Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

Course code and name:	732A97 Multivariate Statistical Methods
Date:	2019/02/14, 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 observe a sample from a three dimensional distribution. The observed sample mean vector is $\bar{X} = (1, 2, 1)^T$ and sample covariance matrix

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

Take

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

and define the linear transformation $f(\vec{x}) = \mathbf{A}\vec{x}$.

(a 2p) Calculate the sample average after transforming each of the original observations by $f(\cdot)$.
(b 3p) Calculate the sample covariance matrix after transforming each of the original observations by $f(\cdot)$.

Problem 2 (5p)

Let $\mathbf{A}$ be a symmetric–positive–definite matrix with eigenvalue decomposition $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, i.e. $\mathbf{P}$ is the matrix of eigenvectors, each column is an eigenvector and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Calculate $\mathbf{P}^T\mathbf{A}\mathbf{P}$.

Problem 3 (5p)

Let $\vec{x}_1$, $\vec{x}_2$, $\vec{x}_3$ and $\vec{x}_4$ be independent $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$ random vectors.
(a 2p) Find the marginal distributions for each of the random vectors
$\vec{v}_1 = \frac{1}{8}\vec{x}_1 - \frac{1}{8}\vec{x}_2 + \frac{1}{8}\vec{x}_3 - \frac{1}{8}\vec{x}_4$ and $\vec{v}_2 = \frac{1}{8}\vec{x}_1 + \frac{1}{8}\vec{x}_2 - \frac{1}{8}\vec{x}_3 - \frac{1}{8}\vec{x}_4$.
(b 2p) Find the joint density of $\vec{v}_1$ and $\vec{v}_2$ defined in (a).
(c 1p) Are $\vec{v}_1$ and $\vec{v}_2$ independent? Justify your answer.

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

A certain cryptographic procedure requires 100 2–dimensional random vectors that come from the normal distribution with mean $\vec{\mu} = (1, 3)^T$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 4 \end{bmatrix}$$

with inverse

$$\mathbf{\Sigma}^{-1} = \frac{1}{3}\begin{bmatrix} 3.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix}$$

and eigendecomposition (matrices rounded to third decimal point)

$$\mathbf{\Sigma} = \begin{bmatrix} 0.16 & -0.987 \\ 0.987 & 0.16 \end{bmatrix}\begin{bmatrix} 4.081 & 0 \\ 0 & 0.919 \end{bmatrix}\begin{bmatrix} 0.16 & 0.987 \\ -0.987 & 0.16 \end{bmatrix}.$$

You are provided with a sample of 100 2–dimensional vectors that have sample mean $\overline{x} = (1.5, 2.5)^T$ and sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} 0.95 & 0.41 \\ 0.41 & 4.05 \end{bmatrix}$$

with inverse (matrix rounded to third decimal point)

$$\mathbf{S}^{-1} = \frac{1}{3}\begin{bmatrix} 1.101 & -0.111 \\ -0.111 & 0.258 \end{bmatrix}$$

and eigendecomposition (matrices rounded to third decimal point)

$$\mathbf{S} = \begin{bmatrix} 0.129 & -0.992 \\ 0.992 & 0.129 \end{bmatrix}\begin{bmatrix} 4.103 & 0 \\ 0 & 0.897 \end{bmatrix}\begin{bmatrix} 0.129 & 0.992 \\ -0.992 & 0.129 \end{bmatrix}.$$

(a 3p) Perform a test at the 5% significance level if the observed sample comes from the desired normal distribution. Justify your choice of test.
(b 2p) Sketch a 95% confidence ellipse for the mean vector. Does $(1, 3)^T$ lie in it? Mark it on the graph.