Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

Course code and name:	732A97 Multivariate Statistical Methods
Date:	2019/08/29, 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)

Assume $p \geq 2$ and $\alpha \geq 0$. Let $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$ be a so–called *equivcorrelation* matrix, that is

$$\mathbf{\Sigma} = (1-\alpha)\mathbf{I} + \alpha \vec{1}_p \vec{1}_p^T,$$

where $\mathbf{I}$ is the identity matrix of appropriate size and $\vec{1}_p$ is a vector of p ones. Show that $\mathbf{\Sigma}$ is symmetric–positive–semi–definite.

Problem 2 (5p)

Let $\vec{X} \sim \mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$. Assume that $\mathbf{G}$ is an orthogonal matrix. Show that if $\vec{\mu} = \vec{0}$ and $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of appropriate size, then $\vec{X}$ and $\mathbf{G}\vec{X}$ have the same distribution.

Problem 3 (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 data set from 1921 contains information on 25 measurements of head lengths of the first and second son. The sample average is $\overline{x} = (185.72, 183.84)^T$ and the sample variance–covariance matrix is

$$\mathbf{S} = \begin{bmatrix} 91.481 & 66.875 \\ 66.875 & 96.775 \end{bmatrix} = \begin{bmatrix} 0.693 & -0.721 \\ 0.721 & 0.693 \end{bmatrix} \cdot \begin{bmatrix} 161.055 & 0 \\ 0 & 27.201 \end{bmatrix} \cdot \begin{bmatrix} 0.693 & -0.721 \\ 0.721 & 0.693 \end{bmatrix}^{-1}$$

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.022 & -0.015 \\ -0.015 & 0.021 \end{bmatrix}.$$

Assume that the measurements come from a normal distribution and that the true variance–covariance matrix is

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}.$$

(a 2p) Under the above assumption perform a test at the 5% significance level if the observed sample comes from a normal distribution with mean vector equalling $\vec{\mu} = (182, 182)^T$. 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 $(182, 182)^T$ lie in it? Mark it on the graph.
(c 1p) What does the assumed covariance matrix claim about the relationship between the two variables? Given the data does it seem reasonable that the measurements come from a distribution with such a variance–covariance matrix, $\mathbf{\Sigma}_1$?

Problem 4 (5p)

Assume now that the variance–covariance matrix for the data set presented in Problem 3 is

$$\mathbf{\Sigma}_2 = \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix} = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix} \cdot \begin{bmatrix} 150 & 0 \\ 0 & 50 \end{bmatrix} \cdot \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}^{-1} \quad \mathbf{\Sigma}_2^{-1} = \begin{bmatrix} 1/75 & -1/150 \\ -1/150 & 1/75 \end{bmatrix}.$$

(a 2p) Under the above assumption again perform a test at the 5% significance level if the observed sample comes from a normal distribution with mean vector equalling $\vec{\mu} = (182, 182)^T$. 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 $(182, 182)^T$ lie in it? Mark it on the graph.
(c 1p) Does the new assumed covariance matrix, $\mathbf{\Sigma}_2$, claim anything different from $\mathbf{\Sigma}_1$ about the relationship between the two variables? Does it seem more reasonable or not?