STAT 245 HW5 Solution

Yuguan Wang

January 2023

1. (a) We have

$$\begin{aligned} E[M] &= \sum_{j=1}^{m} E[I\{\mu_j \notin [\hat{\mu}_{j,\text{left}}, \hat{\mu}_{j,\text{right}}]\}] \\ &= \sum_{j=1}^{m} P(\mu_j \notin [\hat{\mu}_{j,\text{left}}, \hat{\mu}_{j,\text{right}}]) = m\alpha. \end{aligned}$$

(b) We know that

$$P(\mu_j \notin [\hat{\mu}_j + z_{\alpha/2m}, \hat{\mu}_j + z_{1-\alpha/2m}]) = \alpha/m$$

where $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}$. Define

$$\hat{\mu}_{\text{left}} = \begin{pmatrix} \hat{\mu}_1 + z_{\alpha/2m} \\ \hat{\mu}_2 + z_{\alpha/2m} \\ \vdots \\ \hat{\mu}_m + z_{\alpha/2m} \end{pmatrix}, \quad \hat{\mu}_{\text{right}} = \begin{pmatrix} \hat{\mu}_1 + z_{1-\alpha/2m} \\ \hat{\mu}_2 + z_{1-\alpha/2m} \\ \vdots \\ \hat{\mu}_m + z_{1-\alpha/2m} \end{pmatrix}.$$

You can apply union bound to get

$$\begin{aligned} P(\mu \notin [\hat{\mu}_{\text{left}}, \hat{\mu}_{\text{right}}]) &= P(\cup_{j=1}^{m}\{\mu_j \notin [\hat{\mu}_{j,\text{left}}, \hat{\mu}_{j,\text{right}}]\}) \\ &\leq \sum_{j=1}^{m} P(\mu_j \notin [\hat{\mu}_{j,\text{left}}, \hat{\mu}_{j,\text{right}}]) = m\frac{\alpha}{m} = \alpha. \end{aligned}$$

In other words

$$P(\mu \in [\hat{\mu}_{\text{left}}, \hat{\mu}_{\text{right}}]) \geq 1 - \alpha.$$

2. Let $H = X(X^TX)^{-1}X^T$. Check that

- $(I - H)^T = I - H$.
- $(I - H)^2 = I - H$.
- $X^T(I - H) = 0$.

Note that by plugging the formula for $\hat{\beta}$ we have

$$y - X\hat{\beta} = (I - H)y.$$

Thus, we have

$$\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1}), \quad y - X\hat{\beta} \sim N(0, \sigma^2(I - H)).$$

Therefore, to show independence between $\hat{\beta}_j$ and $y_i - X_i\hat{\beta}$ it is enough to show that their covariance is 0. We have

$$\begin{aligned}
E[\hat{\beta}(y - X\hat{\beta})^T] &= E[(X^TX)^{-1}X^Tyy^T(I - H)] \\
&= (X^TX)^{-1}X^TE[yy^T](I - H) \\
&= (X^TX)^{-1}X^T(X\beta\beta^TX^T + \sigma^2 I)(I - H) \\
&= (\beta\beta^TX^T + \sigma^2(X^TX)^{-1}X^T)(I - H) \\
&= ((\beta\beta^T + \sigma^2(X^TX)^{-1})\underbrace{X^T(I - H)}_{=0} \\
&= 0.
\end{aligned}$$

3. (a) Let

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad X_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$$

and

$$\Sigma = \sigma^2 \begin{pmatrix} I_{n_1} & \\ & 2I_{n_2} \end{pmatrix}.$$

The kernel of the log-likelihood is

$$l(\beta) = -\frac{1}{2}(Y - X\beta)^T\Sigma^{-1}(Y - X\beta).$$

Take the derivative with respect to β and set it to zero:

$$-2X^T\Sigma^{-1}Y + 2X^T\Sigma^{-1}X\beta = 0 \Rightarrow \hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y.$$

(b)

$$E[\hat{\beta}] = E[(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y] = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}E[Y] = \beta.$$

(c)

$$\mathrm{Cov}(\hat{\beta}) = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\mathrm{Cov}(Y)[(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}]^T = (X^T\Sigma^{-1}X)^{-1}.$$

(d) $\hat{\beta} \sim N(\beta, (X^T\Sigma^{-1}X)^{-1})$ because $\hat{\beta}$ is linear transform of a normal random vector.