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

TMA4268 Statistical Learning V2020

Stefanie Muff, Department of Mathematical Sciences, NTNU

January xx, 2020

Overview

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

Random vector

- A random vector $\mathbf{X}_{(p\times 1)}$ 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: $f(\mathbf{x})$.
- 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 $f(x_1 \mid x_2) = f_1(x_1)$.

Moments

The moments are important properties about the distribution of $\mathbf{X}$. 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
```

```
dim(corkds)
```

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

- 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:

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

Rules for means

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

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

$\rightarrow$ 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 $\mathbf{X}_{(n\times p)}$ and random matrix $\mathbf{Y}_{(n\times p)}$:

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

(Rules of vector addition)

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

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

Proof: Look at element (i, j) of $\mathbf{AXB}$

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

(where a_{ik} and b_{lj} are elements of $\mathbf{A}$ and $\mathbf{B}$ respectively), and see that $\mathrm{E}(e_{ij})$ is the element (i, j) if $\mathbf{A}\mathrm{E}(\mathbf{X})\mathbf{B}$.

Q: what are the univariate analog to this formula - that you studied in your first introductory course in statistics? What do you think happens if we look at $\mathrm{E}(\mathbf{AXB}) + \mathbf{d}$?

A:

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

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 $i = j$?
- What does it mean when the covariance is
    - negative
    - zero
    - positive?

Make a scatter plot to show this.

Variance-covariance matrix

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

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

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

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

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

Exercise: the variance-covariance matrix

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

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

Explain what this means.

Correlation matrix

Correlation matrix ρ (real and symmetric)

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

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

Exercise: the correlation matrix

Let $\mathbf{X}_{4\times1}$ have variance-covariance matrix

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

Find the correlation matrix.

A:

$$\rho = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix}$$

Linear combinations

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

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

have

$$\mathrm{E}(\mathbf{Z}) = \mathrm{E}(\mathbf{C}\mathbf{X}) = \mathbf{C}\mu$$

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

Proof

Exercise: Study the proof - what are the most important transitions? (todo: study proof in Stahel and perhaps show on board)

Exercise: Linear combinations

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

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

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

```r
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
```

```r
(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
```

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

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

The covariance matrix - more requirements?

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

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

- 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 $\mathbf{b} \neq \mathbf{0}$

$$\mathbf{b}^T \boldsymbol{\Sigma} \mathbf{B} \geq 0 \ .$$

- Why do you think that is?

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

Multiple choice - random vectors

Choose the correct answer - time limit was 30 seconds for each question! Let's go!

Question 1: Mean of sum

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

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

Question 2: Mean of linear combination

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

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

Question 3: Covariance

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

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

Question 4: Mean of linear combinations

$\mathbf{X}$ is a p-dimensional random vector with mean μ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{C}$ is a constant matrix. What is then the mean of the k-dimensional random vector $\mathbf{Y} = \mathbf{C}\mathbf{X}$?

- A: $\mathbf{C}\mu$
- B: $\mathbf{C}\boldsymbol{\Sigma}$
- C: $\mathbf{C}\mu\mathbf{C}^T$
- D: $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$

Question 5: Covariance of linear combinations

$\mathbf{X}$ is a p-dimensional random vector with mean μ and covariance matrix $\mathbf{\Sigma}$. $\mathbf{C}$ is a constant matrix. What is then the covariance of the k-dimensional random vector $\mathbf{Y} = \mathbf{C}\mathbf{X}$?

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

Question 6: Correlation

$\mathbf{X}$ is a 2-dimensional random vector with covariance matrix

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

Then the correlation between the two elements of $\mathbf{X}$ are:

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

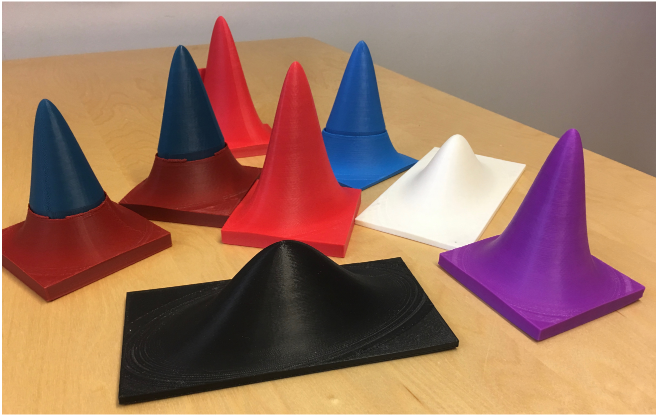
Answers:

1B, 2A, 3B, 4A, 5D, 6C

The multivariate normal distribution

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

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

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

Questions:

- How does this compare to the univariate version?

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

- 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 $|\mathbf{\Sigma}| = 0$?

Six useful properties of the mvN

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

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

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 $\mathbf{x}$ such that

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

  where $b > 0$ is a constant.

- These ellipsoids are centered at μ and have axes $\pm\sqrt{b\lambda_i}\mathbf{e}_i$, where $\mathbf{\Sigma}\mathbf{e}_i = \lambda_i \mathbf{e}_i$, for $i = 1, ..., p$.
- To see this the spectral decomposition of the covariance matrix is useful.
- $(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)$ 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.

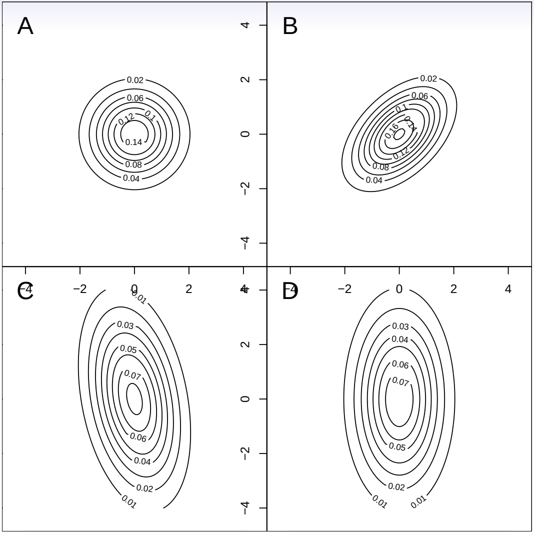
Identify the mvNs from their contours

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

The following four figure contours have been generated:

- 1: $\sigma_x = 1$, $\sigma_y = 2$, $\rho = -0.3$
- 2: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0$
- 3: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0.5$
- 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

Choose the correct answer - time limit was 30 seconds for each question! Let's go!

Question 1: Multivariate normal pdf

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

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

Question 2: Trivariate normal pdf

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

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

Question 3: Multivariate normal distribution

$\mathbf{X}_p \sim N_p(\mu, \mathbf{\Sigma})$, and $\mathbf{C}$ is a $k \times p$ constant matrix. $\mathbf{Y} = \mathbf{CX}$ 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 $\mathbf{C}\mu$

Question 4: Independence

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

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

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

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

Answers:

1A 2C 3D 4B 5C

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.