Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

Course code and name:	732A97 Multivariate Statistical Methods
Date:	2018/04/25, 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 $\vec{\mu}_X = (2, 1, -1)^T$ 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 $\vec{Y} = \mathbf{A}\vec{X}$. You observe a sample of size 100 of $\vec{X}$s, such that

$$\sum_{i=1}^{100} \vec{x}_i = (50, 20, -25)^T.$$

(a 2p) Calculate the sample average $\overline{y}$.
(b 3p) What is the distribution of $\overline{y}$?

Problem 2 (5p)

Let $\mathbf{\Sigma} \in \mathbb{R}^{k \times k}$ be the variance–covariance matrix of the random vector $\vec{X} \in \mathbb{R}^k$.
(a 4p) Show that the covariance between two linear combinations of entries of $\vec{X}$, i.e. $\vec{a} \cdot \vec{X}$ and $\vec{b} \cdot \vec{X}$, for $\vec{a}, \vec{b} \in \mathbb{R}^k$, can be calculated using the matrix vector operation $\vec{a}\mathbf{\Sigma}\vec{b}$.
(a 1p) Using the above proven formula show that

$$\mathrm{Cov}\left[\vec{a} \cdot \vec{X}, \vec{b} \cdot \vec{X}\right] = \mathrm{Cov}\left[\vec{b} \cdot \vec{X}, \vec{a} \cdot \vec{X}\right],$$

i.e. the covariance is symmetric.

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 non–normal distribution that has expectation vector $\vec{\mu} = (3, 4, 0, 2)^T$ and covariance matrix

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

with eigendecomposition

$$\begin{bmatrix} 0.104 & -0.075 & -0.324 & 0.937 \\ -0.394 & 0.662 & 0.567 & 0.292 \\ -0.718 & 0.192 & -0.656 & -0.132 \\ 0.564 & 0.720 & -0.38 & -0.136 \end{bmatrix} \begin{bmatrix} 14.002 & 0 & 0 & 0 \\ 0 & 9.096 & 0 & 0 \\ 0 & 0 & 5.671 & 0 \\ 0 & 0 & 0 & 0.231 \end{bmatrix} \begin{bmatrix} 0.104 & -0.394 & -0.718 & 0.564 \\ -0.075 & 0.662 & 0.192 & 0.720 \\ -0.324 & 0.567 & -0.656 & -0.38 \\ 0.937 & 0.292 & -0.132 & -0.136 \end{bmatrix}.$$

(a 2p) Write all pairs of variables that you know are independent. Justify your choice.
(b 2p) Is $0.104X_1 - 0.394X_2 - 0.718X_3 + 0.564X_4$ correlated with $-0.075X_1 + 0.662X_2 + 0.192X_3 + 0.720X_4$? 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

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

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

  $$(\overline{x}_1 - \overline{x}_2)^T \left(\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \mathbf{S}_{\text{pooled}}\right)^{-1} (\overline{x}_1 - \overline{x}_2) \sim \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1},$$

  where

  $$\mathbf{S}_{\text{pooled}} = \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2$$

– 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. 10 people are assigned to the placebo group and 10 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

$$\mathbf{S}_M = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{S}_P = \begin{bmatrix} 1 & 1 \\ 1 & 3.95 \end{bmatrix}$$

with inverses

$$\mathbf{S}_M^{-1} = \frac{1}{3}\begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{S}_P^{-1} = \frac{1}{2.95}\begin{bmatrix} 3.95 & -1 \\ -1 & 1 \end{bmatrix}, \quad (\mathbf{S}_M + \mathbf{S}_P)^{-1} = \frac{1}{11.9}\begin{bmatrix} 7.95 & -2 \\ -2 & 2 \end{bmatrix}$$

and eigendecompositions

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

$$\mathbf{S}_P = \begin{bmatrix} 0.294 & -0.956 \\ 0.956 & 0.294 \end{bmatrix}\begin{bmatrix} 4.257 & 0 \\ 0 & 0.693 \end{bmatrix}\begin{bmatrix} 0.294 & -0.956 \\ 0.956 & 0.294 \end{bmatrix}^{-1},$$

$$\mathbf{S}_M + \mathbf{S}_P = \begin{bmatrix} 0.292 & -0.957 \\ 0.957 & 0.292 \end{bmatrix}\begin{bmatrix} 8.56 & 0 \\ 0 & 1.390 \end{bmatrix}\begin{bmatrix} 0.292 & -0.957 \\ 0.957 & 0.292 \end{bmatrix}^{-1}.$$

(a 3p) Perform a test at the 5% significance level if the expectation vectors in the two groups are equal. Justify your choice of test.
(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.