STATISTICAL MODELLING Part IIC / Michaelmas 2022
Example Sheet 1 (of 4)

In all the questions that follow, X is an n by p design matrix with full column rank and H is the orthogonal projection on to the column space of X. Also, let X_0 be the matrix formed from the first $p_0 < p$ columns of X and let H_0 be the orthogonal projection on to the column space of X_0. The vector $Y \in \mathbb{R}^n$ will be a vector of responses and we will define $\hat{\beta} := (X^T X)^{-1} X^T Y$, $\hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T Y$ and $\tilde{\sigma}^2 := \|(I-H)Y\|^2/(n-p)$. By normal linear model, we mean the model $Y = X\beta + \epsilon$, $\epsilon \mid X \sim N_n(0, \sigma^2 I)$.

1. Show that

$$\|(H - H_0)Y\|^2 = \|(I - H_0)Y\|^2 - \|(I - H)Y\|^2 = \|HY\|^2 - \|H_0 Y\|^2.$$

2. Let Σ be a known $n \times n$ positive definite matrix. Consider the following linear model

$$Y = X\beta + \varepsilon, \qquad \varepsilon \mid X \sim N(0, \sigma^2\Sigma),$$

where β and σ^2 are unknown. Let $\hat{\beta}_\Sigma$ denote the maximum likelihood estimator of β. What optimisation problem does $\hat{\beta}_\Sigma$ solve? Show that if $(X^T\Sigma^{-1}X)$ is invertible, $\hat{\beta}_\Sigma$ is given by

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

This is called the *generalised least squares* estimator. Show that $\hat{\beta}_\Sigma$ is the best linear unbiased estimator (BLUE) in this model.

3. Consider the model $Y = \mu + \varepsilon$ where $\mathbb{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I$ and $\mu \in \mathbb{R}^n$ is a non-random vector. Suppose we have have performed least squares of Y on a fixed design matrix X, so the fitted values are HY. Show that if $Y^* = \mu + \varepsilon^*$ where $\mathbb{E}(\varepsilon^*) = 0$, $\mathrm{Var}(\varepsilon^*) = \sigma^2 I$ and ε^* is independent of ε, then

$$\frac{1}{n}\mathbb{E}(\|HY - Y^*\|^2) = \sigma^2 + \frac{1}{n}\|(I - H)\mu\|^2 + \frac{\sigma^2 p}{n}.$$

This is known as the *bias-variance tradeoff*, as a larger model reduces the second term ("bias^2") but increases the third term ("variance").

4. (a) Let V and W be linear subspaces of $\mathbb{R}^n$ with $V \subseteq W$. Let Π_V and Π_W denote orthogonal projections on to V and W respectively. Show that for all $x \in \mathbb{R}^n$, $\|x\|^2 \geq \|\Pi_W x\|^2 \geq \|\Pi_V x\|^2$.

   (b) Consider the normal linear model with a fixed design matrix X. Suppose only the first p_0 components of β are non-zero. Show that

$$\mathrm{Var}(\hat{\beta}_{0,j}) \leq \mathrm{Var}(\hat{\beta}_j) \qquad \text{for } j = 1, \ldots, p_0.$$

   Here $\hat{\beta}_{0,j}$ denotes the j^{th} component of $\hat{\beta}_0$. *Hint: Use the partial regression characterisation of $\hat{\beta}_j$ that if $X_j^\perp$ is the orthogonal projection of X_j (the j^{th} column of X) on to the orthogonal complement of the column space of X_{-j} (the matrix formed by removing the j^{th} column from X), then $\hat{\beta}_j = \frac{(X_j^\perp)^T Y}{\|X_j^\perp\|^2}$.*

5. Show that the maximum likelihood estimator of σ^2 in the normal linear model is $\hat{\sigma}^2_{\text{MLE}} = \|(I - H)Y\|^2/n$. Without assuming normality, that is, assuming only $\mathbb{E}(\varepsilon \mid X) = 0$ and $\mathrm{Cov}(\varepsilon \mid X) = \sigma^2 I$, show that $\mathbb{E}(\hat{\sigma}^2_{\text{MLE}}) = (n - p)\sigma^2/n$ (thus $\hat{\sigma}^2_{\text{MLE}}$ is biased).

6. Let the cuboid C be defined $C := \prod_{j=1}^p C_j(\alpha/p)$, where

$$C_j(\alpha) = \left[\hat{\beta}_j - \sqrt{\tilde{\sigma}^2(X^TX)^{-1}_{jj}}\,t_{n-p}(\alpha/2),\ \hat{\beta}_j + \sqrt{\tilde{\sigma}^2(X^TX)^{-1}_{jj}}\,t_{n-p}(\alpha/2)\right].$$

Assuming the normal linear model, show that $\mathbb{P}_{\beta,\sigma^2}(\beta \in C) \geq 1 - \alpha$.

7. Let $\hat{\sigma}^2 = \|(I-H)Y\|^2/(n-p)$. Assuming the normal linear model, show that $(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$. Use this as a pivotal quantity to construct a confidence interval of σ^2 with level $1-\alpha$.

8. Suppose we observe data (X, Y) generated from a normal linear model. Let $x^* \in \mathbb{R}^p$ be a fixed vector and $\epsilon^* \sim N(0, \sigma^2)$ be independent of (X, Y). Denote $Y^* = (x^*)^T\beta + \epsilon^*$. Construct $(1-\alpha)$-confidence intervals for $(x^*)^T\beta$ and Y^*. Which interval is shorter in length? Can you give an intuitive explanation to your answer?

9. In the normal linear model, show that the likelihood ratio statistic for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ can be written as a monotonically increasing function of $\|(I - H_0)Y\|^2 / \|(I - H)Y\|^2$.

10. In the normal linear model, show that $\mathrm{Var}(Y_i - X_i^T\hat{\beta} \mid X) = \sigma^2(1 - H_{ii})$. *Hint: Write the residual $Y_i - X_i^T\hat{\beta}$ as a linear transformation of Y.*

11. Data are available on weights of two groups of three rats at the beginning of a fortnight and at its end. During the fortnight, one group was fed normally, and the other was given a growth inhibitor. The weights of the k^{th} rat in the j^{th} group before and after the fortnight are X_{jk} and y_{jk} respectively. The y_{jk} are taken as realisations of random variables Y_{jk} that follow the model $Y_{jk} = \alpha_j + \beta_j X_{jk} + \varepsilon_{jk}$.

    (a) Let W be the vector of responses, so $W = (Y_{11}, Y_{12}, Y_{13}, Y_{21}, Y_{22}, Y_{23})^T$, and similarly let δ be the vector of random errors. Write down the model above in the form $W = A\theta + \delta$, giving the design matrix A explicitly and expressing the vector of parameters θ in terms of the α_j and β_j.

    (b) The model is to be reparametrised in such a way that it can be specialised to (i) two parallel lines for the two groups, (ii) two lines with the same intercept, (iii) one common line for both groups, just by setting parameters to zero. Give one design matrix that can be made to correspond to (i), (ii) and (iii), just by dropping columns, specifying which columns are to be dropped for which cases.

12. Prove that in the linear model $Y = X\beta + \varepsilon$ with $\varepsilon \mid X \sim N_n(0, \sigma^2 I)$, the F-test and t-test for testing $H_0 : \beta_p = 0$ versus $H_1 : \beta_p \neq 0$ are equivalent. That is, prove that, taking $p_0 = p - 1$, so H_0 is the orthogonal projection on to the first $p-1$ columns of X,

$$\frac{\hat{\beta}_p^2}{\{(X^TX)^{-1}\}_{pp}\hat{\sigma}^2} = \frac{\|(H - H_0)Y\|^2}{\frac{1}{n-p}\|(I - H)Y\|^2}.$$

    Verify the equivalence of the F- and t-tests in an example using R. *Hint: Look at the hint to question 4(b).*