9.1.5 Mean Squared Error (MSE)

Suppose that we would like to estimate the value of an unobserved random variable X given that we have observed $Y = y$. In general, our estimate $\hat{x}$ is a function of y:

$$\hat{x} = g(y).$$

The error in our estimate is given by

$$\begin{aligned}\tilde{X} &= X - \hat{x} \\ &= X - g(y).\end{aligned}$$

Often, we are interested in the *mean squared error* (MSE) given by

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

One way of finding a point estimate $\hat{x} = g(y)$ is to find a function $g(Y)$ that minimizes the *mean squared error* (MSE). Here, we show that $g(y) = E[X|Y = y]$ has the lowest MSE among all possible estimators. That is why it is called the *minimum mean squared error (MMSE) estimate*.

For simplicity, let us first consider the case that we would like to estimate X without observing anything. What would be our best estimate of X in that case? Let a be our estimate of X. Then, the MSE is given by

$$\begin{aligned}h(a) &= E[(X - a)^2] \\ &= EX^2 - 2aEX + a^2.\end{aligned}$$

This is a quadratic function of a, and we can find the minimizing value of a by differentiation:

$$h'(a) = -2EX + 2a.$$

Therefore, we conclude the minimizing value of a is

$$a = EX.$$

Now, if we have observed $Y = y$, we can repeat the above argument. The only difference is that everything is conditioned on $Y = y$. More specifically, the MSE is given by

$$\begin{aligned}h(a) &= E[(X - a)^2 | Y = y] \\ &= E[X^2 | Y = y] - 2aE[X|Y = y] + a^2.\end{aligned}$$