

Again, we obtain a quadratic function of  $a$ , and by differentiation we obtain the MMSE estimate of  $X$  given  $Y = y$  as

$$\hat{x}_M = E[X|Y = y].$$


---

Suppose that we would like to estimate the value of an unobserved random variable  $X$ , by observing the value of a random variable  $Y = y$ . In general, our estimate  $\hat{x}$  is a function of  $y$ , so we can write

$$\hat{X} = g(Y).$$

Note that, since  $Y$  is a random variable, the estimator  $\hat{X} = g(Y)$  is also a random variable. The error in our estimate is given by

$$\begin{aligned}\tilde{X} &= X - \hat{X} \\ &= X - g(Y),\end{aligned}$$

which is also a random variable. We can then define the *mean squared error* (MSE) of this estimator by

$$E[(X - \hat{X})^2] = E[(X - g(Y))^2].$$

From our discussion above we can conclude that the conditional expectation  $\hat{X}_M = E[X|Y]$  has the lowest MSE among all other estimators  $g(Y)$ .

### Mean Squared Error (MSE) of an Estimator

Let  $\hat{X} = g(Y)$  be an estimator of the random variable  $X$ , given that we have observed the random variable  $Y$ . The **mean squared error (MSE)** of this estimator is defined as

$$E[(X - \hat{X})^2] = E[(X - g(Y))^2].$$

The MMSE estimator of  $X$ ,

$$\hat{X}_M = E[X|Y],$$

has the lowest MSE among all possible estimators.

## Properties of the Estimation Error:

Here, we would like to study the MSE of the conditional expectation. First, note that

$$\begin{aligned} E[\hat{X}_M] &= E[E[X|Y]] \\ &= E[X] \quad (\text{by the law of iterated expectations}). \end{aligned}$$

Therefore,  $\hat{X}_M = E[X|Y]$  is an *unbiased* estimator of  $X$ . In other words, for  $\hat{X}_M = E[X|Y]$ , the estimation error,  $\tilde{X}$ , is a zero-mean random variable

$$E[\tilde{X}] = EX - E[\hat{X}_M] = 0.$$

Before going any further, let us state and prove a useful lemma.

---

### Lemma 9.1

Define the random variable  $W = E[\tilde{X}|Y]$ . Let  $\hat{X}_M = E[X|Y]$  be the MMSE estimator of  $X$  given  $Y$ , and let  $\tilde{X} = X - \hat{X}_M$  be the estimation error. Then, we have

- a.  $W = 0$ .
- b. For any function  $g(Y)$ , we have  $E[\tilde{X} \cdot g(Y)] = 0$ .

*Proof:*

- a. We can write

$$\begin{aligned} W &= E[\tilde{X}|Y] \\ &= E[X - \hat{X}_M|Y] \\ &= E[X|Y] - E[\hat{X}_M|Y] \\ &= \hat{X}_M - E[\hat{X}_M|Y] \\ &= \hat{X}_M - \hat{X}_M = 0. \end{aligned}$$

The last line resulted because  $\hat{X}_M$  is a function of  $Y$ , so  $E[\hat{X}_M|Y] = \hat{X}_M$ .

- b. First, note that

$$\begin{aligned} E[\tilde{X} \cdot g(Y)|Y] &= g(Y)E[\tilde{X}|Y] \\ &= g(Y) \cdot W = 0. \end{aligned}$$

Next, by the law of iterated expectations, we have

$$E[\tilde{X} \cdot g(Y)] = E[E[\tilde{X} \cdot g(Y)|Y]] = 0.$$


---

We are now ready to state a very interesting property of the estimation error for the MMSE estimator. Namely, we show that the estimation error,  $\tilde{X}$ , and  $\hat{X}_M$  are uncorrelated. To see this, note that

$$\begin{aligned}
\text{Cov}(\tilde{X}, \hat{X}_M) &= E[\tilde{X} \cdot \hat{X}_M] - E[\tilde{X}]E[\hat{X}_M] \\
&= E[\tilde{X} \cdot \hat{X}_M] \quad (\text{since } E[\tilde{X}] = 0) \\
&= E[\tilde{X} \cdot g(Y)] \quad (\text{since } \hat{X}_M \text{ is a function of } Y) \\
&= 0 \quad (\text{by Lemma 9.1}).
\end{aligned}$$

Now, let us look at  $\text{Var}(X)$ . The estimation error is  $\tilde{X} = X - \hat{X}_M$ , so

$$X = \tilde{X} + \hat{X}_M.$$

Since  $\text{Cov}(\tilde{X}, \hat{X}_M) = 0$ , we conclude

$$\text{Var}(X) = \text{Var}(\hat{X}_M) + \text{Var}(\tilde{X}). \quad (9.3)$$

The above formula can be interpreted as follows. Part of the variance of  $X$  is explained by the variance in  $\hat{X}_M$ . The remaining part is the variance in estimation error. In other words, if  $\hat{X}_M$  captures most of the variation in  $X$ , then the error will be small. Note also that we can rewrite [Equation 9.3](#) as

$$E[X^2] - E[X]^2 = E[\hat{X}_M^2] - E[\hat{X}_M]^2 + E[\tilde{X}^2] - E[\tilde{X}]^2.$$

Note that

$$E[\hat{X}_M] = E[X], \quad E[\tilde{X}] = 0.$$

We conclude

$$E[X^2] = E[\hat{X}_M^2] + E[\tilde{X}^2].$$

### Some Additional Properties of the MMSE Estimator

- The MMSE estimator,  $\hat{X}_M = E[X|Y]$ , is an unbiased estimator of  $X$ , i.e.,

$$E[\hat{X}_M] = EX, \quad E[\tilde{X}] = 0.$$

- The estimation error,  $\tilde{X}$ , and  $\hat{X}_M$  are uncorrelated

$$\text{Cov}(\tilde{X}, \hat{X}_M) = 0.$$

- We have

$$\begin{aligned} \text{Var}(X) &= \text{Var}(\hat{X}_M) + \text{Var}(\tilde{X}), \\ E[X^2] &= E[\hat{X}_M^2] + E[\tilde{X}^2]. \end{aligned}$$

Let us look at an example to practice the above concepts. This is an example involving jointly normal random variables. Thus, before solving the example, it is useful to remember the properties of jointly normal random variables. Remember that two random variables  $X$  and  $Y$  are jointly normal if  $aX + bY$  has a normal distribution for all  $a, b \in \mathbb{R}$ . As we have seen before, if  $X$  and  $Y$  are jointly normal random variables with parameters  $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2$ , and  $\rho$ , then, given  $Y = y$ ,  $X$  is normally distributed with

$$\begin{aligned} E[X|Y = y] &= \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}, \\ \text{Var}(X|Y = y) &= (1 - \rho^2)\sigma_X^2. \end{aligned}$$

---

#### **Example 9.7**

Let  $X \sim N(0, 1)$  and

$$Y = X + W,$$

where  $W \sim N(0, 1)$  is independent of  $X$ .

- Find the MMSE estimator of  $X$  given  $Y$ , ( $\hat{X}_M$ ).
- Find the MSE of this estimator, using  $MSE = E[(X - \hat{X}_M)^2]$ .
- Check that  $E[X^2] = E[\hat{X}_M^2] + E[\tilde{X}^2]$ .

### Solution

Since  $X$  and  $W$  are independent and normal,  $Y$  is also normal. Moreover,  $X$  and  $Y$  are also jointly normal, since for all  $a, b \in \mathbb{R}$ , we have

$$aX + bY = (a + b)X + bW,$$

which is also a normal random variable. Note also,

$$\begin{aligned}\text{Cov}(X, Y) &= \text{Cov}(X, X + W) \\ &= \text{Cov}(X, X) + \text{Cov}(X, W) \\ &= \text{Var}(X) = 1.\end{aligned}$$

Therefore,

$$\begin{aligned}\rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= \frac{1}{1 \cdot \sqrt{2}} = \frac{1}{\sqrt{2}}.\end{aligned}$$

a. The MMSE estimator of  $X$  given  $Y$  is

$$\begin{aligned}\hat{X}_M &= E[X|Y] \\ &= \mu_X + \rho\sigma_X \frac{Y - \mu_Y}{\sigma_Y} \\ &= \frac{Y}{2}.\end{aligned}$$

b. The MSE of this estimator is given by

$$\begin{aligned}E[(X - \hat{X}_M)^2] &= E\left[\left(X - \frac{Y}{2}\right)^2\right] \\ &= E\left[X^2 - XY + \frac{Y^2}{4}\right] \\ &= EX^2 - E[X(X + W)] + \frac{EY^2}{4} \\ &= EX^2 - EX^2 - EXEW + \frac{EY^2}{4} \\ &= \frac{\text{Var}(Y) + (EY)^2}{4} \\ &= \frac{2 + 0}{4} = \frac{1}{2}.\end{aligned}$$

c. Note that  $E[X^2] = 1$ . Also,

$$E[\hat{X}_M^2] = \frac{EY^2}{4} = \frac{1}{2}.$$

In the above, we also found  $MSE = E[\tilde{X}^2] = \frac{1}{2}$ . Therefore, we have

$$E[X^2] = E[\hat{X}_M^2] + E[\tilde{X}^2].$$


---