Time Series (MATH5845)

Dr. Atefeh Zamani
Based on the notes by Prof. William T.M. Dunsmuir

T2 2025

Chapter 1

Introduction to MATH5845

Contents

1.1 Welcome to Time Series (MATH5845) **2**
1.2 About the Lecturer **2**
1.3 What is Time Series Analysis? **2**
1.4 Course Overview **3**
1.5 Assessments **3**
1.5.1 More on Group Project 3
1.6 Course Structure **4**
1.7 Resources **5**
1.8 Expectations and Software **5**
1.8.1 Expectations 5
1.8.2 Software 5
1.9 Review Notes on Multivariate Distributions **6**
1.9.1 Joint Distribution and Density Functions 6
1.9.2 Independence 6
1.9.3 Conditional Distributions. 7
1.9.4 Expected Values. 7
1.9.5 Means and Covariances for Random Vectors 7
1.9.6 The Multivariate Normal Distribution. 8
1.10 Exercises **13**

1.1 Welcome to Time Series (MATH5845)

Time series analysis is a fundamental area of study in statistics and data science, focusing on data collected sequentially over time. The objective of this course, MATH5845 (Time Series), is to equip you with both theoretical knowledge and practical tools for analyzing time-dependent data and building models to uncover insights from temporal patterns.

This chapter introduces the structure of the course, and the resources available to guide you through the term.

1.2 About the Lecturer

Lecturer: Dr. Atefeh Zamani

- **Education**: Ph.D. in Mathematical Statistics and Master in Data Science
- **Research Interests**: Functional Data Analysis, Time Series Analysis, Interval-valued time series, Mathematical Statistics and Data Science
- **Contact Information**: Email: `atefeh.zamani@unsw.edu.au`
  Please include "MATH5845" in the subject line for course-related inquiries.

1.3 What is Time Series Analysis?

A time series is a sequence of data points measured and recorded at successive time intervals, typically spaced uniformly. It allows to analyze how a variable behaves over time and to infer future behavior based on historical patterns.

Why Study Time Series?

Time series analysis is essential for understanding patterns in data and making predictions. Key reasons for studying time series include:

- **A**nalyze Past Trends: Understand historical behavior, seasonality, and cyclicality.
- **F**orecast Future Values: Make predictions about future events based on observed data.
- **D**etect Patterns and Anomalies: Identify unusual behavior or changes in trends.
- **M**odel Relationships: Explore causal connections and interactions between variables.

Forecast vs. Prediction

Although these terms are often used interchangeably, they have distinct meanings:

- **Forecasting** focuses on estimating future values using past data (specific to time series)
- **Prediction** refers more generally to estimating outcomes, which may include past, present, or future values.

Understanding these distinctions is crucial for applying appropriate techniques in practice.

1.4 Course Overview

Topics Covered

This course provides a comprehensive overview of time series analysis. The topics include:

- **Week 1**: Introduction to time series
- **Week 2**: Simple models for time series
- **Week 3**: ARMA models
- **Week 4**: Estimation and prediction for ARMA models
- **Week 5**: ARIMA models
- **Week 6**: Flexibility Week
- **Week 7**: Time series regression
- **Week 8**: Spectral analysis- Part 1
- **Week 9**: Spectral analysis- Part 2
- **Week 10**: Additional topics

1.5 Assessments

- **Quiz (10%)**: Week 5, multiple-choice and short-answer questions.
- **Midterm Exam (15%)**: Week 7, in-person, pen and paper.
- **Group Project (15%)**: Analyze a dataset and submit a report by Week 10.
- **Final Exam (60%)**: In-person, scheduled during the exam period.

Note 1.1 *You must achieve* **at least 50% of the total mark AND at least 40% in the final exam** *to pass this course.*

1.5.1 More on Group Project

- Groups of three/four in the same tutorial (Created randomly in Week 1)
- Each group has a dataset to analyze
- Throughout the term, you will work on this dataset during tutorials and independently
- Submit a comprehensive report summarizing your findings by the end of Week 10.
- Goal is to
  - enhance your problem-solving skills
  - provide hands-on experience with real-world data analysis

    – teamwork

- You need to work as a group and your contribution will be assessed by yourself and your team by the end of the term.
- People in the same group might not receive the same mark.

Week	Topic	Assessment
1	Introduction to time series	
2	Simple models for time series	
3	ARMA models	
4	Estimation and prediction for ARMA models	Quiz (10%)
5	ARIMA models	
6	Flexibility week (self-directed learning)	
7	Time series regression	Midterm Exam (15%)
8	Spectral analysis- Part 1	
9	Spectral analysis- Part 2	
10	Additional topics	Group Project (15%)
		Final Exam (60%)

Table 1.1: Course Schedule and Assessments

1.6 Course Structure

The course includes both lectures and tutorials:
Lectures provide theoretical foundations and examples, with opportunities for interaction and clarification.

Day	**Time**	**Location**	
Tuesday	9:00 - 11:00	F10 June Griffith M18	(K-F10-M18)
Wednesday	9:00 - 10:00	Colombo Theatre A	(K-B16-LG03)

In **Tutorials** the focus is on practical applications using R and theoretical exercises. You should attend **the tutorial you are enrolled in** due to the group project.

Day	**Time**	**Location**	
Wednesday	10:00 - 11:00	UNSW Business School 220	(K-E12-220)
Wednesday	11:00 - 12:00	UNSW Business School 232	(K-E12-232)
Wednesday	12:00 - 13:00	Mathews 105	(K-F23-105)
Wednesday	13:00 - 14:00	Law Library G17	(K-F8-G17)
Friday	11:00 - 12:00	UNSW Business School 105	(K-E12-105)
Friday	12:00 - 13:00	Law Library G17	(K-F8-G17)
Friday	13:00 - 14:00	UNSW Business School 119	(K-E12-119)

1.7 Resources

Moodle Site

The Moodle site serves as the central hub for course materials:

- Lecture slides (uploaded before class).
- Discussion forums for Q&A.
- Assessment instructions and grades.

Lecture Notes

Notes by **William T. M. Dunsmuir (Emeritus Professor in UNSW)** with some modifications

Recommended Textbooks

Main Textbooks:

- Brockwell, P.J., & Davis, R.A. (2009). *Time Series: Theory and Methods.*
- Montgomery, D.C., Jennings, C.L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting.*
- Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications.*

Elementary Resource:

- Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting.*

1.8 Expectations and Software

1.8.1 Expectations

This course integrates both theory and practical applications. Students are expected to:

- Understand the theoretical foundations of time series analysis.
- Apply these concepts to real-world data using statistical software.
- Engage actively in lectures, tutorials, and group work.

Note that assessments and the final exam will include theory, proofs, and practical questions. We believe that understanding of theoretical concepts is needed for application.

1.8.2 Software

The course uses `R` for data analysis. Python users can refer to:

- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python.*

1.9 Review Notes on Multivariate Distributions

These notes will serve as a foundation for understanding basic concepts of joint distributions of random variables and the properties of the multivariate normal distribution. You might remember them from Multivariate analysis (MATH5845). Therefore, we are going to have a quick review.

1.9.1 Joint Distribution and Density Functions

An n-dimensional random vector

$$X = (X_1, \ldots, X_n)'$$

has joint distribution function

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

for all real numbers $x_1, \ldots, x_n$.

The joint distribution of any sub-vector can be obtained by setting $x_i = \infty$ for the other random variables not in the sub-vector. For example, the distribution of X_1 is given by

$$F_{X_1}(x_1) = F(x_1, \infty, \ldots, \infty)$$

and the joint distribution of (X_i, X_j) by

$$F_{X_i,X_j}(x_i, x_j) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty, x_j, \infty, \ldots, \infty).$$

A random vector is said to be continuous if its distribution function can be written in terms of a non-negative density function $f(\cdot)$ as

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \ldots, y_n) dy_1 \ldots dy_n$$

where

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \ldots, y_n) dy_1 \ldots dy_n = 1.$$

Note that we can derive the density by differentiating the distribution function

$$f(x_1, \ldots, x_n) = \frac{\partial F(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

1.9.2 Independence

The random variables $X_1, \ldots X_n$ are said to be independent if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

for all $x_1, \ldots, x_n \in \mathbb{R}$. This is equivalent to

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

or

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

Note 1.2 *For random variables $X_1, \ldots X_n$,*

- *two random variables X_i and X_j, $i \neq j$, are (pairwise) independent* **iff** $f(x_i, x_j) = f(x_i)f(x_j)$.
- *the set of n random variables is said to be pairwise independent if each pair of random variables in the set are independent.*
- *pairwise independence does not imply that the entire set is independent.*

1.9.3 Conditional Distributions.

Let $X = (X_1, \ldots, X_n)'$ and $Y = (Y_1 \ldots, Y_m)'$ be two random vectors with joint density $f_{X,Y}$. The conditional density of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Note that if X and Y are independent, then

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

so that $f_{Y|X}(y|x) = f_Y(y)$, in which case knowledge of $X = x$ does not alter the probabilities assigned to outcomes for Y. Conversely, if $f_{Y|X}(y|x) = f_Y(y)$ then X and Y are independent. Similary properties hold in terms of the distribution functions.

1.9.4 Expected Values.

Let $g(X)$ be a function of the random vector X. The expected value of $g(X)$ is

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n \\ &= \int_{-\infty}^{\infty} g(x)f(x)dx. \end{aligned}$$

Note 1.3 *In the univariate case, the mean $\mu = E(X)$ of a random variable corresponds to setting $g(X) = X$ while the variance $\sigma^2 = var(X) = E(X-\mu)^2$ corresponds to setting $g(X) = (X-\mu)^2$. The linearity property of expectation is*

$$E(aX + b) = aE(X) + b.$$

Note also that

$$var(aX + b) = a^2 var(X).$$

1.9.5 Means and Covariances for Random Vectors

For the multivariate random variable X, the mean vector is

$$\mu_X = E(X) = (E(X_1), \ldots, E(X_n))'$$

and the covariance between X_i and X_j is

$$cov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j).$$

The correlation is

$$corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{var(X_i)var(X_j)}}.$$

For two random vectors X and Y the covariance matrix between them is

$$\Sigma_{XY} = cov(X, Y) = E(X - EX)(Y - EY)' = E(XY') - (EX)(EY)'$$

with (i, j) element

$$(\Sigma_{XY})_{ij} = cov(X_i, Y_j).$$

When $Y = X$, $cov(X, Y)$ reduces to the variance matrix of the random vector X. Note that if X and Y are independent then the covariance between them is the null (zero) matrix and we call them uncorrelated. The converse is not true in general but is true for the multivariate normal distribution - see below.

Let Y and X be linearly related as $Y = a + BX$ where a is a vector and B is a matrix (all with conforming dimensions). Then

$$\mu_Y = E(Y) = a + BE(X) = a + B\mu_X$$

and

$$\Sigma_{YY} = B\Sigma_{XX}B'.$$

Note also that any variance matrix Σ is non-negative definite, that is $b'\Sigma b \geq 0$ for any vector b. **Hint**: To prove, let $Y = b'X$ where X has covariance matrix Σ. Now, find $var(Y)$.

1.9.6 The Multivariate Normal Distribution.

The general multivariate normal density

The random vector X has the multivariate normal distribution with mean μ and non-singular covariance matrix Σ if

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right\}.$$

Notation: $X \sim N(\mu, \Sigma)$.

The Bivariate Normal Density

As special case is the bivariate normal density from which most of the required insight about the multivariate normal is obtained. Let $X = (X_1, X_2)'$, with $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $var(X_1) = \sigma_1^2$, $var(X_2) = \sigma_2^2$ and $corr(X_1, X_2) = \rho$. Then,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

with inverse

$$\Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{bmatrix}$$

and $\det(\Sigma) = \sigma_1^2\sigma_2^2(1-\rho^2)$. Substitution in the general multivariate normal density gives

$$f_X(x) = \frac{1}{2\pi[\sigma_1^2\sigma_2^2(1-\rho^2)]^{1/2}} \exp\left\{-\frac{1}{2}Q(x_1, x_2; \sigma_1, \sigma_2, \rho)\right\}$$

with quadratic form

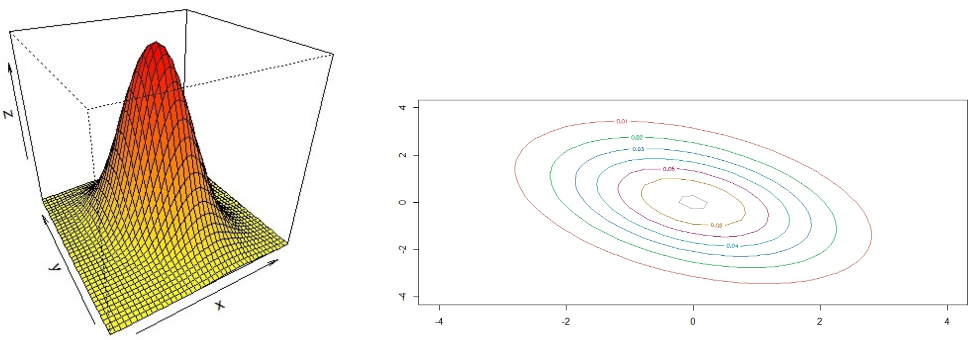
$$Q(x_1, x_2; \sigma_1, \sigma_2, \rho) = \frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

Some important facts about the bivariate normal are:

1. The contours of equal density are ellipses

$$\{(x_1, x_2) : Q(x_1, x_2; \sigma_1, \sigma_2, \rho) = k\}$$

for any constant $k \geq 0$. Figure 1.1 shows the 3-D density function and the contour plot of $X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}\right)$.



(a) (b)

Figure 1.1: (a) 3-D density function (b) contour plot of bivariate normal distribution with $cov(X_1, X_2) = -1$.

2. When the correlation $\rho = 0$ the two random variables X_1 and X_2 are independent. This can easily be concluded from the form of $Q(x_1, x_2; \sigma_1, \sigma_2, \rho = 0)$. Hence for the bivariate normal distribution, independence is equivalent to uncorrelatedness. Figure 1.2 demonstrates the 3-D density function and the contour plot of $X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\right)$.

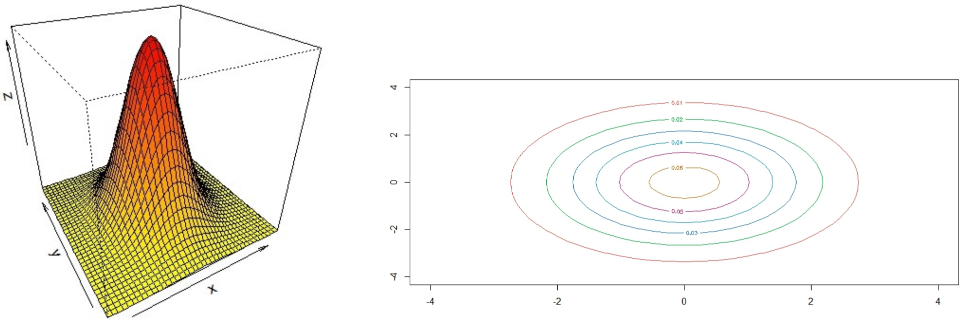
Standardised Bivariate Normal Density

Consider a special case of the bivariate normal density for the two standardised random variables

$$U = \frac{X_1-\mu_1}{\sigma_1}, \quad V = \frac{X_2-\mu_2}{\sigma_2},$$

with joint normal density with

$$\mu_U = \mu_V = 0,$$
$$\sigma_U^2 = \sigma_V^2 = 1,$$



(a) (b)

Figure 1.2: (a) 3-D density function (b) contour plot of bivariate normal distribution with $cov(X_1, X_2) = 0$.

so that

$$f_{U,V}(u,v) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2(1-\rho^2)}[u^2+v^2-2\rho uv]\right). \tag{1.1}$$

Recall, for any pair of continuous random variables, the conditional density of $V|U=u$ is

$$f_{V|U}(v|u) = \frac{f_{U,V}(u,v)}{f_U(u)}$$

so that the joint density can be expressed as

$$f_{U,V}(u,v) = f_U(u) f_{V|U}(v|u). \tag{1.2}$$

Hence if we can find a factorization of the bivariate normal density (1.1) in the form (1.2) then we have derived the marginal density of U and the conditional density of $V|U=u$.

The key to the factorization is the completion of the square in the exponent as follows

$$\begin{aligned}
\frac{u^2+v^2-2\rho uv}{1-\rho^2} &= \frac{(u^2-\rho^2u^2)+(v^2-2\rho uv+\rho^2u^2)}{1-\rho^2}\\
&= \frac{u^2(1-\rho^2)+(v-\rho u)^2}{1-\rho^2}\\
&= u^2 + \frac{(v-\rho u)^2}{1-\rho^2}.
\end{aligned}$$

Substituting this into the exponent in equation (1.1) we get

$$\begin{aligned}
f_{U,V}(u,v) &= \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2}u^2 - \frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right)\\
&= \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)\right]\left[\frac{1}{\sqrt{2\pi}(1-\rho^2)^{1/2}}\exp\left(-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right)\right].
\end{aligned}$$

Now the first factor is the standard normal $N(0,1)$ density. The second factor is a density of a $N(\rho u, (1-\rho^2))$ random variable. We can therefore identify the first factor as the marginal density for U,

$$f_U(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$$

and the second factor as the conditional density for $V|U = u$,

$$f_{V|U}(v|u) = \frac{1}{\sqrt{2\pi}(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right).$$

This proves that the marginal density for U is the standard normal density, and the conditional density for $V|U = u$ is normal with (conditional) mean

$$E(V|U = u) = \rho u$$

and (conditional) variance

$$var(V|U = u) = 1 - \rho^2.$$

In summary

$$U \sim N(0,1)$$

and

$$V|U = u \sim N(\rho u, 1-\rho^2).$$

Notes:

1. The parameter ρ is the correlation between U and V and is easily derived as follows

$$\begin{aligned}
corr(U,V) &= \frac{cov(U,V)}{\sqrt{var(U)var(V)}} \\
&= cov(U,V), \quad \text{(since } U \text{ and } V \text{ have unit variance)} \\
&= \int\int uv f_{U,V}(u,v)dvdu \\
&= \int\int uv f_U(u) f_{V|U}(v|u)dvdu \\
&= \int u f_U(u) \left[\int v f_{V|U}(v|u)dv\right] du
\end{aligned}$$

But $\int v f_{V|U}(v|u)dv$ is the mean value for a random variable with the $N(\rho u, 1-\rho^2)$ density. Hence $\int v f_{V|U}(v|u)dv = \rho u$. Substituting this in the double integral we get

$$\begin{aligned}
corr(U,V) &= \int u f_U(u) \times \rho u \; du \\
&= \rho \int u^2 f_U(u)du.
\end{aligned}$$

But $\int u^2 f_U(u)du$ is $E\left(U^2\right)$ where $U \sim N(0,1)$ and hence $\int u^2 f_U(u)du = 1$ leading to

$$corr(U,V) = \rho.$$

This proves that for the bivariate normal density given by equation (1.1) the parameter ρ is the correlation between U and V.

2. When $\rho = 0$ the conditional density of $V|U$ simplifies to

$$f_{V|U}(v|u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) = f_V(v)$$

so that U, V are independent. You can also verify independence directly by considering what happens in expression (1.1) when $\rho = 0$. That is

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(u^2+v^2)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) \\ &= f_U(u) f_V(v). \end{aligned}$$

3. For both positive and negative $\rho \neq 0$, $var(V|U=u) = 1-\rho^2 < 1$ so that use of $U = u$ information to predict V using the conditional mean $E(V|U=u) = \rho u$ will lead to lower conditional variance for the prediction of V. That is when U and V are not independent, conditioning on one improves precision of prediction of the other.

4. If $|\rho| \to 1$, it means that correlation gets large in absolute value, and consequently the conditional variance $var(V|U=u) \to 0$. Therefore, once $U = u$ is known, we expected V to be equal to u or $-u$.

Properties of the General Multivariate Normal Distribution.

Some important facts about the general multivariate normal distribution are:

1. Any subvector of a multivariate normal vector has a multivariate normal distribution.

2. If $X \sim N(\mu, \Sigma)$ is an n-variate random variable, B is an $m \times n$ matrix of real numbers and a is a real $m \times 1$ vector then $Y = a + BX$ is an m-variate random variable and

$$Y = a + BX \sim N(a + B\mu_X, B\Sigma B').$$

3. In particular, any linear combination $b'X$ has a univariate normal distribution.

In general, consider a multivariate normal random vector $X \sim N(\mu, \Sigma)$. Partition X as

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\mu^{(j)} = E(X^{(j)})$ and $\Sigma_{ij} = E(X^{(i)} - \mu^{(j)})(X^{(j)} - \mu^{(j)})'$. Then

1. $X^{(1)}$ and $X^{(2)}$ are independent if and only if $\Sigma_{12} = 0$.

2. The conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is multivariate normal with conditional mean vector

$$\mu_{1|2} := E(X^{(1)}|X^{(2)} = x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})$$

and covariance matrix

$$\Sigma_{1|2} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

This is a very useful and important result.

1.10 Exercises

Exercise 1.1 (Mood et al. [1974]) *Let random variable X have a density function $f(\cdot)$, cumulative distribution function $F(\cdot)$, mean μ and variance σ^2. Define $Y = \alpha + \beta X$, where α and β are constants satisfying $\alpha \in \mathbb{R}$ and $\beta > 0$.*

i. *Select α and β so that Y has mean zero and variance 1.*

ii. *What is the correlation coefficient between X and Y?*

iii. *Find the cumulative distribution of Y in terms of α, β and $F(\cdot)$.*

iv. *Assume that X is symmetrically distributed about μ, which means $X - \mu$ and $-(X - \mu)$ have the same distribution, or ($X - \mu \stackrel{d}{=} -(X - \mu)$). Is Y necessarily symmetrically distributed about its mean?*

Exercise 1.2 (Härdle et al. [2024]) *Assume that the p-dimensional random vector Y has the following normal distribution: $Y \sim N(0, I)$, where I is the $p \times p$ identity matrix. Transform Y as $X = a + BY$ to create $X \sim N(\mu, \Sigma)$ with mean $\mu = (3, 2)'$ and $\Sigma = \begin{pmatrix} 1 & -1.5 \\ -1.5 & 4 \end{pmatrix}$.*

Exercise 1.3 (Härdle et al. [2024]) *Suppose that $X = (X_1, X_2)'$ has mean zero and covariance $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Let $Y = X_1 + X_2$. Write Y as a linear transformation of X, i.e., find the transformation matrix A, such that $Y = AX$. Then, compute $var(Y)$. Can you obtain the result in another fashion?*

Exercise 1.4 *Let $(X_1, X_2, X_3)^T$ have a multivariate normal distribution with zero mean vector and covariance matrix*

$$\Gamma = \begin{pmatrix} 1 & \phi & \phi^2 \\ \phi & 1 & \phi \\ \phi^2 & \phi & 1 \end{pmatrix}$$

where $|\phi| < 1$.

i. *Fully specify the distributions of*

   (a) $X_3 | X_2, X_1$

   (b) $X_3 | X_2$.

ii. *What can we conclude from Part (i)?*