

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

Multivariate Analysis: End-semester (Backpaper) Examination

Date: ~~FBD~~ 17/12/2025

Time: 3 hours

Maximum marks: 100

Q.1 Let $\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}, \dots, \begin{bmatrix} X_n \\ Y_n \end{bmatrix}$ be a random sample from a normal distribution with parameters $\mathbf{0}$ and $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.
Define

$$R = \sum_i X_i Y_i / \sqrt{\sum_i X_i^2 \sum_i Y_i^2}$$

Find the asymptotic distribution of R . Hence show that $R \xrightarrow{p} \rho$.

[You may use the fact: Let $X \sim N(0, 1)$ then $E(X^4) = 3$.] [10 + 4]

Q.2 Let Σ be a $p \times p$ positive definite covariance matrix, and consider the problem of approximating Σ by a $p \times p$ matrix M of rank k obtained by minimizing

$$\|\Sigma - M\| = \left[\sum_{i=1}^p \sum_{j=1}^p (\sigma_{i,j} - m_{i,j})^2 \right]^{1/2}.$$

Show that

$$\|\Sigma - M\|^2 = \text{trace} \left\{ (Q - \Lambda) (Q - \Lambda)^T \right\},$$

where $\Sigma = \Gamma \Lambda \Gamma^T$ (the diagonal elements of Λ are ordered), and $Q = \Gamma^T M \Gamma$. Hence, show that the matrix M of rank k , which minimizes $\|\Sigma - M\|$ is the k -PC projection matrix. [12]

Q.3 Let C be a psd matrix. Then show that $\delta_{i,j} = c_{i,i} + c_{j,j} - 2c_{i,j} \geq 0$ for all (i, j) . Further, show that the distance matrix formed by the components $d_{i,j} = \sqrt{\delta_{i,j}}$ satisfies the triangle inequality. [4 + 6]

Q.4 Let M and W be independent Wishart variates, $M \sim W_p(\Sigma, q)$ and $W \sim W_p(\Sigma, n)$, $n > p$. Let $T = M + W$, and partition W and T as follows:

$$W = \begin{bmatrix} W_{1,1} & W_{1,2} \\ W_{1,2}^\top & W_{2,2} \end{bmatrix}, \quad \text{and} \quad T = \begin{bmatrix} T_{1,1} & T_{1,2} \\ T_{1,2}^\top & T_{2,2} \end{bmatrix},$$

where both $W_{1,1}$ and $T_{1,1}$ are $r \times r$ matrices, $r < p$. Define the Λ -statistic as

$$\Lambda = \frac{\det(W)}{\det(T)} = \frac{\det(W_{1,1})}{\det(T_{1,1})} \times \frac{\det(W_{22.1})}{\det(T_{22.1})} = \Lambda_1 \times \Lambda_2.$$

(a) Find the distributions of Λ_1 .

(b) Show that Λ_1 and Λ_2 are independently distributed.

(c) Hence, or otherwise, find the expectation of Λ_2 .

[4 + 4 + 7]

Q.5 Consider the normal linear regression model with response y and p non-stochastic covariates $\{x_1, \dots, x_p\}$. There are n observations on $\{y, x_1, \dots, x_n\}$. The i -th observation of y, y_i , is related to the i -th observation of $\{x_1, \dots, x_p\}, \mathbf{x}_i$, as

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + e_i,$$

where $\mathbf{e}_n = [e_1, \dots, e_n]^\top \sim N_n(\mathbf{0}, \sigma^2 I)$. Assume further that the $n \times p$ design matrix $X = [\mathbf{x}_1 : \dots : \mathbf{x}_n]^\top$ has rank p . Consider the problem of testing $H_0 : \mathbf{a}^\top \boldsymbol{\beta} = 0$.

- (a) Find the LR test statistic for testing H_0 .
- (b) Find the exact distribution of the LR test statistic obtained in part (a).
- (c) Next suppose $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1 : \boldsymbol{\beta}'_2]$, where $\boldsymbol{\beta}_1 \in \mathbb{R}^q$ with $q < p$. Find the LR test statistic for testing $H_N : \boldsymbol{\beta}_2 = \mathbf{0}$.
- (d) Find the exact distribution of the LR test statistic obtained in part (c). [4 + 4 + 4 + 3]

Q.6 Consider a sample of size N_i from a $N_m(\mu_i, \Sigma_i)$ distribution ($i = 1, 2, 3$).

- (a) Find the likelihood ratio (LR) statistic (say, Λ_0) for testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$, given $\Sigma_1 = \Sigma_2$.
- (b) Also find the LR test statistics Λ_1 for testing $H_1 : \Sigma_1 = \Sigma_2$. [5 + 5]

Q.7 Let $\mathbf{X}_i = \boldsymbol{\mu} + \mathbf{U}_i + \Lambda \mathbf{f}_i$, $i = 1, \dots, n$, and $\{(\mathbf{f}_i, \mathbf{U}_i), i = 1, \dots, n\}$ are iid samples from a normal distribution with mean $\mathbf{0}$ and variance $\text{diag}(I, \Psi)$.

- (a) Let $F_n = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ and $X_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ be observed. Find the likelihood of $(\boldsymbol{\mu}, \Lambda, \Psi)$ given $\{X_n, F_n\}$.
- (b) Show that, when F_n is included in the data, then the sufficient statistics is $\{\bar{\mathbf{X}}_n, \bar{\mathbf{f}}_n, S_n, S_{x,f} = \text{cov}(X_n, F_n), S_f = \text{var}(F_n)\}$, $S_{x,f}$ and S_f indicate the sample covariance of X_n and F_n , and sample variance of F_n .
- (c) Show that the conditional expectations of the covariances in (b) given $\Theta = \{X_n, \Lambda, \Psi\}$ are

$$E(S_n | \Theta) = S_n, \quad S_{x,f}^* = E(S_{x,f} | \Theta) = S_n B_n, \quad \text{and} \quad S_f^* = E(S_f | \Theta) = B_n^\top S_n B_n + I - \Lambda^\top B_n,$$

$$\text{where } B_n = (\Psi + \Lambda \Lambda')^{-1} \Lambda.$$

- (d) Finally, show that the MLEs of Λ and Ψ are

$$\hat{\Lambda}_{\text{ML}} = S_{x,f}^* S_{f,f}^{*-1}, \quad \hat{\Psi}_{\text{ML}} = S_n - S_{x,f}^* S_f^{*-1} S_{x,f}^{*\top}.$$

[6 + 6 + 6 + 6]