Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

Course code and name:	732A97 Multivariate Statistical Methods
Date:	2020/03/02, 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= $[18-\infty)$ points
	B= $[16-18)$ points
	C= $[14-16)$ 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 random vector $\vec{X}^T = [X_1, X_2, X_3, X_4]$ with mean vector $\vec{\mu}_X^T = [1, 0, 3, 1]$ and variance–covariance matrix

$$\mathbf{\Sigma}_X = \begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{bmatrix}.$$

Partition $\vec{X} = [X_1, X_2 | X_3, X_4]^T \equiv [\vec{X}^{(1)^T}, \vec{X}^{(2)^T}]^T$. Let

$$\mathbf{A} = [2 \;\; 3], \quad \mathbf{B} = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}.$$

Calculate $\mathbb{E}\left[\mathbf{B}\vec{X^{(2)}}\right]$, $\mathrm{Var}\left[\mathbf{A}\vec{X^{(1)}}\right]$ and $\mathrm{Cov}\left[\mathbf{B}\vec{X^{(2)}}, \mathbf{A}\vec{X^{(1)}}\right]$.

Problem 2 (5p)

Consider the covariance matrices

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}, \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0 & 1 & -0.5 \\ 0 & 0 & -0.5 & 1 \end{bmatrix}.$$

Find their eigenvalues and eigenvectors. Provide an interpretation of the principle components.
TIP: For a block diagonal matrix, the eigendecompositions can be done separately for each block and then combined appropriately.

Problem 3 (5p)

Let $\vec{X}$ be distributed as $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma}$ where $\vec{\mu} = [0, 0, 0, 0]^T$ and

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 3 & 0 & 2 & 0 \\ 0 & 4 & -1 & 0 \\ 2 & -1 & 5 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}.$$

Which of the following random variables are independent:
X_1 and X_2
X_3 and X_2
X_4 and $X_1 + 5X_3 - 10X_2$?

Problem 4 (5p)

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

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

– 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
$$(\overline{x}_1 - \overline{x}_2)^T(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2)^{-1}(\overline{x}_1 - \overline{x}_2) \sim \chi_p^2.$$

A data set containing 1000 independent measurements of abundance of two chemicals from a river is provided. The sample mean and variance are $\overline{x} = [10.1, 15.2]^T$ and
$$\mathbf{S} = \begin{bmatrix} 1000 & -0.053 \\ -0.053 & 2000 \end{bmatrix}.$$

TIP: In questions a) and b) you are allowed make some (part of the examination is which ones) approximations in your calculations.
(a 3p) Perform a test at the 5% significance level if the observed sample comes from a distribution with mean vector equalling $\vec{\mu} = [10, 15]^T$. Justify your choice of test. Does the test accept or reject the hypothesis?
(b 2p) Sketch a 95% confidence ellipse for the mean vector. Does $[10, 15]^T$ lie in it? Mark it on the graph.