Qualifying Exam: CAS MA575, Linear Models

Boston University, Spring 2018

1. The theory of optimal linear *prediction* in many ways parallels closely the theory of optimal linear *estimation*. In this question you are asked to derive a fundamental result pertaining to the former.

   Let $Y \in \mathbb{R}$ be a random response, and let $\boldsymbol{x} \in \mathbb{R}^p$ be a *random* predictor. We seek a linear predictor $\alpha + \boldsymbol{x}^T\beta$ that minimizes $\mathbb{E}\left[Y - \alpha - \boldsymbol{x}^T\beta\right]^2$ over all choice of $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^p$. Such a predictor is called a *best linear predictor* of Y.

   Define $\mu_y = \mathbb{E}[Y]$, $\mu_x = \mathbb{E}[\boldsymbol{x}]$, $\sigma_y^2 = \mathrm{Var}(Y)$, $V_{xx} = \mathrm{Cov}(\boldsymbol{x})$, and $V_{xy} = \mathrm{Cov}(\boldsymbol{x}, Y) = V_{yx}^T = \mathrm{Cov}^T(Y, \boldsymbol{x})$. Without loss of generality, we will write an arbitrary linear predictor in the form $\alpha + (\boldsymbol{x} - \mu_x)^T \beta$.

   (a) Show that the optimal choice of α is simply $\alpha = \mu_y$.

   (b) Show that if β_* is a solution to the linear system $V_{xx}\beta = V_{xy}$, then $\mu_y + (\boldsymbol{x} - \mu_x)^T \beta_*$ is a best linear predictor of Y. [Hint: Without loss of generality, you may assume $\mu_y = 0$ and $\mu_x = 0$ for this part.]

   (c) Derive an expression for a 95% prediction interval for y. State any assumptions necessary for the validity of your interval.

2. In many applications, there can be multiple response variables, in which case the linear regression model can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E},$$

where $\boldsymbol{Y} = (y_{ij})_{1\le i\le n, 1\le j\le d}$ is a matrix in $\mathbb{R}^{n\times d}$, $\boldsymbol{X} = (x_{ij})_{1\le i\le n, 1\le j\le p}$ is a matrix in $\mathbb{R}^{n\times p}$, $\boldsymbol{B} = (\beta_{ij})_{1\le i\le p, 1\le j\le d}$ is a matrix in $\mathbb{R}^{p\times d}$ and $\boldsymbol{E} = (e_{ij})_{1\le i\le n, 1\le j\le d}$ is a matrix in $\mathbb{R}^{n\times d}$. Assume that the errors e_{ij}, $1 \le i \le n$, $1 \le i \le d$, are independent normal random variables with mean zero and variance σ^2. In this case, the least squares estimator is defined as

$$\hat{\boldsymbol{B}} = \operatorname*{argmin}_{\boldsymbol{B}\in\mathbb{R}^{p\times d}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}\|^2,$$

where for a matrix $\boldsymbol{A} = (a_{ij})$, the norm is defined as $\|\boldsymbol{A}\| = (\sum_{i,j} a_{ij}^2)^{1/2}$. Suppose you only observe the response matrix $\boldsymbol{Y}$ and the design matrix $\boldsymbol{X}$.

(a) Find the least squares estimate $\hat{\boldsymbol{B}}$.

(b) Find the least squares estimate $\hat{\sigma}^2$.

(c) Construct a test for the null hypothesis that the first column of $\boldsymbol{X}$ is an irrelevant variable.