UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in STK4040/9040 — Multivariate analysis.

Day of examination: 18 December 2009.

Examination hours: 10.00 – 13.00.

This problem set consists of 3 pages.

Appendices: Table of the F-distribution

Permitted aids: Formulas for STK4040/9040. Approved calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

Let $\mathbf{X}$ be a p-variate random vector with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. The eigenvalues of $\boldsymbol{\Sigma}$ are $\lambda_1 \geq \cdots \geq \lambda_p > 0$, and the corresponding orthogonal and normalized eigenvectors are $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. We introduce the matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p]$ with the eigenvectors as columns, and let $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ be the diagonal matrix with the eigenvalues on the diagonal.

a) Use the spectral decomposition of $\boldsymbol{\Sigma}$ to explain that we may write $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$.

The square root matrix $\boldsymbol{\Sigma}^{1/2}$ is given by $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'$.

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

c) The inverse of the square root matrix is denoted $\boldsymbol{\Sigma}^{-1/2}$. Express this inverse in terms of $\mathbf{P}$ and $\boldsymbol{\Lambda}$ and show that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$.

Consider the random vector $\mathbf{Z} = [Z_1, \ldots, Z_p]'$ given by $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$.

d) Show that $E(\mathbf{Z}) = \mathbf{0}$ and $\text{Cov}(\mathbf{Z}) = \mathbf{I}$.

Assume now that the random vector $\mathbf{X}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed.

e) Explain that $Z_1, \ldots, Z_p$ are independent and standard normally distributed.

f) Show that $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ is chi-squared distributed with p degrees of freedom.

(Continued on page 2.)

Problem 2

Let $\mathbf{X} = [X_1, \ldots, X_p]'$ be a random vector with mean vector $\boldsymbol{\mu}$. A factor model with one factor assumes that we may write

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

Here $\mathbf{L} = [l_1, \ldots, l_p]'$ is a vector of loadings, and the common factor F is a random variable with mean 0 and standard deviation 1, assumed to be independent of the vector of specific factors $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_p]'$. Further the specific factors are assumed to be independent with means 0 and variances $\psi_1, \ldots, \psi_p$.

a) Derive an expression for the covariance matrix of $\mathbf{X}$.

b) Express the variances of the variables X_j in terms of the loadings l_i and the specific variances ψ_i.

c) Give an expression for the proportion of the total variance that is explained by the factor F.

Problem 3

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, and let $\mathbf{C}$ be a $q \times p$ matrix of rank q (with $q < p$). We want to test

$$\mathrm{H}_0 \,:\, \mathbf{C}\boldsymbol{\mu} = \mathbf{0} \quad \text{versus} \quad \mathrm{H}_1 \,:\, \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0} \tag{1}$$

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

$$n(\mathbf{C}\overline{\mathbf{X}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\overline{\mathbf{X}}), \tag{2}$$

where $\overline{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_j$ and $\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$.

b) Show that the test 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?

Assume now that we want to test

$$\mathrm{H}_0 \,:\, \mu_1 = \cdots = \mu_p \quad \text{versus} \quad \mathrm{H}_1 \,:\, \text{not all } \mu_j \text{ are equal,}$$

where $\mu_1, \mu_2, \ldots, \mu_p$ are the components of the mean vector $\boldsymbol{\mu}$.

c) Explain how this testing problem is a special case of the testing problem (1).

(Continued on page 3.)

A researcher considered three methods for measuring peak expiratory flow for asthmatic patients. (Peak expiratory flow is a measure of the maximum speed of expiration, and may be measured in liters per minute.) Ten asthmatic patients took part in the study, and for each of the patients the peak expiratory flow was measured using all three methods. The investigation produced the summary statistics:

$$\overline{\mathbf{x}} = [\,400, 410, 400\,]'$$

$$\mathbf{S} = \begin{bmatrix} 100 & 50 & 50 \\ 50 & 100 & 50 \\ 50 & 50 & 100 \end{bmatrix}$$

(The numbers are constructed to make the computations simple.)

d) Test the null hypothesis that there is no difference between the methods for measuring peak expiratory flow.

END