

# AR Model

Masaru Okada

October 8, 2025

## Abstract

An overview of the AR model, which is used for analyzing linear time series data, is provided.

## 1 Linear Prediction Models

There are various types of linear prediction models, including:

1. AR (**A**uto **R**egression) Model
2. MA (**M**oving **A**verage) Model
3. ARMA (**A**utoregressive **M**oving **A**verage) Model, a combination of AR and MA
4. ARIMA (**A**utoregressive **I**ntegrated **M**oving **A**verage) Model, which includes differencing

While these linear prediction models are useful for analyzing time series data within a linear framework, they have a major drawback: they cannot handle nonlinear dynamics, such as chaos.

This note will focus on one of the most fundamental linear prediction models, the AR model.

## 2 Stochastic and Deterministic Processes

A time series consisting of  $N$  observations is represented as:

$$\{x(t)\}_0^{N-1} = \{x(t_0), x(t_0 + \Delta t), x(t_0 + 2\Delta t), \dots, x(t_0 + (N-1)\Delta t)\}$$

Here,  $t$  is an integer value from 0 to  $N-1$ .

$\Delta t$  represents the time interval between adjacent observations.

$t_0$  is the initial time of observation, and the data ends at  $t_0 + (N-1)\Delta t$ .

This means we are observing only a small part of the possible behaviors.  $\{x(t)\}_0^{N-1}$  is just one sample.

Therefore, even if we analyze  $\{x(t)\}_0^{N-1}$  to estimate the time series, there is always a possibility that the results deviate from the true underlying properties.

Evaluating how close the estimation results are to the true properties is a significant challenge in itself.

Each data point in  $\{x(t)\}_0^{N-1}$  is considered a value of a random variable,  $X$ , realized with a certain probability.

A dynamic process of time evolution modeled by probability is called a **stochastic process**.

If the probability of realization at each time is equal to 1, the process is said to be **deterministic**.

The state changes of a real system should inherently be deterministic, following physical laws or some causal principles. It seems the only process that is not fundamentally causal is the collapse of the wave function in quantum mechanics.

However, state changes are often determined by a vast number of factors, or by uncontrollable external influences, making it practically impossible to set the internal factors precisely.

In such cases, a lack of knowledge arises when analyzing the dynamics of time evolution based on data. As knowledge decreases, the observed behavior appears to move further away from a deterministic process.

### 3 Stationary States

A **stationary stochastic process**, or a system in a **stationary state**, means the time series doesn't continuously increase or decrease, but rather remains around a constant level, signifying a statistical equilibrium.

A system becomes a stationary stochastic process when its probability density function doesn't depend on time.

Specifically, a process is called (**strongly**) **stationary** if a time series starting at time  $t_0$ ,  $\{x(t_0)\}_0^{N-1}$ , and a time series starting at a time delayed by  $T$ ,  $\{x(t_0 - T)\}_0^{N-1}$ , have the same joint probability distribution.

$$F[\{x(t_0)\}_0^{N-1}] = F[\{x(t_0 - T)\}_0^{N-1}]$$

### 4 Linear Dynamics

To simplify notation, let's assume each element of  $\{x(t)\}_0^{N-1}$  has a mean of zero,  $\mu = 0$ . In other words, the following processing has already been performed:

$$x(t) \leftarrow x(t) - \mu$$

Next, let's form a vector from the sequence of observations.

$$\mathbf{x}(t) = (x(t), x(t-1), x(t-2), \dots, x(t-(D-1)))$$

$D = \infty$  implies going back to the infinite past.

Suppose the future behavior at  $\tau$  steps from the present,  $x(t+\tau)$ , can be expressed using a mapping  $F$  that represents the state change:

$$x(t+\tau) = F[\mathbf{x}(t)]$$

This  $F$  is a **linear dynamic** if it satisfies:

$$F(a\mathbf{x} + b\mathbf{y}) = aF(\mathbf{x}) + bF(\mathbf{y})$$

This is the dynamic process we will analyze.

If  $F$  is not a linear dynamic, it's called a **nonlinear dynamic**. I'll cover this in a separate note.

## 5 Wold's Decomposition Theorem

Any stationary process  $\{z(t)\}$  can be expressed as the sum of a deterministic stationary process  $\{y(t)\}$  and a non-deterministic stationary process  $\{x(t)\}$  [1].

$$z(t) = x(t) + y(t)$$

This is known as Wold's Decomposition Theorem.

Our interest lies in the non-deterministic stationary process,  $x(t)$ .

A non-deterministic stationary process is given by an infinite series of white noise,  $\xi(t)$ , with a zero mean and a variance of 1.

$$x(t) = \xi(t) + \sum_{i=0}^{\infty} a_i \xi(t-i)$$

The coefficients of this infinite series are absolutely convergent:

$$\sum_{i=0}^{\infty} |a_i| < \infty$$

Using duality, it can also be rewritten as:

$$x(t) = \xi(t) + \sum_{i=0}^{\infty} c_i x(t-i)$$

The coefficients of this infinite series are also absolutely convergent:

$$\sum_{i=0}^{\infty} |c_i| < \infty$$

Both equations contain infinite series, so they cannot be used directly for time series forecasting.

Therefore, we approximate the stochastic process by truncating the series to a finite number of terms. This results in the **AR process**.

## 6 Autoregressive (AR) Model

An **autoregressive process** (AR process) is represented by:

$$x(t) = \xi(t) + c_1 x(t-1) + c_2 x(t-2) + \cdots + c_p x(t-p)$$

A process with  $p$  coefficients like the equation above is called a  $p$ -th order AR process. These  $p$  coefficients are called **AR parameters**. A model approximated by  $p$  AR parameters is called an **AR( $p$ ) model**.

Specifically, an AR(1) process with a coefficient close to 1 is called a **unit root**.

## 7 AR Model Estimation

### 7.1 Yule-Walker Equations

Starting with the AR( $p$ ) process equation:

$$x(t) = \xi(t) + c_1 x(t-1) + c_2 x(t-2) + \cdots + c_p x(t-p)$$

Multiplying both sides by  $x(t - \tau)$  gives:

$$x(t)x(t - \tau) = \xi(t)x(t - \tau) + c_1x(t - 1)x(t - \tau) + \cdots + c_px(t - p)x(t - \tau)$$

We define the autocovariance as:

$$\gamma(\tau) = E[x(t)x(t - \tau)]$$

Taking the expected value  $E[\cdot]$  of both sides for  $\tau = 0$ , we get:

$$x(t)^2 = \sigma + c_1x(t - 1)x(t) + c_2x(t - 2)x(t) + \cdots + c_px(t - p)x(t)$$

which simplifies to:

$$\gamma(0) = c_1\gamma(1) + c_2\gamma(2) + \cdots + c_p\gamma(p)$$

Here, the variance of the white noise is denoted as  $E[\xi(t)x(t)] = \sigma$ .

Using the fact that  $E[\xi(t)x(t - \tau)] = 0$  for  $\tau > 0$ , we obtain:

$$\gamma(\tau) = c_1\gamma(\tau - 1) + c_2\gamma(\tau - 2) + \cdots + c_p\gamma(\tau - p)$$

By substituting  $\tau = 1, 2, \dots, p$  into this equation, we can derive the system of linear equations known as the **Yule-Walker equations**.

$$\begin{aligned} \sigma &= \gamma(0) - c_1\gamma(1) - c_2\gamma(2) - \cdots - c_p\gamma(p) \\ \gamma(1) &= c_1\gamma(0) + c_2\gamma(1) + c_3\gamma(2) + \cdots + c_p\gamma(p - 1) \\ \gamma(2) &= c_1\gamma(1) + c_2\gamma(0) + c_3\gamma(1) + \cdots + c_p\gamma(p - 2) \\ &\vdots \\ \gamma(p) &= c_1\gamma(p - 1) + c_2\gamma(p - 2) + c_3\gamma(p - 3) + \cdots + c_p\gamma(0) \end{aligned}$$

Note that the symmetry  $\gamma(k) = \gamma(-k)$  has been used.

By solving this system of equations to find the AR parameters, we can determine the AR model.

Solving for the AR parameters allows for the best approximation of the **linear predictor**:

$$\hat{x}(t) = c_1x(t - 1) + c_2x(t - 2) + \cdots + c_px(t - p)$$

The prediction error of the linear predictor is given by:

$$H_{\text{lin}} = \frac{1}{N} \sum_{t=p}^{n-1} \left( x(t) - \hat{x}(t) \right)^2$$

The AR coefficients that minimize  $H_{\text{lin}}$  are found by setting the partial derivative with respect to each coefficient to zero,  $\frac{\partial H_{\text{lin}}}{\partial c_i} = 0$ . These coefficients are the same as those obtained by solving the Yule-Walker equations.

When the order  $p$  is high, a fast algorithm is needed to solve the Yule-Walker equations.

Algorithms like the Levinson-Durbin algorithm are known for their efficiency [4], [5].

## References

- [1] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis; Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.
- [2] Yule, G. Udny (1927), “On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers” , Philosophical Transactions of the Royal Society of London, Ser. A 226: 267 – 298
- [3] Walker, Gilbert (1931), “On Periodicity in Series of Related Terms” , Proceedings of the Royal Society of London, Ser. A 131: 518 – 532 Hamilton (1994), p. 59
- [4] Levinson, N. (1947). ”The Wiener RMS error criterion in filter design and prediction.” J. Math. Phys., v. 25, pp. 261 – 278.
- [5] Durbin, J. (1960). ”The fitting of time series models.” Rev. Inst. Int. Stat., v. 28, pp. 233 – 243.