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 Analysis (MATH5845)!

- **Lecturer: Dr. Atefeh Zamani**
  - **Education**
    - Ph.D. in Mathematical Statistics
    - Master in Data Science
  - **Research interest**
    - Functional data analysis
    - Time series analysis
    - Interval-valued time series
  - **Contact**

    atefeh.zamani@unsw.edu.au
    Please make sure to mention "MATH5845" in the subject of the email.

1.1.1 What is time series? Why we need it?

- *Time series* is a type of data that records the values of a variable over time, usually at regular intervals.

- We need time series because it allows us to *analyze the past* behavior and patterns of the variable, and to *forecast* its future values based on historical trends.

- Time series can be used for various purposes, such as understanding the seasonality and cyclicality of a phenomenon, detecting outliers and anomalies, identifying causal relationships between variables, and testing hypotheses and modeling.

Forecast or prediction? Why?

1.2 Structure and Resources

1.2.1 Topics

- **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 7**: Time series regression
- **Week 8**: Spectral analysis- Part 1
- **Week 9**: Spectral analysis- Part 2
- **Week 10**: Additional topics

1.2.2 Assessments

- **Quiz (10%)**: Week 4, 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: To **pass** the course, you must achieve

- at least 50% of the total mark

AND

- at least 40% in the final exam

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 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).

Why not the same as 2023 and 2024??!!

Assessment Item	Calculation	Result Profile	Assessed	Mean	Std Dev
FINAL: Final Result	up to date	FL PS CR DN HD	100	70.3%	16.1%
EXAM: Final Exam		FL PS CR DN HD	99	60.2%	20.3%
ASST1: Assignment 1		PS CR DN HD	99	86.4%	14.5%
ASST2: Assignment 2		FL PS CR DN HD	99	82.3%	18.9%

Assessment Item	Calculation	Result Profile	Assessed	Mean	Std Dev
FINAL: Final Result	up to date	PS CR DN HD	141	70.7%	12.4%
EXAM: Final Exam		FL PS CR DN HD	141	57.3%	18.7%
ASST1: Assignment 1		CR DN HD	142	89.1%	9.3%
ASST2: Assignment 2		CR DN HD	142	89.4%	10.9%

1.2.3 Lectures

3 Hours/Week

- **Lecture**

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)

  - Introduce topics; prove some results.
  - Answer questions as they arise. (Please ask!)
  - Revise topics and provide clarification.
  - Provide general feedback on assessments.

1.2.4 Tutorial

1 Hours/Week

- **Tutors**
  Atefeh Zamani and Rasool Roozegar
- **Tutorial Sessions**
  Attend **the tutorial you are enrolled in** due to the group project.
  - Demonstrate examples in `R`.
  - Some theoretical Exercises.

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)

1.2.5 Consultation Session

Day	Time	Location
Wednesday	15:00-16:00	Online
Friday	15:00-16:00	Online

Links to the online consultation sessions can be found on the Moodle page.

1.2.6 Moodle Site

- Post lecture slides in advance of the lecture.
- Answer questions on the discussion forum.
- Grades

1.2.7 Texts

- **Lecture Notes**
  - Notes by **William T. M. Dunsmuir (Emeritus Professor in UNSW)** with some modifications
- **Main textbooks**
  - Brockwell, P.J., & Davis, R.A. (2009). Time series: theory and methods. Springer science & business media.
  - Shumway, R.H., & Stoffer, D. S. (2017), Time series analysis and its applications. New York: springer.
- **Elementary resource**
  - Brockwell, P.J., & Davis, R.A. (Eds.). (2016). Introduction to time series and forecasting. New York, NY: Springer New York.

1.3 Expectations

1.3.1 Theory and Practice

- This course covers both **THEORY** and **PRACTICE**.
- Assessments and the final exam will include theory, proofs, and practical questions.
- Understanding of theoretical concepts is needed for application.
- Apply these concepts to real-world data using statistical software.
- Engage actively in lectures, tutorials, and group work.

1.3.2 Software

- This course is offered in `R`.
- Software demonstrations during tutorials.
- You can find a wide range of resources online.
  - Codes used by Shumway & Stoffer.
  - Forecasting: Principles and Practice, Hyndman and Athanasopoulos.

What about Python??

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

1.4 Review Notes on Multivariate Distributions

These notes will be relied on for basic facts about joint distributions of random variables as well as properties of the multivariate normal distribution. You might remember them form Multivariate analysis (MATH5845). Therefore, we are going to have a quick review.

1.4.1 Joint Distribution and Density Functions

- n-dimensional random vector: $X = (X_1, \ldots, X_n)'$,
- joint distribution function: $F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$,
- joint distribution of any sub-vector can be obtained by setting $x_i = \infty$ for the other random variables not in the sub-vector.
  - the distribution of X_1: $F_{X_1}(x_1) = F(x_1, \infty, \ldots, \infty)$
  - the joint distribution of (X_i, X_j):

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

- A random vector is said to be **continuous** if F 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

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

Note that

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

1.4.2 Independence

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

$$P(X_1 \le x_1, \dots, X_n \le x_n) = P(X_1 \le x_1) \cdots P(X_n \le x_n)$$

for all $x_1, \dots, x_n \in \mathbb{R}$, or equivalently,

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

or

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

Note 1.1 *For random variables $X_1, \dots X_n$,*

- *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.*

1.4.3 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

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

- If X and Y are independent, then

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

  - $f_{Y|X}(y|x) = f_Y(y)$
  - 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.

1.4.4 Expected Values.

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

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

Let X be a **univariate** random variable. Then,

- Mean of X ($\mu = E(X)$): $g(X) = X$,
- Variance of X ($\sigma^2 = var(X) = E(X - \mu)^2$): $g(X) = (X - \mu)^2$.

Note that

$$\begin{aligned} E(aX + b) &= aE(X) + b, \\ var(aX + b) &= a^2 var(X). \end{aligned}$$

For two **univariate** random variables X and Y,

- Covariance between X and Y: $cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$,
- Correlation between X and Y:

$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}.$$

1.4.5 Means and Covariances for Random Vectors

For the multivariate random variable $X = (X_1, \ldots, X_n)'$:

- Mean vector: $\mu_X = E(X) = (E(X_1), \ldots, E(X_n))'$,

For two random vectors $X = (X_1, \ldots, X_n)'$ and $Y = (Y_1, \ldots, Y_n)'$

- Covariance matrix between X and Y:

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

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

- When $Y = X$, $cov(X, X) = var(X) = E(X - EX)(X - EX)'$.

- If X and Y are independent then $cov(X, Y) = \mathbf{0}$ (null (zero) matrix) and we call them uncorrelated.

  - The converse is not true in general but is true for the multivariate normal distribution.

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

- $\mu_Y = E(Y) = a + BE(X) = a + B\mu_X$,
- $\Sigma_{YY} = B\Sigma_{XX}B'$.
  - Any variance matrix Σ is non-negative definite, i.e., $b'\Sigma b \geq 0$ for any vector b.
    Proof: Let $Y = b'X$ where X has covariance matrix Σ. Then
$$0 \leq var(Y) = b'\Sigma b.$$

1.4.6 The Multivariate Normal Distribution.

The general multivariate normal density

The random vector X has the **multivariate normal distribution** with mean μ and non-singular (square matrix with non-zero determinant/invertible) 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

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

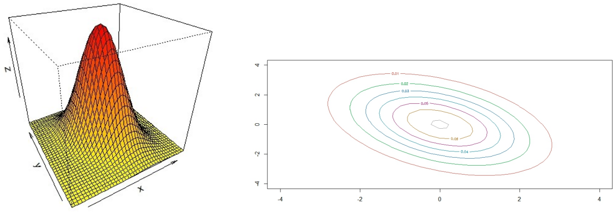
$$f_X(x) = \frac{1}{2\pi[\sigma_1^2\sigma_2^2(1-\rho^2)]^{1/2}} \exp\left\{-\frac{1}{2}Q(x_1, x_2; \sigma_1, \sigma_2, \rho)\right\}$$

with quadratic form

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

Some important facts about the bivariate normal are:

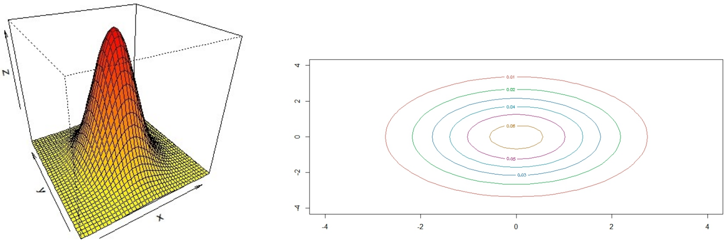
1. The contours of equal density are ellipses $\{(x_1, x_2) : Q(x_1, x_2; \sigma_1, \sigma_2, \rho) = k\}$ for any constant $k \geq 0$.



(a) (b)

(a) 3-D density function (b) contour plot of bivariate normal distribution $X \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 2 & -1 \\ -1 & 3 \end{array}\right)\right)$.

2. When the correlation $\rho = 0$, the two random variables X_1 and X_2 are independent.



(a) (b)

(a) 3-D density function (b) contour plot of bivariate normal distribution $X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\right)$.

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 where

$$\begin{aligned} \mu_U &= \mu_V = 0, \\ \sigma_U^2 &= \sigma_V^2 = 1, \end{aligned}$$

so that

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

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

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

- 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 $V|U = u$.

- Complete the square in the exponent:

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

- Substituting this into the exponent in equation (1.1):

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

– The first factor as the marginal density for U,

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

– The second factor is the conditional density for $V|U = u$,

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

In summary,

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

and

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

Notes:

1. 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 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, i.e.,

$$f_{U,V}(u, v) = f_U(u) f_V(v).$$

3. For both positive and negative $\rho \neq 0$, $var(V|U = u) = 1 - \rho^2 < 1$. So, when U and V are not independent, conditioning on one improves precision of prediction of the other.

4. If $|\rho| \to 1$, it means $var(V|U = u) \to 0$. Therefore, once $U = u$ is known, we expected V to be equal to u or $-u$.

Properties of the General Multivariate Normal Distribution.

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

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

   - Any linear combination $b'X$ has a univariate normal distribution.

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

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

and covariance matrix

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$