Module 2, Part 2: Random vectors, covariance, multivariate Normal distribution

TMA4268 Statistical Learning V2021

Stefanie Muff, Department of Mathematical Sciences, NTNU

January 19, 2021

Last update: January 18, 2021

Overview

- Random vectors
- The covariance and correlation matrix
- The multivariate normal distribution

Random vector

$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$

- A random vector $X_{(p \times 1)}$ is a p-dimensional vector of random variables. For example
  - Weight of cork deposits in $p = 4$ directions (N, E, S, W).
  - Factors to predict body fat: bmi, age, weight, hip circumference,....
    $X_1 \quad X_2 \quad X_3 \quad X_4$
- Joint distribution function: $f(x)$. $f : \mathbb{R}^p \to \mathbb{R}$ $\quad \int f(x)\,dx = 1$
- From joint distribution function to marginal (and conditional distributions).

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

- Cumulative distribution (definite integrals!) used to calculate probabilites.
- Independence: $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ and $f(x_1 \mid x_2) = f_1(x_1)$.

Moments

The moments are important properties of the distribution of X. We will look at:

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

```
corkds = as.matrix(read.table("https://www.math.ntnu.no/emner/TMA4268/2019v/data/corkMKB.txt"))
dimnames(corkds)[[2]] = c("N", "E", "S", "W")
head(corkds)
```

```
##       N  E  S  W
## [1,] 72 66 76 77
## [2,] 60 53 66 63
## [3,] 56 57 64 58
## [4,] 41 29 36 38
## [5,] 32 32 35 36
## [6,] 30 35 34 26
```

tree 1
tree 2
⋮

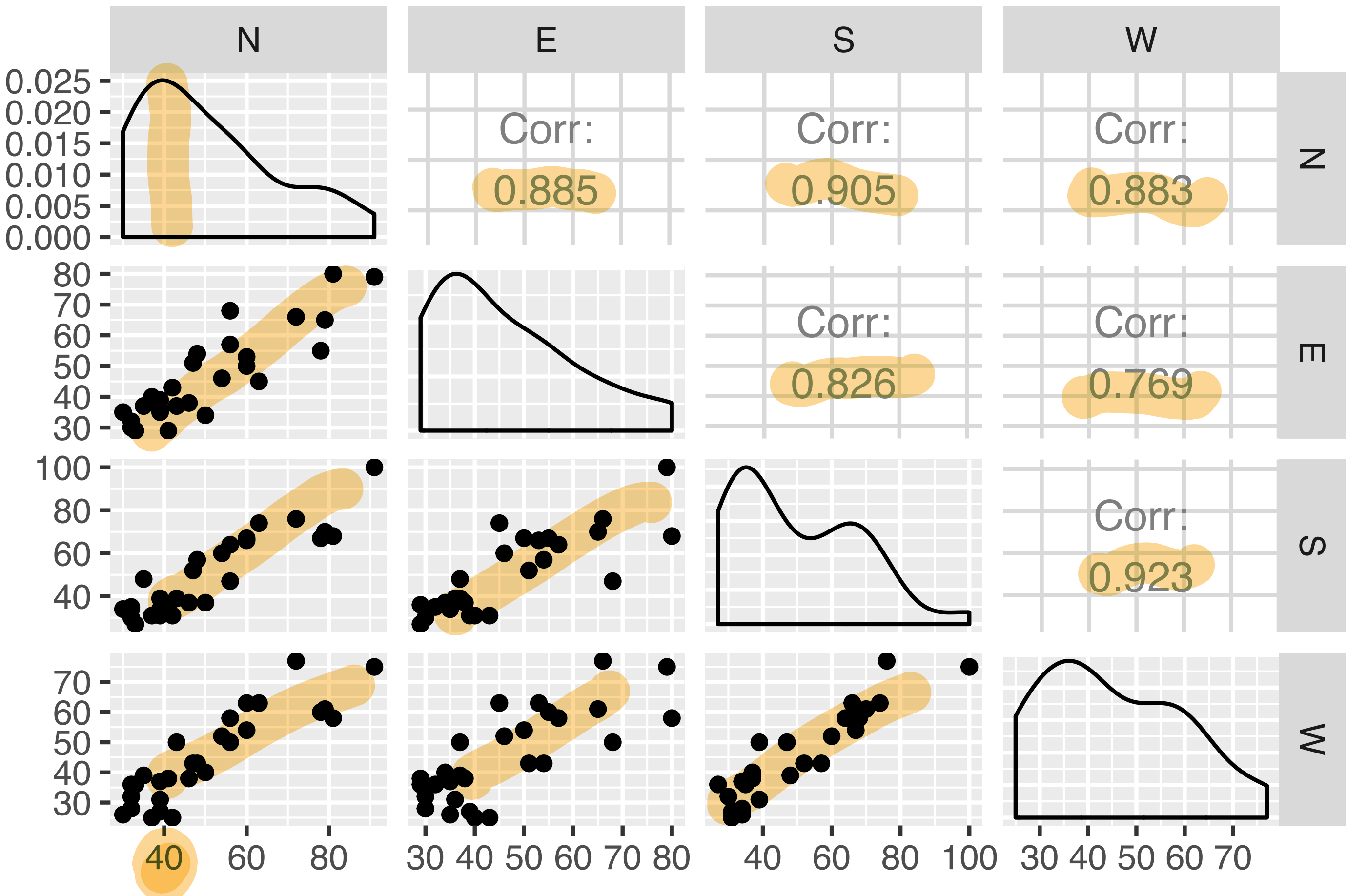
```
dim(corkds)
```

```
## [1] 28  4
```

X_1 X_2 X_3 X_4

Look at the data (always the first thing to do):

```
library(GGally)
corkds <- as.data.frame(corkds)
ggpairs(corkds)
```



- Here we have a random sample of $n = 28$ cork trees from the population and observe a $p = 4$ dimensional random vector for each tree.
- This leads us to the definition of random vectors and a random matrix for cork trees:

$$X_{(28\times 4)} = \begin{array}{c} \begin{array}{cccc} N & E & S & W \end{array} \\ \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ X_{31} & X_{32} & X_{33} & X_{34} \\ \vdots & \vdots & \ddots & \vdots \\ X_{28,1} & X_{28,2} & X_{28,3} & X_{28,4} \end{bmatrix} \\ \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \end{array}$$

Rules for means

- Random vector $X_{(p\times1)}$ with mean vector $\mu_{(p\times1)}$:[1]

$$X_{(p\times1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \text{ and } \mu_{(p\times1)} = \mathrm{E}(X) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix}.$$

- Same rule for random matrices.
- Random matrix $X_{(n\times p)}$ and random matrix $Y_{(n\times p)}$:

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

(Rules of vector addition)

---

[1] Observe that $\mathrm{E}(X_j)$ is calculated from the marginal distribution of X_j and contains no information about dependencies between X_j and X_k, $k \neq j$.

- Random matrix $X_{(n \times p)}$ and conformable constant matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$:

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

Proof: Board

Look at element (i,j) of AXB:

$$e_{ij} = \sum_{k=1}^{n} a_{ik} \sum_{l=1}^{p} x_{kl} \cdot b_{lj}$$

$$E[e_{ij}] = \sum_{k=1}^{n} a_{ik} \sum_{l=1}^{p} E[x_{kl}]\, b_{lj} \rightarrow$$ this is element (i,j) in $A \cdot E[X] \cdot B$

$E[\;]$ is a linear transf.

$\square$

Q:

- What are the univariate analogue to the formulas on the previous two slides (which you studied in your first introductory course in statistics)?

$$E[aX + bY] = aE[X] + bE[Y]$$

scalar a b

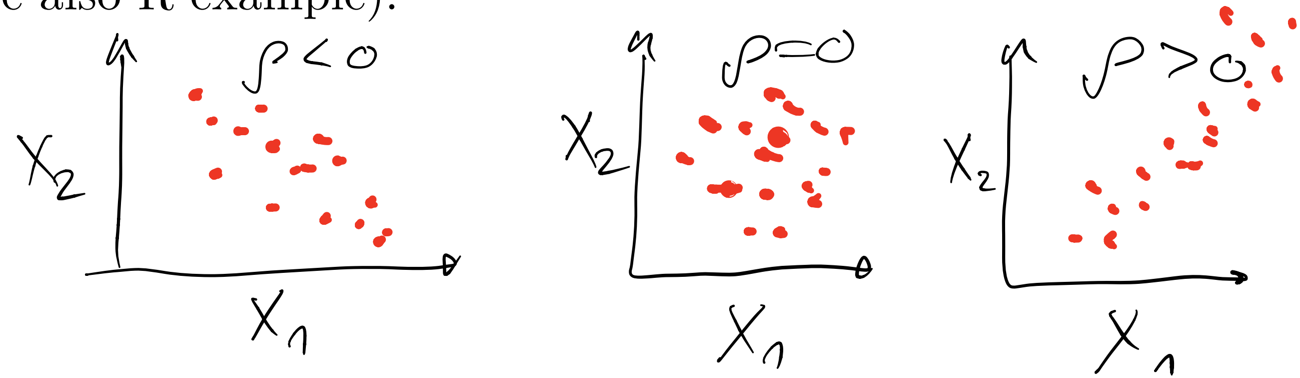
The covariance

In the introductory statistics course we defined the covariance

$$\begin{aligned}\rho_{ij} = \mathrm{Cov}(X_i, X_j) &= \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathrm{E}(X_i \cdot X_j) - \mu_i \mu_j \ .\end{aligned}$$

- What is the covariance called when $i = j$? $E[(X_i - \mu_i)^2] = Var(X_i)$
- What does it mean when the covariance is
  - negative
  - zero
  - positive?

Make a scatter plot for negative, zero and positive correlation (see also R example).



Variance-covariance matrix

- Consider random vector $X_{(p\times 1)}$ with mean vector $\mu_{(p\times 1)}$:

$$X_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \text{ and } \mu_{(p\times 1)} = \mathrm{E}(X) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix}$$

- Variance-covariance matrix Σ (real and symmetric)

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

$\sigma_{11} = \sigma_1^2$

- The diagonal elements in Σ, $\sigma_{ii} = \sigma_i^2$, are variances.
- The off-diagonal elements are covariances $\sigma_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ji}$.
- Σ is called variance, covariance and variance-covariance matrix and denoted both $\mathrm{Var}(X)$ and $\mathrm{Cov}(X)$.

Exercise: the variance-covariance matrix

Let $X_{4\times 1}$ have variance-covariance matrix

$\mathrm{cov}(X_1, X_2) = 1$

$\mathrm{cov}(X_1, X_3) = 0$

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

All X_i have same variance

Explain what this means.

Correlation matrix

Correlation matrix ρ (real and symmetric)

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

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

Correlation between −1 and 1

Exercise: the correlation matrix

Let $X_{4\times 1}$ have variance-covariance matrix

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

Find the correlation matrix.

$\sqrt{\sigma_{ii} \cdot \sigma_{jj}} = \sqrt{2 \cdot 2}$ for all i, j

$$\Rightarrow \rho = \Sigma \cdot \frac{1}{2} = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}$$

A:

Linear combinations

Consider a random vector $X_{(p\times 1)}$ with mean vector $\mu = \mathrm{E}(X)$ and variance-covariance matrix $\Sigma = \mathrm{Cov}(X)$.
The linear combinations

$$Z = CX = \begin{bmatrix} \sum_{j=1}^{p} c_{1j}X_j \\ \sum_{j=1}^{p} c_{2j}X_j \\ \vdots \\ \sum_{j=1}^{p} c_{kj}X_j \end{bmatrix}$$

have

$$\mathrm{E}(Z) = \mathrm{E}(CX) = C\mu = C \cdot \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_p] \end{bmatrix}$$

$$\mathrm{Cov}(Z) = \mathrm{Cov}(CX) = C\Sigma C^T$$

Exercise: Follow the proof (board) - what are the most important transitions?

univariate: $\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{Var}(X)$

Exercise: Linear combinations

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

Scientists would like to compare the following three *contrasts*: N-S, E+W and (E+W)-(N+S), and define a new random vector $Y_{(3\times1)} = C_{(3\times4)}X_{(4\times1)}$ giving the three contrasts.

- Write down C.
- Explain how to find $\text{E}(Y_1)$ and $\text{Cov}(Y_1, Y_3)$.
- Use R to find the mean vector, covariance matrix and correlations matrix of Y, when the mean vector and covariance matrix for X is given below.

Matrix multiplication

```
corkds <- as.matrix(read.table("https://www.math.ntnu.no/emner/TMA4268/2019v/data/corkMKB.txt"))
dimnames(corkds)[[2]] <- c("N", "E", "S", "W")
mu = apply(corkds, 2, mean)
mu
Sigma = var(corkds)
Sigma
```

```
##        N        E        S        W
## 50.53571 46.17857 49.67857 45.17857
##          N        E        S        W
## N 290.4061 223.7526 288.4378 226.2712
## E 223.7526 219.9299 229.0595 171.3743
## S 288.4378 229.0595 350.0040 259.5410
## W 226.2712 171.3743 259.5410 226.0040
```

$\Sigma_x = cov(X)$

```
(C <- matrix(c(1, 0, -1, 0, 0, 1, 0, 1, -1, 1, -1, 1), byrow = T, nrow = 3))
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0   -1    0
## [2,]    0    1    0    1
## [3,]   -1    1   -1    1
```

C

```
C %*% Sigma %*% t(C)
```

C^T

```
##           [,1]       [,2]       [,3]
## [1,]  63.53439  -38.57672   21.02116
## [2,] -38.57672  788.68254 -149.94180
## [3,]  21.02116 -149.94180  128.71958
```

$cov(Y) = C \cdot cov(X) \cdot C^T$

The covariance matrix - more requirements?

Random vector $X_{(p \times 1)}$ with mean vector $\mu_{(p \times 1)}$ and covariance matrix

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

- The covariance matrix is by construction symmetric, and it is common to require that the covariance matrix is positive semidefinite. This means that, for every vector $b \neq 0$

$$\underbrace{b^T \Sigma b}_{\text{scalar}} \geq 0 \; .$$

- Why do you think that is?

Hint: Is it possible that the variance of the linear combination $Y = b^T X$ is negative?

Random vectors - Single-choice exercise

We are doing a poll in zoom.

Question 1: Mean of sum

X and Y are two bivariate random vectors with $\mathrm{E}(X) = (1,2)^T$ and $\mathrm{E}(Y) = (2,0)^T$. What is $\mathrm{E}(X+Y)$?

- A: $(1.5, 1)^T$
- B: $(3, 2)^T$
- C: $(-1, 2)^T$
- D: $(1, -2)^T$

Question 2: Mean of linear combination

X is a 2-dimensional random vector with $\mathrm{E}(X) = (2, 5)^T$, and $b = (0.5, 0.5)^T$ is a constant vector. What is $\mathrm{E}(b^T X)$?

- A: 3.5
- B: 7
- C: 2
- D: 5

Question 3: Covariance

X is a p-dimensional random vector with mean μ. Which of the following defines the covariance matrix?

- A: $E[(X-\mu)^T(X-\mu)]$
- B: $E[(X-\mu)(X-\mu)^T]$
- C: $E[(X-\mu)(X-\mu)]$
- D: $E[(X-\mu)^T(X-\mu)^T]$

Question 4: Mean of linear combinations

X is a p-dimensional random vector with mean μ and covariance matrix Σ. C is a constant matrix. What is then the mean of the k-dimensional random vector $Y = CX$?

- A: $C\mu$
- B: $C\Sigma$
- C: $C\mu C^T$
- D: $C\Sigma C^T$

Question 5: Covariance of linear combinations

X is a p-dimensional random vector with mean μ and covariance matrix Σ. C is a constant matrix. What is then the covariance of the k-dimensional random vector $Y = CX$?

- A: $C\mu$
- B: $C\Sigma$
- C: $C\mu C^T$
- D: $C\Sigma C^T$

Question 6: Correlation

X is a 2-dimensional random vector with covariance matrix

$$\Sigma = \begin{bmatrix} 4 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

Then the correlation between the two elements of X are:

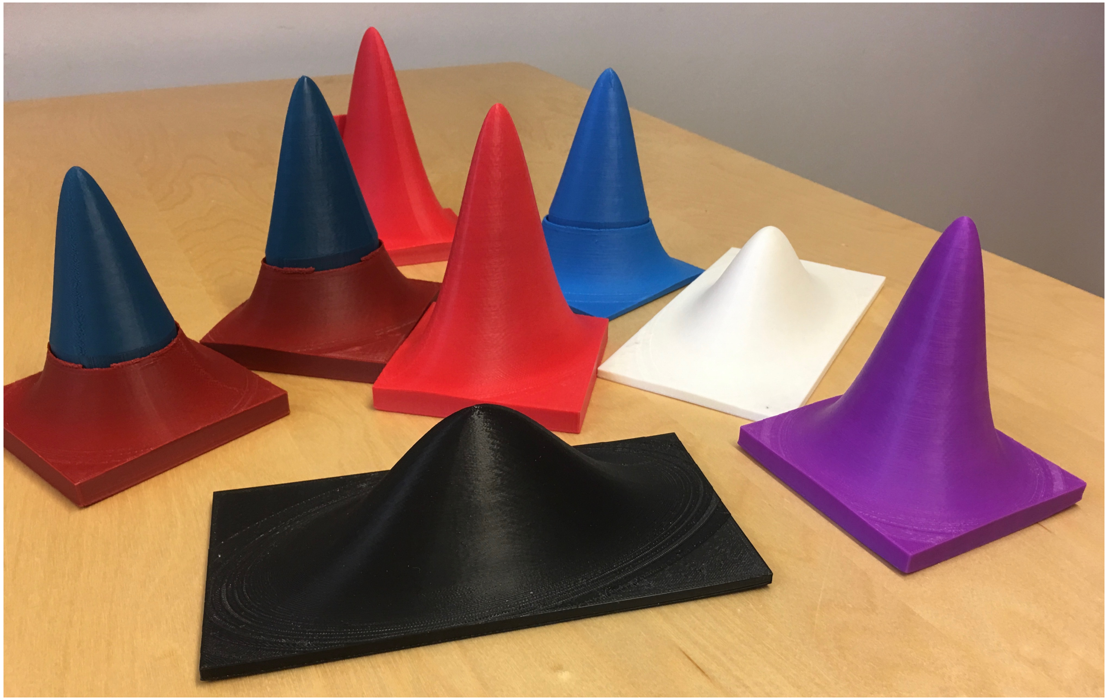
- A: 0.10
- B: 0.25
- C: 0.40
- D: 0.80

The multivariate normal distribution

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \Sigma_p \right)$$

Why is the mvN so popular?

- Many natural phenomena may be modelled using this distribution (just as in the univariate case).
- Multivariate version of the central limit theorem- the sample mean will be approximately multivariate normal for large samples.
- Good interpretability of the covariance.
- Mathematically tractable.
- Building block in many models and methods.



3D multivariate Normal distributions

The multivariate normal (mvN) pdf

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$$

The random vector $X_{p\times 1}$ is multivariate normal N_p with mean μ and (positive definite) covariate matrix Σ. The pdf is:

vector

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

p-variate case

$|\Sigma| = \det(\Sigma)$

Questions:

- How does this compare to the univariate version?

scalar

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\}$$

- Why do we need the constant in front of the exp?
- What is the dimension of the part in exp?
- What happens if the determinant $|\Sigma| = 0$ (degenerate case)?

Four useful properties of the mvN

Let $X_{(p \times 1)}$ be a random vector from $N_p(\mu, \Sigma)$.

1. The grapical contours of the mvN are ellipsoids (can be shown using spectral decomposition).
2. Linear combinations of components of X are (multivariate) normal
3. All subsets of the components of X are (multivariate) normal (special case of the above).
4. Zero covariance implies that the corresponding components are independently distributed (in contrast to general distributions).

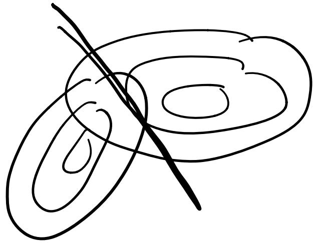
If you need a refresh, you might find that video useful: https://www.youtube.com/watch?v=eho8xH3E6mE

All of these are proven in TMA4267 Linear Statistical Models.

The result 4 is rather useful! If you have a bivariate normal and observed covariance 0, then your variables are independent.

Contours of multivariate normal distribution

- Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by x such that



$$(x-\mu)^T\Sigma^{-1}(x-\mu) = b$$

  where $b > 0$ is a constant.
- These ellipsoids are centered at μ and have axes $\pm\sqrt{b\lambda_i}e_i$, where $\Sigma e_i = \lambda_i e_i$ (eigenvector for λ_i), for $i = 1, \ldots, p$.
- To see this the spectral decomposition of the covariance matrix is useful.
- $(x-\mu)^T\Sigma^{-1}(x-\mu)$ is distributed as χ^2_p.

Note:

In M4: Classification the mvN is very important and we will often draw contours of the mvN as ellipses- and this is the reason why we do that.

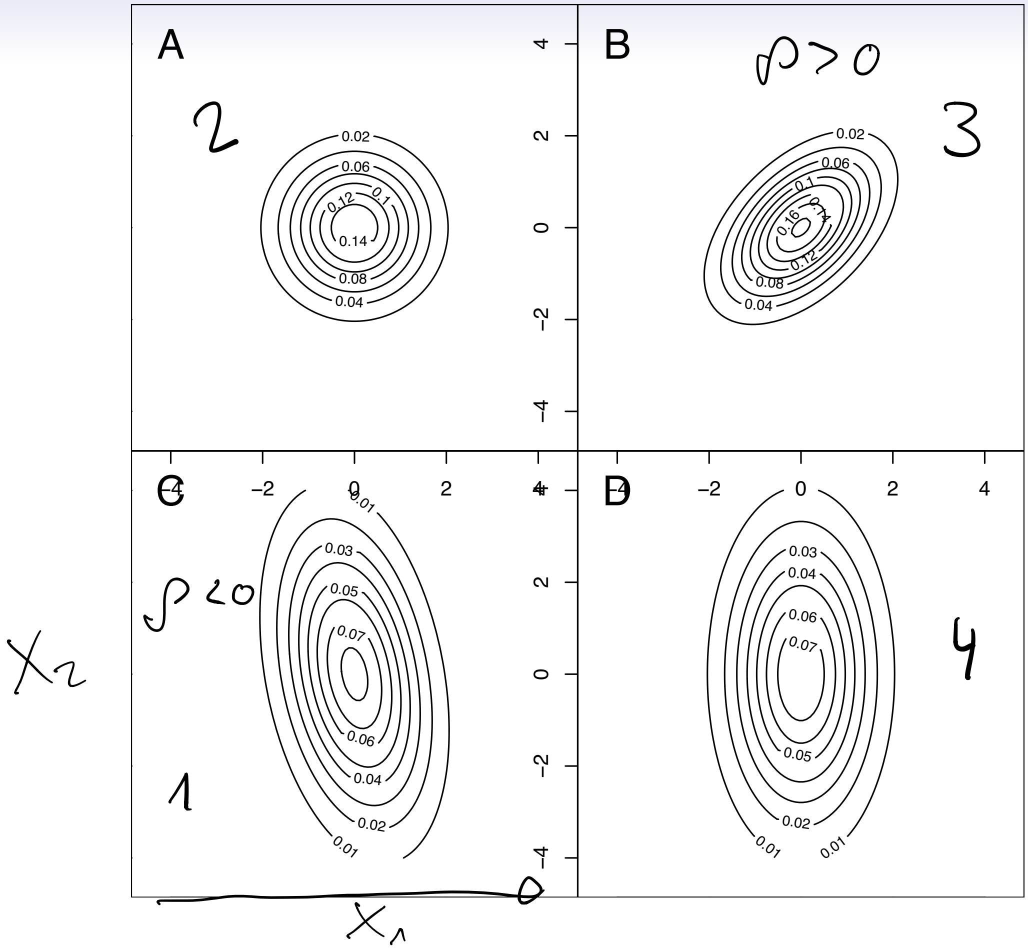
Identify the mvNs from their contours

Let $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$.

The following four figure contours have been generated:

- 1: $\sigma_x = 1$, $\sigma_y = 2$, $\rho = -0.3$ negative corr.
- 2: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0$
- 3: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0.5$ positive corr.
- 4: $\sigma_x = 1$, $\sigma_y = 2$, $\rho = 0$

Match the distributions to the figures on the next slide.



Take a look at the contour plots - when are the contours circles, when ellipses?

Multiple choice - multivariate normal

A second zoom poll.
Choose the correct answer. Let's go!

Question 1: Multivariate normal pdf

The probability density function is $(\frac{1}{2\pi})^{\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\{-\frac{1}{2}Q\}$ where Q is

- A: $(x-\mu)^T\Sigma^{-1}(x-\mu)$
- B: $(x-\mu)\Sigma(x-\mu)^T$
- C: $\Sigma-\mu$

Question 2: Trivariate normal pdf

What graphical form has the solution to $f(x) = \text{constant}$?

- A: Circle
- B: Parabola
- C: Ellipsoid
- D: Bell shape

Question 3: Multivariate normal distribution

$X_p \sim N_p(\mu, \Sigma)$, and C is a $k \times p$ constant matrix. $Y = CX$ is

- A: Chi-squared with k degrees of freedom
- B: Multivariate normal with mean $k\mu$
- C: Chi-squared with p degrees of freedom
- D: Multivariate normal with mean $C\mu$

Question 4: Independence

Let $X \sim N_3(\mu, \Sigma)$, with

$$\Sigma = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & 2 & 5 \end{bmatrix}.$$

Which two variables are independent?

- A: X_1 and X_2
- B: X_1 and X_3
- C: X_2 and X_3
- D: None – but two are uncorrelated.

Question 5: Constructing independent variables?

Let $X \sim N_p(\mu, \Sigma)$. How can I construct a vector of independent standard normal variables from X?

- A: $\Sigma(X - \mu)$
- B: $\Sigma^{-1}(X + \mu)$
- C: $\Sigma^{-\frac{1}{2}}(X - \mu)$
- D: $\Sigma^{\frac{1}{2}}(X + \mu)$

$$\mathrm{cov}\left(\Sigma^{-\frac{1}{2}}(X)\right) = \Sigma^{-\frac{1}{2}} \Sigma \overbrace{\left(\Sigma^{-\frac{1}{2}}\right)^T}^{\Sigma^{-\frac{1}{2}}}$$

$$= I$$

Further reading/resources

- Videoes on YouTube by the authors of ISL, Chapter 2

Acknowledgements

Thanks to Mette Langaas, who developed the first slide set in 2018 and 2019, and to Julia Debik for contributing to this module.