Séries Temporelles
2017-2018

M1 MINT

Marie-Luce Taupin
marie-luce.taupin@univ-evry.fr

Organizational issues

- 5 classes : 17 jan, 24 jan, 31 jan, 14 fev, 21 fev, 07 mars, 21 mars
- 2 Labs.
  - 14 mars
  - 21 mars
- Slides, R examples and labs are on my webpage http ://www.math-evry.cnrs.fr/members/mtaupin
- My email marie-luce.taupin@univ-evry.fr
- Evaluation of the course based on labs 1 and 2 and a final evaluation

Chapter 1 : Time series characteristics

Johnson & Johnson quaterly earnings [SS10] I

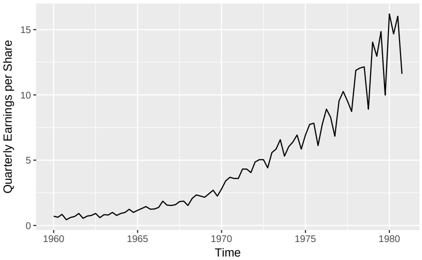


FIG.: Quarterly earnings per share for the U.S. company Johnson & Johnson.

Notice :

- the increasing underlying trend and variability,
- and a somewhat regular oscillation superimposed on the trend that seems to repeat over quarters.

Johnson & Johnson quaterly earnings [SS10] II

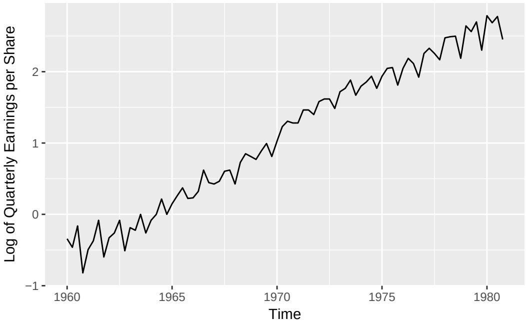


FIG.: Quarterly log(earnings) per share for the U.S. company Johnson & Johnson.

Notice :

- the trend is now (almost) linear

Global temperature index from 1880 to 2015 I

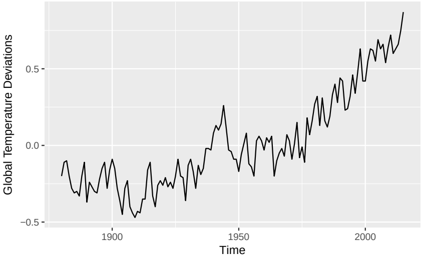


FIG. : Global temperature deviation (in $^\circ C$) from 1880 to 2015, with base period 1951-1980.

Notice :

- the trend is not linear (with periods of leveling off and then sharp upward trends).

Speech data

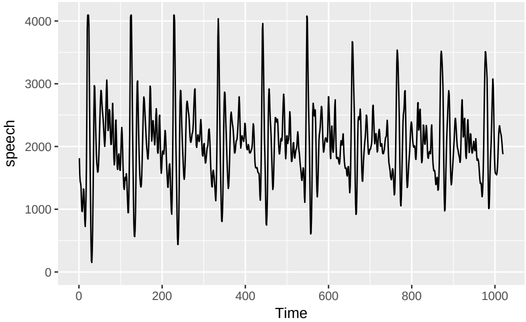


FIG. : Speech recording of the syllable "aaa ... hhh" sampled at 10,000 points per seconds with $n = 1020$ points [SS10]

Notice :

- the repetition of small wavelets.

Dow Jones Industrial Average I

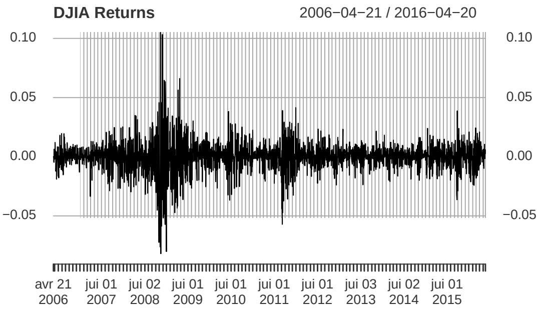


FIG. : Dayly percent change of the Dow Jones Industrial Average from April 20,2006 to April 20,2016 [SS10]

Notice :

- the mean of the series appears to be stable with an average return of approximately zero,
- the volatility (or variability) of data exhibits clustering ; that is, highly volatile periods tend to be clustered together.

Dow Jones Industrial Average I

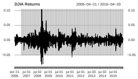


Fig. : Dayly percent change of the Dow Jones Industrial Average from April 20,2006 to April 20,2016 [SS10]

Notice :

- the mean of the series appears to be stable with an average return of approximately zero,
- the volatility (or variability) of data exhibits clustering ; that is, highly volatile periods tend to be clustered together.

ARCH and GARCH models : not covered this year !

El Niño and Fish Population I

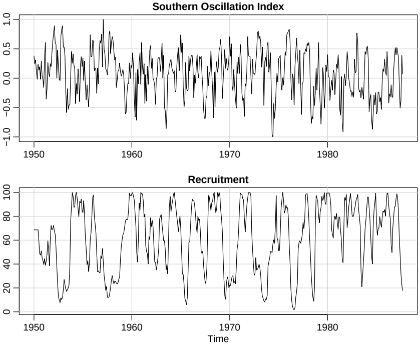


FIG. : Monthly values of an environmental series called the Southern Oscillation Index and associated Recruitment (an index of the number of new fish). [SS10]

Notice :

- SOI measures changes in air pressure related to sea surface temperatures in the central Pacific Ocean.

El Niño and Fish Population II

- The series show two basic oscillations types, an obvious annual cycle (hot in the summer, cold in the winter), and a slower frequency that seems to repeat about every 4 years.
- The two series are also related ; it is easy to imagine the fish population is dependent on the ocean temperature.

fMRI Imaging I

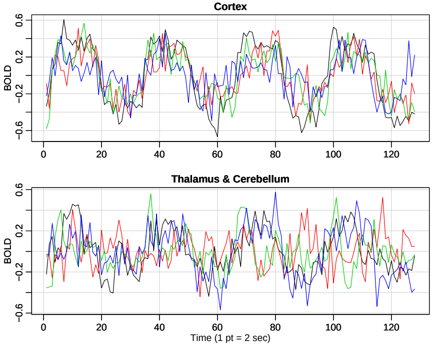


FIG. : Data collected from various locations in the brain via functional magnetic resonance imaging (fMRI) [SS10]

- Notice the periodicities.

What we are seeking

To construct models

- to describe
- to forecast

times series.

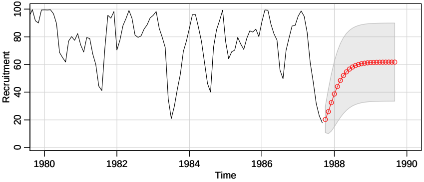


FIG.: Twenty-four month forecast for the Recruitment series shown on slide 10

Times series specificities

Definition of a Time Serie (TS)

A **time series** is a sequence $(X_t)_{t\in\mathbb{Z}}$ of random variables, i.e. a stochastic process.

- Usually the are dependencies between X_t and X_s and $\mathcal{L}(X_t) \neq \mathcal{L}(X_s)$.
- A **time series model** specifies (at least partially) the joint distribution of the sequence.

Notation : when no confusion is possible, we'll write for short X instead of $X = (X_t)_{t\in\mathbb{Z}}$.

Usual tools in statistics for i.i.d variables

- Law of Large Numbers
- Central Limit Theorem

Law of Large Numbers

Let n random variables $X_1, X_2, \cdots, X_n$ independent and identically distributed and satisfying $E(|X_1|) < \infty$ and $Var(X_1) = \sigma^2 < \infty$.
Then

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{\mathbb{P}} \mathbb{E}(X_1).$$

Based on

- identically distributed
- independency

General version of Law of Large Numbers

Let $X_k = f(\cdots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \cdots)$ where $(\varepsilon)_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables and where $f : \mathbb{R}^{\mathbb{Z}} \mapsto \mathbb{R}$. Let g be such that $\mathbb{E}(|g(X_0)|) < \infty$. Then

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(g(X_0)).$$

In this case

- the X_i are not identically distributed, but the sequence X is strictly stationary, and ergodic
- the X_i are not independent $=>$ structure on covariance between variables

Central Limit Theorem

Let n random variables $X_1, X_2, \cdots, X_n$ independent and identically distributed and satisfying $E(X_1) = \mu$ and $Var(X_1) = \sigma^2 < \infty$.
Then

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Based on

- identically distributed
- independent random variables

General version of Central Limit Theorem

Let $X_k = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{k-j}$, for $k, j \in \mathbb{Z}$ with $(\varepsilon)_{i \in \mathbb{Z}}$ being a sequence of independent and identically distributed square integrable random variables. If $\sum_{j \in \mathbb{Z}} |a_j| < \infty$ then $E(|X_0|) < \infty$ and $\mathbb{V}ar(X_0) < \infty$, and

$$Z_n = \sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \mathbb{V}ar(X_0) + 2 \sum_{k=1}^{\infty} \text{Cov}(X_0, X_k).$$

Based on In this case

- the X_i are not identically distributed, but the sequence X is strictly stationary, and ergodic
- the X_i are not independent $=>$ structure on covariance between variables

Statistics tools for inference for time series

- Those versions of LLN and TCL allow to state statistical results for most time series.
- What kind of statistical results : estimation, confidence interval, testing, model selection.
- Concepts that underly to these versions of LLN and TCL
  - Stationarity
  - Covariance and correlation structure

Stationarities

Weak stationarity

A time series (X_t) is **weakly stationary** if

- $\mathbb{E}(X_i^2) = \mathbb{E}(X_0^2) \forall i \in \mathbb{Z}$
- $\mathbb{E}(X_i) = \mathbb{E}(X_0) \forall i \in \mathbb{Z}$
- $\text{Cov}(X_0\,, X_k) = \text{Cov}(X_i\,, X_{i+k}) \forall i, k \in \mathbb{Z}$

Weak White noise (1)

$$(\omega_t) \sim WN(0, \sigma^2) \tag{1}$$

when

$$\left\{ \begin{array}{rcl} \text{Cov}(\omega_s, \omega_t) & = & 0 \; \forall s \neq t \in \mathbb{Z}^2 \\ \mathbb{E}(\omega_t) & = & 0 \; \forall t \in \mathbb{Z} \\ \mathbb{V}ar(\omega_t) & = & \sigma^2 \; \forall t \in \mathbb{Z} \end{array} \right.$$

Notation : for white noises, greek letters will be used ω, η

Stationarities

Strict stationarity

A time series (X_t) is **strictly stationary** if for all $k \geq 1$, $t_1, \ldots, t_k \in \mathbb{Z}^k$ and $h \in \mathbb{Z}$

$$\mathcal{L}(X_{t_1}, \ldots, X_{t_k}) = \mathcal{L}(X_{t_1+h}, \ldots, X_{t_k+h})$$

If $(X_t)_{t\in\mathbb{Z}}$ is **strictly stationary** and $\mathbb{E}(X_0^2) < \infty$ then $(X_t)_{t\in\mathbb{Z}}$ is **weakly stationary**.

⚠The inverse is not true.

If $(X_t)_{t\in\mathbb{Z}}$ is a Gaussian **weakly stationary** then $(X_t)_{t\in\mathbb{Z}}$ is **strictly stationary**
An i.i.d. $(X_i)i \in \mathbb{Z}$ is strictly stationary.

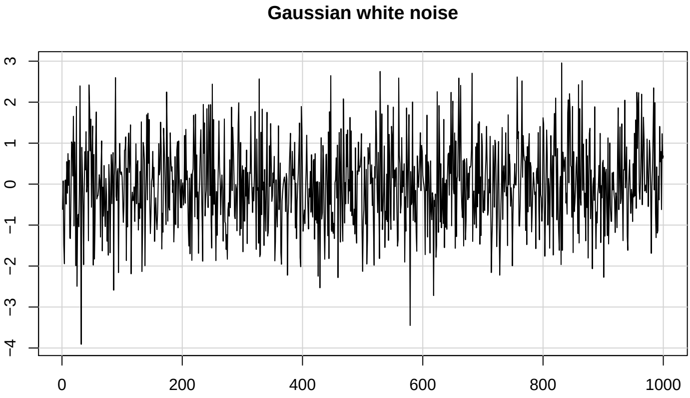
Examples

i.i.d. white noise and i.i.d. Gaussian white noise

$(\omega_t) \sim i.i.d.(0, \sigma^2)$ when

- $(\omega_t) \sim WN(0, \sigma^2)$
- and (ω_t) are i.i.d.

$(\omega_t) \sim i.i.d.\mathcal{N}(0, \sigma^2)$ if, in addition, $\omega_t \sim \mathcal{N}(0, \sigma^2)$ for all $t \in \mathbb{Z}$.



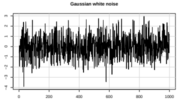
Examples

i.i.d. white noise and i.i.d. Gaussian white noise

$(\omega_t) \sim i.i.d.(0, \sigma^2)$ when

- $(\omega_t) \sim WN(0, \sigma^2)$
- and (ω_t) are i.i.d.

$(\omega_t) \sim i.i.d. \mathcal{N}(0, \sigma^2)$ if, in addition, $\omega_t \sim \mathcal{N}(0, \sigma^2)$ for all $t \in \mathbb{Z}$.



dire que ce n'est pas très intéressant pour la prédiction (cf. Bartlett)

Exercise

Check the stationarity of the following processes :

- the white noise, defined in (1)
- the random walk, defined in (4)

Properties

If $(X_t)_{t\in\mathbb{Z}}$ is stationary and if $(a_i)_{i\in\mathbb{Z}}$ is a sequence of real numbers satisfying

$$\sum_{i\in\mathbb{Z}} |a_i| < \infty,$$

then

$$Y_t = \sum_{i\in\mathbb{Z}} a_i X_{t-i},$$

is stationary.

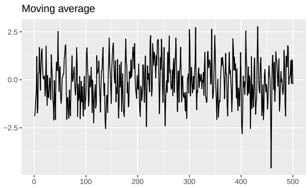
This also holds if we consider a finite sequence of real $(a_i)_{|i|\leq M}$.

Models with serial correlation I

Moving averages

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = \frac{1}{3}(\omega_{t-1} + \omega_t + \omega_{t+1}) \ \forall t \in \mathbb{Z}. \tag{2}$$



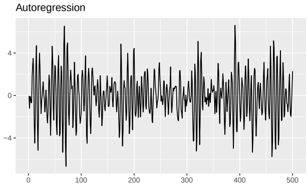
Notice the similarity with SIO and some fMRI series.

Models with serial correlation II

Autoregression

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = X_{t-1} - 0.9X_{t-2} + \omega_t \ \forall t \in \mathbb{Z}. \quad (3)$$



Notice

- the almost periodic behavior and the similarity with the speech series example
- the above definition misses initial conditions, we'll come back on that later.

Models with serial correlation III

Random walk with drift

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$\begin{cases} X_0 &= 0, \\ X_t &= \underbrace{\delta}_{\text{drift}} + \underbrace{X_{t-1}}_{\text{previous position}} + \underbrace{\omega_t}_{\text{step}} \ \forall t \in \mathbb{Z}. \end{cases} \tag{4}$$

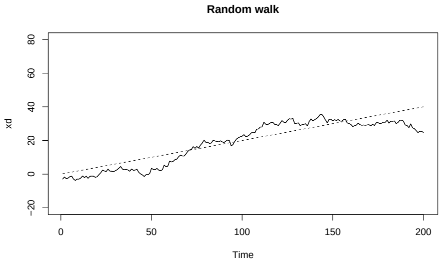


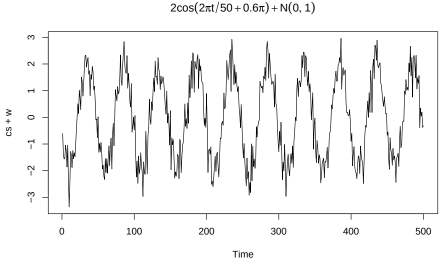
FIG. : Random walk with drift $\delta = 0.2$ (upper jagged line), with $\delta = 0$ (lower jagged line) and line with slope 0.2 (dashed line)

Models with serial correlation IV

Signal plus noise

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = \underbrace{2\cos\left(2\pi\frac{t+15}{50}\right)}_{\text{signal}} + \underbrace{\omega_t}_{\text{white noise}} \tag{5}$$



Notice the similarity with fMRI signals.

Measures of dependence

We now introduce various measures that describe the **general behavior of a process as it evolves over time.**

Mean measure

Define, for a time series $(X_t)_{t\in\mathbb{Z}}$, the mean function

$$\mu_X(t) = \mathbb{E}(X_t)\ \forall t \in \mathbb{Z}$$

when it exists.

Exercise

Compute the mean functions of

- the moving average defined in (2).
- the random walk plus drift defined in (4)
- the signal+noise model in (5)

We now introduce various measures that describe the **general behavior of a process as it evolves over time.**

Mean measure

Define, for a time series $(X_t)_{t\in\mathbb{Z}}$, the mean function

$$\mu_X(t) = \mathbb{E}(X_t)\ \forall t \in \mathbb{Z}$$

when it exists.

Exercise

Compute the mean functions of

- the moving average defined in (2).
- the random walk plus drift defined in (4)
- the signal+noise model in (5)

- $\mathbb{E}(\frac{1}{3}(\omega_{t-1} + \omega_t + \omega_{t+1})) = 0$
- $\mathbb{E}X_t = \mathbb{E}(\delta + X_{t-1} + \omega_t) = \delta + \delta + \mathbb{E}(X_{t-2}) + 0 = \cdots \delta * t$ si $X_0 = 0$.
- $\mathbb{E}X_t = 2\cos\left(2\pi \frac{t+15}{50}\right)$

Autocovariance

We now assume that for all $t \in \mathbb{Z}$, $X_t \in \mathbb{L}^2$.

Autocovariance

The autocovariance function of a time series $(X_t)_{t\in\mathbb{Z}}$ is defined as

$$\gamma_X(s,t) = \text{Cov}(X_s, X_t) = \mathbb{E}\big[(X_s - \mathbb{E}(X_s))(X_t - \mathbb{E}(X_t))\big], \quad \forall s,t \in \mathbb{Z}.$$

Properties

- It is a symmetric function $\gamma_X(s\,,t) = \gamma_X(t\,,s)$.
- It measures the linear dependence between two values of the same series observed at different times.
- In $(X_t)_{t\in\mathbb{Z}}$ is stationary, $\gamma_X(t\,,t+h) = \gamma_X(t+h\,,t) = \gamma_X(0\,,h)$. In this context we write $\gamma_X(h)$ as short for $\gamma_X(0\,,h)$.

Autocovariance of stationary time series

Theorem

The autocovariance function γ_X of a stationary time series X verifies

1. $\gamma_X(0) \geq 0$
2. $|\gamma_X(h)| \leq \gamma_X(0)$
3. $\gamma_X(h) = \gamma_X(-h)$
4. γ_X *is positive-definite.*

Furthermore, any function γ that satisfies (3) and (4) is the autocovariance of some stationary time series.

Reminder :

- A function $f : \mathbb{Z} \mapsto \mathbb{R}$ is positive-definite if for all n, the matrix F_n, with entries $(F_n)_{i,j} = f(i-j)$, is positive definite.
- A matrix $F_n \in \mathbb{R}^{n \times n}$ is positive-definite if, for all vectors $a \in \mathbb{R}^n$, $a^\top F_n a \geq 0$.

Autocovariance of stationary time series

Theorem

The autocovariance function γ_X of a stationary time series X verifies

1. $\gamma_X(0) \geq 0$
2. $|\gamma_X(h)| \leq \gamma_X(0)$
3. $\gamma_X(h) = \gamma_X(-h)$
4. γ_X *is positive-definite.*

Furthermore, any function γ that satisfies (3) and (4) is the autocovariance of some stationary time series.

Reminder :

- A function $f : \mathbb{Z} \mapsto \mathbb{R}$ is positive-definite if for all n, the matrix F_n, with entries $(F_n)_{i,j} = f(i-j)$, is positive definite.
- A matrix $F_n \in \mathbb{R}^{n \times n}$ is positive-definite if, for all vectors $a \in \mathbb{R}^n$, $a^\top F_n a \geq 0$.

Proof

1. $\gamma_X(0) = \mathrm{Cov}(X_t\,, X_t) = \mathbb{V}ar(X_t)$
2. $\gamma_X(h) = \mathbb{E}(X_h X_0) \leq \sqrt{\mathbb{E}(X_h^2)\mathbb{E}(X_0^2)} = \gamma_X(0)$
3. $\gamma_X(h) = \mathrm{Cov}(X_{t+h}\,, X_t) = \mathrm{Cov}(X_t\,, X_{t-h}) = \gamma_X(-h)$
4. For all $n \geq 0$, let $\Gamma_n \in \mathbb{R}^{n \times n}$ denote the matrix with entries

$$\Gamma_n(i,j) = \gamma_X(|i-j|) = \mathrm{Cov}(X_i, X_j).$$

For all $a \in \mathbb{R}^n$

$$a^\top \Gamma_n a = \sum_{i,j=1}^{n} a_i a_j \gamma_X(|i-j|) = \sum_{i,j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$= \mathrm{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j) = \mathbb{V}ar(\sum_{i=1}^{n} a_i X_i) \geq 0.$$

Autocovariance

Exercise

Compute the autocovariance functions of

- the white noise defined in (1)
- the moving average defined in (2)
- a moving average $X_t = \omega_t + \theta\,\omega_{t-1}$ where ω_t is a weak white noise.

Exercise

Compute the autocovariance functions of

- the white noise defined in (1)
- the moving average defined in (2)
- a moving average $X_t = \omega_t + \theta \omega_{t-1}$ where ω_t is a weak white noise.

- $\gamma_X(h) = 0$ si $h \neq 0$ and $\gamma_X(0) = \mathbb{V}ar(\omega_0) = \sigma^2$.
- $\gamma_X(h) = 0$ si $|h| > 2$, $\gamma_X(0) = (1 + 2\sigma^2/9)$, $\gamma_X(1) = 2\sigma^2/9$ and $\gamma_X(2) = \sigma^2/9$.
- $\gamma_X(h) = 0$ si $|h| \geq 2$, $\gamma_X(1) = \theta\sigma^2$ and $\gamma_X(0) = (1 + \theta^2)\sigma^2$.

Autocorrelation function (ACF)

Associated to the autocovariance function, we define the autocorrelation function.

Autocorrelation function

The ACF of a time series $(X_t)_{t\in\mathbb{Z}}$ is defined as

$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}} \quad \forall s,t \in \mathbb{Z}.$$

- It is a symmetric function $\rho_X(s\,,t) = \rho_X(t\,,s)$.
- It measures the correlation between two values of the same series observed at different times.
- In the context of stationarity, $\rho_X(t\,,t+h) = \rho_X(t+h\,,t) = \rho_X(0\,,h)$. In this context we write $\rho_X(h)$ as short for $\rho_X(0\,,h)$.

Moving average MA(1) model

⚠Warning : not to be confused with moving average.

Moving average model MA(1)

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ and construct the MA(1) as

$$X_t = \omega_t + \theta\omega_{t-1} \ \forall t \in \mathbb{Z} \tag{6}$$

Exercise

- Study its stationarity.
- Compute its ACF and its autocorrelation function.

Moving average MA(1) model

⚠Warning : not to be confused with moving average.

Moving average model MA(1)

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ and construct the MA(1) as

$$X_t = \omega_t + \theta\omega_{t-1} \; \forall t \in \mathbb{Z} \tag{6}$$

Exercise

- Study its stationarity.
- Compute its ACF and its autocorrelation function.

- Stationary
- $\gamma_X(0) = (1+\theta^2)\sigma^2$, $\gamma_X(1) = \theta\sigma^2$, $\rho(1) = \theta/(1+\theta^2)$, $\gamma_X(h) = \rho(h) = 0$, if $h \geq 2$.

Autoregressive AR(1) model

Autoregressive AR(1)

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0\,,\sigma^2)$ and construct the AR(1) as

$$X_t = \phi X_{t-1} + \omega_t \; \forall t \in \mathbb{Z}, \quad X_0 = 0. \tag{7}$$

Exercise

Assume $|\phi| < 1$. Show that under this condition X_t is stationary. And compute

- its mean function
- its ACF.

Autoregressive AR(1) model

Autoregressive AR(1)
Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0,\sigma^2)$ and construct the AR(1) as

$$X_t = \phi X_{t-1} + \omega_t \ \forall t \in \mathbb{Z}, \quad X_0 = 0. \tag{7}$$

Exercise
Assume $|\phi| < 1$. Show that under this condition X_t is stationary. And compute
- its mean function
- its ACF.

Stationary, $\mathbb{E}(X_t) = 0$, $\mathbb{E}(X_t^2) = \phi^2\mathbb{E}(X_{t-1}^2) + \sigma^2$, $\mathbb{E}(X_t^2) = \frac{\sigma^2}{1-\phi^2}$ and thus for $h \geq 0$,

$$\gamma_X(h) = \phi\gamma_X(h-1) = \phi^h\gamma_X(0) = \phi^h\frac{\sigma^2}{1-\phi^2}.$$

AR(1) as a linear process

Let $(X_t)_{t\in\mathbb{Z}}$ be the stationary solution to $X_t - \phi X_{t-1} = W_t$ where W_t is a white noise $WN(0\,,\sigma^2)$. If $|\phi| < 1$,

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

is a solution. This infinite sum. converges in mean square, since $|\Phi| < 1$ implies $\sum_{j\geq 0} |\phi^j| < \infty$. Furthermore, X_t is the unique stationary solution, since we can check that any other stationary solution Y_t is the mean square limit :

$$\lim_{n\to\infty} \mathbb{E}\left(Y_t - \sum_{i=0}^{n-1} \Phi^i W_{t-i}\right)^2 = \lim_{n\to\infty} \mathbb{E}\left(\phi^n Y_{t-n}\right)^2 = 0.$$

As a conclusion, if $|\phi| < 1$ then X_t can be written as

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

Now, if $|\Phi| = 1$? or if $|\Phi| > 1$?

AR(1) another formulation

The equation $X_t - \Phi X_{t-1} = W_t$ where W_t is a white noise $WN(0\,,\sigma^2)$ is equivalent to

$$X_t - \Phi X_{t-1} = W_t \Longleftrightarrow \Phi(B)X_t = W_t,$$

where B is the back-shift operator $BX_t = X_{t-1}$ and where $\Phi(z) = 1 - \Phi z$.
Also, we can write

$$X_t = \sum_{j=0}^{\infty} \Phi^j W_{t-j} = \sum_{j=0}^{\infty} \Phi^j B^j W_t = \Pi(B) W_t.$$

AR(1) another formulation

With these notations,

$$\Pi(B) = \sum_{j=0}^{\infty} \Phi^j B^j \text{ and } \Phi(B) = 1 - \Phi(B),$$

we can check that $\Pi(B) = \Phi^{-1}(B)$: thus

$$\Phi(B)X_t = W_t \Longleftrightarrow X_t = \Pi(B)W_t.$$

AR(1) another formulation

Notice that manipulating operators like $\Phi(B)$ or $\Pi(B)$ is like manipulating polynomials with

$$\frac{1}{1-\Phi z} = 1 + \Phi z + \Phi^2 z^2 + \cdots +,$$

provided that $|\Phi| < 1$ and $|z| \leq 1$.
If $|\Phi| > 1$, $\Pi(B)W_t$ does not converge. But we can rearrange and write

$$X_{t-1} = \frac{1}{\Phi}X_t - \frac{1}{\Phi}W_t,$$

so that we can check that the unique stationary solution is

$$X_t = -\sum_{j=1}^{\infty} \Phi^{-j} W_{t+j}.$$

Notice that here X_t depends on the future of W_t.

$\Longleftrightarrow$ notions of causality and invertibility

Linear process

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0\,,\sigma^2)$ and define the linear process X as follows

$$X_t = \mu + \sum_{j\in\mathbb{Z}} \psi_j \omega_{t-j} \ \forall t \in \mathbb{Z} \tag{8}$$

where $\mu \in \mathbb{R}$ and (ψ_j) satisfies $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$. X

Theorem

The series in Equation (8) *converges in* $\mathbb{L}^2$ *and the linear process* X *defined above is stationary. (see Proposition 3.1.2 in [BD13]).*

Exercise

Compute the mean and autocovariance functions of the linear process $(X_t)_{t\in\mathbb{Z}}$.

Linear processes

Linear process

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$ and define the linear process X as follows

$$X_t = \mu + \sum_{j\in\mathbb{Z}} \psi_j \omega_{t-j} \ \forall t \in \mathbb{Z} \tag{8}$$

where $\mu \in \mathbb{R}$ and (ψ_j) satisfies $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$. X

Theorem

The series in Equation (8) converges in $\mathbb{L}^2$ and the linear process X defined above is stationary. (see Proposition 3.1.2 in [BD13]).

Exercise

Compute the mean and autocovariance functions of the linear process $(X_t)_{t\in\mathbb{Z}}$.

- Proof based on Cauchy criterion
- $\gamma_X(h) = \sigma^2 \sum_{j\in\mathbb{Z}} \phi_j \phi_{j+h}$

Examples of linear processes

Exercises

- Show that the following processes are particular linear processes
  - the white noise process
  - the MA(1) process.
- Consider a linear process as defined in (8), put $\mu = 0$,

$$\begin{cases} \psi_j = \phi^j & \text{if } j \geq 0 \\ \psi_j = 0 & \text{if } j < 0 \end{cases}$$

  and suppose $|\phi| < 1$. Show that X is in fact an AR(1) process.

Examples of linear processes

Exercises

- Show that the following processes are particular linear processes
  - the white noise process
  - the MA(1) process.
- Consider a linear process as defined in (8), put $\mu = 0$,

$$\begin{cases} \psi_j = \phi^j & \text{if } j \geq 0 \\ \psi_j = 0 & \text{if } j < 0 \end{cases}$$

  and suppose $|\phi| < 1$. Show that X is in fact an AR(1) process.

a

Linear prediction

Linear predictor

Best least square estimate of Y given X is $\mathbb{E}(Y|X)$. Indeed

$$\min_f \mathbb{E}[(Y - f(X))^2] = \min_f \mathbb{E}\left[\mathbb{E}[(Y - f(X))^2|X]\right] = \mathbb{E}\left[\mathbb{E}[(Y - \mathbb{E}(Y|X))^2|X]\right].$$

Similarly, the best least squares estimate of X_{n+h} given X_n is

$$f(X_n) = \mathbb{E}(X_{n+h}|X_n).$$

For Gaussian and stationary $(X_t)_{t\mathbb{Z}}$, the best estimate of X_{n+h} given $X_n = x_n$ is

$$f(X_n) = \mu + \rho(h)(x_n - \mu),$$

and

$$\mathbb{E}(X_{n+h} - f(X_n))^2 = \sigma^2(1 - \rho(h)^2).$$

Prediction accuracy improves as $|\rho(h)| \underset{n\to\infty}{\longrightarrow} 1$.
Predictor is linear since $f(x_n) = \mu(1 - \rho(h)) + \rho(h)x_n$.

ACF and prediction

Linear predictor and ACF

Let X be a stationary time series with ACF ρ. The linear predictor $\hat{X}_{n+h}^{\{n\}}$ of X_{n+h} given X_n is defined as

$$\hat{X}_{n+h}^{\{n\}} = \operatorname*{argmin}_{a,b} \mathbb{E}\Big(\big(X_{n+h} - (aX_n + b)\big)^2 \Big) = \rho(h)(X_n - \mu) + \mu$$

Exercise

Prove the result.

Notice that

- linear prediction needs only second order statistics, we'll see later that it is a crucial property for forecasting.
- the result extends to longer histories $(X_n, X_{n-1}, \ldots)$.

2018-02-22

ACF and prediction

Linear predictor and ACF

Let X be a stationary time series with ACF ρ. The linear predictor $\hat{X}_{n+h}^{\{n\}}$ of X_{n+h} given X_n is defined as

$$\hat{X}_{n+h}^{\{n\}} = \underset{a,b}{\operatorname{argmin}} \mathbb{E}\Big(\big(X_{n+h} - (aX_n + b)\big)^2\Big) = \rho(h)(X_n - \mu) + \mu$$

Exercise

Prove the result.

Notice that

- linear prediction needs only second order statistics, we'll see later that it is a crucial property for forecasting.
- the result extends to longer histories $(X_n, X_{n-1}, \ldots)$.

a

Chapter 2 : Estimation of the mean and of the ACF

Estimation of the mean

Suppose that X is a stationary time series and recall that for all $t, h \in Z$,

$$\mu_X(t) = \mu, \;\; \gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) \;\text{ and }\; \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Theorem

Let $(X_i)_{i\in\mathbb{Z}}$ is weakly stationary with autocovariance function γ. Then $\bar{X} \underset{n\to\infty}{\overset{\mathbb{P}}{\longrightarrow}} \mathbb{E}(X_0)$ if and only if $\frac{1}{n}\sum_{i=1}^{n}\gamma_X(i) \underset{n\to\infty}{\longrightarrow} 0$.

Estimation of the ACF

Estimation

Consider observations $X_1, \dots, X_n$ (from the strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$), with $\mathbb{E}(X_0^2) < \infty$ and satisfying $\frac{1}{n}\sum_{i=1}^{n} \gamma_X(i) \underset{n\to\infty}{\longrightarrow} 0$. We can compute

- the **sample mean** $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$
- $\frac{1}{n}\sum_{i=1}^{n-p} X_i X_{i+p} \underset{n\to\infty}{\overset{\mathbb{P}}{\longrightarrow}} \mathbb{E}(X_0 X_p)$.
- the **sample autocovariance function**

$$\hat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}) \quad \forall -n < h < n$$

- the **sample autocorrelation function**

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

⚠Warning : $\gamma_X(h) = \text{Cov}(X_t\,, X_{t+h})$ but the sample autocorrelation function is not the corresponding empirical covariance ! !
Indeed

$$\frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}) \neq$$

$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}\left(X_{t+|h|} - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_{t+h}\right)\left(X_t - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_t\right)$$

⚠Warning : $\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h})$ but the sample autocorrelation function is not the corresponding empirical covariance!!
Indeed

$$\frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}) \neq$$

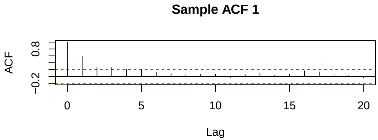
$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_{t+h})(X_t - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_t)$$

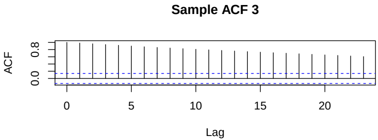
The sum in (1.34) runs over a restricted range because xt+h is not available for t + h > n. The estimator in (1.34) is preferred to the one that would be obtained by dividing by n−h because (1.34) is a non-negative definite function. The autocovariance function, (h), of a stationary process is non-negative definite (see Problem 1.25) ensuring that variances of linear combinations of the variates xt will never be negative. And, because var(a1xt1 + · · · + anxtn) is never negative, the estimate of that variance should also be non-negative. The estimator in (1.34) guarantees this result, but no such guarantee exists if we divide by n − h ; this is explored further in Problem 1.25.

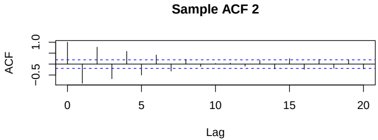
Examples of sample ACF

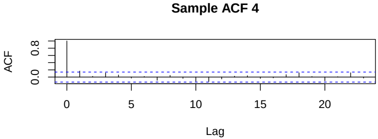
Exercise

Can you find the generating time series models (white noise, MA(1), AR(1), random noise with drift) associated with the sample ACF ?









Examples of sample ACF

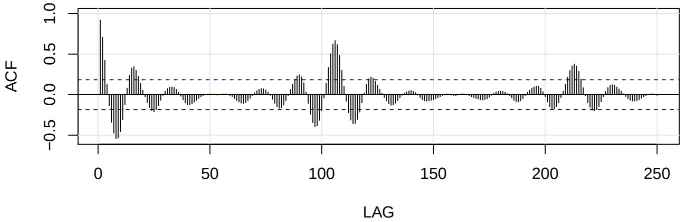


FIG.: The ACF of speech data example on slide 8

Notice :

- the regular repetition of short peaks with decreasing amplitude.

Sample ACF

Behavior

Time Series	**Sample ACF** $\rho_X(h)$
White Noise	zero
Trend	Slow decay
Periodic	Periodic
$MA(q)$	Zero for $|h| > q$
$AR(p)$	Decays to zero exponentially

Properties of empirical sum : Soft version of Law of Large Numbers

Let $X_k = f(\cdots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \cdots)$ where $(\varepsilon)_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables and where $f : \mathbb{R}^{\mathbb{Z}} \mapsto \mathbb{R}$. Let g be such that $\mathbb{E}(|g(X_0)|) < \infty$. Then

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(g(X_0)).$$

<u>Comments:</u>

- If $X_k = f(\cdots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \cdots)$ where $(\varepsilon)_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables, then $(X_k)_{k\in\mathbb{Z}}$ is a strictly stationary time series.
- More generally, for g such that $\mathbb{E}(|g(\cdots, X_{-1}, X_0, X_{+1}, \cdots)|) < \infty$, then

$$\frac{1}{n}\sum_{k=1}^{n} g(\cdots, X_{k-1}, X_k, X_{k+1}, \cdots) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[g(\cdots, X_{-1}, X_0, X_{+1}, \cdots)].$$

- Most time series $(X_k)_{k\in\mathbb{Z}}$ satisfy that there exist $(\varepsilon)_{i\in\mathbb{Z}}$, a sequence of independent and identically distributed random variables such that

$$X_k = f(\cdots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \cdots).$$

Properties of $\bar{X}_n$: Another version of Law of Large Number

If $(X_k)_{k\in\mathbb{Z}}$ is is a stationary time series and the sample mean verifies

$$\mathbb{E}(\bar{X}_n) = \mu \text{ and } \mathbb{V}ar(\bar{X}_n) = \frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)\gamma_X(k).$$

Moreover

$$\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{L}_2} \mu \text{ if and only if } \frac{1}{n}\sum_{k=-n}^{n}\gamma_X(k) \underset{n\to\infty}{\longrightarrow} 0.$$

In that case

$$n\mathbb{V}ar(\bar{X}_n) \underset{n\to\infty}{\longrightarrow} \sum_{k=-\infty}^{\infty}\gamma_X(k) = \sigma^2\sum_{k=-\infty}^{\infty}\rho_X(k).$$

Comments:

- It concerns time series which are not necessary function of i.i.d. variables such as in "soft version".
- Less general that the "soft version" since it holds only for $\bar{X}_n$ and not for function of X_k.
  See Appendix A [SS10]

Properties of $\bar{X}_n$: Another version of Law of Large Number

If $(X_k)_{k\in\mathbb{Z}})$ is is a stationary time series and the sample mean verifies

$$\mathbb{E}(\bar{X}_n) = \mu \text{ and } \mathbb{V}ar(\bar{X}_n) = \frac{1}{n}\sum_{k=-(n-1)}^{n-1}(1 - \frac{|k|}{n})\gamma_X(k).$$

Moreover

$$\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{L}_2} \mu \text{ if and only if } \frac{1}{n}\sum_{k=-n}^{n}\gamma_X(k) \underset{n\to\infty}{\longrightarrow} 0.$$

In that case

$$n\mathbb{V}ar(\bar{X}_n) \underset{n\to\infty}{\longrightarrow} \sum_{k=-\infty}^{\infty}\gamma_X(k) = \sigma^2\sum_{k=-\infty}^{\infty}\rho_X(k).$$

Comments:

- It concerns time series which are not necessary function of i.i.d. variables such as in "soft version".
- Less general that the "soft version" since it holds only for $\bar{X}_n$ and not for function of X_k.
  See Appendix A [SS10]

a

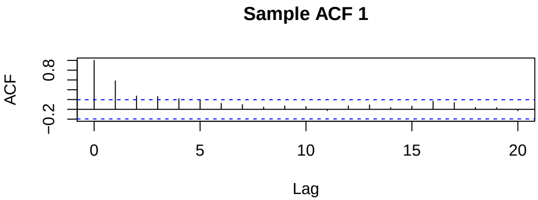
Large sample property : asymptotic normality

Theorem

Under general conditions, if X is a white noise, then for n large, the sample ACF, $\hat{\rho}_X(h)$, for $h = 1, 2, \ldots, H$, where H is fixed but arbitrary, is approximately normally distributed with zero mean and standard deviation given by

$$\sigma_{\hat{\rho}_X(h)} = \frac{1}{\sqrt{n}}$$

Consequence : only the peaks outside of $\pm 2/\sqrt{n}$ may be considered to be significant.



See Appendix A [SS10]

Chapter 3 : ARMA models

Introduction

We now consider that we have estimated the trend and seasonal components of

$$Y_t = T_t + S_t + X_t$$

and focus on X_t. **Aim of the chapter** : to propose to model the time series X via ARMA models. They allow

- to describe this time series
- to forecast.

Key fact : we know that

- for every stationary process with autocovariance function γ verifying $\lim_{h\to\infty} \gamma_X(h) = 0$, it is possible to find an ARMA process with the same autocovariance function, see [BD13].
- The Wold decomposition (see [SS10] Appendix B) also plays an important role. Its says that every stationary process is the sum of a MA(∞) process and a deterministic process.

AR(1)

Exercice

Consider a time series X following the AR(1) model

$$X_t = \phi X_{t-1} + \omega_t \ \forall t \in \mathbb{Z}.$$

1. Show that for all $k > 0$ $X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \omega_{t-j}$.
2. Assume that $|\phi| < 1$ and prove $X_t \overset{\mathbb{L}^2}{=} \sum_{j=0}^{\infty} \phi^j \omega_{t-j}$.
3. Assume now that $|\phi| > 1$ and prove that
   3.1 $\sum_{j=0}^{k-1} \phi^j \omega_{t-j}$ does not converge in $\mathbb{L}^2$
   3.2 one can write $X_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}$
   3.3 Discuss why the case $|\phi| > 1$ is useless.

The case where $|\phi| = 1$ is a random walk (slide 4) and we already proved that this is not a stationary time series.

AR(1)

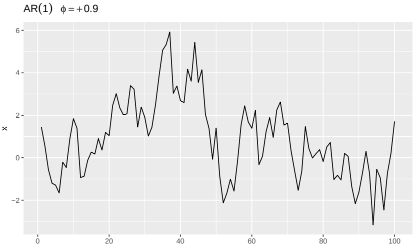
Exercice
Consider a time series X following the AR(1) model

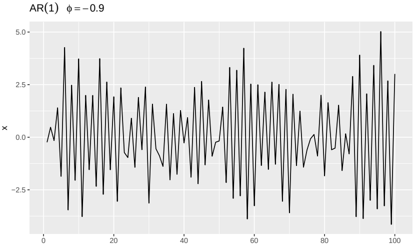
$$X_t = \phi X_{t-1} + \omega_t \ \forall t \in \mathbb{Z}.$$

1. Show that for all $k > 0$ $X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \omega_{t-j}$.
2. Assume that $|\phi| < 1$ and prove $X_t \overset{\mathbb{L}^2}{=} \sum_{j=0}^{\infty} \phi^j \omega_{t-j}$.
3. Assume now that $|\phi| > 1$ and prove that
   - 3.1 $\sum_{j=0}^{k-1} \phi^j \omega_{t-j}$ does not converge in $\mathbb{L}^2$
   - 3.2 one can write $X_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}$
   - 3.3 Discuss why the case $|\phi| > 1$ is useless.

The case where $|\phi| = 1$ is a random walk (slide 4) and we already proved that this is not a stationary time series.

a


AR(1) ϕ = +0.9
x
6
4
2
0
−2
0
20
40
60
80
100


AR(1) ϕ = −0.9
x
5.0
2.5
0.0
−2.5
0
20
40
60
80
100

Note on polynomials in $\mathbb{R}$

Notice that manipulating operators like $\phi(\mathrm{B})$ is like manipulating polynomials with complex variables.
In particular :

$$\frac{1}{1-\phi z} = 1 + \phi z + \phi^2 z + \ldots$$

provided that $|\phi| < 1$ and $|z| \leq 1$. More generally, if the polynome P has not in the unique circle, then for $|z| \leq 1$ we have

$$\frac{1}{P(z)} = \sum_{i=0}^{\infty} \psi_i z^i, \text{ with } \sum_{i=0}^{\infty} |\psi_i| < \infty.$$

For instance

$$\frac{1}{1-\rho z} = \sum_{i=0} \rho^i z^i, \text{ for } |\rho| < 1 \text{ and } |z| \leq 1.$$

Polynomial of the back-shift operator

Consider now $P(B)$ where $BX_t = X_{t-1}$. Can we write

$$\tilde{P}(B) = \sum_{i=0}^{\infty} \psi_i B^i, \text{ with } \tilde{P}(B) \circ P(B) = P(B) \circ \tilde{P}(B) = Id?$$

- If

$$A(B) = \sum_{i \in \mathbb{Z}} a_i B^i, \text{ and } C(B) = \sum_{i \in \mathbb{Z}} c_i B^i,$$

with $\sum_i |a_i| < \infty$ and $\sum_i |c_i| < \infty$ then

$$A(B) \circ C(B)(X_t) = \sum_{k=-\infty}^{+\infty} d_k X_{t-k} = C(B) \circ A(B)(X_t),$$

with

$$d_k = \sum_{i=-\infty}^{+\infty} a_i c_{k-i} = \sum_{j=-\infty}^{+\infty} a_{k-j} c_j = a \star c, \sum_j |d_j| < \infty$$

- If P has not roots in the circle unit, then

$$\tilde{P}(B) \circ P(B) = P(B) \circ \tilde{P}(B) = Id.$$

Consequently $\tilde{P}(B) = \sum_{i=0}^{\infty} \psi_i B^i$ **exits and is the inverse of** $P(B)$.

Causality

Causal linear process

A linear process X is said to be **causal** (a causal function of W_t) when there is

- a power series $\Psi : \Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 \, , \ldots,$
- with $\sum_{j=0}^{\infty} |\psi_j| < \infty$
- and $X_t = \Psi(B) \, \omega_t$

ω is a white noise $WN(0 \, , \sigma^2)$.
In this case X_t is $\sigma\{\, \omega_t \, , \omega_{t-1} \, , \ldots\}$-measurable.

• We will exclude non-causal AR models from consideration. In fact this is not a restriction because we can find causal counterpart to such process.
• Causality is a property of $(X_t)_t$ and $(W_t)_t$.

AR(1) and causality

Consider the $AR(1)$ process defined by $\Phi(B)X_t = W_t = (1 - \phi B)W_t$.
This process $\Phi(B)X_t = W_t$ is **causal** and **stationary**

- iff $|\phi| < 1$
- iff the root z_1 of the polynomial $\Phi(z) = 1 - \phi z$ satisfies $|z_1| > 1$.
- If $|\phi| > 1$ we can define an equivalent causal model

$$X_t - \phi^{-1}X_{t-1} = \tilde{W}_t,$$

where $\tilde{W}_t$ is a new white noise sequence.

- If $|\phi| = 1$ the $AR(1)$ process is not stationnary.
- If X_t is an $MA(1)$, is is always causal.

Exercise

Consider the non-causal AR(1) model $X_t = \phi X_{t-1} + \omega_t$ with $|\phi| > 1$ and suppose that $\omega \sim i.i.d\ \mathcal{N}(0, \sigma^2)$

1. Which distribution has X_t ?
2. Define the time series $Y_t = \phi^{-1}Y_{t-1} + \eta_t$ with $\eta \sim i.i.d\ \mathcal{N}(0, \sigma^2/\phi^2)$. Prove that X_t and Y_t have the same distribution.

AR(1) and causality

Consider the $AR(1)$ process defined by $\Phi(B)X_t = W_t = (1 - \phi B)W_t$.
This process $\Phi(B)X_t = W_t$ is **causal** and **stationary**
- iff $|\phi| < 1$
- iff the root z_1 of the polynomial $\Phi(z) = 1 - \phi z$ satisfies $|z_1| > 1$.
- If $|\phi| > 1$ we can define an equivalent causal model

$$X_t - \phi^{-1}X_{t-1} = \bar{W}_t,$$

where $\bar{W}_t$ is a new white noise sequence.
- If $|\phi| = 1$ the $AR(1)$ process is not stationnary.
- If X_t is an $MA(1)$, is is always causal.

Exercise

Consider the non-causal AR(1) model $X_t = \phi X_{t-1} + \omega_t$ with $|\phi| > 1$ and suppose that $\omega \sim i.i.d\ \mathcal{N}(0, \sigma^2)$

1. Which distribution has X_t ?
2. Define the time series $Y_t = \phi^{-1}Y_{t-1} + \eta_t$ with $\eta \sim i.i.d\ \mathcal{N}(0, \sigma^2/\phi^2)$. Prove that X_t and Y_t have the same distribution.

solve this ex...

Autoregressive model

AR(p)

An autoregressive model of order p is of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \omega_t \ \forall t \in \mathbb{Z}$$

where X is assumed to be stationary and ω is a white noise $WN(0\,,\sigma^2)$. We will write more concisely

$$\Phi(B)X_t = \omega_t \ \forall t \in \mathbb{Z}$$

where Φ is the polynomial of degree p, $\Phi(z) = (1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p)$.

Without loss of generality, we assume that each X_t is centered.

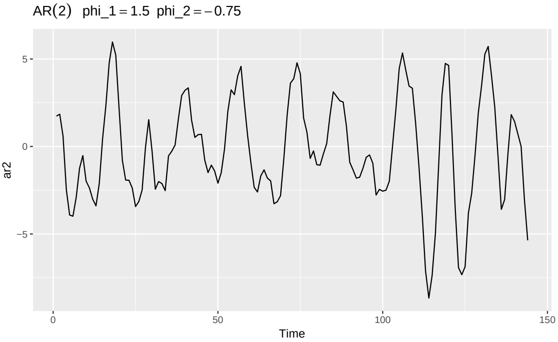
Condition of existence and causality of AR(p)

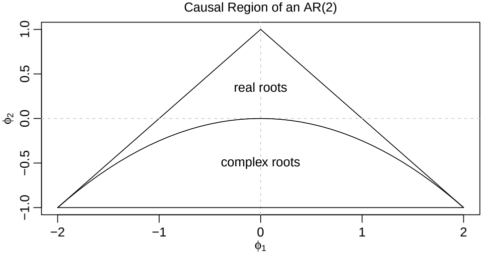
A stationary solution to $\Phi(B)X_t = \omega_t \; \forall t \in \mathbb{Z}$ exists if and only if

$$\Phi(z) = 0 \Longrightarrow |z| \neq 1.$$

In this case, this defines an AR(p) process, which is causal iff in addition

$$\Phi(z) = 0 \Longrightarrow |z| > 1.$$


AR(2) phi_1 = 1.5 phi_2 = −0.75
5
0
−5
ar2
0
50
100
150
Time


Causal Region of an AR(2)
1.0
0.5
0.0
−0.5
−1.0
ϕ_2
real roots
complex roots
−2
−1
0
1
2
ϕ_1

Recall Causality

Causal linear process

A linear process X is said to be **causal** (a causal function of W_t) when there is

- a power series $\Psi : \Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 , \ldots,$
- with $\sum_{j=0}^{\infty} |\psi_j| < \infty$
- and $X_t = \Psi(B)\,\omega_t$

ω is a white noise $WN(0\,,\sigma^2)$.
In this case X_t is $\sigma\{\,\omega_t\,,\omega_{t-1}\,,\ldots\}$-measurable.

- Causality is a property of $(X_t)_t$ and $(W_t)_t$.
- Calculating Ψ for an $AR(p)$?

AR(p) and causality

Consider and $AR(p)$ process $\Phi(B)X_t = W_t \iff X_t = \Psi(B)W_t$ where ω is a white noise $WN(0\,,\sigma^2)$.
We get that

$$1 = \Psi(B)\Phi(B) \iff 1 = (\psi_0 + \psi_1 B + \cdots)(1 - \phi_1 B - \cdots - \phi_p B^p),$$

$\iff$

$$\begin{aligned} 1 &= \psi_0 \\ 0 &= \psi_1 - \phi_1\psi_0 \\ 0 &= \psi_2 - \phi_1\psi_1 - \phi_2\psi_0 \\ \cdots \end{aligned}$$

$\iff$

$$1 = \psi_0$$

$$0 = \psi_j \quad (j < 0), 0 = \Phi(B)\psi_j.$$

We can solve these linear difference equations in several ways :

- numerically
- by guessing the form of a solution and using an inductive proof
- by using the theory of linear difference equations

MA(1) and invertibility

Define

$$X_t = W_t + \theta W_{t-1} = (1+\theta B)W_t.$$

- If $|\theta| < 1$ we can write

$$(1+\theta B)^{-1}X_t = W_t \iff (1-\theta B+\theta^2 B^2-\theta^3 B^3+\cdots)X_t = W_t \iff \sum_{j=0}^{\infty}(-\theta)^j X_{t-j} = W_t.$$

That is, we can write W_t as a causal function of X_t.
We say that this $MA(1)$ is **invertible**.
- if$|\theta| > 1$, the sum $\sum_{j=0}^{\infty}(-\theta)^j X_{t-j}$ diverges, but we can write

$$W_{t-1} = -\theta^{-1}W_t + \theta^{-1}X_t.$$

Just like the noncausal $AR(1)$, we can show that

$$W_t = -\sum_{j=1}^{\infty}(-\theta)^{-j}X_{t+j}.$$

Thus, W_t is written as a linear function of X_t, but it is non causal. We say that this $MA(1)$ is non invertible.

Moving average model

MA(q)

An moving average model of order q is of the form

$$X_t = \omega_t + \theta_1\omega_{t-1} + \theta_2\omega_{t-2} + \ldots + \theta_q\omega_{t-q} \;\forall t \in \mathbb{Z}$$

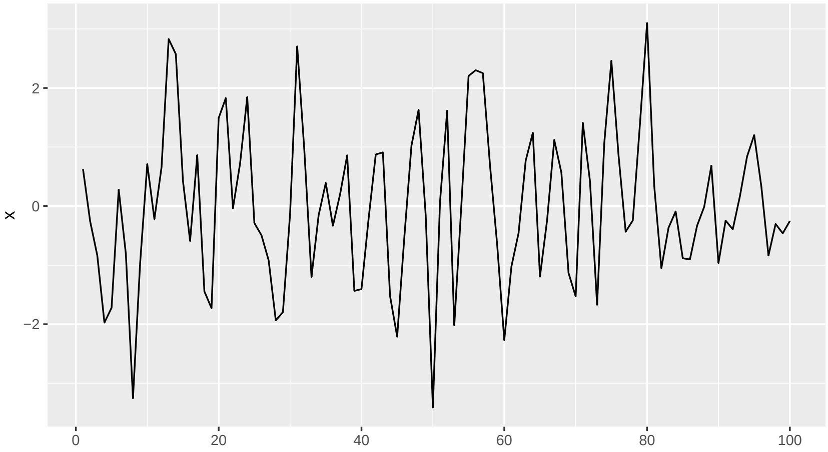
where ω is a white noise $WN(0\,, \sigma^2)$. We will write more concisely

$$X_t = \Theta(B)\omega_t \;\forall t \in \mathbb{Z}$$

where θ is the polynomial of degree q $\theta(x) = (1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q)$.

Unlike the AR model, the **MA model is stationary for any values of the thetas.**

MA(1) $\theta = +0.9$



Invertibility I

Invertibility of a MA(1) process

Consider the MA(1) process

$$X_t = \omega_t + \theta\omega_{t-1} = (1 + \theta B)\omega_t \; \forall t \in \mathbb{Z}$$

where ω is a white noise $WN(0, \sigma^2)$.
Show that

- If $|\theta| < 1$, $\omega_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$
- If $|\theta| > 1$, $\omega_t = -\sum_{j=1}^{\infty} (-\theta)^{-j} X_{t+j}$

In the first case, X is invertible.

Invertibility

A linear process X is **invertible** when there is

- a power series $\Pi : \Pi(x) = \pi_0 + \pi_1 x + \pi_2 x^2, \ldots,$
- with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- and $\omega_t = \Pi(B) X_t$

ω is a white noise $WN(0, \sigma^2)$.

Invertibility I

Invertibility of a MA(1) process

Consider the MA(1) process

$$X_t = \omega_t + \theta\omega_{t-1} = (1 + \theta B)\omega_t \; \forall t \in \mathbb{Z}$$

where ω is a white noise $WN(0, \sigma^2)$.
Show that

- If $|\theta| < 1, \;\omega_t = \sum_{j=0}^{\infty}(-\theta)^j X_{t-j}$
- If $|\theta| > 1, \;\omega_t = -\sum_{j=1}^{\infty}(-\theta)^{-j} X_{t+j}$

In the first case, X is invertible.

Invertibility

A linear process X is **invertible** when there is

- a power series $\Pi : \Pi(x) = \pi_0 + \pi_1 x + \pi_2 x^2, \ldots,$
- with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- and $\omega_t = \Pi(B)X_t$

ω is a white noise $WN(0, \sigma^2)$.

solve this ex...

Invertibility II

Exercise

Consider the non-invertible MA(1) model $X_t = \omega_t + \theta\omega_{t-1}$ with $|\theta| > 1$ and suppose that $\omega \sim i.i.d.\mathcal{N}(0, \sigma^2)$

1. Which distribution has X_t?
2. Can we define an invertible time series Y defined through a new Gaussian white noise η such that X_t and Y_t have the same distribution $(\forall t)$?

Invertibility of a MA(1) process

Consider the MA(1) process

$$X_t = \omega_t + \theta\omega_{t-1} = (1+\theta B)\omega_t \ \forall t \in \mathbb{Z}$$

where ω is a white noise $WN(0\,,\sigma^2)$. $(X_t)_t$ is invertible • iff $|\theta| < 1$
• iff the root z_1 of the polynomial $\Theta(z) = 1 + \theta z$ satisfies $|z_1| > 1$.
• If $|\theta| > 1$ we can define an equivalent invertible model in term of a new white noise sequence.
• Is and $AR(1)$ invertible ?

Autoregressive moving average model

Autoregressive moving average model ARMA(p, q)

An ARMA(p, q) process $(X_t)_{t\in\mathbb{Z}}$ is a stationary process that is defined through

$$\Phi(B)X_t = \Theta(B)\omega_t$$

where $\omega \sim WN(0, \sigma^2)$, Φ is a polynomial of order p, Θ is a polynomial of order q and Φ and Θ have no common factors.

Can accurately approximate many stationary processes

• For any stationary process with autocovariance γ, and any $k > 0$, there is an ARMA process $(X_t)_t$ for which

$$\gamma_X(h) = \gamma(h), \quad h = 0, 1, \cdots, k.$$

• Usually we insist that $\phi_p \neq 0, \theta_q \neq 0$ and that the polynomials Φ and Θ have no common factors. This implies it is not a lower order ARMA model.

Autoregressive moving average model

Examples ARMA(p, q)

- $AR(p) = ARMA(p, 0) : \Theta(B) \equiv 1$.
- $MA(q) = ARMA(0, q) : \Phi(B) \equiv 1$.
- $WN = ARMA(0, 0) : \Theta(B) = \Phi(B) \equiv 1$.

Exercise

Consider the process X defined by $X_t - 0.5X_{t-1} = \omega_t - 0.5\,\omega_{t-1}$. Is it trully an ARMA(1,1) process ?

Autoregressive moving average model

Examples ARMA(p, q)

- $AR(p) = ARMA(p, 0) : \Theta(B) \equiv 1$.
- $MA(q) = ARMA(0, q) : \Phi(B) \equiv 1$.
- $WN = ARMA(0, 0) : \Theta(B) = \Phi(B) \equiv 1$.

Exercise

Consider the process X defined by $X_t - 0.5X_{t-1} = \omega_t - 0.5\omega_{t-1}$. Is it trully an ARMA(1,1) process ?

solve this ex...

Recall Causality and invertibility

Causality

A linear process $(X_t)_t$ is said to be **causal** (a causal function of W_t) when there is

- a power series $\Psi : \Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 , \ldots,$
- with $\sum_{j=0}^{\infty} |\psi_j| < \infty$
- and $X_t = \Psi(B)\,\omega_t$

ω is a white noise $WN(0\,, \sigma^2)$.

Invertibility

A linear process $(X_t)_t$ is **invertible** when there is

- a power series $\Pi : \Pi(x) = \pi_0 + \pi_1 x + \pi_2 x^2 , \ldots,$
- with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- and $\omega_t = \Pi(B) X_t$

ω is a white noise $WN(0\,, \sigma^2)$.

Stationarity, causality and invertibility

Theorem

Consider the equation $\Phi(B)X_t = \Theta(B)\,\omega_t$, where Φ and Θ have no common factors.

- There exists a stationary solution iff

$$\Phi(z) = 0 \Leftrightarrow |z| \neq 1.$$

- This process ARMA(p, q) is causal iff

$$\Phi(z) = 0 \Leftrightarrow |z| > 1.$$

- It is invertible iff the roots of $\Theta(z)$ are outside the unit circle.

Exercise

Discuss the stationarity, causality and invertibility of $(1 - 1\,.5B)X_t = (1 + 0\,.2B)\,\omega_t$.

Stationarity, causality and invertibility

Theorem

Consider the equation $\Phi(B)X_t = \Theta(B)\,\omega_t$, where Φ and Θ have no common factors.

- There exists a stationary solution iff
$$\Phi(z) = 0 \Leftrightarrow |z| \neq 1.$$
- This process ARMA(p, q) is causal iff
$$\Phi(z) = 0 \Leftrightarrow |z| > 1.$$
- It is invertible iff the roots of $\Theta(z)$ are outside the unit circle.

Exercise

Discuss the stationarity, causality and invertibility of $(1 - 1.5B)X_t = (1 + 0.2B)\,\omega_t$.

solve this ex...

Theorem

Let X be an ARMA process defined by $\Phi(B)X_t = \Theta(B)\,\omega_t$.
If

$$\forall |z| = 1 \;\; \theta(z) \neq 0,$$

then there are polynomials $\tilde{\phi}$ and $\tilde{\theta}$ and a white noise sequence $\tilde{\omega}$ such that X satisfies

- $\tilde{\Phi}(B)X_t = \tilde{\Theta}(B)\tilde{\omega}_t$,
- and is a causal,
- invertible ARMA process.

We can now **consider only causal and invertible ARMA processes.**

Theorem

Let X be an ARMA process defined by $\Phi(B)X_t = \Theta(B)\,\omega_t$.

1. If
$$\forall |z| = 1 \;\; \theta(z) \neq 0,$$
then there are polynomials $\tilde{\phi}$ and $\tilde{\theta}$ and a white noise sequence $\tilde{\omega}$ such that X satisfies
    - $\tilde{\Phi}(B)X_t = \tilde{\Theta}(B)\tilde{\omega}_t$,
    - and is a causal,
    - invertible ARMA process.

We can now **consider only causal and invertible ARMA processes.**

The linear process representation of an ARMA

Causal and invertible representations

Consider a causal, invertible ARMA process defined by $\Phi(B)X_t = \Theta(B)\,\omega_t$. It can be rewritten

- as a MA(∞) :

$$X_t = \frac{\Theta(B)}{\Phi(B)}\omega_t = \psi(B)\omega_t = \sum_{k\geq 0} \psi_k \omega_{t-k}$$

- or as an AR((∞))

$$\omega_t = \frac{\Phi(B)}{\Theta(B)}X_t = \pi(B)X_t = \sum_{k\geq 0} \pi_k X_{t-k}$$

Notice that both π_0 and ψ_0 equal 1 and (ψ_k) and (π_k) are entirely determined by (ϕ_k) and (θ_k).

Autocovariance function of an ARMA

Autocovariance of an ARMA

The autocovariance function of an ARMA(p, q) follows from its MA(∞) representation and equals

$$\gamma_X(h) = \sigma^2 \sum_{k \geq 0} \psi_k \psi_{k+h} \ \forall h \geq 0.$$

Exercise

- Compute the ACF of a causal ARMA$(1, 1)$.
- Show that the ACF of this ARMA verifies a linear difference equation of order 1. Solve this equation.
- Compute ϕ and θ from the ACF.

Autocovariance of an ARMA

The autocovariance function of an ARMA(p, q) follows from its MA(∞) representation and equals

$$\gamma_X(h) = \sigma^2 \sum_{k \geq 0} \psi_k \psi_{k+h} \; \forall h \geq 0.$$

Exercise

- Compute the ACF of a causal ARMA$(1, 1)$.
- Show that the ACF of this ARMA verifies a linear difference equation of order 1. Solve this equation.
- Compute ϕ and θ from the ACF.

Solve this ex...

Definition of the spectral measure

Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary process (centered in order to simplify notations). Its autocovariance function satisfies $\forall(s,t)$, $\gamma_X(s-t)=\gamma_X(t-s)$ and for all $k\in\mathbb{N}^*$, $(t_1,\cdots,t_k)\in\mathbb{Z}^k$ and $c_1,\cdots,c_k\in\mathbb{C}^k$,

$$\mathbb{E}\left|\sum_{i=1}^{k} c_i X_{t_i}\right|^2 = \sum_{1\le i;j\le k} c_i\overline{c_j}\gamma_X(t_i-t_j)\ge 0.$$

Theorem

The autocovariance of $(X_t)_{t\in\mathbb{Z}}$ satisfies that for all $k\in\mathbb{Z}$,

$$\gamma_X(k)=\int_{-\pi}^{\pi} e^{ikx}d\mu(x)=\int \cos(kx)d\mu(x)$$

where μ is a non negative measure, symetric and bounded on $[-\pi,\pi]$. The measure μ is unique, with total measure $\gamma_X(0)=Var(X_0)$.
This measure is called the spectral measure of $(X_t)_{t\in\mathbb{Z}}$. Inversely, if μ is a non negative measure, symetric and bounded on $[-\pi,\pi]$ then

$$\gamma(k)=\int_{-\pi}^{\pi} e^{ikx}d\mu(x)=\int \cos(kx)d\mu(x),$$

is the autocovariance function of a weakly stationary process $(X_t)_{t\in\mathbb{Z}}$.

Note that, two weakly stationary processes having the same autocovariance function, have the same spectral measure.

Definition of the spectral density

Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary process (centered in order to simplify notations) with autocovariance function $\gamma_X(k)$ and spectral measure μ.
If μ admits a density f with respect to the Lebesgue measure on $[-\pi\,,\pi]$ then f is called the spectral density of the process $(X_t)_{t\in\mathbb{Z}}$. It satisfies

$$\gamma_X(k) = \int_{-\pi}^{\pi} e^{ikx} d\mu(x),$$

and $\gamma_X(k)$ is the k-th Fourier coefficients of f.
Two sufficient conditions for the existence of f:

C1) If $\sum_{k=0}^{\infty} |\gamma_X(k)| < \infty$, then the spectral density exists et equals

$$f(x) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \gamma(k) e^{-ixk}.$$

Moreover this density is continuous and bounded.

C2) f exists and belongs to $\mathbb{L}^2([-\pi\,,\pi])$ if and only if $\sum_{k=0}^{\infty} \gamma_X(k)^2 < \infty$. In that case, $\frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma(k) e^{-ixk}$ converges in $\mathbb{L}^2([-\pi\,,\pi])$ to f.

Linear Process

Let $(Y_t)_{t\in\mathbb{Z}}$ be a weakly stationary process. A linear process $(X_t)_{t\in\mathbb{Z}}$ in $\mathbb{L}^2$ is a sequence of random variables such that

$$X_t = \sum_{i\in\mathbb{Z}} a_i Y_{t-i},$$

as soon as the sum is well defined in $\mathbb{L}^2$.

Theorem

1. *If $(Y_t)_{t\in\mathbb{Z}}$ is centered and has a bounded spectral density , and if $\sum_{j\in\mathbb{Z}} a_j^2 < \infty$,, then $(X_t)_{t\in\mathbb{Z}}$ is well defined, is centered and weakly stationary. Moreover $(X_t)_{t\in\mathbb{Z}}$ admits a spectral density which statisfies*

$$f = g.f_Y, \text{ where } g(x) = \left|\sum_{k\in\mathbb{Z}} a_k e^{ikx}\right|^2 \in \mathbb{L}^1([-\pi,\pi]).$$

2. *If $\sum_{j\in\mathbb{Z}} |a_j| < \infty$,, then the process $(X_t)_{t\in\mathbb{Z}}$ has the spectral measure μ_x which that the density g with respect to μ_Y. We will then write $d\mu_X = gd\mu_Y$.*

Regular representation of ARMA process (1)

Theorem

Let $\Phi_1, \cdots, \Phi_p \in \mathbb{R}^p$ *and* $\theta_1, \cdots, \theta_q \in \mathbb{R}^q$ *and consider* $(X_t)_{t\in\mathbb{Z}}$ *satisfying*

$$\begin{aligned} \Phi(B)(X_t) &= \Theta(B)(\varepsilon_t), \qquad (9)\\ \text{with } \Phi(z) &= 1 - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p \\ \text{and } \Theta(z) &= 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q. \end{aligned}$$

1. *If* Φ *has no roots in the unite circle, then there exist a unique stationary solution to (9).* $(X_t)_{t\in\mathbb{Z}}$ *is a causal process with* $\varepsilon_t \perp \overline{span}(X_i\,, i \leq t-1)$.
2. *Morevover, if* Θ *has no roots in the unique circle ,* $\varepsilon_t \in \overline{span}(X_i\,, i \leq t)$ *and* $(X_t)_{t\in\mathbb{Z}}$ *is a causal and invertible process.*

In those cases, $(\varepsilon_t)_{t\in\mathbb{Z}}$ *are the innovations of the process* $(X_t)_{t\in\mathbb{Z}}$.

Regular representation of ARMA process (2)

Theorem

Let $\Phi_1, \cdots, \Phi_p \in \mathbb{R}^p$ and $\theta_1, \cdots, \theta_q \in \mathbb{R}^q$ and consider $(X_t)_{t\in\mathbb{Z}}$ satisfying

$$\begin{aligned} \Phi(B)(X_t) &= \Theta(B)(\varepsilon_t), \\ \textit{with } \Phi(z) &= 1 - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p \\ \textit{and } \Theta(z) &= 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q. \end{aligned}$$

If $\Phi(z)\Theta(z) \neq 0$ for $|z| = 1$, then there exist two polynoms $\bar{\Phi}$ and $\bar{\Theta}$ having no roots in the unite circle, of degree p and q such that

$$\bar{\Phi}(B)(X_t) = \bar{\Theta}(B)(\varepsilon_t^*), \tag{10}$$

where ε_n^* is a weak white noise defined as the innovations of $(X_t)_{t\in\mathbb{Z}}$. Moreover we have

$$\bar{\Phi}(z) = \Phi(z) \prod_{r<j\leq p} \frac{1 - a_j z}{1 - z/a_j}, \textit{ and } \bar{\Theta}(z) = \Theta(z) \prod_{s<j\leq q} \frac{1 - b_j z}{1 - z/b_j},$$

where $a_{r+1}, \cdots, a_p$ and $b_{s+1}, \cdots, b_q$ are the roots of Φ and Θ belonging to the unite circle and $a_1, \cdots, a_r$ and $b_1, \cdots, b_s$ are the roots of Φ and Θ being outside of the the unite circle. If $r = p$ and $s = q$, then $(X_t)_{t\in\mathbb{Z}}$ is causal and invertible and $\varepsilon_n^* = \varepsilon_n$

Chapter 4 : Linear prediction and partial autocorrelation function

Recall linear predictor

Linear predictor and ACF

Let X be a stationary time series with ACF ρ. The linear predictor $\hat{X}_{n+h}^{\{n\}}$ of X_{n+h} given X_n is defined as

$$\hat{X}_{n+h}^{\{n\}} = \underset{a,b}{\operatorname{argmin}} \mathbb{E}\Big(\big(X_{n+h} - (aX_n + b)\big)^2 \Big) = \rho(h)(X_n - \mu) + \mu$$

Linear prediction

Given $X_1, \cdots, X_n$, the best linear predictor of X_{n+m} is

$$X_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i X_i,$$

which satisfies the prediction equations

$$\mathbb{E}(X_{n+m} - X_{n+m}^n) = 0$$

$$\text{and } \mathbb{E}\left[(X_{n+m} - X_{n+m}^n)X_i\right] = 0, \text{ for } i = 1, \cdots, n.$$

- Orthogonal projection on the linear span generated by the past $1, X_1, \cdots, X_n$.
- The prediction errors $(X_{n+m} - X_{n+m}^n)$ are uncorrelated with the prediction variables $(1, \cdots, X_n)$.

Introduction

Consider here that $(X_t)_{t\in\mathbb{Z}}$ is and ARMA(p,q) process, which is causal, invertible and stationary and that the coefficients $\Phi_1\ ,\cdots\ ,\Phi_p \in \mathbb{R}^p$ and $\theta_1\ ,\cdots\ ,\theta_q \in \mathbb{R}^q$ are known.

We aim at building predictions and prediction intervals.

Recall

- The linear space $\mathbb{L}^2$ of r.v. with finite variance with the inner-product $\langle X, Y\rangle = \mathbb{E}(XY)$ is an Hilbert space.
- Now considering a time series X with $X_t \in \mathbb{L}^2$ for all t
  - the subspace $\mathcal{H}_n = \text{span}(X_1, \ldots, X_n)$ is a closed subspace of $\mathbb{L}^2$ hence
  - for all $Y \in \mathbb{L}^2$ there exists an unique projection onto $\mathcal{H}_n$ which is denoted by $\Pi_{\mathcal{H}_n}(Y)$.
  - for all $\forall w \in \mathcal{H}_n$

$$\|\Pi_{\mathcal{H}_n}(Y) - Y\| \leq \|w - Y\|$$
$$\langle \Pi_{\mathcal{H}_n}(Y) - Y, w\rangle = 0.$$

Wold decomposition (1)

Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary and centered process.

Notation :

- $\mathcal{M}_n = \overline{\text{span}}(X_i\,, -\infty \leq i \leq n)$ the Hilbert subspace defined as the closed subspace in $\mathbb{L}^2$ consisting in finite linear combination of $(X_i)_{i\leq n}$.
- $\mathcal{M}_\infty = \overline{\text{span}}(X_i\,, i \in \mathbb{Z})$
- $\mathcal{M}_{-\infty} = \cap_{n=-\infty}^{n=+\infty} \mathcal{M}_n$.
- $\Pi_{\mathcal{M}_n}$ the orthogonal projection on $\mathcal{M}_n$.

The best linear prediction of X_{n+1} given $(X_i\,, i \leq n)$ is $\Pi_{\mathcal{M}_n}(X_{n+1})$. And the prediction error is

$$\sigma^2 = \| X_{n+1} - \Pi_{\mathcal{M}_n}(X_{n+1}) \|^2 = \mathbb{E}[X_{n+1} - \Pi_{\mathcal{M}_n}(X_{n+1})]^2.$$

We can easily check that this error does not depend on n, thanks to the weak stationarity of the process $(X_t)_{t\in\mathbb{Z}}$.
If $\sigma^2 > 0$ the process is said <u>regular</u>, if $\sigma^2 = 0$ the process is said <u>deterministic</u>.
For $(X_t)_{t\in\mathbb{Z}}$ regular we set

$$\varepsilon_t = X_t - \Pi_{\mathcal{M}_{t-1}}(X_t).$$

Wold decomposition (2)

Check that

- $\mathbb{E}(\Pi_{\mathcal{M}_{t-1}}(X_t)) = 0$.
- $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$
- $\varepsilon_t \in \mathcal{M}_t$ and $\varepsilon_t \perp \mathcal{M}_{t-1}$, *i.e.* $\forall U \in \mathcal{M}_{t-1}, \mathbb{E}(\varepsilon_t U) = 0$.
- $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$ if $i \neq j$.

Definition

The process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a weak white noise which is called the innovation of $(X_t)_{t\in\mathbb{Z}}$.

Wold decomposition (3)

Theorem

Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary and centered process. Then we can write

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} + Y_t, t \in \mathbb{Z} \tag{11}$$

1. *with $a_0 = 1$, and $\sum_{j\geq 0} a_j^2 < \infty$*
2. *$(\varepsilon_t)_{t\in\mathbb{Z}}$ is a weak white noise such that $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$, $\varepsilon_t \in \mathcal{M}_t$ and $\varepsilon_t \perp \mathcal{M}_{t-1}$.*
3. *$Y_t \in \mathcal{M}_{-\infty}, \forall t \in \mathbb{Z}$.*
4. *Y_t is deterministic.*

Moreover, the sequences (a_j), (ε_j) and (Y_j) are uniquely determined by (11) and the conditions (1)-(4).

The process $Z_t = X_t - Y_t$, has the following wold decomposition $Z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$. Hence $\varepsilon_t = X_t - \mathbb{E}(\Pi_{\mathcal{M}_{t-1}}(X_t)) = Z_t - \mathbb{E}(\Pi_{\mathcal{M}_{t-1}}(Z_t))$.
The process $(Z_t)_{t\in\mathbb{Z}}$ is purely non deterministic.

Best linear predictor (1)

Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary and centered process. We already said that the best linear predictor of X_{n+1} given $(X_i\ , i \leq n)$ is the orthogonal projection denoted by $\Pi_{\mathcal{M}_n}(X_{n+1})$. It is given by the Wold decomposition $\Pi_{\mathcal{M}_n}(X_{n+1}) = \sum_{j=1}^{n} a_j \varepsilon_{n+1-j} + Y_{n+1}$ with $\varepsilon_{n+1} = X_{n+1} - \Pi_{\mathcal{M}_n}(X_{n+1})$. The prediction error is then $\sigma^2 = \mathbb{E}(\varepsilon_0^2)$.

In practice, the important thing is to predict X_{n+1} given $X_1\ , \cdots\ , X_n$.

Notations :

- $\mathcal{H}_n = \overline{\text{span}}(X_i, 1 \leq i \leq n)$ the hilbert subspace consisting in finite linear combination of $(X_i)_{1\leq i\leq n}$.
- $\Pi_{\mathcal{H}_n}$ the orthogonal projection on $\mathcal{H}_n$.
- $\Pi_{\mathcal{H}_n} = \sum_{i=1}^{n} \Phi_{n,i} X_{n+1-i}$

Best linear predictor (2)

It remains to find $\Pi_{\mathcal{H}_n}$ that is to find $\Phi_{n,i}$.
According to Hilbert properties, for all $1 \leq i \leq n$

$$\mathbb{E}[(X_{n+1} - \Pi_{\mathcal{H}_n})X_{n+1-i}] = 0.$$

That is all $1 \leq i \leq n$

$$\mathbb{E}[X_{n+1}X_{n+1-i}] = \mathbb{E}[\Pi_{\mathcal{H}_n}(X_{n+1-i})].$$

$$\mathbb{E}[X_{n+1}X_{n+1-i}] = \sum_{j=1}^{n} \Phi_{n,j}\mathbb{E}[X_{n+1-j}X_{n+1-i}].$$

This is a linear system that can be written as

$$\Gamma_n\Phi_n = \gamma_n, \text{ and } v_n = \mathbb{E}[(X_{n+1} - \mathbb{E}[\Pi_{\mathcal{H}_n}(X_{n+1}))^2].$$

with

$$\Phi_n = \begin{pmatrix} \Phi_{n,1} \\ . \\ . \\ . \\ \Phi_{n,n} \end{pmatrix}, \gamma_n = \begin{pmatrix} \gamma(1) \\ . \\ . \\ . \\ \gamma(n) \end{pmatrix}, \text{ and } \Gamma_n = \begin{pmatrix} \gamma(0) & \cdots & \gamma(n-1) \\ \gamma(1) & \cdots & \gamma(n-2) \\ \gamma(i-1) & \cdots & \gamma(n-i) \\ \gamma(n-1) & \cdots & \gamma(0) \end{pmatrix}$$

Best Linear predictor (3)

Linear system

The best linear predictor of X_{n+1} given $X_1, \cdots, X_n$ is

$$\Pi_{\mathcal{H}_n}(X_{n+1}) = \phi_{n,1}X_n + \phi_{n,2}X_{n-1} + \cdots + \phi_{n,n}X_1$$

satisfying

$$\Gamma_n \phi_n = \gamma_n$$

with

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots \gamma(n-2) \\ \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdots & \cdot \\ \gamma(n-1) & \gamma(n-2) & \cdots \gamma(0) \end{pmatrix}$$

and

$$\phi_n = (\phi_{n,1}, \phi_{n,2}, \cdots, \phi_{n,n})^t, \quad \gamma_n = (\gamma(1), \gamma(2), \cdots, \gamma(n))^t.$$

Mean squared error for one-step-ahead linear prediction

$$v_{n+1} = \mathbb{E}(X_{n+1} - \Pi_{\mathcal{H}_n}(X_{n+1}))^2 = \gamma(0) - \gamma_n^t \Gamma_n^{-1} \gamma_n.$$

Variance is reduced !

Best linear predictor (4) :

Numerical solution

Two methods for solving the system

$$\Gamma_n \Phi_n = \gamma_n, \text{ with } v_{n+1} = \mathbb{E}[(X_{n+1} - \Pi_{\mathcal{H}_n}(X_{n+1}))^2].$$

1. Durbin Levinson algorithm
2. Innovations algorithm

Best linear predictor (4) : Durbin Levinson algorithm

We start from $\Pi_{\mathcal{H}_0}(X_1) = 0$,

$$\Phi_{1,1} = \frac{\gamma(1)}{\gamma(0)}, \cdots, \Phi_{n,n} = \frac{\gamma(n) - \sum_{j=1}^{n-1} \Phi_{n-1,j}\gamma(n-j)}{v_{n-1}},$$

and

$$\begin{pmatrix} \Phi_{n,1} \\ \cdot \\ \cdot \\ \cdot \\ \Phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \Phi_{n-1,1} \\ \cdot \\ \cdot \\ \cdot \\ \Phi_{n-1,n-1} \end{pmatrix} - \Phi_{n,n} \begin{pmatrix} \Phi_{n-1,n-1} \\ \cdot \\ \cdot \\ \cdot \\ \Phi_{n-1,1} \end{pmatrix}$$

with $v_n = v_{n-1}(1 - \Phi_{n,n}^2) = \prod_{k=1}^{n}(1 - \alpha_{k,k}^2)$, $v_0 = \gamma(0)$.
Only for stationary sequence.

Best linear predictor (5) : Innovations algorithm

Suppose that Γ_n is invertible for all n, then

$$\Pi_{\mathcal{H}_n}(X_{n+1}) = \sum_{k\geq 1} c_{n,k}(X_{n+1-k} - \Pi_{\mathcal{H}_{n-k}}(X_{n+1-k})), \text{ if } n \geq 1,$$

$$\Pi_{\mathcal{H}_n}(X_{n+1}) = 0 \text{ if } n = 0,$$

with

- $v_0 = \mathrm{Var}(X_0) = \gamma(0)$
- $c_{n,n} = \frac{\mathrm{Cov}(X_{n+1}, X_1)}{v_0}$
- $c_{n,n-l} = \frac{\mathrm{Cov}(X_{n+1}, X_{l+1}) - \sum_{k=0}^{l-1} c_{l,l-k} c_{n,n-k} v_k}{v_l} \; \forall 1, \cdots, n-1$
- $v_n = \mathrm{Var}(X_{n+1}) - \sum_{k=0}^{n-1} c_{n,n-k}^2 v_k$.

Still holds for non stationary sequences. Solved, v_0, $c_{1,1}$, v_1, $c_{2,2}$, $c_{2,1}$, v_2, $c_{3,3}$, $c_{3,2}$, $c_{3,1}$, etc...

Best linear predictor (6)

In the same way, the prediction of X_{n+h} given $(X_1, \cdots, X_n)$ is given by $\Pi_{\mathcal{H}_n}(X_{n+h})$. But

$$\Pi_{\mathcal{H}_n}(X_{n+h}) = \Pi_{\mathcal{H}_n}(\Pi_{\mathcal{H}_{n+h-1}}(X_{n+h})) = \Pi_{\mathcal{H}_n}(\sum_{j=1}^{n+h-1} c_{n+h-1,j}(X_{n+h-j} - \Pi_{\mathcal{H}_{n+h-j}}(X_{n+h}))$$

Since $X_{n+h-j} - \Pi_{\mathcal{H}_{n+h-j}}(X_{n+h})) \perp \mathcal{H}_n$ for $j < h$ we get

$$\Pi_{\mathcal{H}_n}(X_{n+h})) = \sum_{j=h}^{n+h-1} c_{n+h-1,j}(X_{n+h-j} - \Pi_{\mathcal{H}_{n+h-j-1}}(X_{n+h-j})).$$

And

$$v_{n+h} = \mathrm{Var}(X_{n+h}) - \sum_{k=h}^{n+h-1} c^2_{n,n+h-k-1} v_{n+h-k-1}.$$

Best linear predictor

Given $X_1, X_2, \ldots, X_n$, the **best linear m-step-ahead predictor** of X_{n+m} defined as

$$\Pi_{\mathcal{H}_n}(X_{n+m}) = \alpha_0 + \phi_{n1}^{(m)} X_n + \phi_{n2}^{(m)} X_{n-1} + \phi_{nn}^{(m)} X_1 = \alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j}$$

is the orthogonal projection of X_{n+m} onto $\text{span}\{1, X_1, \ldots, X_n\}$. In particular, it satisfies the **prediction equations**

$$\mathbb{E}(\Pi_{\mathcal{H}_n}(X_{n+m}) - X_{n+m}) = 0$$
$$\mathbb{E}((\Pi_{\mathcal{H}_n}(X_{n+m}) - X_{n+m})X_k) = 0 \quad \forall k = 1, \ldots, n$$

We'll now compute α_0 and the $\phi_{nj}^{(m)}$'s.

Derivation of α_0

We get

$$\Pi_{\mathcal{H}_n}(X_{n+m}) - \mu = \alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j} - \mu = \sum_{j=1}^{n} \phi_{nj}^{(m)} (X_{n+1-j} - \mu)$$

- Thus, we'll ignore α_0 and put $\mu = 0$ until we discuss estimation.
- There are two consequences
  1. the projection of X_{n+m} on onto span$\{1, X_1, \dots, X_n\}$ is in fact the projection onto span$\{X_1, \dots, X_n\}$
  2. $\mathbb{E}(X_k X_l) = \text{Cov}(X_k, X_l)$

Derivation of α_0

We get

$$\Pi_{\mathcal{H}_n}(X_{n+m}) - \mu = \alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j} - \mu = \sum_{j=1}^{n} \phi_{nj}^{(m)} (X_{n+1-j} - \mu)$$

- Thus, we'll ignore α_0 and put $\mu = 0$ until we discuss estimation.
- There are two consequences
  1. the projection of X_{n+m} on onto span$\{1, X_1, \ldots, X_n\}$ is in fact the projection onto span$\{X_1, \ldots, X_n\}$
  2. $\mathbb{E}(X_k X_l) = \mathrm{Cov}(X_k, X_l)$

$$\mathbb{E}(\Pi_{\mathcal{H}_n}(X_{n+m}) - X_{n+m}) = 0 \Longleftrightarrow \mathbb{E}(\alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j} - X_{n+m}) = 0$$

$$\Longleftrightarrow \mu\big(1 - \sum_{j=1}^{n} \phi_{nj}^{(m)}\big) = \alpha_0.$$

Derivation of the $\phi_{nj}^{(m)}$'s

As $X_{n+m}^{(n)}$ satisfies the prediction equations of slide 100, we can write for all $k = 1, \ldots, n$

$$\begin{aligned} & \mathbb{E}((\Pi_{\mathcal{H}_n}(X_{n+m}) - X_{n+m})X_k) = 0 \\ \iff & \sum_{j=1}^{n} \phi_{nj}^{(m)} \mathbb{E}\big(X_{n+1-j}X_{n+1-k}\big) = \mathbb{E}\big(X_{n+m}X_{n+1-k})\big) \\ \iff & \sum_{j=1}^{n} \alpha_j \gamma_X(k-j) = \gamma_X(m+k-1) \end{aligned}$$

This can rewritten in matrix notation.

Prediction

Prediction equations

The $\phi_{nj}^{(m)}$'s verify

$$\Gamma_n \phi_n^{(m)} = \gamma_n^{(m)}$$

where

$$\Gamma_n = \Big(\gamma_X(k-j)\Big)_{1 \le j,k \le n}$$
$$\phi_n^{(m)} = \left(\phi_{n1}^{(m)}, \ldots, \phi_{nn}^{(m)}\right)^\top,$$
$$\gamma_n^{(m)} = \left(\gamma_X(m), \ldots, \gamma_X(m+n-1)\right)^\top.$$

Prediction error

The mean square prediction error is given by

$$v_{n+m} = \mathbb{E}\Big(\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)^2\Big) = \gamma_X(0) - (\gamma_n^{(m)})^\top \Gamma_n^{-1} \gamma_n^{(m)}.$$

Prediction

Prediction equations

The $\phi_{nj}^{(m)}$'s verify

$$\Gamma_n \phi_n^{(m)} = \gamma_n^{(m)}$$

where

$$\begin{aligned}
\Gamma_n &= \Big(\gamma_X(k-j)\Big)_{1 \le j,k \le n} \\
\phi_n^{(m)} &= (\phi_{n1}^{(m)}, \ldots, \phi_{nn}^{(m)})^\top, \\
\gamma_n^{(m)} &= (\gamma_X(m), \ldots, \gamma_X(m+n-1))^\top.
\end{aligned}$$

Prediction error

The mean square prediction error is given by

$$v_{n+m} = \mathbb{E}\Big(\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)^2\Big) = \gamma_X(0) - (\gamma_n^{(m)})^\top \Gamma_n^{-1} \gamma_n^{(m)}.$$

$$\begin{aligned}
\mathbb{E}\Big(\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)^2\Big) &= \mathbb{E}\Big(\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)\Big) \\
&= \mathbb{E}\Big(\big(X_{n+m}\big)\big(X_{n+m} - \Pi_{\mathcal{H}_n}(X_{n+m})\big)\Big) = \gamma_X(0) - \mathbb{E}\Big(X_{n+m}\Pi_{\mathcal{H}_n}(X \\
&= \gamma_X(0) - \mathbb{E}(X_{n+m}\mathbf{X}\phi_n^{(m)}) \\
&= \gamma_X(0) - \mathbb{E}(X_{n+m}\mathbf{X})\phi_n^{(m)} = \gamma_X(0) - (\gamma_n^{(m)})^\top \phi_n^{(m)}
\end{aligned}$$

Forecasting an AR(2)

Exercise

Consider the causal AR(2) model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \omega_t$.

1. Determine the one-step-ahead $X_3^{(2)}$ prediction of X_3 based on X_1 , X_2 from the prediction equations.
2. From causality, determine $X_3^{(2)}$.
3. How $\phi_{21}^{(1)}$, $\phi_{22}^{(1)}$ and ϕ_1 , ϕ_2 are related ?

Forecasting an AR(2)

Exercise

Consider the causal AR(2) model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \omega_t$.

1. Determine the one-step-ahead $X_3^{(2)}$ prediction of X_3 based on X_1, X_2 from the prediction equations.
2. From causality, determine $X_3^{(2)}$.
3. How $\phi_{21}^{(1)}, \phi_{22}^{(1)}$ and ϕ_1, ϕ_2 are related ?

solve this ex.... ! ! ! !

PACF

Partial autocorrelation function

The **partial autocorrelation function** (PACF) of a stationary time series X is defined as

$$\phi_{11} = \text{cor}\big(X_1, X_0\big) = \rho(1)$$
$$\phi_{hh} = \text{cor}\big(X_h - \Pi_{\mathcal{H}_{h-1}}(X_h), X_0 - \Pi_{\mathcal{H}_{h-1}}(X_0)\big) \text{ for } h \geq 2,$$

where $\Pi_{\mathcal{H}_{h-1}}(X_0)$ is the orthogonal projection of X_0 onto $\text{span}\{X_1, \dots, X_{h-1}\}$.

Notice that

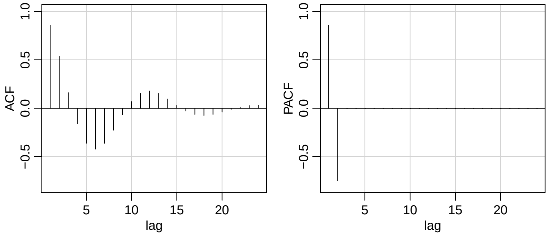
- $X_h - \Pi_{\mathcal{H}_{h-1}}(X_h)$ and $X_0 - \Pi_{\mathcal{H}_{h-1}}(X_0)$ are, by construction, uncorrelated with $\{X_1, \dots, X_{h-1}\}$, so ϕ_{hh} is the correlation between X_h and X_0 with the linear dependence of $X_1, \dots, X_{h-1}$ on each removed.
- The coefficient ϕ_{hh} is also the last coefficient (i.e. $\phi_{hh}^{(1)}$) in the best linear one-step-ahead prediction of X_{h+1} given $X_1, \dots, X_h$.

Forecasting and PACF of causal AR(p) models

PACF of an AR(p) model

Consider the causal AR(p) model $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \omega_t$

1. Consider $p = 2$ and verify that $\Pi_{\mathcal{H}_n}(X_{n+1}) = \phi_1 X_n + \phi_2 X_{n-1}$. Deduce the value of the PACF for $h > 2$
2. In the general case, deduce the value of the PACF for $h > p$.



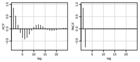
2018-02-22

Forecasting and PACF of causal AR(p) models

PACF of an AR(p) model

Consider the causal AR(p) model $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \omega_t$

1. Consider $p = 2$ and verify that $\Pi_{\mathcal{H}_n}(X_{n+1}) = \phi_1 X_n + \phi_2 X_{n-1}$. Deduce the value of the PACF for $h > 2$
2. In the general case, deduce the value of the PACF for $h > p$.



see page 100 of [BD13]

PACF of invertible MA models

Exercise : PACF of a MA(1) model

Consider the invertible MA(1) model $X_t = \omega_t + \theta \omega_{t-1}$

1. Compute $\hat{X}_3^{(2)}$ and $\hat{X}_1^{(2)}$, the orthogonal projections of X_3 and X_1 onto $\text{span}\{X_2\}$.
2. Deduce the first two values of the PACF.

More calculations (see Problem 3.23 in [BD13]) give

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}.$$

In general, the PACF of a MA(q) model does not vanish for larger lag, it is however bounded by a geometrically decreasing function.

PACF of invertible MA models

Exercise : PACF of a MA(1) model

Consider the invertible MA(1) model $X_t = \omega_t + \theta \omega_{t-1}$

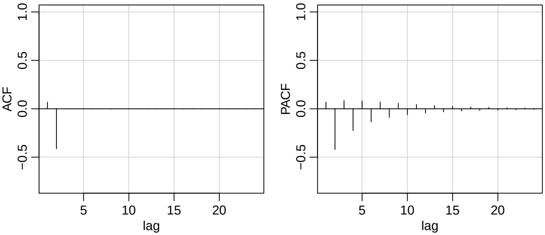
1. Compute $\hat{X}_3^{(2)}$ and $\hat{X}_1^{(2)}$, the orthogonal projections of X_3 and X_1 onto span$\{X_2\}$.
2. Deduce the first two values of the PACF.

More calculations (see Problem 3.23 in [BD13]) give

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}.$$

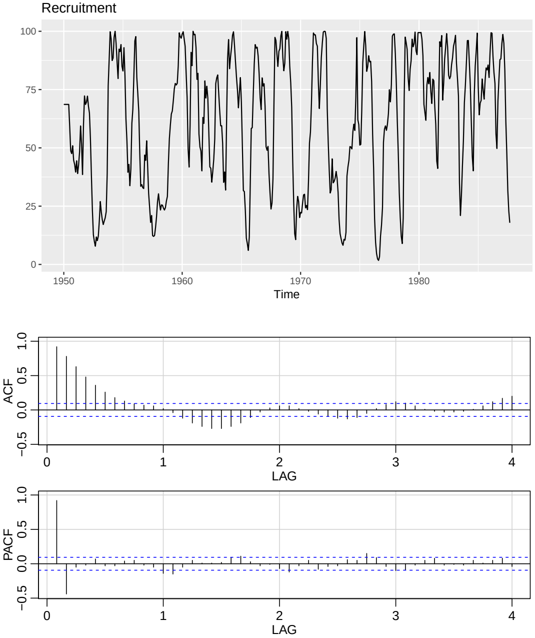
In general, the PACF of a MA(q) model does not vanish for larger lag, it is however bounded by a geometrically decreasing function.

see page 100 of [BD13]



ACF
1.0
0.5
0.0
−0.5
5
10
15
20
lag
PACF
1.0
0.5
0.0
−0.5
5
10
15
20
lag

An AR(2) model for the recruitement series



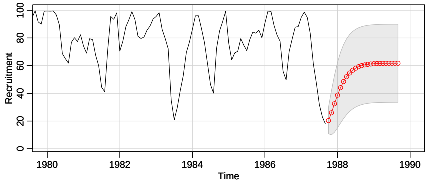


FIG.: Twenty-four month forecats for the Recruitment series shown on slide 10

ACF and PACF

So far, we show that

Model	**ACF**	**PACF**
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA	decays	decays

- We can use these results to **build a model**.
- And we know how to **forecast in an AR(p) model**.
- It remains to give algorithms that will allow to forecast in MA and ARMA models.

Innovations

So far, we have written $\Pi_{\mathcal{H}_n}(X_{n+1})$ as $\sum_{j=1}^{n} \phi_{n,j} X_{n+1-j}$ i.e. as the projection of X_{n+1} onto span$\{X_1, \ldots, X_n\}$ but we clearly have

$$\text{span}\{X_1, X_2 - \Pi_{\mathcal{H}_1}(X_2), X_3 - \Pi_{\mathcal{H}_2}(X_3), \ldots, X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)\}.$$

Approximate Innovations

The values $X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)$ are closed to innovations and verify $X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)$ is orthogonal to $\mathtt{span}\{X_1, \ldots, X_{n-1}\}$.

As a consequence, we can rewrite

$$\Pi_{\mathcal{H}_{n-1}}(X_n) = \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \Pi_{\mathcal{H}_{n-j}}(X_{n+1-j})$$

The one-step-ahead predictors $\Pi_{\mathcal{H}_{n-1}}(X_n)$ and their mean-squared errors $v_n = \mathbb{E}[X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)]^2$ can be calculated iteratively via the innovations algorithm.

The innovations algorithm

The one-step-ahead predictors can be iteratively be computed via

$$X_1^0 = 0, v_1 = \gamma_X(0) \text{ and } t = 1, 2, \dots$$

$$\Pi_{\mathcal{H}_t}(X_{t+1}) = \sum_{j=1}^{t} \theta_{tj}(X_{t+1-j} - \Pi_{\mathcal{H}_{t+j-1}}(X_{t+j}))$$

$$v_{t+1} = \gamma_X(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 v_{j+1} \text{ where}$$

$$\theta_{t,t-h} = \Big(\gamma_X(t-h) - \sum_{k=0}^{h-1} \theta_{h,h-k}\theta_{t,t-k} v_{k+1}\Big)(v_{h+1})^{-1} \quad h = 0, 1, \dots, t-1$$

This can be solve by calculting v_1, θ_{11}, v_2, θ_{22}, θ_{21}, etc.

The innovations algorithm

The innovations algorithm

The one-step-ahead predictors can be iteratively be computed via

$$X_1^0 = 0, v_1 = \gamma_X(0) \text{ and } t = 1, 2, \dots$$

$$\Pi_{\mathcal{H}_t}(X_{t+1}) = \sum_{j=1}^{t} \theta_{tj}(X_{t+1-j} - \Pi_{\mathcal{H}_{t+j-1}}(X_{t+j}))$$

$$v_{t+1} = \gamma_X(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 v_{j+1} \text{ where}$$

$$\theta_{t,t-h} = \Big(\gamma_X(t-h) - \sum_{k=0}^{h-1} \theta_{h,h-k}\theta_{t,t-k}v_{k+1}\Big)(v_{h+1})^{-1} \quad h = 0, 1, \dots, t-1$$

This can be solve by calculting v_1, θ_{11}, v_2, θ_{22}, θ_{21}, etc.

easy proof of p172-173 of [BD13]

Prediction for an MA(1)

Exercise

Consider the MA(1) model $X_t = \omega_t + \theta\,\omega_{t-1}$ with $\omega \sim WN(0\,,\sigma^2)$. We know that $\gamma_X(0) = \sigma^2(1+\theta^2)$, $\gamma_X(1) = \theta\sigma^2$ and $\gamma_X(h) = 0$ for $h \geq 2$.
Show that

$$\Pi_{\mathcal{H}_n}(X_{n+1}) = \theta\frac{X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)}{r_n}$$

with

$$r_n = v_n/\sigma^2.$$

Prediction for an MA(1)

Exercise

Consider the MA(1) model $X_t = \omega_t + \theta \omega_{t-1}$ with $\omega \sim WN(0, \sigma^2)$. We know that $\gamma_X(0) = \sigma^2(1+\theta^2)$, $\gamma_X(1) = \theta\sigma^2$ and $\gamma_X(h) = 0$ for $h \geq 2$.

Show that

$$\Pi_{\mathcal{H}_n}(X_{n+1}) = \theta \frac{X_n - \Pi_{\mathcal{H}_{n-1}}(X_n)}{r_n}$$

with

$$r_n = v_n/\sigma^2.$$

begin with $\theta_{nn} = 0$ (the sum is empty), that implies that $\theta_{n,n-1} = 0$ and so on, till θ_{n1}

The innovations algorithm for the ARMA(p, q) model

Consider an ARMA(p, q) model

$$\Phi(B)X_t = \Theta(B)\omega_t \text{ with } \omega \sim WN(0, \sigma^2).$$

Let $m = \max(p, q)$, to simplify calculations, the innovation algorithm is not applied directly to X but to

$$\begin{cases} W_t = \sigma^{-1}X_t & t = 1, \ldots, m \\ W_t = \sigma^{-1}\Phi(B)X_t & t > m. \end{cases}$$

see page 175 of [BD13]

Infinite past

We will now show that it is easier for a **causal, invertible** ARMA process

$$\Phi(B)X_t = \Theta(B)\omega_t$$

to approximate $\Pi_{\mathcal{H}_n}(X_{n+h})$ by a **truncation** of the projection of X_{n+h} onto the infinite past

$$\bar{\mathcal{H}}_n = \bar{\text{span}}\{X_n, X_{n-1}, \ldots\} = \bar{\text{span}}\{X_k, k \leq n\}.$$

The projection onto $\bar{\mathcal{H}}_n = \bar{\text{span}}(X_k\ , k \leq n)$ can be defined as

$$\lim_{k\to\infty} P_{\text{span}(X_{n-k},\ldots,X_n)}$$

We will define

$$\tilde{X}_{n+h} \text{ and } \tilde{\omega}_{n+h}$$

as the projections of X_{n+h} and ω_{n+h} onto $\bar{\mathcal{H}}_n$.

Causal and invertible

Recall (see slide 79) that since X is causal and invertible, we may write

- $X_{n+h} = \sum_{k \geq 0} \psi_k \, \omega_{n+h-k}$ (MA(∞) representation)
- $\omega_{n+h} = \sum_{k \geq 0} \pi_k X_{n+h-k}$ (AR((∞)) representation).

Now, applying the projection operator onto $\mathcal{M}_n$ on both sides of both equations, we get

$$\tilde{X}_{n+h} = \sum_{k \geq 0} \psi_k \tilde{\omega}_{n+h-k} \tag{12}$$

$$\tilde{\omega}_{n+h} = \sum_{k \geq 0} \pi_k \tilde{X}_{n+h-k}. \tag{13}$$

Causal and invertible

Recall (see slide 79) that since X is causal and invertible, we may write

- $X_{n+h} = \sum_{k\geq 0} \psi_k \omega_{n+h-k}$ (MA(∞) representation)
- $\omega_{n+h} = \sum_{k\geq 0} \pi_k X_{n+h-k}$ (AR((∞)) representation).

Now, applying the projection operator onto $\mathcal{M}_n$ on both sides of both equations, we get

$$\tilde{X}_{n+h} = \sum_{k\geq 0} \psi_k \tilde{\omega}_{n+h-k} \tag{12}$$

$$\tilde{\omega}_{n+h} = \sum_{k\geq 0} \pi_k \tilde{X}_{n+h-k}. \tag{13}$$

Let's work with 13 :

$$\tilde{\omega}_{n+h} = 0 = \tilde{X}_{n+h} + \sum_{k\geq 1} \pi_k \tilde{X}_{n+h-k}$$

(put $\pi_0 = 1$).
Now, 12. As ω is a WN, $\tilde{\omega}_{n+h-k} = 0$ as long as $n+h-k > n \Leftrightarrow k < h$. If $k \geq h$, $\tilde{\omega}_{n+h-k} = \omega_{n+h-k}$ by linearity+causality $\bar{\text{span}}\{X_n, X_{n-1}, \ldots,\} = \bar{\text{span}}\{\omega_n, \omega_{n-1}, \ldots,\}$ so

$$\tilde{X}_{n+h} = \sum_{k\geq 0} \psi_k \tilde{\omega}_{n+h-k} = \sum_{k\geq h} \psi_k \tilde{\omega}_{n+h-k}$$

Subtracting this from the MA(∞) representation, we get

$$X_{n+h} - \tilde{X}_{n+h} = \sum_{k<h} \psi_k \tilde{\omega}_{n+h-k}.$$

Iteration

We get

$$\tilde{X}_{n+h} = -\sum_{k\geq 1} \pi_k \tilde{X}_{n+h-k} \text{ and}$$

$$\mathbb{E}\big((X_{n+h} - \tilde{X}_{n+h})^2\big) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 .$$

As $\tilde{X}_t = X_t$ for all $t \leq n$, we can define recursively

$$\tilde{X}_{n+1} = -\sum_{k\geq 1} \pi_k X_{n+h-k}$$

$$\tilde{X}_{n+2} = -\pi_1 \tilde{X}_{n+1} - \sum_{k\geq 2} \pi_k X_{n+h-k}$$

$$\ldots .$$

Truncation

In practice, we do not observe the past from $-\infty$ but only $X_1, \ldots, X_n$, but we can use a truncated version

$$\tilde{X}^T_{n+1} = -\sum_{k=1}^{n} \pi_k X_{n+h-k}$$

$$\tilde{X}^T_{n+2} = -\pi_1 \tilde{X}^T_{n+1} - \sum_{k=2}^{n+1} \pi_k X_{n+h-k}$$

$$\ldots$$

and $\mathbb{E}\big((X_{n+h} - \tilde{X}_{n+h})^2\big) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ is used an approximation of the predictor error.

Chapter 5 : Estimation and model selection

Introduction

We saw in the last chapter, that if we know

- the **orders (p and q)** and
- the **coefficients**

of the ARMA model under consideration, we can build predictions and prediction intervals.

The aim of this chapter is to present

- methods for **estimating the coefficients** when the orders (p and q) are known
- **model selection methods**, i.e. methods for selecting p and q

Caution :

- To avoid confusion, **true parameters now wear a star** :

$$\sigma^{2,\star}, \phi_1^\star, \ldots, \phi_p^\star, \theta_1^\star, \ldots, \theta_q^\star$$

- we have a sample $(X_1, \ldots, X_n)$ to build estimators.

Moment estimations

We assume that $\mu^\star = 0$ (without loss of generality) in Chapter 4 and consider causal and invertible ARMA processes of the form

$$\Phi(B)(X_t - \mu^\star) = \Theta(B)\omega_t$$

where $\mathbb{E}(X_t) = \mu^\star$

Estimation of the mean

For a stationary time series, the moment estimator of $\mu^\star$ is the sample mean $\bar{X}_n$.

AR(1) model

Give the moment estimators in a stationary AR(1) model.

Moment estimators for AR(p) models

Yule-Walker equations for an AR(p)

The autocovariance function and parameters of the AR(p) model verify

$$\Gamma_p \phi^\star = \gamma_p \quad \text{and} \quad \sigma^{2,\star} = \gamma_X(0) - (\phi^\star)^\top \gamma_p$$

where

$$\Gamma_p = \left(\gamma_X(k-j)\right)_{1\le j,k\le p} \quad \phi^\star \quad = (\phi_1^\star, \dots, \phi_p^\star)^\top, \text{ and } \gamma_p = \left(\gamma_X(1), \dots, \gamma_X(p)\right)^\top.$$

This leads to

$$\widehat{\phi} = \widehat{\Gamma}_p^{-1}\widehat{\gamma}_p \quad \text{and} \quad \hat{\sigma}^2 = \hat{\gamma}_X(0) - \widehat{\phi}^\top \widehat{\gamma}_p.$$

AR(2)

Verify the Yule Walker equation for a causal AR(2) model.

Moment estimators for AR(p) models

Yule-Walker equations for an AR(p)

The autocovariance function and parameters of the AR(p) model verify

$$\Gamma_p \phi^\star = \gamma_p \quad \text{and} \quad \sigma^{2,\star} = \gamma_X(0) - (\phi^\star)^\top \gamma_p$$

where

$$\Gamma_p = \left(\gamma_X(k-j)\right)_{1\le j,k\le p} \quad \phi^\star = \left(\phi_1^\star, \dots, \phi_p^\star\right)^\top, \text{ and } \gamma_p = \left(\gamma_X(1), \dots, \gamma_X(p)\right)^\top.$$

This leads to

$$\widehat{\phi} = \widehat{\Gamma}_p^{-1} \widehat{\gamma}_p \quad \text{and} \quad \hat{\sigma}^2 = \hat{\gamma}_X(0) - \widehat{\phi}^\top \widehat{\gamma}_p.$$

AR(2)

Verify the Yule Walker equation for a causal AR(2) model.

For the causal AR(2) : $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \omega_t$. So

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_t, \phi_1 X_{t+h-1} + \phi_2 X_{t+h-2} + \omega_{t+h}) = \phi_1 \gamma_X(h-1) + \phi_2 \gamma_X(h-2).$$

$$\begin{pmatrix} \gamma_X(0) & \gamma_X(-1) \\ \gamma_X(1) & \gamma_X(0) \end{pmatrix} \begin{pmatrix} \phi 1 \\ \phi 2 \end{pmatrix} = \begin{pmatrix} \gamma_X(1) \\ \gamma_X(2) \end{pmatrix}$$

Asymptotics

The only case in which the moment method is (asymptotically) efficient is the AR(p) model.

Asymptotic distribution of moment estimators

Under mild conditions on ω, and if the AR(p) is causal, the Yule-Walker estimators verify

$$\sqrt{n}(\widehat{\phi} - \phi^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^{2,\star}\Gamma_p^{-1})$$
$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^{2,\star}.$$

Likelihood of an causal AR(1) model I

We now deal with maximum likelihood estimation, we assume that

$$\omega \sim i.i.d. \mathcal{N}(0, \sigma^{2,\star}).$$

The likelihood the causal AR(1) model

$$X_t = \mu^\star + \phi^\star (X_{t-1} - \mu^\star) + \omega_t$$

is given by

$$\mathcal{L}_n(\mu, \phi, \sigma^2) = f_{\mu,\phi,\sigma^2}(X_1, \ldots, X_n)$$
$$= f_{\mu,\phi,\sigma^2}(X_1) f_{\mu,\phi,\sigma^2}(X_2|X_1) f_{\mu,\phi,\sigma^2}(X_3|X_1, X_2) \ldots f_{\mu,\phi,\sigma^2}(X_n|X_1, X_2, \ldots, X_{n-1})$$

Likelihood of an causal AR(1) model I

We now deal with maximum likelihood estimation, we assume that

$$\omega \sim i.i.d. \mathcal{N}(0, \sigma^{2,\star}).$$

The likelihood the causal AR(1) model

$$X_t = \mu^\star + \phi^\star(X_{t-1} - \mu^\star) + \omega_t$$

is given by

$$\mathcal{L}_n(\mu, \phi, \sigma^2) = f_{\mu,\phi,\sigma^2}(X_1, \ldots, X_n)$$
$$= f_{\mu,\phi,\sigma^2}(X_1) f_{\mu,\phi,\sigma^2}(X_2|X_1) f_{\mu,\phi,\sigma^2}(X_3|X_1, X_2) \ldots f_{\mu,\phi,\sigma^2}(X_n|X_1, X_2, \ldots, X_{n-1})$$

As the causal AR(1) model has the strong Markov property,

$$f_{\mu,\phi,\sigma^2}(X_k|X_1, X_2, \ldots, X_{k-1}) = f_{\mu,\phi,\sigma^2}(X_k|X_{k-1}).$$

Having assumed that $\omega_k \sim \mathcal{N}(0\,, \sigma^{2,\star})$, we deduce that

$$X_k \sim \mathcal{N}(\mu^\star + \phi^\star(X_{k-1} - \mu^\star), \sigma^{2,\star})$$

and, as a consequence

$$\mathcal{L}_n(\mu, \phi, \sigma^2) = f_{\mu,\phi,\sigma^2}(X_1) \prod_{k=2}^{n} \varphi\big(\big(X_k - \mu + \phi(X_{k-1} - \mu)\big)/\sigma\big).$$

In addition, using the MA(∞) representation, $X_1 = \mu^\star + \sum_{j\geq 0} \phi^{j,\star}\,\omega_{1-j}$, hence

$$X_1 \sim \mathcal{N}(\mu^\star, \frac{\sigma^{2,\star}}{1 - \phi^{2,\star}}).$$

Likelihood of an causal AR(1) model II

We can now write the log-likelihood

$$\ell_n(\mu, \phi, \sigma^2) = \log \mathcal{L}_n(\mu, \phi, \sigma^2)$$
$$= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}S(\mu,\ \phi)$$

with

$$S(\mu,\ \phi) = (1-\phi^2)(X_1-\mu) + \sum_{k=2}^{n}\left(X_k - \mu + \phi(X_{k-1}-\mu)\right)^2.$$

It is straightforward to see that

$$\hat{\sigma}^2 = \frac{1}{n}S(\hat{\mu},\ \hat{\phi})$$

where

$$\hat{\mu}, \hat{\phi} = \text{argmin}_{\mu,\phi} \log(S(\mu,\ \phi)/n) - \frac{1}{n}\log(1-\phi^2).$$

Likelihood of an causal AR(1) model II

We can now write the log-likelihood

$$\begin{aligned}\ell_n(\mu, \phi, \sigma^2) &= \log \mathcal{L}_n(\mu, \phi, \sigma^2)\\ &= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2} S(\mu,\ \phi)\end{aligned}$$

with

$$S(\mu,\ \phi) = (1-\phi^2)(X_1 - \mu) + \sum_{k=2}^{n} \left(X_k - \mu + \phi(X_{k-1} - \mu)\right)^2.$$

It is straightforward to see that

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\mu},\ \hat{\phi})$$

where

$$\hat{\mu}, \hat{\phi} = \text{argmin}_{\mu,\phi} \log(S(\mu,\ \phi)/n) - \frac{1}{n}\log(1-\phi^2).$$

replace σ^2 by $\sigma^{2,\star}$ in the likelihood to obtain $\hat{\mu}$ and $\hat{\phi}$

Likelihood for causal, invertible ARMA model I

Consider the causal and invertible ARMA(p, q)

$$\Phi(B)X_t = \Theta(B)\omega_t,$$

when $\omega \sim i.i.d.\mathcal{N}(0, \sigma^{2,\star})$, one can show that

$$X_t|X_1, \ldots, X_{t-1} \sim \mathcal{N}(X_t^{(t-1)}, P_t^{t-1}) \quad \text{with}$$

$$P_t^{t-1} = \sigma^{2,\star} \sum_{j\geq 0} \psi_j^{2,\star} \prod_{k=1}^{t-1} (1 - \phi_{kk}^{2,\star}) := \sigma^{2,\star} r_t$$

see the details on pages 126 and following of [SS10] and the Durbin-Levinson algorithm (see page 112).

Likelihood for causal, invertible ARMA model II

Log-likelihood of an Gaussian ARMA(p, q) process

Denoting by β the vector $(\mu, \phi_1, \phi_p, \theta_1, \dots, \theta_q)$, we have

$$\ell_n(\beta, \sigma^2) = \log \mathcal{L}_n(\beta, \sigma^2)$$
$$= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{k=1}^{n}\log(r_k(\beta)) - \frac{1}{2\sigma^2}S(\beta)$$

with

$$S(\beta) = \sum_{k=1}^{n}\left(\frac{X_k - X_k^{k-1}}{r_k(\beta)}\right)^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}S(\widehat{\beta})$$

where

$$\widehat{\beta} = \text{argmin}_\beta \log(S(\beta)/n) - \frac{1}{n}\sum_{k=1}^{n}\log(r_k(\beta)).$$

The minimization problem is usually solve via Newton-Raphson algorithm.

Asymptotics

Asymptotic distribution of maximum likelihood estimators

Under appropriate conditions, and if the ARMA(p, q) is causal and invertible, the maximum likelihood estimators verify

$$\sqrt{n}(\widehat{\beta} - \beta^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^{2,\star}\Gamma_{p,q}^{-1,\star})$$

$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^{2,\star}$$

where the matrix $\Gamma_{p,q}^\star$ depends on $(\phi_1^\star, \dots, \phi_p^\star, \theta_1^\star, \dots, \theta_q^\star)$.

Other options involve in particular conditional sum of squares and the Gauss-Newton algorithm (details may be found on page 129 and following in [SS10]).

Model selection

Once the likelihood is given, model selection for the choice of parameters p and q can be performed via usual criteria.

To avoid confusion, we now denote by

$$\widehat{\beta}_{p,q} = (\hat{\mu}, \hat{\phi}_1, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q) \quad \text{and} \quad \hat{\sigma}^2_{p,q}$$

the maximum likelihood estimators in the ARMA(p, q) model, that is

$$\widehat{\beta}_{p,q}, \hat{\sigma}^2_{p,q} = \text{argmin}_{\beta_{p,q},\sigma^2} - 2\ell_n(\beta_{p,q}, \sigma^2)$$

AICc and BIC

AICc

The corrected AIC (Akaike Information Criterion) choose p and q that minimize

$$-2\ell_n(\widehat{\beta}_{p,q}, \hat{\sigma}^2) + 2\frac{(p+q+1)n}{n-p-q-2}.$$

BIC

The BIC (Bayesian Information Critetion) choose p and q that minimize

$$-2\ell_n(\widehat{\beta}_{p,q}, \hat{\sigma}^2) + \log(n)(p+q+1).$$

Residuals

The final step of model building is diagnostics on residuals.

Standardized innovations (residuals)

In a given model, the standardized innovations are given by

$$\frac{X_i - X_i^{(i-1)}}{\sqrt{P_i^{(i-1)}}}$$

for $i = 1, \ldots, n$.

Diagnostics on residuals

In the model is correct standardized innovations should behave like a white noise (even a Gaussian white noise if Gaussian maximum likelihood has been used.)

1. Plot the standardized innovations and their ACFs.
2. To check for normality, plot a histogram or a QQ-plot.
3. Verify that the ACF coefficients stay in the confidence interval for $h \geq 1$
4. Use a Ljung-Box test

Ljung-Box test

To test $\mathcal{H}_0 : \omega$ is a white noise, use the test statistic

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}^2_{p,q}(h)}{n-h}$$

where $\hat{\rho}_{p,q}$ is the sample ACF of the residuals in a given ARMA(p, q) model.

Q is asymptotically (under mild conditions) of χ^2_{H-p-q} distribution.

Diagnostics on residuals

In the model is correct standardized innovations should behave like a white noise (even a Gaussian white noise if Gaussian maximum likelihood has been used.)

1. Plot the standardized innovations and their ACFs.
2. To check for normality, plot a histogram or a QQ-plot.
3. Verify that the ACF coefficients stay in the confidence interval for $h \geq 1$
4. Use a Ljung-Box test

Ljung-Box test

To test $\mathcal{H}_0 : \omega$ is a white noise, use the test statistic

$$Q = n(n+2)\sum_{h=1}^{H} \frac{\hat{\rho}^2_{p,q}(h)}{n-h}$$

where $\hat{\rho}_{p,q}$ is the sample ACF of the residuals in a given ARMA(p,q) model.

Q is asymptotically (under mild conditions) of χ^2_{H-p-q} distribution.

We know that for the sample acf of a white noise

$$\sqrt{n}(\hat{\rho}(1), \dots, \hat{\rho}(h)) \xrightarrow{\mathcal{L}} \mathcal{N}(\vec{0}, I_H)$$

so that

$$n \sum_{h=1}^{H} \hat{\rho}^2(h) \xrightarrow{\mathcal{L}} \chi^2_n$$

see G. M. Ljung ; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". Biometrika. for the end...

This is my link

Chapter 6 : Chasing stationarity, exploratory data analysis

- **Why do we need to chase stationarity ?**
  Because we want to do statistics : averaging lagged products over time, as in the previous section, has to be a sensible thing to do.
- **But....**
  Real time series are often non-stationary, so we need methods to "stationarize" the series.

- Plot the time series.
- Look for trends, seasonal components, step changes, outliers
- Transform data so that residuals are stationary
  1. Estimate and substracts T_t, and S_t
  2. Differencing
  3. Nonlinear transofrmations $(\log ; \surd, \ldots)$
- Fit model to residuals

An example I

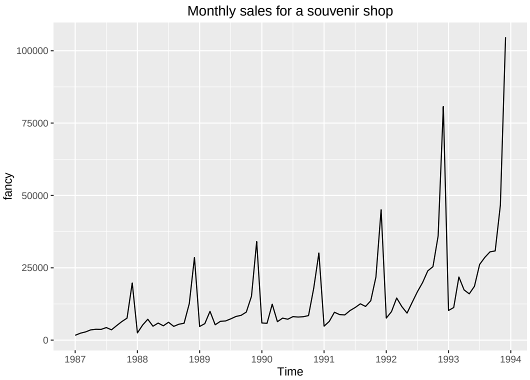


FIG. : Monthly sales for a souvenir shop on the wharf at a beach resort town in Queensland, Australia. [MWH08]

An example II

Notice that the variance grows with the mean, this usually calls for a log transformation ($X \to \log(X)$), which is part of the general family of Box-Cox transformation

$$\begin{cases} X \to X^{\lambda-1}/\lambda & \lambda \neq 0 \\ X \to \log(X) & \lambda = 0 \end{cases}$$

An example III

FIG.: Log of monthly sales.

The series is not yet stationary because there are a trend and a seasonal components.

An example IV

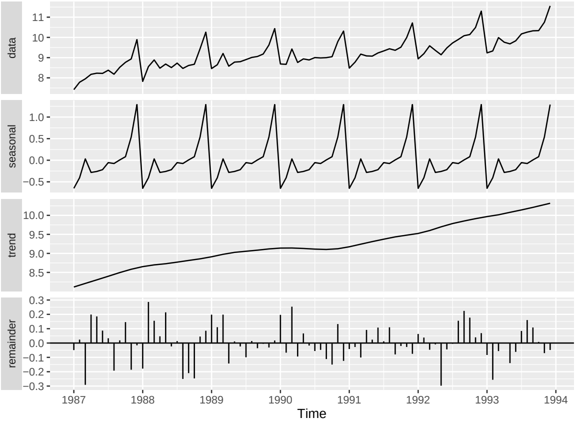


FIG.: Decomposition of monthly sales with `slt` function in R

Classical decomposition of a time series

$$Y_t = T_t + S_t + X_t$$

où

- $T = (T_t)_{t\in\mathbb{Z}}$ is the trend
- $S = (S_t)_{t\in\mathbb{Z}}$ is the seasonality
- $X = (X_t)_{t\in\mathbb{Z}}$ is a stationary centered time series.

Back to the global temperature I

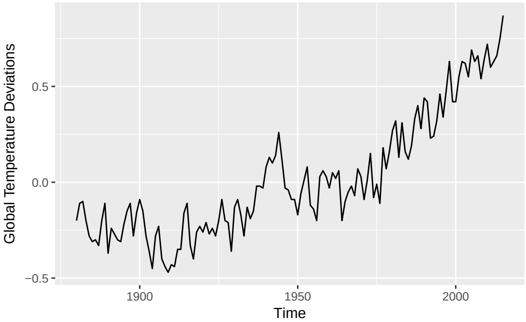


FIG. : Global temperature deviation (in $^\circ C$) from 1880 to 2015, with base period 1951-1980 - see slide 7

Back to the global temperature II

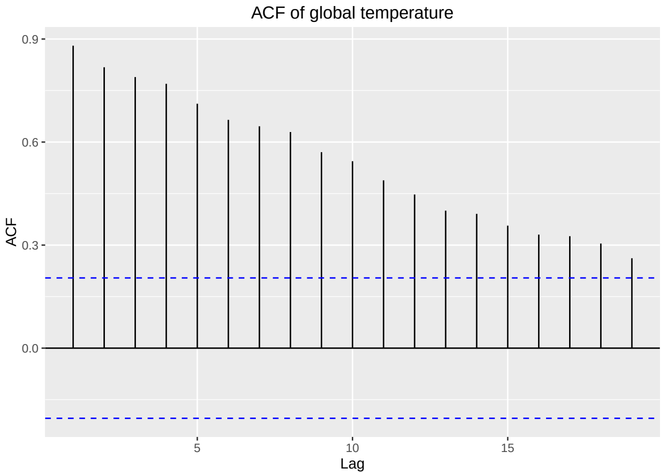


FIG.: ACF of global temperature deviation

Back to the global temperature III

We model this time series as

$$Y_t = T_t + X_t$$

and are now looking for a model for $T = (T_t)_{t \in \mathbb{Z}}$.
Looking at the series, two possible models for T are

- (model 1) a linear function of t $T_t = \beta_1 + \beta_2 t$
- (model 2) a random walk with drift $T_t = \delta + T_{t-1} + \eta_t$, where η is a white noise (see slide 4).

In both models, we notice that

$$Y_t - Y_{t-1} = T_t - T_{t-1} + \omega_t - \omega_{t-1} = \beta_2 + \omega_t - \omega_{t-1} \text{ (model 1)}$$
$$Y_t - Y_{t-1} = T_t - T_{t-1} + \omega_t - \omega_{t-1} = \delta + \eta_t + \omega_t - \omega_{t-1} \text{ (model 2)}$$

are stationary time series (check this fact as an exercise).

Back to the global temperature IV

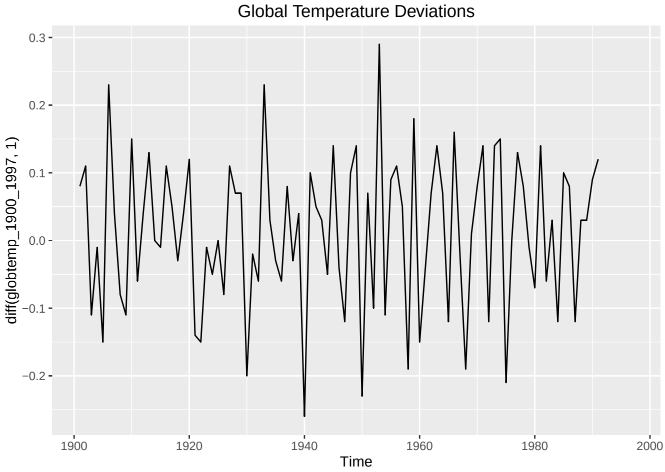


FIG.: Differenced global temperature deviation

Back to the global temperature V

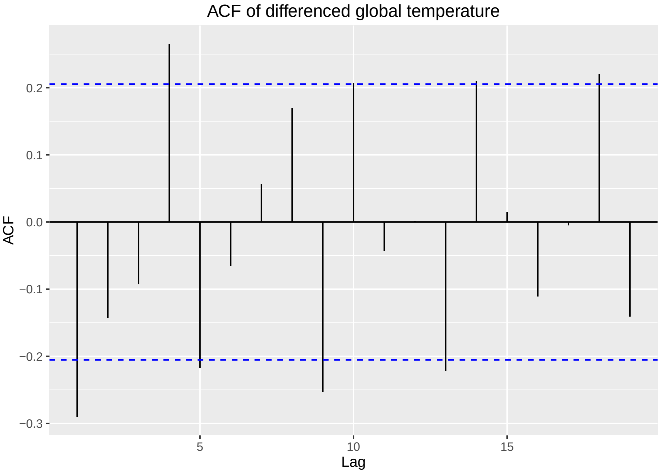


FIG.: Differenced global temperature deviation

Not far from a white noise!!

Backshift operator

Backshift operator

For a time series X, we define the **backshift operator** as

$$BX_t = X_{t-1},$$

similary

$$B^k X_t = X_{t-k}.$$

Difference of order d

Differences of order d are defined as

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

To stationarize the global temperature series, we applied the 1st order difference to it.
See http ://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html for an example of 2nd order integrated ts.

Moving average smoother

Moving average smoother

For a time series X,

$$M_t = \sum_{j=-k}^{k} a_j X_{t-j}$$

with $a_j = a_{-j} \geq 0$ and $\sum_{j=-k}^{k} a_j = 1$ is a symmetric moving average.

Note : `slt` function in R uses loess regression, the moving average smoother is just a loess regression with polynomials of order 1. More details on this on http ://www.wessa.net/download/stl.pdf, [CCT90].

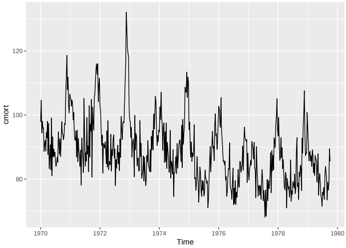


FIG. : Average daily cardiovascular mortality in Los Angeles county over the 10 year period 1970-1979.

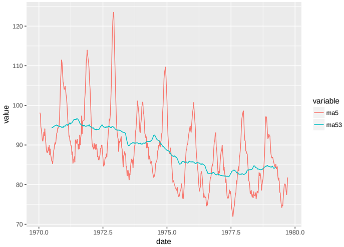


FIG.: Smoothed (ma 5 and 53) mortality

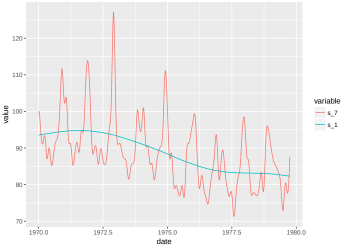


FIG.: Smoothed (splines) mortality

Chapter 7 : Non-stationarity and seasonality

Integrated models

We now introduce a new class of models, which, based on ARMA models, incorporates a wide range of **non-stationary series**.

ARIMA(p,d,q) model

A process X is said to be ARIMA(p, d, q) if

$$(1-B)^d X_t$$

is an ARMA(p,q). We can rewrite

$$\Phi(B)(1-B)^d X_t = \Theta(B)\omega_t.$$

If $\mathbb{E}\big((1-B)^d X_t\big) = \mu$, we write

$$\Phi(B)(1-B)^d X_t = \delta + \Theta(B)\omega_t,$$

where $\delta = \mu(1-\phi_1-\phi_2-\ldots-\phi_p)$.
Caution : X is not stationary but $(1-B)^d X$ is, provided that the roots of Φ are outside the unit circle.

A special example

Random walk with drift

Consider the model of slide 4

$$X_t = \delta + X_{t-1} + \omega_t$$

with $X_0 = 0$

1. Check that X is an ARIMA(0,1,0).
2. Given data $X_1, \ldots, X_n$, give the one-step-ahead prediction of X_{n+1}
3. Deduce the *m*-step-ahead prediction $X_{n+m}^{(n)}$
4. Compute the prediction error $P_{n+m}^{(n)}$.

A special example

Random walk with drift

Consider the model of slide 4

$$X_t = \delta + X_{t-1} + \omega_t$$

with $X_0 = 0$

1. Check that X is an ARIMA(0,1,0).
2. Given data $X_1, \dots, X_n$, give the one-step-ahead prediction of X_{n+1}
3. Deduce the m-step-ahead prediction $X_{n+m}^{(n)}$
4. Compute the prediction error $P_{n+m}^{(n)}$.

correction pages 142 and following

Forecasting in ARIMA models

Exponentially weighted moving averages

Consider the process :

$$X_t = X_{t-1} + \omega_t - \theta\omega_{t-1},$$

with $|\theta| < 1$

1. Write it as an ARIMA(0,1,1).
2. Define $Y_t = \omega_t - \theta\,\omega_{t-1}$ and verify that Y is invertible.
3. Deduce that

$$X_t = \sum_{j=1}^{\infty}(1-\theta)\theta^{j-1}X_{t-j} + \omega_t$$

4. and finally that, based on the data $X_1, \ldots, X_n$

$$X_{n+1}^{(n)} = (1-\theta)X_n + \theta X_n^{(n-1)}.$$

Forecasting in ARIMA models

Exponentially weighted moving averages

Consider the process :

$$X_t = X_{t-1} + \omega_t - \theta\omega_{t-1},$$

with $|\theta| < 1$

1. Write it as an ARIMA(0,1,1).
2. Define $Y_t = \omega_t - \theta\,\omega_{t-1}$ and verify that Y is invertible.
3. Deduce that
$$X_t = \sum_{j=1}^{\infty}(1-\theta)\theta^{j-1}X_{t-j} + \omega_t$$
4. and finally that, based on the data $X_1, \dots, X_n$
$$X_{n+1}^{(n)} = (1-\theta)X_n + \theta X_n^{(n-1)}.$$

see https ://www.stat.berkeley.edu/ bartlett/courses/153-fall2010/lectures/14.pdf p13

Building ARIMA model I

Here a the steps you should follow to build an ARIMA model

1. Construct a time plot of the data and inspect the graph for anomalies. For example, if the variability in the data depends upon time, you need to stabilize the variance via a Box-Cox transformation
2. Transform the data and construct a new time plot.
3. Choice of d : a look at the new time plot will help you determine if a differentiation is needed. If it is the case
   - Differentiate the series and inspect the time plot
   - If additional differentiating is required, apply the operator $(1 - B)^2$ and so on
   - Do not forget that a ACF decreasing too slowly is also a sign of non-stationarity
   - Caution : do not differentiate too many times

Counter example

Show that if ω_t is a white noise, $\omega_t - \omega_{t-1}$ is a MA(1) ! !

Building ARIMA model II

4. The next step is to identify reasonable values (or a set of reasonable values) for q, p
   4.1 Represent the ACFs and PACFs of the differentiated series (they can be more than one if you hesitate between two values for d) and
   4.2 choose few reasonable values for q and p
5. At this stage, you should have few preliminary reasonable values for d, q and p. Estimate the parameters in the different models and compute their AICc and BIC.
6. Choose one model and conduct diagnostic tests on its residuals

Back to seasonality

Can you you think of a model for data with these sample ACF and PACF ?

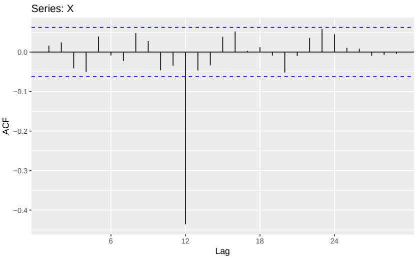


FIG.: ACF

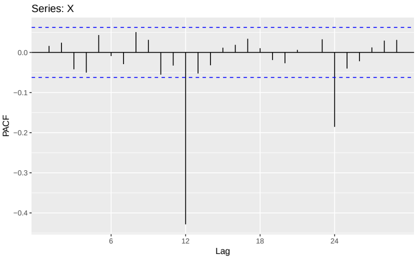


FIG.: PACF

Pure seasonal ARMA model

Pure seasonal ARMA$(P, Q)_s$

A pure seasonal **ARMA**$(P, Q)_s$ process X is a stationary process that is defined through

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)\omega_t$$

where $\omega \sim WN(0, \sigma^2)$, Φ_P is a polynomial of order P, Θ_Q is a polynomial of order Q and Φ_P and Θ_Q have no common factors.

Exercise

1. Verify that the pure seasonal ($s = 12$) ARMA$(0\,,1)_{12}$ (this is an MA$(1)_{12}$) has a ACF given by

$$\rho(12) = \theta/(1+\theta^2)$$
$$\rho(h) = 0 \text{ otherwise.}$$

2. Verify that the pure seasonal ($s = 12$) ARMA$(1\,,0)_{12}$ (this is an AR$(1)_{12}$) has a ACF given by

$$\rho(12k) = \phi^k \quad \text{for } k = 1, \dots$$
$$\rho(h) = 0 \text{ otherwise.}$$

Exercise

1. Verify that the pure seasonal ($s = 12$) ARMA$(0,1)_{12}$ (this is an MA$(1)_{12}$) has a ACF given by

$$\rho(12) = \theta/(1+\theta^2)$$
$$\rho(h) = 0 \text{ otherwise.}$$

2. Verify that the pure seasonal ($s = 12$) ARMA$(1,0)_{12}$ (this is an AR$(1)_{12}$) has a ACF given by

$$\rho(12k) = \phi^k \quad \text{for } k = 1, \dots$$
$$\rho(h) = 0 \text{ otherwise.}$$

correction for this exercise

ARMA$(p, q) \times (P, Q)_s$

In general, we will mix seasonal and non-seasonal operators to build **multiplicative seasonal ARMA** : ARMA$(p, q) \times (P, Q)_s$.

ARMA$(0, 1) \times (1, 0)_{12}$

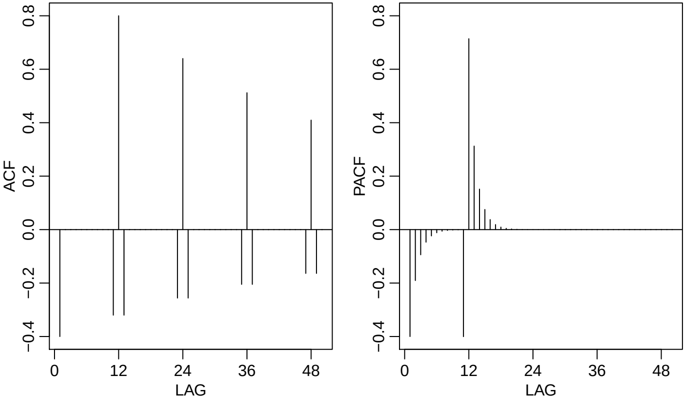


FIG.: ACF and PACF of the process $(1 - 0.8B^{12})X_t = (1 - 0.5B)\omega_t$ see [SS10]

ARMA$(p, q) \times (P, Q)_s$

In general, we will mix seasonal and non-seasonal operators to build **multiplicative seasonal ARMA** : ARMA$(p, q) \times (P, Q)_s$.

ARMA$(0, 1) \times (1, 0)_{12}$

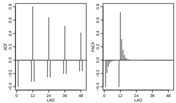


FIG.: ACF and PACF of the process $(1 - 0.8B^{12})X_t = (1 - 0.5B)\omega_t$ see [SS10]

correction for this exercise

SARIMA

Multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$

A multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ process X is a process that is defined through

$$\Phi_P(B^s)\Phi(B)(1 - B^s)^D(1 - B)^d X_t = \Theta_Q(B^s)\Theta(B)\omega_t$$

where

- $\omega \sim WN(0, \sigma^2)$,
- Φ_P is a polynomial of order P
- Θ_Q is a polynomial of order Q
- Φ is a polynomial of order p
- Θ is a polynomial of order q

Model building

To choose p,q,P,Q,d,D

- First difference sufficiently to get to stationarity.
- Then find suitable orders for ARMA or seasonal ARMA models for the differenced time series. The ACF and PACF is again a useful tool here.
- Select few models, compare their AICc and BIC
- Finally conduct a diagnosis check for the residuals of the select model.

Peter J Brockwell and Richard A Davis, *Time series : theory and methods*, Springer Science & Business Media, 2013.

Robert B Cleveland, William S Cleveland, and Irma Terpenning, *Stl : A seasonal-trend decomposition procedure based on loess*, Journal of Official Statistics **6** (1990), no. 1, 3.

Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman, *Forecasting methods and applications*, John Wiley & Sons, 2008.

Robert H Shumway and David S Stoffer, *Time series analysis and its applications : with r examples*, Springer Science & Business Media, 2010.