The Minimum Mean Square Error Linear Estimate of a Random Vector from Another

We have random variables X and Y defined by the joint probability distribution function p, where $p(x, y)$ is the probability that X turns out to be x AND Y turns out to be y. We want to find a real number b such that we will open Y, get y, and claim by is our prediction of X. We want to decide b before either X or Y are opened. Once we have done our prediction, we will open X, get x, and see how well we fared. Our error will be $x - by$. We will be as sad as $(x - by)^2$. We want to minimize *today* our average future sadness. I.e. we want to find a b today, that minimizes

$$E[(X - bY)^2] \tag{1}$$

We have to understand the above equation. $(x - by)^2$ is the actual sadness we will have to bear. Since we do not know *today* how sad we will be *then*, today our future sadness is a random variable, namely $(X - bY)^2$, which is what we are average-futuring in expression (1). We easily differentiate the above expression and equate it to zero.

$$\frac{d}{db}E[(X - bY)^2] = 0$$

$$\frac{d}{db}E[X^2] - 2\frac{d}{db}bE[XY] + \frac{d}{db}b^2E[Y^2] = 0 \tag{2}$$

$$-2E[XY] + 2bE[Y^2] = 0$$

Thus, our optimum b is

$$b = \frac{E[XY]}{E[Y^2]} \tag{3}$$

This completes the proof.

Proof Using Error After Predictor is Opened

Though the above proof is enough, it might be interesting to see a variant. Let us say that Y has come out as y, and X is still inside the box. Our predictor is by. Since Y is now known, our understanding of X has changed to the "conditional" random variable $(X \mid y)$. Our predictor error is a random variable $(X \mid y) - by$, and our sadness is a random variable $((X \mid y) - by)^2$. Our average future sadness is

$$E_{X|y}[((X \mid y) - by)^2] \tag{4}$$

We want to minimize the above. But, we want to find the b even before the fact $Y = y$ is known, i.e. the above *average future* itself is a random variable. This random variable is $E_{X|Y}[((X \mid Y) - bY)^2]$. Now, while designing b, we want to find b that will minimize the *average future* average future sadness! I.e., the one that will minimize

$$\begin{aligned} & E_Y[E_{X|Y}[((X \mid Y) - bY)^2]] \\ & = E[((X \mid Y) - bY)^2] \end{aligned} \tag{5}$$

We have to understand that the random variable $X \mid Y$ is now the random variable X, since our expectation doesn't anymore talk about the intermediate future when Y is opened and X is still closed. This means we are back to the expression (1).