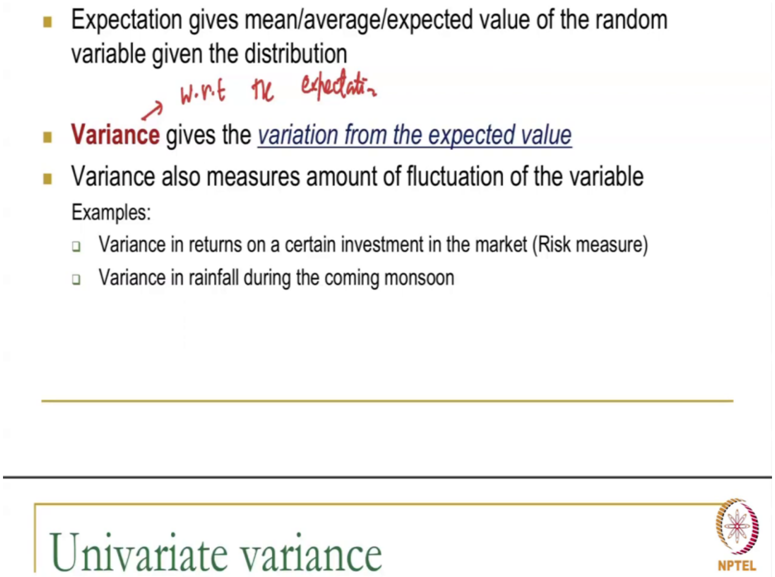
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology – Madras

Lecture – 34
Variance and Covariance

(Refer Slide Time: 00:21)



Welcome back, we are in week six of inverse methods in heat transfer. In the last video we saw the idea of expectation which is just like a mean or an average which we can extract out of a probability density function. Now we are going to look at a second quantity which is similar to the expectation. The expectation sort of gives you the average mean of a quantity whereas variance tells you a fluctuation over the mean, obviously no quantity is going to look fully like the mean.

So, if it looks fully like the mean all it will look like is this. Usually, our random variables are defined by their fluctuation. if we are not going to get fully mean. So, at the very least in order to understand a variable, random variable, you need these first two things you need an expectation and you need a variance. As it turns out you might require a whole lot more but at least we need these couple of things.

So, Random variables as we saw earlier, they return result in different outcomes. Their variation is captured completely I will emphasize this again only by their distribution. So, this is the most important thing which has captured completely as I wrote even in the last video captured completely. So, a full sort of simple example not an example a simple reduction of this random variable is given by the expectation.

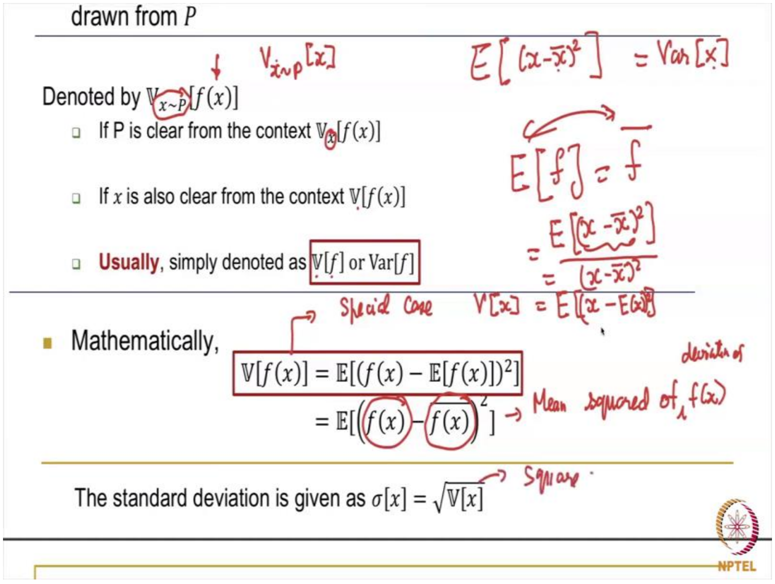
So, instead of asking how well did each boy do in the exam. you can ask a boy scored 60 percent in 10 standard board examinations, whereas girls scored on an average 70 percent. So, that is the expectation but you also should look at the variance. So, for example if uh there is a large amount of variance that say something if there is a small amount of variance if every boy scored only around 60, 61, 50, 59, 61 that says something.

But whereas a lot of boy’s code 0 and a lot of y score 100 and in the middle, it was completely different that also tells you something else. So, expectation by itself is insufficient for us to find out what a distribution looks like, which is why we look at the first kind of quantity which is the variance. So, the variance gives us variation from the expected value. So, notice this variance is always with respect to the expectation.

So, variation from the mean. so, for example variance as I told you, you can talk about mutual fund what is the average expected return? let us say somebody says 10 percent but what is the risk? the risk is how much will this fluctuate. Just like we have electricity fluctuations it is not just important to know what the actual or the average voltages but you want to know what the fluctuations.

Similarly, temperature average temperature in my cell phone is let us say 20 degrees Centigrade that is less important than the variance, if the variance is like 50 degrees up uh then we are in trouble because it can hit 70 at any time and you basically can it can blow up. Similarly, variance in rainfall tells you how heavy or a rainfall can get even if average is slightly low. So, variance is an important quantity.

(Refer Slide Time: 04:04)



So, the first thing we will look at is what is known as Univariate variance which is a simple single variable x. So, what it measures is how much the value of $f(x)$ varies for various samples when x is drawn from P very similar to what we did with expectation. we are I have written the general formula for $f(x)$ you can simply think of $var[x]$ also. So, that is easy but the same thing as last time I might write x drawn from P similar simply write x when P is obvious drop that x also.

When x is obvious power, you can simply write V $var[f]$. Now mathematically $var[f(x)]$ is given by this kind of complicated looking thing, which is,

$$V[f(x)] = \mathbb{E}[f(x) - (\mathbb{E}[f(x)])^2]$$

which is fine I am going to use the notation like before the expectation of f I am going to call,

$$= \mathbb{E}[f(x) - (\overline{f(x)})^2]$$

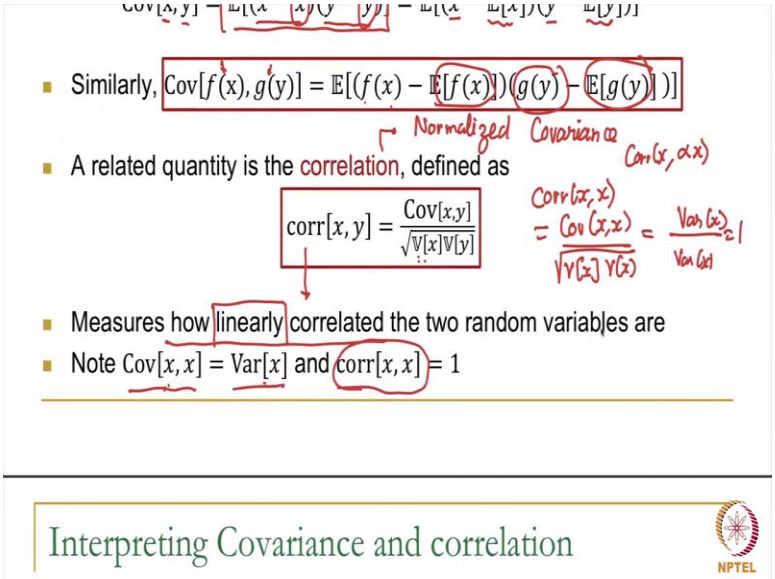
So, then it is easier to see you find out what the function is.

Find out what its deviation from its averages square that and then average it once more and that gives us the variance. So, variance is basically like mean squared of f of x. Now a special case is what is $V[f]$? $V[f]$ is simply $E[(X - \bar{X})^2]$ another way of writing it which we did earlier is remember when we did coefficient of determination $E[(X - \bar{X})^2]$.

So, Square again mean squared deviation mean squared of deviation of f of x from its mean. So, mean square error where error is defined as how much x varies from its average is what this is. another way of writing it is $\overline{(X - \bar{X})^2}$, but this is a little bit more confusing. So, we can use a mixture of these two notations. So, expectation once again of x minus x bar squared is basically variance of x.

How much does x vary over its mean that is what it means the standard deviation of is of course root mean square which is given as square root of the variance as you can see here that is the standard deviation it is given by Sigma.

(Refer Slide Time: 07:25)



Now let us come to covariance. This covariance is usually for a pair of variables x and y. So, two variable cases. we first look at what variance is? the variance is expectation of as I just wrote $\overline{(X-\bar{X})^2}$, so you can write it as a product of two things $(X-\bar{X})(X-\bar{X})$. Similarly, we define the covariance of two variables x and y. Now what does this mean we will come to this. This is a very interesting meaning we will come to this meaning shortly.

But what it is, is this $E(X-\bar{X})$ and E $(Y-\bar{Y})$ what this says is how much does the fluctuation in x move along with the fluctuation in y. do these two align with each other or are these two completely misaligned. So, another way of writing it is of course the more formal way of writing is $E[(x-E(x))(y-E(y))]$. But like I said it is preferable to use this when you write it because you might get confused when you write it on paper.

Similarly, covariance of here I wrote $cov(x,y)$ more generally if you look at $cov[f(x), g(y)]$ which is,

$$cov[f(x), g(y)] = E[(f(x) - E(f(x))])(g(y) - E[g(y)])]$$

Now a very important related quantity is what is known as correlation, which you can think of as normalized covariance. So, covariance by itself might be high or low we cannot make much out of it but if we divide by this quantity $\sqrt{V(x)V(y)}$, why is that?

So, if you look at this look at the size of this, this is expectation of Δx multiplied by Δy let us call it that. So, instead of expectation of Δx, Δy we want to find out a normalized way. I divide this by expectation of Δx, Δx you know expectation of Δy, Δy but of course there are two of these. So, I put a square root just to get the same size here. So, this becomes expectation of a $\sqrt{V(x)V(y)}$.

Now what this correlation measures are? this is really important. So, please notice this it measures how linearly correlated the word linearly is very important, how linearly correlated these two random variables are. So, if you notice what $cov(X, X)$ is. So,

$$cov(X, X) = E[(x - \bar{x})(x - \bar{x})]$$

which is simply Var[x]. So, that is what is written here,

$$Cov[x, x] = var[x]$$

But what is $Corr[X, X]$?

$$Corr[X, X] = \frac{Cov[X, X]}{\sqrt{V[X]V[X]}}$$

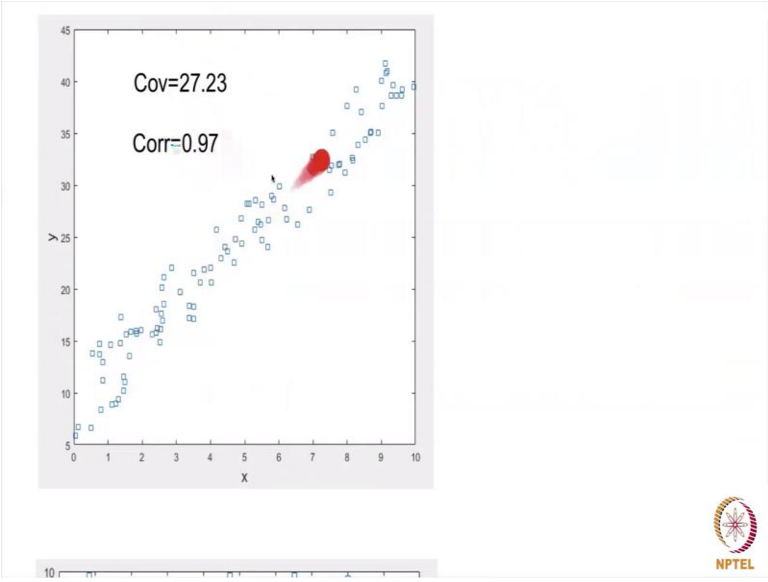
which is,

$$Corr[X, X] = \frac{Var[X]}{Var[X]} = 1$$

$Cov[X, X]$ is $Var[X]$ divided by of course this is also $Var[X]$ which is 1. An interesting exercise that you can do is what is correlation of x, some constant multiplied by x please find this out and see what it is that is an interesting exercise for you to do.

So, coming back to this notice two things $Cov[X, X]$, x is $Var[X]$ and $Corr[X, X]$ with itself is one. So, we are now going to use these relations to interpret what covariance and correlation actually mean.

(Refer Slide Time: 11:48)

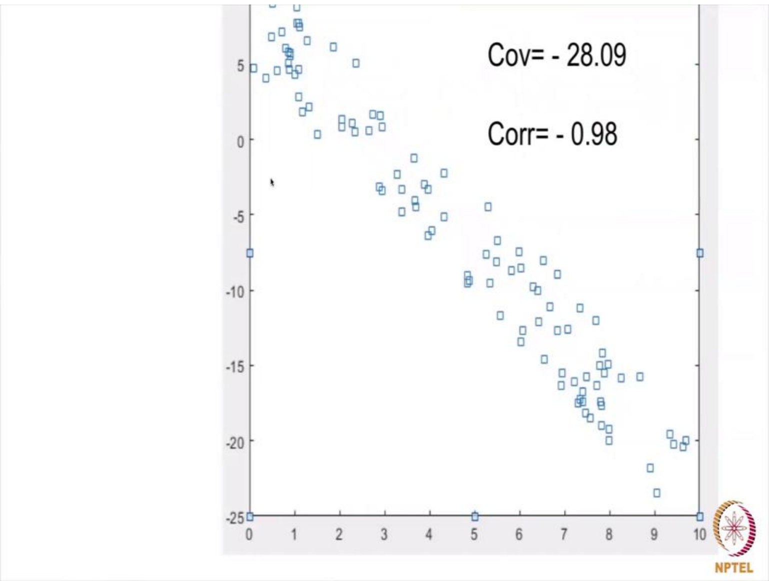


So, here I have drawn a few plots. let us concentrate on each one of these plots one by one these are some random variables x and y. So, x is some random variable, y is some other random variable and we collected data notice a lot of this joint data, let us say there are about 100 points here, can randomly generated points for x and y based on some relation. Now overall you can see just like our slab temperature there is an increasing trend that its x increases overall you can say y increases.

But that does not always happen it is possible for x to increase and y to decrease. So, for example this x is lower this x is higher but this y is higher and this y is low sorry y does decrease occasionally but overall, it increases. So, how do we capture this idea that there seems to be some relationship between x and y they are not just totally randomly related. So, that is what covariance and correlation to covariance is positive it is a positive value.

But is this value high, is this low we do not know without normalizing it. So, correlation is normalized we know that if x was plotted with x, this would give us exactly one. this is 0.97 which is pretty close to one which is these are very nicely linearly correlated you can see that here. And this positive value in fact indicates that whenever x increases in general on an average y will increase it is a powerful quantity.

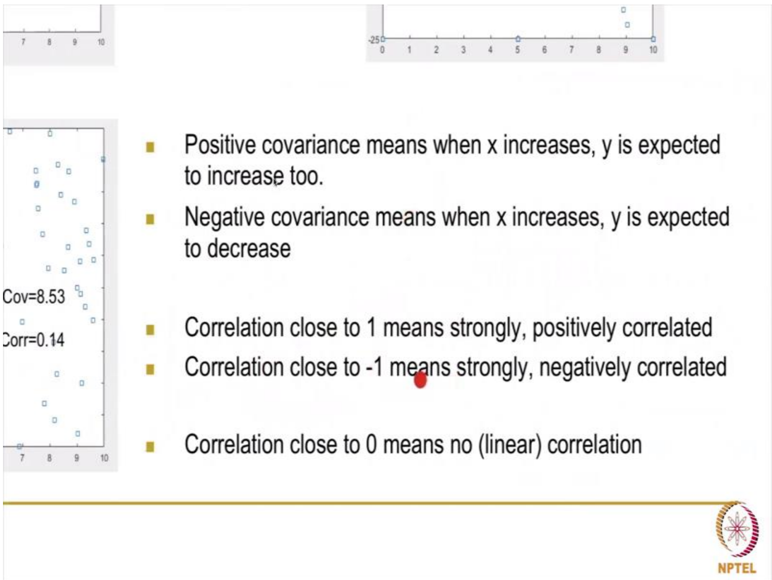
(Refer Slide Time: 13:21)



Look at this here exactly the opposite relationship when x increases y decreases though not always again and like I said sometimes it can happen that x increases but y increases also. But overall, the trend is a decreasing trend in such a case, we look at the covariance, covariance is negative again is this big or small we do not know but we can look at the correlation, normalized correlation is minus 0.98 which means that it is very close to plotting something like x versus minus x.

So, this is perfectly almost perfectly negatively correlated which means it will decrease but the two variables are correlated.

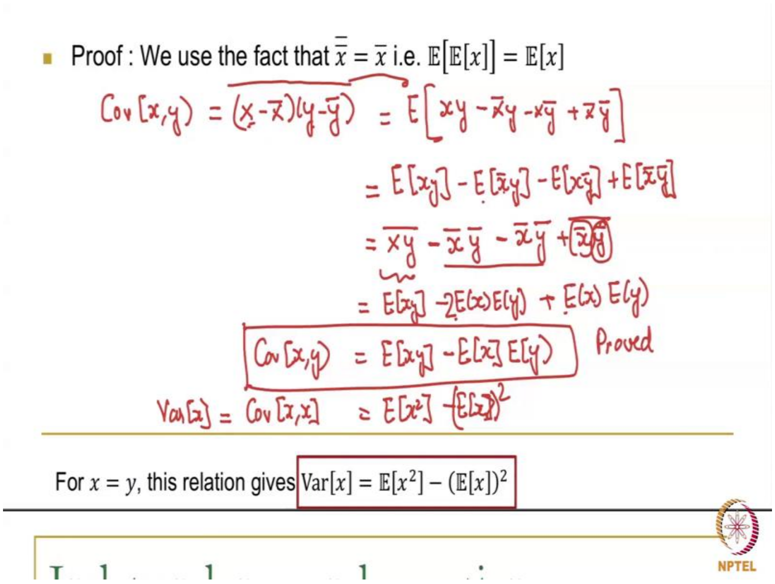
(Refer Slide Time: 14:05)



Now compare this covariance is 8.53. we cannot interpret covariance as except that it is positive, but it is a meaningless positive. correlation is 0.14 what correlation is 0.14 says is that there is practically zero linear correlation between x and y you can see these are just randomly distributed you cannot see any kind of trend here. So, what we can say is positive covariance means x increases y is expected to increase, negative means when x increases y is expected to decrease correlation close to 1 means strongly positively correlated.

Correlation close to minus one like the minus 0.98 case means strongly negatively correlated and correlation 0 to close to zero like here means no linear correlation what does linear correlation mean?

(Refer Slide Time: 15:00)



Let me show that in the next couple of slides let us just do a quick simplification here which we will use in future also. So, that is this nice relationship which says that,

$$Cov[x, y] = \overline{xy} - \bar{x}\bar{y}$$

So, we had seen this earlier also when we were looking at the linear regression that $E[xy]$ is not the same of $E[x]E[y]$ in fact their difference is exactly the covariance.

So, that is what we are going to look at right now. We are going to use a simple fact here, that the average of the average is the same as the average because the average is no longer a random variable. So, $E[E[x]]$ is the same as E[x]. So, we will use that. So, let us start here. So, the $Cov[x, y]$ is,

$$Cov[x, y] = \overline{(X - \bar{X})(Y - \bar{Y})}$$

So, I am going to use this expression. So, let us open this up I call this expectation for now. So, this is,

$$= E(xy - \bar{x}y - x\bar{y} + \bar{x}\bar{y})$$

So, we have our linearity of expectations so we can add this up this is,

$$= E(xy) - E(\bar{x}y) - E(x\bar{y}) + E(\bar{x}\bar{y})$$

So, this is x y average, minus x bar is a constant.

So, $\bar{x}\bar{y}$ minus again $\bar{x}\bar{y}$ plus $\overline{\bar{x}\bar{y}}$. So, this is of course this term I will just leave it as expectation of xy. This term is these two terms are the same and they are minus expectation of x expectation of y and of course there are two of these plus x bars is a constant y bar is a constant it can be taken out of the expectation. So, really this is just average of one.

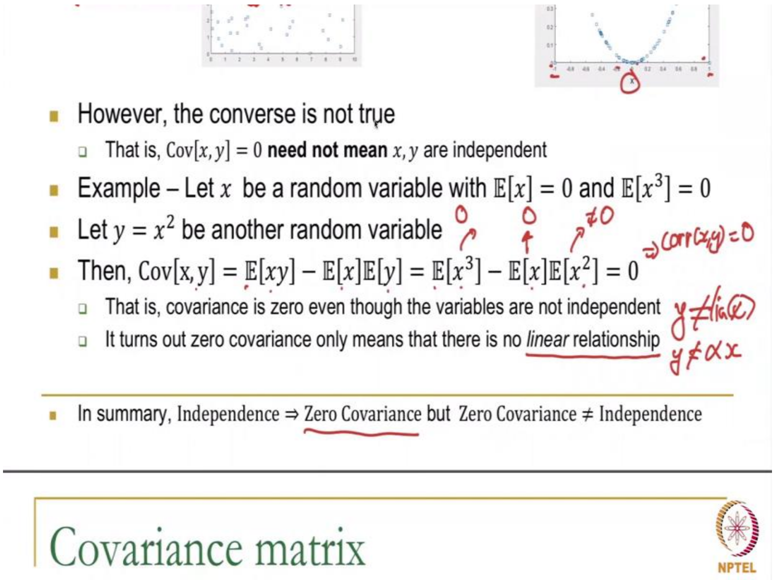
So, this is once again expectation of x multiply the expectation of y. So, this is expectation of x y minus expectation of x. So, these two add up multiplied by expectation of y is covariance of x, y. So, that is the proof there is a special case here when x equal to y this tells us that Cov[X,X] is,

$$Cov[X,X] = E(X^2) - (E(X))^2$$

but $Cov[X,X]$ as we saw earlier is simply $Var[X]$.

So, we have this relationship that $Var[X]$ is the difference between the $\overline{X^2} - \bar{X}^2$. So, this is a very useful relationship. So, we often use this to find out the standard deviation and many other things you would have again you would have seen this in school also. x square bar minus x bar square. Another way of writing it square bar minus x bar Square you might remember this also sitting in our linear regression formulas which are there they are there for a reason is because it is a physically very important quantity.

(Refer Slide Time: 19:29)



Now I talked about how covariance means linearly sorry correlation equal to zero, simply means that these are not linearly correlated or they are not linearly dependent. So, that is a very important statement that was made that statement being linearly. So, let us look at the implications of that statement. So, let me show this with an example. So, there are two statements the statements are this if x and y are independent.

So, take two independent variables and write or find out their covariance. Their covariance will always be zero or close to if they are truly independent then fully zero if not, they will be close to zero. So, here is an example in this case covariance has to be approximately zero. Especially correlation has to be close to zero. but the opposite is not true. that is if covariance is 0 it does not mean that the variables are independent it could mean they are independent but you cannot always infer that they will be independent.

So, that is what I have written here covariance xy is 0 does not mean that x and y are independent. So, let us look at one such case here x and y are random variables. So, x is a random variable you can see it is randomly distributed, y I defined as x square. So, clearly x and y are not independent, you can see, y it depends on x there is a nice clear trend then should covariance be zero or non-zero.

Now if we think that correlation simply means or more than quarter covariance let us talk about correlation should correlation be zero or not looks like but they are dependent on each other correlation should not be zero and in fact I should get a nice value. But unfortunately, it turns out that correlation actually will be zero, actually will be close to 0 here, if you plot use the formula that we had even though x and y are not independent.

So, covariance equal to 0 does not mean independent why is that. So, let us look at this

$$Cov[X, Y] = E[XY] - E[X]E[Y]$$

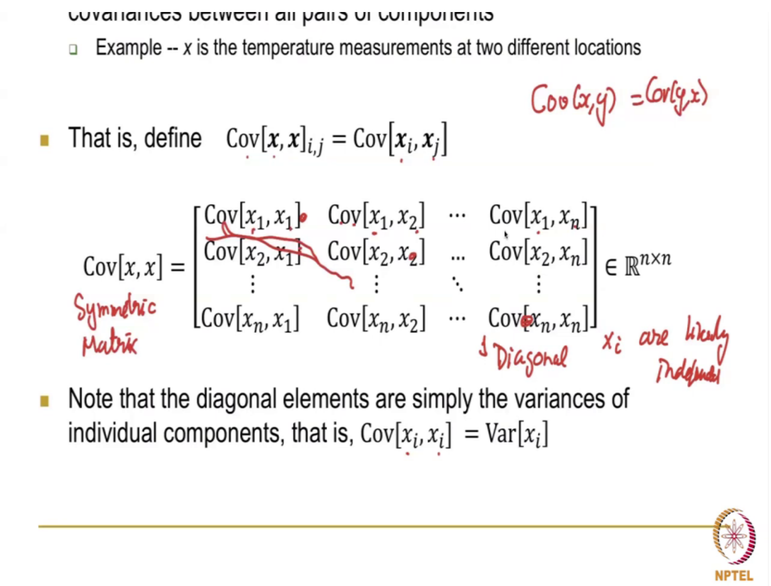
This is expectation of x Cube minus expectation of x multiplied by expectation of x square. Now look at this variable x this variable x goes from -1 to 1 it is an odd variable. So, equally likely it is expected to take positive values and negative values.

So, expectation of x cubed will be zero, expectation of x square will not be zero, but expectation of x is 0, which means covariance of xy and correlation of xy if you calculate it will turn out to be zero. So, zero covariance only means that there is no linear relationship, that is why y is not a linear function of x. So, y will not look like some alpha x, but it could look like as in this case it could look like x square.

So, this is a very important counter example. So, if two variables give you correlation of 0, all you know is that they are not linearly dependent on each other but you cannot say anything more than that. Nonetheless correlation is an important quantity in order to figure out linear dependence of one variable on the other. So, we do use that. So, in summary zero covariance, Independence means zero covariance, but zero covariance does not mean Independence.

We now come to the next important idea; this is a very central idea when we go to the multivariable case a multivariate case which is called The Covariance Matrix. we have just been looking at univariate covariance. Now we will look at multivariate covariance.

(Refer Slide Time: 23:48)



So, in many cases we actually have multiple variables just like in the case of a slab, I can think of x as all the temperatures put together as I said earlier. So, x could be the random Vector T_1 through T_6 and now we can start looking at pairwise covariance's that is; is the temperature here independent of or linearly independent of T_2 is T_1 and independent of T_3 etcetera. So, we can now think of what is called a covariance matrix.

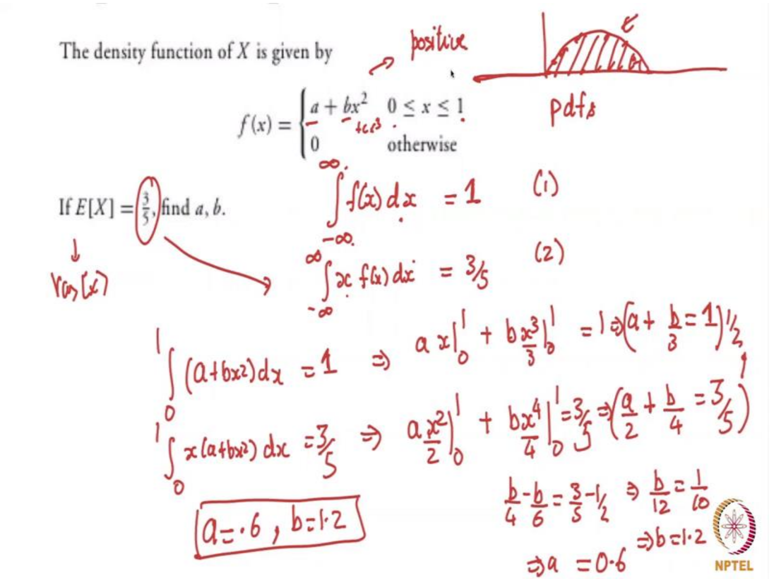
So, you can make a full Matrix each of them is the pairwise covariance of any two of these variables. So, basically simple idea you write $Cov[x,x]_{i,j}$ which is $Cov[x_i,x_j]$ So, first variable with first variable, first variable with second variable, first variable with nth variable of course it is a symmetric Matrix because covariance of x, y is the same as covariance of y, x.

So, it is a symmetric matrix. The second thing to note is all the diagonal elements. So, if we look at just the diagonal elements here, these are simply the variances right because covariance of as we saw xi, xi is simply variance of xi. So, what this measure is if this thing is diagonal, then we can say that at least all the xi are linearly independent. So, the closer this Matrix looks to diagonal and we will often make this assumption while making inverse problems.

Is that we are assuming that linearly the measurements are independent. I am going to show a more specific case when we come to the next week. But notice that the diagonal elements are the variances and the off-diagonal elements are the pairwise covariances which we had defined earlier. Now we have looked at a series of ideas expectation variance covariance etcetera. So, just to sort of round this off I will show you a couple of very simple examples or something related is also sitting in your exercise for this week.

And after that we will move to the next topic which is Gaussian distributions, which is in the next video but for. Now let us just take a couple of simple examples.

(Refer Slide Time: 26:22)



So, let us take this example there is a density function of x which is the probability density function of x it is given by this simple formula,

$$f(x) = \begin{cases} a + bx^2 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

So, the probability distribution function would look something of this sort. So, this 0 here 0 here remember PDF always has to be positive. So, you are asked to find out what a and b are given that E(x) = $^3/_5$.

So, this is also given to you and you are asked to find them. So, you have two things to find out and you are given one piece of information. So, let us go ahead and do that and see how to think about this clear. Now you might think that this is an absurd problem like no such problem would occur in reality but as it turns out this is a very important problem in approximation of PDFs.

As I told you we do not usually have actual probability distribution functions, but we will have some of their properties like we will have their expectation, we will have their variance and it is an important exercise to model these probability density functions as some simple PDFs and then go ahead and recalculate what these functions are on the basis of this. So, let us come back here.

So, let us say expectation of x is 3 by 5, what else do I know? I know that the probability density function has to integrate one. why because the net probability always has to be one. So, this is one equation the next equation is from here which tells us,

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} xf(x)dx = \frac{3}{5}$$

Now we will use these two pieces of information to figure out these two variables a and b.

So, let us take the first. So, the first one says,

$$\int_{0}^{1} (a + bx^2)dx = 1$$

which when you integrate this gives us,

$$ax|_0^1 + b\frac{x^3}{3}|_0^1 = 1$$

So, this gives us,

$$a + \frac{b}{3} = 1$$

Similarly, the second equation says,

$$\int_{0}^{1} x(a + bx^2)dx = \frac{3}{5}$$

So, this gives us,

$$a\frac{x^2}{2}|_0^1 + b\frac{x^4}{4}|_0^1 = \frac{3}{5}$$

So, this tells us,

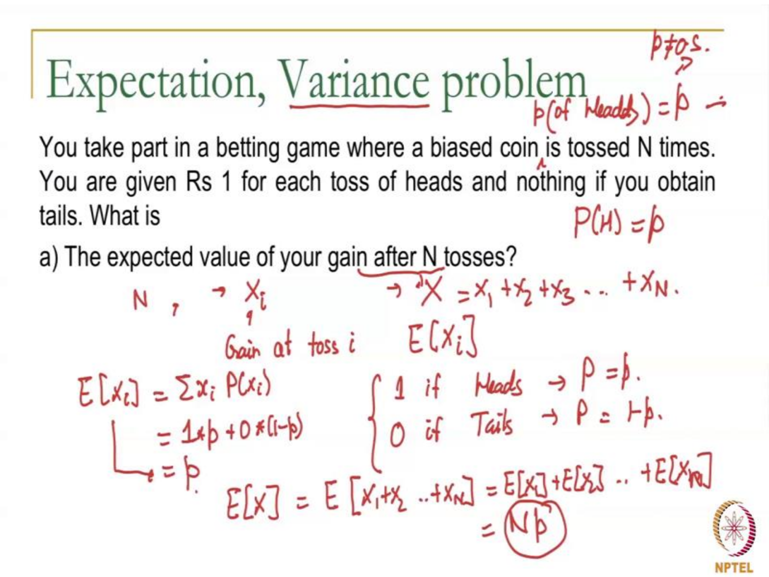
$$\frac{a}{2} + \frac{b}{4} = \frac{3}{5}$$

Now we can easily solve these two equations together in order to find out a and b. So, we can simply multiply this equation by half and subtract it from this equation. So, that gives us b by 4 minus b by 6 equals to 3 by 5 minus 1 over 2. So, this gives us b by 12 equals to 6 minus right 1 by 10. So, b is 1.2 and this gives us is 0.6. So, a equal to 0.6, b equal to 1.2 and this is a simple way of solving this.

Of course, we could have extended this problem to more general problem and given let us say a c. So, I could give a c x cube and then I would need some extra piece of information I might ask you to give the variance of x also. So, as I give more and more information you can find out more and more parameters for this probability density function. So, that is a powerful way of using our these are called moments.

So, for example you can see this is integral of f of x, this is the integral of x f of x, the third one will have terms which involve inter integral of x square f of x, because it is x minus x bar but generally it will involve terms involving x, x square f of x and that is what makes this whole thing very powerful. So, I mean very flexible in finding out probability density functions.

(Refer Slide Time: 31:34)



Let me give one more problem this is sort of related to the previous problems that I gave or examples that I did. So, let us say you are taking part you know betting game and you have a biased coin which is tossed n times. So, let us say the biased coin has probability of heads equal to P. So, P could be something that is a point six, point seven, it is not necessarily equal to 0.5.

So, once again it is sort of a flip of what I told earlier or maybe the same that you're given rupees one for each toss of heads and nothing if you obtain tails. So, what is your expected value of gain after N tosses. So, that is the question. So, remember we now have to define carefully what our random variable is, whatever expected value of a random variable is etcetera. So, let us do that and then solve this problem.

So, let us define our random variables carefully. So, let us say we are making N tosses and each task corresponds to a random variable xi or the gain that you gain. So, gain a toss i is xi. So, you get 0 if you toss ahead and you get ah sorry you get one if you toss ahead and zero if you toss a day remember that probability of a head is P as we defined here and x your gain after N tosses is x1 what you will get after the first toss what will get up until xN, a gain that you get only at toss i is this and x is this.

So, now what is expectation of xi, that is at the ith toss, remember you get one if you get heads. So, the probability of this is p when you get 0 if tails the probability of this is 1 minus P. So,

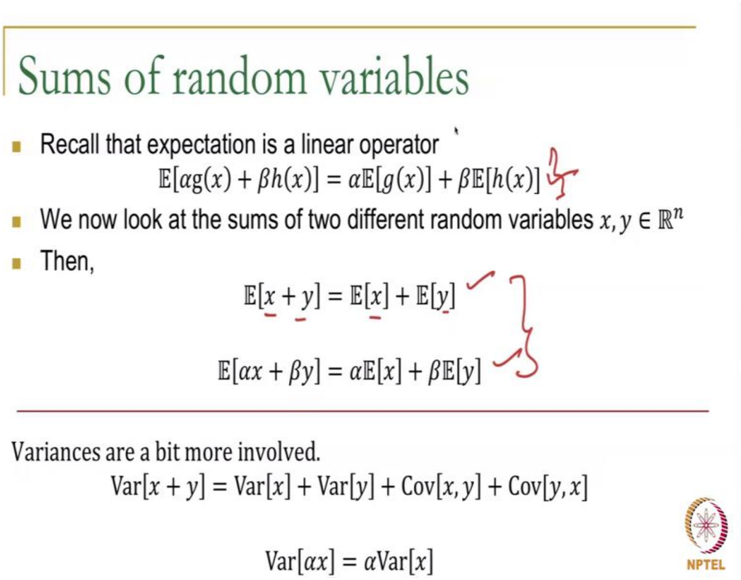
$$E(X_i) = \sum x_i P(x_i)$$

So, this is basically one with probability p and 0 with probability 1 minus p. So, the expected gain at the x ith toss is p. So, therefore the expectation after x at x is sigma actually, I should do it in two steps $E[x_1 + x_2 + \cdots + x_N]$ which if you remember is the same as expectation of x1 because expectation of x2 the expectation of xN.

So, which basically is NP. So, our expected value of our gain after N tosses is simply N times whatever is the probability of getting heads. So, this is very obvious if you say that the probability of getting heads is 0.6 and I did a hundred tosses what do you expect to gain after 100 tosses you will say of course 60 rupees but we have a more important or interesting question, what is the variance in your game.

So, one person tossed 100 and got 60 rupees, another person is everybody going to get 60 rupees, no obviously different people are going to get different number of heads and depending on their luck they might get slightly lower or slightly higher, what is the variance? can we actually say something about the variance in this game. So, we are going to ask that question and uh we will answer it in a slightly roundabout fashion.

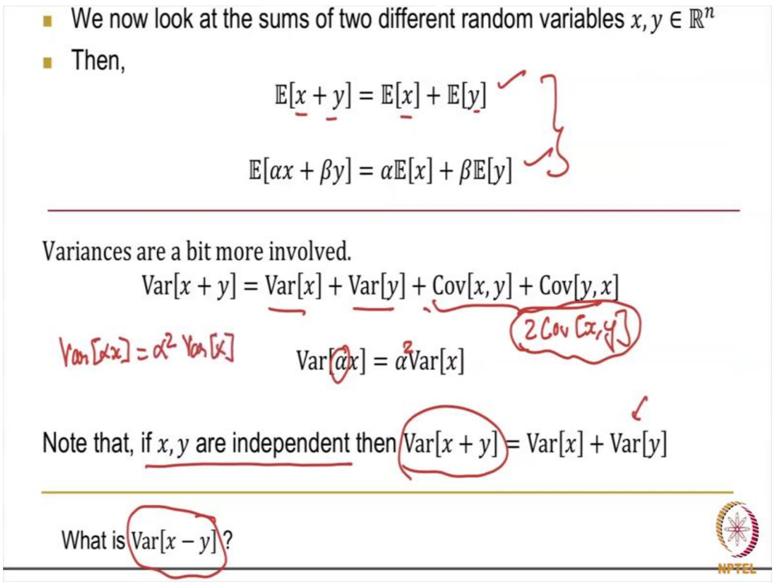
(Refer Slide Time: 36:08)



To answer that question, we have to go back to when we did sums of random variables. So, I cheated a little bit, I said that expectation is a linear operator which is true, but it was a linear operator in one variable and I kind of applied it so, far to two different random variables, where it also holds true. it turns out that this is also true. There are subtle differences between these two cases but I am not going to get into that this also happens to be true.

So, we are going to use these forms rather than this form rather than taking two different functions of the same variable, we are going to take two different random variables all together. So, x is a different random variable and y is a different random variable and we are going to take sums of that. it turns out expectation is linear there also however variance is a little bit more involved.

(Refer Slide Time: 37:09)



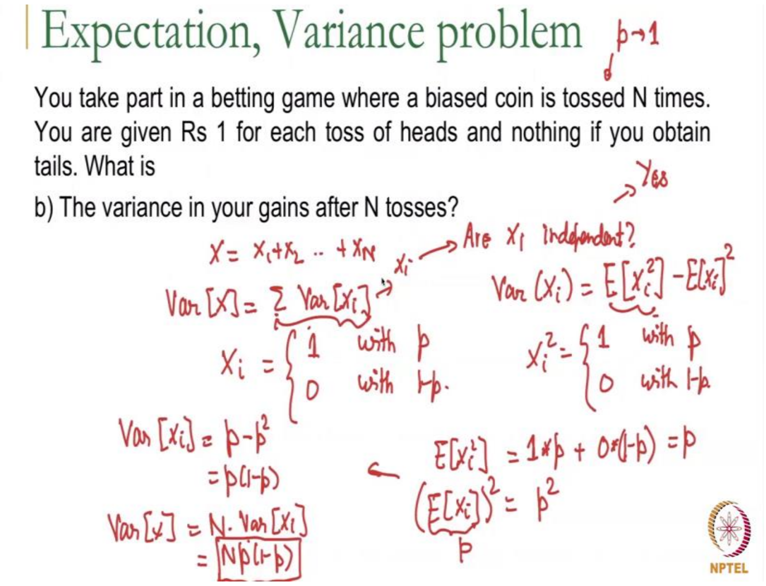
So, variance is it turns out variance of x plus, variance of y plus, 2 times covariance of x, y. you have to be careful if x and y are vectors which is why I wrote it separately but in case they are just scalars you can treat this as two times covariance of x, y. So, notice that if x and y are dependent linear variables, then this does not cancel out automatically. Only If x and y are independent then you see that variance of x plus y is variance of x plus a variance of y similarly variance of alpha times x is Alpha times variance of Y.

So, you have to be a little bit careful about how you do these things when variances are involved. So, we will see that shortly in a in the continuation of this video please also think about what happens to variance of x minus y. So, this is an interesting question that you can answer for yourself as an exercise. So, remember these Expressions variance of x plus y is equal to variance of x plus variance of y plus 2 times covariance of x, y in general variance of x plus y is equal to variance of only if x and y are independent.

Independent in every way not just linearly independent; independent in every single way and variance of alpha x, there is a small error here which will be Alpha Square. So, I will just write that again variance of alpha x is Alpha Square variance of x. So, all these formulas are things that you can prove, I am not going to prove these. So, proof has been skipped. You can prove this by yourself it is not hard but that is beyond the point of the current code.

So, please try and prove it for yourself very similar to the proofs that I have done so far. So, we will use this formula here.

(Refer Slide Time: 39:26)



So, now once again we are taking part in that betting game, where biased coin is tossing N times what is the variance in your gains after N tosses. So, now let us write that x once again is x1 plus x2 obtain xN. Now the question is are the xi independent and the answer is yes, if you toss the coin two times separately the outcome of one does not depend on the outcome of the other. therefore, you can basically write variance of x as summation of variance of xi.

Now there are various ways of calculating variance of xi. So, let us do that. So, xi once again is the same thing it is 1 with probability p and 0 with probability 1 minus p. So, we want to since xi are independently, because of that you can write this as a summation of variances. So, variance of xi we can use our formula expectation of xi square minus expectation of xi the whole Square.

So, what does expectation of xi square mean. So, xi square is also 1 with probability p and 0 with probability 1 minus p. So, $E({x_i}^2)$ is $1 * P + 0 * (1 - P)$. So, which is a what about $(E[x_i])^{2}$ we had already calculated this in the last part this was P. So, this is P^2. So, this tells us that $Var[x_i]$ is,

$$\begin{aligned} Var[x_i] &= P - P^2 \\ &= P(1-P) \end{aligned}$$

So,

$$\begin{aligned} Var[x] &= N.Var[x_i] \\ &= NP(1-P) \end{aligned}$$

So, you can see that there is a amount of variance in how much gains you will get. So, let us say your biased coin has probability p which is very close to 1. So, very close to one your variance, you will actually notice what happens to it actually I will suggest this as an exercise I was going to say this.

So, plot what the variance looks like with varying bees and see if you can find out what the meaning is you know what values it takes. why is the variance high in a few cases and why is the variance low in a few cases that gives you sort of some intuition for how this thing works. So, what we did in this video was continued from the previous video we looked at the variance, we looked at covariance and we looked at the covariance Matrix.

And we looked at the sum rule for variants which said that variance of x plus y is variance of x plus variance of y, in case x and y are independent and we did a couple of problems. So, in the next video I will move to uh one of the fundamental distributions. So, we talked about probability distributions in general here, but we look at one of the fundamental distributions, probably the most fundamental distribution in all of Science and Engineering which is the Gaussian distribution also called the normal distribution. So, we will see that in the next video, thank you.