TMA4270 Multivariate Analysis H2008
Selected Theoretical Exercises

Chapters 1, 2, 3: Introduction to multivariate analysis

Problem 1: Matrix operations as sums
Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be p-dimensional vectors, and $\boldsymbol{C}$ a $p \times p$ matrix.

a) What is the dimension of $\boldsymbol{a}^T\boldsymbol{b}$? Write $\boldsymbol{a}^T\boldsymbol{b}$ using sums.

b) What is the dimension of $\boldsymbol{a}^T\boldsymbol{a}$? Write $\boldsymbol{a}^T\boldsymbol{a}$ using sums.

c) What is the dimension of $\boldsymbol{a}\boldsymbol{b}^T$? Write a general element of $\boldsymbol{a}\boldsymbol{b}^T$ using sums.

d) What is the dimension of $\boldsymbol{a}^T\boldsymbol{C}$ and $\boldsymbol{C}\boldsymbol{b}$? Write the general element of these two matrices using sums.

e) What is the dimension of $\boldsymbol{a}^T\boldsymbol{C}\boldsymbol{b}$? Write $\boldsymbol{a}^T\boldsymbol{C}\boldsymbol{b}$ using sums.

Problem 2: Bivariate Pareto distribution
The bivariate Pareto distribution is defined by the joint probability density of (X, Y)

$$f(x, y) = \frac{c}{(x + y - 1)^{a+2}} \text{ for } x > 1, y > 1$$

where $a > 0$ is a parameter.

a) Show that $c = a(a + 1)$.

b) Find the marginal densities for X and Y.

c) Show that (for $a > 1$)

$$\mathrm{E}(X) = \mathrm{E}(Y) = \frac{a}{a - 1}$$

d) Show that (for $a > 2$) the covariance matrix is given as

$$\frac{1}{(a-1)^2(a-2)} \begin{bmatrix} a & 1 \\ 1 & a \end{bmatrix}$$

e) Find the generalized variance (determinant of covariance matrix) and the total variance (trace of covariance matrix).

Problem 3: The Square Root Matrix
Let the expectation (mean) and covariance matrix for a p-variate random vector $\boldsymbol{X}$ be $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X})$.

Let further $(\lambda_i, \boldsymbol{e}_i)$, $i = 1, ..., p$ be the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$. Let $\boldsymbol{P}$ be the matrix of eigenvector,

$$\boldsymbol{P} = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_p]$$

and $\boldsymbol{\Lambda}$ be a diagonal matrix with the eigenvalues $\lambda_1, \lambda_1, ..., \lambda_p$ on the diagonal.

a) Define the matrices $\boldsymbol{\Sigma}^{\frac{1}{2}}$ and $\boldsymbol{\Sigma}^{\frac{1}{2}}$ by

$$\begin{aligned} \boldsymbol{\Sigma}^{\frac{1}{2}} &= \boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{P}^T \\ \boldsymbol{\Sigma}^{-\frac{1}{2}} &= \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T \end{aligned}$$

Show that both matrices are symmetric and that the following is true:

$$\begin{aligned} \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} &= \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}} &= \boldsymbol{\Sigma}^{-1} \\ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}} &= \boldsymbol{I} \end{aligned}$$

where $\boldsymbol{I}$ is the identity matrix.

b) The transform

$$\boldsymbol{Y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$$

is called the Mahalanobis transform. Show that $\mathrm{E}(\boldsymbol{Y}) = 0$ and $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{I}$.

Problem 4: Matrix Calculations and Partitions

These tasks are to be done by hand (but use R to check the answers if you want).

Suppose

$$\boldsymbol{A} = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 5 & 2 \\ 2 & 2 & 4 \end{bmatrix}$$

Partition $\boldsymbol{A}$ into

$$\boldsymbol{A}_{11} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix} \quad \boldsymbol{A}_{12} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \boldsymbol{A}_{21} = \begin{bmatrix} 2 & 2 \end{bmatrix} \quad \boldsymbol{A}_{22} = \begin{bmatrix} 4 \end{bmatrix}$$

a) Evaluate the eigenvalues and corresponding unit length eigenvectors of $\boldsymbol{A}_{11}$.

b) Evaluate the symmetric square root $\boldsymbol{A}_{11}^{\frac{1}{2}}$.

c) Find $\boldsymbol{A}_{11.2}$ and evaluate $\boldsymbol{A}_{11.2}^{-1}$. Hence find the partitions $\boldsymbol{A}^{11}$, $\boldsymbol{A}^{12}$, $\boldsymbol{A}^{21}$, and $\boldsymbol{A}^{22}$ of $\boldsymbol{A}^{-1}$. Check your answer by showing $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$.

d) Check that the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 8$, $\lambda_2 = 4$, and $\lambda_3 = 2$.

e) Check that $tr\boldsymbol{A} = \sum_{\forall i} \lambda_i$.

f) Check that $|\boldsymbol{A}| = |\boldsymbol{A}_{11.2}||\boldsymbol{A}_{22}| = |\boldsymbol{A}_{22.1}||\boldsymbol{A}_{11}| = \prod_{\forall i} \lambda_i$.

Problem 5: **Joint, Marginal and Conditional Densities**
Two random variables X and Y have joint density

$$f(x,y) = \begin{cases} kx, & 0 < y < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the marginal densities of X and Y and the conditional densities of $X|Y$ and $Y|X$. Hence deduce that the two random variables are not independent. Explain how his conclusion can hold, given that the pdf appearently is a product of a function of X and a function of Y.

Problem 6: **Covariance and Correlation of Linear Combinations**
Suppose X_1, X_2, X_3 are three independent random variables each with variance 1. Construct new random variables $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1 - X_2$ and $Y_3 = X_1 - X_3$. Find the covariance matrix and correlation matrix of $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^T$.

Problem 7: **Covariance**
Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be random vectors with dimensions respectively p and q. Let further $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\boldsymbol{\eta} = \mathrm{E}(\boldsymbol{Y})$. We define

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) =_{\text{def}} \mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\eta})']$$

a) What is the dimension of $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$?

What is the interpretation of entry (i,j) in this matrix?

What is $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})$?

b) Prove the formula

$$\mathrm{Cov}(\boldsymbol{CX}, \boldsymbol{DY}) = \boldsymbol{C}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{D}'$$

c) Let $\boldsymbol{c}$ and $\boldsymbol{d}$ be vectors of dimension p and let $\boldsymbol{C}$ be a $q \times p$ matrix. Use the formula in (b) to show

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{c}'\boldsymbol{X}, \boldsymbol{d}'\boldsymbol{X}) &= \boldsymbol{c}'\mathrm{Cov}(\boldsymbol{X})\boldsymbol{d} \\ \mathrm{Cov}(\boldsymbol{CX}) &= \boldsymbol{C}\mathrm{Cov}(\boldsymbol{X})\boldsymbol{C}' \end{aligned}$$

d) Define a $(p+q)$-dimensional random vector by letting the p first coordinates be $\boldsymbol{X}$ and the q last be $\boldsymbol{Y}$, i.e. consider

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix}$$

What is $\mathrm{E}(\boldsymbol{Z})$?

Prove that

$$\mathrm{Cov}(\boldsymbol{Z}) = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{11} &= \mathrm{Cov}(\boldsymbol{X}) \\ \boldsymbol{\Sigma}_{22} &= \mathrm{Cov}(\boldsymbol{Y}) \\ \boldsymbol{\Sigma}_{12} &= \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) \\ \boldsymbol{\Sigma}_{21} &= \boldsymbol{\Sigma}_{12}' = \mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{X}) \end{aligned}$$

Problem 8: Empirical distribution
Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be independent and identically distributed observations of the random (p-dimensional) vector $\boldsymbol{X}$.

The *empirical distribution* of $\boldsymbol{X}$ based on the n observations is defined to be the discrete distribution which gives probability $1/n$ to each of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

Let E_* and Cov_* mean expectation and covariance in the *empirical* distribution.

a) Show that if $\boldsymbol{X}^*$ is a random vector with the empirical probability distribution, then

$$\mathrm{E}_*(\boldsymbol{X}^*) = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{x}_j \equiv \bar{\boldsymbol{x}}$$

$$\mathrm{Cov}_*(\boldsymbol{X}^*) = \frac{1}{n}\sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})' \equiv \boldsymbol{S}_n$$

What the book calls *sample-mean* and *sample covariance matrix* are thus the usual expectation and covariance matrix of the *empirical* distribution.

b) Assume that we are interested in a linear function $\boldsymbol{c}'\boldsymbol{X}$. Our n observations then give a sample of n values $Y_1 = \boldsymbol{c}'\boldsymbol{x}_1, \ldots, Y_n = \boldsymbol{c}'\boldsymbol{x}_n$).

Show by direct computation that

$$\bar{Y} = \boldsymbol{c}'\bar{\boldsymbol{x}}$$

$$\frac{1}{n}\sum_{j=1}^{n} (Y_j - \bar{Y})^2 = \boldsymbol{c}'\boldsymbol{S}_n\boldsymbol{c}$$

Discuss in light of (a) how this also follows directly from the corresponding population-formulae

$$\mathrm{E}(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'\boldsymbol{\mu}$$

$$\mathrm{Var}(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are expectation and covariance matrix of $\boldsymbol{X}$ with respect to the *true* (but as a rule unknown) distribution for $\boldsymbol{X}$.

Chapter 4: Multivariate Normal Distribution

Exams: December 2003 # 1, December 2004 # 1, December 2005 #1.

Problem 9: **Mahalanobis distance**
Let $\boldsymbol{X}$ be a random vector with dimension p. Let further $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X})$.

The squared *Mahalanobis-distance* from $\boldsymbol{x}$ to $\boldsymbol{\mu}$ is defined as

$$(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

Write down an expression (not the matrix expression since that is already given) for the squared Mahalanobis-distance when

a) $p = 1$

b) $\boldsymbol{\Sigma}$ is a diagonal matrix

c) $p = 2$. Find out how the squared Mahalanobis-distance depends on the correlation coefficient ρ_{12} in this case. (Example 4.1 on page 151 in Johnsen & Wichern is relevant.)

Problem 10: **Bivariate Normal and Correlation**
Consider the bivariate random vector

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

Expand the matrix form of the density function to get the usual bivariate normal density involving σ_1, σ_2, ρ, and exponential terms in $(x_1 - \mu_1)^2$, $(x_1 - \mu_1)(x_2 - \mu_2)$ and $(x_2 - \mu_2)^2$. What happens for:

a) $\rho = 0$.

b) $\rho \to 1$.

c) $\rho \to -1$.

Problem 11: **MGF of Multivariate Normal**
Given that the density function of a $p \times 1$ normal random vector $\boldsymbol{X}$ is

$$f(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}, \quad \boldsymbol{x} \in R^p$$

show that the moment generating function (mgf) of $\boldsymbol{X}$ is

$$\mathrm{E}\left\{ \exp(\boldsymbol{t}^T \boldsymbol{x}) \right\} = \exp\left\{ \boldsymbol{t}^T \boldsymbol{\mu} + \frac{1}{2} \boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t} \right\}, \quad \boldsymbol{t} \in R^p.$$

Hint: Make a linear transformation to $\boldsymbol{y} = \boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{x} - \boldsymbol{\mu})$ and set $\boldsymbol{s} = \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{t}$.

Problem 12: Independence

a) Let

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

Show that if $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$, then

- $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are stochastically independent.
- $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are each multinormally distributed.

(Hint: This is essentially Exercise 4.14 in Johnson & Wichern.)

b) Show that if you only know that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are each separately multivariate normally distributed, and that they are stochastically independent, then the vector $\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$ is multivariate normally distributed.

Problem 13: Independence of $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let as usual

$$\begin{aligned} \bar{\boldsymbol{X}} &= \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{X}_j \\ \boldsymbol{S} &= \frac{1}{n-1}\sum_{j=1}^{n} (\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})^T \end{aligned}$$

It is known from the lectures that $\bar{\boldsymbol{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.

In this exercise we shall prove that $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are stochastically independent, and that

$$(n-1)\boldsymbol{S} \sim W_{n-1}(\boldsymbol{\Sigma})$$

(where $W_m(\boldsymbol{\Sigma})$ means the Wishart-distribution with m degrees of freedom and parameter $\boldsymbol{\Sigma}$).

As an aid we shall need an orthonormal $n \times n$-matrix $\boldsymbol{T} = (t_{jk})$ where the bottom line is

$$(t_{n1}, t_{n2}, \ldots, t_{nn}) \equiv (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$$

Remember that $\boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}^T = \boldsymbol{I}$, i.e. the Euclidean length of each line and each column is 1, while the inner product of two rows or of two columns are always 0.

a) Why can we find an orthogonal $\boldsymbol{T}$ with the specified bottom line? Are there more than one such $\boldsymbol{T}$?

b) Define, for $j = 1, 2, \ldots, n$,

$$\boldsymbol{Z}_j = \sum_{k=1}^{n} t_{jk}\boldsymbol{X}_k$$

Show that $\boldsymbol{Z}_n = \sqrt{n}\bar{\boldsymbol{X}}$.

Show that

$$\begin{aligned} \mathrm{E}(\boldsymbol{Z}_j) &= \boldsymbol{0} \qquad \text{for } j = 1, \ldots, n-1 \\ \mathrm{E}(\boldsymbol{Z}_n) &= \sqrt{n}\boldsymbol{\mu} \end{aligned}$$

c) Show that

$$\begin{aligned} \text{Cov}(\boldsymbol{Z}_j, \boldsymbol{Z}_r) &= \mathbf{0} \quad \text{if } j \neq r \\ \text{Cov}(\boldsymbol{Z}_j) &= \boldsymbol{\Sigma} \quad \text{for } j = 1, \ldots, n \end{aligned}$$

(See e.g. the Theoretical Exercise named Covariance for definition of $\text{Cov}(\boldsymbol{X}, \boldsymbol{Y})$.)

d) Show that

$$\sum_{j=1}^{n} \boldsymbol{Z}_j \boldsymbol{Z}_j' = \sum_{j=1}^{n} \boldsymbol{X}_j \boldsymbol{X}_j'$$

Derive from this that

$$\begin{aligned} (n-1)\boldsymbol{S} &= \sum_{j=1}^{n} \boldsymbol{X}_j \boldsymbol{X}_j' - n \bar{\boldsymbol{X}} \bar{\boldsymbol{X}}' \\ &= \sum_{j=1}^{n-1} \boldsymbol{Z}_j \boldsymbol{Z}_j' \end{aligned}$$

e) Explain how the desired results on independence and Wishart-distribution (see beginning of Problem) now will follow if we can show that $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent and multivariate normal (this is to be shown in the next question).

f) Explain first why the np-dimensional random vector

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix}$$

is multivariate normally distributed.

Verify then that we for some matrix $\boldsymbol{A}$ defined from $\boldsymbol{T}$ can write

$$\begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \\ \vdots \\ \boldsymbol{Z}_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix}$$

Finally, explain why this implies that $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent and multivariate normally distributed

Chapter 5, 6: Inference on mean or means

Exams: August 1998 #4, November 1999 #2, December 2003 #2, December 2005 #2ab.

Problem 14: **Hypothesis testing using contrasts**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let further $\boldsymbol{A}$ be a $q \times p$-matrix with $q \leq p$ and $\text{rank}(A) = q$ (i.e. the rows of $\boldsymbol{A}$ are linearly independent).

We shall test

$$H_0: \boldsymbol{A\mu} = \mathbf{0} \text{ vs. } H_1: \boldsymbol{A\mu} \neq \mathbf{0}$$

a) Explain why it is reasonable to reject when the statistic

$$n(\boldsymbol{A}\bar{\boldsymbol{X}})'(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}')^{-1}(\boldsymbol{A}\bar{\boldsymbol{X}}) \tag{1}$$

is bigger than some constant.

Show that this statistic, under H_0, is distributed as a constant times a Fisher-distribution. What is the constant, and what are the degrees of freedom of the Fisher-distribution?

(*Hint:* Consider the usual Hotelling's T^2 for the transformed variables

$$\boldsymbol{Y}_j = \boldsymbol{A}\boldsymbol{X}_j$$

where $j = 1, 2, \ldots, n$).

b) As an application of the above, assume that we shall test

$$H_0: \ \mu_1 = \mu_2 = \cdots = \mu_p \text{ vs. } H_1: \ \text{ at least one} \neq$$

Show that this can be done by choosing $\boldsymbol{A}$ to be the $(p-1) \times p$-matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & & \ddots & & \\ 1 & 0 & & \cdots & -1 \end{pmatrix}$$

May other matrices $\boldsymbol{A}$ be used as well? (In this connection, consider pages 278-280 in Johnsen & Wichern "A Repeated-Measures Design for Comparing Treatments". Note that q has been used there instead of p).

c) The matrix $\boldsymbol{A}$ in the last point is an example of a so called *contrast matrix*. This is per definition a $(p-1) \times p$-matrix with linearly independent columns (i.e. $\text{rank}(A) = p - 1$) and with all row sums equal to 0 (i.e. $\boldsymbol{A}\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is a vector of 1's).

Prove that the test statistic in (1) gives the same value for all $(p-1) \times p$ contrast matrices $\boldsymbol{A}$. (See also footnote on p. 279 in Johnsen & Wichern.)

Problem 15: CLT and Wilks lambda [Chapter 5.3, NOT in the curriculum]

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

$$H_0: \ \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ mot } H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

The *likelihood ratio* test does this based on the likelihoodfunction $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ from paragraph 5.3 (not in curriculum).

The likelihood ratio test rejects H_0 for small values of (see e.g. (5-12))

$$\begin{aligned} \Lambda &= \frac{\text{maximum of L when } H_0 \text{ holds}}{\text{maximum of L in the full model}} \\ &= \frac{\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \\ &= \frac{L(\boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}_0)}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})} \end{aligned}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the usual MLE, i.e. respectively $\bar{\boldsymbol{X}}$ and $\frac{n-1}{n}\boldsymbol{S}$, while $\hat{\boldsymbol{\Sigma}}_0$ is MLE for $\boldsymbol{\Sigma}$ when it is known that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

a) Find $\hat{\mathbf{\Sigma}}_0$. (See Chapter 5.3 in Johnsen & Wichern.)

b) Show that

$$\Lambda = \left(\frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_0|} \right)^{n/2}$$

Hint: see (5-12).

c) Explain why Hotelling's test is equivalent to rejecting H_0 when $\Lambda^{2/n} <$ konstant, where $\Lambda^{2/n} = \frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_0|}$ is called *Wilk's lambda.*

Chapter 7: Linear Regression

Exams: January 1994 # 1+3, December 2004 # 2.

Problem 16: **Weighted least squares**

a) Start by solving Problem (exercise) 7.3 on page 417 of Johnsen & Wichern.

b) Then, show that in this situation the maximum likelihood estimator for $\boldsymbol{\beta}$ is found by minimizing

$$(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}) \tag{2}$$

c) Also show that the given $\hat{\boldsymbol{\beta}}_W$ is the unique minimum of (2).

d) Solve Problem (exercise) 7.4 on page 417 of Johnsen & Wichern.

Problem 17: **Likelihood Ratio Tests for the Regression Parameters**
Let the situation be as in the paragraph *Likelihood Ratio Tests for the Regression Parameters* p. 370-372 in Johnsen & Wichern. Remark that one there assumes that $\boldsymbol{\varepsilon}$ is *multinormal.*

Introduce in addition

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T \\ \boldsymbol{H}_1 &= \boldsymbol{Z}_1(\boldsymbol{Z}_1^T\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1^T \end{aligned}$$

and

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}} &= \boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\varepsilon}}_1 &= \boldsymbol{Y} - \boldsymbol{Z}_1\hat{\boldsymbol{\beta}}_1 \end{aligned}$$

We can now write

$$\begin{aligned} F &= \frac{(SS_{res}(\boldsymbol{Z}_1) - SS_{res}(\boldsymbol{Z})))/(r-q)}{s^2} \\ &= \frac{(\hat{\boldsymbol{\varepsilon}}_1^T\hat{\boldsymbol{\varepsilon}}_1 - \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}})/(r-q)}{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/(n-r-1)} \end{aligned}$$

a) Show that we have

$$\begin{aligned} \hat{\varepsilon} &= (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon} \\ \hat{\varepsilon}_1 &= (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{\varepsilon} \end{aligned}$$

b) Show that

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} &= \boldsymbol{\varepsilon}^T (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon} = ((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon})^T ((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon}) \\ \hat{\boldsymbol{\varepsilon}}_1^T \hat{\boldsymbol{\varepsilon}}_1 - \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} &= \boldsymbol{\varepsilon}^T (\boldsymbol{H} - \boldsymbol{H}_1)\boldsymbol{\varepsilon} = ((\boldsymbol{H} - \boldsymbol{H}_1)\boldsymbol{\varepsilon})^T ((\boldsymbol{H} - \boldsymbol{H}_1)\boldsymbol{\varepsilon}) \end{aligned}$$

(*Hint:* Note that $\boldsymbol{H} - \boldsymbol{H}_1$ is idempotent.)

c) In Johnsen & Wichern it is shown that

$$\frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{\sigma^2} \sim \chi^2_{n-r-1}$$

Now show that

$$\frac{\hat{\boldsymbol{\varepsilon}}_1^T \hat{\boldsymbol{\varepsilon}}_1 - \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{\sigma^2} \sim \chi^2_{r-q}$$

(*Hint:* Again use that $\boldsymbol{H} - \boldsymbol{H}_1$ is idempotent and show that it has trace equal to $r - q$.)

d) In order to show that F is Fisher-distributed it remains to prove that the numerator and denominator of F are stochastically independent. This will follow if we can prove that $(\boldsymbol{H} - \boldsymbol{H}_1)\boldsymbol{\varepsilon}$ and $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\varepsilon}$ are stochastically independent (why?).

Show that they *are* indeed independent.

(*Hint:* Show that $(\boldsymbol{H} - \boldsymbol{H}_1)(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$).

Chapter 8: Principal Components

Exams: January 1994 # 2, August 1999 # 3

Chapter 9: Factor Analysis

Exams: December 2002 # 2+3, December 2005 # 3.

Chapter 11: Discrimination and Classification

Exams: December 1995 # 3, November 1996 # 1, December 2004 # 3, December 2005 #2c.

Problem 18: **Fisher**
The problem is to classify into one of two multinormal distributions, $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ og $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$.

Assume that the p variables in $\boldsymbol{X}$ are *equicorrelated* with the same variance, i.e. the covariance matrix has the form

$$\boldsymbol{\Sigma} = (1-\rho)\sigma^2\boldsymbol{I} + \rho\sigma^2\mathbf{1}\mathbf{1}^T$$

where $\boldsymbol{I}$ is the identity matrix and $\mathbf{1}$ is a vector of 1's.

Show that Fisher's linear discriminant function applied to an observation $\boldsymbol{X}$ is proportional to

$$\boldsymbol{\delta}^T\boldsymbol{X} - \frac{p^2\rho}{1+(p-1)\rho}\bar{\delta}\bar{X}$$

where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, $\bar{\delta} = \frac{1}{p}\sum_{i=1}^{p}\delta_i$, $\bar{X} = \frac{1}{p}\sum_{i=1}^{p}X_i$,

Hint: Prove first that

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{(1-\rho)\sigma^2}\left(\boldsymbol{I} - \frac{\rho}{1+(p-1)\rho}\mathbf{1}\mathbf{1}^T\right)$$

Chapter 12: Clustering

Exams: August 1998 # 1, November 1999 # 1, December 2005 # 4.