Homework 6 Solutions

1. As $\hat{\beta}_1$ is an unbiased estimator of β_1, the best accuracy is equivalent to the smallest $\mathrm{var}(\hat{\beta}_1)$. Recall that

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^4 (x_i - \bar{x})^2}.$$

Therefore, the smallest variance is achieved when $\sum_{i=1}^4 (x_i - \bar{x})^2$ is the largest. Since we have $x_1, \ldots, x_4 \in [-1, 1]$, notice that

$$\sum_{i=1}^4 (x_i - \bar{x})^2 = \sum_{i=1}^4 x_i^2 - 4\bar{x}^2 \le \sum_{i=1}^4 x_i^2 \le 4,$$

where the inequalities become equality when any two of $x_1, \ldots, x_4$ are 1 and the other two are -1, e.g., $x_1 = x_2 = 1$ and $x_3 = x_4 = -1$.

2. When $p = 2$, we have

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Hence, letting $\bar{x} = \sum_{i=1}^n x_i / n$ and $\bar{y} = \sum_{i=1}^n y_i / n$, we have

$$X^T X = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix} \quad \text{and} \quad X^T y = \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix}.$$

The inverse of $X^T X$ is

$$(X^T X)^{-1} = \frac{1}{n \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}. \tag{1}$$

Hence,

$$\begin{aligned} \hat{\beta} &= \frac{1}{n \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2} \begin{bmatrix} n \sum_{i=1}^n x_i^2 \bar{y} - n\bar{x} \sum_{i=1}^n x_i y_i \\ -n^2 \bar{x}\bar{y} + n \sum_{i=1}^n x_i y_i \end{bmatrix} \\ &= \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 \bar{y} - n\bar{x}^2 \bar{y} + n\bar{x}^2 \bar{y} - \bar{x} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{bmatrix}. \end{aligned}$$

From this, we can see that the second component of $\hat{\beta}$ is

$$\frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

which is $\hat{\beta}_1$ derived from solving the linear equations. Also, the first component of $\hat{\beta}$ is

$$\frac{\sum_{i=1}^n x_i^2 \bar{y} - n\bar{x}^2 \bar{y} + n\bar{x}^2 \bar{y} - \bar{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \bar{y} - \hat{\beta}_1 \bar{x},$$

which is $\hat{\beta}_0$ derived from solving the linear equations. Hence, $\hat{\beta}$ leads to the same formula.

3. From $\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1})$, we first have $\mathbb{E}\hat{\beta} = \beta$, which coincides with the straightforward calculations from $\hat{\beta}_0, \hat{\beta}_1$. Using the above equation (1), the covariance matrix of $\hat{\beta}$ is

$$\sigma^2(X^\top X)^{-1} = \frac{\sigma^2}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2}\begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}.$$

From this, we can see that the variance of the first component of $\hat{\beta}$ is

$$\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} = \frac{\sigma^2(\sum_{i=1}^n x_i^2 - n\bar{x}^2 + n\bar{x}^2)}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} = \frac{\sigma^2}{n} + \frac{\sigma^2\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Meanwhile, the variance of the second component of $\hat{\beta}$ is

$$\frac{\sigma^2 n}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Lastly, the covariance of the first and second components of $\hat{\beta}$ is

$$\frac{-\sigma^2 n\bar{x}}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} = -\frac{\sigma^2\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

These match the following results based on straightforward calculations from $\hat{\beta}_0, \hat{\beta}_1$:

$$\begin{aligned} \mathrm{var}(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \frac{\sigma^2\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \mathrm{var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= -\frac{\sigma^2\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

4-(a). As $\frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p}$, we have

$$\mathbb{E}\hat{\sigma}_c^2 = c\mathbb{E}\|y - X\hat{\beta}\|^2 = c\sigma^2(n-p).$$

Hence, $c = \frac{1}{n-p}$ leads to an unbiased estimator.

4-(b). The log-likelihood is given by

$$\ell(\beta, \sigma^2) = -\frac{\|y - X\beta\|^2}{2\sigma^2} - \frac{n\log\sigma^2}{2} + \text{ some constant.}$$

From this, we can deduce that the MLE of β is the LSE $\hat{\beta}$ and the MLE of σ^2 is obtained by solving

$$\frac{\|y - X\hat{\beta}\|^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0,$$

which leads to

$$\frac{\|y - X\hat{\beta}\|^2}{n}.$$

Hence, $c = \frac{1}{n}$ gives the MLE of σ^2.

4-(c). By the bias-variance decomposition,

$$\mathbb{E}(\hat{\sigma}_c^2 - \sigma^2)^2 = \mathrm{var}(\hat{\sigma}_c^2) + (\mathbb{E}\hat{\sigma}_c^2 - \sigma^2)^2 = \mathrm{var}(\hat{\sigma}_c^2) + \sigma^4(c(n-p)-1)^2.$$

As $\frac{\|y-X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p}$, we have

$$\mathrm{var}(\hat{\sigma}_c^2) = c^2\mathrm{var}(\|y - X\hat{\beta}\|^2) = 2c^2\sigma^4(n-p).$$

Hence,

$$\mathbb{E}(\hat{\sigma}_c^2 - \sigma^2)^2 = \sigma^4(2(n-p)c^2 + (c(n-p)-1)^2),$$

which is minimized at $c = \frac{1}{n-p+2}$.

4-(d). We have

$$\mathbb{P}\left(\chi^2_{n-p,\frac{\alpha}{2}} \le \frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \le \chi^2_{n-p,1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Hence,

$$\mathbb{P}\left(\frac{\|y - X\hat{\beta}\|^2}{\chi^2_{n-p,1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{\|y - X\hat{\beta}\|^2}{\chi^2_{n-p,\frac{\alpha}{2}}}\right) = 1-\alpha.$$

Therefore, we have a confidence interval

$$\left[\frac{\|y - X\hat{\beta}\|^2}{\chi^2_{n-p,1-\frac{\alpha}{2}}}, \frac{\|y - X\hat{\beta}\|^2}{\chi^2_{n-p,\frac{\alpha}{2}}}\right] = \left[\frac{\hat{\sigma}_c^2/c}{\chi^2_{n-p,1-\frac{\alpha}{2}}}, \frac{\hat{\sigma}_c^2/c}{\chi^2_{n-p,\frac{\alpha}{2}}}\right].$$

5-(a). Since $\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$, we have

$$(x^*)^T\hat{\beta} \sim N\left((x^*)^T\beta, \sigma^2(x^*)^T(X^TX)^{-1}x^*\right).$$

5-(b). Note that

$$\frac{(x^*)^T\beta - (x^*)^T\hat{\beta}}{\sigma\sqrt{(x^*)^T(X^TX)^{-1}x^*}} \sim N(0,1).$$

Define $\hat{\sigma}^2 = \|y - X\hat{\beta}\|^2/(n-p)$. Then, from Q4, we have

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}.$$

Also, we have learned that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent. Hence,

$$\frac{\frac{(x^*)^T\beta - (x^*)^T\hat{\beta}}{\sigma\sqrt{(x^*)^T(X^TX)^{-1}x^*}}}{\sqrt{\hat{\sigma}^2/\sigma^2}} = \frac{(x^*)^T\beta - (x^*)^T\hat{\beta}}{\hat{\sigma}\sqrt{(x^*)^T(X^TX)^{-1}x^*}} \sim t_{n-p}.$$

Therefore, we have a confidence interval

$$\left((x^*)^T\hat{\beta} - t_{n-p,1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{(x^*)^T(X^TX)^{-1}x^*}, (x^*)^T\hat{\beta} + t_{n-p,1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{(x^*)^T(X^TX)^{-1}x^*}\right).$$

6-(a). We have

$$\begin{aligned}\mathbb{E}(\hat{\theta}-\theta)^2 &= \mathbb{E}(\hat{\theta}-\mathbb{E}\hat{\theta}+\mathbb{E}\hat{\theta}-\theta)^2\\ &= \mathbb{E}(\hat{\theta}-\mathbb{E}\hat{\theta})^2 + 2\mathbb{E}(\hat{\theta}-\mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta}-\theta) + (\mathbb{E}\hat{\theta}-\theta)^2\\ &= \mathbb{E}(\hat{\theta}-\mathbb{E}\hat{\theta})^2 + (\mathbb{E}\hat{\theta}-\theta)^2\\ &= \mathrm{var}(\hat{\theta}) + (\mathbb{E}\hat{\theta}-\theta)^2,\end{aligned}$$

where the third equality uses $\mathbb{E}(\hat{\theta}-\mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta}-\theta) = (\mathbb{E}\hat{\theta}-\mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta}-\theta) = 0$.

6-(b). We have

$$\begin{aligned}\sum_{i=1}^n (x_i-\theta)^2 &= \sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\theta)^2\\ &= \sum_{i=1}^n (x_i-\bar{x})^2 + 2\sum_{i=1}^n (x_i-\bar{x})(\bar{x}-\theta) + \sum_{i=1}^n (\bar{x}-\theta)^2\\ &= \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2,\end{aligned}$$

where the third equality uses $\sum_{i=1}^n (x_i-\bar{x})(\bar{x}-\theta) = (\sum_{i=1}^n (x_i-\bar{x}))(\bar{x}-\theta) = 0$.

6-(c). Similarly, we have

$$\begin{aligned}\sum_{i=1}^n \|X_i-\theta\|^2 &= \sum_{i=1}^n \|X_i-\bar{X}+\bar{X}-\theta\|^2\\ &= \sum_{i=1}^n \|X_i-\bar{X}\|^2 + 2\sum_{i=1}^n (X_i-\bar{X})^T(\bar{X}-\theta) + \sum_{i=1}^n \|\bar{X}-\theta\|^2\\ &= \sum_{i=1}^n \|X_i-\bar{X}\|^2 + n\|\bar{X}-\theta\|^2,\end{aligned}$$

where the third equality uses $\sum_{i=1}^n (X_i-\bar{X})^T(\bar{X}-\theta) = (\sum_{i=1}^n (X_i-\bar{X}))^T(\bar{X}-\theta) = 0$.

6-(d). We have

$$\begin{aligned}\|y-X\beta\|^2 &= \|y-X\hat{\beta}+X\hat{\beta}-X\beta\|^2\\ &= \|y-X\hat{\beta}\|^2 + 2(y-X\hat{\beta})^T(X\hat{\beta}-X\beta) + \|X\hat{\beta}-X\beta\|^2\\ &= \|y-X\hat{\beta}\|^2 + \|X\hat{\beta}-X\beta\|^2,\end{aligned}$$

where the third equality uses

$$\begin{aligned}(y-X\hat{\beta})^T(X\hat{\beta}-X\beta) &= y^T(I_n - X(X^TX)^{-1}X^T)^T X(\hat{\beta}-\beta)\\ &= y^T(X - X(X^TX)^{-1}X^TX)(\hat{\beta}-\beta)\\ &= 0.\end{aligned}$$

6-(e). Basically, all the identities above can be written as $\|u-v\|^2 = \|u-w\|^2 + \|v-w\|^2$, where $\|\cdot\|$ is a suitable norm based on the inner product with an appropriate w such that $(u-w) \perp (v-w)$.t