Mid-Sem Exam: MTL766

Semester I (Academic year 2024-25)

Name: .   Entry number:

Time: 2 Hours

- Read instructions and questions carefully.
- Maximum you can score= 25   score = min (25, score)
- Questions 1-5 carry 2 marks each. For the remaining questions, the marks are indicated in the right margin, inside boxes next to each question.
- You can use results proved in the class directly.

Questions 1-5: State true or false and justify your answer

Question 1. For any $\boldsymbol{X} \in \mathbb{R}^p$, the Mahalanobis distance $M_D(\boldsymbol{X}, \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)) = (\boldsymbol{X} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu})$ is non-negative.

Question 2. For any fixed $\boldsymbol{d} \in \mathbb{R}^p$, the distribution of Mahalanobis distance $M_D(\boldsymbol{d}, \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)) = (\boldsymbol{d} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{d} - \boldsymbol{\mu})$ has a non-central chi-square distribution.

Question 3. Let X and Y be two univariate random variables. If $aX + bY$ follows a normal distribution for all fixed $a, b \in \mathbb{R}$ then joint distribution of (X, Y) is bivariate normal.

Question 4. Let $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ and $g : \mathbb{R}^3 \to \mathbb{R}$ be a continuous function. Then the distribution of $g(\boldsymbol{X})$ is univariate normal.

Question 5A Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$; $\overline{\boldsymbol{X}} = \sum_{i=1}^n \boldsymbol{X}_i / n$. Then $\overline{\boldsymbol{X}}$ and $\boldsymbol{X}_1 - \overline{\boldsymbol{X}}$ are independently distributed.

Descriptive solution questions

Question 5B Let $\boldsymbol{A} \sim W_p(n, \Sigma)$. Then using the definition of Wishart distribution or otherwise, prove or disprove $\mathbb{E}(\boldsymbol{A}) = n\Sigma$. $\boxed{3}$

Question 6. Explain the steps of detecting outliers, assuming the data follows a multivariate normal distribution. $\boxed{3}$

Question 7. Three observations from $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$ are $\{(1,2), (1.2, 1.8), (2,3)\}$. Find a 95% minimum volume confidence region for $\boldsymbol{\mu}$. Critical values of the statistics can be found in the table provided at the end of the book. $\boxed{3}$

Question 8. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$; $\overline{\boldsymbol{X}} = \sum_{i=1}^{n} \boldsymbol{X}_i/n$ and $\boldsymbol{S} = \sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})^\top/(n-1)$. Then prove or disprove that the distribution of $(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$ converges to a chi-square distribution as $n \to \infty$. $\boxed{3}$

Question 9. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from $\mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$, where $\Sigma > 0$ and $\boldsymbol{\mu}$ are unknown. For testing $H_0 : \boldsymbol{\mu} = (1,2,3)^\top$ against $H_A :$ H_0 is not true, construct a test based on Hotelling's T^2 statistic and write its distribution under H_0. $\boxed{3+1}$

Question 10. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from $\mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$, where $\Sigma > 0$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top$ are unknown . For testing $H_0 : \mu_1 = \mu_2 - \mu_3$ against $H_A :$ H_0 is not true, construct a test based on Hotelling's T^2 statistic and write its distribution under H_0. $\boxed{3+1}$