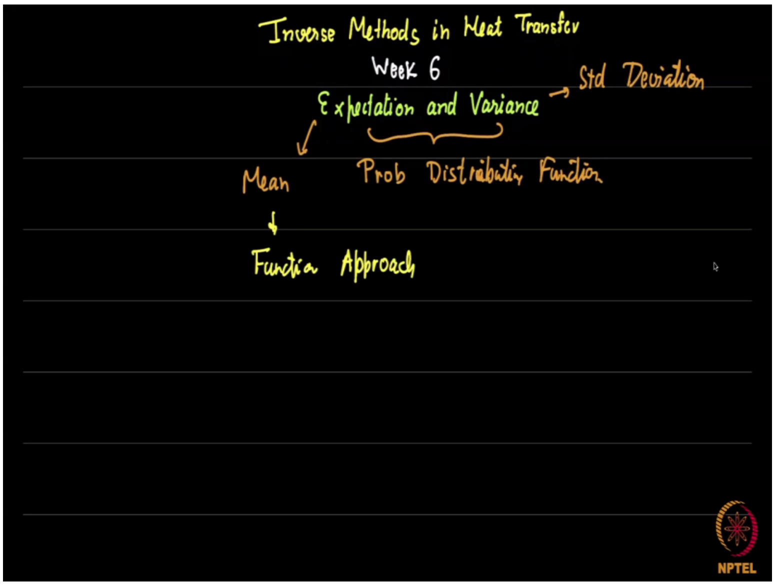
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology – Madras

Lecture – 33
Independence and Expectation

(Refer Slide Time: 00:21)

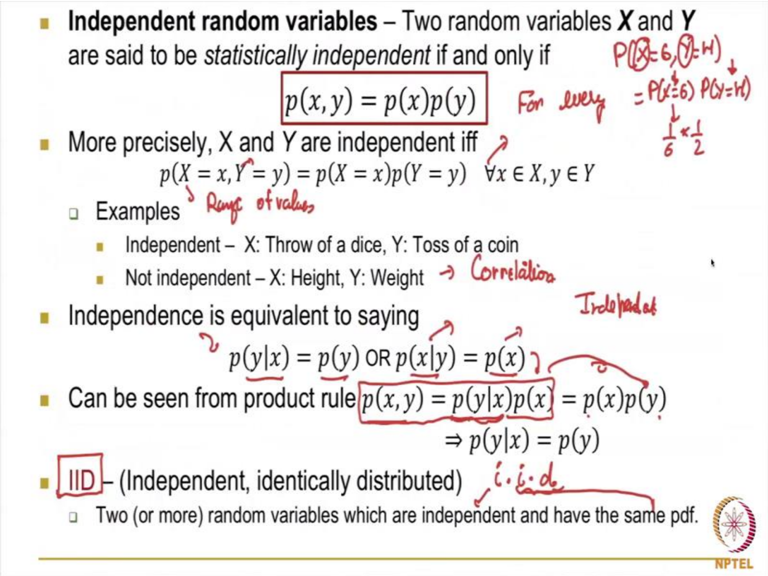


Welcome back, this is week six of inverse methods and heat transfer. We are doing a review of a basic probability theory. In this week in this video, I will primarily be talking about two ideas or two things that we extract from a probability distribution. These are called expectation variance and both these arise out of the probability distribution function or of course, the probability Mass function in case it is a discrete variable.

But our primary idea is to extract these two main things. Now you would have seen expectation as the idea of mean in the usual statistics class and variance of course on standard deviation etcetera are related to these ideas only. There is a slight twist to these and how these become important once we start looking at continuous distributions as well as display distributions when we look at the entire distribution.

So, I will go over that these are very central ideas again. we will use them very regularly within the next week within inverse methods. it will be seen in the next week that it is the expectation, which we were predicting in the usual function approach. We will talk about this more or this kind of point prediction when we come to the next week. So, what we are doing right. Now directly feeds in into what we will be doing next week which will be applying probability techniques directly to inverse methods. So, let us look at a couple of ideas.

(Refer Slide Time: 02:08)



The first idea is that of Independence, you would have seen this again earlier within school within college etcetera. So, two random variables X and Y are said to be statistically independent. If and only if so, this is the definition of independence,

$$p(x, y) = p(x)p(y)$$

Now you have to be really careful when you say this you have to be careful because remember X can take a range of values.

So, if it is continuous all possible ranges of X and if Y is continuous all possible ranges of Y or if let us say X is throwing a dice and Y is my toss of a coin then acts as a sample space of six y has a sample space of two what should be true is $p(X = x, Y = y)$ should be this for every possible value for all values of X and all values of small y. For example, if I ask the question what is the probability that the dice through a 6 and the coin was a head.

You will say this is the same as the dice being 6 multiplied by the probability of the coin being a head. So, this of course is 1 by 6 into 1 by 2, but this is not enough to established that the throwing of a dice and the tossing of a coin are independent. what you have to do is? it should be true for every possible value of x and every possible value of y. So, if x is 5 and Y is etcetera it should be true.

Now an example of independent variable are the things that I have shown physically we realize that the throwing of a dice and the tossing of a coin are independent. Non-independent characteristics are things like height and weight. it does not mean that everybody who is taller will weigh heavier, but there is some correlation and we will come to this a little bit later within this video what this correlation means and how we measure it.

Now independence it is more obviously seen is the equivalent of saying $p(y|x)$ is the same as $p(y)$ or that $p(x|y)$ is $p(x)$. For example, what is the probability that I will throw a dice given that I toss the coin and got a head like why should it depend at all on the tossing of a coin does it. So, $p(x|y)$ is the same as $p(x)$ and similarly $p(y|x)$ is the same as $p(y)$, X has no effect whatsoever on it.

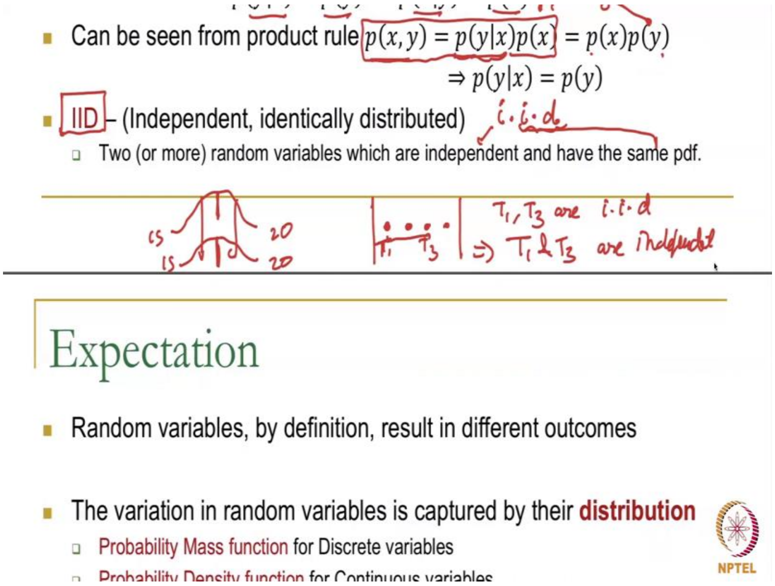
So, now if we apply the product rule here, we say that $p(x, y)$ occurring is the same as,

$$p(x, y) = p(y|x)p(x)$$

which is the product rule. Now we already know of $p(y|x)$, if it is independent is $p(y)$. So, $p(x)p(y)$. so, we can derive this product rule from here or we can derive this rule, $p(x, y) = p(x)p(y)$ from here as well. So, both of these are interchangeable.

Now an important characteristic is this IID often written in small letters iid you will see this multiple times within again especially within the inverse methods literature this means independent and identically distributed. So, what it means is two or more random variables which are independent of course the first line is that and how the same PDF is identically distributed.

(Refer Slide Time: 06:13)



Let us take our slab example. Now if I say that the measurements at 1 and 3 are iid, what this means is that T_1 and T_3 are independent random variables, that is if I find out the probability of T_1 occurring and T_3 occurring together it will be the probability the product of the individual probabilities. Now this might or might not be true in a slab but generally often in inverse methods we assume this to be true.

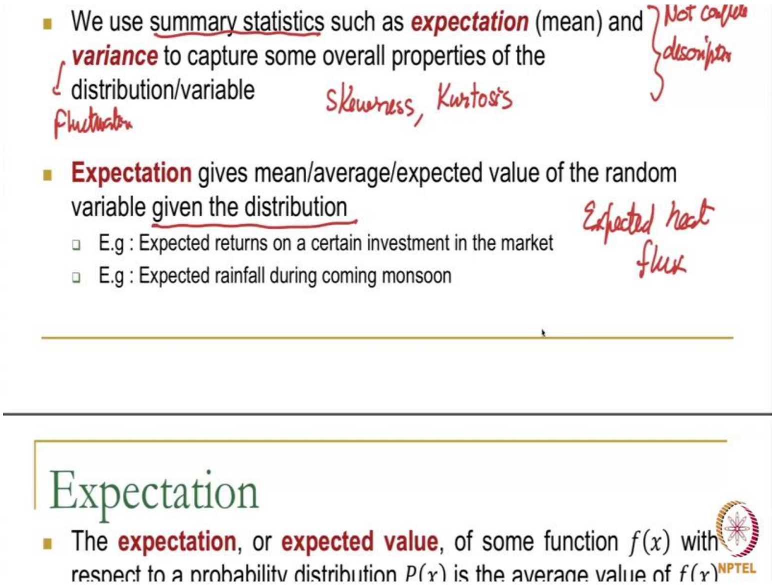
And we will use this in the next week. but why we assume this why it could not be true is of course like the temperature here, let us say it affects the temperature there then maybe the random variables might not be actually independent of each other. So, if one affects the other, they might not be independent. second part is identically distributed what that means is that if both of them have a probability distribution function they look the same.

Now this does not mean that both of them have exactly the same values at each time, obviously that would mean they are not independent. But what it means is when I draw the probability distribution function so they are identically would distribute would mean that a T_1 if it lies between 15 and 20 and has a certain shape then T_2 also will lie between 15 and 20 and have the same shape.

So, the probability of getting 17.5 here and probability of getting 17.5 here would be the same. A Simple example in discrete cases would be let us say you have two coins; you have two friends. both of them are tossing a coin. So, both these events are IID events Q toss a coin and your friend tosses a coin and both of them are independent your toss has nothing to do with the friends toss and similarly both of you will get approximately heads as well as tails with the same probability provided that coins are actually both fair coins.

So, that is the meaning of an IID variables. So, you could have this for two variables, three variables, n number of variables all of these random variables which are independent and have the same PDF are called IID variables.

(Refer Slide Time: 08:49)



Now we come to the idea important idea of expectation. The expectation is simply an extension of the idea of statistical idea of mean. Except mean generally applies to finite samples, but effectively expectation is what you expect in the infinite limit or the very large limit. So, to give you an example, let us say you have two people in within a room and one of them is four feet high and another is six feet high then the mean is simply 5 feet.

But if you ask what is the expected height of a human being within India let us say. So, that would be expectation, it is effectively taking the mean of every person in India. we usually cannot calculate expectation with simple statistics. We can calculate an approximation of that expectation, whereas mean is for a finite sample again to give you an example let us say an election takes place and you wish to find out what is the expected outcome.

So, you would have seen these polls they take polls exit polls from various elections and based on the mean. Mean is what they calculate but expectation is what they estimate expectation is the expected outcome of this election is 53 percent vote for XYZ party and 40 vote for PQR party etcetera. So, expectation is what would happen in case everything was taken to the full sample, whereas mean is what happens for an actual small sample.

So, we will call it infinite limit versus finite limit is what the mean is. So, we call that random variables by definition result in different outcomes and the variation in these random variables basically what happens is captured completely in fact I should say as far as probability is concerned complete information about the randomness in this variable is captured by the distribution.

Now expectation is simply what is called as summary statistic. So, if I ask what is the temperature in Chennai at this point of time, I would say the average temperature is 17 degrees Centigrade actually throughout the day it might have never really hit 17 but I have actually jumped a couple of degrees here there up and down at your actual measurement. But what it is? it is a sample or it is a summary of the sort of temperature variation you saw through the day.

So, it is what is known as a summary statistic. So, 20 of the time you could have been at a certain place, forty percent of the time could have been somewhere else. So, that summary is an expectation that is the mean. The variance tells you how much fluctuation is there as we will see later on. So, some overall properties so, these two are not complete descriptions expectation and variance. What is a complete description of course is the distribution you say how much each variable is likely to take.

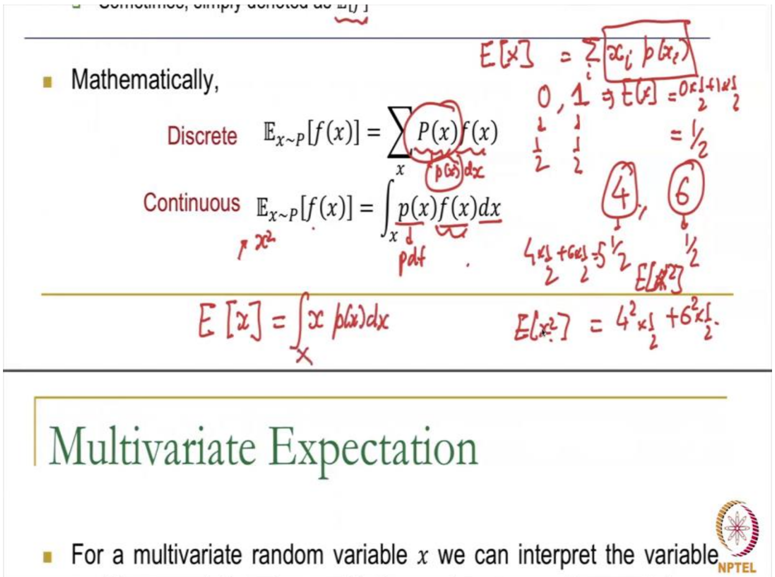
Now that is often difficult to capture you might more easily capture at least what is called the first moment the expectation, the second moment which is variance. Then you have other things called skewness which is the third moment at kurtosis. Here we will not cover these but anyway I have talked about moments etcetera we will see what they mean shortly. So, what is the expectation the expectation gives you the mean average expected value of a random variable given the distribution. So, this is important.

If you knew the full distribution you could calculate the expectation but as you will see well, I will not discuss this in too much detail, that would belong to a full course in statistics. you will not have the full distribution; you will have only some samples of the distribution. So, samples like I said give mean and the distribution gives the expectation. So, you could ask questions like what is the expected return on a certain investment in the market I give this what are the expected returns.

What are the expected returns of this mutual fund, what is the expected rainfall during the coming monsoon, what is the expected heat flux in this configuration. So, similarly you could ask that question. So, typically when we ask for heat flux before in the last four weeks, we were actually asking for the expected heat flux. As you will see the values that we got out of linear regression were actually the expected values of the heat flux obviously we know that given there is the problem is ill-defined it is ill posed.

There will actually be a variation heat flux could vary with various probabilities what we got was the expected value of the heat flux as I will show you in the next week.

(Refer Slide Time: 14:21)



So, what is the definition I am just going to talk a little bit about the notation and then we will come to the definition. The expectation of a function is simply the average value of that function when x is drawn from P. So, here is the notation $x{\sim}P$ means X is drawn from the probability distribution P. what does is drawn from you can just assume it is a box. So, I say the orange is drawn from the box blue and therefore the probability of x being drawn from this distribution of six oranges or two apples is six over eight.

That will seem like a very complex notation, but that is usually the way it is done I am going to try to skip this notation. because it is more important when we do very formal probability and I will just lead you through it but you might see this in books. So, it is my duty to actually tell you what it means. If p is clear where you are drawing from which probability distribution you are drawing from, if it is clear then we simply say $\mathbb{E}_x[f(x)]$.

If x is also clear then we simply say $\mathbb{E}[f(x)]$ or simply we simply say $\mathbb{E}[f]$. Like I said a simple calculation of expectation is simply the mean. So, we will see what that means. So, how do we define expectation mathematically. Mathematically it is very simple I will show you a few examples here, when it becomes clear, but the formula is $P(x)f(x)$.

The simplest expectation is,

$$\mathbb{E}(x) = \sum_i x_i p(x_i)$$

For example, if I toss heads which is 0 and Tails which is 1 and each of these have probability half, this is the probability mass function. then expectation of getting heads or a net expected value of this variable x of this random variable X is 0 times of plus 1 times half which is 1 by 2.

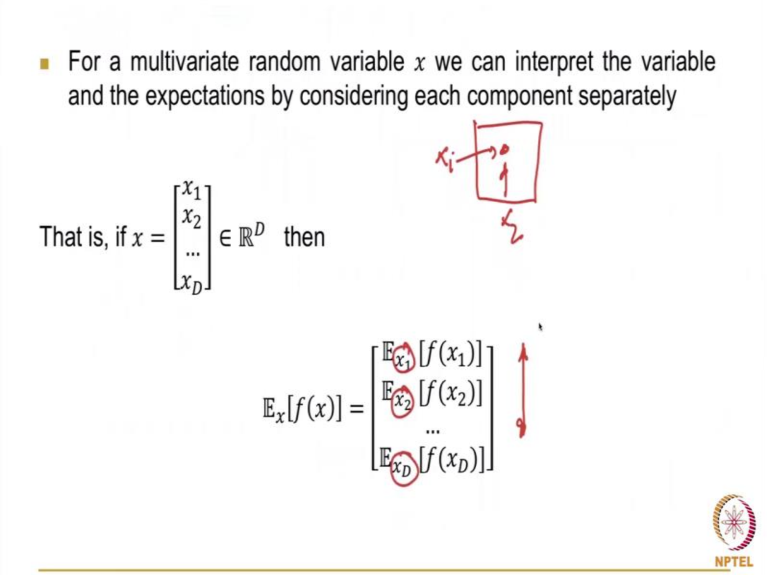
Or there are two people one has height or one has four sweets another person has six sweets; I randomly pick a person with probability half each. What is the expected number of sweets that I will get four into half plus six into half which is 5. Now all these are trivial you can see that this becomes exactly the mean. For example, if something repeats multiple times that is basically why these counts as $xp(x_i)$.

In the continuous case $P(x)$ of course is replaced by $P(x)dx$ as you remember $P(x)$ is capital $P(x)$ per unit length. So, in the continuous case expectation becomes $\int P(x)f(x)dx$, where $P(x)$ is the probability density function and $f(x)$ is the function, you are approximated. More specifically if you want $\mathbb{E}(x)$ just x, not a $f(x)$.

$f(x)$ for example, would be x^2. So, suppose one person has four dollars another person has six dollars and I want expectation of the amount square dollar Square then it will be 4 Square times half plus 6 Square Times half. So, this would be the $\mathbb{E}(x^2)$ where $f(x)$ is x^2. But simplest expectation is $\mathbb{E}(x)$ itself. So,

$$\mathbb{E}(x) = \int_x xP(x)dx$$

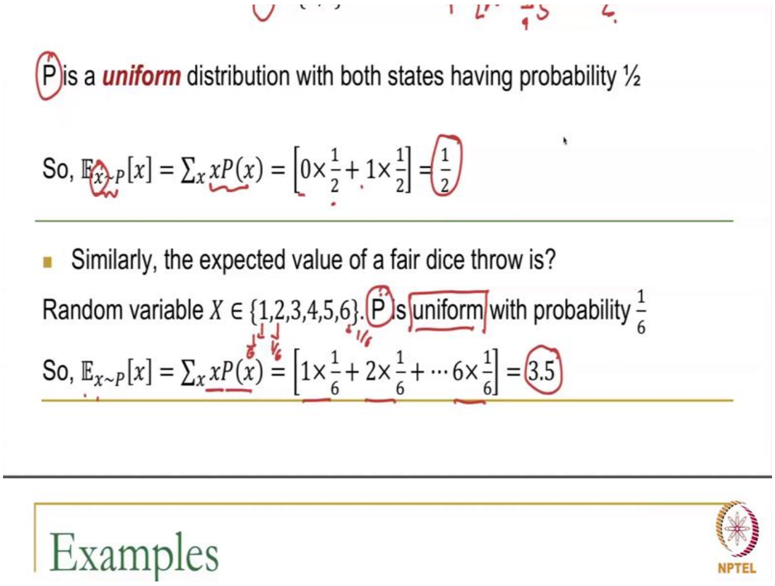
(Refer Slide Time: 19:01)



So, if we have multiple variables and x like in the temperature case is multiple variables, it is not just a temperature at one point but its temperature at six points. Then we just take expectation over all those six variables. Now you can do this in multiple ways similarly x could have multiple components. For example, in a slab at a point this could depend on x_1 and x_2 that is another example.

So, you can look at it as multiple variables or multiple independent variables or multiple dependent variables. In all these cases you can actually take expectations along independent directions. So, you simply list it you just list these number of expectations separately expectation along x and expectation along y. So, you do a partial Independence.

(Refer Slide Time: 19:57)



So, here is a simple example, this is the example that I did a little bit earlier. I am just doing it a little bit more formally here. A good way to think about expectation is as a value. So, notice this word value. So, I cannot say you get heads or tails what is expected you cannot say head and a half. but we can think of it as a betting game. So, a betting game would be something like this you get one rupee if you toss a head and you get zero if you toss a tail.

Then if you do a lot of tosses what is the average amount of earning you will get per toss, that would be the expected value of a single point toss for a fair point. So, now notice I am going to do this formally, the random variable we are choosing is X the probability P from which we are drawing is a uniform distribution it looks like this $P\{X = 0\} = {}^{1}/_{2}$ this is the distribution remember $P\{X = 1\} = {}^{1}/_{2}$.

So, when I say X is drawn from P, X is drawn from here. So, now this x is randomly drawn when it is randomly drawn it either becomes 0 or it becomes one that is the mathematical way of thinking about it. And as we have our formula for expectation as $xP(x)$ this is 0 multiplied by the probability of 0 which is half plus one multiplied by the probability of one which is half. So, the next expected value is half.

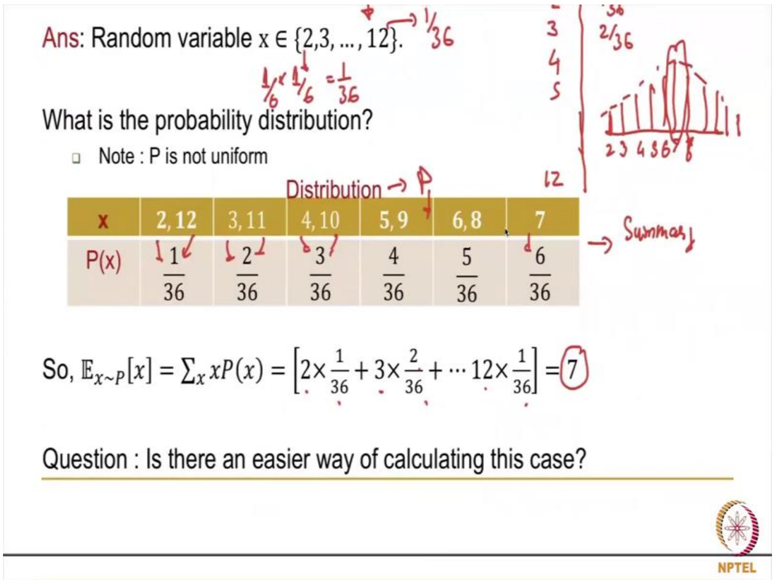
Now similarly you can ask another question if you have a fair dice what is the expected value of a pair dice throw. again, you will never get this value exactly, but if I were to earn one rupee for throwing one two rupees for throwing two etcetera, this is the expected value of earning with one throw. So, that is the mean earning per throw if you average over a large number of throws that is one way of thinking about its P is again a uniform distribution.

So, notice this word uniform distribution means each one of these has the same probability which obviously has to be one over six. So, we now write,

$$\mathbb{E}_{x\sim P}[x] = \sum_{x} xP(x)$$

$P(x)$ which is 1 times 1 by 6 plus 2 times 1 by 6 which is basically a 1 plus 2 plus 3 plus 4 plus 5 plus 6, 21 divided by 6 which is 3.5. So, these are some simple examples of expectation.

(Refer Slide Time: 22:46)



So, let us make the example slightly more complicated and have a 2 it is like we can now look at the joint distribution. Now and see how that affects this expectation. Now let us take the case where two dice are thrown together and once again the random variable is X and X has this sample space. So, X is the sample space which goes from 2 to 12 and what is X? X is the sum of the two dice.

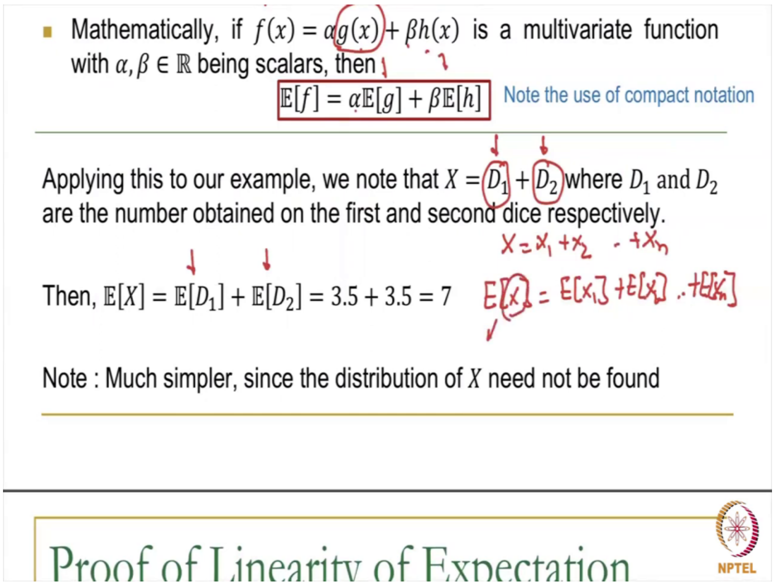
There is no one there because it only will vary between 2 and 12. Now the random variable X has several possibilities you get two only if you get one and one and the possibility of that of course is one by six multiplied by one by six which is 136. Similarly, 12 also has a probability of 1 by 36. So, you can now draw sort of a probability table of value of x versus P of x. So, 2 the probability is 1 by 36 3 can occur as 1, 2 or 2, 1.

So, the probability now becomes 1 by 36 for this and 1 by 36 for this basically 2 by 36. 4 can occur in 3 ways 1 3, 3 1 and 2, 2 so on and so, forth. You can now draw a table so, 2, 3, 4, 5 up until 12 which is what I have written here as a summary of the distribution for P. So, the P distribution is 2 is 1 by 36, 12 is 1 by 36 you can see 7 is 6 by 36 and similarly you have this. If you draw it, it will look like this to 3, 4, 5, 6 and then 7 and then 8, 9, 10, 11 and 12.

So, you will see some such rough shape of course it is not continuous, it is a discrete probability mass function and that is the distribution. Now once again if I ask what is the $\mathbb{E}_{x\sim P}[x] = \sum_x xP(x)$. So, 2 times 1 by 36 plus 3 times 2 by 36 up until 12 times 1 by 36 and in some sense not surprisingly you get 7 which is actually the exact middle of this entire value.

Now as it turns out there was an easier way of calculating this. you might see that there is something suspicious here for one dice it was 3.5 and for two die the sum is actually 7, which also happens to be the sum of their individual distributions. Now if it is 3 is it actually 10.5 or do we have to make this gigantic table with joint probabilities again, turns out we do not because there is something called the linearity of expectation which is an extremely useful property to utilize.

(Refer Slide Time: 26:13)



So, as it turns out, there is an important property of the expectation of random variables, which is that the expectation operator is linear. what does linear mean? linear means this, if you have a function f prefix and it is. So, for example let us say $f(x) = \alpha x^2 + \beta x$. Then if I want to find out $\mathbb{E}[f(x)]$ this will be,

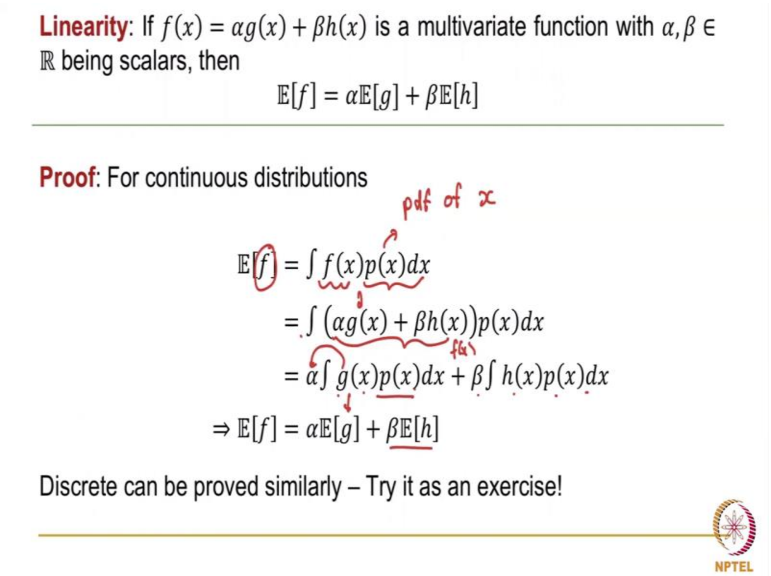
$$\mathbb{E}[f(x)] = \alpha\mathbb{E}[x^2] + \beta\mathbb{E}[x]$$

Another way of writing it is the way I have written it here in the most general case, if we have $\mathbb{E}[g(x)]\alpha$, α is a scalar it is just a multiplying thing this is a function plus $\beta\mathbb{E}[H]$. this turns out to be very useful. So, for example I have x as a function of two random variables D_1 and D_2. Remember D_1 was the outcome of the first dice, D_2 is the outcome of the second dice and X is the sum of these two.

Then using my expectation property, I can simply say expectation of D_1 plus expectation of D_2 will be the expected value of the net sum of the dial which is very powerful. So, similarly if you have $X = x_1 + x_2 + \cdots + x_n$ then $\mathbb{E}[X] = \mathbb{E}[x_1] + \mathbb{E}[x_2] + \cdots + \mathbb{E}[x_n]$. Now this is much simpler to use this because you do not need to find out the distribution of x.

So, remember when we were doing it together, we actually had to find out how many times does 2 occur, how many times does 3 occur, you know how many times 7 will occur three, four, four, three, one, six, six you have to do none of those counts. All you need to do is when one dice came my expectation was 3.5 and 2 dice comes regardless of how these things arrange themselves it is still going to be seven. So, that is a remarkable property.

(Refer Slide Time: 28:47)



Here is a quick proof of this, this will look like it is a complicated proof but it is not. So, we just go directly with the definition. So, the definition was this if I have a function f then expectation was defined as $f(x)P(x)dx$, where $P(x)$ is the PDF of x. Now we already know,

$$f(x) = \alpha g(x) + \beta h(x)$$

So, this is f of x. So, I have just substituted that here open these up, because integral is a linear in α you can take this α out,

$$\alpha \int g(x)P(x)dx + \beta \int h(x)P(x)dx$$

And this of course is expectation of g because this always is PDF of x, not PDF of g of x, similarly this is beta expectation of h. You can prove this case discretely also I mean I just showed it for a continuous case because integrals are a little bit easier to look at rather than summations in my opinion at least. So, but this can be proved similarly you can try this as a very simple exercise.

So, what we have seen. So, far is expectation. I had intended to show variance also in the same video because but this video has already gotten a little bit longer. So, I will show you variance what variances are starting in the next video. So, I will see in the next video, thank you.