Midterm SMI Mathematical Statistics, Summer 2023

Each part of each problem is worth 20 points, for a total of 80 points.

1. Consider the standard linear model $Y = X\beta + \epsilon$ where $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$ and $r(X) = p \le n$. Take $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I)$. Let also $D \in \mathbb{R}^{q\times p}$ with $r(D) = q \le p$. In addition, assume that the columns of X are mutually orthogonal and have Euclidean length equal to 1.

1.1 Write out the formula for $\widehat{\beta}_H$, the least squares estimator of β subject to the constraint that $D\beta = c$ in this special case.

Solution. Recall that in general we have

$$\widehat{\beta}_H = \widehat{\beta} + (X^T X)^{-1} D^T [D(X^T X)^{-1} D^T]^{-1} (c - D\widehat{\beta}).$$

As the columns of X are orthogonal and have length one, we obtain $X^\top X = I_p$, the identity matrix in $\mathbb{R}^p$, and the formula simplifies to

$$\widehat{\beta}_H = \widehat{\beta} + D^T [DD^T]^{-1} (c - D\widehat{\beta}) \quad \text{where} \quad \widehat{\beta} = X^\top Y. \tag{1}$$

1.2 Consider the regression model

$$Y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \epsilon_i$$

and assume the conditions stated above are satisfied. Use the previous problem to find the least squares estimate $\widehat{\beta}_H$ of β subject to the constraint that $\beta_2 = \beta_3$, that is, under the condition that each of these two predictors have the same influence on the outcome. Express $\widehat{\beta}_H$ in terms of (only) $\widehat{\beta}$ and verify that it satisfies the given constraint.

Solution. In this case $D = [0, 1, -1]$ and $c = 0$, and (1) specializes to

$$\widehat{\beta}_H = \widehat{\beta} - D^T [DD^T]^{-1} D\widehat{\beta} = (I - D^T [DD^T]^{-1} D)\widehat{\beta} = (I - D^T D/2)\widehat{\beta},$$

since $DD^T = 2$. Further

$$DD^\top = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} [0, 1, -1] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

so

$$I - D^T D/2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

hence

$$\widehat{\beta}_H = (I - D^T D/2)\widehat{\beta} = (\widehat{\beta}_1, (\widehat{\beta}_2 + \widehat{\beta}_3)/2, (\widehat{\beta}_3 + \widehat{\beta}_2)/2)^\top.$$

The estimate of β_1 is unchanged, and the two other slope estimates are equal, as desired.

2. Again consider the linear model under the assumptions that $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}$ and $r(X) = p \leq n$. Take $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I)$.

2.1 Compute the Information lower bound for the unbiased estimation of β, and determine if the least squares estimator achieves it.

Solution: As $\mathbf{Y}$ is a linear transformation of a multivariate normal it is also multivariate normal, and as it has mean $X\beta$ and variance $\sigma^2 I$, we see that $\mathbf{Y} \sim \mathcal{N}(X\beta, \sigma^2 I)$ and therefore has density

$$p(y;\beta) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2}\|Y - X\beta\|^2.\right)$$

As the normalizing constant does not depend on β, taking the partial derivative of the log with respect to β yields

$$\begin{aligned} \partial_\beta \log p(y;\beta) = \partial_\beta \left(-\frac{1}{2\sigma^2}\|Y - X\beta\|^2\right) &= -\frac{1}{2\sigma^2}\left(-2X^T Y + 2X^T X\beta\right) \\ &= -\frac{1}{\sigma^2}\left(-X^T Y + X^T X\beta\right). \end{aligned}$$

Hence, the information matrix is given by

$$\begin{aligned} \frac{1}{\sigma^4}\mathrm{Var}_\beta\left(-X^T Y + X^T X\beta\right) &= \frac{1}{\sigma^2}\mathrm{Var}_\beta\left(X^T Y\right) \\ &= \frac{1}{\sigma^4}X^\top \mathrm{Var}_\beta\left(Y\right)X = \frac{1}{\sigma^2}X^\top X. \end{aligned}$$

As the variance of the least squares estimator is also equal to $\sigma^2(X^T X)^{-1}$, it achieves the information bound.

2.2 Now specialize to the one dimensional linear regression model

$$Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$$

and assume the conditions above are satisfied. In general, does one achieve better estimates of β_2 by having knowledge of β_1? If so, are there any conditions one can impose on the model so that the estimation of β_2 is not affected by the lack of knowledge of the value of β_1?

Solution: The design matrix here is given by $X = (\mathbf{1}, \mathbf{x})$ where the first column is all ones, and the second column is $\mathbf{x} = (x_1, \ldots, x_n)^T$. Therefore

$$X^\top X = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} =: n \begin{bmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{bmatrix}$$

With ρ the correlation between the components of the score function, the effective information I_{22}^* for estimating β_2 when β_1 is unknown is given by

$$I_{11}^* = I_{11}(1 - \rho^2) = n \left(1 - \frac{\overline{x}^2}{\overline{x^2}}\right).$$

Hence, in general this quantity takes on values strictly less than I_{11}, showing that ignorance of β_1 will degrade the estimation of β_2, except only in the special case where $\sum_i x_i = 0$.