ARIMA

1. Introduction

The ARIMA model, which stands for **AutoRegressive Integrated Moving Average**, is a foundational statistical approach used for modeling and forecasting univariate time series data. It is particularly effective when the underlying process exhibits temporal dependencies such as autocorrelation and trend.

2. Model Notation and Structure

An ARIMA model is formally denoted as ARIMA(p, d, q), where:

- p is the order of the autoregressive (AR) part,
- d is the degree of differencing required to make the series stationary,
- q is the order of the moving average (MA) part.

The general form of an ARIMA(p, d, q) model is:

$$\Phi_p(B)\,\nabla^d y_t = \Theta_q(B)\varepsilon_t$$

where:

- B is the backshift operator, such that $By_t = y_{t-1}$,
- $\nabla^d = (1 - B)^d$ is the differencing operator,
- $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the AR polynomial,
- $\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ is the MA polynomial,
- ε_t is white noise: $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

3. Components of the Model

3.1 Autoregressive (AR) Component

This part models the current value of the time series as a linear function of its previous values:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

The coefficients ϕ_i capture the influence of lagged observations.

3.2 Integrated (I) Component

Differencing is used to transform a non-stationary time series into a stationary one. The d parameter indicates the number of times the series is differenced.

First-order differencing:

$$\nabla y_t = y_t - y_{t-1}$$

Second-order differencing:

$$\nabla^2 y_t = \nabla(\nabla y_t) = y_t - 2y_{t-1} + y_{t-2}$$

3.3 Moving Average (MA) Component

This component models the current value as a function of past forecast errors:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

The θ_j parameters represent the impact of prior noise terms on the current value.

4. Stationarity and Invertibility

- The series must be made stationary via differencing before applying AR or MA models.
- The AR polynomial must be stationary: roots of $\Phi_p(B) = 0$ must lie outside the unit circle.
- The MA polynomial must be invertible: roots of $\Theta_q(B) = 0$ must also lie outside the unit circle.

5. Model Estimation and Selection

ARIMA model parameters are usually estimated via **Maximum Likelihood Estimation (MLE)** or conditional sum-of-squares.

To select optimal values of (p, d, q), the following criteria are commonly used:

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

6. Forecasting with ARIMA

Once the model is fitted, it can be used to make:

- **Point forecasts:** future values based on estimated model parameters.
- **Prediction intervals:** confidence intervals based on forecast error variance.

Forecasting is recursive and accounts for both autoregressive structure and error terms.

7. Extensions of ARIMA

- **SARIMA (Seasonal ARIMA):** Adds seasonal AR, I, and MA terms.
- **ARIMAX:** Includes exogenous variables (external regressors).
- **VARIMA:** Generalization for multivariate time series.

8. Applications

ARIMA models are widely used in:

- Financial market prediction (e.g., stock prices, returns)
- Economic forecasting (e.g., GDP, inflation)
- Demand forecasting (e.g., sales, inventory)
- Environmental data modeling (e.g., temperature, rainfall)

9. Advantages and Limitations

Advantages

- Interpretable linear model
- Effective for short-term forecasting
- Applicable to many real-world stationary processes

Limitations

- Requires data to be stationary
- Cannot model nonlinear relationships
- Performance may degrade on long-term or volatile forecasts

10. Conclusion

The ARIMA model remains one of the most powerful and interpretable tools for analyzing and forecasting univariate time series data. With proper stationarity checks and parameter tuning, it can provide reliable short- to medium-term forecasts in a wide range of applications.