Introduction to Time Series Modelling

Sean Hellingman ©

Regression for Applied Data Science (ADSC2020)

shellingman@tru.ca

Winter 2025

THOMPSON RIVERS UNIVERSITY

Topics

Introduction

- So far, we have not covered any methods for dealing with data that have been observed at different points in time.
- In regression approaches, we have assumed observations are independent and identically distributed.
- By introducing time into our models, the independence assumption is in serious jeopardy.

Definitions

Stochastic Process

- A **stochastic process** is a collection of random variables indexed by time.
- Analysis often focuses on the evolution of this process over time.
- Some common methodology:
  - Markov Chains
  - **ARIMA** (time series)
  - Machine learning

Time Series Data

- A **time series** is sequence of observations ordered by an index.
- Series of data measurements on the same entity.
- Used in many fields:
    - Stock prices
    - Sales
    - Environmental readings
    - Social statistics
    - Birth rates

Example 1

1. Import the *tempdub* time series object from the TSA package into R.
2. Use base R to plot the time series.
3. What do you notice?
4. Will linear regression work?

Mean Function

- The **mean function** is the expected value of the process at time t.

$$\mu_t = E(Y_t) \quad \text{for } t = 0 \pm 1, \pm 2, \ldots \tag{1}$$

- In general, μ_t can be different at each time point t.

Autocorrelation

- The **autocorrelation function** is a measure of the linear correlation between values of the process at different times.

$$Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} \quad \text{for } t, s = 0 \pm 1, \pm 2, ... \tag{2}$$

- In other words, how linearly related values are across time.

White Noise Processes

- A **white noise** process is a random variable indexed in time that has a constant expected value and variance.
- Properties:
  - $E(e_t) = \mu_e$
  - $Var(e_t) = \sigma_e^2$
  - $\rho_{t,s} = 0 \quad \text{if } t \neq s$
- *Similar to the error term we find in regression.*

Example 2

1. Set your seed and simulate a white noise process from a normal distribution.
2. Use base R to plot the time series.
3. What do you notice?

Random Walk Process

- A **random walk** process is a random variable indexed in time that depends on the previous observation and a white noise term.

$$Y_t = Y_{t-1} + e_t \tag{3}$$

- Where e_t follows a white noise process.

Example 3

1. Set your seed and simulate a random walk process using your white noise term from Example 2.
2. Use base R to plot the time series.
3. What do you notice?

Moving Average Process

- A **moving average** process is a random variable indexed in time that partially depends on the current and previous white noise terms.

$$Y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + e_t \tag{4}$$

- Where e_t follows a white noise process.
- MA(3) Example: $Y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + e_t$

Example 4

1. Set your seed and adjust your code to simulate the following MA(1) process:

   $Y_t = 2 + \theta_1 e_{t-1} + e_t$

   Where $\theta_1 = 0.7$

2. Use base R to plot the time series.

3. What do you notice?

Autoregressive Process

- An **autoregressive** process is a random variable indexed in time that partially depends on the previous values and a white noise terms.

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t \tag{5}$$

- Where e_t follows a white noise process.
- AR(3) Example: $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$

Example 5

1. Set your seed and adjust your code to simulate the following AR(2) process:

   $Y_t = 1 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$

   Where $\phi_1 = 0.7$ and $\phi_2 = -0.2$

2. Use base R to plot the time series.

3. What do you notice?

Stationarity

- A process is said to be **strictly stationary** if the joint distribution of Y_{t_1}, Y_{t_2}, ..., Y_{t_n} is the same as Y_{t_1-k}, Y_{t_2-k}, ..., Y_{t_n-k}.
  - For all time points t_1, t_2, ..., t_n and all time lags k.

- In other words, the nature of the process does not change over time.

- If a process is strictly stationary and has finite variance, then the variance function must depend only on the time lag.

Weak Stationarity

- A stochastic process is said to be **weakly stationary** if:
  1. The mean function is constant over time.
  2. $\gamma_{t,t-k} = \gamma_{0,k}$ for all t and k.
- *We will refer to weak stationarity during this course.*

Example 6

1. Import the AirPassengers from the *TSA* package and take a moment to understand the data.
2. Use base R to plot the time series.
3. Compare this plot with other plots that you have made.
4. What do you notice?

Exercise 1

- Take some time to examine time series data (real and simulated).
- Do any of these processes seem to be stationary?

References & Resources

1. Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Link

- arima.sim()
- TSA
- zoo