INDIAN STATISTICAL INSTITUTE
M.Stat. I (Year 2025-26, Semester I)

Multivariate Analysis: End-semester Examination

Date: 17-11-2025 Time: 3 hours Maximum achievable marks: 100

Q.1 Let the joint distribution of $\mathbf{X}$ be an equal mixture of $N_4(\mathbf{0}, \Sigma_1)$ and $N_4(\mathbf{0}, \Sigma_2)$, where $\Sigma_1 = \begin{bmatrix} E_{2,\rho} & 0 \\ 0 & E_{2,\rho} \end{bmatrix}$ and $\Sigma_2 = \begin{bmatrix} E_{2,\rho} & 0 \\ 0 & E_{2,-\rho} \end{bmatrix}$, and $E_{2,\rho}$ is the 2×2 equi-correlation matrix with $-1 < \rho < 1$.

(a) Find the joint pdfs of $\boldsymbol{X}_{(1)} = (X_1, X_2)^\top$, and that of $\boldsymbol{X}_{(2)} = (X_3, X_4)^\top$.

(b) Prove that X_4 and X_3 are uncorrelated but not independent.

(c) Are (X_1, X_2) and (X_3, X_4) independent? [5 + 5 + 5]

Q.2 Let $\mathbf{X}$ follow a p-dimensional elliptical distribution with density function

$$f_{\mathbf{X}}(\mathbf{x}) = \det(\Sigma)^{-1/2} g\left[(\mathbf{x} - \mu\mathbf{1})^T \Sigma^{-1} (\mathbf{x} - \mu\mathbf{1})\right],$$

where $\Sigma = \sigma^2\left[(1-\rho)I + \rho\mathbf{1}\mathbf{1}^T\right]$. Find an appropriate orthogonal transformation $\mathbf{Y} = H\mathbf{X}$, such that $\mathbf{Y}$ has the density

$$f_{\mathbf{Y}}(\mathbf{y}) \propto g\left[c_1\left(y_1 - \sqrt{p}\mu\right)^2 + c_2 \sum_{j=2}^{p} y_j^2\right],$$

where c_1, c_2 are constants. Also find the values of c_1, c_2 and the proportionality constant in $f_{\mathbf{Y}}(\cdot)$. [15]

Q.3 Let $M = X^\top X$, where $X \in \mathbb{R}^{m \times p}$ is a normal data matrix (NDM) with parameters $(\mathbf{0}, \theta I)$, $m \geq p$, $\theta > 0$, and $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$, where M_{11} is of order $k \times k$ $(k < p)$. Define $M_{11.2} = M_{11} - M_{12}M_{22}^{-1}M_{21}$.

(a) Show that the distribution $M_{12}M_{22}^{-1/2}$, $M_{11.2}$ and M_{22} are mutually independent.

(b) Define

$$\tilde{\Lambda} = \frac{\det(M)}{[\mathrm{tr}(M)]^p}.$$

Using Basu's theorem, show that $\tilde{\Lambda}$ and $\mathrm{tr}(M)$ are independently distributed.

(c) Hence or otherwise, find the $E\left(\tilde{\Lambda}\right)$. [6 + 4 + 5]

Q.4 Let $F(\lambda, \Psi; S_n) = \mathrm{trace}(\Sigma^{-1}S_n) - \log\left(\det(\Sigma^{-1}S_n)\right) - p$, where $\Sigma = \Lambda\Lambda^\top + \Psi$, $\Psi = \mathrm{Diag}(\psi_1, \ldots, \psi_p)$, $\psi_j > 0$ for all j, and Λ is a $p \times k$ matrix. Let $(\widehat{\Psi}, \widehat{\Lambda})$ denote a pair that minimizes $F(\lambda, \Psi; S_n)$ with respect to (Ψ, Λ), and let $D = \mathrm{Diag}(d_1, \ldots, d_p)$ with $d_j > 0$ for all j.

(a) Show that $(\tilde{\Psi}, \tilde{\Lambda}) = (D\Psi D^\top, D\widehat{\Lambda})$ minimizes $F(\lambda, \Psi; DS_nD^\top)$ with respect to (Ψ, Λ).

(b) Hence deduce that the of maximum likelihood estimates of the factor model are scale invariant. [7 + 8]

Q.5 Let D_n be a distance matrix (not necessarily Euclidean). Show that there exists some real number a, such that the matrix $\Delta_n = ((\delta_{i,j}))$ defined as

$$\delta_{i,j}^2 = d_{i,j}^2 - 2a, \text{ for all } i \neq j, \quad \text{and} \quad \delta_{i,i} = 0, \text{ for all } i,$$

is Euclidean. [10]

Q.6 Let $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_p)$ with real numbers $\lambda_1 > \ldots > \lambda_p$ and $\Delta = \mathrm{Diag}(\delta_1, \ldots, \delta_p)$ with non-negative numbers $\delta_1 > \ldots > \delta_p > 0$.

(a) Show that, minimizing

$$\mathrm{trace}\left(\Lambda - G\Delta G^\top\right)^2$$

over orthogonal matrices G is equivalent to maximizing $\mathrm{trace}\left(\Lambda G\Delta G^\top\right)$.

(b) Hence or otherwise show that the minimizer is $G = \mathrm{Diag}(\pm 1, \ldots, \pm 1)$. [3 + 7]

Q.7 Consider the normal linear regression model

$$\mathbf{Y}_n = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I), \text{ and } X \text{ is a non-stochastic } n \times p \text{ matrix of rank } p.$$

Consider the problem of testing $H_0 : A\boldsymbol{\beta} = \boldsymbol{c}$, where A is an $r \times p$ matrix of rank r.

(a) Find the first order solution of $(\boldsymbol{\beta}, \sigma^2)$ of the Lagrangian function under the constraint H_0. Let the solution be $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$.

(b) Next, show that, for any given $\sigma^2 > 0$, the squared error $\|\mathbf{Y} - X\boldsymbol{\beta}\|^2$ is minimized under H_0 if $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_0$.

[Hint: Towards this, you may first show that

$$\|\mathbf{Y} - X\boldsymbol{\beta}\|^2 = \|\mathbf{Y} - X\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}\|^2 + \|X(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \boldsymbol{\beta})\|^2,$$

and then show that, under H_0,

$$\|X(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \boldsymbol{\beta})\|^2 = \|X(\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})\|^2 + \|X(\widehat{\boldsymbol{\beta}}_0 - \widehat{\boldsymbol{\beta}}_{\mathrm{LS}})\|^2,$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ is the unconstrained least square of $\boldsymbol{\beta}$.]

(c) Hence, or otherwise show that, under H_0, the likelihood is maximized if $(\boldsymbol{\beta}, \sigma^2) = (\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$.

[Hint: Towards this, you may show that, under H_0

$$l(\widehat{\boldsymbol{\beta}}_0, \sigma^2) - l(\boldsymbol{\beta}, \sigma^2) \geq 0, \quad \text{and} \quad l(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2) - l(\widehat{\boldsymbol{\beta}}_0, \sigma^2) \geq 0,$$

where $l_0(\boldsymbol{\beta}, \sigma^2)$ is the log-likelihood of $(\boldsymbol{\beta}, \sigma^2)$ given $\mathbf{Y}_n$.]

(d) Find the LRT test statistic for testing H_0 and its exact distribution under H_0. [7 + 7 + 7 + (4 + 10)]