TMA4267 Linear Statistical Models V2017 [L2]

Part 1: Multivariate RVs, and the multivariate normal distribution
Moments: mean and covariance [H:4.2]

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: January 13, 2017

Last lecture

- A random vector $\boldsymbol{X}_{(p \times 1)}$ is ... a p-dimensional vector of random variables.
  - Weight of cork deposits in $p = 4$ directions (N, E, S, W).
  - Rent index in Munich: rent, area, year of construction, location, bath condition, kitchen condition, central heating, district.
- Joint distribution function:$f(\boldsymbol{x})$.
- From joint distribution function to marginal (and conditional distributions).

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_p) dx_2 \cdots dx_p$$

- Cumulative distribution (definite integrals!) used to calculate probabilites.
- Independence: $f(x_1, x_2) = f_1(x_1) \cdot f(x_2)$ and $f(x_1 \mid x_2) = f_1(x_1)$.
- From marginal cumulative distribution functions to joint using copula.

Word cloud: Probability

Today

- Moments: important properties about the distribution of $\boldsymbol{X}$.
- E: Mean of random vector and random matrices.
- Cov: Covariance matrix.
- Corr: Correlation matrix.
- E and Cov of multiple linear combinations.

The Cork deposit data

- Classical data set from Rao (1948).
- Weigth of bark deposits of $n = 28$ cork trees in $p = 4$ directions (N, E, S, W).

Tree	N	E	S	W
1	72	66	76	77
2	60	53	66	63
3	56	57	64	58
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
28	48	54	57	43

How may we define a random vectors and random matrices for cork trees?

The Cork deposit data

Draw a random sample of size $n = 28$ from the population of cork treed and observe a $p = 4$ dimensional random vector for each tree.

$$\boldsymbol{X}_{(28\times 4)} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ X_{31} & X_{32} & X_{33} & X_{34} \\ \vdots & \vdots & \ddots & \vdots \\ X_{28,1} & X_{28,2} & X_{28,3} & X_{28,4} \end{bmatrix}$$

and $\mathrm{E}(\boldsymbol{X}) = \{\mathrm{E}(X_{ij})\}$.

Random vectors and matrices: rules for means

- Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{(p\times 1)}$:

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \qquad \boldsymbol{\mu}_{(p\times 1)} = \mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix}$$

- 1) Random matrix $\boldsymbol{X}_{(n\times p)}$ and random matrix $\boldsymbol{Y}_{(n\times p)}$:

$$\mathrm{E}(\boldsymbol{X}+\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$$

- 2) Random matrix $\boldsymbol{X}_{(n\times p)}$ and conformable constant matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\mathrm{E}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$$

Variance-covariance matrix

- Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{(p\times 1)}$:

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad \boldsymbol{\mu}_{(p\times 1)} = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

- Variance-covariance matrix $\boldsymbol{\Sigma}$ (real and symmetric)

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

- $\sigma_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)]$

Hands-on

Let $\boldsymbol{X}_{4\times 1}$ have variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}.$$

Explain to your neighbour what this means.

Correlation matrix

Correlation matrix $\boldsymbol{\rho}$ (real and symmetric)

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{\rho} = (\boldsymbol{V}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}^{\frac{1}{2}})^{-1}, \text{ where } \boldsymbol{V}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

Hands-on

Let $\boldsymbol{X}_{4\times 1}$ have variance-covariance matrix

$$\boldsymbol{\Sigma} = \left[\begin{array}{cccc} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{array}\right].$$

Find the correlation matrix.

Linear combinations

- Random vector $\boldsymbol{X}_{(p \times 1)}$ with mean vector $\boldsymbol{\mu_X} = \mathrm{E}(\boldsymbol{X})$ and variance-covariance matrix $\boldsymbol{\Sigma_X} = \mathrm{Cov}(\boldsymbol{X})$.
- The linear combinations $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{X}$ have

$$\begin{aligned} \boldsymbol{\mu_Z} &= \mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\mu_X} \\ \boldsymbol{\Sigma_Z} &= \mathrm{Cov}(\boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\Sigma_X}\boldsymbol{C}^T \end{aligned}$$

Hands-on: Focus on $\boldsymbol{C}$?

$$\boldsymbol{X} = \begin{bmatrix} X_N \\ X_E \\ X_S \\ X_W \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \mu_N \\ \mu_E \\ \mu_S \\ \mu_W \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{NN} & \sigma_{NE} & \sigma_{NS} & \sigma_{NW} \\ \sigma_{NE} & \sigma_{EE} & \sigma_{ES} & \sigma_{EW} \\ \sigma_{NS} & \sigma_{EE} & \sigma_{SS} & \sigma_{SW} \\ \sigma_{NW} & \sigma_{EW} & \sigma_{SW} & \sigma_{WW} \end{bmatrix}$$

- Scientists would like to compare the following three contrasts: N-S, E-W and (E+W)-(N+S),
- and define a new random vector $\boldsymbol{Y}_{(3\times1)} = \boldsymbol{C}_{(3\times4)}\boldsymbol{X}_{(4\times1)}$ giving the three contrasts.
- Write down $\boldsymbol{C}$.
- Use the formulas we just developed and *explain* how to find $\mathrm{E}(Y_1)$ and $\mathrm{Cov}(Y_1, Y_3)$.

Exam V2014: Problem 1a

Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ be a random vector with mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Further, let

$\boldsymbol{A} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$ be a matrix of constants.

Define $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \boldsymbol{AX}$.

Find $\mathrm{E}(\boldsymbol{Y})$ and $\mathrm{Cov}(\boldsymbol{Y})$.

Are X_1 and X_2 independent?

Are Y_1 and Y_2 independent? Justify your answers.

Find the mean of $\boldsymbol{X}^T\boldsymbol{AX}$.

The covariance matrix

Random vector $\boldsymbol{X}_{(p\times1)}$ with mean vector $\boldsymbol{\mu}_{(p\times1)}$ and covariance matrix

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

The covariance matrix is by construction symmetric, and we would only consider covariance matrices that are positive definite (PD).
Why would we only consider PD matrices?
Homework for next lecture: Read H.Chapter 2.1-2.2 to remind yourself of spectral decomposition (diagonalization), positive definite matrix, eigenvalues and eigenvectors.

What have we worked with today?

- Mean: $\boldsymbol{\mu}_X = E(\boldsymbol{X}) = \mathrm{E}(X_j)$
- Covariance: $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}((\boldsymbol{X} - \boldsymbol{\mu}_X)(\boldsymbol{Y} - \boldsymbol{\mu}_Y)^T)$.
- Variance-covariance:
  $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}((\boldsymbol{X} - \boldsymbol{\mu}_X)(\boldsymbol{X} - \boldsymbol{\mu}_X)^T)$, also sometimes denoted $\mathrm{Var}(\boldsymbol{X})$.
- Correlation: $\mathrm{Corr}(\boldsymbol{X}) = \boldsymbol{V}^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{V}^{-\frac{1}{2}}$.
- $\boldsymbol{CX}$: $\mathrm{E}(\boldsymbol{CX}) = \boldsymbol{C}\boldsymbol{\mu}_X$ and $\mathrm{Cov}(\boldsymbol{CX}) = \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T$.

Next lecture: First work with the covariance matrix and positive definiteness, then start with the multivariate normal distribution (where we use moment generating functions and a multivariate version of the transformation formula).