

We will now develop the solution to the linear least mean squares estimation problem. We want to find coefficients a and b that make this expression, this mean squared error, as small as possible. We will approach this problem by proceeding in two stages. In the first stage, we assume that the choice of a has already been found and concentrate on the question of choosing b .

Now if a has been found and has been fixed to a specific number, then this quantity here is a specific random variable. And what do we have? We have a random variable minus a constant. And we want to choose that constant, so that this difference squared is as small as possible in the expected value sense.

So essentially, we're trying to choose a constant b that estimates this random variable in the best possible way. But this is a problem that's familiar to us, and we know that the best choice of b is equal to the expected value of that random variable. And using also linearity, this expected value can be written in this form. So if we know a , this is how b should be chosen.

Let us now move on to the choice of a . Since we know what b should be equal to, we can rewrite this expression that we're trying to minimize by substituting our choice of b , which is the expected value of Θ minus aX . And this is the quantity we want to minimize.

What is it? We have a random variable minus the expected value of that random variable squared. And then we take expected value. This is just the variance of the random variable Θ minus aX . This is the variance of the difference of two random variables.

Because the two random variables are dependent, this is not just the sum of the individual variances. But we do have a formula, even for the general case. And the formula tells us that the variance of the difference of two random variables is the variance of the first random variable plus the variance of the second. And when we pull a outside the variance, that gives us a contribution of a squared.

And then we have a cross-term. Because of the minus sign here, the cross-term will have a minus sign. We have a factor of 2. And then we want the covariance of Θ with aX . And because we have seen that covariances behave in a linear manner, we can pull a outside the covariance. And this is what we are left with.

This is the quantity we want to optimize with respect to a . And to do this optimization, we just set the derivative with respect to a to 0. And that's going to give us $2a$ times the variance of X minus twice the covariance of Θ with X , equal to 0. From which it follows that a should be equal to the covariance of Θ with X divided by the variance of X .

So we have found what a should be. Once we know what a is, we know what b should be equal to. So we have solved the problem.

And here's the form of the solution. This coefficient here is the coefficient a . So we have here the term aX . And then this term together with a times the expected value of X , this corresponds to the coefficient b .

It's also instructive to rewrite this solution in a slightly different form involving the correlation coefficient. Recall that the correlation coefficient between two random variables is defined as the covariance between the two random variables divided by the product of their standard deviations. Using this relation, we can now write the coefficient a as the covariance which is $\text{Cov}(\Theta, X)$ divided by the variance of X , which is σ_X^2 .

And after we cancel a factor of σ_X from the numerator and the denominator, we see that a is also equal to ρ times σ_Θ divided by σ_X . And this gives us this alternative form for the solution. What we will be doing next will be to interpret this solution and also to give a number of examples.