SOLUTIONS TO THE FINAL EXAM
SMI MATHEMATICAL STATISTICS, AUGUST 13, 2026

1. Consider

$$Y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \qquad i = 1, \dots, n,$$

where the x_i are fixed and the errors are independent $\mathcal{N}(0, \sigma^2)$ random variables. Put

$$S_1 = \sum_{i=1}^n x_i, \qquad S_2 = \sum_{i=1}^n x_i^2, \qquad \overline{x} = \frac{S_1}{n}.$$

The log likelihood, up to an additive constant, is

$$\ell(\beta_1, \beta_2) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_1 - \beta_2 x_i)^2.$$

Hence the score vector is

$$U(\beta) = \begin{pmatrix} U_1(\beta) \\ U_2(\beta) \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} \sum_{i=1}^n (Y_i - \beta_1 - \beta_2 x_i) \\ \sum_{i=1}^n x_i (Y_i - \beta_1 - \beta_2 x_i) \end{pmatrix}.$$

The Fisher information matrix in the two-parameter model is therefore

$$I(\beta) = \frac{1}{\sigma^2} \begin{pmatrix} n & S_1 \\ S_1 & S_2 \end{pmatrix}.$$

(a) If β_1 is known, then

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

is a one-parameter normal model. Its score is

$$U_2(\beta_2) = \frac{1}{\sigma^2} \sum_{i=1}^n x_i (Y_i - \beta_1 - \beta_2 x_i),$$

so

$$I_{22}(\beta) = \operatorname{Var}\big(U_2(\beta_2)\big) = \frac{S_2}{\sigma^2}.$$

Consequently, the information lower bound for the variance of an unbiased estimator of β_2 is

$$\frac{1}{I_{22}(\beta)} = \frac{\sigma^2}{S_2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

(b) When β_1 is unknown, it is a nuisance parameter. The inverse information matrix is

$$I(\beta)^{-1} = \frac{\sigma^2}{nS_2 - S_1^2} \begin{pmatrix} S_2 & -S_1 \\ -S_1 & n \end{pmatrix}.$$

Thus the information lower bound for estimating β_2 is

$$\big(I(\beta)^{-1}\big)_{22} = \frac{\sigma^2 n}{nS_2 - S_1^2} = \frac{\sigma^2}{S_2 - S_1^2/n} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}.$$

It follows that the effective information is

$$I_{22}^*(\beta) = \frac{1}{\big(I(\beta)^{-1}\big)_{22}} = \frac{1}{\sigma^2} \left(S_2 - \frac{S_1^2}{n} \right) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \overline{x})^2.$$

Equivalently, this is the Schur complement

$$I_{22}^*(\beta) = I_{22} - I_{21} I_{11}^{-1} I_{12}.$$

(c) Since the covariance matrix of the score is the information matrix,

$$\rho = \operatorname{Corr}\big(U_1(\beta), U_2(\beta)\big) = \frac{I_{12}}{\sqrt{I_{11} I_{22}}} = \frac{S_1}{\sqrt{n S_2}}.$$

Therefore

$$\begin{aligned} I_{22}(1-\rho^2) &= \frac{S_2}{\sigma^2}\left(1 - \frac{S_1^2}{n S_2}\right) \\ &= \frac{1}{\sigma^2}\left(S_2 - \frac{S_1^2}{n}\right) = I_{22}^*. \end{aligned}$$

Thus the loss of information from not knowing the intercept is exactly controlled by the squared correlation between the two score components. There is no loss precisely when $S_1 = 0$, that is, when the covariates are centered.

(d) When both parameters are unknown, the design matrix is

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad X^\top X = \begin{pmatrix} n & S_1 \\ S_1 & S_2 \end{pmatrix}.$$

The least-squares estimator satisfies

$$\widehat{\beta} = (X^\top X)^{-1} X^\top Y,$$

and hence

$$\operatorname{Cov}(\widehat{\beta}) = \sigma^2 (X^\top X)^{-1} = \frac{\sigma^2}{n S_2 - S_1^2} \begin{pmatrix} S_2 & -S_1 \\ -S_1 & n \end{pmatrix}.$$

In particular,

$$\operatorname{Var}(\widehat{\beta}_2) = \frac{\sigma^2 n}{n S_2 - S_1^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2},$$

which is exactly the information lower bound in part (b).

If β_1 is known, subtract it from the response and fit the one-parameter model. The least-squares estimator is

$$\widetilde{\beta}_2 = \frac{\sum_{i=1}^n x_i (Y_i - \beta_1)}{\sum_{i=1}^n x_i^2}.$$

It is unbiased, and

$$\begin{aligned} \operatorname{Var}(\widetilde{\beta}_2) &= \operatorname{Var}\left(\frac{\sum_{i=1}^n x_i \epsilon_i}{S_2}\right) \\ &= \frac{\sigma^2 S_2}{S_2^2} = \frac{\sigma^2}{S_2}, \end{aligned}$$

which is the information lower bound in part (a). Thus the usual least-squares estimators attain the appropriate bounds in both cases.

2. Maximum likelihood estimation.

(a) In the Gaussian linear model

$$Y = X\beta + \epsilon, \qquad \epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n),$$

the density of Y is

$$p_\beta(y) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|y - X\beta\|_2^2\right\}.$$

Therefore the log likelihood is

$$\ell(\beta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Y - X\beta\|_2^2.$$

The first term does not depend on β, so maximizing the likelihood is equivalent to minimizing $\|Y - X\beta\|_2^2$. The score equation is

$$X^\top(Y - X\beta) = 0,$$

or

$$X^\top X\beta = X^\top Y.$$

Since $r(X) = p$, the matrix $X^\top X$ is invertible, and both the MLE and the least-squares estimator are

$$\widehat{\beta}_{\mathrm{MLE}} = \widehat{\beta}_{\mathrm{LS}} = (X^\top X)^{-1}X^\top Y.$$

(b) The log likelihood is

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{n}\left[-(X_i - \theta) - e^{-(X_i-\theta)}\right] \\ &= -\sum_{i=1}^{n} X_i + n\theta - e^{\theta}\sum_{i=1}^{n} e^{-X_i}.\end{aligned}$$

Hence

$$\ell'(\theta) = n - e^{\theta}\sum_{i=1}^{n} e^{-X_i},$$

and the likelihood equation gives

$$e^{\widehat{\theta}} = \frac{n}{\sum_{i=1}^{n} e^{-X_i}}.$$

Therefore

$$\widehat{\theta} = \log n - \log\left(\sum_{i=1}^{n} e^{-X_i}\right) = -\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{-X_i}\right).$$

Moreover,

$$\ell''(\theta) = -e^{\theta}\sum_{i=1}^{n} e^{-X_i} < 0 \qquad \text{for every } \theta,$$

so ℓ is strictly concave and has at most one maximizer. Since $\ell(\theta) \to -\infty$ as $\theta \to \pm\infty$, the critical point above is the unique global maximizer.

3. Weighted least squares, generalized ridge regression, and kernel ridge regression.

(a) Let 0_p denote the zero vector in $\mathbb{R}^p$. Define

$$Z_\Gamma = \begin{pmatrix}\Gamma^{1/2}Y \\ 0_p\end{pmatrix}, \qquad W_{\Gamma,K} = \begin{pmatrix}\Gamma^{1/2}X \\ \sqrt{\lambda}\,K^{1/2}\end{pmatrix}.$$

Then

$$\begin{aligned}\|Z_\Gamma - W_{\Gamma,K}\beta\|_2^2 &= \|\Gamma^{1/2}(Y - X\beta)\|_2^2 + \|\sqrt{\lambda}\,K^{1/2}\beta\|_2^2 \\ &= (Y - X\beta)^\top\Gamma(Y - X\beta) + \lambda\beta^\top K\beta \\ &= \|Y - X\beta\|_\Gamma^2 + \lambda\|\beta\|_K^2.\end{aligned}$$

Thus the generalized ridge problem is an ordinary least-squares problem with this augmented response and design matrix.

(b) We have

$$W_{\Gamma,K}^\top W_{\Gamma,K} = X^\top \Gamma X + \lambda K$$

and

$$W_{\Gamma,K}^\top Z_\Gamma = X^\top \Gamma Y.$$

Therefore

$$\widehat{\beta} = \left(X^\top \Gamma X + \lambda K\right)^{-1} X^\top \Gamma Y.$$

The inverse exists because, for every nonzero $v \in \mathbb{R}^p$,

$$\begin{aligned} v^\top \left(X^\top \Gamma X + \lambda K\right) v &= (Xv)^\top \Gamma (Xv) + \lambda v^\top K v \\ &= \|Xv\|_\Gamma^2 + \lambda \|v\|_K^2 > 0. \end{aligned}$$

Thus $X^\top \Gamma X + \lambda K$ is positive definite.

(c) Let $\mathbf{K}$ denote the Gram matrix with entries $\mathbf{K}_{ij} = \mathcal{K}(x_i, x_j)$.

(i) For each i,

$$f_\alpha(x_i) = \sum_{j=1}^n \alpha_j \mathcal{K}(x_i, x_j) = (\mathbf{K}\alpha)_i.$$

Hence

$$(f_\alpha)_X = \mathbf{K}\alpha.$$

Also, by the reproducing property,

$$\begin{aligned} \|f_\alpha\|_H^2 &= \left\langle \sum_{i=1}^n \alpha_i \mathcal{K}(\cdot, x_i), \sum_{j=1}^n \alpha_j \mathcal{K}(\cdot, x_j) \right\rangle_H \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathcal{K}(x_i, x_j) \\ &= \alpha^\top \mathbf{K}\alpha. \end{aligned}$$

Substitution gives the finite-dimensional problem

$$\underset{\alpha \in \mathbb{R}^n}{\text{minimize}} \left\{ (Y - \mathbf{K}\alpha)^\top \Gamma (Y - \mathbf{K}\alpha) + \lambda \alpha^\top \mathbf{K}\alpha \right\}.$$

(ii) Apply part (b) with response Y, design matrix $\mathbf{K}$, weight matrix Γ, penalty matrix $\mathbf{K}$, and parameter vector α. Since $\mathbf{K}$ is symmetric positive definite,

$$\widehat{\alpha} = (\mathbf{K}\Gamma\mathbf{K} + \lambda\mathbf{K})^{-1} \mathbf{K}\Gamma Y.$$

Factoring $\mathbf{K}$ on the left gives

$$\mathbf{K}\Gamma\mathbf{K} + \lambda\mathbf{K} = \mathbf{K}(\Gamma\mathbf{K} + \lambda I_n),$$

so

$$\begin{aligned} \widehat{\alpha} &= (\Gamma\mathbf{K} + \lambda I_n)^{-1} \Gamma Y \\ &= (\mathbf{K} + \lambda\Gamma^{-1})^{-1} Y. \end{aligned}$$

Therefore the fitted function is

$$\widehat{f}(\cdot) = \sum_{j=1}^n \widehat{\alpha}_j \mathcal{K}(\cdot, x_j), \qquad \widehat{\alpha} = (\mathbf{K} + \lambda\Gamma^{-1})^{-1} Y.$$