Time Series Modeling: AR, MA, ARIMA — Worksheet

Learning Objectives

By the end of this worksheet, you should be able to:

- Understand **lag**, **difference of lag**, and **autocorrelation**.
- Explain what **ACF** and **PACF** represent.
- Use ACF and PACF conceptually to reason about **AR**, **MA**, and **ARIMA**.
- Reason about time-dependent data without writing code.

1. Time Series Basics

Why do we need time series models?

Time series models are used when the order of observations matters. Past values influence future values, so shuffling the data breaks this relationship.

What is a time series?

A time series is a sequence of observations recorded over time.

Important terms:

- **Lag**: value from a previous time step (e.g., X_{t-1})
- **Trend**: long-term increase or decrease
- **Seasonality**: repeating pattern over fixed intervals
- **Noise**: random variation
- **Stationarity**: stable mean and variance over time

Autocorrelation

Autocorrelation measures how strongly a time series is related to its own past values.

If autocorrelation at lag k is high, it means:

$$X_t \text{ is strongly related to } X_{t-k}$$

Most time series models are built by studying this dependence on past values.

ACF (Autocorrelation Function)

The **Autocorrelation Function (ACF)** shows the correlation between the series and its lagged versions for different lag values.

In simple terms:

- ACF answers: *"How similar is the series to itself after k time steps?"*
- Each spike in the ACF corresponds to correlation at a specific lag

ACF is mainly used to:

- Understand the overall dependence structure
- Identify the order q in MA models

Basic workflow

- Visualize the data
- Check whether the series is stationary
- Examine autocorrelation using ACF
- Use differencing if required
- Select a suitable model

2. AR Models (Autoregressive)

Why use AR?

AR models are useful when the present value depends directly on past values.

Model form

AR(p):

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

AR, ACF, and PACF

For AR models:

- Dependence is on past values
- ACF typically decreases gradually
- PACF shows a sharp cutoff after lag p

PACF (Partial Autocorrelation Function)

PACF measures the direct relationship between X_t and X_{t-k} after removing the effect of intermediate lags.

PACF is mainly used to decide the value of p in AR models.

3. MA Models (Moving Average)

Why use MA?

MA models are useful when the series reacts mainly to sudden, short-term shocks.

Model form

MA(q):

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

MA, ACF, and PACF

For MA models:

- Dependence is on past errors (shocks)
- ACF cuts off sharply after lag q
- PACF decreases gradually

4. ARIMA Models

Why ARIMA?

Many real-world time series are non-stationary due to trends or changing levels. ARIMA handles this by combining differencing with AR and MA components.

Model form

ARIMA(p, d, q):

- p: AR order (from PACF)
- d: number of differences
- q: MA order (from ACF)

Differencing:

$$Y_t = X_t - X_{t-1}$$

How ARIMA is applied

- Difference the series until it becomes stationary
- Use ACF and PACF on the differenced series
- Fit AR and MA components together

5. Application Questions

Q1. Daily electricity consumption in a hostel is usually similar to the previous day. Which model (AR or MA) is more suitable? Explain using lag.

Q2. A machine produces stable output most days, but sudden power failures affect output for a few days. Which model would capture this behaviour better? Explain.

Q3. "The value at time t depends on how much the value changed from time $t-1$ to t."

Does this statement refer to a lag or a lag difference? Write the corresponding mathematical expression.

Q4. A time series shows a steady upward trend. You compute:

$$Y_t = X_t - X_{t-1}$$

(a) What does this operation represent?

(b) Why is it useful before applying AR or MA?

Q5. The ACF of a stationary time series shows significant spikes up to lag 2 and then becomes insignificant.

(a) What does this indicate about the dependence structure?

(b) Which type of model does this behaviour suggest?

Q6. A non-stationary series becomes stationary after first differencing. The ACF cuts off at lag 1, while the PACF decays slowly.

(a) Why is differencing needed?

(b) Which ARIMA model family would be appropriate?

Q7. Why does shuffling time series data make ACF and lag-based reasoning meaningless?