Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

Course code and name:	732A97 Multivariate Statistical Methods
Date:	2018/03/05, 8–12
Examinator:	Krzysztof Bartoszek phone 013–281 885
Allowed aids:	Pocket calculator
	Table with common formulae and moment generating functions (distributed with the exam)
	Table of integrals (distributed with the exam)
	Table with distributions from Appendix in the course book (distributed with the exam)
	One double sided A4 page with own hand written notes
Grades:	A= $[19 - \infty)$ points
	B= $[17 - 19)$ points
	C= $[14 - 17)$ points
	D= $[12 - 14)$ points
	E= $[10 - 12)$ points
	F= $[0 - 10)$ points
Instructions:	Write clear and concise answers to the questions.

Problem 1 (5p)

You are given the normally distributed vector $\vec{X} = (X_1, X_2, X_3)^T$ with mean vector $\mu_X = (2, 1, -1)^T$ and covariance matrix

$$\mathbf{\Sigma}_X = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Take

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

and define $\vec{Y} = \mathbf{A}\vec{X}$.

(a 2p) Find $\mathbb{E}\left[\vec{Y}\right]$ and $\mathrm{Var}\left[\vec{Y}\right]$.
(b 2p) Calculate the total variance and the generalized variance of both $\vec{X}$ and $\vec{Y}$.
(c 1p) What is the distribution of $\vec{Y}$?

Problem 2 (5p)

Let $\mathbf{A}$ be a square symmetric matrix and have the block representation

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Assume that both $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are invertible, square matrices, i.e. $\mathbf{A}_{11}^{-1}$ and $\mathbf{A}_{22}^{-1}$ exist.
(a 1p) Is there any relationship between $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$?
(b 4p) Show that

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

where $\mathbf{I}$ are identity matrices of appropriate sizes and $\mathbf{0}$ are matrices of 0s of appropriate sizes.

Problem 3 (4p)

Let the random vector $\mathbb{R}^4 \ni \vec{X} = (X_1, X_2, X_3, X_4)^T$ be distributed according to some distribution that has expectation vector $\vec{\mu} = (3, 4, 0, 2)^T$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 & 1 \\ -2 & 8 & 3 & 0 \\ 0 & 3 & 10 & -3 \\ 1 & 0 & -3 & 10 \end{bmatrix}.$$

(a 2p) Write all pairs of variables that you know are independent. Justify your choice.
(b 2p) Is $3X_1 + X_2$ correlated with $2(X_3 - X_4)$? Justify your answer.

Problem 4 (6p)

You are provided with the following distributional results.

- Let $\mathbb{R}^p \ni \vec{X} \sim \mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$, then

$$(\vec{X} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{X} - \vec{\mu}) \sim \chi_p^2,$$

- Let $\mathbb{R}^p \ni \overline{x}$ be the sample mean of n normal observations and $\mathbf{S}$ the sample covariance. If the population expectation is $\vec{\mu}$, then

$$(\overline{x} - \vec{\mu})^T \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\overline{x} - \vec{\mu}) \sim \frac{(n-1)p}{n-p} F_{p,n-p},$$

- If we have two independent samples, both of dimension p, first of size n_1 from $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma}_1)$ and second of sizes n_2 from $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma}_2)$, then denoting by $\overline{x}_1$, $\mathbf{S}_1$ and $\overline{x}_2$, $\mathbf{S}_2$ the respective sample averages and covariances

  –
  $$(\overline{x}_1 - \overline{x}_2)^T(\frac{1}{n_1}\mathbf{\Sigma}_1 + \frac{1}{n_2}\mathbf{\Sigma}_2)^{-1}(\overline{x}_1 - \overline{x}_2) \sim \chi^2_p,$$

  – if $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$
  $$(\overline{x}_1 - \overline{x}_2)^T(\frac{n_1-1}{n_1+n_2-2}\mathbf{S}_1 + \frac{n_2-1}{n_1+n_2-2}\mathbf{S}_2)^{-1}(\overline{x}_1 - \overline{x}_2) \sim \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1},$$

  – if $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ and n is large, then approximately
  $$(\overline{x}_1 - \overline{x}_2)^T(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2)^{-1}(\overline{x}_1 - \overline{x}_2) \sim \chi^2_p.$$

In a medical study the content of two different substances in the blood are measured in two groups, one taking medication (group M) and the other placebo (group P). The substances can be treated as a two–dimensional random vector $\vec{X} = (X_1, X_2)^T$ which can be assumed to be normally distributed. 100 people are assigned to the placebo group and 100 to the medication group. After the measurements the results are: sample mean vectors $\overline{x}_M = (3, 3.4)^T$, $\overline{x}_P = (2.6, 3)^T$ and sample covariance matrices

$$\mathbf{S}_M = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{S}_P = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

with inverses

$$\mathbf{S}_M^{-1} = \frac{1}{3}\begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{S}_P^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}, \quad (\mathbf{S}_M + \mathbf{S}_P)^{-1} = \frac{1}{10}\begin{bmatrix} 7 & -2 \\ -2 & 2 \end{bmatrix}$$

and eigendecompositions

$$\mathbf{S}_M = \begin{bmatrix} 0.29 & -0.957 \\ 0.957 & 0.29 \end{bmatrix}\begin{bmatrix} 4.303 & 0 \\ 0 & 0.697 \end{bmatrix}\begin{bmatrix} 0.29 & -0.957 \\ 0.957 & 0.29 \end{bmatrix}^{-1},$$

$$\mathbf{S}_P = \begin{bmatrix} 0.383 & -0.924 \\ 0.924 & 0.383 \end{bmatrix}\begin{bmatrix} 3.414 & 0 \\ 0 & 0.586 \end{bmatrix}\begin{bmatrix} 0.383 & -0.924 \\ 0.924 & 0.383 \end{bmatrix}^{-1},$$

$$\mathbf{S}_M + \mathbf{S}_P = \begin{bmatrix} 0.331 & -0.943 \\ 0.943 & 0.331 \end{bmatrix}\begin{bmatrix} 7.702 & 0 \\ 0 & 1.298 \end{bmatrix}\begin{bmatrix} 0.331 & -0.943 \\ 0.943 & 0.331 \end{bmatrix}^{-1}.$$

(a 3p) Perform a test at the 5% significance level if the expectation vectors in the two groups are equal.
(b 3p) Sketch a 95% confidence ellipse for the difference between the mean vectors. Does $(0,0)^T$ lie in it? Mark it on the graph.