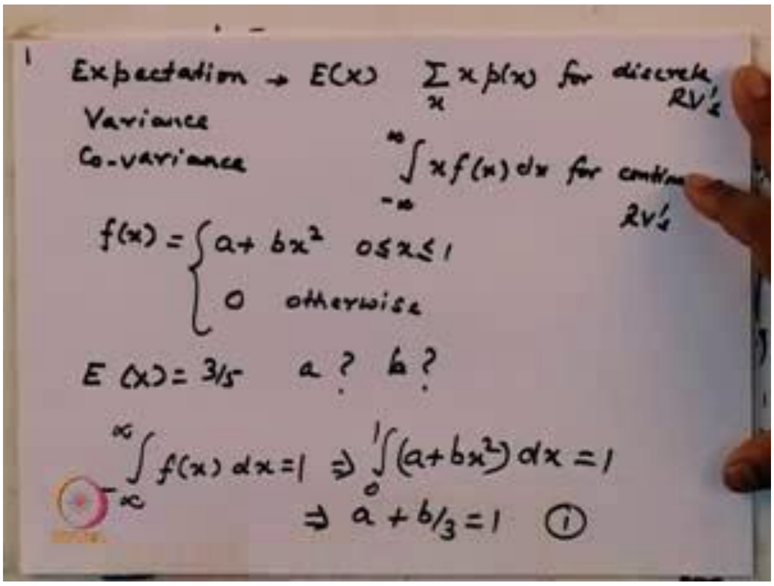
Introduction to Biostatistics
Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture - 21
Expectation, Variance and Covariance Part-II

Hello and welcome to today's class. We will begin by having a brief recap of what was covered in the previous class. So, one of the most important concepts, we had introduced in last class was the concept of expectation and variance and covariance.

(Refer Slide Time: 00:31)

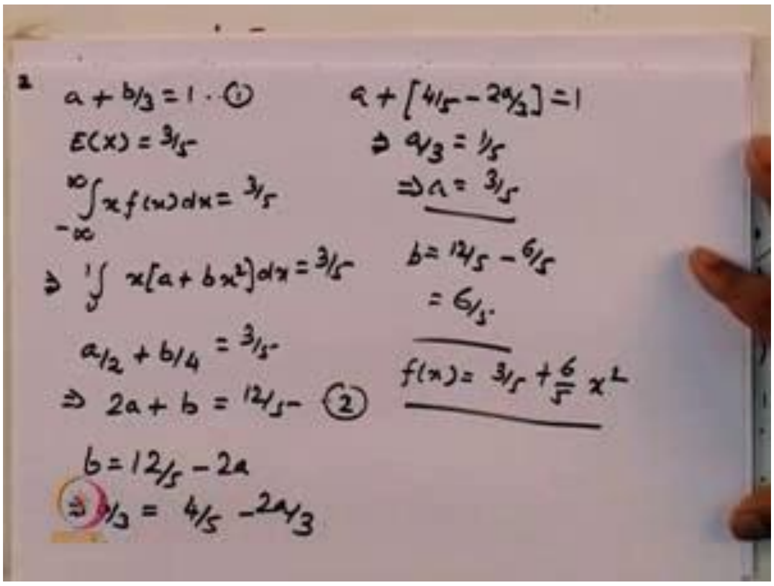


So, expectation is written as E of X for a random variable X and is defined as either summation x p of x over all x for discrete RVs and for continuous RVs. So, let us take the following example, let us say f x; the probability density function f x is defined as follows. So, define the probability density function f x of random variable as a plus b x square for 0 less equal to x less equal to 1 and 0 otherwise. Now if for this random variable the given E of X is equal to 3 by 5 then what is the value of a and what is the value of b?

So, we essentially have two unknowns in this problem; a and b and we are given this condition E of X equal to 3 by 5. So, how do we go about it? So, of course, since f x is the probability density function we know that minus infinity to infinity f x d x is equal to

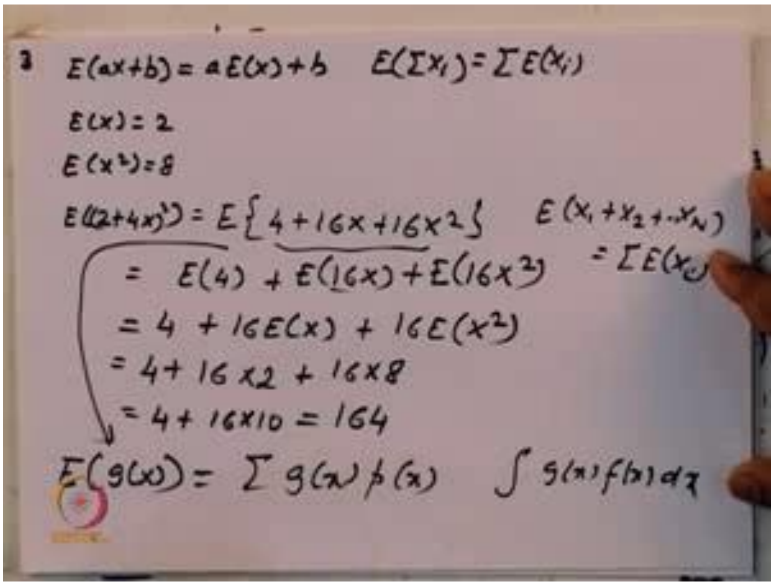
1. This gives us this one equation. So, if we go through it, this implies implying. So, this is equation 1.

(Refer Slide Time: 02:38)



So, I can write a plus b by 3 equal to 1. Now the other equation we have is E of X is equal to 3 by 5. So, E of X is defined as x f x d x, this is 3 by 5. So, this will boil down to x. So, this would be a by 2 plus b by 4 x square by 2 is equal to 3 by 5 applying 2 a plus b is equal to 12 by 5. If 2 a plus b is 12 by 5 then this equation 2. So, from this equation I can write b is equal to 12 by 5 minus 2 a implying b by 3 equal to 4 by 5 minus 2 a by 3 and if we plug this value in equation 1 then, we have a plus 4 by 5 minus 2 a by 3 equal to 1 implying a by 3 equal to 1 by 5 is 3 by 5. So, then b becomes 6 by 5. So, thus f x becomes 3 by 5 plus 6 by 5 into x square.

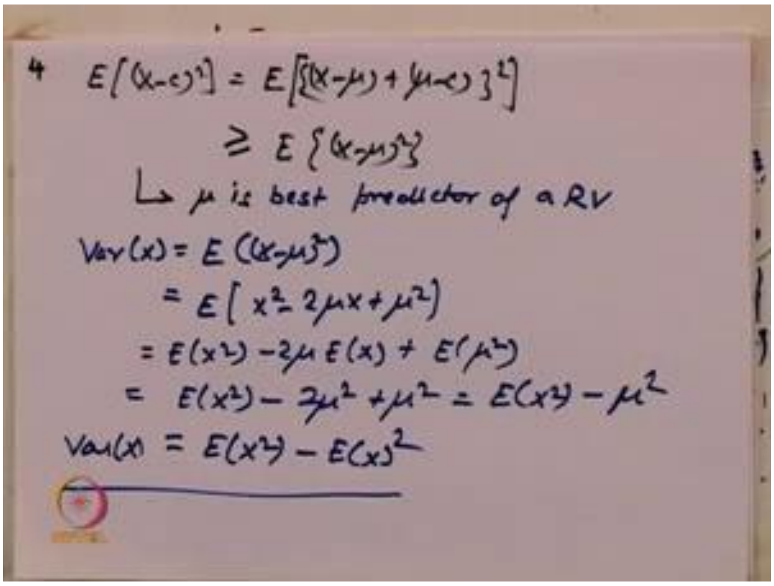
(Refer Slide Time: 04:42)



So, we also have the following you know variations of expectation. So, we know that E of a x plus b can be written as a expectation of x plus b. Now let us say, we have a random variable such that E of x is 2, E of x square is 8. So, we want to find out E of 2 plus 4 x whole square. What would be the expectation of this particular variable? So, what is it, I can expand this equation; expectation of 4, 16 x plus 16 x square then, I can write expectation of 4 plus expectation of 16 x plus expectation of 16 x square. So, here in this equation I have invoked E of X 1 plus X 2 plus dot, dot X N is equal to summation E of X i.

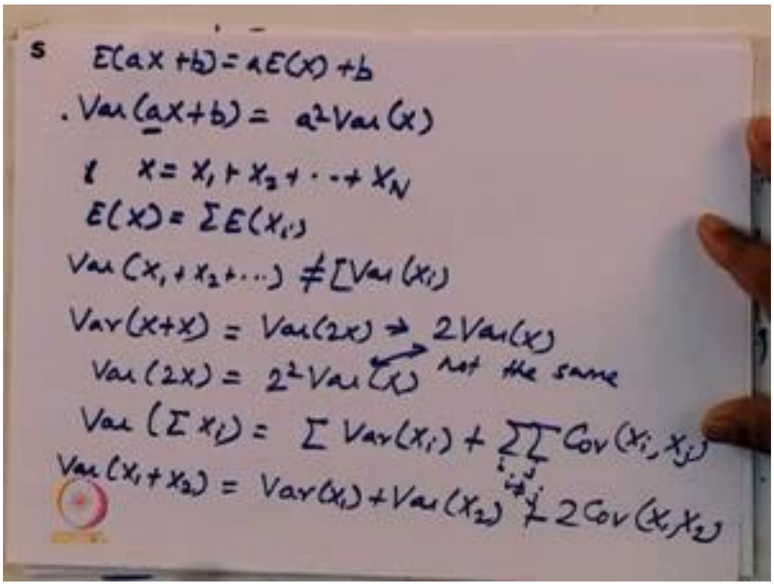
So, now E of 4 is a constant it will be 4. I can take 16 out. So, I can made you of this expression. So, it is 16 expectation of X, I can take expectation 16 out here also and I get expectation of X square. So, 4 plus 16 into expectation of X is 2 plus 16 into 8. So, is 4 plus 16 into 10 is equal to 164. So, I can make use of this and the identity E of summation of X i is equal to summation E of X i or I can break it down. I also have the following thing, if you have a function E of g x. So, actually this we you know broke it down, but this is more like E of g x is summation of g x p x d x or integral g x f x d x.

(Refer Slide Time: 07:05)



So, we had earlier determined. So, if I do E of X minus C whole square, I can expand this as E of X minus mu plus mu minus C whole square and this we had shown that this is greater equal to E of X minus mu whole square. So, this is why, mu is the best predictor of a RV. So, this will give you the best value of expectation, it minimizes this. So, E of X minus mu whole square is nothing but the variance of X. So, I can expand this. So, variance of X has this identity.

(Refer Slide Time: 08:43)

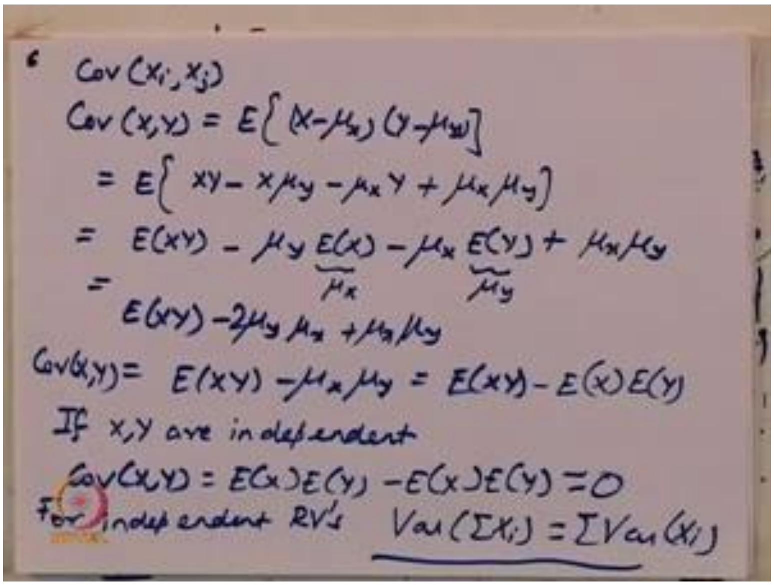


Now, so I have this, I already showed that E a X plus b is equal to a E of X plus b. So, what do we get for variance of a X plus b? So, what we found was I do not need to derive this equation, but what we found was variance of a X plus b is nothing, but a square variance of X, variance of a constant is 0 and when you have a pre factor scalar multiple it just takes the square form. So, when you have multiple independent random variables X as defined as X 1, plus X 2 plus dot, dot X N.

So, you can have, you can write E of X is equal to summation of E of X i; however, you cannot write variance of X 1 plus X 2 plus dot, dot, dot you cannot write for two variants of summation variance of X i you cannot write. Let us see why? So, imagine I have variance of X plus X right. So, if I use this particular formula I should get. So, equal to this is variance of 2 X should return me a value of two variance of X; however, variance of 2 X if I use this formula. So, variance of 2 X, 2 square variance of X. So, this and this are not the same. So, you cannot write this equation in the general case, variance of X summation is not the summation of the variances.

However so, in the general case I am not going to derive this equation, you can write variation of summation of X i is given by summation variance of X i plus, this is our i this is our j, f and this expression has to be computed for i not equal to j. So, this is the general case. So, if I have two variables I can write X 1 plus X 2 becomes variance of X 1 plus variance of X 2 plus twice covariance of X 1 comma X 2. You have two of covariance of X 1 comma X 2.

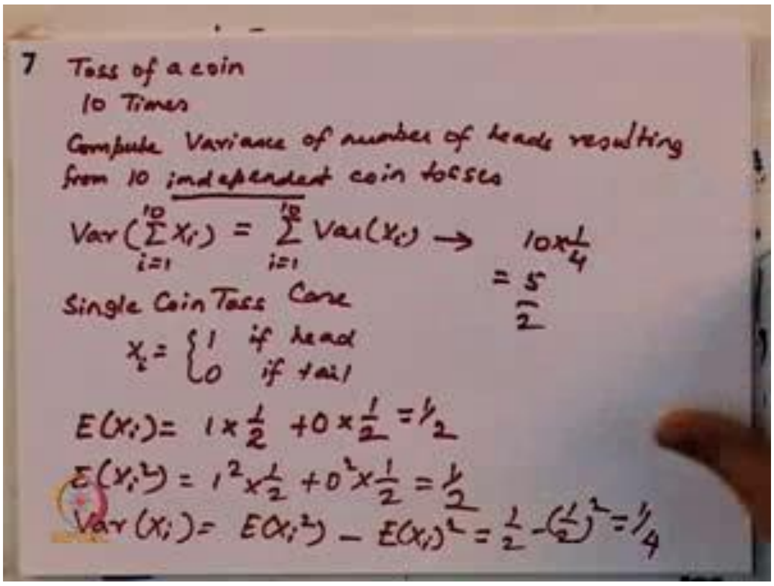
(Refer Slide Time: 11:27)



How do we define covariance? So, covariance of X i comma X j or let us say covariance of X Y is defined as expectation. So, I can expand this.

I can write this as, E of X Y minus mu y E of X minus mu x E of Y plus. So, E of x is mu x and E of y is mu y. So, this nothing becomes nothing, but X Y minus 2 mu y mu x plus mu x mu y. So, this finally, you can write it as E X Y minus mu x into mu y equal to E of X Y minus E X into E of Y. So, this is your definition of covariance of X comma Y. Now what you see is if X and Y were independent. So, if X and Y are independent then, covariance of X comma Y is equal to E X into E Y. So, I can expand E X Y as E X into E Y minus E X into E Y, this gives me a value of 0. So, for independent variables only I can write. So, this equation is applicable when your random variables are independent of each other. So, in general this equation is not valid, but when the random variables are independent, you can write this equation.

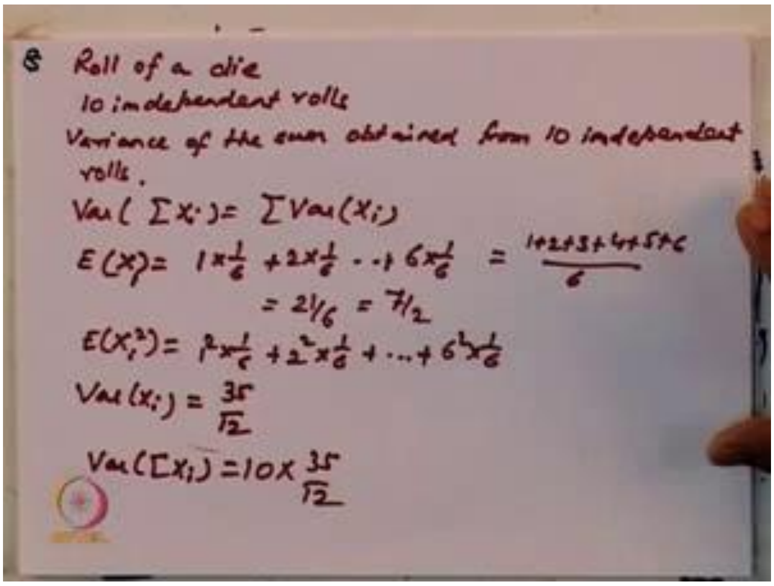
(Refer Slide Time: 13:52)



So, let us solve two sample cases where I can use this equation. So, imagine you are taking a coin toss case, you have toss of a coin and you are tossing it 10 times. So, you want to know compute, you want to compute variance of number of heads. So, 10 times of heads resulting from, 10 independent coin tosses. So, what you see here? Independent coin tosses right. So, I can write variance of summation X I, i equal to 1 to 10 as nothing, but summation i equal to 1 to 10 variance of X i. So, covariance is 0 because the events are independent of each other.

Now, let us take a single coin toss. So, I can write X i if head 0, if tail right. So, in this case my E of X i becomes 1 into probability of 1 which is head, which is half plus 0 into probability of tail which is half. So, this is simply half, I can write E of X i square also which is 1 square into half plus 0 square into half simply half. So, my variance of X i is E of X i square minus E X i whole square equal to half minus half square equal to one-forth. So, thus my variance this becomes 10 into one-fourth is equal to 5 by 2.

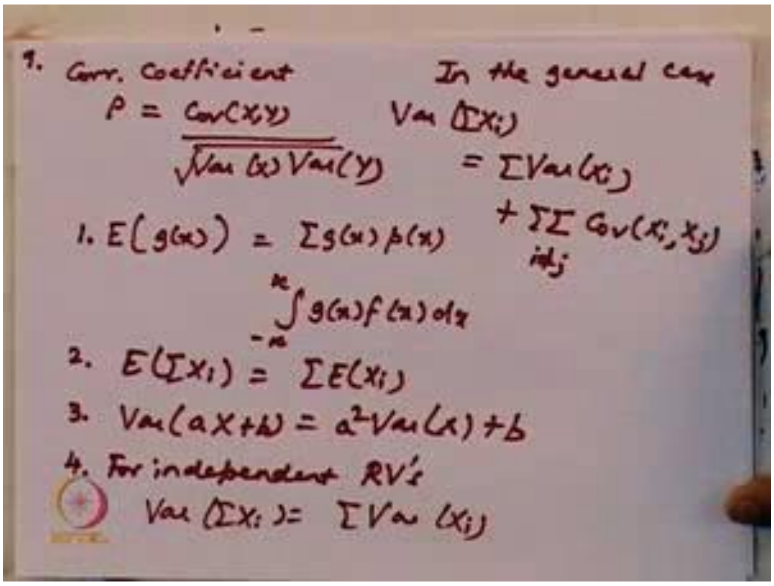
(Refer Slide Time: 16:27)



So, let us take another example, you take the example of role of a die and in this case again; let say 10 independent rolls of the die and you want to know what is the variance of the sum obtained from 10 independent rolls. So, as in the previous case, we can again have the same thing summation X i is summation variance of X i. So, this would give me. So, I want to compute variance of X i right. So, for rolling of a die right for rolling of a die you know we had the previous case we calculated E of X is equal to 1 into 1 by 6 plus 2 into 1 by 6 plus 6 into 1 by 6 and this will come to be. So, 1 plus 2 plus 3 plus 4 plus 5 plus 6 by 6, which is equal to 7 plus 7, 14 plus 7, 21 by 6, 7 by 2 and E of X i square f x square is equal to 1 square into 1 by 6 plus 2 square into 1 by 6.

So, finally, you will see variance of X i. So, I can write this as X i variance of X i will come to be. So, you can compute this value and see I think it will come to around 35 by 12 point something, but just check. So, if this is for a single variable then, I have to multiply. So, variance of summation X i then, becomes 10 into 35 by 12.

(Refer Slide Time: 18:54)



So, the correlation coefficient rho is given by nothing but covariance of X comma Y by root of var of X into var of Y. So, you can see how you can make use of expectation and variance for calculating various quantities and use this to estimate the variance of a population depending on the events. So, two important things from take away from this class, you use this general expression E of g x is nothing, but summation g x p x d x (Refer Time: 19:35) or integral g x f x d x. This is one thing to remember, you can write E of summation X i as summation E of X I, you can write variance of a X plus b is a square variance of X plus b. For independent RVs, you can have variance of summation X i is equal to summation variance of X i, but in the general case. So, you have to have variance of summation of X i is summation of var X i plus.

So, with that I thank you for your attention and will meet again in next class.