UPPSALA UNIVERSITET
Matematiska institutionen
Krzysztof Bartoszek
Tel. 471 3226

TENTAMEN I MATEMATISK STATISTIK
Multivariate Methods 1MS003
2014 XII 19

Maximum sum for exam: 20points.
Final grade calculated from exam, assignments and presentation.
Permitted aids: One double sided A4 sheet with handwritten notes, pocket calculator.
I will visit the exam room about 1 hour after the start of the exam.

1. (4p) Assume $\mathbb{R}^p \ni \vec{X}$ comes from a distribution with mean $\vec{\mu} \in \mathbb{R}^p$ and variance–covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ (full rank). Let $\mathbf{A} \in \mathbb{R}^{p\times p}$ be a non–singular matrix and $\vec{a} \in \mathbb{R}^p$ a fixed vector. Put $\vec{T} = \mathbf{A}\vec{X} + \vec{a}$ and derive expressions for
   a) The mean and variance–covariance matrix of $\vec{T}$.
   b) The mean and variance–covariance matrix of the principal components of $\vec{T}$.

2. (2p) Are the following valid distance functions on the set of positive real numbers? Are they metrics? Justify your answers.
   a) $d(x, y) = |\log(x/y)|$
   b) $d(x, y) = |(x - y)^3((x - y)^4 - 1)^2((x - y)^8 - 2)^2|$

3. (2p) Let $X \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ where

$$\mu = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}.$$

   Define further

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

   Are $A^T X$ and $B^T X$ independent? Justify.

4. (4p)
   a) (3p) Let $\mathbf{X}$ denote a $n \times p$ data matrix (we have a sample of n independent observations of a p–dimensional vector). Assume that the unbiased sample covariance matrix ($\vec{X}_i$ is the i–th column of $\mathbf{X}$ and $\overline{\mathbf{X}}$ is the sample average):

$$\mathbf{S}_X = \frac{1}{n-1}\sum_{i=1}^{n}(\vec{X}_i - \overline{\mathbf{X}})(\vec{X}_i - \overline{\mathbf{X}})^T$$

   has eigendecomposition $\mathbf{S}_X = \mathbf{P\Lambda P}^T$ where the entries of $\mathbf{\Lambda}$ are from highest to lowest. Define $\mathbf{Y} = \mathbf{HXP}$ where

$$\mathbf{H} = \mathbf{I} - \frac{1}{n}\vec{1}\vec{1}^T,$$

   $\mathbf{I}$ is the unit $n \times n$ matrix and $\vec{1}$ is a vector of n ones. Show that the unbiased sample covariance matrix of $\mathbf{Y}$ satisfies, $\mathbf{S}_Y = \mathbf{\Lambda}$.
   b) (1p) Show that the determinant of a eigendecomposable matrix $\mathbf{A}$ can be expressed as a product of its eigenvalues (for simplicity assume they are all real).

5. (4p) Let $\mathbb{R}^p \ni \vec{X} \sim \mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$ and $\mathbf{X}$ be a sample of n independent observations from the law of (distributed as) $\vec{X}$. Let $\overline{\mathbf{X}}$ be the sample average.

a) (2p) What is the distribution of $\sqrt{n}(\overline{\mathbf{X}} - \vec{\mu})$?

b) (2p) Let now $\vec{Y}$ come from a distribution with finite mean $\vec{\mu}$ and finite covariance $\mathbf{\Sigma}$. Let $\mathbf{Y}$ be a sample of n independent observations from the law of (distributed as) $\vec{Y}$. Let $\overline{\mathbf{Y}}$ be the sample average. What can be said about the distribution of $\sqrt{n}(\overline{\mathbf{Y}} - \vec{\mu})$ as the sample size increases? What implications for inference does this have?

6. (4p) A cohort of 98 patients underwent radiotherapy and measurements on their symptoms, activity, sleep, eating patterns, appetite and skin reaction were taken to see how they respond to the therapy. The sample covariance and correlation matrices with their eigendecompositions are,

$$\mathbf{S} = \begin{bmatrix} 4.65 & 0.93 & 0.59 & 0.28 & 1.07 & 0.16 \\ 0.93 & 0.61 & 0.11 & 0.12 & 0.39 & -0.02 \\ 0.59 & 0.11 & 0.57 & 0.09 & 0.35 & 0.11 \\ 0.28 & 0.12 & 0.09 & 0.11 & 0.22 & 0.02 \\ 1.07 & 0.39 & 0.35 & 0.22 & 0.86 & -0.01 \\ 0.16 & -0.02 & 0.11 & 0.02 & -0.01 & 0.86 \end{bmatrix} = \mathbf{P}_S \begin{bmatrix} 5.28 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.89 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.77 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.43 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 \end{bmatrix} \mathbf{P}_S^T$$

$$\mathbf{P}_S = \begin{bmatrix} 0.93 & -0.02 & 0.35 & -0.11 & -0.06 & 0.01 \\ 0.21 & 0.17 & -0.23 & 0.71 & 0.61 & -0.05 \\ 0.14 & -0.15 & -0.50 & -0.64 & 0.55 & 0.02 \\ 0.07 & 0.01 & -0.20 & 0.09 & -0.13 & 0.96 \\ 0.26 & 0.14 & -0.73 & 0.11 & -0.55 & -0.26 \\ 0.04 & -0.96 & -0.08 & 0.24 & -0.06 & -0.03 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0.55 & 0.36 & 0.39 & 0.54 & 0.08 \\ 0.55 & 1 & 0.19 & 0.46 & 0.54 & -0.03 \\ 0.36 & 0.19 & 1 & 0.35 & 0.50 & 0.16 \\ 0.39 & 0.46 & 0.35 & 1 & 0.70 & 0.07 \\ 0.54 & 0.54 & 0.50 & 0.70 & 1 & -0.01 \\ 0.08 & -0.03 & 0.16 & 0.07 & -0.01 & 1 \end{bmatrix} = \mathbf{P}_R \begin{bmatrix} 2.86 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.08 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.78 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.39 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.24 \end{bmatrix} \mathbf{P}_R^T$$

$$\mathbf{P}_R = \begin{bmatrix} -0.44 & -0.03 & -0.34 & 0.55 & 0.60 & 0.15 \\ -0.43 & -0.29 & -0.50 & 0.06 & -0.69 & 0.08 \\ -0.36 & 0.38 & 0.63 & 0.42 & -0.33 & 0.21 \\ -0.46 & -0.02 & 0.12 & -0.67 & 0.21 & 0.53 \\ -0.52 & -0.07 & 0.20 & -0.20 & 0.10 & -0.79 \\ -0.06 & 0.87 & -0.43 & -0.18 & -0.05 & -0.12 \end{bmatrix}.$$

In the above matrices the rows and columns are ordered as measurement on: symptoms, activity, sleep, eat, appetite, skin reaction.

a) (1p) Calculate the total sample variance?

b) (1p) How much does each eigenvalue contribute to the total sample variance?

c) (1p) Decide on the number of important principle components. Are you able to summarize the data with a single reaction–index?

d) (1p) Look at the principle components and provide interpretations of some or all of them.

Good luck!