# Time Series Analysis and Modelling

## Part 3: Time Series Models (AR and MA)


Jonathan Mwaura

Khoury College of Computer Sciences


August 1, 2024


The Roux Institute
at Northeastern University

# Introduction

## Acknowledgements


- These slides have been adapted from Achraf Cohen, University of West Florida - from the class class STA6856
- Lecture Notes from Dewei Wang, Department of Statistics, University of South Carolina (See notes in Canvas)
- C. Chatfield, The Analysis of Time Series: Theory and Practice, Chapman and Hall (1975).

# Stationary Processes

## Remarks

1. A key role in time series analysis is given by processes whose properties **do not vary with time**.
2. If we wish to make predictions then clearly we must assume that *something* does not vary with time
3. In time series analysis, our goal is to predict a series that contains a random component, if this random component is stationary (*weakly*) then we can develop powerful techniques to forecast its future values.

# Stationary Processes

## Basic Properties

1. The autocovariance function (ACVF)
   $\gamma(h) = Cov(X_{t+h}, X_t), \quad h = 0, \pm 1, \pm 2, \ldots$
2. The autocorrelation function (ACF) $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$
3. $\gamma(0) \geq 0$
4. $|\gamma(h)| \leq \gamma(0)$, for all h ($\rho(h) \leq 1$) or (Cauchy-Schwarz inequality $E(XY)^2 \leq E(X^2)E(Y^2)$)
5. $\gamma(h) = \gamma(-h)$

# Stationary Processes

## Prediction

The ACF and ACVF provide a useful measure of **the degree of dependence** among the values of a time series at different times => Very important if we consider the prediction of future values of the series in terms of the past and present values.

# Stationary Processes

## Prediction

The ACF and ACVF provide a useful measure of **the degree of dependence** among the values of a time series at different times => Very important if we consider the prediction of future values of the series in terms of the past and present values.

## Question

What is the role of the autocorrelation function in prediction?

# Stationary Processes

## The role of autocorrelation in prediction

- Consider $X_t$ a stationary Gaussian time series (all of its joint distributions are Multivariate Normal)
- We observed $X_n$ and we would like to find the function of $X_n$ that gives us *the best predictor* of $X_{n+h}$ (the value of the series after h time units).
- **The best predictor** will be given by the function of $X_n$ that minimizes the mean squared error (MSE).

## Question

What is function of $X_n$ that gives us the best predictor of $X_{n+h}$?

## Answer

The best predictor of $X_{n+h}$ in terms of MSE is given by
$E(X_{n+h}|X_n) = \mu + \rho(h)(X_n - \mu)$

# Stationary Processes

## The role of autocorrelation in prediction

For time series with non-normal joint distributions the calculation are in general more complicated => we look at **the best linear predictor** ($l(X_n) = aX_n + b$), then our problem becomes finding a and b that minimize $E((X_{n+h} - aX_n - b)^2)$.

## Answer

The best linear predictor of $l(X_n)$ in terms of MSE is given by

$$l(X_n) = \mu + \rho(h)(X_n - \mu)$$

The fact that the best linear predictor depends only on the mean and the ACF of the series $X_t$ means that it can be calculated without more detailed knowledge of the series $X_t$

# Stationary Processes: Examples

## The MA(q) process

$X_t$ is a moving-average process of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

## Remarks

- If $X_t$ is a stationary q-correlated time series with mean 0, then it can be represented as the MA(q) process

# Stationary Processes: Examples

## The ARMA(1,1) process

The time series $X_t$ is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$

# Stationary Processes: Properties

## The MA(q) process

The simplest ways to construct a time series that is strictly stationary is to "filter" an iid sequence of random variables. Consider $Z_t \sim IID$, we define:

$$X_t = g(Z_t, Z_{t-1}, \ldots, Z_{t-q})$$

for some real-valued function $g(., \ldots, .)$. We can say that $X_t$ is **q-dependent**.

## Remarks

- IID is 0-dependent
- WN is 0-correlated
- A stationary time series is q-correlated if $\gamma(h) = 0$ whenever $lhl > q$
- MA(1) is 1-correlated
- MA(q) is q-correlated

# Stationary Processes: Properties

**The MA(q) process**

$X_t$ is a **moving-average process** of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

If $X_t$ is a stationary q-correlated time series with mean 0, then it can be represented as the MA(q) process given by the equation above.

# Stationary Processes: Properties

The class of Linear time series models, which includes the class of Autoregressive Moving-Average (ARMA) models, provides a general framework for studying stationary processes.

## Linear processes

The time series $X_t$ is a **Linear process** if it has the representation:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \tag{2}$$

for all t, where $Z_t \sim WN(0, \sigma^2)$ and $\psi_j$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$

1. The condition $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$ ensures the infinite sum in (2) converges
2. If $\psi_j = 0$ for all $j < 0$ then a linear process is called a moving average or MA($\infty$)

# Stationary Processes: Linear Process

## Proposition

Let $Y_t$ be a stationary series with mean 0 and autocovariance function $\gamma_y$. If $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Y_t = \psi(B) Y_t \tag{3}$$

is stationary with mean zero and autocorrelation function $\gamma_x(h)$

1. $\gamma_x(h) = \sum \sum \psi_j \psi_k \gamma_X (h - k + j)$

# Stationary Processes: Properties

## The AR(1) process

The time series $X_t$ is an AR(1) process was found to be stationary:

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1} \tag{4}$$

if $Z_t \sim WN(0, \sigma^2)$ and $\mid \Phi \mid < 1$, and $Z_t$ is uncorrelated with $X_s$ for each $s \leq t$

1. $X_t$ is called a causal or future-independent function of $Z_t$.
2. If $\mid \Phi \mid > 1$ (the series does not converge, we can rewrite then AR(1))
3. If $\Phi = \pm 1$, there is no stationary solution of (4)

# Stationary Processes: Properties

## The ARMA(1,1) process

The time series $X_t$ is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta + \Phi \neq 0$

1. A stationary the ARMA(1,1) exists if and only if $\phi \neq \pm 1$
2. If $\mid \phi \mid < 1$, then the unique stationary solution is $X_t = Z_t + (\theta + \phi) \sum \phi^{j-1} Z_{t-j}$. This $X_t$ is causal, since $X_t$ can be expressed in terms of the current and past values of $Z_s$, $s \leq t$
3. If $\mid \phi \mid > 1$, then the unique stationary solution is $X_t = -\theta \psi^{-1} Z_t + (\theta + \phi) \sum \phi^{-j-1} Z_{t+j}$. This $X_t$ is noncausal, since $X_t$ can be expressed in terms of the current and future values of $Z_s$, $s \geq t$. (unatural solution)

# Stationary Processes: Properties

## The ARMA(1,1) process

The time series $X_t$ is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta + \Phi \neq 0$

1. if $\theta = \pm 1$ , then ARMA(1,1) process is invertible in the more general sense that $Z_t$ is a mean square limit of finite linear combinations of $X_s$, $s \leq t$.
2. If $\mid \theta \mid < 1$, then the ARMA(1,1) is invertible, since $Z_t$ can be expressed in terms of the current and past values of $X_s$, $s \leq t$
3. If $\mid \theta \mid > 1$, then the ARMA(1,1) is noninvertible, since $Z_t$ can be expressed in terms of the current and future values of $X_s$, $s \geq t$.

# Stationary Processes: Properties

## The Sample Mean and Autocorrelation Function

A weakly stationary time series $X_t$ is charachterized by its mean $\mu$, its autocovariance $\gamma(.)$, and its autocorrelation $\rho(.)$. The estimation of these statistics play a crucial role in problems of inference => constructing an appropriate model for the data. We examine here some properties of the sample estimates $\overline{X}$ and $\hat{\rho}(.)$.

1. $E(\overline{X}_n) = \mu$
2. $Var(\overline{X}_n) = n^{-1} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$
3. If the time series is Gaussian, then $(\overline{X}_n - \mu)/\sqrt{n} \sim N(0, (1 - \frac{|h|}{n})\gamma(h))$
4. The confidence Intervals for $\mu$ are given $\overline{X}_n \pm Z_{1-\alpha/2} \frac{\sqrt{\hat{v}}}{\sqrt{n}}$, where $\hat{v} = \sum_{|h|<\sqrt{n}} (1 - \frac{|h|}{\sqrt{n}}) \hat{\gamma}(h)$. For ARMA processes, this is a good approximation of v for large n.

# Stationary Processes: Properties

## The Sample Mean: Asymptotic distribution

If we know the asymptotic distribution of $\overline{X}_n$, we can use it to infer about $\mu$ (e.g. is $\mu$=0?). Similarly for $\hat{\rho}(h)$

1. $$(\overline{X}_n - \mu)\sqrt{n} \sim AN(0, \sum_{|h|<n} (1 - \frac{|h|}{n})\gamma(h))$$

2. In this case, the confidence Intervals for $\mu$ are given $\overline{X}_n \pm Z_{1-\alpha/2}\frac{\sqrt{\hat{v}}}{\sqrt{n}}$, where $\hat{v} = \sum_{|h|<\sqrt{n}} (1 - \frac{|h|}{\sqrt{n}})\hat{\gamma}(h)$. For ARMA processes, this is a good approximation of v for large n.

## Example

What are the approximate 95% confidence intervals for the mean of AR(1)?
"AN" means Asymptotically Normal

# Stationary Processes: Properties

## The Estimation of $\gamma(.)$ and $\rho(.)$

The sample autocovariance and autocorrelation functions are defined by:

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-|h|}(X_{t+|h|} - \overline{X}_n)(X_t - \overline{X}_n)}{n}$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

1. For $h$ slightly smaller than n, the estimate of $\gamma(h)$ and $\rho(h)$ are unreliable. Since there few pairs $(X_{t+h}, X_t)$ available (only one if h=n-1). **A practical guide is to have at least n=50 and h $\leq$ n/4**

# Stationary Processes: Properties

## The Estimation of $\gamma(.)$ and $\rho(.)$: Asymptotic distribution

The sampling distribution of $\rho(.)$ can usually be approximated by a normal distribution for large sample sizes. For Linear Models (ARMA):

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \ldots, \hat{\rho}(k))' \sim AN(\boldsymbol{\rho}, \frac{W}{n})$$

where $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(k))$, and $W$ is the covariance matrix whose $(i, j)$ element is given by **Bartlett's formula**:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

1. IID Noise?

# Stationary Processes: Properties ACF

## The sample ACF Examination - Results

- If $\mid \hat{\rho}(h) \mid < \frac{1.96}{\sqrt{n}}$, for all $h \geq 1$, then assume $MA(0)$, WN sequence.
- If $\mid \hat{\rho}(1) \mid > \frac{1.96}{\sqrt{n}}$, then we should look at the rest $\hat{\rho}(h)$ with $\pm \frac{1.96\sqrt{1+2\rho^2(1)}}{\sqrt{n}}$; we can replace $\rho(1)$ by its estimate (you can also remark that $2\hat{\rho}^2(1)/n \sim 0$, for large n).
- In general, if $\mid \hat{\rho}(h_0) \mid > \frac{1.96}{\sqrt{n}}$ and $\mid \hat{\rho}(h) \mid < \frac{1.96}{\sqrt{n}}$, for $h \geq h_0$, then assume MA(q) model with $q = h_0$.

# Stationary Processes: Forecasting

## Forecasting $P_n X_{n+h}$

Now we consider the problem of predicting the values $X_{n+h}$; $h > 0$. Let's assume $X_t$ is a stationary time series with $\mu$ and $\gamma$.

*The goal is to find the linear combination* of $1, X_n, \ldots, X_1$ that minimizes the mean squared error. We will denote

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1$$

It remains to find the coefficients $a_i$ that minimizes:

$$E((X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2)$$

# Stationary Processes: Forecasting

## Forecasting $P_n X_{n+h}$

We can show that $P_n X_{n+h}$ is given by:

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$$

and

$$E((X_{n+h} - P_n X_{n+h})^2) = \gamma(0) - \boldsymbol{a}_n' \gamma_n(h)$$

where $\boldsymbol{a}_n$ satisfies

$$\Gamma_n \boldsymbol{a}_n = \gamma_n(h)$$

where $\boldsymbol{a}_n = (a_1, \dots, a_n)'$, $\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{bmatrix}$;

$\gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))'$; and $a_0 = \mu(1 - \sum_{i=1}^{n} a_i)$

# Stationary Processes: Forecasting

## Prediction Algorithms

The following prediction algorithms use the idea of one-step predictor $P_n X_{n+1}$ based on n previous observations would be used to calculate $P_{n+1} X_{n+2}$. This said to be recursive.

1. The Durbin-Levinson Algorithm (well suited to forecasting $AR(p)$)
2. The Innovations Algorithm (well suited to forecasting $MA(q)$)