A collection of problems for STK4040/9040

This collection gives theoretical problems that it may be useful to work through when preparing for the written exam in STK4040/9040.

Problem 1

Let $\mathbf{X}$ be a p-variate random vector with mean vector $E(\mathbf{X}) = \boldsymbol{\mu}$ and covariance matrix

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\} \tag{1}$$

a) Let $\mathbf{a}$ be a p-dimensional vector. Use (1) to find an expression for the variance of $Y = \mathbf{a}'\mathbf{X}$.

b) Show that the covariance matrix $\boldsymbol{\Sigma}$ is non-negative definite.

Problem 2

Let $\boldsymbol{\Sigma}$ be a positive definite $p \times p$ matrix, and let λ be an eigenvalue of $\boldsymbol{\Sigma}$ with corresponding eigenvector $\mathbf{e}$.

a) Show that $\lambda > 0$.

b) Show that $1/\lambda$ is an eigenvalue of $\boldsymbol{\Sigma}^{-1}$ and determine the corresponding eigenvector.

Problem 3

Let $\boldsymbol{\Sigma}$ be a positive definite $p \times p$ matrix, and let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of $\boldsymbol{\Sigma}$ with corresponding orthogonal and normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Introduce the matrix $\mathbf{P}$ with the eigenvectors as columns, $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p]$, and let $\boldsymbol{\Lambda}$ be the diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ on the diagonal.

a) Show that $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$. (*Hint:* Use the spectral decomposition of $\boldsymbol{\Sigma}$.)

b) Show that $|\boldsymbol{\Sigma}| = \prod_{i=1}^{n} \lambda_i$.

Let $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'$ be the square root matrix.

c) Show that $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$.

d) Define $\boldsymbol{\Sigma}^{-1/2} = \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}'$ and show that $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} = \mathbf{I}$.

Problem 4

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. p-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

a) Show that $\overline{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_j$ is an unbiased estimator of $\boldsymbol{\mu}$.

b) Show that $\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$ is an unbiased estimator of $\boldsymbol{\Sigma}$.

Assume now that the $\mathbf{X}_j$'s are $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed.

c) Give the distributions of $\overline{\mathbf{X}}$ and $(n-1)\mathbf{S}$.

d) Derive the distribution of $n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})$.

e) Describe how the result in d) may be used to obtain confidence regions for $\boldsymbol{\mu}$.

Problem 5

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. and $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random vectors.

a) Show that the likelihood takes the form

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}\,|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_j - \boldsymbol{\mu})\right\}.$$

b) Show that

$$\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_j - \boldsymbol{\mu})$$
$$= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'\right)\right] + n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}).$$

c) Derive the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
Are these estimators unbiased?

Problem 6

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed. We want to test

$$\text{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \qquad \text{versus} \qquad \text{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

where $\boldsymbol{\mu}_0$ is a known vector.

a) Explain why it is reasonable to reject H_0 for large values of the test statistic

$$T^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0). \tag{2}$$

b) Explain why the statistic (2), under H_0, is distributed as a constant times a F-distributed random variable. What is the constant, and what are the degrees of freedom of the F-distribution?

Problem 7

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. p-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

a) Let $\mathbf{a}$ be a p-dimensional vector. Explain why

$$T_{\mathbf{a}} = \frac{\sqrt{n}(\mathbf{a}'\overline{\mathbf{X}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}$$

follows a t-distribution with $n-1$ degrees of freedom.

The result in a) may be used to determine a confidence interval for $\mathbf{a}'\boldsymbol{\mu}$ for a given vector $\mathbf{a}$. In order to derive a confidence interval that is valid simultaneously for all choices of $\mathbf{a}$, we make use of the following result

$$\max_{\mathbf{a}} T_{\mathbf{a}}^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \tag{3}$$

b) Prove (3) and use the result to derive a confidence interval for $\mathbf{a}'\boldsymbol{\mu}$ that is valid simultaneously for all choices of $\mathbf{a}$.

c) Give the confidence interval in b) for the special case that $\mathbf{a} = (1, 0, 0, \ldots, 0)'$.

Problem 8

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. and $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random vectors. We will consider the likelihood ratio test for testing

$$\mathrm{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \qquad \text{versus} \qquad \mathrm{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

where $\boldsymbol{\mu}_0$ is a known vector. We denote the likelihood by $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; cf. Problem 5.

a) Explain that

$$\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} \, |\widehat{\boldsymbol{\Sigma}}|^{n/2}} e^{-np/2},$$

where $\widehat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimator of $\boldsymbol{\Sigma}$.

b) Derive an expression for the maximum likelihood estimator $\widehat{\boldsymbol{\Sigma}}_0$ of $\boldsymbol{\Sigma}$ under the null hypothesis, and show that the likelihood ratio takes the form

$$\Lambda = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_0|} \right)^{n/2}.$$

c) Describe the relation between the likelihood ratio test and the test discussed in Problem 6.

Problem 9

Let $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$ be $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$-distributed, and let $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$ be $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$-distributed. Further assume that all the variables are independent.

a) Describe how one may obtain an estimator $\mathbf{S}_{\text{pooled}}$ for $\boldsymbol{\Sigma}$ by combining the sample covariance matrices from the two samples.

b) Obtain the distribution of

$$\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \mathbf{S}_{\text{pooled}}\right]^{-1} \left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)$$

c) Use the result of b) to obtain a test for

$$\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \qquad \text{versus} \qquad \mathrm{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

d) Derive a confidence interval for $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ that is valid simultaneously for all choices of $\mathbf{a}$.

Problem 10

Let $\mathbf{X}$ be a p-variate random vector with positive definite covariance matrix $\boldsymbol{\Sigma}$, and let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of $\boldsymbol{\Sigma}$ with corresponding orthogonal and normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$. The first population principal component is the linear combination $\mathbf{a}'\mathbf{X}$ that maximizes $\mathrm{Var}(\mathbf{a}'\mathbf{X})$ subject to $\mathbf{a}'\mathbf{a} = 1$

a) Show that the first population principal component is given by $Y_1 = \mathbf{e}_1'\mathbf{X}$.

The i-th population principal component is given by $Y_i = \mathbf{e}_i'\mathbf{X}$, $i = 1, \dots, p$.

b) Give a description of the properties of the second and third population principal components.

The total variance of $\mathbf{X}$ is given by $\mathrm{tr}(\boldsymbol{\Sigma})$.

c) Show that $\mathrm{tr}(\boldsymbol{\Sigma}) = \sum_{i=1}^{p} \lambda_i$ (*Hint:* Use the result in Problem 3a.)

d) Explain that $\sum_{i=1}^{q} \lambda_i / \sum_{i=1}^{p} \lambda_i$ is an expression for proportion of the total variance explained by the first q principal components.

Problem 11

Let $\mathbf{X}$ be a p-variate random vector with mean vector $\boldsymbol{\mu}$. The orthogonal factor model assumes that we may write

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}. \tag{4}$$

Here the unobservable vectors $\mathbf{F}$ and $\boldsymbol{\epsilon}$ are independent and satisfy

$$E(\mathbf{F}) = \mathbf{0} \qquad \text{Cov}(\mathbf{F}) = \mathbf{I}, \tag{5}$$

and

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \qquad \text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \text{diag}\{\psi_1, \psi_2, \ldots, \psi_p\}. \tag{6}$$

a) Show that the covariance matrix of $\mathbf{X}$ may be given as $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$.

b) Give an interpretation of the factor model (4) and explain what we mean by communalities and specific variances.

c) The factor model is not uniquely determined by the equations (4)–(6). Discuss the nature of this nonuniqueness.

c) Discuss briefly two method that may be used for estimation of the factor model.

Problem 12

We have two populations, denoted π_1 and π_2 with prior probabilities p_1 and p_2, respectively. If a random vector $\mathbf{X}$ is selected from π_1 it has density $f_1(\mathbf{x})$, while it has density $f_2(\mathbf{x})$ if it is selected from π_2. A classification rule assign an observation to population π_1 provided that $\mathbf{X} \in R_1$. If $\mathbf{X} \in R_2$, the observation is classified to population π_2. We assume that R_1 and R_2 are disjoint and that their union is the sample space of $\mathbf{X}$.

a) Explain that the total probability of misclassification may be given as

$$TPM = p_1 \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x}) d\mathbf{x}.$$

b) Show that we may write

$$TPM = \int_{R_1} \left[p_2 f_2(\mathbf{x}) - p_1 f_1(\mathbf{x}) \right] d\mathbf{x} + p_1.$$

c) Show that the total probability of misclassification is minimized if we choose

$$R_1 = \left\{ \mathbf{x} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \right\}. \tag{7}$$

d) Derive the classification rule (7) when $f_i(\mathbf{x})$ is the multivariate normal density with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}$, $i = 1, 2$.