2 events and when we want both of these events to occur then we will have to take the intersection of these 2 at this.

(Refer Slide Time: 08:55)



Let us proceed and let us expand this expression out and we will. So, what we are going to do now is separate the - i plus j term in to be term with i and a term with j in it. Now is these 2 summations, summation over i summation over j, can be interchanged. So, in the first term we simply take the i out of the summation over change the turn inside the summation of j is takes the same, here we actually note is that we interchange in j that is because these are 2 summations can be interchanged, we get the turn of the j outside of the summation over i and we have the probably turn inside.

Now, shares something that is interesting. Let us look at this particular expression form. (Refer Time: 09:58), going on this is a summation over several j values that these are the value that y can take, and inside of this summation, the events x is equal to i, is unchanged and this event x equal to i is intersected with the events y equal to j for various values of j. So, and y equal to j for various values of j are all destroyed because y can we only take one of them at any time.

So, we can apply the law of total probability. If we apply the law of total probability, this whole term simply becomes the probability of x equal to i, and in similar fraction this whole term becomes the probability that x equals, sorry y equals j, and the by applying the law of total probability. So, now, what we get summation over I, let us look at this term now.

This is summation over i i times that probability there x is equal to i that is nothing, but e of x and here we have summation over j, the j times the probability that y equals j that is nothing, but e of y and which is exactly what we wanted. Now we can extend this proof to more number of random variables, just by using induction. So, that is brings us to the end of our discussion on random variables.