Lecture 4

Autoregressive models

Review

- **Time series**
- **Stochastic process**
- **Measure of dependence**

  Mean function

  Autocovariance function
- **Stationarity**

  Strict stationarity

  Weak stationarity

WHY?

It makes "prediction" possible.

The Guarantee of Stationarity: the statistics from single time series samples can become valid estimates of the true population properties

For stationary processes, we can construct models such as AR, MA, and ARMA models.

Review

- $\{x_t\}$ **is strictly (or strongly) stationary** if

  $(x_{t_1}, \ldots, x_{t_k})$ and $(x_{t_1+h}, \ldots, x_{t_k+h})$ have the same joint distribution

  for all $k, t_1, \ldots, t_k, h$

Selecting an arbitrary time window:

Suppose we arbitrarily choose k time points $(t_1, t_2, \ldots, t_k)$

- **That is**

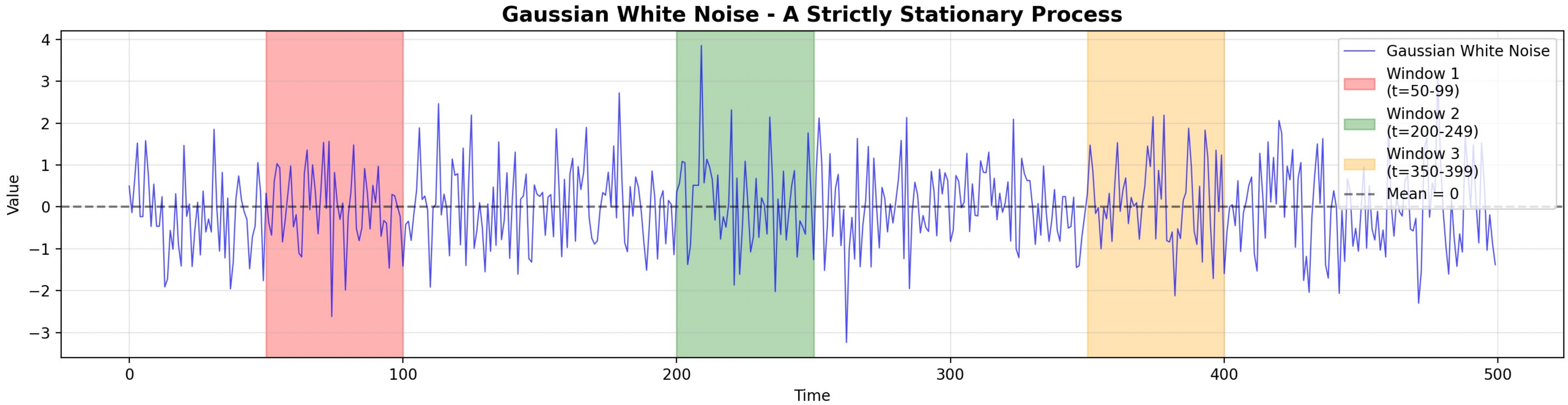
$$\mathrm{P}\{x_{t_1} \le c_1, \ldots, x_{t_k} \le c_k\} = \mathrm{P}\{x_{t_1+h} \le c_1, \ldots, x_{t_k+h} \le c_k\}$$

$$\Downarrow$$

$$\mu_t = \mu_s, \quad \forall s, t$$

$$\gamma(s,t) = \gamma(s+h, t+h)$$

Review

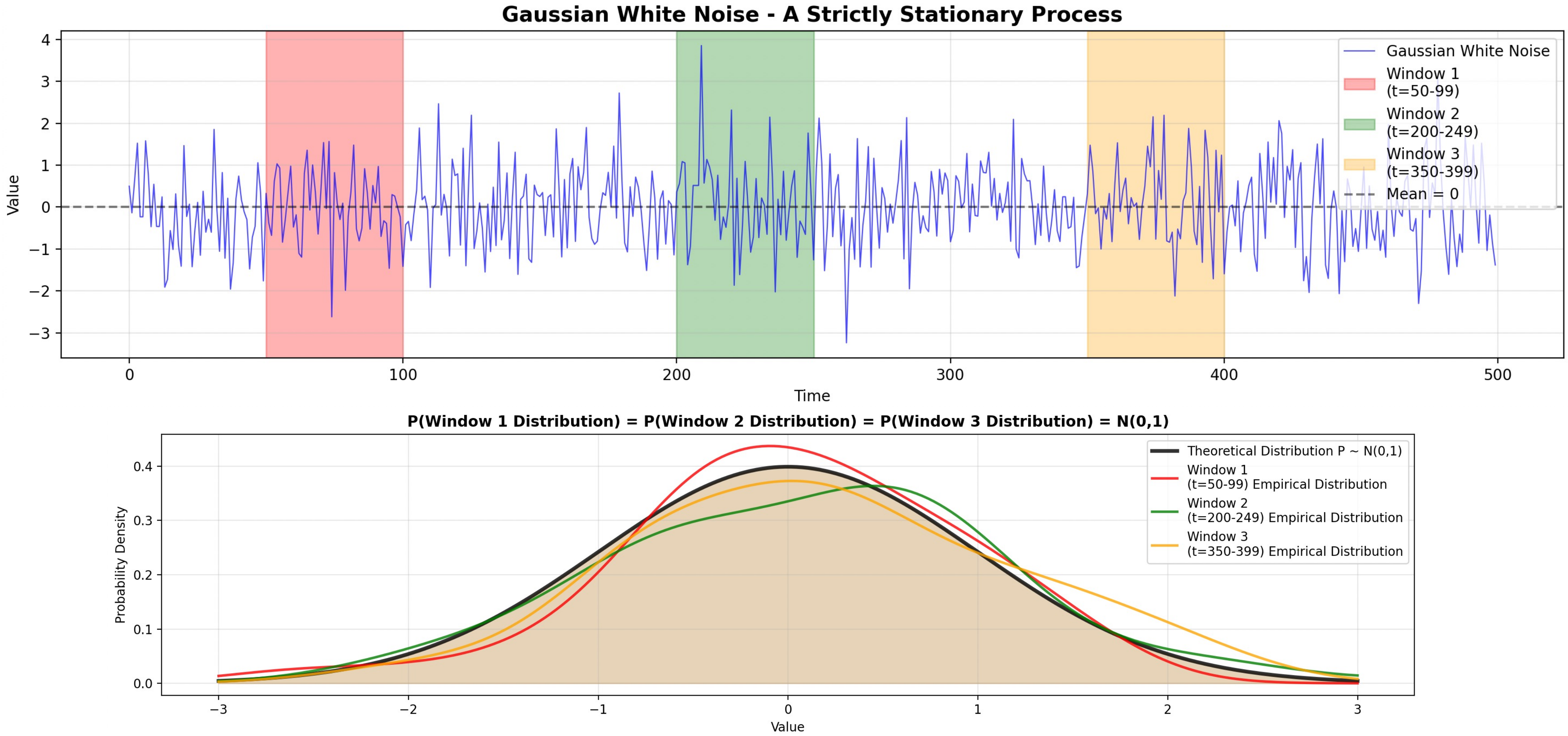


$(x_{t_1}, \ldots, x_{t_k})$ and $(x_{t_1+h}, \ldots, x_{t_k+h})$ have the same joint distribution

for all $k, t_1, \ldots, t_k, h$

Performing an arbitrary time shift: Now, we shift this entire time window backward (or forward) by h units, obtaining a new set of time points (t_1+h, t_2+h, ..., t_k+h). The data at this new set of points (x_{t1+h}, x_{t2+h}, ..., x_{tk+h}) also has a joint probability distribution.

Review



Gaussian White Noise - A Strictly Stationary Process
Gaussian White Noise
Window 1 (t=50-99)
Window 2 (t=200-249)
Window 3 (t=350-399)
Mean = 0
Value
Time
4
3
2
1
0
−1
−2
−3
0
100
200
300
400
500
P(Window 1 Distribution) = P(Window 2 Distribution) = P(Window 3 Distribution) = N(0,1)
Theoretical Distribution P ~ N(0,1)
Window 1 (t=50-99) Empirical Distribution
Window 2 (t=200-249) Empirical Distribution
Window 3 (t=350-399) Empirical Distribution
Probability Density
Value
0.4
0.3
0.2
0.1
0.0
−3
−2
−1
0
1
2
3

Review

- **That is**

$$P\{x_{t_1} \leq c_1, \ldots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \ldots, x_{t_k+h} \leq c_k\}$$

Arbitrarily chosen thresholds ($c_1, c_2, \ldots, c_k$)

For any possible k (window size), any set of starting time points ($t_1, \ldots, t_k$), and any shift amount h, the joint distributions of these two sets of data (before and after the shift) must be identical.

Core Idea:

Statistical properties remain entirely unchanged over time.

(mean, variance, skewness, kurtosis, quantiles, extreme value probabilities, etc.)

Review

$$\mu_t = \mu_s, \quad \forall s, t$$

$$\gamma(s,t) = \gamma(s+h, t+h)$$

The mean is constant for all time points s and t （**constant over time**）

The covariance depends only on the time lag

Here, $\gamma(s, t)$ typically represents the covariance between x_t and x_s, that is, $\mathrm{Cov}(x_t, x_s)$.

The covariance $\gamma(t, s)$ between two time points (e.g., t and s) depends only on the time difference $|t - s|$, not on their absolute positions on the time axis.

Review

- $\{x_t\}$ **is weakly stationary** if

  μ_t is independent of t; and

  $\gamma(t+h,t)$ is independent of t for each h

$$\mu_t = \mu$$

$$\gamma(t+h,t) = \gamma(h,0) = \gamma(h)$$

Review

- $\{x_t\}$ **is weakly stationary** if

  μ_t is independent of t; and

  $\gamma(t+h, t)$ is independent of t for each h

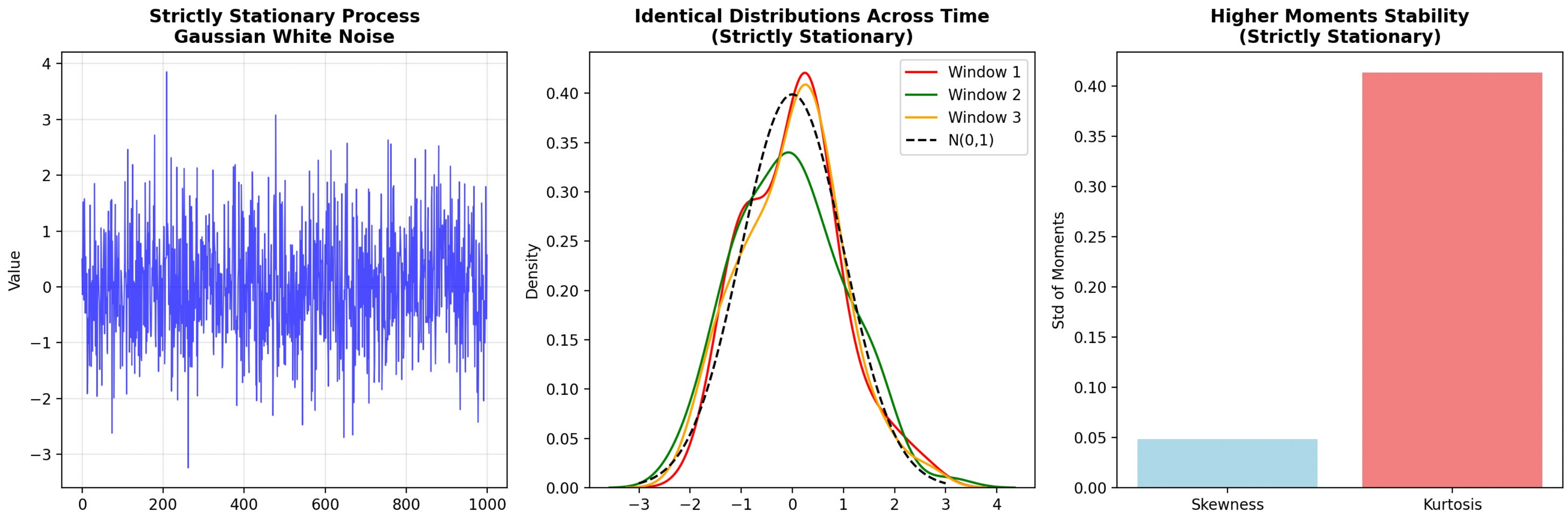
$$\Downarrow$$

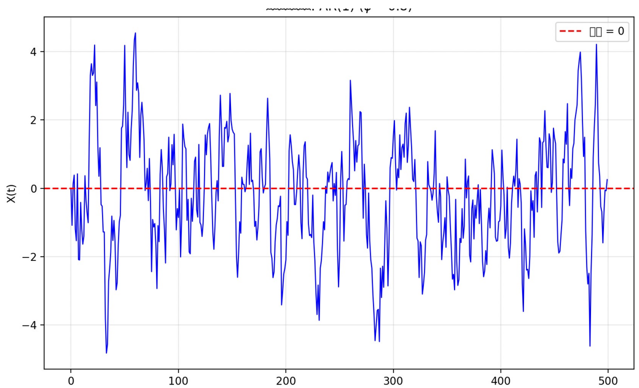
$$\mu_t = \mu$$

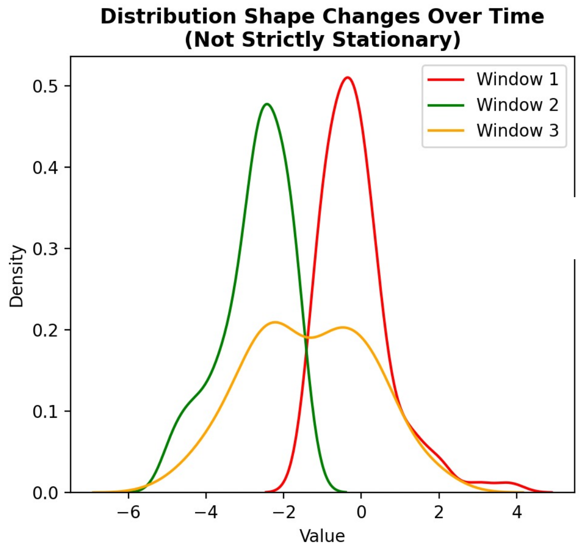
$$\gamma(t+h, t) = \gamma(h, 0) = \gamma(h)$$

the covariance depends only on the time interval *h*, and not on the specific time point *t*.

For example, the covariance between "today and yesterday" should be equal to the covariance between "yesterday and the day before yesterday," and also equal to the covariance between "tomorrow and today." This is because their time interval *h* is 1 in all cases.







Only the first two moments (mean, variance

Feature	Strict Stationarity	Weak Stationarity
Mean	Not explicitly required, but as a consequence, the mean is constant (if it exists).	$E[X_t] = \mu$ (constant for all t).
Variance	Not explicitly required, but as a consequence, the variance is constant (if it exists).	$\mathrm{Var}(X_t) = \sigma^2$ (constant for all t).
Core Definition	For any set of times $t_1, t_2, ..., t_n$ and any time shift k, the joint distribution satisfies: $F_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_1, x_2, ..., x_n) = F_{X_{t_1+k}, X_{t_2+k}, ..., X_{t_n+k}}(x_1, x_2, ..., x_n)$ (All finite-dimensional joint distributions are unchanged).	1. $E[X_t] = \mu$ 2. $\mathrm{Cov}(X_t, X_{t+k}) = \gamma(k)$ (depends only on the lag k, not on t).
Focus	The complete probability distribution.	Only the first two moments (mean, variance, covariance).

Review

- **Basic properties of** $\gamma(\cdot)$

  1. $\gamma(0) \geq 0$
  2. $|\gamma(h)| \leq \gamma(0), \ \forall h$
  3. $\gamma(h) = \gamma(-h), \ \forall h$

Review

- **Basic properties of $\gamma(\cdot)$**

1. $\gamma(0) \geq 0$
2. $|\gamma(h)| \leq \gamma(0), \ \forall h$
3. $\gamma(h) = \gamma(-h), \ \forall h$

r(0) denotes the autocorrelation of a time series with itself at lag zero.

$$\gamma(0) = \mathrm{Cov}(X_t, X_t) = E[(X_t - \mu)(X_t - \mu)] = E[(X_t - \mu)^2] = \mathrm{Var}(X_t)$$

$$\gamma(0) = \mathrm{Var}(X_t) = \sigma^2$$

Review

- **Basic properties of** $\gamma(\cdot)$

1. $\gamma(0) \geq 0$
2. $|\gamma(h)| \leq \gamma(0), \ \forall h$
3. $\gamma(h) = \gamma(-h), \ \forall h$

The autocovariance function γ(h) is an even function.

Review

- **The autoccorelation function (ACF) of a stationary time series $\{x_t\}$ is defined as**

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\mathrm{Cov}(x_{t+h}, x_t)}{\mathrm{Cov}(x_t, x_t)} = \mathrm{Corr}(x_{t+h}, x_t)$$

$$\Downarrow$$

$$-1 \leq \rho(h) \leq 1$$

γ(h) is the **autocovariance function**, namely **Cov(X_t, X_{t+h})**.

γ(0) is the **variance** of the series, namely **Var(X_t)**.

By standardizing through division by the variance γ(0), the range of the ACF is constrained to **[-1, 1]**.
This enables the comparison of correlation strengths across different time series.

Review

- **For observations** $x_1, \ldots, x_n$ **of a time series, the sample mean**

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

- **The sample autocovariance function is**

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad \text{for } -n < h < n$$

- **The sample autocorrelation function (sample ACF) is**

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad \text{for } -n < h < n$$

Example

Assume we have a very simple time series sample: **[2, 4, 6, 8]**, meaning N=4.

1. **Calculate the mean:**

   $\bar{x} = (2 + 4 + 6 + 8) / 4 = 5$

2. **Calculate $\hat{\gamma}(0)$ (variance):**

   $\hat{\gamma}(0) = (1/4) * [(2-5)(2-5) + (4-5)(4-5) + (6-5)(6-5) + (8-5)(8-5)]$

   $= (1/4) * [(-3)(-3) + (-1)(-1) + (1)(1) + (3)(3)]$

   $= (1/4) * [9 + 1 + 1 + 9] = (1/4) * 20 = \mathbf{5}$

3. **Calculate $\hat{\gamma}(1)$:**

   $\hat{\gamma}(1) = (1/4) * [(2-5)(4-5) + (4-5)(6-5) + (6-5)(8-5)]$

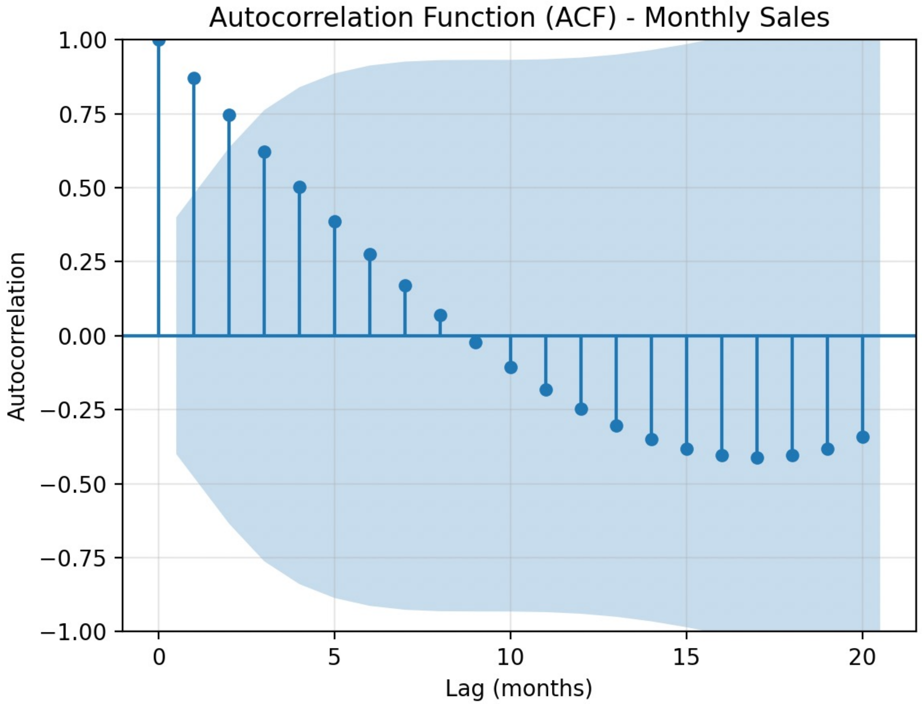
   $= (1/4) * [(-3)(-1) + (-1)(1) + (1)(3)]$

   $= (1/4) * [3 - 1 + 3] = (1/4) * 5 = \mathbf{1.25}$

- $\hat{\rho}(1) = \hat{\gamma}(1) / \hat{\gamma}(0) = 1.25 / 5.0 = \mathbf{0.25}$

A company's sales over 24 consecutive months, showing steady growth.

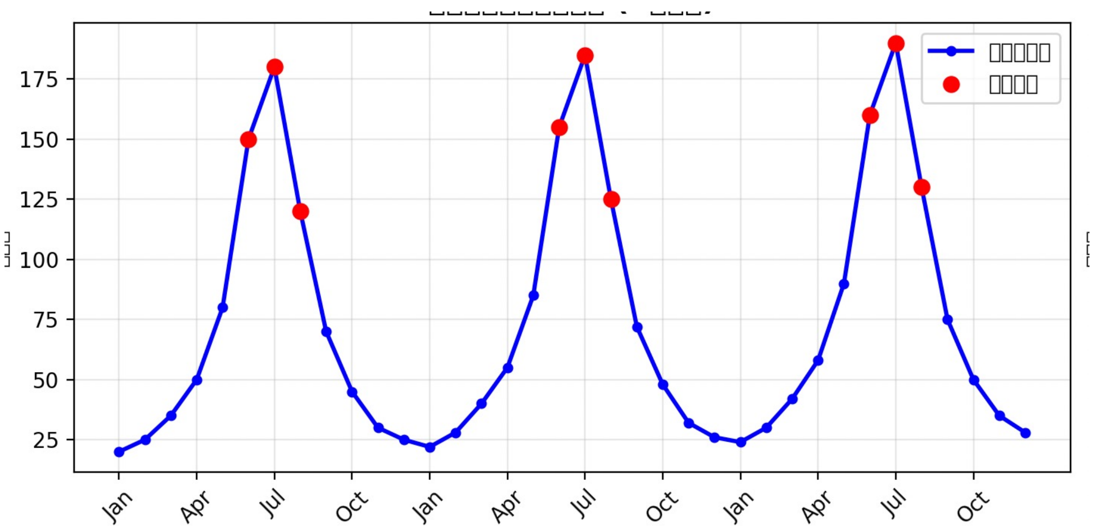
Sales = [100, 112, 125, 138, 150, 163, 177, 190, 205, 220, 235, 250, 265, 281, 298, 315, 333, 351, 370, 389, 409, 430, 451, 473]



A company's sales over 24 consecutive months, showing steady growth.

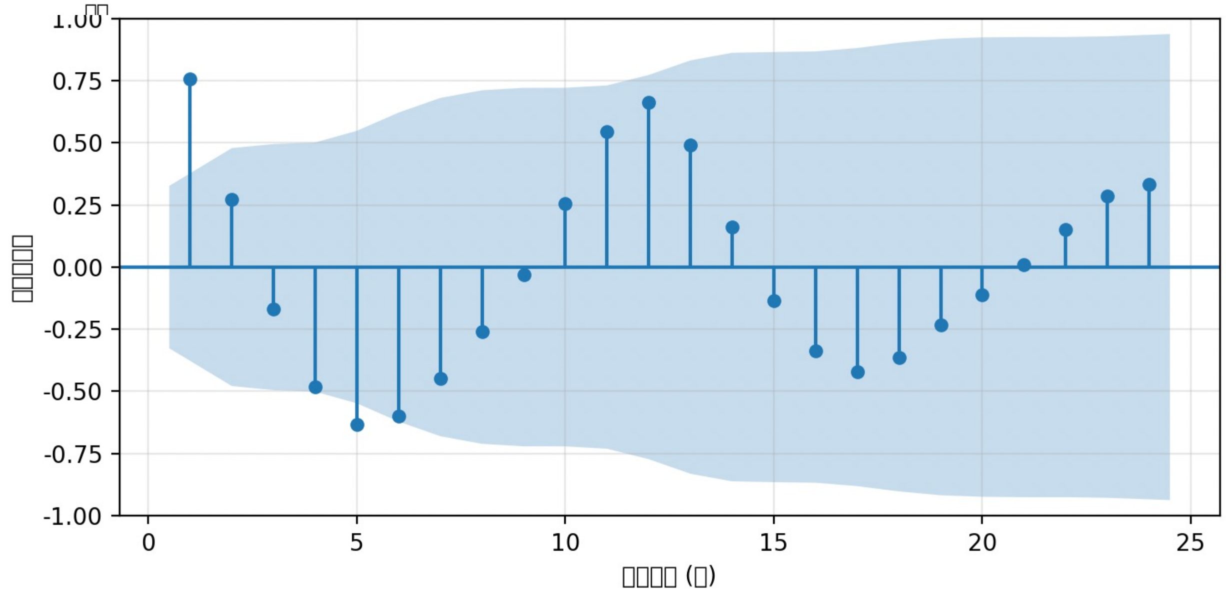
Sales = [100, 112, 125, 138, 150, 163, 177, 190, 205, 220, 235, 250, 265, 281, 298, 315, 333, 351, 370, 389, 409, 430, 451, 473]

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.stattools import acf

# Sales data
sales = [100, 112, 125, 138, 150, 163, 177, 190, 205, 220, 235, 250,
         265, 281, 298, 315, 333, 351, 370, 389, 409, 430, 451, 473]

# Create ACF plot
plt.figure(figsize=(12, 6))
plot_acf(sales, lags=20, alpha=0.05, title='Autocorrelation Function (ACF) - Monthly Sales')
plt.xlabel('Lag (months)')
plt.ylabel('Autocorrelation')
plt.grid(True, alpha=0.3)
plt.show()

# Calculate ACF values
acf_values = acf(sales, nlags=20)
print("ACF Values for first 10 lags:")
for lag, value in enumerate(acf_values[:11]):
    print(f"Lag {lag}: {value:.4f}")
```



Time Series Description: Monthly sales of a beverage over N months (3 years). Summer is the peak sales season each year.

Sales = [20, 25, 35, 50, 80, 150, 180, 120, 70, 45, 30, 25, 22, 28, 40, 55, 85, 155, 185, 125, 72, 48, 32, 26, 24, 30, 42, 58, 90, 160, 190, 130, 75, 50, 35, 28]



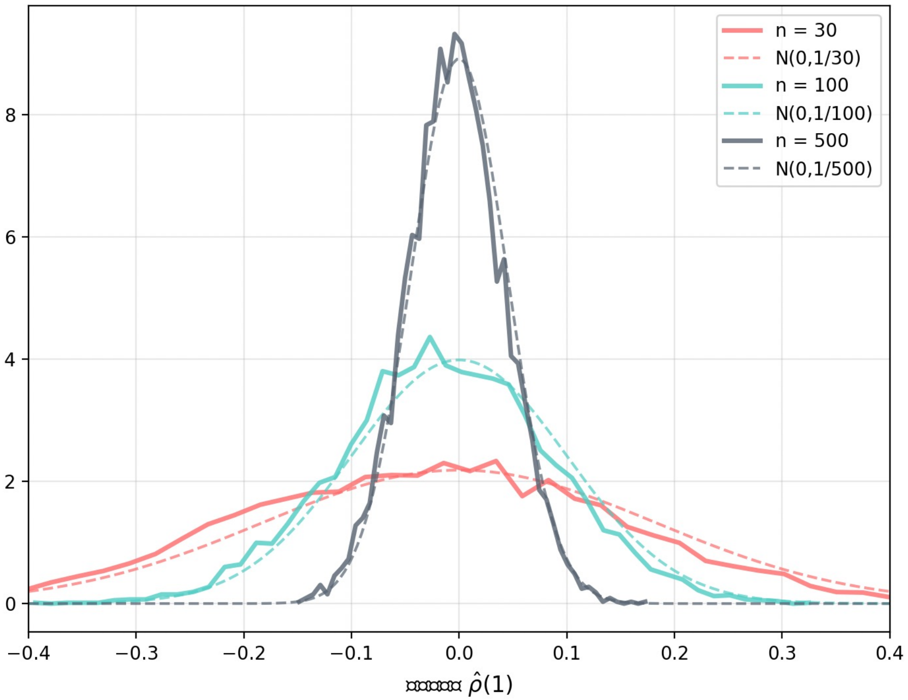
Review

- **For a white noise process w_t, if $E(w_t^4) < \infty$,**

$$\begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(K) \end{pmatrix} \sim AN\left(0, \frac{1}{n} I\right)$$

Asymptotically Normal

w_t^4 refers to the fourth power of the white noise process w_t.



As the sample size *n* increases, the shape of the distribution approaches a normal distribution.

Review

- **Backshift operator:**

$$Bx_t = x_{t-1}$$

- **Forward-shift operator:**

$$x_t = B^{-1}x_{t-1}$$

- **First difference operator: (eliminate linear trend)**

$$\nabla x_t = (1-B)x_t$$

- **Differences with order d: (eliminate higher order trend)**

$$\nabla^d = (1-B)^d$$

Review

- **Backshift operator:**

$$Bx_t = x_{t-1}$$

Backshift Operator

$B\,x_t = x_{t-1}$ (Backward shift by 1 step)

$B^2\,x_t = B(B\,x_t) = B\,x_{t-1} = x_{t-2}$ (Backward shift by 2 steps)

$B^p\,x_t = x_{t-p}$ (Backward shift by p steps)

Time (t)	Observation (X_t)
1	5
2	8
3	6
4	9
5	7

1. Applying the backshift operator once (B)

- $BX_2 = X_1 = 5$
- $BX_3 = X_2 = 8$
- $BX_4 = X_3 = 6$
- $BX_5 = X_4 = 9$

2. Applying the backshift operator twice (B^2)

- $B^2X_3 = X_1 = 5$
- $B^2X_4 = X_2 = 8$
- $B^2X_5 = X_3 = 6$

Review

- **Forward-shift operator:**

$$x_t = B^{-1}x_{t-1}$$

$$FX_t = X_{t+1}$$

More generally, applying the operator **k** times:

$$F^k X_t = X_{t+k}$$

Time (t)	Observation (X_t)
1	5
2	8
3	6
4	9
5	7

1. Applying the forward shift operator once (F)

- $FX_1 = X_2 = 8$
- $FX_2 = X_3 = 6$
- $FX_3 = X_4 = 9$
- $FX_4 = X_5 = 7$

2. Applying the forward shift operator twice (F^2)

- $F^2X_1 = X_3 = 6$
- $F^2X_2 = X_4 = 9$
- $F^2X_3 = X_5 = 7$

- **First difference operator: (eliminate linear trend)**

$$\nabla x_t = (1 - B)x_t$$

- **Purpose**: It calculates the change between **consecutive observations** in a time series.
- **Formula**:

$$\nabla x_t = x_t - x_{t-1} = (1 - B)x_t$$

How It Works

From the formula $(1 - B)x_t$:

1. 1 represents the "current value": $1 \cdot x_t = x_t$
2. $-B$ represents "minus the previous value": $-Bx_t = -x_{t-1}$
3. Combined: $(1 - B)x_t = x_t - x_{t-1}$

- **First difference operator: (eliminate linear trend)**

$$\nabla x_t = (1 - B)x_t$$

Time (t)	Observation (x_t)
1	10
2	12
3	14
4	16
5	18

Apply the first difference operator ∇x_t:

- $\nabla x_2 = (1 - B)x_2 = x_2 - x_1 = 12 - 10 = 2$
- $\nabla x_3 = (1 - B)x_3 = x_3 - x_2 = 14 - 12 = 2$
- $\nabla x_4 = (1 - B)x_4 = x_4 - x_3 = 16 - 14 = 2$
- $\nabla x_5 = (1 - B)x_5 = x_5 - x_4 = 18 - 16 = 2$

Resulting Series: `[?, 2, 2, 2, 2]`

- **Differences with order d: (eliminate higher order trend)**

$$\nabla^d = (1-B)^d$$

$$\nabla x_t = x_t - x_{t-1}$$

1	$2+1(1)+0.5(1)^2$	3.5
2	$2+1(2)+0.5(2)^2$	6.0
3	$2+1(3)+0.5(3)^2$	9.5
4	$2+1(4)+0.5(4)^2$	14.0
5	$2+1(5)+0.5(5)^2$	19.5
6	$2+1(6)+0.5(6)^2$	26.0

- $\nabla x_2 = 6.0 - 3.5 = 2.5$
- $\nabla x_3 = 9.5 - 6.0 = 3.5$
- $\nabla x_4 = 14.0 - 9.5 = 4.5$
- $\nabla x_5 = 19.5 - 14.0 = 5.5$
- $\nabla x_6 = 26.0 - 19.5 = 6.5$

(∇x_t): `[?, 2.5, 3.5, 4.5, 5.5, 6.5]`

- **Differences with order d: (eliminate higher order trend)**

$$\nabla^d = (1-B)^d$$

1	$2+1(1)+0.5(1)^2$	3.5
2	$2+1(2)+0.5(2)^2$	6.0
3	$2+1(3)+0.5(3)^2$	9.5
4	$2+1(4)+0.5(4)^2$	14.0
5	$2+1(5)+0.5(5)^2$	19.5
6	$2+1(6)+0.5(6)^2$	26.0

(∇x_t): `[?, 2.5, 3.5, 4.5, 5.5, 6.5]`

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$$

- $\nabla^2 x_3 = \nabla x_3 - \nabla x_2 = 3.5 - 2.5 = 1.0$
- $\nabla^2 x_4 = \nabla x_4 - \nabla x_3 = 4.5 - 3.5 = 1.0$
- $\nabla^2 x_5 = \nabla x_5 - \nabla x_4 = 5.5 - 4.5 = 1.0$
- $\nabla^2 x_6 = \nabla x_6 - \nabla x_5 = 6.5 - 5.5 = 1.0$

($\nabla^2 x_t$): `[?, ?, 1.0, 1.0, 1.0, 1.0]`

Difference operator

- **The first difference eliminates a linear trend:**

$$x_t = \beta_0 + \beta_1 t + y_t$$

- **The second order difference eliminates a quadratic trend:**

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + y_t$$

Difference operator

- **The first difference eliminates a linear trend:**

$$x_t = \beta_0 + \beta_1 t + y_t$$

$$\begin{aligned}\nabla x_t &= x_t - x_{t-1} \\ &= \beta_0 + \beta_1 t + y_t - (\beta_0 + \beta_1(t-1) + y_{t-1}) \\ &= \beta_1 + y_t - y_{t-1}\end{aligned}$$

- **The second order difference eliminates a quadratic trend:**

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + y_t$$

$$\begin{aligned}\nabla x_t &= x_t - x_{t-1} \\ &= \beta_1 - \beta_2 + 2\beta_2 t + y_t - y_{t-1}\end{aligned}$$

$$\begin{aligned}\nabla^2 x_t &= \nabla(\nabla x_t) \\ &= 2\beta_2 + y_t - 2y_{t-1} + y_{t-2}\end{aligned}$$

Difference operator

- **Higher order trend:**

$$x_t = \sum_{i=0}^{k} \beta_i t^i + y_t$$

- **Seasonal trend:**

$$x_t = s_t + y_t \text{ (where } s_t = s_{t-p} \text{ for all } t\text{)}$$

Linear process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

where $\{w_t\} \sim \mathrm{wn}(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying $\displaystyle\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Linear Process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

"Linear" in time series specifically means that the current value x_t can be expressed as a weighted sum (i.e., a linear combination) of white noise shocks $w_{t\text{-}j}$.

With the weight ψ_j and the shock w_{t-j}

where $\{w_t\} \sim \text{wn}(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying $\displaystyle\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Linear Process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

Causal part: Composed of shocks from the **past and present** $(j \geq 0)$

Non-causal part: Composed of shocks from the **future** $(j < 0)$

$$x_t = \mu + \dots + \psi_2 w_{t-2} + \psi_1 w_{t-1} + \psi_0 w_t + \psi_{-1} w_{t+1} + \dots$$

where $\{w_t\} \sim \mathrm{wn}(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Linear Process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

- μ is the mean of the process.

where $\{w_t\} \sim \text{wn}(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Linear Process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

where $\{w_t\} \sim \mathrm{wn}(0, \sigma_w^2)$

- $\{w_t\}$ is a **white noise** sequence with:
  - $E[w_t] = 0$
  - $\mathrm{Var}(w_t) = \sigma_w^2 < \infty$
  - $\mathrm{Cov}(w_t, w_s) = 0$ for $t \neq s$

and μ, ψ_j are parameters satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Linear Process

- **The time series $\{x_t\}$ is a linear process if it has the representation**

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

The condition $\sum|\psi_j| < \infty$ implies a crucial property: lim_{|j|→∞} $|\psi_j| = 0$. That is, the weights eventually decay to zero.

where $\{w_t\} \sim \mathrm{wn}(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Weak Stationarity

Linear process

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

$$\gamma(h) = \sigma^2 \sum_{-\infty}^{\infty} \psi_{j+h} \psi_j$$

Derivation

$$\gamma(h) = \text{Cov}(\tilde{x}_t, \tilde{x}_{t+h}) = E\left[\sum_{j=1}^{\infty} \psi_j w_{t-j} \sum_{k=1}^{\infty} \psi_k w_{t+h-k}\right]$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_j \psi_k E[w_{t-j} w_{t+h-k}]$$

Linear Process and AR Model

Core Relationship

An AR model is a specific type of Linear Process.

Think of it like this:

- **Linear Process** is a broad **category** or a **family** of models.
- **AR Model** is a specific, widely-used **member** of that family.

All (stationary) AR models are linear processes, but not all linear processes are AR models.

- **AR Model**: Memory is stored in **past values of itself (x_{t-1}, x_{t-2}...)**, defining the current value through a linear combination of historical observations

Autoregressive models

- **Intuition: the current value of the series, x_t, can be explained as a function of p past values, x_{t-1}, x_{t-2}, …, x_{t-p}**

- **Further assume linear relationship between current and past values**

- **Allows for the forecast of future values based on the observed data (current and past values)**

Autoregressive models

- **An autoregressive model of order *p* (denoted as AR(*p*)), is of the form**

  $$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

  $$w_t \sim \mathrm{wn}(0, \sigma_w^2)$$

  $\phi_1, \phi_2, \ldots, \phi_p$ are constants $(\phi_p \neq 0)$

AR Model (of order p)

For an **AR(p)** model, the formula is:

```
Current Value = φ₁ * (1st Lag Value) + φ₂ * (2nd Lag Value) + ... + φp * (p-th Lag Value) + Random Shock
```

- `p = 1` **(1st Order)**:
  - Model: `Today's Temperature = φ₁ * Yesterday's Temperature + Random Shock`
  - Meaning: Only considers the influence of **yesterday** on today.
- `p = 2` **(2nd Order)**:
  - Model:
    ```
    Today's Temperature = φ₁ * Yesterday's Temperature + φ₂ * Day Before Yesterday's Temperature + Random Shock
    ```
  - Meaning: Considers the influence of both **yesterday** and **the day before yesterday** on today.

AR Model

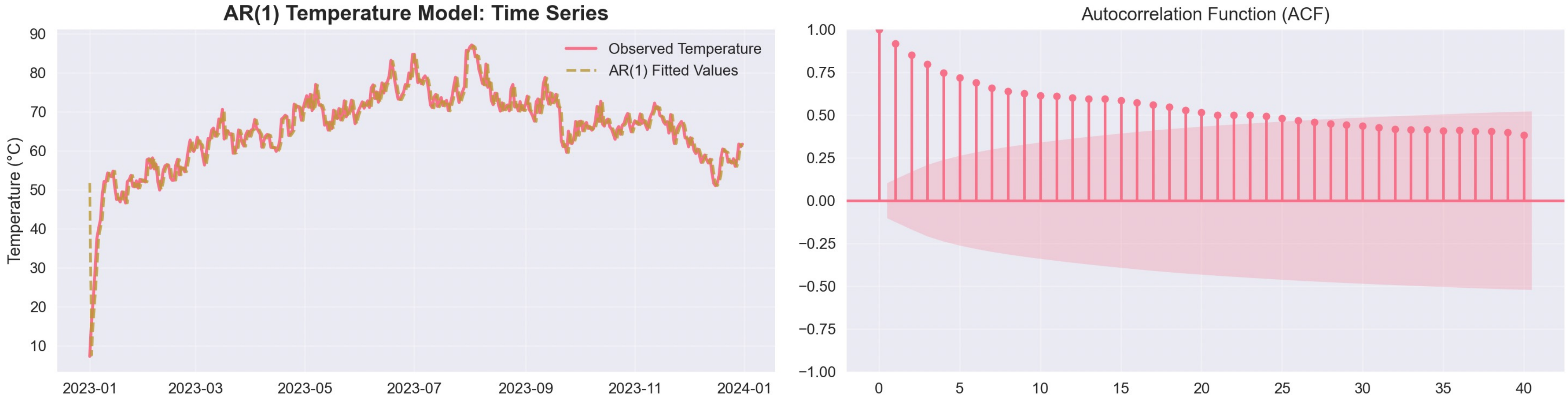
- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

The value at time *t* is related to the series values from the previous *p* periods, and the error term is the random disturbance at the current period, which is a white noise sequence with a mean of zero.

The equation of an AR model is quite similar to that of a multiple linear regression, except that the independent variables (features) are replaced by time-lagged values of the series itself (x_{t-1}, x_{t-2}, ..., x_{t-p}).

$$T_t = 10.0 + 0.85 \times T_{t-1} + \varepsilon_t$$



Practical Example: Stock Market Prices
Suppose the daily closing price of a certain stock follows an AR(2) model:

$$x_t = 0.6x_{t-1} + 0.3x_{t-2} + \varepsilon_t$$

This implies:

- Today's stock price is 60% influenced by the previous day's price.
- 30% is influenced by the price from two days ago.
- The remaining fluctuations come from white noise.

Summary: The AR model is suitable for data with a **trend**, meaning the current value depends on past observations.

Autoregressive models

- **Autoregressive operator**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$\Big(\boxed{1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p}\Big) x_t = w_t$$

$$\phi(B) x_t = w_t$$

AR Model

- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

Backshift Operator

shift the values of a time series backward by one time unit.

$B\ x_t = x_{t-1}$ (Backward shift by 1 step)

$B^2\ x_t = B(B\ x_t) = B\ x_{t-1} = x_{t-2}$ (Backward shift by 2 steps)

$B^p\ x_t = x_{t-p}$ (Backward shift by p steps)

- $x_{t-1} = B\ x_t$
- $x_{t-2} = B^2\ x_t$
- $x_{t-p} = B^p\ x_t$

$$x_t = \varphi_1(B\ x_t) + \varphi_2(B^2\ x_t) + \ldots + \varphi_p(B^p\ x_t) + w_t$$

AR Model

$$x_t = \varphi_1(B\ x_t) + \varphi_2(B^2\ x_t) + ... + \varphi_p(B^p\ x_t) + w_t$$

Move all terms containing x_t to the left side of the equation from above:

$$x_t - \phi_1 B\, x_t - \phi_2 B^2\, x_t - ... - \phi_p B^p\, x_t = w_t$$

Factor out the common factor x_t:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)\, x_t = w_t$$

Look! This entire expression within the parentheses is precisely **$\phi(B)$**:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$$

Therefore, the entire equation becomes the extremely concise form:

$$\phi(B)\, x_t = w_t$$

AR Model

$$\phi(B)x_t = w_t, \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

$\phi(B)$ is not merely an abbreviation; it represents a system or a filter.

- **Input:** Time series `x_t`
- **System:** `φ(B)` (It is defined by the model's parameters $\phi_1, \phi_2, \ldots, \phi_p$)
- **Output:** White noise `w_t`

The meaning of the equation $\phi(B)x_t = w_t$ is:

> "If we filter the original series `x_t` through this `φ(B)` system defined by our parameters, all the predictable patterns that can be captured by past values will be removed, and the final output will be a purely random white noise `w_t`."

AR(1) model

- **AR(1) model:**

$$x_t = \phi x_{t-1} + w_t, \quad 0 < |\phi| < 1$$

- **Solution:**

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

- **Mean and autocovariance function:**

$$\mu = 0; \quad \gamma(h) = \frac{\sigma_w^2 \phi^h}{1 - \phi^2}$$

AR(1) model

- **AR(1) model:**

$$x_t = \phi x_{t-1} + w_t, \quad 0 < |\phi| < 1$$

- **Solution:**

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

Substitute for x_{t-1}:

$x_{t-1} = \phi x_{t-2} + w_{t-1}$. Plugging this in:

$$x_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t = \phi^2 x_{t-2} + \phi w_{t-1} + w_t$$

Substitute for x_{t-2}:

$x_{t-2} = \phi x_{t-3} + w_{t-2}$. Plugging this in:

$$x_t = \phi^2(\phi x_{t-3} + w_{t-2}) + \phi w_{t-1} + w_t = \phi^3 x_{t-3} + \phi^2 w_{t-2} + \phi w_{t-1} + w_t$$

Continue the pattern: If we continue this process n times, we get:

$$x_t = \phi^n x_{t-n} + \sum_{j=0}^{n-1} \phi^j w_{t-j} \qquad \lim_{n\to\infty} \phi^n x_{t-n} = 0$$

AR(1) model

- **In terms of the backshift operator:**

$$x_t - \phi x_{t-1} = w_t \quad \Rightarrow \quad \begin{aligned} \phi(B)x_t &= w_t \\ \phi(B) &= 1 - \phi B \end{aligned}$$

Inverse operator

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} \quad \Rightarrow \quad \begin{aligned} x_t &= \psi(B)w_t \\ \psi(B) &= \sum_{j=0}^{\infty} \phi^j B^j \end{aligned}$$

Consider the AR(1) model:

$$(1 - \phi B)X_t = W_t$$

To solve for X_t, we apply the inverse operator $(1 - \phi B)^{-1}$ to both sides of the equation:

$$X_t = (1 - \phi B)^{-1} W_t$$

Substituting the infinite series expansion for the inverse operator (valid when $|\phi| < 1$):

$$X_t = \left(\sum_{j=0}^{\infty} \phi^j B^j \right) W_t$$

Now, applying the lag operator B^j to W_t gives $B^j W_t = W_{t-j}$. Therefore, the solution becomes:

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$$

In operator theory, we formally treat the operator B as if it were a numerical variable (while being careful about convergence conditions).

Therefore, for the operator expression $(1 - \phi B)$, if we want to find its "inverse", we analogously write:

$$(1 - \phi B)^{-1} = \frac{1}{1 - \phi B}$$

Then, formally applying the geometric series expansion:

$$\frac{1}{1 - \phi B} = 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \cdots = \sum_{j=0}^{\infty} \phi^j B^j$$

Causal process

- **Causality** is a fundamental concept in time series analysis and signal processing.

A time series process is called **causal** if its current value depends **only** on:

- Present and **past** values of the input/shock process
- **Not** on future values

AR(1) model

- **An AR(1) process defined by**

  $\phi(B)x_t = w_t$ with $\phi(B) = 1 - \phi B$

  is causal if and only if

  $|\phi| < 1$

  or

  the root z_1 of the polynomial $\phi(z) = 1 - \phi z$ satisfies $|z_1| > 1$

Causality Condition 1:

$$|\phi| < 1$$

This is what we discussed earlier: when $|\phi| < 1$, the process can be expressed as

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j},$$

which depends only on past and present noise.

Non-causal case

When $|\phi| > 1$, the process is non-causal and depends on future noise:

$$x_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}$$

- **Autoregressive characteristic polynomial**

$$(1 - \phi B)x_t = w_t$$

$$\phi(B) = 1 - \phi B \qquad \phi(z) = 1 - \phi z$$

the root z_1 of the polynomial $\phi(z) = 1 - \phi z$ satisfies

$|z_1| > 1$

AR(p) model

- **Example: check causality**

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

$$x_t = -0.7x_{t-1} - 0.6x_{t-2} + w_t$$

AR(p) model

- **AR(p) model:**

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

$$\phi(B)x_t = w_t, \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

- **Stationary solution exists if and only if**

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| \neq 1$$

AR(p) model

- **Example: check causality**

$$x_t = 0.7x_{t-1} + 0.6x_{t-2} + w_t$$

$$\phi(B) = 1 - 0.7B - 0.6B^2$$

$$\phi(z) = 1 - 0.7z - 0.6z^2 = 0$$

$$\Downarrow$$

$$z_1 = -2, \quad z_2 = 0.8333$$

$$\Downarrow$$

Not Causal

AR(p) model

- **Example: check causality**

$$x_t = -0.7x_{t-1} - 0.6x_{t-2} + w_t$$

$$\phi(B) = 1 + 0.7B + 0.6B^2$$

$$\phi(z) = 1 + 0.7z + 0.6z^2 = 0$$

$$\Downarrow$$

$$z_{1,2} = -0.5833 \pm 1.1517\mathrm{i}$$

$$\Downarrow$$

Causal