M.Sc. in Stochastic and Data Science

# Statistics for Stocastic Processes

A.Y. 2020/2021

Prof. Elvira DI NARDO - Prof. Luis Alberiko Gil-Alaña

Written by Gianluca MITTONE

# Contents

# 1 Lecture 1

### Content summary of the course

Why study statistics specifically for stochastic processes? Because the first assumption in any statistics we have seen in the undergraduate course is that the data is iid (independent and identically distributed); this is not always the case (e.g., stock market data, epidemiology...). When data are ordered (usually through time) and are correlated between themselves, the standard statistical procedures cannot work correctly; we need a tool called **Time Series Analisys**. However, what is a time series?

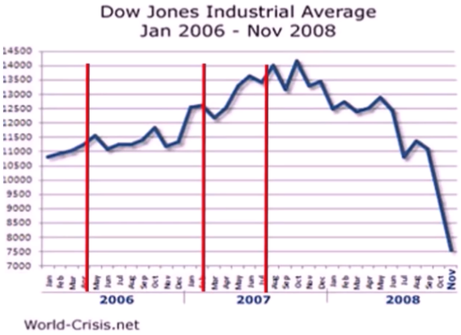
Data are observed at discrete time from a stochastic dynamical system, usually described with a stochastic process.

**Definition 1.1.** *An **observed times series** are observation of a sequence of random variables indexed by a set of real numbers:*

$$(x_t)_{t \in T}, x_t \in \mathbb{R}, T \subset \mathbb{R}$$

*where $T = \{t_1, ..., t_n, ..\}$ is assumed to be discrete. If there is only one variable the then we have a **univariate** time series ($k = 1$), otherwise we have a **multivariate** time series ($k > 1$).*

In general we also assume that $t \in \mathbb{Z}$, making the sampling point **equidistant**. The goal of a time series analisys is then to develop mathematical models that provide reasonable description of simple data.



World-Crisis.net

Figure 1: example of a simple time series

There are two main approaches to time series analysis, and they are not to be seen as exclusive of one another:

- **Time domain approach**: assumes that correlation among adjacent points is best explained as the dependence of current value on past values; the present is seen as a parametric function of the past. The simplest of these models is the linear one: the current value of a variable is calculated as a linear regression of its past values. ARIMA (Autoregressive integrated moving average models) are an example of this approach that we will see later in the course. A more modern approach is the additive one, in which the observed variables are seen as a sum of different series.

- **Frequency domain approach**: this approach focuses on periodicity and sinusoidal variation of the interested data.

Typically the two approaches lead to very similar results, especially with much data available. Nevertheless, what distinguishes statistical analysis of iid data from time series analysis?

Recall the notion of joint pdf (probability density function): Knowledge of $(X_t)_{t\in\mathbb{Z}} \Leftrightarrow$ joint distribution function $\mathbb{P}\left(X_{t_1} \leq x_1, ..., X_{t_n} \leq x_n\right), (x_1, ..., x_n) \in \mathbb{R}^n$ of $X_{t_1}, ..., X_{t_n}$ for any choiche of $t_1, ..., t_n \in \mathbb{Z}$ and for all $n \in \mathbb{N}$. How is possible to estimate $\mathbb{P}$? If $(x_t)_{t\in\mathbb{Z}}$ are iid then $\mathbb{P}$ is fully specified by the density function $F_{X_0}(x)$ that can be calculated as the empirical marginal distribution:

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}_{\{x_t \leq x\} \approx F_{X_0}(x)}$$

with $\{x_1, ..., x_n\}$ sampled (Glivenko-Cantelli theorem). If $X_t$ are not iid there is no practical way to estimate $\mathbb{P}$ and the sample size has no more meaning; this is due to the **lack of stability**: distribution of $X_{t+1}, X_{t+2}, ..., X_{t_n}$ changes too much as a function of t so that the observed time series $x_1, ..., x_n$ is not sufficiently representative of $\mathbb{P}$ on $\mathbb{R}^{\mathbb{Z}}$. This implies that also the classical estimator for the mean $\mathbb{E}\left[X_t\right]$ and the variance $Var(X_t)$ may not be significant.

**Example 1.1.** *Consider the sequence $(X_t)_{t\in\mathbb{Z}}$ such that $X_t = \mu t + W_t$ where $\mu > 0$, $(W_t)_{t\in\mathbb{Z}}$ are iid and distributed as $\mathcal{N}(0, \sigma^2)$. If we compute the sample mean we obtain:*

$$\begin{aligned}
\bar{X}_n &= \frac{1}{n}\sum_{t=1}^{n} X_t \\
&= \frac{1}{n}\sum_{t=1}^{n}(\mu t + W_t) \\
&= \frac{\mu}{n}\sum_{t=1}^{n} t + \frac{1}{n}\sum_{t=1}^{n} W_t \\
&= \frac{\mu}{n}\frac{n(n+1)}{2} + \bar{W}_t \\
&= \mu\frac{n+1}{n} + \bar{W}_t
\end{aligned}$$

*Now, for the Law of Large Numbers:*

$$\lim_{n\to+\infty}(\mu\frac{n+1}{n} + \bar{W}_t) \xrightarrow{a.s.} +\infty$$

*This implies that it will be more useful to work with sequences of random variable with constant means; we can achieve this by considering a transformation like $X_t - \mu t = W_t$.*

**Example 1.2.** *Consider the sequence $(X_t)_{t\in\mathbb{Z}}$ such that*

$$X_t = \begin{cases} 0 & t \leq 0 \\ \sum_{s=1}^{t} W_s & t > 0 \end{cases}$$

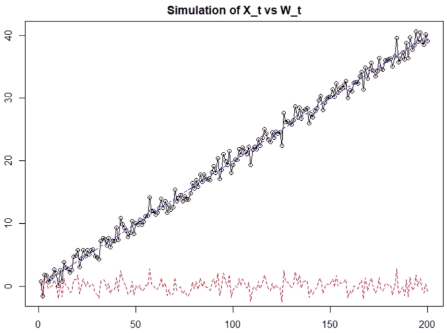


Figure 2: example of the result of a transformation like $X_t - \mu t = W_t$

*where* $(W_t)_{t\in\mathbb{Z}}$ *are iid and distributed as* $\mathcal{N}(0,\sigma^2)$. *The sample mean is then:*

$$\begin{aligned}
\bar{X}_n &= \frac{1}{n}\sum_{t=1}^{n} X_t \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t} W_s \\
&= \frac{1}{n}[W_1 + (W_1 + W_2) + ... + (W_1 + ... + W_n)] \\
&= \frac{1}{n}[nW_1 + (n-1)W_2 + ... + W_n] \\
&= \frac{1}{n}\sum_{t=1}^{n}(n-t+1)W_t
\end{aligned}$$

*Clearly* $\mathbb{E}\left[\bar{X}_t\right] = 0$, *since* $\mathbb{E}\left[W_t\right] = 0$. *Now let us compute the variance of the sample mean:*

$$\begin{aligned}
Var(\bar{X}_n) &= \frac{1}{n^2}Var(\sum_{j=1}^{n} jW_{n+1-j}) \leftarrow \textit{we made an index charge here} \\
&= \frac{1}{n^2}\sum_{j=1}^{n} j^2 Var(W_{n+1-j}) \\
&= \frac{\sigma^2}{n^2}\sum_{j=1}^{n} j^2 \\
&= \frac{\sigma^2}{n^2}\frac{n(n+1)(2n+1)}{6} \\
&= \frac{\sigma^2}{n}\frac{(n+1)(2n+1)}{6}
\end{aligned}$$

*But it can be seen that:*

$$\lim_{n\to+\infty}\frac{\sigma^2}{n}\frac{(n+1)(2n+1)}{6} \to +\infty$$

*This explains why classical statistical tools are not fit to time series: usually the sample mean is used because its variance tends to 0 as the sample size grows* *large, but this does not happen in the context of time series. It is clearly more useful to work with random variables that have a constant variance.*

**Remark 1.1.** *Applying the transformation $X_t - X_{t-1} = W_t$, with $t \in \mathbb{Z}$, it is possible to repeat the same calculation and obtain something more useful. We will study this later on.*

Usually, the first step for a correct time series analysis is to produce a visual description of the data, like a plot with the observed values on the y-axis and the time on the x-axis. This could help to seek properties of the data like trends or seasonality. A common tool for doing this is the **scatter diagram**, which shows the correlations in the data. Another technique is the **decomposition**, which allows separating the noise from the signal, allowing, for example, looking for components explicable with different dynamics. These techniques allow us to separate a remainder (stochastic) component from the data. We will have to guess a model for it employing various techniques (Shapiro test, QQ plot, histogram...). If the data appears to be uncorrelated, it can be a case of **white noise**, that is a sequence of uncorrelated random variables. To assess the quality of the obtained model we can use **hypotesis testing** or the **AIC** and **BIC** criterion. The **forecasting** techniques exploit the stochastic model proposed to explain the data behavior up the given time to predict its dynamics after it.

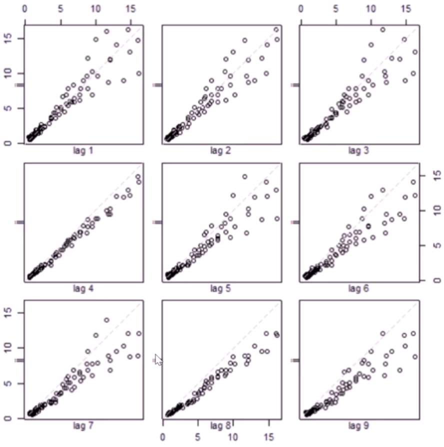


Figure 3: example of a scatter plot

The **frequency domain approach** analyses the time series as a sum of sinusoidal components with uncorrelated random coefficients. This is called the **spectral representation** of the time series. The main tools for doing this are the **discrete Fourier transformations** and their statistical properties. An example of this approach could be

$$X_t = \sum_k a_k \sin(2\pi f_k t) + \sum_k b_k \cos(2\pi f_k t)$$

This approach clearly applyies naturally to data that have a periodic (cyclic) nature. A tool used in this context is the **periodogram** that graphs a measure of the relative importance of possible frequency values that might explain the oscillation pattern of the observed data.

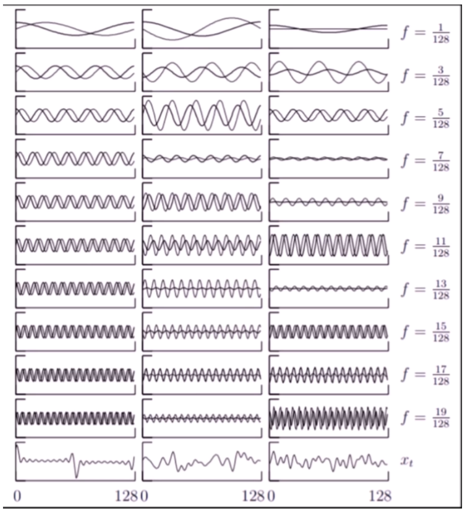


Figure 4: example of how a time series can be seen as a summation of trigonometrical functions

# 2 Lecture 2

**Gaussian time series. Mean, variance, autocovariance, autocorrelation. Examples: random walk, periodic signals. Properties of autocovariance functions.**

The Gaussian model is the first that a data scientist typically tries when he thinks that the observation result from a superposition of many factors, occurring independently from each other, at least asymptotically. The Central Limit Theorem justifies this assumption.

To work with **Gaussian time series**, we have to recall the notion of **multivariate Gaussian distribution**.

**Definition 2.1.** *The **bivariate Gaussian distribution** joint pdf is defined as:*

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

*where* $\mathbb{E}\left[X\right] = \mu_X$, $\mathbb{E}\left[Y\right] = \mu_Y$, $D[X] = \sigma_X$, $D[Y] = \sigma_Y$ *and* $\rho = \frac{\mathbb{E}[[X-\mu_x][Y-\mu_Y]]}{\sigma_X\sigma_Y}$ *is the correlation coefficient. It is clear that this distribution is completely specified by the means vector and the covariance matrix:*

$$\boldsymbol{\mu}^\mathsf{T} = \begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix} = \begin{pmatrix}\sigma_X^2 & cov(X,Y)\\ cov(X,Y) & \sigma_Y^2\end{pmatrix}$$

*allowing us to write the same formula but in matrix notation:*

$$f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{det(\boldsymbol{\Sigma})}}\exp\left\{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\right\}$$

*Using this generalisation it is straight-forward to generalise the expression to an* $n > 2$ *random Gaussian vector:*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{det(\boldsymbol{\Sigma})}}\exp\left\{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\right\}$$

The marginal and conditional distributions of a multivariate Gaussian distribution are still Gaussian. Remember also that this family distribution is closed under linear transformations. Before introducing the notion of Gaussian time series, we need to recall few more notions:

**Definition 2.2.** $\boldsymbol{z} = (Z_1, ..., Z_m)$ *is a m-dimensional standard Gaussian random vector iif* $Z_1, ..., Z_m$ *are iid random variables* $\sim \mathcal{N}(0, \sigma^2)$*. Let us consider another standard Gaussian random vector* $\boldsymbol{x} = (X_1, ..., X_n)$*. Then there exists a* $\boldsymbol{b} \in \mathcal{R}^n$ *and a* $\boldsymbol{A} \in \mathcal{R}^{n\times m}$ *such that:*

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{Z}^\mathsf{T} + \boldsymbol{b}^\mathsf{T}$$

*We than have that:*

$$\boldsymbol{\mu} = \mathbb{E}\left[\boldsymbol{x}\right] = \boldsymbol{b} \text{ and } \boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{x}^\mathsf{T} - \boldsymbol{b}^\mathsf{T})(\boldsymbol{x}-\boldsymbol{b})\right] = \boldsymbol{A}\boldsymbol{A}^\mathsf{T} \in \mathbb{R}^{n\times n}$$

*We can note a multivariate Gaussian distribution in matrix notation as:*

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Remind that:

**Definition 2.3.** *A matrix is non-singular when:*

$$det(\boldsymbol{\Sigma}) > 0$$

*This condition is equivalent to say that the inverse matrix exists ($\boldsymbol{\Sigma}^{-1} \in \mathbb{R}^{n \times n}$) or that the $\boldsymbol{\Sigma}$ is positive definite ($\boldsymbol{\Sigma} > 0 \Rightarrow \boldsymbol{x}\boldsymbol{\Sigma}\boldsymbol{x}^\intercal > 0 \; \forall \boldsymbol{x} \in \mathbb{R}^n$).*

*If $det(\boldsymbol{\Sigma}) \geq 0$ then the matrix is said to be non negative definite.*

**Definition 2.4.** *A multivariate Gaussian distribution is said to be singular when:*

$$det(\boldsymbol{\Sigma}) = 0$$

*In such a case the possible value of the vector $\boldsymbol{x}$ are contrained to lie in a subspace of $\mathbb{R}$ with dimension equal to the rank of $\boldsymbol{\Sigma}$, that is:*

$$rank(\boldsymbol{\Sigma}) = k < n \Rightarrow \exists \boldsymbol{C} \in \mathbb{R}^{n \times k} | \boldsymbol{x} = \boldsymbol{C}\boldsymbol{z}^\intercal + \boldsymbol{b}^\intercal$$

*with $\boldsymbol{z} \in \mathcal{N}(0, \boldsymbol{I}_k)$.*

We are now ready to define Gaussian time series.

**Definition 2.5.** *A time series $(X_t)_{t \in \mathbb{Z}}$ is said to be a* ***Gaussian time series*** *iif*

$$\forall n \in \mathbb{N}, \; \forall t_1, ..., t_n \in \mathbb{Z}$$

*there exists*

$$\boldsymbol{b}_{(t_1,...,t_n)} \in \mathbb{R}^n, \; \boldsymbol{\Sigma}_{(t_1,...,t_n)} \geq 0$$

*such that*

$$\boldsymbol{X}_{(t_1,...,t_n)} \equiv (X_{t_1}, ..., X_{t_n}) \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{\Sigma}_{(t_1,...,t_n)})$$

*That is, the random vector of observation is distributed according to a multivariate Gaussian distribution.*

One more tool that we will use with Gaussian time series is the **characteristic function**, which in this case is defined as:

$$\Phi(\boldsymbol{z}) \equiv \mathbb{E}\left[e^{i\boldsymbol{z}\boldsymbol{x}^\intercal_{(t_1,...,t_n)}}\right] = \exp\left\{i\boldsymbol{z}\boldsymbol{b}^\intercal_{(t_1,...,t_n)} - \frac{1}{2}\boldsymbol{z}\boldsymbol{\Sigma}_{(t_1,...,t_n)}\boldsymbol{z}^\intercal\right\}$$

Now we will introduce the **special functions** of time series. We assume that $(X_t)_{t \in \mathbb{Z}} \in \mathcal{L}^2$, meaning that all the observede variables have finite second moment ($\mathbb{E}\left[{}^{2}_{t}\right] < \infty, \forall t \in \mathbb{Z}$).

**Definition 2.6.** *The* ***mean*** *of a time series is defined as*

$$\mu_t = \mathbb{E}\left[X_t\right], \; t \in \mathbb{Z}$$

**Definition 2.7.** *The* ***Auto Covariance Function (ACF)*** *of a time series is defined as*

$$\gamma(s,t) \equiv cov(X_s, X_t) = \mathbb{E}\left[(X_t - \mu_t)(X_s - \mu_s)\right], \; t, s \in \mathbb{Z}$$

*Observe that $\gamma : \mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{R}$.*

**Definition 2.8.** *The **Auto Correlation Function (ACF)** of a time series is defined as*

$$\rho(s,t) \equiv \frac{\gamma s,t}{\sqrt{\gamma(s,s)\gamma(t,t)}},\ t,s \in \mathbb{Z}$$

*Observe that*

$$\gamma(s,s) = cov(X_s,X_s) = \mathbb{E}\left[(X_s-\mu_s)^2\right] = Var(X_s)$$

*making it possible to rewrite the previous definition as*

$$\rho(s,t) \equiv \frac{\gamma s,t}{\sqrt{Var(X_s)Var(X_t)}},\ t,s \in \mathbb{Z}$$

**Lemma 2.1.** $\rho(s,t) \in [-1,1]$*,* $t,s \in \mathbb{Z}$

*Proof.* we will use the Cauchy-Schwart's inequality:

$$\mathbb{E}\left[XY\right] \leq \mathbb{E}\left[|XY|\right] \leq \sqrt{\mathbb{E}\left[X^2\right]}\sqrt{\mathbb{E}\left[Y^2\right]}$$

Squaring both external sides:

$$\mathbb{E}\left[XY\right]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

Now replace $X = X_t - \mu_t$ and $Y = X_s - \mu_s$:

$$\mathbb{E}\left[(X_t-\mu_t)(X_s-\mu_s)\right]^2 \leq \mathbb{E}\left[(X_t-\mu_t)^2\right]\mathbb{E}\left[(X_s-\mu_s)^2\right]$$

Remembering definition 2.8:

$$\begin{aligned}\mathbb{E}\left[(X_t-\mu_t)(X_s-\mu_s)\right]^2 &\leq Var(X_t)Var(X_s)\\ \left(\frac{\mathbb{E}\left[(X_t-\mu_t)(X_s-\mu_s)\right]}{\sqrt{Var(X_t)Var(X_s)}}\right)^2 &\leq 1\\ \rho(s,t)^2 &\leq 1\end{aligned}$$

□

**Remark 2.1.** *If* $\rho(s,t) = \pm 1$ *then there is a **strong linear dependence** between* $X_t$ *and* $X_s$*, that it*

$$\mathbb{P}\left(X_t = aX_s + b\right) = 1$$

**Definition 2.9.** *A time series* $(X_t)_{t\in\mathbb{Z}}$ *is called **stationary** if:*

- $X_t \in \mathbb{L}^2\ \forall t \in \mathbb{Z}$
- $\mu_t = \mu\ \forall t \in \mathbb{Z}$
- $\gamma(s,t) = \gamma(0,t-s)\ \forall t \in \mathbb{Z}$ *(this is also notated as* $\gamma(h)$*, assuming the **lag window*** $h = t - s$*)*

**Corollary 2.1.** *If* $(X_t)_{t\in\mathbb{Z}}$ *is stationary then* $Var(X_t)$ *is constant* $\forall t \in \mathbb{Z}$*.*

*Proof.* Observe that $\gamma(t,t) = Var(X_t) = \gamma(0,t-t) = \gamma(0) = const \geq 0$. □

What about the correlation function? In this case we have:

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

A less powerful property if the **weak stationarity**: a weak stationary time series can be stationary in *mean* is it has a constant mean or stationary in *variance* is the variance is constant.

**Example 2.1.** *Now we will see a classical example of non-stationary time series, the random walk. Consider* $X_t = \mu + X_{t-1} + W_t$ *for* $t = 1, 2, ...$ *where:*

- $\mu > 0$
- $(W_t)_{t \in 1,2,...}$ *iid with* $\mathbb{E}\left[W_t\right] = 0,\ \mathbb{E}\left[W_t^2\right] = \sigma^2$
- $\mathbb{P}\left(X_0 = 0\right) = 1$

*Is* $X_t$ *stationary? First of all, let's prove that* $X_t = \mu t + X_0 + \sum_{j=1}^{n} W_j$. *This is true for* $t = 1$ *since* $X_1 = \mu + X_0 + W_1$. *Proceeding by induction, we assume that the expression is true for an arbitrary* $t = s$*, obtaining:*

$$\begin{aligned} X_{s+1} &= \mu + X_s + W_{s+1} \quad X_s = s\mu + X_0 + \sum_{j=1}^{s} W_j \\ &= (s+1)\mu + X_0 + \sum_{j=1}^{s+1} W_j \end{aligned}$$

*This is useful beacuse we can now demonstrate that* $\mathbb{P}\left(X_0 = 0\right) = 1 \implies \mathbb{E}\left[X_0\right] = 0$*:*

$$\mathbb{E}\left[X_t\right] = \mu t + \mathbb{E}\left[X_0\right] + \sum_{j=1}^{s+1} \mathbb{E}\left[W_j\right] = \mu t \quad t = 0, 1, 2$$

*Then, since the mean is not constant, the time series is not stationary.*

**Exercise 2.1.** *Prove that* $\gamma(s,t) = \sigma^2 min\{s,t\}$ *for* $s, t = 1, 2, ....$

Now we will study some examples of periodic signals.

**Example 2.2.** *Assume*

$$X_t = R\sin(2\pi\omega t + \phi) + W_t \quad t \in \mathbb{Z}$$

*where*

- $(W_t)_{t \in \mathbb{Z}}$ *iid* $\sim \mathcal{N}(0, \sigma^2)$
- $R > 0$

*In the context of periodic signal all these constants have special meaning;* $R$ *is the* ***amplitude****,* $\omega$ *is the* ***frequency****,* $\phi$ *is the* ***phase shift*** *and* $p = \frac{1}{\omega}$ *is the* ***period****. Suppose* $R = 1$*,* $\phi = 0$ *and* $p = 2\pi$ *(so* $\omega = \frac{1}{2}\pi$*). Then we obtain:*

$$X_t = R\sin t + W_t$$

*The shift of the wave with respect to the crossing of the y-axis is given by* $-\frac{\phi}{2\pi\omega}$*. To better understand this, consider the following examples:*

$$2\sin\left(2\pi\frac{4}{2\pi}t-2\right)$$

*In this case,* $R=2$*,* $p=\frac{\pi}{2}$ *and* $-\frac{\phi}{2\pi\omega}=\frac{1}{2}$*. This mean that the beginning of the curve in this case is shifted toward left by* $\frac{1}{2}$*. In general:*

- $\phi > 0$ *implies a shift to the left;*
- $\phi < 0$ *implies a shift to the right;*

*Now the question is:* $X_t$ *is stationary? Well, start by observing that* $X_t \in \mathcal{L}^2$ *since* $W_t \in \mathcal{L}^2$ *and that* $\mathbb{E}[X_t] = R\sin(2\pi\omega t+\phi)$*. Specifically, this last term results in a non-constant* $\mu_t$*, implying that* $X_t$ *is not stationary. As an exercise we further calculate*

$$\gamma(s,t)=\mathbb{E}\left[(X_t-\mu_t)(X_s-\mu_s)\right]=\mathbb{E}\left[W_tW_s\right]=\begin{cases}0 & s\neq t\\ \sigma^2 & s=t\end{cases}$$

**Example 2.3.** *Consider:*

$$X_t = R\cos(2\pi\omega t+\phi) \quad t\in\mathbb{Z}$$

*with:*

- $R\in\mathcal{L}^2$
- $\phi\sim\mathcal{U}(-\pi,\pi) \implies f_\phi(\theta)=\begin{cases}\frac{1}{2\pi} & \theta\in[-\pi,\pi]\\ 0 & \theta\notin[-\pi,\pi]\end{cases}$
- $R\perp\phi$

*Again, the question is: is* $X_t$ *stationary? First, observe that* $X_t\in\mathcal{L}^2$ *since* $R,\phi\in\mathcal{L}^2$*. Let us now compute the mean of* $X_t$*:*

$$\mathbb{E}[X_t]=\mathbb{E}[R]\,\mathbb{E}[\cos(2\pi\omega t+\phi)]$$

*Since the two terms of this equation are independent, let su compute the value of the second one:*

$$\mathbb{E}[\cos(2\pi\omega t+\phi)]=\int_{-\pi}^{\pi}\cos(2\pi\omega t+\theta)\frac{1}{2\pi}d\theta=0$$

*This proves that:*

$$\mathbb{E}[X_t]=\mu_t=0 \quad \forall t\in\mathbb{Z}$$

*Now let us compute the covariance function:*

$$\begin{aligned}
\gamma(t,s) &= \gamma(t,t+h) \\
&= \mathbb{E}\left[(X_t-\mu_t)(X_{t+h}-\mu_{t+h})\right] \\
&= \mathbb{E}\left[X_t X_{t+h}\right] \\
&= \mathbb{E}\left[R^2\cos(2\pi\omega t+\phi)\cos(2\pi\omega t+2\pi\omega h+\phi)\right] \\
&= \mathbb{E}\left[R^2\cos(\alpha)\cos(\alpha+\beta)\right] \quad \alpha=2\pi\omega t+\phi,\ \beta=2\pi\omega h \\
&= \mathbb{E}\left[R^2\frac{1}{2}[\cos(2\alpha+\beta)+\cos\beta]\right] \\
&= \mathbb{E}\left[\frac{R^2}{2}[\cos(2\pi\omega h)+\cos(2(2\pi\omega t+\phi)+2\pi\omega h)]\right] \\
&= \mathbb{E}\left[\frac{R^2}{2}\right]\{\cos(2\pi\omega h)+\mathbb{E}\left[\cos[2\pi\omega(2t+h)+2\phi]\right]\} \\
&= \mathbb{E}\left[\frac{R^2}{2}\right]\cos(2\pi\omega h)
\end{aligned}$$

*Proving $X_t$ stationary.*

**Example 2.4.** *Consider*

$$X_t = A\cos(2\pi\omega t)+B\sin(2\pi\omega t) \quad t\in\mathbb{Z}$$

*Assume that:*

- $\mathbb{E}[A]=\mathbb{E}[B]=0$
- $\mathbb{E}\left[A^2\right]=\mathbb{E}\left[B^2\right]=\sigma^2$
- $A\perp B \implies \mathbb{E}[AB]]=0$

*Is $X_t$ stationary? Start by observing that $X_t\in\mathcal{L}^2$ since $A,B\in\mathcal{L}^2$ and that $\mathbb{E}[X_t]=0\ \forall t\in\mathbb{Z}$. Now let us compute the covariance function:*

$$\begin{aligned}
\gamma(t,t+h) &= \mathbb{E}\left[\{A\cos(2\pi\omega t)+B\sin(2\pi\omega t)\}\{A\cos(2\pi\omega(t+h))+B\sin(2\pi\omega(t+h))\}\right] \\
&= \mathbb{E}\left[A^2\right]\cos(2\pi\omega t)\cos(2\pi\omega(t+h))+\mathbb{E}\left[B^2\right]\sin(2\pi\omega t)\sin(2\pi\omega(t+h)) \\
&= \sigma^2[[A^2]\cos(2\pi\omega t)\cos(2\pi\omega(t+h))+\sin(2\pi\omega t)\sin(2\pi\omega(t+h))] \\
&= \sigma^2(\cos\beta\cos\alpha+\sin\beta\sin\alpha) \\
&= \sigma^2\cos(\alpha-\beta) \\
&= \sigma^2\cos[2\pi\omega(t+h)-2\pi\omega t] \\
&= \sigma^2\cos 2\pi\omega h
\end{aligned}$$

*Since $\gamma$ depends only on $h$, the time series is stationary.*

**Remark 2.2.** *Consider example 2.3 vs example 2.4. They have the same $\mathbb{E}[X_t]=0$ and the same $\gamma(t,t+h)=const$, so are they equal? This is true only* *when $A = R\cos\phi$ and $B = -R\sin\phi$. Check the hypotesis on $A$ and $B$:*

$$\mathbb{E}\left[A\right] = \mathbb{E}\left[R\cos\phi\right] = \mathbb{E}\left[R\right]\mathbb{E}\left[\cos\phi\right]]\,\mathbb{E}\left[R\right]\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos\theta d\theta = 0$$

$$\mathbb{E}\left[B\right] = \mathbb{E}\left[R\right]\mathbb{E}\left[\sin\phi\right] = 0$$

$$cov(A,B) = \mathbb{E}\left[AB\right] = \mathbb{E}\left[-R^2\cos\phi\sin\phi\right] = -\frac{1}{2}\mathbb{E}\left[R^2\right]\mathbb{E}\left[\sin 2\phi\right] = 0$$

$$\mathbb{E}\left[A^2\right] = \mathbb{E}\left[R^2\right]\sin^{\nvDash}\phi = \mathbb{E}\left[R^2\right]\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin^2\theta d\theta =$$

$$\mathbb{E}\left[R^2\right]\left[\frac{1}{2\pi}\int_{-pi}^{\pi}\frac{1}{2}d\theta - \frac{1}{4\pi}\int_{-\pi}^{\pi}\cos(2\theta)d\theta\right] = \frac{1}{2}\mathbb{E}\left[R^2\right] = \sigma^2 \textit{ in 2.4}$$

**Exercise 2.2.** *State if the following time series $(X_t)_{t\in\mathbb{Z}}$ are stationary:*

- $X_t = Z \quad \forall t \in \mathbb{Z} \quad \mathbb{E}\left[Z\right] = 0 \quad \mathbb{E}\left[Z^2\right] = 1$
- $X_t = (-1)^t Z \quad \forall t \in \mathbb{Z} \quad \mathbb{E}\left[Z\right] = 0 \quad \mathbb{E}\left[Z^2\right] = 1$

Now we will see some properties of the covariance function when the time series is stationary.

**Theorem 2.1.** *Consider $(X_t)_{t\in\mathbb{Z}}$ stationary. Then we have that:*

1. $\gamma(0) \geq 0$
2. $\gamma(h) = \gamma(-h) \quad \forall h \in \mathbb{Z}$
3. $|\gamma(h)| \leq \gamma(0) \quad \forall h \in \mathbb{Z}$

*Proof.*

1. $\gamma(0) = cov(X-t, X_t) = Var(X_t) \geq 0 \quad \forall t \in \mathbb{Z}$
2. $\gamma(s,t) = \gamma(t,s) \implies \gamma(0, s-t) = \gamma(0, t-s) \implies \gamma(\pm h) = \gamma(\mp h)$
3. $|\rho(h)| \leq 1 \Leftrightarrow |\gamma(s,t)| \leq \sqrt{Var(X_t)Var(X_s)} = \gamma(0) \implies |\gamma(h)| \leq \gamma(0)$

□

Of course, all these properties are mirrored by the correlation function (always on stationary time series):

**Corollary 2.2.** *Consider $(X_t)_{t\in\mathbb{Z}}$ stationary. Then we have that:*

- $\rho(0) = 1$
- $\rho(h) = \rho(-h) \quad \forall h \in \mathbb{Z}$
- $|\rho(h)| \leq 1 \quad \forall h \in \mathbb{Z}$

*Note that in this case we know the value of the correlation function in 0.*

**Exercise 2.3.** *State if $(X_t)_{t\in\mathbb{Z}}$ is stationary:*

$$X_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1}) \quad \forall t \in \mathbb{Z}$$

*where $(W_t)_{t\in\mathbb{Z}}$ are iid and $\mathbb{E}\left[W_t\right] = 0$ and $\mathbb{E}\left[W_t^2\right] = \sigma^2$ $\forall t \in \mathbb{Z}$.*

One more property of a stationary time series is to have a covariance function which is a non-negative definite function. Let us first define a non-negative definite function:

**Definition 2.10.** *A function $R : \mathbb{Z} \mapsto \mathbb{R}$ is said to be **non-negative definite** iif:*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i R(t_i - t_j) a_j \geq 0$$

$\forall n \in \mathbb{N}, \forall \boldsymbol{a} = (a_1, ..., a_n) \in \mathbb{R}^n, \forall \boldsymbol{t} = (t_1, ..., t_n) \in \mathbb{Z}^n$.

**Theorem 2.2.** *Consider $\gamma : \mathbb{Z} \mapsto \mathbb{R}$ be an ACF of a stationary time series. Then:*

1. $\gamma(h) = \gamma(-h) \quad \forall h \in \mathbb{Z}$
2. $\gamma()$ *is non-negative definite*

*Proof.* Let us start from the $\Longrightarrow$: $\gamma()$ is an ACF of a stationary time series, implying that *1.* is true. Now we have to prove that $\gamma$ is non-negative definite. Consider $\boldsymbol{X}_{\boldsymbol{t}}^{(n)} = (X_{t_1 - \mu}, ..., X_{t_n - \mu})$ with $\mu = \mathbb{E}\left[X_t\right]$. Observe that $\mathbb{E}\left[\boldsymbol{X}_{\boldsymbol{t}}^{(n)}\right] = 0$. Consider now a vector $\boldsymbol{a} \in \mathbb{R}^n$:

$$\begin{aligned}
Var(\boldsymbol{a}\boldsymbol{X}_{\boldsymbol{t}}^{(n)\intercal}) &= \mathbb{E}\left[\boldsymbol{a}\boldsymbol{X}_{\boldsymbol{t}}^{(n)\intercal}\boldsymbol{a}\boldsymbol{X}_{\boldsymbol{t}}^{(n)\intercal}\right] \\
&= \mathbb{E}\left[\boldsymbol{a}\boldsymbol{X}_{\boldsymbol{t}}^{(n)\intercal}\sum_{i=1}^{n} a_i(X_{t_i} - \mu)\right] \\
&= \mathbb{E}\left[\boldsymbol{a}\boldsymbol{X}_{\boldsymbol{t}}^{(n)\intercal}\boldsymbol{X}_{\boldsymbol{t}}^{(n)}\boldsymbol{a}^\intercal\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbb{E}\left[(X_{t_1} - \mu)(X_{t_j} - \mu)\right] a_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(t_i, t_j) a_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(t_i - t_j) a_j
\end{aligned}$$

So the covariance function $\gamma$ is a non-negative definite function because of the arbitrary vector $\boldsymbol{a} \in \mathbb{R}^n$ that we have chosen. No we will prove the $\Longleftarrow$ part. Remember the **Kolmogrov extension theorem**: suppose to have a system of finite-dimensional distributions $\mathcal{P} = \{\pi_{t_1,...,t_n} | t_1, ..., t_n \in T, n \in \mathcal{N}\}$ (a set of measures, usually $T \subset \mathbb{R}$, where $\pi_{t_1,...,t_n}$ are probability measures on the measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$ $\forall n \in \mathbb{N}$ $\forall t_1, ..., t_n \in T$). Now suppose this system consistent, which means that:

$$\pi_{t_1,...,t_n}(B_1 \times ... \times B_n) = \pi_{t_1,...,t_n,t_{n+1},...,t_{n+m}}(B_1 \times ... \times B_n \times m\mathbb{R})$$

$\forall B_1, ..., B_n \in \mathcal{B}(\mathbb{R}),\ m \in \mathbb{N}$ and:

$$\pi_{t_1,...,t_n}(B_1 \times ... \times B_n) = \pi_{\sigma(t_1),...,\sigma(t_n)}(B_{\sigma(1)} \times ... \times B_{\sigma(n)})$$

$\forall B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$, $\sigma$ permutation of $\{1, ..., n\}$. The Kolmogorov theorem states that there exist a probabilistic space $(\Omega, \mathcal{H}, \mathbb{P})$ and a stocastic process $(X_t)_{t\in\mathbb{Z}}$ such that the family $\mathbb{P}$ is the system of finite distribution of $X$, that is:

$$\mathbb{P}_{X_{t_1},...,X_{t_n}}(B) = \pi_{t_1,...,t_n}(B) \quad \forall B \in \mathcal{B}(\mathbb{R})^n,\ n \in \mathbb{N}$$

For example:

- $n = 2 \implies \pi_{t_1}(B_1) = \pi_{t_1,t_2}(B_1 \times \mathbb{R})$
- $n = 2 \implies \pi_{t_1,t_2}(B_1 \times B_2) = \pi_{t_2,t_1}(B_2 \times B_1)$

So the Kolmogorov theorem states that if the system of finite dimensional distribution satifies the two consistency requirements than there exist a probability space and a stochastic process matching the finite dimensional system. In particular, two results follow from this theorem:

1. A system $\mathcal{P}_N$ of finite-dimensional Gaussian distributions is consisten (beacuse the vector of means and the covariance matrix completely determins all the finite dimenal Guassian distributions)
2. If $(X_t)$ and $(Y_t)$ are Gaussian time series such that $\forall n \in \mathbb{N}$, $\forall t_1, ..., t_n \in T$ and $(X_{t_1}, ..., X_{t_n}) \overset{d}{=} (Y_{t_1}, ..., Y_{t_n})$ then $\mathbb{P}(X_t = X_t,\ \forall t \in T) = 1$.

Now suppose that $\gamma : \mathbb{Z} \mapsto \mathbb{R}$ with *1.* and *2.* true. Seòect $t_1, ..., t_n \in \mathbb{Z}$ and define:

$$(\Gamma(n))_{ij} = \gamma(t_i - t_j) \quad i, j = 1, 2, ..., n$$

where $\Gamma(n) \geq 0$. Consider:

$$\Phi(\boldsymbol{z}) := \exp\left\{-\frac{1}{2}\boldsymbol{z}\Gamma(n)\boldsymbol{z}^\intercal\right\} \quad \boldsymbol{z} \in \mathbb{R}^n$$

which is the characteristic function of $\mathcal{N}(0, \Gamma(n))$. Remember that this function uniquely determines the distribution of a function. Denote with $\mathbb{N}(0, \Gamma(n))$ the distribution corresponding to $\Phi$ and denote with $\mathcal{P}_N$ the system of finite dimensional distribution. Then the Kolmogorov extension theorem states that:

$$\exists(\Omega, \mathcal{H}, \mathbb{P}),\ X = (X_t)_{t\in\mathbb{Z}}$$

having $\mathcal{P}_N$ as system of finite distributions. Now suppose to fix $n \in \mathbb{N}$, $t_1, ..., t_n \in \mathbb{Z}$; then:

- $X_t \in \mathcal{L}^2 \quad \forall t \in \{t_1, ..., t_n\}$
- $\mathbb{E}[X_t] = 0 \quad \forall t \in \{t_1, ..., t_n\}$

Whe have to check that $cov(X_{t_i}, X_{t_j}) = function(t_i - t_j)\ \forall t_i, t_j \in \{t_1, ..., t_n\}$. Observe that:

$$\mathbb{E}\left[e^{i(z_i X_{t_i} + z_j X_{t_j})}\right] = \lim_{\boldsymbol{z}(i,j)\to 0} \mathbb{E}\left[e^i(z_1 X_{t_1} + ... + z_n X_{t_n})\right]$$

where $\boldsymbol{z}(i, j) = n-2$-dimensional vector obtained from $\boldsymbol{z} = (z_1, ..., z_n)$ deleting the $i$-th and $j$-th components. We can now write that:

$$\mathbb{E}\left[e^{i(z_i X_{t_i} + z_j X_{t_j})}\right] = \exp\left\{-\frac{1}{2}(z_i, z_j)\Gamma^{(i,j)}(n)\begin{pmatrix} z_i \\ z_j \end{pmatrix}\right\}$$

where

$$\Gamma^{(i,j)}(n) = \begin{pmatrix} \gamma(0) & \gamma(t_i - t_j) \\ \gamma(t_i - t_j) & \gamma(0) \end{pmatrix}$$

Now we are ready to conclude the proof of this theorem, since $\gamma(t_i - t_j) = cov(X_{t_i}, X_{t_j})$. □

**Exercise 2.4.** *Suppose $\gamma(h) = h \ \forall h \in \mathbb{Z}$. Is $\gamma$ an ACF?*

**Exercise 2.5.**

- *Consider $\gamma(h) = \cos(h), \ h \in \mathbb{Z}$. Is $\gamma(h)$ a ACF?*
- *Find $(X_h)_{h \in \mathbb{Z}}$ having $\gamma(h)$ as ACF.*

# 3 Lecture 3

**Existence of a time series having a fixed autocovariance function. Strong and weak stationary time series, IID sequence, q-dependent time series, white noise, gaussian white noise. Simulation of GWN(0,1) and GWN(0,10) and modulated GWN(0,1).**

We have just seen the property referred to as the **weak stationarity property**. Now we will see the strong one; from now on, we will write $(X_t)$ instead of $(X_t)_{t\in\mathbb{Z}}$ to shorten the notation.

**Definition 3.1.** *A time series $(X_t)$ is said to be **strongly stationary** if*

$$(X_{t_1}, X_{t_2}, ..., X_{t_k}) \overset{d}{=} (X_{t_1+h}, X_{t_2+h}, ..., X_{t_k+h})$$

$\forall k \in \mathbb{N},\ h \in \mathbb{Z}$ *and* $t_1, ..., t_n \in \mathbb{Z}$*.*

**Example 3.1.** *A time series $(X_t)$ of iid random variables is strong stationary. To prove this statement, fix $k \in \mathbb{N}$ and $t_1, ..., t_k \in \mathbb{Z}$ such that*

$$\begin{aligned}
\mathbb{P}\left(X_{t_1} \le x_1, ..., X_{t_k} \le x_k\right) &\overset{IND}{=} \prod_{i=1}^{k} \mathbb{P}_{X_{t_i}}(-\infty, x_i] \\
&= \prod_{i=1}^{k} \mathbb{P}(-\infty, x_i) \\
&= \prod_{i=1}^{k} \mathbb{P}_{t_i+h}(-\infty, x_i] \\
&\overset{IND}{=} \mathbb{P}\left(X_{t_1+h} \le x_1, X_{t_2+h} \le x_2, ..., X_{t_k+h} \le x_k\right)
\end{aligned}$$

*for $h \in \mathbb{Z}$.*

Now let us see a little R example on these topics:

```
####################################
# Generating Gaussian time series #
####################################

# generate a sample path of values from N(0,1):
# set the seed to 154
set.seed(154)
w=rnorm(1000,0,1)
help('rnorm')
# two plots in the same window
par(mar=c(2,2,2,2))
par(mfrow=c(2,1))
# plot of w: points on the window
plot(w,main='A path of gaussian time series N(0,1)')
# a red line to show where the zero mean is located
abline(h=0,col=c('red'),lwd=3)
# plot.ts to have a line trought the dots
```

```
plot.ts(w,type='o',main='The same path')
abline(h=0,col=c('red'),lwd=3)
# generate a sample path of 1000 values from N(0,10)
w1=rnorm(1000,0,sqrt(10))
# plot.ts to have a line trought the dots
plot.ts(w,type='o',main='A path of gaussian time series N(0,1)')
abline(h=0,col=c('red'),lwd=3)
# plot.ts to have a line trought the dots
plot.ts(w1,type='o',main='A path of gaussian time series N(0,10)')
abline(h=0,col=c('red'),lwd=3)
# Be careful to y-axis.
plot.ts(w,type='o',main='A path of gaussian time series N(0,1)',
    ylim=c(-10,10))
abline(h=0,col=c('red'),lwd=3)
plot.ts(w1,type='o',main='A path of gaussian time series N(0,10)',
    ylim=c(-10,10))
abline(h=0,col=c('red'),lwd=3)
```

Strong stationarity is very difficult to achieve, so a weakened version is often more useful.

**Definition 3.2.** *A time series $(X_t)$ is said to be* ***strong stationary of order*** *$k$ if*

$$(X_{t_1}, X_{t_2}, ..., X_{t_k}) \overset{d}{=} (X_{t_1+h}, X_{t_2+h}, ..., X_{t_k+h})$$

*for a fixed $k \in \mathbb{N}$, $\forall t_1, ..., t_k \in \mathbb{Z}$ and $h \in \mathbb{Z}$.*

**Example 3.2.** *Plot a path of $X_t = W_t \sin(2\pi\omega t)$ with $\omega = \frac{1}{500}$, $t = 1, ..., 1000$ and $(W_t)$ iid $\sim \mathcal{N}(0, 1)$.*

```
################################################
# Generating a modulated Gaussian time series #
################################################

# generate a sample path of values from N(0,1): set the seed to 154,
# so we reproduce the same results
set.seed(154)
w=rnorm(1000,0,1)
#
plot.ts(w*sin(2*pi*(1:1000)/500),main='Gaussian time series N(0,1)
    modulated by a sin', ylab='X_t')
lines(sin(2*pi*(1:1000)/500),col=c('red'),lwd=3)
```

So, what is the relationship between the strong and weak stationary time series? Despite their name, the strong stationary property does not always imply the weak one. Let us see an example.

**Example 3.3.** *Consider a time series $(X_t)$ made up of iid random variables with Cauchy distribution:*

$$f(x) = \frac{1}{\pi(1 + x^2)} \quad x \in \mathbb{R}$$

*then $(X_t)$ is obiovsly strong stationary, but it is not weak stationary since $X_t \notin \mathcal{L}^2$.*

**Theorem 3.1.** *Suppose $(X_t)$ strong stationary and $(X_t) \in \mathcal{L}^2$ $\forall t \in \mathbb{Z}$. Then $(X_t)$ is weak stationary.*

*Proof.* $X_t \in \mathcal{L}^2$ $\forall t \in \mathbb{Z} \implies \mu_t = \mathbb{E}[X_t] < \infty$ $\forall t \in \mathbb{Z}$. Then, from the strong stationarity it follows that $X_t \overset{d}{=} X_{t+h}$ $\forall t, h \in \mathbb{Z} \implies \mu_t = \mu_{t+h} \implies \mathbb{E}[X_t] = const$. Now, considering the Cauchy-Schwarts inequality we have that $\mathbb{E}[X_t X_s] < \infty$ $\forall t, s \in \mathbb{Z}$, so:

$$\begin{aligned}\mathbb{E}[X_t X_s] &= \int_{\mathbb{R}^2} xy \mathbb{P}_{X_s X_t}(dx, dy) \\ &= \int_{\mathbb{R}^2} xy \mathbb{P}_{X_{s+h} X_{t+h}}(dx, dy) \\ &= \mathbb{E}[X_{s+h} X_{t+h}]\end{aligned}$$

Consider than the covariance:

$$\begin{aligned}cov(X_s, X_t) &= \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)] \\ &= \mathbb{E}[X_s X_t] - \mu_s \mu_t \\ &= \mathbb{E}[X_{s+h} X_{t+h}] - \mu_{s+h} \mu_{t+h} \\ &= cov(X_{s+h}, X_{t+h})\end{aligned}$$

implying that $cov(X_s, X_t) = cov(X_{s+h}, X_{t+h}) = \gamma(0, h)$ $\forall h \in \mathbb{Z}$. Then, since the covariance depends on the lag window we can conclude that the time series is weak stationary. □

Of course, also the weak stationarity property does not imply the strong one.

**Exercise 3.1.** *Consider $(X_t)_{t \in \mathbb{Z}}$ iid random variables $\sim \mathcal{N}(0, 1)$. Define:*

$$X_t = \begin{cases} Z_t & t \text{ odd} \\ \frac{Z_{t-1}^2 - 1}{\sqrt{2}} & t \text{ even} \end{cases}$$

- *Is $(X_t)$ strong stationary?*
- *Is $(X_t)$ weak stationary?*

**Definition 3.3.** *A time series $(X_t)$ is said to be **white noise** if*

- *$(X_t)$ is stationary*
- $\mathbb{E}[X_t] = 0 \quad \forall t \in \mathbb{Z}$
- $\gamma(0) = \sigma^2, \ \gamma(h) = 0 \quad \forall h \in \mathbb{Z} - \{0\}$

*Notation:*

- *white noise: $(X_t) \sim \mathcal{WN}(0, \sigma^2)$*
- *iid random variables: $(X_t) \sim \mathcal{IID}(0, \sigma^2)$*

- *Gaussian white noise:* $(X_t) \sim \mathcal{GWN}(0, \sigma^2) \implies X_t \sim \mathcal{N}(0, \sigma^2)$

**Remark 3.1.** *While $\mathcal{IID}(0, \sigma^2) \implies \mathcal{WN}(0, \sigma^2)$, the other way is not always true. It also true that $\mathcal{GWN}(0, \sigma^2) \implies \mathcal{IID}(0, \sigma^2)$ since:*

$$\mathbf{\Sigma}_n = diag(\sigma^2, ..., \sigma^2)$$

*implying that:*

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\mathbf{\Sigma}_n|}} \exp\left\{ -\frac{1}{2} \boldsymbol{x} \mathbf{\Sigma}_n^{-1} \boldsymbol{x}^\intercal \right\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{(\sigma^2)^n}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n \frac{x_i^2}{\sigma^2} \right\} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \prod_{i=1}^n \exp\left\{ -\frac{1}{2} \frac{x_i^2}{\sigma^2} \right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \frac{x_i^2}{\sigma^2} \right\} \end{aligned}$$

*This gives independence among the random variables of the series.*

Since the strong and weak stationarity are very strong assumptions, seldom encountered in real situations, we can relax a little the requirements.

**Definition 3.4.** *A time series $(X_t)$ is said to be **q-dependet** if the random variable $X_t$ and $X_s$ are independent whenever $|t - s| > q$, with $q \in \mathbb{N}$.*

**Example 3.4.** *A $\mathcal{IID}$ time series is 0-dependant.*

**Proposition 3.1.** *Consider two time series $(Y_t) \sim \mathcal{IID}$ and $(X_t)$ such that $X_t = g(Y_t, ..., Y_{t-q})$ $\forall t \in \mathbb{Z}$ with $g$ being a measurable function.*

1. *$(X_t)$ is q-dependent;*
2. *$(X_t)$ is strong stationary.*

*Proof.* 1. Set $t = s + h$ and $h > q$. Consider

$$\begin{gathered} Y_{s-q}, ..., Y_{s-1}, Y_s, ..., Y_{s+h-q}, ..., Y_{s+h-1}, Y_{s+h} \\ g(Y_s, ..., Y_{s-q}) \rightarrow IND \leftarrow g(Y_{s+h}, Y_{s+h-q}) \\ X_s \rightarrow IND \leftarrow X_{s+h} \end{gathered}$$

$\forall s \in \mathbb{Z}$.

2. Fix $n \in \mathbb{N}$, $t_1, ..., t_n \in \mathbb{Z}$:

$$(Y_{t_1}, ..., Y_{t_1-q}) \overset{d}{=} (Y_{t_1+h}, ..., Y_{t_1+h-q})$$

$$(Y_{t_n}, ..., Y_{t_n-q}) \overset{d}{=} (Y_{t_n+h}, ..., Y_{t_n+h-q})$$

$$\begin{aligned} (X_{t_1}, ..., X_{t_n}) &= (g(Y_{t_1}, ..., Y_{t_1-q}), ..., g(Y_{t-n}, ..., Y_{t_n-q})) \\ &\overset{d}{=} (g(Y_{t_1+h}, ..., Y_{t_1+h-q}), ..., g(Y_{t-n+h}, ..., Y_{t_n+h-q})) \\ &= (X_{t_1+h, ..., X_{t_n+h}}) \end{aligned}$$

$\forall h \in \mathbb{Z}$. □

**Remark 3.2.** $(Y_t) \in \mathcal{L}^2 \;\not\Longrightarrow\; (X_t) \in \mathcal{L}^2$

**Example 3.5.** $X_1, X_2 \sim \mathcal{C} \implies X_1, X_2 \notin \mathcal{L}^2,\; \frac{X_1}{X_2} \sim \mathcal{N}(0,1) \in \mathcal{L}^2$

A special case of the $g$ function is when it is a polinomia function ($g(x_0, ..., x_q) = x_0 + \phi_1 x_1 + ... + \phi_q x_q$).

**Example 3.6.** *A very popular time series is $(W_t) \sim \mathcal{IID}(0, \sigma^2)$ with $(X_t)$ such that:*

$$X_t = W_t + \phi_1 W_{t-1} + ... + \phi_q W_{t-q} \quad t \in \mathbb{Z},\; q \in \mathbb{N}$$

*then $(X_t)$ is q-dependant and strong stationary. This property is frequently exploited to build strong stationary time series.*

# 4 Lecture 4

**Generation and study in R of simulated time series. Time average and almost sure convergence for strong stationary t.s. Ergodic time series and tail sigma-algebra. The employment of the Kolmogorov law 0-1. Strong ergodic theorem. Examples and exercises**

We now introduce a property similar to the q-dependence, the q-correlation.

**Definition 4.1.** *A stationary time series $(X_t)$ is said to be* ***q-correlated*** *if*

$$\gamma(h) = 0 \quad for \ \ |h| > q$$

**Example 4.1.** $(W_t) \sim \mathcal{WN} \implies (W_t)$ *0-correlated.*

**Exercise 4.1.** *Consider a stationary time series $(X_t)$ and $(Y_t)$ such that*

$$Y_t = \begin{cases} X_t & t \ odd \\ X_{t+1} & t \ even \end{cases}$$

*Check if $(Y_t)$ is stationary.*

**Proposition 4.1.** *Consider $(W_t) \sim \mathcal{WN}(0, \sigma^2)$ and $(X_t)$ such that*

$$X_t = W_t + \phi_1 W_{t-1} + ... + \phi_q W_{t-q} \quad t \in \mathbb{Z}$$

*where $q \in \mathbb{N}$ and $\phi_1, ..., \phi_q \in \mathbb{R}$. Then $(X_t)$ is weak stationary and q-correlated.*

**Remark 4.1.** *$(X_t)$ is a* ***moving average*** *time series of order q.*

*Proof.*

$$X_t \in \mathcal{L}^2 \ \forall t \in \mathbb{Z} \ \Leftarrow W_t \in \mathcal{L}^2 \ \forall t \in \mathbb{Z}$$

$$\mathbb{E}\left[X_t\right] = 0 \ \forall t \in \mathbb{Z} \ \Leftarrow \mathbb{E}\left[X_t\right] = \sum_{j=0}^{q} \phi_j \mathbb{E}\left[W_{t-j}\right] = 0 \quad (\phi_0 = 1)$$

The next step would be the calculation of $cov(X_s, X_{s+h})$. This is left as an exercise, but the solution is provided:

$$cov(X_s, X_{s+h}) = \begin{cases} 0 & |h| > q \\ \sum_{j=0}^{q-|h|} \phi_j \phi_{j+|h|} & |h| \leq q \end{cases}$$

We observe that $(X_t)$ as a covariance function depending only on tha lag window $h$, so it is a weak statinary time series; moreover we have proved that the covariance function is 0 when $h > q$, then $(X_t)$ is a q-correlated time series. □

**Example 4.2.** *Simulate (plot) a path of 500 steps of a time series*

$$X_t = \begin{cases} Z_t & t \ odd \\ \frac{Z_{t-1} - 1}{\sqrt{2}} & t \ even \end{cases}$$

*where $(Z_t) \sim \mathcal{GWN}(0, 1)$ (seed=100). Compare the two plots of $(X_t)$ and $(Z_t)$.*

*The solution is the following:*

```
#######################################
# Generation of a WN(0,1)
#######################################
# Generate a sample path of GWN(0,1) with seed 100
set.seed(100)
w=rnorm(500,0,1)
# initialize the vector x with w
x=w
# change the elements in x corresponding to even steps.
for(i in 1:250) {x[2*i]=(x[2*i-1]^2-1)/sqrt(2)}
# set up the graphic window
par(mar=c(2,2,2,2))
par(mfrow=c(2,1))
# difference between plot and plot.ts
plot(w, main=’A path of gaussian white noise’)
plot.ts(w, main=’A path of gaussian white noise’)
# comparing plots
plot.ts(x, type=’o’, main=’A path of a white noise’)
plot.ts(w, type=’o’, main=’A path of gaussian white noise’)
# the two plots do not have the same axes. To set the same axes:
plot.ts(x, main=’A path of white noise’,ylim=range(-3,6))
plot.ts(w, main=’A path of gaussian white noise’, ylim=range(-3,6))
```

**Example 4.3.** *Simulate a path of a random walk*

$$X_t = \mu + X_{t-1} + W_t \quad t = 1, ..., 200$$

*with* $W_t \sim \mathcal{GWN}(0,1),\ X_0 = 0,\ \mu = 0.2.$

1. *compare with* $\mu = 0$*;*

2. *for* $\mu = 0.2$ *fit the trend with a line.*

*The solution is the following:*

```
#########################################
# Generation of paths of a random walk
#########################################
# Generate 200 values from N(0,1) with seed 154
set.seed(154)
w=rnorm(200,0,1)
# do the cumulative summation of the 200 values
x = cumsum(w)
# set the times (1:200)
times=seq(1,200,1)
# set mu=0.2: xd <-random walk
mu = 0.2
xd = mu*times + x
# set up the graphic window
par(mar=c(2,2,2,2))
# Do the plot
```

```
plot.ts(xd, main=’On the random walk’, type=’o’)
# To add the path generated from the random walk with no drift
lines(x, col=’red’)
# To estimate the trend component of xd
fit=lm(xd˜times)
# To validate the estimation
summary(fit)
# To add the regression line to the plot
(coeff=fit$coefficients)
lines(coeff[2]*times+coeff[1], col=’blue’, lty="dashed")
```

Now we will explore briefly the **Ergodic theory**, which gives results on asymptotic behaviors of systems evolving over time. It can allow us to work with the classical concepts of statistics, like sample mean and sample variance, on time series.

**Definition 4.2.** *An* ***ergodic theorem*** *gives condition under which*

$$\bar{X}_n = \frac{1}{n}(X_1 + ... + X_n)$$

*converges as n becomes larger.*

In the classical framework, $\bar{X}$ is the sample mean, since all the variables are iid, while in the time series context this is not true anymore; so the value $\bar{X}$ is called **time average**. Note that also the strong law of large numbers is an ergodic theorem. Note also that stationary time series have a constant mean $\mu$. If we have more time series in the same sample space, the mean of all the sample means gives the so-called **ensemble average**.

**Example 4.4.** *Consider a strong stationary time series* $(X_t) \implies (X_t) \sim \mathcal{ID}$ *with* $x_1, ..., x_n \in range(X_t)$*. These information are sufficient to estimate* $\mu$ *from* $\bar{X}_n$*?*

*Consider two coins; the first one is fair* ($\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$) *and the second one is not* ($\mathbb{P}(H) = 1,\ \mathbb{P}(T) = 0$)*. Choose a coin randomly and then flip it infinitely times. The sample space of this experiment is* $\Omega = \{Coin1, Coin2\} \times \{H, T\}^{\mathbb{N}}$*. Now partition* $\Omega = \{A, B\}$ *as:*

$$\begin{aligned} A &= (Coin1, \omega_1, \omega_2, ...) \in \Omega \\ B &= (Coin2, \omega_1, \omega_2, ...) \in \Omega = (Coin2, H, H, H, ...) \end{aligned}$$

*We have that* $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$ *due to the random selection. Now consider the coordinate random variables*

$$X_n(\omega) = \begin{cases} 1 & \textit{if the n-th toss gives head} \\ 0 & \textit{if the n-th toss gives tail} \end{cases}$$

*with* $\omega \in \Omega$*. Let us study this sequence of random variables.*

*First of all,* $(X_n) \sim \mathcal{ID}$*, since*

$$\begin{aligned} &\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 1|A)\,\mathbb{P}(A) + \mathbb{P}(X_n = 1|B)\,\mathbb{P}(B) = \frac{3}{4} \\ &\mathbb{P}(X_n = 0) = 1 - \mathbb{P}(X_n = 1) = \frac{1}{4} \end{aligned}$$

*As you can see, these distributions do not depend on $n$, so all the random variables $X_n$ have the same distribution; in particular, the mean is*

$$\mathbb{E}\left[X_n\right] = 1\frac{3}{4} + 0\frac{1}{4} = \frac{3}{4}$$

*The random variables $(X_n)$ are not independent: indeed, if*

$$X_n = \begin{cases} \mathcal{B}e(1) & \textit{if B is selected} \\ \mathcal{B}e(\frac{1}{2}) & \textit{if A is selected} \end{cases} \implies X_n \sim \mathcal{B}e(P)$$

*with $P$ random variable such that $\mathbb{P}\left(P = 1\right) = \mathbb{P}\left(P = 0.5\right) = \frac{1}{2}$. Since random variables $X_n$ depend on the same random variable $P$ they are dependent.*

*The time series $(X_n)$ is strong stationary: if $n, m \in \mathbb{N}$:*

$$\mathbb{P}\left(X_n = 0, X_m = 0\right) = \mathbb{P}\left(X_n = 0, X_m = 0|A\right)\mathbb{P}\left(A\right) + \mathbb{P}\left(X_n = 0, X_m = 0|B\right)\mathbb{P}\left(B\right) = \frac{1}{8}$$

$$\mathbb{P}\left(X_n = 0, X_m = 1\right) = \mathbb{P}\left(X_n = 1, X_m = 0\right) = \frac{1}{8}$$

$$\mathbb{P}\left(X_n = 1, X_m = 1\right) = \mathbb{P}\left(X_n = 1, X_m = 1|A\right)\mathbb{P}\left(A\right) + \mathbb{P}\left(X_n = 1, X_m = 1|B\right)\mathbb{P}\left(B\right) = \frac{5}{8}$$

*Also, these distributions do not depend on $n, m$, making the sequence strong stationary.*

*Let us now compute the covariance:*

$$cov(X_n, X_m) = \mathbb{E}\left[X_n X_m\right] - \mathbb{E}\left[X_n\right]\mathbb{E}\left[X_m\right] = \frac{5}{8} - \frac{9}{16} = \frac{1}{16} \neq 0$$

*Can we state*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = const \quad \forall\omega \in \bar{\Omega} \ s.t.\mathbb{P}\left(\bar{\Omega}\right) = 1$$

*? Consider:*

$$\tilde{\omega} = (B, T, T, ...) \implies X_i(\tilde{\omega}) = 1 \implies \frac{1}{n}\sum_{I=1}^{n} X_i(\tilde{\omega}) = 1$$

$$\omega^* = (A, \omega) \ with \ \omega \in \{H, T\}^{\mathbb{N}} \implies X_i(\omega^*) = X_i^*(\omega)$$

*i-th coordinate random variable of Bernulli trials. Computing the limit:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega^*) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^*(\omega) \overset{a.s.}{=} \frac{1}{2}$$

*by the strong law of large numbers. So no, it is not possible to find a subset of the sample space $\bar{\Omega}$ such that the limit of the time average is equal to a constant. Also, we have proved that the time average does not give information on the mean of a random variable sequence, since the limits have different values from the mean of the sequence.*

*This is an example of a strong stationary time series with constant mean such that the time average does not give information on the overall mean of the process. This is why we need something more.*

**Theorem 4.1.** *If a time series $(X_t)$ is strong stationary and $X_n \in \mathcal{L}^1$ $\forall n$ then*

$$\bar{X}_n \overset{a.s.}{\Longrightarrow} \bar{X} \in \mathcal{L}^1$$

*This is the* ***strong ergodic theorem****, and we will not see the proof.*

The strong ergodic theorem is a generalization of the strong law of large numbers, because we have random variables having the same distribution but are not (necessarily) independent.

**Corollary 4.1.** $\bar{X}_n \overset{\mathcal{L}^1}{\Longrightarrow} \bar{X}$*, allowing us to obtain information on the mean of the sequence by means of the time average.*

**Remark 4.2.** *As* $X_n \sim \mathcal{ID} \implies \mathbb{E}\left[X_n\right] = \mu \; \forall n$*, then*

$$\mathbb{E}\left[\bar{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_n\right] = \mu$$

*therefore, as* $\bar{X}_n \overset{\mathcal{L}^1}{\Longrightarrow} \bar{X}$ *then*

$$\lim_{n\to\infty} \mathbb{E}\left[\bar{X}_n\right] = \mathbb{E}\left[\bar{X}\right] \implies \mathbb{E}\left[\bar{X}\right] = \mu$$

*Therefore, if we add to the strong stationarity property the property that these random variables have finite mean, we have that the time average converges almost surely to a random variable $\bar{X}$ that has finite mean, which is the same as the one in the sequence. Therefore, while we recover the time average from observation, this gives information on the overall mean. How good is this information depends on the variance.*

Now let us go back to the previous example to see if this property is verified.

**Example 4.5.** *If*

$$\tilde{\omega} = (B, H, H, ...) \implies \lim_{n\to\infty} \bar{X}_n(\tilde{\omega}) = 1$$
$$\omega^* \in \{\tilde{\omega}\}^c \implies \lim_{n\to\infty} \bar{X}_n(\omega^*) = \frac{1}{2}$$

*so, consider a random variable $\bar{X}$ such that*

$$\mathbb{P}\left(B\right) = \mathbb{P}\left(\{\tilde{\omega}\}\right) = \mathbb{P}\left(\bar{X} = 1\right) \quad \textit{and} \quad \mathbb{P}\left(\bar{X} = \frac{1}{2}\right) = \frac{1}{2}$$

*Therefore, we have found a random variable $\bar{X}$ such that*

$$\bar{X}_n \overset{a.s.}{\Longrightarrow} \bar{X}$$

*Note that* $\mathbb{E}\left[\bar{X}\right] = 1\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{3}{4} = \mathbb{E}\left[X_n\right]$ *and* $\bar{X} \overset{d}{=} P$.

Since it is rare to work with a strong stationary time series, we will need more tools that do not require this property.

**Definition 4.3.** *Suppose a time series* $(X_t)$ *defined on the probability space* $(\Omega, \mathcal{H}, \mathbb{P})$*. Let us consider the following* $\sigma$*-algebra:*

$$\mathcal{G}_t = \sigma X_t = \left\{X_t^{-1}(A),\ A \in \mathcal{B}(\mathbb{R})\right\}$$

*This* $\sigma$*-algebra contains the information of the system at time* $t$*. To collect all the information after a certain time, we consider*

$$\mathrm{T}_s = \sigma\left(\bigcup_{t>s} \mathcal{G}_t\right) = \bigvee_{t>s} \mathcal{G}_t$$

*This time series contains information about the future of the system after* $s$*.*

*Since* $\{\mathrm{T}_s\}$ *is such that* $\mathrm{T}_s \subseteq \mathrm{T}_{s+1}$ $\forall s$*:*

$$\mathrm{T} = \bigcap_s \mathrm{T}_s$$

*this is the* ***tail*** $\sigma$***-algebra*** *and contains information about the remote future and events that are not influenced by finite events.*

**Example 4.6.** *Suppose to consider the set of all sequences alternating -1 and 1:*

$$\Sigma = \{-1, 1\}^{\mathbb{N}} \quad \textit{with } \mathbb{N} = \{1, 2, ...\}$$
$$(X_n) \textit{ coordinate random variables}$$
$$S_n = \sum_{i=1}^{n} X_i \quad \textit{for } n \geq 1$$
$$H = \{\omega \in \Omega / S_n(\omega) \geq 0\ \forall n \geq 1\}$$

*Is the event* $H \in \mathrm{T}$*?*

*Consider* $\omega_1 = (1, -1, 1, -1, ...) \in \Omega$*. Then*

$$S_1(\omega_1) = 1,\ S_2(\omega_1) = 0,\ S_3(\omega_1) = 1,\ S_4(\omega_4) = 0,\ ... \implies S_n(\omega_1) \geq 0$$

$\forall n \geq 1$*, so* $\omega_1 \in H$*.*
*Consider now* $\omega_2 = (-1, -1, 1, -1, 1, -1...) \in \Omega$*. Then*

$$S_1(\omega_2) = -1,\ S_2(\omega_2) = -2,\ S_3(\omega_2) = -1,\ S_4(\omega_2) = -2,\ ... \implies S_n(\omega_2) \leq 0$$

$\forall n \geq 1$*, so* $\omega_2 \notin H$*.*
*Note that if an observer looks at the sequence from the third element of* $\omega_2$ *he can believe to watch the sequence* $\omega_1$*, thus believing that* $\omega_2 \in H$*, but this is not true; that is because* $H$ *is not a* ***tail event****. So, if an event* $H \in \mathrm{T}$ *we can state if* $\omega \in H$ *or* $\omega \notin H$*,* $\forall \omega \in \Omega$*.*

**Definition 4.4.** *A time series* $(X_t)$ *is an* ***ergodic time series*** *if* $\forall H \in \mathrm{T}\ \mathbb{P}(H) = 0 \vee 1$*.*

**Remark 4.3.** *(0-1 Kolmogrov law) If the sequence of* $\sigma$*-algebras* $\{\mathcal{G}_t\}$ *are independent then* $\forall H \in \mathrm{T}\ \mathbb{P}(H) = 0 \vee 1$*.*

**Lemma 4.1.** *If the time series* $(X_t)$ *is a sequence of independent random variables,* $(X_t)$ *is ergodic.*

Now let us get back to the coin example.

**Example 4.7.** *Consider the set*

$$H = \left\{ \omega \in \Omega / \lim_{n \to \infty} \bar{X}_m(\omega) = 1 \right\} = \{(B, H, H, ...)\}$$

*we know that* $\mathbb{P}(H) = \frac{1}{2} \neq 1$, *therefore* $H \notin \tau$.

**Exercise 4.2.** *State if the following time series are ergodic:*

1. $\mathcal{IID}(0, \sigma^2)$
2. $\mathcal{WN}(0, \sigma^2)$ *of independent random variables (notation* $\mathcal{IWN}(0, \sigma^2)$*)*
3. $\mathcal{GWN}(0, \sigma^2)$
4. $\mathcal{IID}$ *with standard Cauchy distribution* $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$
5. $X_t = U \sim \mathcal{B}e(p)$ *with* $p \in (0, 1)$

# 5 Lecture 5

**Elements of ergodic theory**

There is one more property of time series that implies ergodicity that sometimes is easier to verify than the previous property that we have seen; this is the **mixing** property. For stochastic processes, mixing means asymptotically independence between the random variables of the sequence. This requires an explicit way to measure the dependence between two different random variables, and we have no time to deal with this. Instead, we will focus on properties similar to the strong law of large numbers.

**Definition 5.1.** *A time series $(X_t)$ has the* ***mean-ergodic*** *property (almost surely) if*

$$\exists \mu \in \mathbb{R} \text{ } s.t. \text{ } \bar{X}_n \overset{a.s.}{\Longrightarrow} \mu$$

Suppose that we have a strong stationary time series $(X_t)$ with $X_t \in \mathcal{L}^1$ $\forall t$. Then $\bar{X}_n \overset{a.s.}{\Longrightarrow} \bar{X} \in \mathcal{L}^1$ by the strong ergodic theorem. But if we add that the time series is also ergodic we recover the mean-ergodic property in the almost sure way: $\bar{X}_n \overset{a.s}{\Longrightarrow} \mu = \mathbb{E}\left[X_t\right]$. We will now formalize and prove this result.

**Theorem 5.1.** ***(Birrhoff ergodic theorem)*** *Consider a time series $(X_t)$ which is:*

- *strong stationary;*
- *$X_t \in \mathcal{L}^1$, $\forall t$;*
- *ergodic;*

*then*

$$\bar{X}_n \overset{a.s.}{\Longrightarrow} \mu = \mathbb{E}\left[X_t\right]$$

*Proof.* As the time series is strong stationary and has finite mean, we can apply the strong ergodic theorem:

$$\bar{X}_n \overset{a.s.}{\Longrightarrow} \bar{X} \in \mathcal{L}^1$$

Consider

$$\begin{aligned} A &= \left\{\omega \in \Omega / \bar{X} \leq a\right\}, \ a \in \mathbb{R} \\ &= \left\{\omega \in \Omega / \lim_n \bar{X} \leq a\right\} \in \mathrm{T} \end{aligned}$$

As the time series is ergodic by hypotesis $\mathbb{P}(A) = 0 \vee \mathbb{P}(A) = 1$. Since $\mathbb{P}\left(\bar{X} \leq a\right) = F_{\bar{X}}(a)$, we have that $F_{\bar{X}}()$ is a step function. This means that the random variable $\bar{X} \overset{a.s.}{=} const$, but which constant? As $\bar{X} \overset{\mathcal{L}^1}{\Longrightarrow} \bar{X}$, from the corollary of the strong ergodic theorem, we have that $\mathbb{P}\left(\bar{X}\right) = \mu$, which is the constant we were searching for. Therefore $\bar{X} \overset{a.s.}{\Longrightarrow} \mu$. □

**Exercise 5.1.** *Consider the time series $(X_t)$ with $X_t = \mu t + Wt$ with $\mu > 0$ and $(X_t) \sim \mathcal{IID}(0, \sigma^2)$. Check if $X_t$ has the mean-ergodic property (a.s.).*

Now we will see some results on ergodic time series without proving them.

**Theorem 5.2.** *Consider a time series* $(W_t)$ *such that:*

- $(X_t) \sim \mathcal{IID}$*;*
- $f : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}$ *measurable;*
- $X_t \overset{a.s.}{=} f(W_s,\ s \leq t)\ \forall t \in \mathbb{Z}$*;*

*then* $(X_t)$ *is ergodic.*

**Theorem 5.3.** *Consider a time series* $(W_t)$ *such that:*

- $(W_t)$ *is strong stationary;*
- $(W_t)$ *is ergodic;*
- $f : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}$ *measurable;*
- $X_t \overset{a.s.}{=} f(W_s,\ s \geq t)\ \forall t \in \mathbb{Z}$*;*

*then* $(X_t)$ *is ergodic.*

**Theorem 5.4.** *Consider a strong stationary time series* $(X_t)$*. Then*

$$(X_t) \iff \forall k \geq 1,\ \forall A \in \mathcal{B}(\mathbb{R}^k)\ \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(X_j,\dots,X_{j+k})\in A} = \mathbb{P}\left((X_0,\dots,X_k) \in A\right)$$

*is sufficient and necessary for* $(X_t)$ *to be ergodic.*

**Definition 5.2.** *A time series* $(X_t)$ *as the mean-ergodic property in the* $\mathcal{L}^2$ *sense if*

$$\exists \mu \in \mathbb{R}\ s.t.\ \bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu$$

**Remark 5.1.** *The constant* $\mu$ *could also not be the mean of the process; this depends on the process.*

**Remark 5.2.** $\bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu \iff \lim_{n\to\infty} \mathbb{E}\left[(\bar{X}_n - \mu)^2\right] = 0$

**Remark 5.3.** *When* $\mu = \mathbb{E}\left[\bar{X}_n\right]$ *then* $\bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu \iff \lim_{\mu\to\infty} Var(\bar{X}_n) = 0$

**Exercise 5.2.** *Consider a random walk* $(X_t)$ *such that*

$$X_t = \begin{cases} 0 & t \leq 0 \\ X_{t-1} + W_t & t \geq 1 \end{cases}$$

*with* $(W_t) \sim \mathcal{IID}(0, \sigma^2)$*. Check if* $(X_t)$ *has the mean-ergodic property in* $\mathcal{L}^2$*.*

**Theorem 5.5.** *Consider a time series* $(X_t)$ *such that:*

- $(X_t)$ *stationary;*
- $\lim_{h\to\infty} \gamma(h) = 0$*;*

*then* $(X_t) \overset{\mathcal{L}^2}{\Longrightarrow} \mu = \mathbb{E}[X_t]$*.*

*Proof.* Start by calculate the variance of the sequence:

$$
\begin{aligned}
Var(\bar{X}_n) &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\left(\sum_{t=1}^{n} X_t - n\mu\right)\right]^2 \\
&= \mathbb{E}\left[\frac{1}{n^2}\left(\sum_{t=1}^{n}(X_t - \mu)\right)^2\right] \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathbb{E}\left[(X_t-\mu)(X_s-\mu)\right] \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} cov(X_t, X_s) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma(t-s) \\
&= \frac{1}{n^2}\sum_{h=-(n-1)}^{n-1}(n-|h|)\gamma(h) \\
&= \frac{1}{n}\sum_{h=-(n-1)}^{n-1}\left(1-\frac{|h|}{n}\right)\gamma(h) \\
&= \frac{1}{n}\sum_{|h|<n}\left(1-\frac{|h|}{n}\right)\gamma(h) \\
&\leq \frac{1}{n}\sum_{|h|<n}|\gamma(h)| \\
&= \frac{\gamma(0)}{n} + \frac{1}{n}\left(\sum_{h=-(n-1)}^{-1}|\gamma(h)| + \sum_{h=1}^{n-1}|\gamma(h)|\right) \\
&= \frac{\gamma(0)}{n} + \frac{1}{n}\left(\sum_{j=(n-1)}^{1}|\gamma(-j)| + \sum_{h=1}^{n-1}|\gamma(h)|\right) \qquad h=-j \\
&= \frac{\gamma(0)}{n} + \frac{1}{n}\left(\sum_{j=1}^{n-1}|\gamma(j)| + \sum_{h=1}^{n-1}|\gamma(h)|\right) \\
&= \frac{\gamma(0)}{n} + \frac{2}{n}\sum_{h=1}^{n-1}|\gamma(h)| \\
&= \frac{\gamma(0)}{n} - 2\frac{|\gamma(n)|}{n} + 2\frac{|\gamma(n)|}{n} + \frac{2}{n}\sum_{h=1}^{n-1}|\gamma(h)| \\
&= \frac{\gamma(0)}{n} - 2\frac{|\gamma(n)|}{n} + \frac{2}{n}\sum_{h=1}^{n}|\gamma(h)|
\end{aligned}
$$

Therefore

$$\lim_{n\to\infty} Var(\bar{X}_n) = \lim_{n\to\infty} \left( \frac{\gamma(0)}{n} - 2\frac{|\gamma(n)|}{n} + \frac{2}{n}\sum_{h=1}^{n} |\gamma(h)| \right)$$

Thanks to the corollary of Cesaro's lemma, we know that

$$\{b_i\} \ s.t. \ \lim_{i\to\infty} b_i = b \in \mathbb{R} \implies \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} b_i = b$$

Assume that $b_i \Leftarrow |\gamma(h)|$. We know that $\lim_{h\to\infty} = 0 \implies \lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} |\gamma(h)| = 0$. Now we can finally state that

$$\lim_{n\to\infty} Var(\bar{X}_n) = 0$$

ending the proof of this theorem. □

**Corollary 5.1.** *If $(X_t)$ is a stationary time series and $\lim_{h\to\infty}\gamma(h) = 0$ then $\bar{X}_n \overset{\mathbb{P}}{\Longrightarrow} \mu = \mathbb{E}[X_t]$. This means that $\bar{X}_n$ is a consisten estimator of $\mu$.*

**Definition 5.3.** *The ACF is said to be* ***absolutely sommable*** *if*

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < \infty$$

**Theorem 5.6.** *Consider a stationary time series $(X_t)$ with $\gamma()$ absolutely sommable. Then*

$$\lim_{n\to\infty} nVar(\bar{X}_n) = \sum_{h=-\infty}^{+\infty} \gamma(h)$$

*which means that*

$$Var(\bar{X}_n) \sim \frac{1}{n}\sum_{h=-\infty}^{+\infty} \gamma(h)$$

*Proof.* From the previous proof:

$$nVar(\bar{X}_n) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right)\gamma(h) \leq \sum_{|h|<n} |\gamma(h)|$$

Considering now the limit:

$$\lim_{n} nVar(\bar{X}_n) \leq \lim_{n} \sum_{|h|<n} |\gamma(h)| = \sum_{h=-\infty}^{+\infty} |\gamma(h)| < \infty$$

Consider a function:

$$f_n(h) = \begin{cases} \left(1 - \frac{|h|}{n}\right)\gamma(h) & h = 0, \pm 1, \pm 2, ... \pm (n-1) \\ 0 & otherwise \end{cases}$$

Suppose to denote with $m$ a counting measure on $\mathbb{Z}$ such that $m(\{h\}) = 1 \ \forall h \in \mathbb{Z}$. Then

$$\sum_{|h<n|} \left(1 - \frac{|h|}{n}\right)\gamma(h) = \int_{\mathbb{Z}} f_n(h) m \ dh$$

Consider $\{f_n\}$:

- $|f_n| \leq |\gamma| \ \forall m \in \mathbb{N} \Leftarrow \left|1 - \frac{|h|}{n}\right| \leq 1$
- $|\gamma|$ integrable with respect to m $\Leftarrow \int_{\mathbb{Z}} |\gamma(h)| \, m \ dh = \sum_{h=\infty}^{\infty} |\gamma(h)| < \infty$
- $\lim_n f_n(h) = \gamma(h) \ \forall h \in \mathbb{Z}$

From the dominated convergence theorem:

$$\lim_n \int_{\mathbb{Z}} f_n(h) m \ dh = \int_{\mathbb{Z}} \lim_n f_n(h) m \ dh = \int_{\mathbb{Z}} \gamma(h) m \ dh = \sum_{h=-\infty}^{+\infty} \gamma(h)$$

but, since

$$\lim_n \int_{\mathbb{Z}} f_n(h) m \ dh = \lim_n \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$$

we have successfully demonstrated that

$$\lim_n \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sum_{h=-\infty}^{+\infty} \gamma(h)$$

□

**Exercise 5.3.** *State if the following time series have the mean-ergodic property in $\mathcal{L}^2$ sense:*

1. $\mathcal{IID}(0, \sigma^2)$
2. $\mathcal{WN}(0, \sigma^2)$
3. $X_t = Y$ *with* $Var(Y) < \infty$

**Definition 5.4.** *(From the mean-ergodic theorem) If $(X_t)$ is a stationary time series with $\lim_{h\to\infty} \gamma(h) \neq 0$, the it is said to have a* ***long term dependence****.*

In case we have a long-term dependent time series, are we able to say something on the asymptotic time average?

**Theorem 5.7.** *If $(X_t)$ is a stationary time series, then $\exists \bar{X} \in \mathcal{L}^2$ such that $\bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \bar{X}$.*

If we further remove the stationary property, are we able to say something about the strong ergodic property in $\mathcal{L}^2$?

**Theorem 5.8.** *Consider a time series $(X_t)$ such that:*

- $\mathbb{E}[X_t] = \mu_t \ \forall t \in \mathbb{Z}$
- $Var(X_t) = \sigma_t^2 < \infty \ \forall t \in \mathbb{Z}$
- $\exists \mu \in \mathbb{R} \ s.t. \ \lim_{t\to\infty} \mu_t = \mu$
- $\lim_{t\to\infty)} cov(\bar{X}_n, X_n) = 0$

*then $\bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu$.*

**Corollary 5.2.** *If $(X_t)$ is stationary and* $\lim_{n\to\infty} cov(\bar{X}_n, X_n) = 0$ *then* $\bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu = \mathbb{E}[X_t]$.

**Remark 5.4.** $\lim_{n\to\infty} cov(\bar{X}_n, X_n) = 0 \iff \lim_{n\to\infty} \frac{1}{n}\sum_{h=0}^{n-1}\gamma(h) = 0.$

*Proof.*

$$\begin{aligned}
cov(\bar{X}_n, X_n) &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right)(X_n * \mu)\right] \\
&= \mathbb{E}\left[\frac{1}{n}\left(\sum_{t=1}^{n}(X_t-\mu)(X_n-\mu)\right)\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[(X_t-\mu)(X_n-\mu)\right] \\
&= \frac{1}{n}\left(\gamma(n-1)+\gamma(n-2)+...+\gamma(0)\right) \\
&= \frac{1}{n}\sum_{h=0}^{n-1}\gamma(h)
\end{aligned}$$

□

**Theorem 5.9.** *(Slutsky) If a time series $(X_t)$ is stationary then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{h=0}^{n-1}\gamma(h) = 0 \iff \bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu$$

Then, summing up, the sufficient condition for the mean-ergodic property in $\mathcal{L}^2$ are:

1. $\mathbb{E}[X_t] = \mu_t \ \forall t \in \mathbb{Z}$
2. $Var(X_t) = \sigma_t^2 < \infty \ \forall t \in \mathbb{Z}$
3. $\exists \mu \in \mathbb{R} \ s.t. \ \lim_{t\to\infty}\mu_t = \mu$
4. $\lim_{n\to\infty} cov(\bar{X}_n, X_n) = 0$

**Example 5.1.** *Suppose $(W_t) \sim \mathcal{IID}(0,\sigma^2)$. Consider:*

- $X_t = \mu t + W_t, \ \forall t \Leftarrow$ *3. is not true n this case, since* $\lim_{t\to\infty}\mathbb{E}[X_t] = +\infty$.
- $X_t = \begin{cases} 0 & t \le 0 \\ \sum_{s=1}^{t} W_s & t \ge 1 \end{cases}$

  *recall that* $\mathbb{E}[X_t] = 0 \ \forall t$ *and* $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} tW_{n-t+1} \implies \mathbb{E}[\bar{X}_n] = 0$. *Consider also*

  $$\begin{aligned}
  cov(\bar{X}_n, X_n) &= \mathbb{E}\left[\bar{X}_n, X_n\right] \\
  &= \mathbb{E}\left[\frac{1}{n}(W_n + 2W_{n-1} + ... + nW_1)(W_1 + W_2 + ... + W_n)\right] \\
  &= \frac{1}{n}\left\{\sigma^2 + 2\sigma^2 + ... + n\sigma^2\right\} \\
  &= \frac{\sigma^2}{n}\frac{n(n+1)}{2} \overset{n\to\infty}{\longrightarrow} +\infty
  \end{aligned}$$

*so the condition 4. is not true.*

**Exercise 5.4.** *Suppose* $(X_t)$ *a stationary time series with* $\mathbb{E}[X_t] = 0\ \forall t$ *and* $\lim_{h\to\infty}\gamma(h) = 0$. *Consider* $(Y_t)$ *such that* $Y_t = X_t + W$, *where* $W$ *is a random variable independent of* $(X_t)$ *ans such that* $\mathbb{E}[W] = 0$ *and* $Var(W) = 1$. *Check if the time series* $(Y_t)$ *exhibits a long term dependence.*

Let us end the lecture with some specific notion of ergodicity for Gaussian time series.

**Theorem 5.10.** *Consider a time series* $(X_t)$ *which is Gaussian and stationary. Then*

$$\lim_{h\to\infty}\gamma(h) = 0 \iff \bar{X}_n \overset{\mathcal{L}^2}{\Longrightarrow} \mu$$

Heuristically, a Gaussian time series is ergodic if and only if any two random variables positioned far apart in the sequence are almost independently distributed.

In conclusion, for stationary time series we do not need to observe separate independent trajectories of the time series to obtain a consistent estimation of its overall mean; a good estimation can be obtained by observing a sufficiently long trajectory of the time series.s

**Exercise 5.5.** *Check if the following time series has the mean-ergodic property in* $\mathcal{L}^2$*:*

$$(X_t) \text{ with } X_t = \frac{\mu}{t} + W_t\ t \in \mathbb{Z} - \{0\} \text{ and } (W_t) \sim \mathcal{IID}(0,\sigma^2)$$

# 6 Lecture 6

**Estimation of ACF's functions: statistical properties, examples in R, exercises. Ergodicity covariance property in $\mathcal{L}^2$. Ljung-Box test. Transformations of data and difference operators: examples in R.**

**Definition 6.1.** *Given the observations $X_1, x_2, ..., x_n$ the sample ACF is*

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x}_n)(x_t - \bar{x}_n)$$

*with $\hat{\gamma}(-h) := \hat{\gamma}(h)$, for $h = 0, 1, ..., n-1$ and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.*

**Remark 6.1.** *This result holds for any data set $\{x_1, ..., x_n\}$.*

**Definition 6.2.** *The sample (auto) correlation function is*

$$\hat{\rho} := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad |h| < n$$

**Remark 6.2.** $|\hat{\rho}(h)| \leq 1$

**Example 6.1.** *Plot the sample covariance function and the sample correlation function of 500 observations generated from $\mathcal{GWN}(0, 10)$ (seed=154). Recall that*

$$\gamma(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases} \quad \rho(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

*Plot also the correlation among data up to $h = 9$ by using a **lag plot** (which is a scatter diagram of $\{x_t, x_{t+h}\}$).*

```
####################################
# Generating Gaussian WN(0,10)- ACF#
####################################
# generate a sample path of values from N(0,10) with the seed 154
set.seed(154)
w10=rnorm(500,0,sqrt(10))
# plot the autocorrelation function
acf(w10)
# the blue lines refer to 95%-confidence interval with amplitude 2 /sqrt(n)
# plot the autocovariance function
acf(w10,type='covariance')
# default lag
10*log10(500)
# change the lag window (up to n-1<--499)
acf(w10,lag.max=40)
# to remove the zero lag of the ACF (which is equal to rho(0)=1), let us use the
# library astsa (see first lecture).
# Please, install the package.
install.packages('astsa')
#Load the library
library(astsa)
```

```
# to remove the zero lag value of the ACF
acf1(w10,max.lag=40)
# a different way to see uncorrelation
lag.plot(w10,9)
# lag 1: X_t versus X_t+1--> (x1,x2),(x2,x3),...
# lag 2: X_t versus X_t+2--> (x1,x3),(x2,x4),...
# up to lag 9
# lag 9: X_t versus X_t+9--> (x1,x10),(x2,x11),...
# a different number of plots
lag.plot(w10,4)
# to see how the correlation moves with the lagged time: set n=10
w10=rnorm(10,0,sqrt(10))
lag.plot(w10,4,do.lines=TRUE)
# depending on the version of R, as default do.lines=TRUE. To remove this option use
lag.plot(w10,4,do.lines=FALSE)
# Ljung-Box test
help('Box.test')
#
Box.test(w10,type="Ljung",lag=20,fitdf=0)
```

**Exercise 6.1.** *Plot the sample covariance function and the sample correlation function of $(X_t)$ such that*

$$X_t = \begin{cases} Z_t & t \text{ odd} \\ \frac{Z_{t-1}^2-1}{\sqrt{2}} & t \text{ even} \end{cases}$$

*with $(Z_t) \sim \mathcal{IID}(0,1)$ Gaussian random variables. Set seed=154, t=1 and n=100.*

**Example 6.2.** 1. *Plot the sample covariance function of a random walk with $\mu = 0.2,\ seed = 154, t_0 = 1, n = 200$.*

2. *Study the behaviour of the sample correlation function for the following time windows: (0,40), (0,80), (0,150).*

3. *Produce a lag plot for $h = 9$.*

```
########################################
# Exercise in R
########################################
# Generate a sample path of GWN(0,1) with seed 154
set.seed(154)
w=rnorm(100,0,1)
# initialize the vector x with w
x=w
# change the elements in x corresponding to even steps.
for(i in 1:50) {x[2*i]=(x[2*i-1]^2-1)/sqrt(2)}
# plot the covariance function
acf(x,50,type='covariance')
# load the library
library(astsa)
```

```
# plot ACF without the value in zero
acf1(x,50)
#########################################
# ACF of a random walk
#########################################
# Generate 200 values from N(0,1) with seed 154
set.seed(154)
w=rnorm(200,0,1)# the random walk
x = cumsum(w)
times=seq(1,200,1)
mu = 0.2
xd = mu*times + x
# plot the sample covariance function
acf(xd,type='covariance')
10*log10(200)
#load the library
library(astsa)
# plot the sample correlation function with different time windows
acf1(xd,max.lag=40)
acf1(xd,max.lag=80)
acf1(xd,max.lag=150)
# plot the lag plot
lag.plot(xd,9)
```

As estimators of $\gamma(h)$ (and $\rho(h)$) we can use $x_{t+h} \leftarrow X_{t+h}$ and $\bar{x}_n \leftarrow \bar{X}_n$, obtaining

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

for $h = 0, 1, ..., n-1$ and $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. The same goes for $\rho(h)$.

**Remark 6.3.** *For $h = 0$:*

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2 = \frac{1}{n}\sum_{t=1}^{n} Y_t^2$$

*where $Y_t = X_t - \bar{X}_n$ $t = 1, 2, ..., n$. For $h = 1$:*

$$\hat{\gamma}(1) = \frac{1}{n}\sum_{t=1}^{n-1}(X_{t+1} - \bar{X}_n)(X_t - \bar{X}_n) = \frac{1}{n}\sum_{t=1}^{n-1} Y_{t+1}Y_t$$

*In general:*

$$\hat{\gamma}(h) = \frac{1}{n}\left(Y_{h+1}Y_1 + ... + Y_nY_{n-h}\right)$$

*for $h = 0, 1, ..., n-1$.*

Unfortunately, these estimators are not unbiased.

**Remark 6.4.** *Consider $h = 0$ and suppose $X_1, X_2, ..., X_n$ iid:*

$$\hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2 \neq \sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2$$

*$\hat{\gamma}(h)$ is then a biased estimator of the variance of the random variables involved in the sequence.*

So why use $\hat{\gamma}(h)$?

1. under certain conditions $\lim_{n\to\infty} \mathbb{E}\left[\hat{\gamma}(h)\right] = \gamma(h)$;
2. According to theorem 2.2, $\hat{\gamma}(h)$ is an ACF beacuse:
   - $\hat{\gamma}(h)$ is an even function (by definition);
   - $\hat{\gamma}(h)$ is non-negative definite.

Now we will prove this last assumption.

*Proof.* Consider

$$\left(\hat{T}_{(n)}\right)_{ij} = \hat{\gamma}(i-j)$$

Note that $\hat{T}_{(n)}$ is non-negative definite. Recall that if $Y_i = X_i - \bar{X}_n$ $i = 1, 2, ..., n$. Then

$$\hat{\Gamma}_{(n)} = \frac{1}{n} T T^{\intercal}$$

$\forall \boldsymbol{a} \in \mathbb{R}$ we have

$$\boldsymbol{a}\hat{\Gamma}_{(n)}\boldsymbol{a}^{\intercal} = \frac{1}{n}\boldsymbol{a}TT^{\intercal}\boldsymbol{a}^{\intercal} = \frac{1}{n}\left(\boldsymbol{a}T\right)\left(\boldsymbol{a}T\right)^{\intercal} \geq 0$$

so $\hat{\gamma}$ is a non-negative definite function. □

**Remark 6.5.** *We might check if*

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

*is an unbiased estimator of $\gamma(h)$, but in this scenario is not anymore possible to say that $\hat{\Gamma} = TT^{\intercal}$.*

**Definition 6.3.** *The time series $(X_t)$ has the covariance ergodic property in $\mathcal{L}^2$ if*

$$\hat{\gamma}(h) \overset{\mathcal{L}^2}{\Longrightarrow} \gamma(h) \quad \forall h \in \mathbb{Z}$$

*For Gaussian stationary time series:*

$$\sum_{h=-\infty}^{+} \infty\left|\gamma(h)\right| < \infty$$

*is a sufficient condition to recover the ergodic covariance property in $\mathcal{L}^2$.*

**Example 6.3.** `##################################################`

```
# Generation of sinusoidal ACF
##################################################
#load the library astsa
library(astsa)
# load the data
```

```
data(speech)
# plot the speech dataset
plot(speech)
#plot ACF up to 250, check the sample size
length(speech)
acf1(speech,250)
# plot the lag.plot
lag.plot(speech,9,do.lines=FALSE)
# Test
Box.test(speech,type="Ljung",lag=20,fitdf=0)
#require the library TSA
install.packages('TSA')
library(TSA)
#load the dataset
data(tempdub)
#plot the dataset
plot(tempdub,type='o',ylab='temperature')
#plot ACF up to ?, check the sample size
length(tempdub)
acf1(tempdub,100)
# plot the lag.plot
lag.plot(tempdub,9,do.lines=FALSE)
# Test
Box.test(tempdub,type="Ljung",lag=20,fitdf=0)
```

**Definition 6.4.** *Given the observation $x_1, x_2, ..., x_n$ the sample ACF is*

$$\begin{aligned}\hat{\gamma}(h) &:= \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x}_n)(X_t - \bar{x}) \\ \hat{\gamma}(-h) &:= \hat{\gamma}(h)\end{aligned}$$

*for $h = 0, 1, ..., n-1$ and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_n$.*

It is not possible to estimate the covariance function for $h \geq n$, and it is not recommended to use values of $h$ near to $n$, since the amount of information we can retrieve is very small. For example, if we fix $h = n-1$ we have that that $\sum_{t=1}^{n} -h \leftarrow (x_n, x_1)$. An empirical rule adopted in the literature is, for $n \geq 50$, to use $h \leq \frac{n}{4}$.

It is useful to have a correlation function decreasing to zero to use all these estimators. For doing this, we can use a specific hypothesis test.

**Definition 6.5.** *The **Ljung-Box test** (∼Box-Pierce test) is a statistical test that allows studying the correlation function of a time series. It is defined as*

$$Q = n(n+2)\sum_{h=1}^{N}\frac{(\hat{\rho}(h))^2}{n-h}$$

*with $N \simeq 20$. If the independence between the random variables of the sequence is reasonable, then $Q \simeq \chi^2_N$.*

Examples of the usage of the Ljung-Box test are available in the R code snippets in this Lecture. It can be used to assess the independence between the random variables of the time series.

It is sometimes useful to apply some transformation to a time series in order to investigate it better. The most common transformations are:

- to linearize exponential growth:
- to stabilize the variance (square-root transformations);
- to transform multiplicative pattern in additive ones (logarithmic transformations);
- to make data normally distributed.

The logaritmic and square root transformation are very popular and are special cases of the **Box-Cox transformations**:

$$y_t = \begin{cases} \frac{x_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

**Example 6.4.** *Consider $P_r$ price of a risk asset at some time $t$. Then $\frac{P_t}{P_{t-1}}$ is the relative change of the price over $(t-1,t)$. A very popular model to fit financial data is the sthocastic process $X_t = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1}$.*

**Definition 6.6.** *The **backshift** operator is defined as*

$$B : \mathbb{R}^\mathbb{Z} \mapsto \mathbb{R}^\mathbb{Z}$$

*such that*

$$x = (x_t)_{t\in\mathbb{Z}} \implies Bx = y = (y_t)_{t\in\mathbb{Z}} \; y_t = x_{t-1} \; \forall t \in \mathbb{Z}$$

*and is notated as $BX_t = X_{t-1}$. This operator can be iterated in the following way:*

$$B^J X_t = X_{t-j} \; for \; j \geq 1$$

**Example 6.5.** $B^2X_t = B(BX_t) = B(X_{t-1}) = X_{t-2}$

**Definition 6.7.** *The **difference operator** is defined as*

$$\nabla \overset{def}{=} 1 - B$$

*such that*

$$\nabla X_t = (1-B)X_t = X_t - BX_t = X_t - X_t$$

*Its iterated version is*

$$\nabla^J X_t \overset{def}{=} \nabla(\nabla^{j-1}X_t) \; for \; j \geq 1$$

*assuming*

$$\nabla^0 X_t = X_t$$

*An other version of this operator is the difference operator at lag $d$:*

$$\nabla_d \overset{def}{=} (1 - B^d)$$

*such that*

$$\nabla_d X_t = (1-B^d)X_t = X_t - B^dX_t = X_t - X_{t-d}$$

**Example 6.6.** *Suppose that we want to evaluate*

$$\nabla^2 X_t$$

*We can proceed in two ways:*

*1.*

$$\begin{aligned}\nabla(\nabla X_t) &= (1-B)[(1-B)X_t] \\ &= (1-B)^2 X_t \\ &= (1-2B+B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

*2.*

$$\begin{aligned}\nabla(\nabla X_t) &= \nabla(X_t - X_{t-1}) \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

**Example 6.7.** *Suppose to consider a random walk $X_t = \mu + X_{t-1} + W_t$. We know that this time series is non stationary, but if we apply the difference operator we obtain that:*

$$X_t - X_{t-1} = \mu + W_t \implies \nabla X_t = \mu + W_t$$

*we find a white noise, or a sequence of iid random variables, shifted of $\mu$ which are, in both cases, stationary. This is especially useful for removing trends, and we can be repeatedly applied to obtain a stationary time series.*

**Example 6.8.** *Consider a time series $X_t = \mu_t + Y_t$ with $\mu_t = \delta + \mu_{t-1} + W_t$ random walk and $Y_t$ a stationary time series. Then we have that*

$$\nabla X_t = X_t - X_{t-1} = \mu_t - \mu_{t-1} + Y_t - Y_{t-1} = \sigma + W_t + Y_t - Y_{t-1}$$

*which is a stationary time series (the proof is left as an exercise).*

**Exercise 6.2.** *Suppose $W_t \sim \mathcal{WN}(0, \sigma^2)$. Rewrite the following time series using $B$:*

1. $X_t = W_t + \phi_1 W_{t-1} + ... + \phi_q W_{t-q},\ t \in \mathbb{Z},\ q \in \mathbb{N}$
2. $X_t = \theta_1 X_{t-1} + ... + \theta_p X_{t-p},\ t \in \mathbb{Z},\ p \in \mathbb{N}$

# 7 Lecture 7

**Transformations of data and difference operators: examples in R. Handling time series in R. Additive models: decomposition in trend and seasonal components and remainder term. The STL procedure in R: examples and exercises. Global and local trend: linear trend, regression and approximation. Filtering: two sided moving average, asymmetric filter. The function filter in R.**

Let us start with some examples about the transformations of time series in R.

**Example 7.1.** #########################################

```
# Transformation of the dataset jj
#########################################

# load the library astsa
library (astsa)

# load the dataset jj
data(jj)

# plot the time series in a new window
windows()
plot(jj, type=’o’,ylab=’Earning per Share’,main=’J&J’)

# apply the log-transformation
windows()
plot(log(jj), type=’o’, ylab=’J&J’,main=’Log-transformation’)

# apply the diff-transformation
windows()
plot(diff(jj),type=’o’, ylab=1J&J’,main=1Diff-transformation’)

# apply the iterated diff-transformation
windows()
plot(diff(jj,2),type=’o’, ylab=’J&J’,main=’Diff2-transformation’)

# apply both transformations
windows()
plot(diff(log(jj)), type=’o’, ylab=1J&J’,main=1Diff-Log-transformation’)

# apply both transformations + iterated difference operator
windows()
plot(diff(log(jj),2), type=’o’, ylab=’J&J’,main=’Diff2-Log-transformation’)

# dev.off() to close the windows
# apply the Box-Cox transformation
# install the packages forecast
install.packages(’forecast’)
```

```
#load the library
library(forecast)

# to find optimal lambda
(lambda = BoxCox.lambda(jj))

# now to transform jj
trans.jj = BoxCox( jj, lambda)

# comparisons with the log transformation
windows()
par(mfrow=c(2,1)) plot(log(jj), type=’o’, ylab=’J&J’,main=’Log-transformation’)
plot(trans.jj, type=101,ylab=1J&J’,main=1Power transformation’)
```

How to handle time series is R?

**Example 7.2.** ##########################################

```
# handling time series
##########################################

# create a vector
(mydata = c(l,2,3,2,l))

# transform mydata in a ts object
(mydata = ts(mydata))

# Is mydata a ts?
is.ts(mydata)
is.ts(c(l,2,3,2,l))

# set the initial time
(mydata = ts(c(1,2,3,2,1), start=1950))

# check the time
time(mydata)

# set a different initial time and a different frequency
(mydata=ts(c(1,2,3,2,1), start=c(1950,3), frequency=4))

# check the time
time(mydata)

# useful functions
start(mydata)
end(mydata)
frequency(mydata)
deltat(mydata)

# select a part of ts
(x = window(mydata, start=c(1951,1), end=c(1951,3)))
```

```
# shift ts
cbind(x,lag(x),lag(x,1),lag(x,2),lag(x,0),lag(x,-1),lag(x,-2))

# load jj
library(astsa)
data (jj)
is.ts (jj)
start(j j)
end(j j)
frequency(j j)

# select a window
(x = window(jj, start=c(1965,3), end=c(1970,1)))
# ts.plot (same frequency)
windows()
ts.plot(cmort,tempr,col=c(1 red1,’blue 1),main=1 Two plots’)
legend("topright", legend = c("cmort", "tempr "), lty = 1, col=c(’red’,’blue’) ,
    title = "Line colors", cex = 1.0)
```

Traditional methods, in the analysis of time series, are mainly based on an addtive model:

$$X_t = m_t + s_t + Y_t$$

where $m_t$ is the **trend**, $s_t$ the **seasonal component** and $Y_t$ the **remainder**. The first two quatities are deterministic, while the last one is stochastic. This is not by any means the best model available for time series, but is the first one that scientists use with new data. In short words:

- trend: identifies long term change in the mean leavel;
- seasonal component: short term regular pattern on fixed period (yearly/monthly);
- cyclic component: like the seasonal one, but may vary in lenght.

One method for removing the trend from a time series is by using $\nabla$.

**Example 7.3.**

$$\begin{cases} X_t = m_t + Y_t & \textit{with } m_t = at + b,\ a, b \in \mathbb{R} \\ X_{t-1} = m_{t-1} + Y_{t-1} \end{cases}$$

*In this case, appling $\nabla$, results in*

$$\nabla X_t = X_t - X_{t-1} = m_t - m_{t-1} + (Y_t - Y_{t-1})$$

*which is a stationary time series.*
*If $m_t = p(t)$ polynomial of degree $k$ then we can apply $\nabla^k$.*

On the other hand, $\nabla_d$ can remove the seasonal component.

**Example 7.4.** *Consider*

$$X_t = s_t + Y_t$$

*with period $d$ such that $s_t = s_{t-d}$ $\forall t$. Then we have that*

$$\nabla_d X_t = s_t + Y_{t-d} - s_{t-d} - Y_{t-d}$$

*which has period $d = 0$.*

A useful R function for working with the additive model is the Sesonal and Trend Loess decomposition (**STL procedure**). Loess stands for locally weighted scatterplot smoothing: a kind of piecewise regression.

**Example 7.5.**
```
#########################################
# Decomposition
#########################################

# check the class
class (j j)
frequency(j j )

# the STL decomposition
comp=stl(j j,1 per1)

# The procedure gives in output a matrix comp$time.series with 3 columns
head(comp$time.series, 10)

# To see the decomposition
windows()
plot(comp,main=’J&J decomposition’)

# Plot all the extracted subseries from a time series in one frame
windows()
par(mfrow = c(2,2))
monthplot(jj, main=’datal)
monthplot(comp, choice = "trend", main=’trend’)
monthplot(comp, choice = "seasonal", main=’seasonal’)
monthplot(comp, choice = "remainder", type = "h",main=’remainder’)

# to understand the plot
(x=window(comp$time.series[,1],start=start(comp$time.series[,
1]),end=c(1968,4)))

# compare the decomposition with the transformed dataset
# the STL decomposition to diff(log(jj),2)
compl=stl(diff(log(jj),2),’per’)

# To see the decomposition
windows()
plot(compl,main=’Transformed J&J decomposition’)
```

So, when we fit a time series with an additive model we aim to approximate the trend and the seasonal component and to get residuals. Now we will see how we can work with these elements.

One of the simplest way to fit a trend component is the linear model:

$$m_t = \alpha + \beta t$$

with $\alpha$ and $\beta$ obtained with a regression procedure. This approach works for gloabl trends, bu we can adopt some modifications for fitting local trend:

$$m_t = \alpha_t + \beta_t t$$

In this case we can retrieve $\alpha_t$ and $\beta_t$ by approximation or interpolation.

**Example 7.6.** *Suppose we want to fit an exponential trend (like the one of the dataset jj):*

$$\begin{aligned} m_t &= \alpha \exp\{\beta t\} \\ \log m_t &= \log \alpha + \beta t \quad regressionprocedure \\ m_t &= \exp\{A\} \exp\{Bt\} \end{aligned}$$

**Example 7.7.**
```
#################################################
# Analysis of the trend: jj dataset
#################################################

# assign to the variable "trend" the column with the trend component
trend=comp$time.series[,2]

# take the time
time.trend = time(trend)

# ask for an exponential fitting, that is y=a*exp(b*t)=> log y = log a + b*t
fit=lm(log(trend) ˜ time.trend, na.action=NULL)

# plot the results: if A and B are the output of the lm procedure the exponential
# model is exp(A)*exp(B*t)
plot(trend,main='Fitting the trend component')
y=exp(fit$coefficients[1])*exp(fit$coefficients[2]*time.trend)
lines(y, col='red')
```

**Example 7.8.**
```
#################################################
# Analysis of the trend: Cardiovascular Mortality
#################################################

#load the library
library(astsa)

# plot the time series
plot(cmort, main="Cardiovascular Mortality", xlab="years", ylab="age")
```

```
# check the frequency and tie class
class(cmort)
frequency(cmort)

# more information
start(cmort)
end(cmort)

# ask for the decomposition
comp=stl(cmort,'per')
plot(comp)

# monthplot to understand the dominance
monthplot(cmort, main='data')

# extrapolate the trend component with the time
trend=comp$time.series[,2]
time.trend=time(comp$time.series[,2])

# plot in a new window the trend component with the spline approximation
plot(time.trend, trend, main = "Approximation with a cubic spline")
lines(spline(time.trend,trend),col='red')

#to get the spline function
splinefun(time.trend,trend)
```

An other technique to deal with trends is to use a **filter** that transform a time series in an other time series by using a linear transformations (**moving average**):

$$\tilde{x}_t = \sum_{j=-k}^{p} a_j x_{t+j}$$

with $\{a_j\}$ such that $\sum a_j = 1$ (often $a_j \in [0,1]$). A variation of this technique is the **symmetric moving average**, where $p = k$ and $a_j = a_{-j}$. The simplest example of a symmetric smoothing filter is the **two-sided moving average**:

$$a_j = \frac{1}{2k+1},\ j = -k, ..., k \implies \tilde{x}_t = \frac{1}{2k+1} \sum_{j=-k}^{k} x_{t+j}$$

This will allow us to smooth the local fluctuations of the data, but only works at the center of the dataset, due to his construction.

There exists also **asymmetric filters**. A simple one is the **one-sided moving average**:

$$a_j = \frac{1}{k+1},\ j = 0, ..., k \implies \tilde{x}_t = \frac{1}{k+1} \sum_{j=0}^{k} x_{t-j}$$

This filter can be used for a variety of operations:

- forecasting: $\tilde{x}_{n+h} = \frac{1}{k+1} \sum_{j=0}^{k} x_{n-j}$

- naive forecasting: $k = 0 \implies \tilde{x}_{n+h} = x_n$
- two-term forecast: $k = 1 \implies \tilde{x}_{n+h} = \frac{1}{2}(x_n + x_{n-1})$

One more asymmetric filter is the **exponential smoothing**:

$$\tilde{x}_t = \sum_{j=0}^{m} \alpha(1-\alpha)^j x_{t-j}, \ t - m > 0, \ \alpha \in (0,1)$$

**Example 7.9.** `#################################################`

```
# Filtering the trend of jj dataset
#################################################

# use a filter to smooth trend
(k=rep(l,5)) # weights=l
(k=k/sum(k)) # normalized weights

# two sided moving average: sides=2
# fjj(t)=0.2*trend(t-2)+0.2*trend(t-1)+0.2*trend(t)+0.2*trend(t+1)+0.2*trend(t+2)
(fjj=filter(trend,methodic('convolution'),sides=2, k))

# Not Available elements appear for t=l and 2 and at the end points
# plot the original dataset and the filtered dataset
windows()
plot(trend,main=1 Filtering the trend component: 2 sided1)
lines(fjj, col='redl)

# one-sided moving average: sides=l
#
fjj(t)=0.2*trend(t-4)+0.2*trend(t-3)+0.2*trend(t-2)+0.2*trend(t-1)+0.2*trend(t)
(fjjl=filter(trend,method=c(1conv'),sides=l, k))

# Plot
windows()
plot(trend, ylab='Earning per Share',main='Filtering the trend component: 1
sided',ylim=c(0,12) )
lines(fjjl, col='blue')
```

Under suitable conditions, the two sided moving averrage filter may be used to approximate a linear trend.

**Example 7.10.** *In the absence of a seasonal component and in the presence of a linear trend, prove that the two sided moving average of lag $x$ is approximately equal to the trend. Assume the average og the remainder term over $[t-k, t+k]$ appeoximately $0$ for all $k$ such that $n - 2k > 0$.*

*Consider the two sided moving average filter:*

$$\tilde{x}_t = \frac{1}{2k+1} \sum_{j=-k}^{k} x_{t+j}$$

*with* $1+k \leq t \leq n-k$*. This because the first index of the summation* $1 \leq t-k \leq t+k \leq n$ *the last index of the summation. Now, the additive model is*

$$X_t = m_t + Y_t$$

*since we do not have a seasonal component. Substituting* $x_{t+j}$ *with* $m_{t+j} + y_{t+j}$ *we obtain*

$$\begin{aligned}
\tilde{x}_t &= \frac{1}{2k+1} \sum_{j=-k}^{k} m_{t+j} + \frac{1}{2k+1} \sum_{j=-k}^{k} y_{t+j} \\
&\simeq \frac{1}{2k+1} \sum_{j=-k}^{k} m_{t+j} \\
&= \frac{1}{2k+1} [a(t-k) + b + a(t-k+1) + b + ... + a(t+k-1) + b + a(t-k) + b] \\
&= \frac{1}{2k+1} [(2k+1)(at+b) + (-ak - ak + a + ... + ak - a + ak)] \\
&= \frac{1}{2k+1} (2k+1)(at+b) \\
&= at + b
\end{aligned} \tag{1}$$

*Thus proving that the two sided moving average approximately gives an approximation of the trend.*

**Exercise 7.1.** *Let* $Y_t$ *be a stationary time series with zero mean. Consider also a time series* $X_t = a + bt + s_t + Y_t$ *with* $a, b \in \mathbb{R}$*, where* $s_t$ *is a seasonal component with period* $d = 12$*. Show that*

$$\nabla \nabla_{12} X_t$$

*is stationary, and express its ACF in terms of the ACF of* $\{Y_t\}$*.*

# 8 Lecture 8

**Seasonal component: fitting with a regression. Periodogram. Analysis of the remainder term. Examples, exercises and case studies in R. Linear filter, time-invariance, MA(q) t.s. Laurent series and operators: convergence a.s.**

A seasonal component is typically represented as a discrete fourier summation:

$$s_t = \alpha_0 + \alpha_1 \cos(2\pi\omega_1 t) + \beta_1 \sin(2\pi\omega_1 t) + ... + \alpha_p \cos(2\pi\omega_p t) + \beta_p \sin(2\pi\omega_p t)$$

**Example 8.1.** *A case study in R: generate a path of*

$$X_t = 2\sin\left(\frac{2\pi}{50}t + 0.6\pi\right) + 5W_t \quad W_t = \mathcal{GWN}(0,1)$$

*and fit the periodic deterministic model underlying the time series*

$$s_t = 2\sin\left(\frac{2\pi}{50}t + 0.6\pi\right) \implies s_t = A\sin(2\pi\omega t + \phi)$$

*Since it is difficult to fit the data with these functions, it is more useful to reparametrize the formula as*

$$s_t = A\cos(2\pi\omega t)\sin\phi + A\sin(2\pi\omega t)\cos\phi = B_1\cos(2\pi\omega t) + B_2\sin(2\pi\omega t)$$

```
#################################################
# Dealing with a periodic signal
#################################################

(omega=1/50) # the frequency

(A=2) # the amplitude

(phi=0.6*pi) # the phase

(start.cycle = -phi/(2*pi*omega))
# the shift of the first cycle with respect to the y-axis

# to see the shift, plot the deterministic function s_t
times=seq(start.cycle,100,1)
wave = A*sin(2*pi*omega*times + phi)
plot(times,wave, type='l',main='Deterministic function s_t')
abline(v=0,col='red')
abline(h=0,col='blue')
abline(v=50,col='purple')

# generate the GWN noise
set.seed(154)
w = rnorm(200,0,1)
wave = A*sin(2*pi*omega*(1:200) + phi)
```

```
plot((1:200),wave+5*w, type='b',main='Deterministic s_t+ 5*N(0,1)',ylab='X_t')
abline(h=0,col='red')

#estimate s_t by using B1 cos(2*pi*omega*times)+B2 sin(2*pi*omega*times)
z1 = cos(2*pi*omega*(1:200))
z2 = sin(2*pi*omega*(1:200))

# run the linear regression
fit <- lm(wave~z1+z2)

# the coefficients of the regression
fit$coefficients

# compare with B1 and B2
A*sin(phi);A*cos(phi);
(coeff=fit$coefficients[2:3])

#fitting the wave with the noise
fit1 <- lm(wave+5*w~z1+z2)
(coeff=fit1$coefficients[2:3])
```

**Example 8.2.** *We will try to fit the seasonal component of TEMPDUB with the same technique we have just seen.*

```
########################################
# The dataset TEMPDUB
########################################

# load the library and the dataset
library(TSA)
data(tempdub)

#plot the dataset
plot(tempdub,type='o',ylab='temperature')

# do a regression on data with sin and con functions
timetemp=time(tempdub)
z1 = cos(2*pi*timetemp)
z2 = sin(2*pi*timetemp)
fit <- lm(tempdub~z1+z2)

# plot tempdub versus the regression model
windows()
par(mfrow=c(2,1))
plot(tempdub,ylab='temperature',type='o',main='Data')
is.ts(timetemp)
timef=as.numeric(timetemp)
plot(timef,fit$fitted.values,type='b', main='Fitting', xlab='times',
    ylab='fitted values')
```

```
# same range on the y-axis
(tdmin=min(as.numeric(tempdub)))
(tdmax=max(as.numeric(tempdub)))
windows()
par(mfrow=c(2,1))
plot(tempdub,ylab='temperature',type='o',main='Data')
plot(timef,fit$fitted.values,type='b', main='Fitting', xlab='times',
    ylab='fitted values', ylim=range(tdmin,tdmax))

# transform the fitting in a time series
is.ts(fit$fitted.values)
fit.ts=ts(fit$fitted.values,start=start(tempdub),end=end(tempdub),
    frequency=frequency(tempdub))

# a different plot
windows()
ts.plot(tempdub,fit.ts,col=c('red','blue'),main='Tempdub vs the fitting',
    ylim=range(8,90))
legend("topright", legend = c("tempdub", "fit"),
lty = 1, col=c('red','blue'),
title = "Line colors", cex = 1.0)

# exercise: fit the seasonal component obtained using the STL procedure
# applied to the dataset jj
```

Any time series can be expressed as a combination of sine and cosine functions with different frequencies. We can use the **periodogram** to graph a measure of the relative importance of possible frequency values that might explain the oscillation pattern of the observed data. Suppose we the series of harmonic frequencies:

$$\omega_j = \frac{j}{n} \quad for j = 1, 2, ..., \frac{n}{2}$$

we can represent this series as

$$X_t = \sum_{j=1}^{\frac{n}{2}} \left[ \beta_1^{j,n} \cos(2\pi\omega_j t) + \beta_2^{j,n} \sin(2\pi\omega_j t) \right]$$

implying $n$ parameters. The first step in setting up the periodogram is estimating these parameters $(\beta_1, \beta_2)$. The mathematical tool used for doing this is the fast Fourier transformations. The value of the periodogram at $\frac{j}{n}$ is

$$P\left(\frac{j}{n}\right) \approx (\beta_1^{j,n})^2 + (\beta_2^{j,n})^2$$

The dominant frequencies might be used to fit cosine and sine wawes to the data or to describe the important periodicities in the series.

**Example 8.3.** ############################################

```
# Periodogram
###########################################
```

```
# first example
# set the frequency, the amplitude, the phase
(omega=1/50)
(A=2)
(phi=0.6*pi)

# define the wave
times=seq(1,100,1)
wave = A*sin(2*pi*omega*times + phi)

#load the library
library('TSA')

# run the periodogram
out=periodogram(wave)

# check harmonic frequencies (equally spaced with step=1/sample.size)
(sample.size=length(times))
(step=1/sample.size)
head(out$freq,6)

# where the maximum is located
which.max(out$spec)

# the related frequency
out$freq[which.max(out$spec)]

# the period
1/out$freq[which.max(out$spec)]

# second example
#There are 6 waves over (0,100)
x1 = 2*cos(2*pi*1:100*6/100) + 3*sin(2*pi*1:100*6/100)

#There are 10 waves over (0,100)
x2 = 4*cos(2*pi*1:100*10/100) + 5*sin(2*pi*1:100*10/100)

#There are 40 waves over (0,100)
x3 = 6*cos(2*pi*1:100*40/100) + 7*sin(2*pi*1:100*40/100)
x = x1 + x2 + x3

# run the periodogram
out=periodogram(x)

# the dominant period
1/out$freq[which.max(out$spec)]

# why?
#frequencies
```

```
100/6

#amplitude 2
100/10 #amplitude 10
100/40 #amplitude 40

# third example
data(tempdub)
plot(tempdub,ylab='temperature',type='o')
out=periodogram(tempdub)

# the increment of the harmonic frequencies
(step=1/length(time(tempdub)))

# check
head(out$freq,6)

# the period
1/out$freq[which.max(out$spec)]

# compare with the output of the frequency function
frequency(tempdub)
```

Let us make a summary of what we have seen. The techniques proposed are useful to analyze the main features of a time series, like trend and seasonal component, but the next step is to fit them with suitable deterministic functions. The last step is to recover and analyze the residuals, which is better if it is stationary (we can use transformation for doing this), and fit it with a classical theoretical model.

**Example 8.4.**
```
####################################################
# Analysis of the reminder term for the dataset jj
###################################################

#load the library and the dataset
library(astsa)
data(jj)

# ask for the decomposition in additive model of the log(jj)
comp=stl(log(jj),'per')

#select the residuals
residualjj=comp$time.series[,3]

# do a plot around the zero level
plot(residualjj,type='b',main='Residuals of the additive model',ylab='residuals')
abline(h=0,col='red')

# Are the residuals observations of a gaussian ts?
# produce a qqnorm of the residuals
```

```
qqnorm(residualjj,col='blue',main='QQplot residuals')
qqline(residualjj,col='red')

# create an histogram with a kernel stimator
hist(residualjj,prob=TRUE,12,main='Histogram residuals')
lines(density(residualjj),col='red')

# ACF of residualjj
acf(residualjj,60)

# Box.test: H0: the ts is aymptotically uncorrelated
# pag.310 in Time series: theory and methods, Brockwell and Davis
library(tseries)
Box.test(residualjj,type=c('Ljung-Box'))

# for testing normality: skewness and kurtosis
install.packages(moments)
library(moments)
skewness(residualjj)
kurtosis(residualjj)

# do the jarque.bera test--> H0: data from a normal distribution
jarque.bera.test(residualjj)

# The KPSS test
# H0: the contribution of the random walk is zero
kpss.test(residualjj)
```

The **KPSS test** is used to test the null hypothesis that time series observations result to be stationary around a deterministic trend (i.e. trend-stationary). The model is

$$X_t = \textit{deterministic trend} + \textit{random walk} + \textit{stationary noise}$$

where the random walk has a zero drift. The null hypothesis of the test is $H_0$: the variance of the random walk is zero.

The stochastic model we employ is based on the idea that observations of a time series $X_t$ in which successive values are highly dependent can be frequently regarded as generated by a sequence of uncorrelated shocks. These shocks are randomly drawn from a fixed distribution usually assumed to be white noise. So the white noise is supposed to transform to a sequence $X_t$ by a linear filter.

**Definition 8.1.** *A **linear filter** $\mathcal{L} : \mathcal{E} \mapsto \mathcal{E}$ ($\mathcal{E}$ is the set of all time series) such that $\forall \alpha_1, \alpha_2 \in \mathbb{R}$ and $\forall (W_{t,1}), (W_{t,2} \in \mathcal{E})$:*

$$\mathcal{L}\left[\alpha_1(W_{t,1}) + \alpha_2(W_{t,2})\right] = \alpha_1\mathcal{L}\left[(W_{t,1})\right] + \alpha_2\mathcal{L}\left[(W_{t,2})\right]$$

**Example 8.5.** *Consider:*

- $\mathcal{L}\left[(W_t)\right] = (3W_t)$*: this is a linear filter*
- $\mathcal{L}\left[(W_t)\right] = (W_t^2)$*: this is not a linear filter*

- $X_t - 2X_{t-1} + 3X_{t-2} = 4W_t + 5W_{t-1}$: *this is a linear filter*

**Example 8.6.** *Consider* $(X_t) = \mathcal{L}[(W_t)]$ *such that*

$$X_t = \sum_{j \in I_t} \psi_{t,j} W_j$$

*where* $I_t \subset \mathbb{Z}$ *and* $|I_t| < \infty$. $\psi_{t,j}$ *are the weights. This linear filter is a special case of two sided moving average filter:*

$$\psi_{t,j} = \frac{1}{2k+1}, \ j \in I_t \quad and \quad I_t = \{-k, ..., k\} \subseteq \mathbb{Z}$$

**Definition 8.2.** *A linear filter is **time invariant** if dealying the input of any constant step n delays the output of the same constant step n:*

$$\mathcal{L}[(W_{t-n})] = (X_{t-n})$$

$\forall n \in \mathbb{N}$ *and* $W_{t-n} = B^n W_t, \ X_{t-n} = B^n X_t$ *with B backshift operator.*

**Example 8.7.** *Consider the linear filter*

$$\mathcal{L}[(W_t)] = (X_t)$$

*with* $X_t = \nabla W_t$. *Is this filter time invariant? Let us check the linear property:*

$$\begin{aligned}
\nabla(\alpha_1 W_{t,1} + \alpha_2 W_{t,2}) &= (\alpha_1 W_{t,1} + \alpha_2 W_{t,2}) - (\alpha_1 W_{t-1,1} + \alpha_2 W_{t-1,2}) \\
&= \alpha_1 \nabla W_{t,1} + \alpha_2 \nabla W_{t,2} \\
\mathcal{L}[\alpha_1 (W_{t,1}) + \alpha_2 (W_{t,2})] &= (\nabla(\alpha_1 W_{t,1} + \alpha_2 W_{t,2})) \\
&= (\alpha_1 \nabla W_{t,1} + \alpha_2 \nabla W_{t,2}) \\
&= \alpha_1 (\nabla W_{t,1}) + \alpha_2 (\nabla W_{t,2}) \\
&= \alpha_1 \mathcal{L}[(W_{t,1})] + \alpha_2 \mathcal{L}[(W_{t,2})]
\end{aligned}$$

*Confirming the linearity of the filter. Observe that*

$$\nabla(B^n W_t) = (1-B)B^n W_t = (B^n - B^{n-1})W_t = B^n(1-B)W_t = B^n \nabla W_t$$

*Then:*

$$\mathcal{L}[(W_{t-n})] = (\nabla B^n W_t) = (B^n \nabla W_t) = (X_{t-n})$$

**Exercise 8.1.** *Check if* $\mathcal{L}[(W_t)] = (aW_{-t}), \ a \in \mathbb{R}$ *is a time invariant linear filter.*

**Proposition 8.1.** *If* $\psi(z) = \sum_{j=-k}^{k} \psi_j z^j, \ \{\psi_j\} \in \mathbb{R}$ *then*

$$\mathcal{L}[(W_t)] = (X_t)$$

*with* $X_t = \psi(B)W_t$, *is a time invariant time series.*

**Remark 8.1.** $\psi(z)$ *is a Laurent polynomial.*

*Proof.* We have that

$$X_t = \psi(B)W_t = \sum_{j=-k}^{k} \psi_j B^j W_t = \sum_{j=-k}^{k} \psi_j W_{t-j}$$

Observe that:

$$\psi(B)B^n = B^n\psi(B)$$
$$\psi(B)B^nW_t = B^n\psi(B)W_t$$

Then

$$\mathcal{L}\left[(W_{t-n})\right] = (\psi(B)W_{t-n}) = (\psi(B)B^nW_t) = (B^n\psi(B)W_t) = (X_{t-n})$$

□

**Definition 8.3.** *One more example of this class of time invariant linear filter is the* ***moving average of order*** $q \in \mathbb{N}$*. In this case* $(W_t) \sim \mathcal{WN}(0, \sigma^2)$*,* $k = q$ *and*

$$\phi_j = \begin{cases} 0 & j < 0 \\ 1 & j = 0 \\ \phi_j & j = 1, 2, ..., q \end{cases}$$

*implying that*

$$\psi(z) = 1 + \phi_1 z + ... + \phi_1 z^q = \phi(z)$$

*Therefore* $\mathcal{L}\left[(W_t)\right] = (X_t)$ *with:*

$$X_t = \phi(B)W_t = W_t + \phi_1 W_{t-1} + ... + \phi_q W_{t-q}$$

$\forall t \in \mathbb{Z}$*. The moving average of order* $q$ *is a weighted linear combination of the present value* $W_t$ *and its* $q$ *past values.*

Now we will try to give meaning to the operator

$$\phi(B) = \sum_{j\in\mathbb{Z}} \psi_j B^j$$

Consider

$$\tilde{X}_{t,n} = \sum_{j=-n}^{n} \psi_j W_{t-j}$$

with $t \in \mathbb{Z}, n \in \mathbb{N} \implies (\tilde{X}_{t,n})_{n,t\in\mathbb{N}}$ random variables. What happends when $n \to \infty$?

**Lemma 8.1.** *If* $(X_k) \in (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ *then*

1. $\mathbb{E}\left[\sum_k |X_k|\right] = \sum_k \mathbb{E}\left[|X_k|\right]$
2. *If* $\sum_k \mathbb{E}\left[|X_k|\right] < \infty$ *then*
   (a) $\sum_k |X_k| \implies \mathbb{P}\left(\{\omega \in \Omega / \sum_k |X_t(\omega)| < \infty\}\right) = 1$
   (b) $\exists \lim_n S_n$ *almost surely with* $S_n$ *the n-th partial sum of* $\sum_k X_t$
   (c) $\mathbb{E}\left[\sum_k X_k\right] = \sum_k \mathbb{E}\left[X_t\right]$

**Theorem 8.1.** *Suppose* $\{\psi_j\}_{j\in\mathbb{Z}} \in \mathbb{R}$ *and* $(W_t)$ *such that* $\sup_{t\in\mathbb{Z}} \mathbb{E}\left[|W_t|\right] < \infty$, *then*

$$\tilde{X}_{t,n} \overset{a.s.}{\Longrightarrow} X_t \quad \forall t \in \mathbb{Z}$$

*implying that* $X_t = \sum_{j=\mathbb{Z}} \psi_j W_{t-j}$.

**Corollary 8.1.**

$$\mathbb{P}\left(\left\{\omega \in \Omega / \sum_{j\in\mathbb{Z}} |\phi_j|\,|W_{t-j}(\omega)| < \infty\right\}\right) = 1$$

*Proof.* To prove that $\exists \lim_n \tilde{X}_{t,n}$ we have to prove

$$\sum_{j\in\mathbb{Z}} \mathbb{E}\left[|\phi_j W_{t-j}|\right] < \infty \quad \forall t \in \mathbb{Z}$$

Fix $t \in \mathbb{Z}$ and $M = \sup_{t\in\mathbb{Z}} \mathbb{E}\left[|W_t|\right]$:

$$\begin{aligned}\sum_{j\in\mathbb{Z}} \mathbb{E}\left[|\phi_j W_{t-j}|\right] &= \lim_n \sum_{j=-n}^{n} \mathbb{E}\left[|\phi_j W_{t-j}|\right] \\ &= \lim_n \sum_{j=-n}^{n} |\phi_j|\, \mathbb{E}\left[|W_{t-j}|\right] \\ &< M \lim_n \sum_{j=-n}^{n} |\phi_j| < \infty\end{aligned}$$

proving theorem 8.1. Now we will prove corollary 8.1.
If $S$ is a positive random variable $\mathbb{E}[S] < \infty \implies \mathbb{P}(S = +\infty) = 0$. From the lemma we have that

$$\mathbb{E}\left[\sum_{j\in\mathbb{Z}} |\phi_j|\,|W_{t-j}|\right] = \sum_{j\in\mathbb{Z}} |\phi_j|\,\mathbb{E}\left[|W_{t-j}|\right] < \infty$$

$$\mathbb{P}\left(\left\{\omega \in \Omega / \sum_{j\in\mathbb{Z}} |\phi_j|\,|W_{t-j}(\omega)| = +\infty\right\}\right) = 0$$

$$\mathbb{P}\left(\left\{\omega \in \Omega / \sum j \in \mathbb{Z}\, |\phi_j|\,|W_{t-j}(\omega)| < \infty\right\}\right) = 1$$

□

# 9 Lecture 9

**Laurent series and operators: convergence in mean squared, linear time series, causality. MA of order infinite. Stationary t.s. obtained by applying linear filters to stationary t.s. ACF of linear time series and MA. Properties of MA(q).**

**Exercise 9.1.** *Suppose $\mathcal{E}$ the set of all $(W_t) \in \mathcal{L}^2$ bounded. Consider*

$$\mathcal{L}\left[(W_t)\right] = (X_t)$$

*with $X_t = \phi(B)W_t \; \forall t \in \mathbb{Z}$ where $\psi(Z) = \sum_{j\in\mathbb{Z}} \psi_j Z^j$. Check if $\mathcal{L}$ is a time invariant linear filter.*

**Definition 9.1.** *If $(W_t) \sim \mathcal{WN}(0, \sigma^2)$ them $(X_t)$ with*

$$X_t = \psi(B)W_t \quad \forall t \in \mathbb{Z}$$

*is said a **linear time series**.*

**Remark 9.1.** *If $\mathcal{E}$ is the set of all $\mathcal{WN}(0, \sigma^2)$ and $\mathcal{L}\left[(W_t)\right] = (X_t)$ such that the previous theorem's condition holds, then $\mathcal{L}$ is a time-invariant linear filter.*

**Definition 9.2.** *If $\psi_j = 0$ for $j < 0$ then $(X_t)$ is said to be a **causal linear time series**:*

$$X_t = \sum_{j\geq 0} \psi_j W_{t-j} = \psi_0 W_t + \psi_1 W_{t-1} + \psi_2 W_{t-2} + ...$$

*In other words, it does not depend on the future.*

**Exercise 9.2.** *Check if $X_t = a_{-1}W_{t+1} + a_0 W_t + a_1 W_{t-1}$ with $a_{-1}, a_0, a_1 \in \mathbb{R}$ and $(W_t) \sim \mathcal{WN}(0, \sigma^2)$ is a causal linear time series.*

**Definition 9.3.** *If $(W_T) \sim \mathcal{WN}(0, \sigma^2)$, $(X_t) \sim MA(\infty)$ of $(W_t)$ if $\exists \{\psi_j\}$ for $j > 0$ with $\psi_0 = 1$ and $\sum_{j\geq 0} |\psi_j| < \infty$ such that*

$$X_t = \sum_{j\geq 0} \psi_j W_{t-j} \quad \forall t \in \mathbb{Z}$$

**Example 9.1.** *Consider $\psi_j = \rho^j$ with $|\rho| < 1$ and $j \geq 0$. Then*

$$X_t = \sum_{j\geq 0} \rho^j W_{t-j} \sim MA(\infty)$$

*with $(W_t) \sim \mathcal{WN}(0, \sigma^2)$.*

**Proposition 9.1.** *If $(W_t) \sim \mathcal{IID}(0, \sigma^2)$ then $MA(\infty)$ is strong stationary.*

*Proof.* $\forall g$ measurable function, if $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are such that $\boldsymbol{z}_1 \overset{d}{=} \boldsymbol{z}_2 \implies g(\boldsymbol{z}_1) \overset{d}{=} g(\boldsymbol{z}_2)$. Fix $K \in \mathbb{N}$ and $(t_1, ..., t_k) = \boldsymbol{t} \in \mathbb{Z}^k$:

$$\tilde{X}_{boldsymbolt,n} = \left(\tilde{X}_{t_1,n}, ..., \tilde{X}_{t_k,n}\right) = \left(\sum_{j=0}^{n} \psi_j W_{t_1-j}, ..., \sum_{j=0}^{n} \psi_j W_{t_k-j}\right)$$

as $(W_t) \sim \mathcal{IID}(0, \sigma^2)$:

$$(W_{t_1}, ..., W_{t_i - n}) \overset{d}{=} (W_{t_i + h}, ..., W_{t_i + h - n})$$

for $i = 1, ..., k,\ \forall h \in \mathbb{Z}$. From the first statement of the proof, we obtain that

$$\tilde{X}_{\boldsymbol{t},n} \overset{d}{=} \left( \sum_{j=0}^{n} \psi_j W_{t_1 + h - j}, ..., \sum_{j=0}^{n} \psi_j W_{t_k + h - j} \right) = \left( \tilde{X}_{t_1 + h, n}, ..., \tilde{X}_{t_k + h, n} \right) = \tilde{X}_{\boldsymbol{t+h},n}$$

From theorem 8.1 we obtain that

$$\tilde{X}_{\boldsymbol{t},n} = \left( \tilde{X}_{t_1,n}, ..., \tilde{X}_{t_k,n} \right) \overset{a.s.}{\Longrightarrow} (X_{t_1}, ..., X_{t_k}) = X_{\boldsymbol{t}}$$

as well as

$$\tilde{X}_{\boldsymbol{t+h},n} = \left( \tilde{X}_{t_1 + h,n}, ..., \tilde{X}_{t_k + h,n} \right) \overset{a.s.}{\Longrightarrow} (X_{t_1 + h}, ..., X_{t_k + h}) = X_{\boldsymbol{t+h}}$$

as $\tilde{X}_{\boldsymbol{t},n} \overset{d}{=} \tilde{X}_{\boldsymbol{t+h},n} \implies X_{\boldsymbol{t} \overset{d}{=} X_{\boldsymbol{t+h}}}$ □

Given proposition 9.1 and the following result:

**Theorem 9.1.** *If $(X_t)$ is a time series such that $X_t = f(W_s, s \leq t)$, $\forall t \in \mathbb{Z}$ and $(W_t) \sim \mathcal{IID}$ with $f$ measurable function, then $(X_t)$ is ergodic.*

we can say that $MA(\infty)$ is ergodic.

**Theorem 9.2.** *Let $\tilde{X}_{t,n} = \sum_{j=-n}^{n} \psi_{t-j} W_{t-j}$ with $\{\psi_j\}$ such that $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$. If $\sup_{t \in \mathbb{Z}} \mathbb{E}\left[W_t^2\right] < \infty$ then*

$$\tilde{X}_{t,n} \overset{\mathcal{L}^2}{\Longrightarrow} \tilde{X}_t \quad \forall t \in \mathbb{Z}$$

*with $X_t = \sum_{j \in \mathbb{Z}} \psi_j W_{t-j}$.*

*Proof.* As first step, let us prove that $\left\{ \tilde{X}_{t,n} \right\}$ is a $\mathcal{L}^2$ Cauchy sequence, that is

$$\lim_n \mathbb{E}\left[ \left( \tilde{X}_{t,n} - \tilde{X}_{t,m} \right)^2 \right] = 0 \quad \forall t \in \mathbb{Z}, n, m \in \mathbb{N}$$

Fix $0 < m < n,\ t \in \mathbb{Z}$:

$$\begin{aligned} \mathbb{E}\left[ \left( \tilde{X}_{t,n} - \tilde{X}_{t,m} \right)^2 \right] &= \mathbb{E}\left[ \left( \sum_{j=-n}^{n} \psi_j W_{t-j} - \sum_{j=-m}^{m} \psi_j W_{t-j} \right)^2 \right] \\ &= \mathbb{E}\left[ \left( \sum_{m < |j| \leq n} \psi_j W_{t-j} \right)^2 \right] \\ &= \sum_{m < |j| \leq n} \sum_{m < |k| \leq n} \psi_j \psi_k \mathbb{E}\left[ W_{t-j} W_{t-k} \right] \end{aligned}$$

Set $M = \sup_t \mathbb{E}\left[W_t^2\right] < \infty$. From Cauchy-Schwartz:

$$\begin{aligned}
\mathbb{E}\left[W_{t-j}W_{t-k}\right] &\leq \left(\mathbb{E}\left[W_{t-j}^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[W_{t-k}^2\right]\right)^{\frac{1}{2}} \\
&\leq \left(\sup_t \mathbb{E}\left[W_t^2\right]\right)^{\frac{1}{2}} \left(\sup_t \mathbb{E}\left[W_t^2\right]\right)^{\frac{1}{2}} \\
&= M < \infty
\end{aligned}$$

We now have that

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{X}_{t,n} - \tilde{X}_{t,m}\right)^2\right] &\leq M \sum_{m<|j|\leq n} \sum_{m<|k|\leq n} \psi_j \psi_k \\
&= \left(\sum_{m<|j|\leq n} |\phi_j|\right)^2 \\
&= \left(\sum_{|j|\leq n} |\psi_j| - \sum_{|k|\leq m} |\psi_j|\right)^2 \\
&= \left(S_n - S_m\right)^2
\end{aligned}$$

$S_n, \ S_m$ are partial sums of

$$\sum_{j\in\mathbb{Z}} |\psi_j| \leq \infty$$

So $\{S_n\}$ is a Cauchy sequence, implying that $\lim_{n,m} \left(S_n - S_m\right)^2 = 0$. We can now state that

$$\lim_{n,m} \mathbb{E}\left[\left(\tilde{X}_{t,n} - \tilde{X}_{t,m}\right)^2\right] \leq M \lim_{n,m} \left(S_n - S_m\right)^2$$

but, since we kwnow that the right member of the equation is equal to 0, we also know that the first one is. Now recall the following result: if $\{X_n\}$ are real valued random variables then the following statements are equivalent:

1. $\{X_n\}$ converges is $\mathcal{L}^2$
2. $\{X_n\}$ is a $\mathcal{L}^2$ Cauchy sequence
3. $\{X_n\}$ converges in probability and is uniformly integrable

As $\left\{\tilde{X}_{t,n}\right\}$ is a $\mathcal{L}^2$ Cauchy sequence then $\exists S_t \in \mathcal{L}^2$ such that

$$\tilde{X}_{t,n} \overset{\mathcal{L}^2}{\Longrightarrow} S_t \textit{ and } \tilde{X}_{t,n} \overset{P}{\Longrightarrow} S_t$$

As $(W_t) \in \mathcal{L}^2$-bounded, then $(W_t) \in \mathcal{L}^1$-bounded, proving theorem 9.1, and

$$\tilde{X}_{t,n} \overset{a.s.}{\Longrightarrow} X_t \textit{ and } \tilde{X}_{t,n} \overset{P}{\Longrightarrow}$$

then $S_t \overset{a.s.}{=} X_t$. □

**Theorem 9.3.** *Consider a time series $(W_t)$ stationary, with zero mean and ACF $\gamma_W()$ and $(X_t)$ such that $X_t = X_t\psi(B)W_t \ \forall t \in \mathbb{Z}$. Then*

*1.* $\mathbb{E}[X_t] = 0 \ \forall t \in \mathbb{Z}$

*2.* $(X_t)$ *is stationary*

*3. the ACF of* $X_t$ *is such that*

$$\gamma_X(h) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \psi_i\psi_k\gamma_W(h+k-j) \quad \forall h \in \mathbb{Z}$$

*Proof.* We know that $(W_t)$ is stationary with zero mean, so $Var(W_t) = \mathbb{E}\left[W_t^2\right] = \sigma^2 > 0$ and $\mathbb{E}\left[|W_t|\right] \leq \left(\mathbb{E}\left[W_t^2\right]\right)\frac{1}{2} = \sigma$. Then

$$\sum_{j\in\mathbb{Z}} \mathbb{E}\left[|\psi_j W_{t-j}|\right] = \sum_{j\in\mathbb{Z}} |\psi_j|\,\mathbb{E}\left[|W_{t-j}|\right] \leq \sum_{j\in\mathbb{Z}} |\psi_j|$$

From the lemma of theorem 9.1:

$$\mathbb{E}\left[\sum_{j\in\mathbb{Z}}\psi_j W_{t-j}\right] = \sum_{j\in\mathbb{Z}}\psi_j\mathbb{E}\left[W_{t-j}\right] = 0$$

proving the first point of the theorem. Now we will prove the second and third points together. Consider

$$\begin{aligned}\gamma_X(t+h,t) &= \mathbb{E}\left[X_{t+h}X_t\right] - \mathbb{E}\left[X_{t+h}\mathbb{E}\left[X_t\right]\right] \\ &= \mathbb{E}\left[\left(\sum_{j\in\mathbb{Z}}\psi_j W_{t+h-j}\right)\left(\sum_{k\in\mathbb{Z}}\psi_k W_{t-k}\right)\right] \\ &= \mathbb{E}\left[\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\psi_j\psi_k W_{t+h-j}W_{t-k}\right]\end{aligned}$$

We can exchange the exèected value and the sums if

$$\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} |\psi_j|\,|\psi_k|\,\mathbb{E}\left[|W_{t+h-j}W_{t-h}|\right] < \infty$$

Applying Cauchy-Schwartz:

$$\begin{aligned}\mathbb{E}\left[|W_{t+h-j}W_{t-k}|\right] &\leq \left(\mathbb{E}\left[W_{t+h-j}^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[W_{t-k}^2\right]\right)^{\frac{1}{2}} \\ &\leq \sigma^2\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}|\psi_j|\,|\psi_k| \\ &= \sigma^2\left(\sum_{j\in\mathbb{Z}}|\psi_j|\right)\left(\sum_{k\in\mathbb{Z}}|\psi_k|\right) \\ &= \sigma^2\left(\sum_{j\in\mathbb{Z}}|\psi_j|\right)^2 < \infty\end{aligned}$$

We can now exhange the expectation and the summations:

$$\mathbb{E}\left[X_{t+h}X_t\right] = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\psi_j\psi_k\mathbb{E}\left[W_{t+h-j}W_{t-k}\right] = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\psi_j\psi_k\gamma_W(h-j+k)$$

for $h = 0$:

$$\begin{aligned}
\mathbb{E}\left[X_t^2\right] &= \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \psi_j\psi_k\gamma_W(h-j+k) \\
&\leq \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} |\psi_j|\,|\psi_k|\,|\gamma_W(h-j+k)| \\
&\leq \gamma_W(0) \\
&= \sigma^2 < \infty
\end{aligned}$$

Thus showing that $X_t \in \mathcal{L}^2$, concluding the proof. □

**Remark 9.2.** *Suppose $(Y_t)$ such that $Y_t = \mu + \psi(B)W_t\ \forall t \in \mathbb{Z},\ \mu \in \mathbb{R}$ and $(W_t)$ stationary, with zero mean and ACF $\gamma_W$. Then theorem 9.3 applies to $(X_t)$ such that*

$$X_t = Y_t - \mu$$

*and*

1. *$Y_t \in \mathcal{L}^2 \Leftarrow X_t \in \mathcal{L}^2\ \forall t \in \mathbb{Z}$*
2. *$\mathbb{E}[Y_t] = \mu\ \forall t \in \mathbb{Z}$*
3. *$cov(Y_{t+h}, Y_t) = \mathbb{E}[(Y_{t+h}-\mu)(Y_t-\mu)] = \mathbb{E}[X_{t+h}X_t] =$ point 3. of 9.3*

**Corollary 9.1.** *1. A linear time series is stationary, zero mean and has ACF*

$$\gamma(h) = \sigma^2\sum_{j\in\mathbb{Z}}\psi_j\psi_{j-h} \quad \forall h \in \mathbb{Z}$$

*2. A $MA(\infty)$ is stationary, zero mean and has ACF*

$$\gamma(h) = \sigma^2\sum_{j\geq 0}\psi_j\psi_{j+|h|} \quad \forall h \in \mathbb{Z}$$

*Proof.* Denote with $(X_t)$ the linear time series. Then we know that $X_t = \psi(B)W_t\ \forall t \in \mathbb{Z}$ with $(W_t) \sim \mathcal{WN}(0, \sigma^2)$. From theorem 9.3 we know that it is stationary, has zero mean and has ACF

$$\gamma_X(h) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \psi_j\psi_k\gamma_W(h-j+k)$$

As we know that

$$\gamma_W(h-j+k) = \begin{cases} \sigma^2 & if\ k = j-h \\ 0 & if\ k \neq j-h \end{cases}$$

then we can rewrite the ACF as

$$\begin{aligned}
\gamma_X(h) &= \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \psi_j\psi_k\gamma_W(h-j+k) \\
&= \sum_{j\in\mathbb{Z}} \psi_j\psi_{j-h}\gamma_W 0 \\
&= \sigma^2\sum_{j\in\mathbb{Z}}\psi_j\psi_{j-h}
\end{aligned}$$

proving point 1. Passing to point 2, now $(X_t) \sim MA(\infty)$; we know that this is a linear time series from the previous point but, for $h \geq 0$, $\psi_j \psi_{j-h} \neq 0 \iff j \geq 0$ *and* $j - h \geq 0$. Then the ACF is

$$\gamma_X(h) = \sigma^2 \sum_{j \geq h} \psi_j \psi_{j-h} = \sigma^2 \sum_{k \geq 0} \psi_{k+h} \psi_k = \sigma^2 \sum_{k \geq 0} \psi_k \psi_{k+|h|}$$

and for $h \leq 0$, $\psi_j \psi_{j-h} \neq 0 \iff j \geq 0$:

$$\gamma_X(h) = \sigma^2 \sum_{j \geq 0} \psi_j \psi_{j-h} = \sigma^2 \sum_{j \geq 0} \psi_j \psi_{j+|h|}$$

concluding the proof. □

Note that the moving average of order q is a special case of $MA(\infty)$ and that:

- it is a causal model;
- $(W_t) \sim \mathcal{IID}(0, \sigma^2) \implies MA(q)$ is strong stationary;
- has ACF
$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \phi_j \phi_{j+|h|} & |h| \leq q \\ 0 & |h| > q \end{cases}$$

  *Proof.* If $(X_t) \sim MA(\infty)$ then
$$\gamma_X(h) = \sigma^2 \sum_{j \geq 0} \psi_j \psi_{j+h}$$

  For $h \geq 0$ $\psi_j \psi_{j+h} \neq 0 \iff 0 \leq j \leq q$ *and* $0 \leq j + h \leq q$ then
$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \phi_j \phi_{j+h} & h = 0, ..., q \\ 0 & h > q \end{cases}$$

  For $h \leq 0$ $\psi_j \psi_{j-h} \neq 0 \iff 0 \leq j \leq q$ *and* $0 \leq j - h \leq q$ then
$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q+h} \psi_j \psi_{j-h} & h = -q, ..., 0 \\ 0 & h < -q \end{cases}$$

  □

- has the mean-ergodic property as $\lim_h \gamma_X(h) = 0 \Leftarrow \gamma_X(h) = 0 \;\; |h| > q$
- If $(X_t) \sim MA(\infty) \Leftarrow \{\psi_j\}$ from data

**Exercise 9.3.** *Check these ACFs $\rho_X(h)$ of $MA(q)$:*

1. $q = 1$: $\rho_X(\pm 1) = \frac{\phi_1}{1+\phi_1^2}$ *and* $\rho_X(h) = 0 \;\; |h| > 1$
2. $q = 2$: $\rho_X(h) = 0 \;\; |h| > 2$
   - $\rho_X(\pm 2) = \frac{\phi_2}{1+\phi_1^2+\phi_2^2}$

- $\rho_X(\pm 1) = \frac{\phi_1(1+\phi_2)}{1+\phi_1^2+\phi_2^2}$

*3.* $q = 3$*:* $\rho_X(h) = 0 \;\; |h| > 3$

- $\rho_X(\pm 3) = \frac{\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2}$
- $\rho_X(\pm 2) = \frac{\phi_2+\phi_1\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2}$
- $\rho_X(\pm 1) = \frac{\phi_1+\phi_1\phi_2+\phi_2\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2}$

# 10 Lecture 10

**Case studies in R of MA(1). The Wold's decomposition: deterministic t.s. and its meaning within forecasting. Invertible t.s. and its relation with identifiability of a model. Algebraic rules to work with the multiplicative inverse of an operator. Autoregressive t.s. of order p: the case p=1**

**Example 10.1.** *Simulate a path of*

- $MA(1) : X_t = W_t - 0.2W_{t-1}$
- $MA(2) : X_t = W_t - 0.2W_{t-1} + 0.6W_{t-2}$
- $MA(3) : X_t = W_t - 0.2W_{t-1} + 0.6W_{t-2} + 0.9W_{t-3}$

*Suppose that* $(W_t) \sim \mathcal{GWN}(0, \sigma^2)$*. Set the seed=154, n=200 and compare the estimated ACF with the theoretical one.*

```
####################################
# Simulation of paths from MA(q)
####################################

set.seed(154)
phi1=-0.2
phi2=0.6
phi3=0.9

# simulation of MA(-0.2), MA(-0.2,0.6), MA(-0.2,0.6,0.9)
ma.sim1=arima.sim(list(ma=phi1),200)
ma.sim2=arima.sim(list(ma=c(phi1,phi2)),200)
ma.sim3=arima.sim(list(ma=c(phi1,phi2,phi3)),200)

# plot the three MA models
windows()
plot(cbind(ma.sim1,ma.sim2,ma.sim3),type='o',main='MA(1),
MA(2),
MA(3)')

# ACF plots
library(astsa)
windows()
par(mfrow=c(3,1))
out1=acf1(ma.sim1,max.lag=100,main='MA(1)')
out2=acf1(ma.sim2,max.lag=100,main='MA(2)')
out3=acf1(ma.sim3,max.lag=100,main='MA(3)')

# MA(1)
(rho1MA1=phi1/(1+phi1^2))
out1[1:3]

#MA(2)
```

```
den1=1+phi1^2+phi2^2
cbind(rho1MA2=(phi1*(1+phi2))/den1, rho2MA2=phi2/den1)
out2[1:4]

#MA(3)
den2=1+phi1^2+phi2^2+phi3^2
cbind(rho1MA3=(phi1+phi1*phi2+phi2*phi3)/den2,
rho2MA3=(phi2+phi1*phi3)/den2, rho3MA3=phi3/den2)
out3[1:5]
```

**Definition 10.1.** *The **Wold's decomposition** states that any stationary time series $(X_t)$ can be decomposed in*

$$X_t = \sum_{j\geq 0} \psi_j W_{t-j} + Z_j$$

*where $\sum_{j\geq 0} \psi_j W_{t-j} = MA(\infty)$ and $Z_t$ a deterministic time series. Note that $cov(W_t, Z_s) = 0 \ \forall t, s \in \mathbb{Z}$.*

However, what is a deterministic time series? Suppose to want to predict the value of $Z_{t+h}$ based on $z_{T-1}, ..., Z_{t-n}$; an idea would be to use a linear combination fo the available data in order to predict the unknown one: chose a set $\left\{\alpha_{i,n}^{(h)}\right\}_{i=0}^{n}$ that minimizes

$$\mathbb{E}\left[\left(Z_{t+h} - \alpha_{0,n}^{(h)} - \alpha_{1,n}^{(h)} Z_{t-1} + ... + \alpha_{n,n}^{(h)} Z_{t-n}\right)^2\right]$$

then the variable

$$P_{\{Z_{t-1},...,Z_{t-n}\}}(Z_{t+h}) = \alpha_{0,n}^{(h)} - \alpha_{1,n}^{(h)} Z_{t-1} + ... + \alpha_{n,n}^{(h)} Z_{t-n}$$

is called the **orthogonal projection**.

**Definition 10.2.** *If $(Z_t)$ stationary, the orthogonal projection of $Z_{t+h}$ on $Z_{t-1}, Z_{t-2}, ...$ is*

$$P_{\{Z_{t-1},Z_{t-2},...\}}(Z_{t+h}) \overset{def}{=} \lim_n P_{\{Z_{t-1},...,Z_{t-n}\}}(Z_{t+h})$$

**Definition 10.3.** *A time series $(Z_t)$ is **deterministic** if*

$$P_{\{Z_{t-1},Z_{t-2},...\}}(Z_t) = Z_t$$

**Example 10.2.** *Consider a time series $(Z_t)$ such that $Z_t = A\cos(\omega t) + B\sin(\omega t)$ with $\omega \in (0, \pi)$ and $A, B$ uncorrelated random variables such that $\mathbb{E}[A] = \mathbb{E}[B] = 0$, $\mathbb{E}\left[A^2\right] = \mathbb{E}\left[B^2\right] = \sigma^2$. Then:*

- *$(Z_t)$ is stationary*
- *$P_{\{Z_{t-1},Z_{t-2},...\}}(Z_t) = 2\cos\omega Z_{t-1} - Z_{t-2} = Z_t \implies (Z_t)$ is deterministic*

**Definition 10.4.** *Suppose $(X_t)$ a zero mean, stationary non deterministic time series. $(X_t)$ is said **purely non deterministic** if $Z_t = 0 \ \forall t \in \mathbb{Z}$.*

An example of a purely non-deterministic time series is $MA(\infty)$.

**Exercise 10.1.** *Consider $(X_t)$ such that $X_t = W_t + Y$ where $(W_t) \sim \mathcal{WN}(0, \sigma^2)$, $Y$ is a random variable with $\mathbb{E}[Y] = 0$, $\mathbb{E}\left[Y^2\right] = \sigma^2$ and $\cos(W_t, Y) = 0 \; \forall t \in \mathbb{Z}$. Give the Wold decomposition of $X_t$.*

**Theorem 10.1.** *Any zero-mean non deterministic stationary time series $(X_t)$ can be decomposed as*

$$X_t = \sum_{j \geq 0} \psi_j W_{t-j} + Z_t$$

*where*

1. $\psi_0 = 1, \; \sum_{j \geq 0} \psi_j^2 < \infty$
2. $(W_t) \sim \mathcal{WN}(0, \sigma^2)$
3. *$cov(W_t, Z_s) = 0 \; \forall t, s \in \mathbb{Z}$*
4. *$W_t$ is the limit of the linear combination of $X_s, \; s < t$*
5. *$(Z_t)$ is a deterministic time series*

**Example 10.3.** *Consider a function*

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\phi_1}{1+\phi_1^2} & h = \pm 1 \qquad \phi_1 \in \mathbb{R} \\ 0 & |h| > 1 \end{cases}$$

*This is the ACF of $X_t = W_t + \phi_1 W_{t-1}$ with $(W_t) \sim \mathcal{WN}(0, \sigma^2)$, but also of $X_t' = W_t + \frac{1}{\phi_1} W_{t-1}$ with $(W_t) \sim \prime, \sigma^\in$, since*

$$\frac{\phi_1}{1 + \phi_1^2} = \frac{\frac{\phi_1}{\phi_1^2}}{\frac{1}{\phi_1^2} + 1} = \frac{\frac{1}{\phi_1}}{1 + \frac{1}{\phi_1^2}}$$

*Note that $(X_t)$ is stationary and $\mathbb{E}[X_t] = 0$, and also $(X_t')$ is stationary and $\mathbb{E}[X_t'] = 0$. Can we state state that*

$$Var(X_t) \overset{?}{=} Var(X_t')$$

*Start by noting that*

$$\gamma_X(0) = \sigma^2(1 + \phi_1^2) \neq \gamma_{X'}(0) = \sigma^2(1 + \frac{1}{\phi_1^2})$$

*that are clearly different; by we can try to work on this a bit. Introduce a new time series $X_t'' = W_t' + \frac{1}{\phi_1} W_{t-1}'$ with $(W_t) \sim \mathcal{WN}(0, \sigma^2)$. In order to have*

$$Var(X_t) = \sigma^2(1 + \phi_1^2) = (\sigma')^2(1 + \frac{1}{\phi_1^2}) = Var(X_t'')$$

*some algebra shows us that $(\sigma')^2 = \sigma^2 \phi_1^2$. So the function $\rho(h)$ is the ACF of $X_t$ and $X_t''$, so there is not a unique model for the ACF, and this is a problem: if we estimate the model from the data, which one we have to chose? Consider the model A:*

$$X_t = W_t + \phi_1 W_{t-1}$$

*we recover that*

$$\begin{aligned}
W_t &= X_t - \phi_1 W_{t-1} \\
&= X_t - \phi_1(X_{t-1} - \phi_1 W_{t-2}) \\
&= X_t - \phi_{t-1} + \phi_1^2 W_{t-2} \\
&\dots \\
&= \sum_{j=0}^{n-1} (-\phi_1)^j X_{t-j} + (-\phi_1)^n W_{t-n}
\end{aligned}$$

*If* $|\phi_1| < \infty$ *then* $\sum_{j\geq 0} \left|(\phi_1)^j\right| < \infty$*. We might then consider*

$$\lim_n \sum_{j=0}^{n-1} (-\phi_1)^j X_{t-j} + \lim_n (-\phi_1)^n W_{t-n} = \Phi(B) X_t$$

*with* $\Phi(Z) = \sum_{j\geq 0} (-\phi_1)^j Z^j$*. Thus* $W_t = \Phi(B)X_t$*. Consider now the series B:*

$$X_t'' = W_t' + \frac{1}{\phi_1} W_{t-1}'$$

*Applying the same arguments:*

$$W_t' = \sum_{j=0}^{n-1} \left(-\frac{1}{\phi_1}\right)^j X_{t-j}'' + \left(-\frac{1}{\phi_1}\right)^n W_{t-n}'$$

*If* $|\phi_1| > 1$ *then* $\left|\frac{1}{\phi_1}\right| < 1$*, so* $\sum_{j\geq 0} \left|\left(-\frac{1}{\phi_1}\right)^j\right| < \infty$*. Studying the limit:*

$$\lim_n \sum_{j=0}^{n-1} \left(-\frac{1}{\phi_1}\right)^j X_{t-j}'' + \lim_n \left(-\frac{1}{\phi_1}\right)^n W_{t-n}' = \tilde{\Phi}(B) X_t''$$

*with* $\tilde{\Phi}(Z) = \sum_{j\geq 0} \left(-\frac{1}{\phi_1}\right)^j Z^j$*. Therefore for* $|\phi_1| < 1$ *then* $(W_t)$ *is perfectly predictable from* $(X_s)$ *for* $s \leq t$*, so* $(X_t)$ *is invertible. If* $|\phi_1| > 1$ *this is no longer true, and we have proved that there exists a different model* $(X_t'')$ *which is invertible having the same statistical features of* $(X_t)$ *up to the second order. But why the first case is better than the second? Why is better to work with* $|\phi_1| < 1$*? The answer is the following:*

- *if* $|\phi_1| < 1$ *the most recent observations have higher weight than the mode distant ones;*
- *if* $|\phi_1| = 1$ *all observations have the same weight;*
- *if* $|\phi_1| > 1$ *the more distant observations have proportionally more influence.*

*That is why it is better to work with invertible time series.*

Before continuing, rememtber some properties of the Laurent power series $\sum_{j\in\mathbb{Z}} \psi_j z^j$. Suppose:

- $\alpha(Z) = \sum_{j\in\mathbb{Z}} \alpha_j Z^j$ such that $\sum_{j\in\mathbb{Z}} |\alpha_j| < \infty$
- $\beta(Z) = \sum_{j\in\mathbb{Z}} \beta^j$ such that $\sum_{j\in\mathbb{Z}} |\beta| < \infty$

then

$$\psi(Z) = \alpha(Z)\beta(Z) = \sum_{j\in\mathbb{Z}} \psi_j Z^j$$

with $\psi_j = \sum_{k\in\mathbb{Z}} \alpha_k \beta_{j-k} = \sum_{k\in\mathbb{Z}} \beta_k \alpha_{j-k}$, $\forall j \in \mathbb{Z}$. The set $\{\psi_j\}$ is the convolution of $\{\alpha_j\}, \{\beta_j\}$ and is noted as $\{\psi_j\} = \{\alpha_j\} * \{\beta_j\}$. When exists, the multiplicative inverse of $\alpha(Z)$ is $\beta(Z)$ such that

$$\alpha(Z)\beta(Z) = \beta(Z)\alpha(Z) = 1$$

and is can be notated as $\beta(Z) = \alpha^{-1}(Z)$. We might apply this rules to the operations in between operators:

$$\psi(B) = \alpha(B)\beta(B) \implies \psi(B) = \sum_{j\in\mathbb{Z}} \psi_j B^j$$

with B the backshift operator. In this case $\psi_j$ stands for $\{\psi_j\} = \{\alpha_j\} * \{\beta_j\}$. Then $\alpha^1(B)$ is such that $\alpha^{-1}(B)\alpha(B) = \alpha(B)\alpha^{-1}(B) = 1$.

**Example 10.4.** *Consider $(X_t) \sim MA(\infty)$, then $X_t = \phi(B)W_t$ with $\phi(B) = 1 + \phi_1 B$. Recover $(W_t)$ from $(X_t)$. Consider $\phi(Z) = 1 + \phi_1 Z$. If $|\phi_1| < 1$ then $\phi^{-1} = \frac{1}{1+\phi_1 Z} = \sum_{j\geq 0}(-\phi_1)^j Z^j$. The multiplicative inverse of $\phi(B)$ is $\phi^{-1}(B)$ such that $\phi^{-1}(B) = \sum_{j\geq 0}(-\phi_1)^j B^J$. Consider now that $X_t = \phi(B)W_t$ implies that $\phi^{-1}(B)X_t = \phi^{-1}(B)\phi(B)W_t = W_t = \phi^{-1}(B)X_t = \sum_{j\geq 0}(-\phi_1)^j X_{t-j}$.*

**Definition 10.5.** *A linear time series $(X_t)$ such that $X_t = \psi(B)W_t$ is* ***invertible*** *if $\exists \psi^{-1}(Z)$*

$$\psi^{-1}(Z) = \sum_{j\in\mathbb{Z}} \psi'_j Z^j$$

*such that $\sum_{j\in\mathbb{Z}} |\psi'_j| < \infty$.*

A widely used stochastic process is the **autoregressive time series** (AR(p)), in which the current value of the process is expressed as a finite linear aggregate of previous values of the process plus a random shock.

**Definition 10.6.** *A time series $(X_t) \sim AR(p)$ is a stationary solution of*

$$X_t = \theta_1 X_{t-1} + ... + \theta_p X_{t-p} + W_t$$

*where $(W_t) \sim \mathcal{WN}(0, \sigma^2)$, $p \in \mathbb{N}$, $\theta_1, ..., \theta_p \in \mathbb{R}$ and $\theta_p \neq 0$.*

**Example 10.5.** *An example of a non-stationary solution of autoregression: consider $p = 1$ and a random walk $X_t = X_{t-1} + W_t$; this stochastic process is a solution to the autoregression ($\theta_1 = 1$) but is not stationary, since it is a random walk.*

If we define the autoregressive polinomial

$$\theta(Z) = 1 - \theta_1 Z + ... - \theta_p Z^p$$

then we can use the notation

$$X_t = \theta_1 X_{t-1} + ... + \theta_p X_{t-p} + W_t \implies \theta(B)X_t = W_t$$

We assume that $\mathbb{E}[X_t] = 0 \ \forall t \in \mathbb{Z}$ since $\mathbb{E}[X_t] = \mu \neq 0 \implies Y_t = X_t - \mu$ leading to

$$\begin{aligned} X_t &= \mu + \theta_1(X_{t-1} - \mu) + ... + \theta_p(X_{t-p} - \mu) + W_t \\ &= (\mu + \theta_1\mu + ... + \theta_p\mu) + \theta_1 X_{t-1} + ... + \theta_p X_{t-p} + W_t \\ &= a + \theta_1 X_{t-1} + ... + \theta_p X_{t-p} + W_t \end{aligned}$$

with our assumption, $a = 0$.

**Remark 10.1.** *Suppose $p = 1$, $(X_t) \sim AR(p)$ with initial condition $X_0$ such that $cov(X_0, W_t) = 0 \ \forall t \in \mathbb{Z}$. We have that*

$$X_1 = \theta_1 X_0 + W_1, \ X_2 = \theta_1 X_1 + W_2 = \theta_1^2 X_0 + \theta_1 W_1 + W_2$$

*in general:*

$$X_n = \theta_1^n X_0 + \theta_1^{n-1} W_1 + ... + \theta_1 W_{n-1} + W_n$$

*So:*

$$\mathbb{E}[X_n] = \theta_1^n \mathbb{E}[X_0] \Leftarrow \mathbb{E}[X_0] = 0, \ \mathbb{E}[X_n] = 0 \quad n = 1, 2, 3...$$

*Now let us evaluate the variance:*

$$\begin{aligned} Var(X_n) &= \theta_1^n Var(X_0) + \theta_1^{n-1} Var(W_1) + ... + Var(W_n) \\ &= \theta_1^n Var(X_0) + \sigma^2(1 + \theta_1 + ... + \theta_1^{n-1}) \end{aligned}$$

*As you can see the variance depends on n, so $(X_n)$ is not stationary. Also if we ask $Var(X_0) = 0$ the other term of the equation is different from zero.*

Let us consider the $AR(1)$ model $X_t = \theta_1 X_{t-1} + W_t$ with $\theta_1 \neq 0, 1$. Apply the substitution $X_{t-1} = \theta_1 X_{t-2} + W_{t-1}$ iteratively:

$$X_t = \theta_1^n X_{t-n} + \sum_{j=0}^{n-1} \theta_1^j Wt - j$$

If $|\theta_1| < 1$ then $\sum_{j \geq 0} |\theta_1|^j < \infty$. Also, $(W_t)$ is $\mathcal{L}^2$ bounded since

$$\begin{aligned} X_t &= \lim_n \sum_{j=0}^{n-1} \theta_1^j W_{t-j} + \lim_n \theta_1^n X_{t-n} \\ &= \sum_{j \geq 0} \theta_1^j W_{t-j} \Leftarrow MA(\infty) \end{aligned}$$

So $(X_t)$ is causal and its ACF is

$$\gamma_X(h) = \sigma^2 \sum_{j \geq 0} \theta_1^j \theta_1^{j+|h|} = \sigma^2 \theta_1^{|h|} \sum_{j \geq 0} (\theta_1^2)^j = \sigma^2 \theta_1^{|h|} \frac{1}{1 - \phi_1^2}$$

which tends to 0 as $h \to \infty$, so $(X_t)$ has the mean ergodicity property. This is the expression of the ACF of an AR(1) when $|\theta_1| < 1$. Moreover, $(X_t)$ is the only solution of the different equation since it is the limit in almost sure convergence of the partial sums. It is also the limit in $\mathcal{L}^2$ convergence of a filtering of white noise.

**Proposition 10.1.** *$X_t = \theta^{-1}(B)W_t$ with $\theta^{-1}(Z)$ such that*

$$\theta^{-1}(Z)\theta(Z) = \theta(Z)\theta^{-1}(Z) = 1$$

*Proof.* Consider the autoregressive polinomial

$$\theta(Z) = 1 - \theta_1 Z \implies \theta^{-1}(Z) = \frac{1}{1 - \phi_1 Z} \quad z \neq \frac{1}{\phi_1}$$

If $|\theta_1 Z| < 1$ then we can state that

$$\theta^{-1}(Z) = \sum_{j \geq 0} (\theta_1 Z)^j$$

Thus $\theta^{-1}(B)W_t = \sum_{j\geq 0} \theta_1^j B^j W_t = \sum_{j \geq 0} \theta_1^j W_{t-j} = X_t$. □

**Remark 10.2.** *If $|\theta_1| < 1$ then $\frac{1}{|\theta_1|} > 1$, then the root of $\theta(Z)(Z_0 = \frac{1}{\phi_1})$ is out of the interval $(-1, 1)$.*

What happens if $|\theta_1| > 1$? The theorems we have seen would not hold anymore; would there be nonetheless solutions to the difference equation that are stationary? Let us consider the difference equation

$$X_{t+1} = \theta_1 X_t + W_{t+1} \implies X_t = \frac{1}{\theta_1} X_{t+1} - \frac{1}{\theta_1} W_{t+1}$$

Applying iteratively the substitution $X_{t+1} = \frac{1}{\theta_1} X_{t+2} - \frac{1}{\theta_1} W_{t+2}$ we obtain that

$$X_t = \frac{1}{\theta_1^n} X_{t+n} - \sum_{j=1}^{n} \frac{1}{\theta_1^j} W_{t-j}$$

As $\left|\frac{1}{\theta_1}\right| < 1$ then $\sum_{j \geq 1} \left|\frac{1}{\theta_1}\right|^j < \infty$ and $(W_t) \in \mathcal{L}^2$-bounded we can consider the limit

$$X_t = \lim_n \frac{1}{\theta_1^n} X_{t+n} - \lim_n \sum_{j=1}^{n} \frac{1}{\theta_1^j} W_{t-j} = -\sum_{j \geq 1} \sum_{j=1}^{n} \frac{1}{\theta_1^j} W_{t-j}$$

Then $(X_t)$ has this representation as a linear filter of $(W_t)$; it is also stationary, but it is not causal because in this case it depends on the future.

**Exercise 10.2.** *If $|\theta_1| > 1$ find the ACF of $(X_t) \sim AR(1)$.*

**Exercise 10.3.** *If $|\theta_1| < 1$ prove that*

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

*is the solutin of $X_t = \theta_1 X_{t-1} + W_t$. If $|\theta_1| > 1$ find the solution of $X_t = \theta_1 X_{t-1} + W_t$ in terms of $B$ backshift operator and check that $X_t$ is the solution of $X_t = \theta_1 X_{t-1} + W_t$.*

# 11 Lecture 11

**Simulation and analysis of AR(1) models. Multivariate representation of AR(p) model and causality. Yule-Walker's equations for the covariance function of AR(p) models. Case study: p=2. Simulation and analysis of AR(2) models. Introduction to ARMA models.**

Suppose to recover an estimation of the parameter $|\theta_1| > 1$ from the data. Using a suitable update of the data is possible to use the model

$$X_t = \frac{1}{\theta_1} X_{t-1} + W'_t \quad \forall t \in \mathbb{Z}$$

?

**Exercise 11.1.** *Suppose $|\theta_1| > 1$. Define*

$$W'_t = X_t - \frac{1}{\theta_1} X_{t-1}$$

*Show that $(W'_t) \sim \mathcal{WN}(0, \sigma^2_W)$ and express $\sigma^2_W$ in terms of $\sigma^2$ and $\theta_1$.*

**Example 11.1.** *Simulate a path for the following $AR(1)$ models:*

- $X_t = 0.9X_{t-1} + W_t$
- $X_t = 0.4X_{t-1} + W_t$
- $X_t = -0.9X_{t-1} + W_t$

*with $(W_t) \sim \mathcal{GWN}(0, 1)$, $n = 200$ and seed $= 154$. Comment the resutls.*

```
####################################
# Simulation of AR(1) paths
####################################

# simulation of AR(0.9), AR(0.4), AR(-0.9)
set.seed(154)
ar.0.9=arima.sim(list(ar=0.9),200)
ar.0.4=arima.sim(list(ar=c(0.4)),200)
ar.m0.9=arima.sim(list(ar=c(-0.9)),200)

# plot the paths
windows()
par(mfrow=c(3,1))
plot(ar.0.9,type='o',main='AR(0.9)',ylim=c(-6,6))
abline(h=0,col='red')
plot(ar.0.4,type='o',main='AR(0.4)',ylim=c(-6,6))
abline(h=0,col='red')
plot(ar.m0.9,type='o',main='AR(-0.9)',ylim=c(-6,6))
abline(h=0,col='red')

# plot ACF
```

```
windows()
par(mfrow=c(1,3))
acf(ar.0.9, main='AR(0.9)',ylim=c(-1,1))
acf(ar.0.4, main='AR(0.4)',ylim=c(-1,1))
acf(ar.m0.9, main='AR(-0.9)',ylim=c(-1,1))

# plot lag.plot
graphics.off()
windows()
lag.plot(ar.0.9,9,main='AR(0.9)')
windows()
lag.plot(ar.m0.9,9,main='AR(-0.9)')

#
# adf test (augmented Dickey-Fuller test) in the library tseries
#
library(tseries)
adf.test(ar.0.9,k=0)
adf.test(ar.0.4,k=0)
adf.test(ar.m0.9,k=0)
```

The **augmented Dickey-Fuller (ADF) test** tests the null hypothesis that the autoregressive polynomial has a unit root. If $k = 1$ the null hypothesis is that the time series is a random walk with zero drift, and the alternative hypothesis is that the series is a $AR(1)$ with $|\theta_1| < 1$; otherwise, the test refers to a $AR(p)$ with the null hypothesis that 1 is a root of the autoregressive polynomial $(1 = \hat{\theta_1} + ... + \hat{\theta_p})$.

What about the autoregressive models of order p greater than one? We will see that if $(X_t) \sim AR(p)$ then it has a causal representation if and only if $\theta(Z) \neq 0 \ \forall Z \in \mathbb{C}$ such that $|Z| \leq 1$.

*Proof.* Consider $p = 3$ and the different equation $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \theta_3 X_{t-3} + W_t$. Suppose to add two more equations to this one:

$$\begin{cases} X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \theta_3 X_{t-3} + W_t \\ X_{t-1} = X_{t-1} \\ X_{t-2} = X_{t-2} \end{cases}$$

This can be rewritten in matrix notation:

$$\begin{pmatrix} X_t \\ X_{t-1} \\ X_{t-2} \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ X_{t-3} \end{pmatrix} + \begin{pmatrix} X_t \\ 0 \\ 0 \end{pmatrix}$$

In general:

- $\boldsymbol{x}_{t,p} = (X_t, ..., X_{t-p+1})$
- $\boldsymbol{w}_{t,p} = (W_t, 0, ..., 0)$
- $\boldsymbol{A}_{p\times p} = \left(\begin{array}{ccc|c} \theta_1 & ... & \theta_{p-1} & \theta_p \\ \hline & \boldsymbol{I}_{p-1} & & 0 \end{array}\right)$

Then the following set of equations can be written as

$$\begin{cases} X_t = \theta_1 X_{t-1} + ... + \theta_p X_{t-p} + W_t \\ X_{t-1} = X_{t-1} \\ ... \\ X_{t-p+1} = X_{t-p+1} \end{cases} \qquad \implies \boldsymbol{x}_{t,p}^\intercal = \boldsymbol{A}\boldsymbol{x}_{t-1,p}^\intercal + \boldsymbol{w}_{t,p}^\intercal$$

Note that $det(\boldsymbol{I}_p - Z\boldsymbol{A}) = \theta(Z)$ (check as exercise for p=3) and $P_{\boldsymbol{A}}(Z) = det(Z\boldsymbol{I}_p - \boldsymbol{A})$. Observe that $z_0 = 0$ is such that $\theta(z_0) \neq 0 \Leftarrow \theta(0) = 1$. So we can rewrite the autoregressive polynomial as

$$\theta(Z) = det(\boldsymbol{I}_p - Z\boldsymbol{A}) = Z^p det\left(\frac{1}{Z}\boldsymbol{I}_p - \boldsymbol{A}\right) = Z^p P_{\boldsymbol{A}}\left(\frac{1}{Z}\right)$$

substituting $Z$ with $\frac{1}{Z}$:

$$\theta\left(\frac{1}{Z}\right) = \frac{1}{Z^p} P_{\boldsymbol{A}}(Z)$$

This implies that the eigenvalues of $\boldsymbol{A}$ are equal to the inverse of the roots of $\theta(Z)$. Now consider to iterate the difference equation $n$ times:

$$\begin{aligned} \boldsymbol{x}_{t,p}^\intercal &= \boldsymbol{A}\boldsymbol{x}_{t-1,p}^\intercal + \boldsymbol{w}_{t,p}^\intercal \\ &= \boldsymbol{A}(\boldsymbol{A}\boldsymbol{x}_{t-2,p}^\intercal + \boldsymbol{w}_{t-1,p}^\intercal) + \boldsymbol{w}_{t,p}^\intercal \\ &= \boldsymbol{A}^2\boldsymbol{x}_{t-2,p}^\intercal + \boldsymbol{A}\boldsymbol{w}_{t-1}^\intercal + \boldsymbol{w}_{t,p}^\intercal \\ &... \\ &= \boldsymbol{A}^n\boldsymbol{x}_{t-n,p}^\intercal + \sum_{j=0}^{n-1} \boldsymbol{A}^j \boldsymbol{w}_{t-j,p}^\intercal \end{aligned}$$

For $n \to \infty$ this expression depends on the eigenvalues of $\boldsymbol{A}$. In particular, if we set

$$\tilde{\rho}(\boldsymbol{A}) = \max\{|\lambda_1|, ..., |\lambda_p|\}$$

then it is possible to prove that the spectral radius ($\tilde{\rho}$) of $\boldsymbol{A}$ is $< 1$ if and only if $\lim_n \boldsymbol{A}^n = 0$. In this case it is possible to prove that

$$\lim_{n\to\infty} \sum_{j=0}^{n-1} \boldsymbol{A}^j \boldsymbol{w}_{t-j,p}^\intercal = \boldsymbol{x}_{t,p}^\intercal$$

is well defined. So in this case the vector admits a causal representation; this is true only when

$$\tilde{\rho}(\boldsymbol{A}) < 1 \iff |\lambda_1|, ..., |\lambda_p| < 1 \iff \frac{1}{|\lambda_1|}, ..., \frac{1}{|\lambda_p|} > 1$$

but among these inverses there are the roots of $theta(Z)$, but this is the statement from which we have started. □

But what about the ACF of $AR(p)$?

**Theorem 11.1.** *If $(X_t) \sim AR(p)$, then*

$$\gamma_X(h) - \theta_1\gamma_X(h-1) + ... - \theta_p\gamma_X(h-p) = \begin{cases} \sigma^2 & h \neq 0 \\ 0 & h = 0 \end{cases}$$

*Proof.* Consider $(X_t) \sim AR(p)$. Then

$$X_t - \theta_1 X_{t-1} + ... - \theta_p X_{t-p} = W_t$$
$$\mathbb{E}\left[X_t X_{t-h}\right] - \theta_1\mathbb{E}\left[X_{t-1}X_{t-h}\right] + ... - \theta_p\mathbb{E}\left[X_{t-p}X_{t-h}\right] = \mathbb{E}\left[W_t X_{t-h}\right]$$
$$\gamma_X(h) - \theta_1\gamma_X(h-1) + ... - \theta_p\gamma_X(h-p) = ?$$

In order to find the missing result, observe that $X_{t-h}$ depends on

$$X_{t-h-1}, X_{t-h-2}, ..., X_{t-h-p}, W_{t-h}$$

in which we have the contribution respectively of

$$W_{t-h-1}, W_{t-h-2}, ..., W_{t-h-p}$$

but these values are in the past when $h \neq 0$, so $X_{t-h}, W_t$ are uncorrelated when $h \neq 0$. In case $h = 0$:

$$W_t X_t = \theta_1 W_t X_{t-1} + ... + \theta_p W_t X_{t-p} + W_t^2$$

taking the expectation of this, we found that only contribution of this is given by $W_t^2$, since all the other shocks are in the past. We find that

$$\mathbb{E}\left[W_t X_{t-h}\right] = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

substituting this result in the previous equation, we retrieve the Yule-Walker's equation, completing the proof. □

The **Yule-Walker's equations** are special cases of homogeneous linear difference equations with constant coefficients:

$$f(h) - \alpha_1 f(h-1) + ... - \alpha_p f(h-p) = 0$$

with $h \neq 0$ and $\alpha_1, ..., \alpha_p \in \mathbb{R}$.

**Example 11.2.** *If $\xi_1, ..., \xi_p \in \mathbb{R}$ are roots of $\theta(Z)$ with $\xi_1 \neq ... \neq \xi_p$ then*

$$\gamma_X(h) = b_1\xi_1^{-h} + ... + b_p\xi_p^{-h}$$

*with $b_1, ..., b_p$ constants depending on initial conditions.*

**Example 11.3.** *Consider the Yule-Walker's equation with $p = 2$:*

$$\gamma_X(h) = \theta_1\gamma_X(h-1) + \theta_2\gamma_X(h-2)$$

*with $h \geq 1$. Dividing every member of the equation by $\gamma_X(0)$ and considering the initial condition $\gamma_X(0) = 1$ we obtain the following system:*

$$\begin{cases} \rho_X(h) = \theta_1\rho_X(h-1) + \theta_2\rho_X(h-2) \\ \rho_X(0) = 1 \end{cases}$$

*Suppose $\xi_1, \xi_2$ be the roots of the autoregressive polynomial $\theta(Z) = 1 - \theta_1 Z - \theta_2 Z^2 = \left(1 - \frac{Z}{\xi_1}\right)\left(1 - \frac{Z}{\xi_2}\right)$. Suppose also that $\left|\xi_1\right|, \left|\xi_2\right| > 1$ (this hypotesis implies that $(X_t)$ is causal). Then*

- *if $\xi_1 \neq \xi_2 \in \mathbb{R}$ then $\rho_X(h) = b_1\xi_1^{-h} + b_2\xi_2^{-h}$ with $b_1, b_2$ depending on initial conditions, since*

  $$\begin{cases} h = 0 \quad \rho_X(0) = 1 \implies b_1 + b_2 = 1 \\ h = 1 \quad \rho_X(1) = \theta_1\rho_X(0) + \theta_2\rho_X(-1) \iff \rho_X(-1) = \rho_X(1) = \frac{\theta_1}{1-\theta_2} = \frac{b_1}{\xi_1} + \frac{b_2}{\xi_2} \end{cases}$$

  *So we can find $b_1$ and $b_2$ by solving the system*

  $$\begin{cases} b_1 + b_2 = 1 \\ \frac{b_1}{\xi_1} + \frac{b_2}{\xi_2} = \frac{\theta_1}{1-\theta_2} \end{cases}$$

  *This is left as exercise.*

- *if $\xi_1 = \xi_2 \in \mathbb{R}$ then $\rho_X(h) = \xi_1^{-h}(b_{10}h^0 + b_{11}h^1) = \xi_1^{-h}(b_{10} + b_{11}h^1)$. To find the constant we have to solve:*

  $$\begin{cases} b_{10} + b_{11} = 1 \\ \xi_1^{-1}(b_{10} + b_{11}) = \frac{\theta_1}{1-\theta_2} \end{cases}$$

*In both cases, for $h \to \infty \implies \rho_X(h) = 0$. In reality, there is one more case:*

- *if $\xi_1 = \bar{\xi}_2 \in \mathbb{C}$ then $\xi_1 = de^{-i\tilde{\theta}}, \xi_2 = de^{i\tilde{\theta}}$ with $\tilde{\theta} \in (0, 2\pi)$. In this case we have that*

  $$\rho_X(h) = b_1d^{-h}e^{i\tilde{\theta}h} + b_2d^{-h}e^{-i\tilde{\theta}h}$$

  *with $b_1 = ae^{ib}, b_2 = ae^{-ib}$. After some algebra we find that*

  $$\rho_X(h) = 2ad^{-h}\cos(b + \tilde{\theta}h)$$

  *which goes to zero as $h \to \infty$.*

**Remark 11.1.** *$\theta_2 \neq 1$. Suppose $\theta_2 = 1$, then $\rho_X(1) = \theta_1\rho_X(0) + \theta_2\rho_X(1) \implies \theta_1 = 0 \implies \theta(Z) = 1 - Z^2$, but in this case it does not exist a stationary solution. This will be demonstrated in the next lecture, but also the following exercise will help in understanding this.*

**Exercise 11.2.** *Suppose $(X_t) = AR(2)$.*

1. *prove that $\theta(Z) \neq 0 \ \forall Z \in \mathbb{C}$ such that $|Z| \leq 1$ if $\theta_1 + \theta_2 < 1, \theta_2 - \theta_1 < 1, |\theta_2| < 1$.*
2. *find a causal representation of $(X_t)$ when $\theta(Z) \neq 0 \ \forall Z \in \mathbb{C} suchthat |Z| \leq 1$.*

**Example 11.4.**

1. *Find $\theta(Z) = 1 - \theta_1Z - \theta_2Z^2$ such that $\xi_1 = -1.5$ and $\xi_2 = 2$ are roots of $\theta(Z)$;*
2. *Simulate a path of $(X_t) \sim AR(2)$ with $\theta(Z)$ given in 1. (seed=154,n=20) and plot the correlation function;*
3. *Check if $(X_t)$ has a causal representation:*
   (a) *$X_t = -1.9X_{t-1} + 0.88X_{t-2} + W_t$*
   (b) *$X_t = X_{t-1} - 0.25X_{t-2} + W_t$*

*(c)* $X_t = 1.5X_{t-1} - 0.75X_{t-2} + W_t$

*Plot the correlation functions and comment the results.*

4. *Check if 1 is a root of* $\theta(Z)$ *for* $(X_t)$ *in 3.a, 3.b, 3.c using the ADF test.*

```
####################################
# Exercise on AR(p)
####################################

# to work with polynomials
install.packages('polynom')
library(polynom)

# a)
# construct a polynomial from its roots
(c1=coef(poly.from.roots(c(-1.5, 2))))
(polynomial(c1))

# for the autoregressive polynomial 1 - theta1 x - theta2 x^2
# normalize with the constant term c1[1]
(c2=c1/c1[1])

# update the polynomial
(p=polynomial(c2))

# check the roots
(pz=solve(p))

# test on the coefficients theta1 and theta2
(coeffAR2=coef(p))
(theta1=-coeffAR2[2])
(theta2=-coeffAR2[3])
theta1+theta2; theta2-theta1; abs(theta2)

# b)
# simulate the corresponding AR(2)
set.seed(154)
ar.sim1=arima.sim(list(ar=c(theta1,theta2)),200)
windows()
plot(ar.sim1,type='o',main='AR(2): roots -1.5 and 2',ylim=c(-6,6))
abline(h=0,col='red')
windows()
acf(ar.sim1, lag.max=50, main='AR(2) roots -1.5 and 2',ylim=c(-1,1))

# c.1)
# find the roots of the AR(2) polynomial: 1 + 1.9 z - 0.88 z^2
# CAR is the vector with coefficients
CAR=c(1,1.9,-0.88)

# check
```

```
(p=polynomial(CAR))

# find the roots
(roots=solve(p))

# test on the coefficients theta1 and theta2: extract the coefficients
(coeffAR2=coef(p))
(theta1=-coeffAR2[2])
(theta2=-coeffAR2[3])
theta1+theta2; theta2-theta1; abs(theta2)

# simulate the AR(2)?
ar.sim2=arima.sim(list(ar=c(1.9,-0.88)),200)

# c.2)
# simulation of AR(2) with polynomial: 1 - z + 0.25 z^2
# check the roots
p=polynomial(c(1,-1.0,0.25))
(roots=solve(p))

# test on the coefficients theta1 and theta2
(coeffAR2=coef(p))
(theta1=-coeffAR2[2])
(theta2=-coeffAR2[3])
theta1+theta2; theta2-theta1; abs(theta2)

# simulate the AR(2)
ar.sim2=arima.sim(list(ar=c(1,-0.25)),200)
windows()
plot(ar.sim2,type=’o’,main=’AR(2): roots 2 and 2’,ylim=c(-6,6))
abline(h=0,col=’red’)

# ACF
windows()
acf(ar.sim2, lag.max=50, main=’AR(2): roots 2 and 2’,ylim=c(-1,1))

# c.3)
# simulation of AR(2) with polynomial: 1 - 1.5 z + 0.75 z^2
# check the roots
(p=polynomial(c(1,-1.5,0.75)))
(roots=solve(p))
abs(roots)

# test on the coefficients theta1 and theta2
(coeffAR2=coef(p))
(theta1=-coeffAR2[2])
(theta2=-coeffAR2[3])
theta1+theta2; theta2-theta1; abs(theta2)

# simulate the AR(2)
```

```
ar.sim3=arima.sim(list(ar=c(1.5,-0.75)),200)windows()
plot(ar.sim3,type='o',main='AR(2): complex conjugate roots',ylim=c(-6,6))
abline(h=0,col='red')

# ACF
windows()
acf(ar.sim3, lag.max=50, main='AR(2): complex conjugate roots',ylim=c(-1,1))

# plot ACF
windows()
par(mfrow=c(1,3))
acf(ar.sim1, main='AR(2): different roots',ylim=c(-1,1))
acf(ar.sim2, main='AR(2): equal roots',ylim=c(-1,1))
acf(ar.sim3, main='AR(2): complex conjugate roots',ylim=c(-1,1))

# d)
# adf test (augmented Dickey-Fuller test)
#
library(tseries)
adf.test(ar.sim2)
adf.test(ar.sim3)
```

**Definition 11.1.** *$(X_t) \sim$ **ARMA(p,q)** is a stationary solution to the difference equation*

$$X_t - \theta_1 X_{t-1} + ... - \theta_p X_p = W_t + \phi_1 W_{t-1} + ... + \phi_q W_{t-q}$$

*where $p, q \in \{0, 1, 2, ...\}$, $\theta_1, ...\theta_p, \phi_1, ..., \phi_p \in \mathbb{R}$, $\theta_p, \phi_p \neq 0$ and $(W_t) \sim \mathcal{WN}(0, \sigma^2)$. The short notation for the previous equation is*

$$\theta(B)X_t = \phi(B)W_t$$

*where the first therm is the autoregressive polynomial and the second one the moving average polynomial. Note that*

- *when $p = 0$ then $\theta(Z) = 1$ $(\theta_0 = 1) \implies ARMA(0, q) = MA(q)$*
- *when $q = 0$ then $\phi(Z) = 1$ $(\phi_0 = 1) \implies ARMA(p, 0) = AR(p)$*

*The case $p = q = 0$ is generally not considered since it reduces to the white noise. We Suppose that $\mathbb{E}[X_t] = 0$ $\forall i \in \mathbb{Z}$.*

**Exercise 11.3.** *Suppose $(X_t) \sim ARMA(2, 2)$ such that*

$$X_t = 2 + 1.3X_{t-1} - 0.4X_{t-2} + W_t + W_{t-1}$$

*Compute $\mathbb{E}[X_t]$.*

# 12 Lecture 12

**ARMA time series. Redundance. Causality of ARMA t.s. Necessary and sufficient conditions to get causality. Recursive relations to get the MA representation of infinite order. The function ARMAtoMA in R. Simulations of ARMA models in R.Invertibility. The ACF generating function. Transformation of the ARMA equation to recover a stationary solution which is causale and invertible.**

**Definition 12.1.** *$(X_t) \sim ARMA(p,q)$ not redundant if and only if*

$$\theta(Z) = 0 \quad and \quad \phi(Z) = 0$$

*have no common roots.*

**Example 12.1.** *Consider $(W_t) \sim \mathcal{WN}(0, \sigma^2)$ and the equation $X_t = W_t$ $\forall t \in \mathbb{Z}$. Then the solution of this equation is equal to the solution of the equation $0.5X_{t-1} = 0.5W_{t-1}$. Subtracting these two equations from each other we obtain that*

$$X_t - 0.5X_{t-1} = W_t - 0.5W_{t-1}$$

*which is an $ARMA(1,1)$ model. So the initial white noise has an ARMA representation, since $\theta(Z) = \phi(Z) = 1 - 0.5Z$. However, this is false: common factors between AR and MA polynomial have to be removed since there are redundant parameters in the model.*

**Example 12.2.** *Consider $X_t = 0.4X_{t-1} + 0.45X_{t-2} + W_t + W_{t-1} + 0.25W_{t-2}$ which is equal to*

$$(1 - 0.4B - 0.45B^2)X_t = (1 + B + 0.25B^2)W_t$$

*We then have that*

$$\theta(Z) = 1 - 0.4Z - 0.45Z^2 = (1 + 0.5Z)(1 - 0.9Z)$$

*and*

$$\phi(Z) = 1 + Z + 0.25Z^2 = (1 + 0.5Z)^2$$

*which have a common factor. By removing it, we obtain*

$$(1 - 0.9B)X_t = (1 + 0.5B)W_t$$

*which corresponds to $X_t = 0.9X_{t-1} + W_t + 0.5W_{t-1}$, which is an ARMA(1,1) model.*

**Exercise 12.1.** *Suppose*

1. *$A$ a random variable with $\mathbb{E}[A] = 0$ and $\mathbb{E}\left[A^2\right] < \infty$*

2. *$(X_t) \sim ARMA(p,q)$ not redundant*

3. *$cov(A, X_t) = 0$ $\forall t \in \mathbb{Z}$*

4. *$Z_0 \in \mathbb{C}$ such that $|Z_0| = 1$*

*Prove that* $(X_t)$ *and* $(X_t + AZ_0^t)$ *are both stationary solutions of*

$$(1 - Z_0B)\theta(B)X_t = (1 - Z_0B)\phi(B)W_t$$

In the following, we will assume that the ARMA models are not redundant.

**Definition 12.2.** $(X_t) \sim ARMA(p, q)$ *is causal if*

$$\exists \{\psi_j\}_{j\geq 0} \text{ with } \sum_{j\geq 0} |\psi_j| < \infty \text{ such that } X_t = \sum_{j\geq 0} \psi_j W_{t-j}$$

*with* $(W_t) \sim \mathcal{WN}(0, \sigma^2)$.

The ARMA model have a number of useful properties: as $X_t = \psi(B)W_t$ and

- $(W_t)$ is $\mathcal{L}^2$-bounded
- $\{\psi_j\}$ such that $\sum_{j\geq 0} |\psi_j| < \infty$

we have that

- $(X_t)$ is the unique solution of the ARMA equation
- ACF: $\gamma_X(h) = \sigma^2 \sum_{j\geq 0} \psi_j \psi_{j+|h|}$

**Proposition 12.1.** *If* $(X_t)$ *causal then* $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty$

*Proof.* Consider

$$\sigma^2 \sum_{h\in\mathbb{Z}} \left| \sum_{j\geq 0} \psi_j \psi_{j+|h|} \right| \leq \sigma^2 \sum_{h\in\mathbb{Z}} \sum_{j\geq 0} |\psi_j| \left|\psi_{j-|h|}\right|$$

Now observe that

- $h = 0$: $\sum_{j\geq 0} |\psi_j| \left|\psi_{j-|h|}\right| = \sum_{j\geq 0} \psi_j^2$
- $h = \pm 1$: $\sum_{j\geq 0} |\psi_j| \left|\psi_{j-|h|}\right| = \sum_{j\geq 0} |\psi_j| |\psi_{j+1}|$
- $h = \pm 2$: $\sum_{j\geq 0} |\psi_j| \left|\psi_{j-|h|}\right| = \sum_{j\geq 0} |\psi_j| |\psi_{j+2}|$
- ...

the sum of all these terms results then in $\left(\sum_{j\geq 0} |\psi_j|\right)^2$ that, since $(X_t)$ is causal, is finite. □

**Corollary 12.1.** $\lim_h \gamma_X(h) = 0 \implies$ *ergodicity in mean; also, if* $(X_t)$ *is Gaussian then it has the covariance ergodic property.*

Recall that

- $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty \implies (X_t)$ is short memory
- $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| = +\infty \implies (X_t)$ is long memory

**Theorem 12.1.** *Suppose* $(X_t) \sim AMRA(p, q)$*, then*

$$(X_t) \text{ causal} \iff \theta(Z) \neq 0 \ \forall Z \in \mathbb{C} \text{ such that } |Z| \leq 1$$

**Corollary 12.2.** *$X_t = \psi(B)W_t$ with $W_t\mathcal{WN}(0,\sigma^2)$ with $\psi(Z) = \sum_{j\geq 0}\psi_j Z^j = \frac{\phi(Z)}{\theta(Z)}$ with $|Z| \leq 1$.*

*Proof.* We will start with the $\Leftarrow$ part. By hypothesis, we know that

$$\exists\delta > 0 \; such\ that\ \theta^{-1}(Z) = \frac{1}{\theta(Z)} = \sum_{j\geq 0}\theta^*_j Z^j \; for\ \ |Z| < 1+\delta$$

in particular this is true as $1+\frac{\delta}{2} < 1+\delta$. Computing the limit:

$$\lim_j \theta^*_j(1+\frac{\delta}{2})^j = 0$$

so

$$\exists k \in (0,+\infty) \; and \; \tilde{j} \geq 0$$

such that

$$|\theta^*_j|\left(1+\frac{\delta}{2}\right)^j < k \implies |\theta^*_j| < k\left(1+\frac{\theta}{2}\right)^{-j}$$

with $j \geq \tilde{j}$. This lead to

$$\sum_{j\geq\tilde{j}}|\theta_j *| < k\sum_{j\geq\tilde{j}}\left(1+\frac{\delta}{2}\right)^{-j} < \infty$$

since $1+\frac{\delta}{2} > 1$. Now consider

$$\psi(Z) = \theta^{-1}(Z)\phi(Z)$$

and, in particular

$$\phi(Z) = \sum_{j\geq 0}\tilde{\phi}_j Z^j \; where \begin{cases} 1 & j=0 \\ \phi_j & j=1,...,q \\ 0 & j>q \end{cases}$$

Therefore

$$\psi(Z) = \sum_{j\geq 0}\psi_j Z^j \implies \{\psi_j\} = \{\theta^*_j\} * \left\{\tilde{\phi}_j\right\}$$

this means that $\psi_j = \sum_{k=0}^{j}\theta^*_k\tilde{\theta}_{j-k}$. Therefore

$$|\psi_j| \leq \sum_{k=0}^{j}|\theta^*_k|\left|\tilde{\phi}_{j-k}\right| \implies \sum_{j\geq 0}|\psi_j| \leq \sum_{j\geq 0}\left(\sum_{k=0}^{j}|\theta^*_k|\left|\tilde{\phi}_{j-k}\right|\right)$$

where the last term is the coefficient of order j of $\left\{|\theta^*_j|\right\} * \left\{\left|\tilde{\phi}_j\right|\right\}$; so it is equal to

$$\left(\sum_{k=0}^{j}|\theta^*_k|\left|\tilde{\phi}_{j-k}\right|\right) = \left(\sum_{j\geq 0}|\theta^*_j|\right)\left(\sum_{j\geq 0}\left|\tilde{\phi}_j\right|\right) < \infty\infty \implies \{\psi_j\} \; absolutely\ sommable$$

Now consider $\theta(B)X_t = \phi(B)W_t$:

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

with $\psi_(B) = \theta^{-1}(B)\phi(B)$. We obtained a causal representation of $X_t$ and also proved the corollary. Now we will see the $\implies$ part. We know that $X_t = \psi(B)W_t$, where

$$\psi(Z) = \sum_{j\geq 0}\psi_j Z^j \quad with \quad \sum_{j\geq 0}|\psi_j| < \infty$$

Apply $\theta(B)$ to both sides of the equation:

$$\theta(B)X_t = \theta(B)\psi(B)W_t = \theta(Z)\psi(Z)W_t = \eta(Z)W_t$$

We have to prove that $\eta(Z) = \phi(Z)$. We know that $\eta(B)W_t = \theta(B)X_t = \phi(B)W_t$ (the last equal by hypothesis. We then know that the first and last members are equal and the equality is equal to

$$\sum_{j\geq 0}\eta_j W_{t-j} = \sum_{j=0}^{q}\phi_j W_{t-j}$$

Multiply both sides by $W_{t-k}$ and take the expectation:

$$\mathbb{E}\left[\sum_{j\geq 0}\eta_j W_{t-j}W_{t-k}\right] = \sum_{j=0}^{q}\phi_j\mathbb{E}\left[W_{t-j}W_{t-k}\right]$$

We know that we can exchange summation ad expectation if $\sum_{j\geq 0}|\eta_j|\,\mathbb{E}\left[|W_{t-j}W_{t-k}|\right] < \infty$. We already know that $\mathbb{E}\left[|W_{t-j}W_{t-k}|\right] \leq \sigma^2$, so we must prove that $\sum_{j\geq 0}|\eta_j| < \infty$. Consider that $\eta(Z) = \theta(Z)\psi(Z)$ and

$$\sum_{j\geq 0}\tilde{\theta}_j Z^j \quad with \quad \begin{cases} 1 & j = 0 \\ -\theta_j & J = 1, ..., p \\ 0 & j > p \end{cases}$$

then $\{\eta_j\} = \left\{\tilde{\theta}_j\right\} * \{\psi_j\}$. As before

$$\sum_{j\geq 0}|\eta_j| \leq \sum_{j\geq 0}\left(\sum_{k=0}^{j}\left|\tilde{\theta}_k\right||\psi_{j-k}|\right) = \left(\sum_{j\geq 0}\left|\tilde{\theta}_j\right|\right)\left(\sum_{j\geq 0}|\psi_j|\right) < \infty\infty$$

So now we can exchange summation and expectation:

$$\sum_{j\geq 0}\eta_j\mathbb{E}\left[W_{t-j}W_{t-k}\right] = \sum_{j=0}^{q}\phi_j\mathbb{E}\left[W_{t-j}W_{t-k}\right]$$

We have that

- $k = 0 \implies \mathbb{E}\left[W_{t-j}W_t\right] \neq 0 \iff j = 0 \implies \eta_0 = \phi_0 = 1$

- $k = 1 \implies \mathbb{E}\left[W_{t-j}W_{t-1}\right] \neq 0 \iff j = 1 \implies \eta_1 = \phi_1$
- ...
- $k = q \implies \mathbb{E}\left[W_{t-j}W_{t-q}\right] \neq 0 \iff j = q \implies \eta_q = \phi_q$
- $k > q \implies \mathbb{E}\left[W_{t-j}W_{t-k}\right] = 0 \implies \eta_k = 0$

implying that $\eta(Z) = \phi(Z)$ and $\psi(Z) = \frac{\phi(Z)}{\theta(Z)}$ for $Z \in \mathbb{C}$ such that $\theta(Z) \neq 0$. Now suppose $\exists Z_0 \in \mathbb{C}$ such that $|Z_0| \leq 1$ and $\theta(Z_0) = 0$. Then

$$\theta(Z_0)\psi(Z_0) = \eta(Z_0) = \phi(Z_0)$$

knowing that $\theta(Z_0) = 0$ and that

$$|\psi(Z_0)| \leq \sum_{j\geq 0} |\psi_j| \, |Z_0|^0 \leq \sum_{j \geq 0} |\psi_j| < \infty$$

we can state that $\theta(Z_0) = 0 = \phi(Z_0)$, implying that $Z_0$ is a common root, but this impossible since $(X_t)$ is not redundant. □

How to compute $\{\psi_j\}$ in $X_t = \psi(B)W_t$? We have that $\theta(Z)\psi(Z) = \phi(Z)$, where

$$\theta(Z) = \sum_{k\geq 0} \tilde{\theta}_k Z^k \quad with \quad \tilde{\theta}_k = \begin{cases} 1 & k = 0 \\ -\theta_k & k = 1, ..., p \\ 0 & k > p \end{cases}$$

and

$$\phi(Z) = \sum_{k\geq 0} \tilde{\phi}_k Z^k \quad with \quad \tilde{\phi}_k = \begin{cases} 1 & k = 0 \\ \phi_k & k = 1, ..., p \\ 0 & k > p \end{cases}$$

implying that $\left\{\tilde{\phi}_k\right\} = \left\{\tilde{\theta}_k\right\} * \{\psi_k\}$, from which we observe that

$$\begin{aligned} \tilde{\phi}_k &= \sum_{j=0}^{k} \tilde{\theta}_j \psi_{k-j} = \psi_k + \sum_{j=1}^{k} \tilde{\theta}_j \psi_{k-j} \\ \psi_j &= \tilde{\phi}_k - \sum_{j=1}^{k} \tilde{\theta}_j \psi_{k-j} \; for \; k = 0, 1, ... \end{aligned} \tag{2}$$

So, for example, if $p, q > 2$ then $\psi_0 = \tilde{\phi}_0 = 1, \; \psi_1 = \tilde{\phi}_1 - \tilde{\phi}_1\psi_0 = \phi_1 + \theta_1\psi_0, \; \phi_2 = \tilde{\phi}_2 - \tilde{\theta}_1\psi_0 - \tilde{\theta}_2\psi_1 = \phi_2 + \theta_1\psi_0 + \theta_2\psi_1$.

**Exercise 12.2.** *Suppose $(X_t) \sim ARMA(2,1)$ such that*

$$X_t = X_{t-1} - \frac{1}{4}X_{t-2} + W_t + W_{t-1}$$

1. *Prove that $(X_t)$ is not reduntant and that it is causal;*
2. *find $\{\psi_j\}$ of the causal representation of $(X_t)$;*
3. *find the ACF.*

**Exercise 12.3.** *Let $(X_t) \sim ARMA(1,1)$:*

1. *what is the condition on $\theta_1$ such that $(X_t)$ is causal?*
2. *Find the ACF of $(X_t)$.*

**Theorem 12.2.** *If $\theta(Z) \neq 0$ $\forall Z \in \mathbb{C}$ such that $|Z| = 1$ then $\exists!(X_t)$ such that $Z_t = \psi(B)W_t$ with $\psi(Z) = \frac{\phi(Z)}{\theta(Z)}$ for $\frac{1}{r} < |Z| < r$ and $r > 1$.*

**Example 12.3.** *There is a function in R allowing us to retrieve the coefficients $\psi_j$:*

```
#####################################
# ARMA to MA
#####################################

# AR(0.9)
psi=ARMAtoMA(ar=c(0.9),lag.max=80)
# theoretical coefficients psi_i=theta^i
theta= numeric(80)
theta[1]=0.9
for (i in 2:80){theta[i]=theta[i-1]*0.91

# plot and compare
windows()
plot(psi,ylab=’weights’,main=’coeff. LF’,type=’o’)
lines(theta,col=’red’)

# MA(0.4)
psi=ARMAtoMA(ma=c(0.4),lag.max=80)

# Example (1-B+0.2513^2)X_t=(l+B)W_t
#
# ar=c(1.0,-0.25),ma=c(1)
#
# redundance?
library(polynom)

# for AR(1.0,-0.25)
(roots=solve(polynomial(c(1,-1.0,+0.25))))
# for MA(1)
(roots=solve(polynomial(c(1,1))))

# ARMA to MA
psi=ARMAtoMA(ar=c(1.0,-0.25),ma=c(1),lag.max=80)
plot(psi,ylab=’weights’,main=’coeff. LF’,type=’o’)
```

**Example 12.4.** *Simulate and plot a path of*

1. $X_t = 0.9X_{t-1} + W_t - 0.5W_{t-1}$
2. $X_t = 0.55X_{t-1} + W_t - 0.5W_{t-1}$

*3.* $X_t = X_{t-1} - 0.25X_{t-2} + W_t - 0.5W_{t-1}$

*with* $n = 200$ *and seed* $= 154$*. Plot also the ACF.*

```
#####################################
# Simulation of paths from ARMA(p,q)
#####################################

set.seed(154)
# simulate ARMA(1,1) with theta_1=0.9 and phi_1=-0.5
arma.siml=arima.sim(list(ar=c(0.9),ma=c(-0.5)),200)
# simulate ARMA(1,1) with theta_1=0.55 and phi_1=-0.5
arma.sim2=arima.sim(list(ar=c(0.55),ma=c(-0.5)),200)
# simulate ARMA(2,1) with theta 1=1, theta2=-0.25 and phi 1=-0.5
arma.sim3=arima.sim(list(ar=c(1.0,-0.25),ma=c(-0.5)),200)

# plot the paths
minr=min(cbind(arma.siml,arma.sim2,arma.sim3))
maxr=max(cbind(arma.siml,arma.sim2,arma.sim3))
windows()
par(mfrow=c(3,1))
plot(arma.sim1,type=’o’,main=’ARMA(1,1): theta_1=0.9 and
phi_1=-0.5’,ylim=c(minr,maxr))
plot(arma.sim2,type=’o’,main=’ARMA(1,1): theta_1=0.55 and
phi_1=-0.5’,ylim=c(minr,maxr))
plot(arma.sim3,type=’o’,main=’ARMA(2,1): theta_1=1,theta2=-0.25 and
phi1=-0.5’,ylim=c(minr,maxr))

#plot the ACF
windows()
par(mfrow=c(1,3))
acf(arma.sim1, lag.max=50, main=’ARMA (0.9,-0.5)’, ylim=c(-1,1))
acf(arma.sim2, lag.max=50, main=’ARMA (0.55,-0.5)’, ylim=c(-1,1))
acf(arma.sim3, lag.max=50, main=’ARMA (1, -0.25; -0.5)’, ylim=c(-1,1))
```

**Definition 12.3.** *Suppose* $(X_t) \sim ARMA(p,q)$ *is invertible if*

$$\exists\,\{\pi_j\} \text{ such that } \sum_{j\geq 0} |\pi_j| < \infty \text{ such that } W_t = \pi(B)X_t \ \forall i \in \mathbb{Z}$$

*with* $\pi(Z) = \sum_{j\geq 0} \pi_j Z^j$*.*

**Theorem 12.3.** *Let* $(X_t) \sim ARMA(p,q)$ *not redundant; then*

$$(X_t) \text{ is invertible } \iff \phi(Z) \neq 0 \ \forall Z \in \mathbb{C} \text{ such that } |Z| \leq 1$$

**Corollary 12.3.** $\{\pi_j\}$ *in* $\pi(Z) = \sum_{j\geq 0} \pi_j Z^j$ *are such that* $\pi(Z) = \frac{\theta(Z)}{\phi(Z)}$ *with* $|Z| \leq 1$*.*

**Corollary 12.4.** *If* $(X_t) \sim ARMA(p,q)$ *is not redundant and*

$$\theta(Z)\phi(Z) \neq 0 \ \forall z \in \mathbb{C} \text{ such that } |Z| \leq 1$$

*then* $(X_t)$ *is causal and invertible.*

**Definition 12.4.** *Suppose $(X_t)$ stationary. The* ***ACF generating function*** *is $G(Z) = \sum_{h\in\mathbb{Z}} \gamma_X(h)Z^h$ provided that $G(Z) < \infty$ $\forall Z \in \mathbb{C}$ such that $\frac{1}{r} < |Z| < r$ with $r > 1$.*

**Example 12.5.** $(W_t) \sim \mathcal{WN}(0, \sigma^2) \implies G(Z) = \sigma^2$

**Lemma 12.1.** *Let $(X_t)$ be a linear time series, $X_t = \psi(B)W_t$. If $\exists r > 1$ such that $\sum_{j\in\mathbb{Z}} |\psi_j| Z^j < \infty$ $\forall Z \in \mathbb{C}$ such that $\frac{1}{r} < |Z| < r$ then*

$$G(Z) = \sigma^2\psi(Z)\psi(Z^{-1})$$

*Proof.* Consider

$$\psi(Z)\psi(Z^{-1}) = \left(\sum_{j\in\mathbb{Z}} \psi_j Z^j\right)\left(\sum_{j\in\mathbb{Z}} \psi_j Z^{-j}\right) = \left(\sum_{j\in\mathbb{Z}} \psi_j Z^j\right)\left(\sum_{k\in\mathbb{Z}} \psi_k Z^k\right)$$

This is a new Laurent series:

$$= \sum_{i\in\mathbb{Z}} \psi_i^* Z^i \quad with \quad \{\psi_j^*\} = \{\psi_j\} * \{\psi_{-k}\}$$

with

$$\psi_j^* = \sum_{k\in\mathbb{Z}} \psi_k\psi_{(i-k)} = \sum_{k\in\mathbb{Z}} \psi_k\psi_{k-i} = \frac{\gamma_X(i)}{\sigma^2}$$

Then

$$G(Z) = \sum_{h\in\mathbb{Z}} \gamma_X(h)Z^h = \sigma^2 \sum_{h\in\mathbb{Z}} \psi_h^* Z^h = \sigma^2\psi(Z)\psi(Z^{-1})$$

□

**Example 12.6.** *Suppose $(X_t) \sim ARMA(p, q)$ with $\theta(0) \neq 0$ $\forall Z \in \mathbb{C}$ such that $|Z| = 1$. Then $\exists!(X_t)$ such that $X_t = \psi(B)W_t$ with $\psi(Z) = \frac{\phi(Z)}{\theta(Z)}$ for $\frac{1}{r} < |Z| < r$ with $r > 1$. Therefore, from the lemma, we have that*

$$G(Z) = \sigma^2 \frac{\phi(Z)}{\theta(Z)} \frac{\phi(Z^{-1})}{\theta(Z^{-1})}$$

**Theorem 12.4.** *Suppose $(X_t) \sim ARMA(p, q)$ with $\theta(Z) \neq 0$, $\phi(Z) \neq 0$ $\forall Z \in \mathbb{C}$ such that $|Z| = 1$; then $\exists\tilde{\phi}(Z)$ and $\tilde{\theta}(Z)$ with*

$$degree(\tilde{\phi}) = q \quad degree(\tilde{\theta}) = p$$

*such that*

$$\tilde{\theta}(Z)\tilde{\phi}(Z) \neq 0 \quad for \;\; |Z| \leq 1$$

*ans $\exists(W_t^*) \sim \mathcal{WN}(0, \sigma^2_{W^*})$ such that $(X_t)$ is the causal and invertible solution to the equation $\tilde{\theta}(B)X_t = \tilde{\phi})(B)W_t^*$ $\forall t \in \mathbb{Z}$.*

*Proof.* Suppose $a_{r+1}, ..., a_p$ roots of $\theta(Z)$ such that $|a_{r+1}|, ..., |a_p| < 1$; then $a_{r+1}, ..., a_p$ are roots of $\prod_{j=r+1}^{p}\left(1 - \frac{Z}{a_j}\right)$. Consider

$$degree\left(\frac{\theta(Z)}{\prod_{j=r+1}^{p}\left(1 - \frac{Z}{a_j}\right)}\right) < p \implies \tilde{\theta(Z)} = \frac{\theta(Z)}{\prod_{j=r+1}^{p}\left(1 - \frac{Z}{a_j}\right)} \prod_{j=r+1}^{p}(1 - Za_j)$$

Note that the roots of $\prod_{j=s+1}^{p}(1-Za_j)$ are $\frac{1}{a_j}$ for $j=s+1,...,p$. Now consider $b_{j+1},...,b_q$ the roots of $\phi(Z)$ such that $|b_{j+1},...,|b_1|| < 1$. Consider

$$\tilde{\phi}(Z) = \frac{\phi(Z)}{\prod_{j=s+1}^{q}\left(1-\frac{Z}{b_j}\right)}\prod_{j=s+1}^{q}(1-b_jZ)$$

We have that $\tilde{\theta}(Z)\tilde{\phi}(0) \neq 0$ for $|Z| \leq 1$. Define now $W_t^* = \frac{\tilde{\theta}(B)}{\tilde{\phi}(B)}X_t \ \forall t \in \mathbb{Z}$. We have to prove that $(W_t^*) \sim \mathcal{WN}$. Observe that

$$\begin{aligned}
W_t^* &= \frac{\frac{\prod_{j=r+1}^{p}(1-a_jB)}{\prod_{j=r+1}^{p}\left(1-\frac{B}{a_j}\right)}}{\frac{\prod_{j=s+1}^{q}(1-b_jB)}{\prod_{j=s+1}^{q}\left(1-\frac{B}{b_j}\right)}}\frac{\theta(B)}{\phi(B)}X_t \\
&= \frac{\frac{\prod_{j=r+1}^{p}(1-a_jB)}{\prod_{j=r+1}^{p}\left(1-\frac{B}{a_j}\right)}}{\frac{\prod_{j=s+1}^{q}(1-b_jB)}{\prod_{j=s+1}^{q}\left(1-\frac{B}{b_j}\right)}}W_t \\
&= \frac{\prod_{j=r+1}^{p}(1-a_jB)\prod_{j=s+1}^{q}\left(1-\frac{B}{b_j}\right)}{\prod_{j=r+1}^{p}\left(1-\frac{B}{a_j}\right)\prod_{j=s+1}^{q}(1-b_jB)}W_t \\
&= \tilde{\psi}(B)W_t
\end{aligned}$$

Then $(W_t^*)$ is a linear time series, since

$$\tilde{\psi}(Z) = \frac{\prod_{j=r+1}^{p}(1-a_jZ)\prod_{j=s+1}^{q}\left(1-\frac{Z}{b_j}\right)}{\prod_{j=r+1}^{p}\left(1-\frac{Z}{a_j}\right)\prod_{j=s+1}^{q}(1-b_jZ)}$$

implies that $\exists \tilde{r} > 1$ such that $\left|\tilde{\psi}(Z)\right| < \infty$ for $\frac{1}{\tilde{r}} < |Z| < \tilde{r}$, further implying that $\sum_{j\geq 0}\left|\tilde{\psi}_j\right| < \infty$. From the lemma of the covariance generating function, we can state that

$$\begin{aligned}
G_{W^*}(Z) &= \sigma^2\tilde{\psi}(Z)\tilde{\psi}(Z^{-1}) \\
&= ... \\
&= \sigma^2\frac{\prod_{j=r+1}^{p}|a_j|^2}{\prod_{j=s+1}^{q}|b_j|^2} \\
&= \sigma_{W^*}^2
\end{aligned}$$

completing the proof. □

**Remark 12.1.** $\psi(Z) = \frac{\phi(Z)}{\theta(Z)} = \frac{\tilde{\phi}(Z)}{\tilde{\theta}(Z)}\tilde{\psi}(Z)$

**Exercise 12.4.** *Apply 12.4 to*

1. $(X_t) \sim MA(1)$ *with* $|\phi_1| > 1$
2. $(X_t) \sim AR(1)$ *with* $|\theta_1| > 1$

# 13 Lecture 13

**ARMA models. From data to models. The Yule-Walker estimators. Invertible covariance matrix. Partial autocorrelation function. Best linear predictors: projection theorem and mean square error.**

Now we will look how to fit the ARMA model to the data.

If the ACF shows a cutoff after a certain lag, we can guess a MA model; after having choosed the order $q$, we can estimate the coefficients $\phi_j$ is to equate the first $q$ theoretical ACF values to the sample correlation values, and find the solution of the implies syste of equations:

$$\begin{cases} \rho_X(1) = \frac{\phi_1+\phi_1\phi_2+\phi_2\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2} = \hat{\rho}_1 \\ \rho_X(2) = \frac{\phi_2+\phi_1\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2} = \hat{\rho}_2 \\ \rho_X(3) = \frac{\phi_3}{1+\phi_1^2+\phi_2^2+\phi_3^2} = \hat{\rho}_3 \end{cases}$$

but this is not usually the best method.

If the ACF goes to 0 when the lags goes to infinity, a short term memory model is usually a good choiche. If we suppose a AR model of order $q = 2$, we can estimate the parameters of the model by using the YW equations:

$$\begin{aligned} \gamma_X(0) - \theta_1\gamma_X(-1) - \theta_2\gamma_X(-2) = \sigma^2 \\ \gamma_X(1) - \theta_1\gamma_X(0) - \theta_2\gamma_X(-1) = 0 \\ \gamma_X(2) - \theta_1\gamma_X(1) - \theta_2\gamma_X(0) = 0 \\ \ldots \end{aligned}$$

Considering the property $\gamma_X(h) = \gamma_X(-h)$, we obtain the system

$$\begin{cases} \gamma_X(1) = \theta_1\gamma_X(0) + \theta_2\gamma_X(1) \\ \gamma_X(2) = \theta_1\gamma_X(1) + \theta_2\gamma_X(0) \end{cases}$$

that can be rewritten as

$$\begin{pmatrix} \gamma_X(1) \\ \gamma_X(2) \end{pmatrix} = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) \\ \gamma_X(1) & \gamma_X(0) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

or, equvalently

$$\boldsymbol{\gamma}_{(2)} = \boldsymbol{\Gamma}_2\boldsymbol{\theta}_2$$

Then, the Yule-Walker's estimator of $(\theta_1, \theta_2)$ is

$$\hat{\boldsymbol{\gamma}}_{(2)} = \hat{\boldsymbol{\Gamma}}_2\hat{\boldsymbol{\theta}}_2$$

What about $\sigma^2$?

$$\sigma^2 = \gamma_X(0) - \theta_1\gamma_X(1) - \theta_2\gamma_X(2) = \gamma_X(0) - \begin{pmatrix} \gamma_X(1) & \gamma_X(2) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

So the YW estimator of $\sigma^2$ is

$$\hat{\sigma}^2 = \hat{\gamma}_X(0) - \hat{\boldsymbol{\gamma}}_{(2)}^\intercal\hat{\boldsymbol{\theta}}_2$$

For $p \in \mathbb{N}$ the YW estimators of $\theta_1, ..., \theta_p$ is $\hat{\boldsymbol{\gamma}}_{(p)} = \hat{\boldsymbol{\Gamma}}_p \hat{\boldsymbol{\theta}}_p$ and for $\sigma^2$ is $\hat{\sigma}^2 = \hat{\gamma}_X(0) - \hat{\boldsymbol{\gamma}}_{(p)}^\intercal \hat{\boldsymbol{\theta}}_p$. Suppose to have $n$ observations, then we can compute the value of $\hat{\gamma}(h)$ for $h = 0, ..., n-1$, and we can build

$$\hat{\boldsymbol{\Gamma}}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$$

then the YW estimation of $\theta_1, ..., \theta_n$ is

$$\hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_{(p)} \quad with \quad \hat{\boldsymbol{\gamma}}_{(p)} = (\hat{\gamma}(1), ..., \hat{\gamma}(p))^\intercal$$

and of $\sigma^2$ is

$$\hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_{(p)}^\intercal \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_{(p)}$$

If the sample size is large, then the YW estimator are approximately normally distributed, and the value of $\hat{\sigma}^2$ is close to the true value. In this sense, the YW estimators are optimal.

**Theorem 13.1.** *Suppose $(X_t)$ stationary, $\gamma(0) > 0$ and $\lim_h \gamma(h) = 0$. Then the covariance matrix of $(X_1, ..., X_n)^\intercal$ $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$ is non-singular $\forall n \geq 1$.*

**Lemma 13.1.** *Suppose $\Sigma$ is the covariance matrix of a column matrix $(X_1, ..., X_n)^\intercal$. Then $\Sigma$ is singular if and only if $\exists \boldsymbol{b} = (b_1, ..., b_n)^\intercal \neq 0$ such that $Var(\boldsymbol{b}^\intercal \boldsymbol{X}) = 0$.*

*Proof.* Suppose the covariance matrix $\Gamma_n$ singular for some $n$. As $\gamma(0) > 0$ then $\exists r \geq 1$ such that $\Gamma_r$ is non-singular and $\Gamma_{r+1}$ is singular $(r+1 \leq n)$. According to the previous lemma:

$$X_{r+1} = \sum_{j=1}^r a_j X_j$$

but from stationarity we also know that

$$X_{r+(h+1)+1} = \sum_{j=1}^r a_j X_{j+(h-1)} \quad for \quad h \geq 1$$

Note that

$$\begin{aligned} X_{r+1} &\implies X_1, ..., X_r \\ X_{r+2} &\implies X_2, ..., X_r, X_{r+1} \implies X_1, ..., X_r \\ &\cdots \\ X_n &\implies X_1, ..., X_r \quad as\ r \geq r+1 \\ X_n &= \sum_{j=1}^r a_j^{(n)} X_j \\ X_n &= (\boldsymbol{a}^{(n)})^\intercal \boldsymbol{X}_r \end{aligned}$$

with $\boldsymbol{a}^{(n)} = (a_1^{(n)}, ..., a_r^{(n)})^\intercal$ and $\boldsymbol{X}_r = (X_1, ..., X_r)^\intercal$. Now, consider

$$\begin{aligned} \gamma(0) &= \mathbb{E}\left[X_n^2\right] \\ &= \mathbb{E}\left[\left((\boldsymbol{a}^{(n)})^\intercal \boldsymbol{X}_r\right)\left(\boldsymbol{X}_r^\intercal \boldsymbol{a}^{(n)}\right)\right] \\ &= (\boldsymbol{a}^{(n)})^\intercal \mathbb{E}\left[\boldsymbol{X}_r \boldsymbol{X}_r^\intercal\right] \boldsymbol{a}^{(n)} \\ &= (\boldsymbol{a}^{(n)})^\intercal \boldsymbol{\Gamma}_r \boldsymbol{a}^{(n)} \end{aligned}$$

We know that $\mathbf{\Gamma}_r$ is non-singular, so the decomposition

$$\mathbf{\Gamma}_r = \boldsymbol{P}\mathbf{\Lambda}\boldsymbol{P}^\intercal$$

where $\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_r)$ is the vector af eigenvalues and $P$ is the orthogonal matrix of eigenvectors, so $\boldsymbol{P}\boldsymbol{P}^\intercal = \boldsymbol{I}_r = \boldsymbol{P}^\intercal\boldsymbol{P}$. Moreover, $\mathbf{\Gamma}_r$ is non-negative definite, meaning that $0 \leq \lambda_1 \leq ... \leq \gamma_r$. Now we can state that

$$\begin{aligned}\gamma(0) &= (\boldsymbol{a}^{(n)})^\intercal\boldsymbol{P}\mathbf{\Lambda}\boldsymbol{P}^\intercal\boldsymbol{a}^{(n)} \\ &\geq \lambda_1(\boldsymbol{a}^{(n)})^\intercal\boldsymbol{P}\boldsymbol{I}_r\boldsymbol{P}^\intercal\boldsymbol{a}^{(n)} \\ &= \lambda_1(\boldsymbol{a}^{(n)})^\intercal\boldsymbol{I}_r\boldsymbol{a}^{(n)} \\ &= \lambda_1(\boldsymbol{a}^{(n)})^\intercal\boldsymbol{a}^{(n)} \\ &= \lambda_1\sum_{j=1}^{r}(a_j^{(n)})^2\end{aligned}$$

But we also know that

$$\begin{aligned}\gamma(0) &= \mathbb{E}\left[X_n X_n\right] \\ &= \mathbb{E}\left[X_n\sum_{j=1}^{r}a_j^{(n)}X_j\right] \\ &= \sum_{j=1}^{r}a_j^{(n)}\mathbb{E}\left[X_n X_j\right] \\ &= \sum_{j=1}^{r}a_j^{(n)}\gamma(n-j) \\ &\leq \sum_{j=1}^{r}\left|a_j^{(n)}\right|\left|\gamma(n-j)\right|\end{aligned}$$

and, if we suppose that $n \to \infty$ then this last summation goes to 0, which is a contradiction, since $\gamma(0)$ is strictly positive, completing the proof. □

**Proposition 13.1.** *If $\left\{\tilde{\theta}_i\right\}_{i=1}^{p}$ are the YW estimates of $\theta_1, ..., \theta_p$ then $\tilde{\theta}(Z) = 1 - \tilde{\theta}_1 Z + ... - \tilde{\theta}_p Z^p \neq 0$ for $|Z| \leq 1$*

**Remark 13.1.** *If $\gamma(h)$ is any ACF such that $\gamma(0) > 0$ and $\lim_h \gamma(h) = 0$ then, for any fixed $p$, there is a causal $AR(p)$ whose ACF at lags $h = 0, 1, ..., p$ is equal to $\gamma(h)$.*

**Exercise 13.1.** *Suppose $\gamma(0) = 1$ and $\gamma(\pm 1) = \beta$. Find $(X_t) \sim AR(1)$ having $\gamma(0)$ and $\gamma(\pm 1)$ as ACF at $h = 0, h = \pm 1$. Is it possible to find a $MA(1)$ with the same ACF?*

**Remark 13.2.** *Due to numerical reason, it is more convinient to use the correlation function in order to estimare the AR parameters. Remember that*

$$\boldsymbol{\rho}_{(p)} = \boldsymbol{R}_p\boldsymbol{\theta}_p \quad where \quad \begin{cases}\boldsymbol{\rho}_{(p)} = (\rho_X(1), ..., \rho_X(p)^\intercal) \\ \boldsymbol{R}_p = \left[\rho_X(i-j)\right]_{i,j=1}^{p} \\ \boldsymbol{\theta}_p = (\theta_1, ..., \theta_p)^\intercal\end{cases}$$

*and that*

$$\sigma^2 = \gamma_X(0) - (\boldsymbol{\gamma}_{(p)})^\intercal \boldsymbol{\theta}_p = \gamma_X(0)\left(1 - \frac{\boldsymbol{\gamma}_{(p)}}{\gamma_X(0)}\boldsymbol{\theta}_p\right) = \gamma_X(0)(1 - \boldsymbol{\rho}_{(p)}\boldsymbol{\theta}_p)$$

*Suppose to have* $n$ *observation* ($n \geq p$) *and to estimate* $\hat{\rho}(h)$ *for* $h = 0, ..., n-1$. *Set* $\hat{\boldsymbol{\rho}}_{(p)} = (\hat{\rho}(1), ..., \hat{\rho}(p))^\intercal$ *and* $\hat{\boldsymbol{R}}_p = [\hat{\rho}(i-j)]_{i,j=1}^p$. *As* $\hat{\rho}(0) = 1 \neq 0$ *and* $\hat{\rho}(h) = 0$ *for* $h > n$, *from 13.1, there exists* $\hat{\boldsymbol{R}}_p^{-1}$ *such that*

$$\begin{aligned} \hat{\boldsymbol{R}}_p^{-1}\hat{\boldsymbol{\rho}}_{(p)} \quad & \textit{estimates} \quad (\theta_1, ..., \theta_p) \\ \hat{\gamma}(0)(1 - \hat{\boldsymbol{\rho}}_{(p)}^\intercal \hat{\boldsymbol{R}}_p \hat{\boldsymbol{\rho}}_{(p)}) \quad & \textit{estimates} \quad \sigma^2 \end{aligned}$$

*Observe that the product* $\boldsymbol{R}_k^{-1}\boldsymbol{\rho}_{(k)}$ *with* $k \geq 1$ *gives the column vector of partial autocorrelation function.*

**Definition 13.1.** *The* ***partial auto correlation function (PACF)*** *of* $(X_t)$ *stationary is defined as*

$$\begin{aligned} \pi_{11} &= corr(X_{t+1}, X_t) = \rho_X(1) \\ \pi_{hh} &= corr(X_{t+h} - \tilde{X}_{t+h}, X_t - \tilde{X}_t) \end{aligned}$$

*with* $h \geq 2$ *and* $\tilde{X}_{t+h}, \tilde{X}_t$ *are the best linear predictors (BLP) of* $X_{t+h}, X_t$ *given* $\{X_{t+1}, ..., X_{t+h-1}\}$.

BLP refers with respect to the mean square error.

**Remark 13.3.** $(X_t)$ *stationary* $\implies$ *with* $t = 1$

$$\begin{aligned} &\pi_{11} = corr(X_2, X_1) \\ &\pi_{hh} = corr(X_{h+1} - \tilde{X}_{h+1}, X_1 - \tilde{X}_1) \quad h \geq 2 \\ &(A) = \{X_2, ..., X_h\} \end{aligned}$$

Why use BLP and not conditional expectation? Consider $X, Z_1, ..., Z_h \in \mathcal{L}^2(\Omega, H, P) = \mathcal{L}^2$. Denote with $F = \sigma(Z_1, ..., Z_h)$ the generated $\sigma$-algebra. Then $\bar{X} = \mathbb{E}[X|F] = \mathbb{E}[X|Z_1, ..., Z_h]$ is a random variable such that

$$\mathbb{E}\left[(X - \tilde{X})^2\right] = \inf_{Y \in \mathcal{L}^2(\Omega, F, \mathbb{P})} \mathbb{E}\left[(X - Y)^2\right]$$

where $\mathcal{L}^2(\Omega, F, \mathbb{P}) = \left\{Y \in \mathcal{L}^2 / Y = f(Z_1, ..., Z_h) \textit{ with } f \textit{ measurable}\right\}$. $\bar{X}$ is in this set since $\mathbb{E}[(X|Z_1, ..., Z_h)] = \tilde{f}(Z_1, ..., Z_h)$. This conditional expectation is the best function of $Z_1, ..., Z_h$ to predict $X$.

**Example 13.1.** *If* $Z_1, ..., Z_h$ *is distributed as a multivariate Gaussian distribution, then*

$$\mathbb{E}[(X|Z_1, ..., Z_h)] = \sum_{j=0}^{h} \alpha_j Z_j$$

*with* $Z_0 \overset{a.s.}{=} 1$ *and* $\alpha_j$ *minimize the MSE. So, for Gaussian time series, the conditional expectation gives the BLP.*

Suppose to replace $\mathcal{L}^2(\Omega, F, \mathbb{P})$ with $sp1, Z_1, ..., Z_h \subset \mathcal{L}^2 = \left\{Y \in \mathcal{L}^2 / Y = \sum_{j=0}^{h} \tilde{\alpha}_j Z_j\right\}$. The existence of a unique random variable that minimizes the MSE computed on the subspace of linear combinations of $Z_1, ..., Z_h$ is granted by the projection theorem.

**Theorem 13.2.** *(From the **projection theorem**). If*

- $M = \bar{sp}\left\{Z_1, ..., Z_h\right\} \subset \mathcal{L}^2$
- $X \in \mathcal{L}^2$

*then*

$$\exists! \tilde{X} \in M \textit{ such that } \mathbb{E}\left[(X - \tilde{X})^2\right] = \inf_{Y \in M} \mathbb{E}\left[(X - Y)^2\right]$$

*$\tilde{X}$ is denoted as $P_M X$ is said the orthogonal projection of $X$ into $M$.*

One of the main property of $P_M X$ is that $X - \tilde{X} \in M^{\perp} = \left\{\tilde{Y} \in \mathcal{L}^2 / \mathbb{E}\left[Y\tilde{Y}\right] = 0 \ \forall Y \in M\right\}$. $M^{\perp}$ is the orthogonal complement of $M$. If we substitute $X - \tilde{X} \in \mathcal{L}^2$ to $\tilde{Y}$ we obtain that $\mathbb{E}\left[Y(X - \tilde{X})\right] = 0 \ \forall Y \in M$. In general this equation is called **predicion equation**. In general, as $Z_1, ..., Z_h \in M$ then $\mathbb{E}\left[(X - \tilde{X})Z_i\right] = 0$ with $i = 1, ..., h$ are called predicion equations. Observe that these equations are the same employed to compute the coefficients of a linear regression over $Z_1, ..., Z_h$ with the least square method.

**Exercise 13.2.** *Consider $(X_t)$ stationary with constant mean ($\mathbb{E}[X_t] = \mu \ \forall t \in \mathbb{Z}$). Prove that*

$$P_{\bar{sp}\{1, X_t, ..., X_{t+h-1}\}}(X_{t+h}) = \mu + P_{\bar{sp}\{Y_t, ..., Y_{t+h-1}\}}(Y_{t+h})$$

*where $Y_t = X_t - \mu \ \forall t \in \mathbb{Z}$. This exercise is interesting, beacause if $\mu = 0$ then*

$$P_{\bar{sp}\{1, X_t, ..., X_{t+h-1}\}}(X_{t+h}) = P_{\bar{sp}\{X_t, ..., X_{t+h-1}\}}(X_{t+h})$$

*The second memeber of the equation is of particular interest, since we will work with time series with zero mean.*

Now we will see how to compute the coefficient of the BLP. Suppose

1. $\tilde{X}_{h+1} = \alpha_{h-1,1}X_h + ... + \alpha_{h-1,h-1}X_2 = P_{sp\{X_2,\bar{...},X_h\}}(X_{h+1})$
2. $\tilde{X}_1 = \beta_{h-1,1}X_2 + ... + \beta_{h-1,h-1}X_h = P_{sp\{X_2,\bar{...},X_h\}}(X_1)$

Denote $\boldsymbol{\alpha}_{(h-1)} = (\alpha_{h-1,1}; ...; \alpha_{h-1,h-1})^{\intercal}$, $\boldsymbol{\beta}_{(h-1)} = (\beta_{h-1,1}; ...; \beta_{h-1,h-1})^{\intercal}$ and $\boldsymbol{\Gamma}_{h-1} = [\gamma(i-j)]_{i,j=1}^{h-1}$ where $\gamma$ is the ACF of $(X_t)$.

**Proposition 13.2.** *If $\Gamma_{h-1}$ is non-singular then*

1. $\boldsymbol{\alpha}_{(h-1)} = \boldsymbol{\Gamma}_{h-1}^{-1}\boldsymbol{\gamma}_{h-1}$
2. $\boldsymbol{\beta}_{h-1} = \boldsymbol{\alpha}_{h-1}$

*Proof.* From the prediciton theorem we have that

$$\mathbb{E}\left[(X_{h+1} - \tilde{X}_{h+1})X_k\right] = 0$$

Remembering the condition

$$\tilde{X}_{h+1} = \alpha_{h-1,1}X_h + ... + \alpha_{h-1,h-1}X_2 = P_{sp\{X_2,...,X_h\}}(X_{h+1})$$

we have that

$$\begin{aligned}\mathbb{E}\left[X_{h+1}X_k\right] &= \sum_{j=1}^{h-1} \alpha_{h-1,j}\mathbb{E}\left[X_{h+1-j}, X_k\right] \\ \gamma(h+1-k) &= \sum_{j=1}^{h-1} \alpha_{h-1,j}\gamma(h+1-k-j)\end{aligned}$$

Set $i = h+1-k$, then for $k = 2$ we have that $i = h-1$ and for $k = h$ we have that $i = 1$. We can now rewrite the previous equation as

$$\gamma(i) = \sum_{j=1}^{h-1} \alpha_{h-1,j}\gamma(i-j)$$

for $i = 1, ..., h-1$. In matrix notation:

$$\boldsymbol{\gamma}_{(h-1)} = \boldsymbol{\Gamma}_{h-1}\boldsymbol{\alpha}_{h-1}$$

Then the fist statement follows by multiplying both sides of the equation by $\boldsymbol{\Gamma}_{h-1}^{-1}$. The second statemnet is proved in the same way and is thus left as an exercise. □

But how can we compute the auto correlation in $h$? From the definition: multivariate integral with joint distribution of $X_2, ..., X_h$; but we can also use the $h$-th coefficient in a linear regression of $X_{h+1}$ on $X_h, ..., X_2, X_1$.

# 14 Lecture 14

**The computation of the partial autocorrelation function. Mean square error. The special case of AR(p). Examples of partial autocorrelation functions in R. Forecasting: best linear predictors (BLP's). Recursive methods for the computation of the coefficients of the BLP: the Durbin-Levinson algorithm.**

**Theorem 14.1.** *If $(X_t) \sim AR(p)$ causal then*

1. $\pi_{pp} = \theta_p$
2. $\pi_{hh} = 0 \quad h > p$

*Proof.* Compare

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ ... \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & ... & \rho(p-1) \\ \rho(1) & 1 & ... & \rho(p-2) \\ ... & ... & ... & ... \\ \rho(p-1) & \rho(p-2) & ... & 1 \end{pmatrix} \begin{pmatrix} \alpha_{p,1} \\ \alpha_{p,2} \\ ... \\ \alpha_{p,p} \end{pmatrix}$$

versus

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ ... \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & ... & \rho(p-1) \\ \rho(1) & 1 & ... & \rho(p-2) \\ ... & ... & ... & ... \\ \rho(p-1) & \rho(p-2) & ... & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ ... \\ \theta_p \end{pmatrix}$$

The first system of equations necessary to compute the PACF such that $\pi_{p,p} = \alpha_{p,p}$ while the second is the YW set of equations for $AR(p)$. The two sets of equations are equal, since $(\alpha_{p,1}), ...\alpha_{p,p} = (\theta_1, ..., \theta_p)$, so *1.* follows. For the point *2.*, as $p \geq 1$ and $h > p$ then $h \geq 2$. Consider $\theta(B)X_t = W_t$ for $t = h+1$; we have that

$$\begin{aligned} W_{h+1} &= X_{h+1} - \theta_1 X_h + ... - \theta_p X_{h+1-p} \\ &= X_{h-1} - \sum_{j=1}^{p} \theta_j X_{h+1-j} \end{aligned}$$

We have to prove that $X_{h-1} - \sum_{j=1}^{p} \theta_j X_{h+1-j} \in \bar{sp}\left\{X_2, ..., X_h\right\}^{\perp}$, that is equivalent to prove

$$\mathbb{E}\left[ Y(X_{h+1} * \sum_{j=1}^{p}) \theta_j X_{h+1-j} \right] = 0$$

with $Y \in \bar{sp}\left\{X_2, ..., X_h\right\}$. As $(X_t)$ is causal we have that

$$\begin{aligned} X_2 &\in \bar{sp}\left\{W_2, W_1, ...\right\} \\ &... \\ X_h &\in \bar{sp}\left\{W_h, W_{h-1}, ...\right\} \end{aligned}$$

implying that $Y \in \bar{sp}\left\{W_j, j \leq h\right\}$ with $h \geq 2$, which in turn imply that $\mathbb{E}\left[YW_{h+1}\right] = 0$, proving the èrevious equation. Moreover

$$\sum_{J=1}^{p} \theta_j X_{h+1-j} = P_{\bar{sp}\left\{X_2, ..., X_h\right\}}(X_{h+1})$$

|  | $AR(p)$ | $MA(1)$ | $ARMA(p,q)$ |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ | Tails off |
| PACF | Cuts off after lag $p$ | Tails off | Tails off |

Table 1: Behaviour of the ACF and PACF for causal and invertible $ARMA$ models

Consider

$$\pi_{hh} = corr\left(X_{h+1} - \sum_{j=1}^{p} X_{h+1-j}, X_1 - \tilde{X}_1\right)$$

but we can observe that

$$\pi_{hh} = corr\left(W_{h+1}, f(X_1, X_2, ..., X_h) \in \bar{sp}\left\{W_j, j \leq h\right\}\right) = 0$$

concluding the proof. □

**Exercise 14.1.** *Suppose*

1. *$(X_t)$ stationary, prove that $\pi_{22} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2}$*
2. *$(X_t) \sim MA(1)$, find $\pi_{22}$ in terms of $\phi_1$.*

What about $\pi_{hh}$ for $MA(1)$? Suppose $(X_t) \sim MA(1)$ invertible; then we know that $W_t = \psi(B)X_t$ with $\psi(Z) = \sum_{j\geq 0}\psi_j Z^j$ and $\sum_{j\geq 0}|\psi_j| < \infty$. As $\phi_0 = 0$ we have that $\psi_0 = 1$, and so we can write

$$W_t = X_t + \sum_{j\geq 1}\psi_j X_{t-j}$$

which is a $AR(\infty)$ representation, implying that $\pi_{hh} \neq 0$ for $h > q$. A similar statement can be proven for $ARMA(p,q)$. More details on this can be found in the table 14.

How to estimate the PACF? One method consists in fitting and AR model starting from order 1 and continuing with 2,3,... using the least-squares method and picking out the last coefficient at each step, so that $\hat{\theta}_1 = \hat{\alpha}_{11}, \hat{\theta}_2 = \hat{\alpha}_{22}, \hat{\theta}_3 = \hat{\alpha}_{33}$ and so on. Another method exploits the set linear equations

$$\boldsymbol{\rho}_{(h)} = \boldsymbol{R}_h \boldsymbol{\alpha}_{(h)}$$

by substituting the theoretical ACF with the approximated one. The standard error for the PACF is $\simeq \frac{1}{\sqrt{n}}$ for $h > p$.

**Example 14.1.** *Now we will study in R the PACF.*

```
#########################################################
# The behavior of the PACF
#########################################################

library(astsa)
library(polynom)
#
# AR(1)
#
set.seed(154)

windows()
ar1.sim1=arima.sim(list(ar=c(0.9)),200)
acf2(ar1.sim1, main='AR(1) with th1=0.9')
windows()
ar1.sim2=arima.sim(list(ar=c(-0.9)),200)
acf2(ar1.sim2, main='AR(1) with th1=-0.9')
#
# AR(2)
#
set.seed(154)

windows()
(roots=solve(polynomial(c(1,-0.5,-0.4))))
ar2.sim1=arima.sim(list(ar=c(0.5,0.4)),200)
acf2(ar2.sim1,main='AR(2) with th1=0.5, th2=0.4')
#
windows()
(roots=solve(polynomial(c(1,0.5,-0.4))))
ar2.sim2=arima.sim(list(ar=c(-0.5,0.4)),200)
acf2(ar2.sim2,main='AR(2) with th1=-0.5, th2=0.4')

#
# MA(1)
#
set.seed(154)
windows()
ma1.sim1=arima.sim(list(ma=c(0.9)),200)
acf2(ma1.sim1,main='MA(1) with phi1=0.9')
#
windows()
ma1.sim2=arima.sim(list(ma=c(-0.9)),200)
acf2(ma1.sim2,main='MA(1) with phi1=-0.9')

#
# MA(2)
#
set.seed(154)
windows()ma2.sim1=arima.sim(list(ma=c(2.1,0.9)),200)
acf2(ma2.sim1,main='MA(2) with ph1=2.1, ph2=0.9')
windows()
ma2.sim2=arima.sim(list(ma=c(-2.1,0.9)),200)
acf2(ma2.sim2,main='MA(2) with ph1=-2.1, ph2=0.9')

#
```

```
# ARMA(1,1)
#
set.seed(154)
windows()
arma.sim1=arima.sim(list(ma=c(0.9),ar=c(0.6)),200)
acf2(arma.sim1,main='ARMA(1,1) with ph1=0.9, ph2=0.6')
#
windows()
arma.sim2=arima.sim(list(ma=c(-0.9),ar=c(-0.6)),200)
acf2(arma.sim2,main='ARMA(1,1) with ph1=-0.9, ph2=-0.6')

# the dataset jj
#
data(jj)

# decomposition in an additive model
#
comp=stl(jj,'per')
# To see the decomposition
#
windows()
plot(comp)
# The procedure gives in output a matrix comp$time.series: the residuals are
# in the third column
head(comp$time.series,6)
# ACF and PACF of residuals
windows()
acf2(comp$time.series[,3],max.lag=40,main='Residuals JJ')
```

Now we will introduce forecasting: in this case the goal is to predict the future value of a time series($x_{n+k}$) based on the observed value up to the present $(x_1, ..., x_n)$. $k$ is said **lead time**. Forecasting is a case of **extrapolation**.

How much an event is predictable depends on several factors, like

- the understanding of the factors;
- the available data;
- the influence of the forecast on the event;

**Example 14.2.** *Electricity demand well suits forecast, since we know very well what are the factors that influence it (temperature, season, economic conditions...), we have much historical data and the forecasting does not influence the future electricity demand.*

*On the other hand, the currency exchange rate satisfies only one of these conditions: the historical data. We do not fully understand all the factors that lead to a change in these rates, and the forecast itself has a direct impact on them, meaning that the forecast influence itself.*

There are two different types of forecasting methods:

- **qualitative**: which is based on subjective assumption and, in many cases, it is the only available option;
- **quantitative**: it is based on numerical information about the past and on the assumption that the past patterns will continue in the future.

The choice of the method depends on various factors, like the aim of the forecast, the property of the time series, the available data and the long/short term aim of the forecast. It is not unusual to use multiple methods and then compare the results.

When choosing a model, it is usually necessary to partition the dataset into two subsets: the **training set** (typically 80% of the original dataset)is used to estimate the parameters of the forecasting model, and the **test set** (typically 20% of the original dataset) to evaluate its accuracy. Of course, the training set contains the data from the beginning up to a specific time, and the test set the remaining data. It is not obvious that a model that fits well the training set will have good accuracy on the test set; how can we choose the best model to forecast? There is no exact answer to this question; typically, a good model has a good forecasting accuracy and a Gaussian residual. Remember the **principle of parsimony**: choose the simplest model (fewest parameters) that fits the evidence. In the following, we will assume that $(X_t)$ is stationary and that the model's parameters are known.

We have seen that the minimum mean square error predictor of $\tilde{X}_{n+k}$ is the conditional expectation:

$$\tilde{X}_{n+k} = \mathbb{E}\left[X_{n+k}|X_1, X_2, ..., X_n\right]$$

since

$$\mathbb{E}\left[X_{n+k} - \tilde{X}_{n+k}\right]^2 = \inf_{Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})} \mathbb{E}\left[X_{n+k} - Y\right]^2$$

where $\mathcal{F} = \sigma(X_1, X_2, ..., X_n)$. Remember that $\tilde{X}_{n+k}$ is the BLP of $X_1, ..., X_n$, that is

$$\tilde{X}_{n+k} = \alpha_{n,1}^{(k)} X_n + ... + \alpha_{n,n}^{(k)} X_1 = \sum_{j=1}^{n} \alpha_{n,j}^{(k)} X_{n+1-j}$$

where $\tilde{X}_{n+k} = P_{\bar{sp}\{X_1,...,X_n\}}(X_{n+k})$ is the **k-step predictor**. How to compute $\left\{\alpha_{n,j}^{(k)}\right\}$? Using the prediction equations:

$$\mathbb{E}\left[\left(X_{n+k} - \sum_{j=1}^{n} \alpha_{n,j}^{(k)} X_{n+1-j}\right) X_{n+1-i}\right] = 0$$

for $i = 1, ..., n$. From some computation we find that

$$\mathbb{E}\left[X_{n+k} X_{n+1-i}\right] = \sum_{j=1}^{n} \alpha_{n,j}^{(k)} \mathbb{E}\left[X_{n+1-j} X_{n+1-i}\right]$$

which can be written in matrix notation as

$$\boldsymbol{\gamma}_{(n)}^{(k)} = \boldsymbol{\Gamma}_n \boldsymbol{\alpha}_{(n)}^{(k)}$$

which (explicitly) is

$$\begin{pmatrix} \gamma(k) \\ \gamma(k+1) \\ ... \\ \gamma(k+n-1) \end{pmatrix} = [\gamma(i-j)]_{i,j=1}^{n} \begin{pmatrix} \alpha_{n,1}^{(k)} \\ ... \\ \alpha_{n,n}^{(k)} \end{pmatrix}$$

For $k = 1$, we have the 1-step predictor

$$\boldsymbol{\gamma}_{(n)}^{(1)} = \boldsymbol{\Gamma}_n \boldsymbol{\alpha}_{(n)}^{(1)}$$

which is the usual system of equations that we have seen in the PACF. In particular, we can use the correlation matric by normalizing by the correlation of the time series and recover that

$$\boldsymbol{\rho}_{(n)} = \boldsymbol{R}_n \boldsymbol{\alpha}_{(n)}$$

that are

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ ... \\ \rho(n) \end{pmatrix} = [\rho(i-j)]_{i,j=1}^{n} \begin{pmatrix} \alpha_{n,1} \\ ... \\ \alpha_{n,n} \end{pmatrix}$$

As a special case, the 1-step predictor of a causal $AR(p)$ is

$$\tilde{X}_{n+1} = \theta_1 X_n + ... + \theta_p X_{n+1-p} \quad n > p$$

since with the PACF we have seen that

$$\alpha_{n,j} = \begin{cases} \theta_j & j = 1, ..., p \\ 0 & j = p+1, ..., n \end{cases}$$

What about the error of replacing $X_{n+k}$ with its BLP $\tilde{X}_{n+k}$? As always we can use the mean squared error:

$$\begin{aligned} MSE(\tilde{X}_{n+k}) &= \mathbb{E}\left[(X_{n+k} - \tilde{X}_{n+k})^2\right] \\ &= \mathbb{E}\left[(X_{n+k} - \tilde{X}_{n+k})(X_{n+k} - \tilde{X}_{n+k})\right] \\ &= \mathbb{E}\left[X_{n+k}(X_{n+k} - \tilde{X}_{n+k})\right] - \mathbb{E}\left[\tilde{X}_{n+k}(X_{n+k} - \tilde{X}_{n+k})\right] \end{aligned}$$

note that $\tilde{X}_{n+k} \in \bar{sp}\{X_1, ..., X_n\}$ and $(X_{n+k} - \tilde{X}_{n+k}) \in \bar{sp}\{X_1, ..., X_n\}$, so the second expectation is equal to zero. Continuing with the computation

$$
\begin{aligned}
MSE(\tilde{X}_{n+k}) &= \mathbb{E}\left[X_{n+k}(X_{n+k}-\tilde{X}_{n+k})\right] - \mathbb{E}\left[\tilde{X}_{n+k}(X_{n+k}-\tilde{X}_{n+k})\right] \\
&= \mathbb{E}\left[X_{n+k}(X_{n+k}-\tilde{X}_{n+k})\right] \\
&= \mathbb{E}\left[X_{n+k}^2\right] - \mathbb{E}\left[\sum_{j=1}^{n} \alpha_{n,j}^{(k)} X_{n+k}X_{n+1-j}\right] \\
&= \mathbb{E}\left[X_{n+k}^2\right] - \sum_{j=1}^{n} \alpha_{n,j}^{(k)} \mathbb{E}\left[X_{n+k}X_{n+1-j}\right] \\
&= \gamma(0) - \sum_{j=1}^{n} \alpha_{n,j}^{(k)} \gamma(k-1+j) \\
&= \gamma(0) - \left(\boldsymbol{\gamma}_{(n)}^{(k)}\right)^{\intercal} \boldsymbol{\alpha}_{(n)}^{(k)}
\end{aligned}
$$

Now let us get back to the resolution of the previous system of equations

$$\boldsymbol{\gamma}_{(n)}^{(k)} = \boldsymbol{\Gamma}_n \boldsymbol{\alpha}_{(n)}^{(k)}$$

If $\boldsymbol{\Gamma}_n$ is non-singular, then $\exists! \boldsymbol{\alpha}_{(n)}^{(k)}$ such that $\boldsymbol{\alpha}_{(n)}^{(k)} = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_{(n)}^{(k)}$. Otherwise, if $\boldsymbol{\Gamma}_n$ is singular then there may be many solutions, but all of them will lead to the same predictor due to the projection theorem.

**Exercise 14.2.** *Suppose* $(X_t) = A\cos(\omega t) + B\sin(\omega t),\ \omega \in (0,\pi))$ *where* $A,\ B$ *are uncorrelated random variables such that* $\mathbb{E}[A] = \mathbb{E}[B] = 0$ *and* $\mathbb{E}\left[A^2\right] = \mathbb{E}\left[B^2\right] = \sigma^2$.

- *Prove that* $\alpha_{1,1} = \cos(\omega),\ \boldsymbol{\alpha}_{(2)} = (2\cos(\omega), -1)^{\intercal}$ *with* $k = 1$.
- *Prove that* $X_3 = 2\cos(\omega)X_2 - X_1$ *and compute* $MSE(\tilde{X}_3)$.
- *Find two expression of* $\tilde{X}_4$ *in terms of* $X_1, X_2, X_3$.
- *Check if* $\boldsymbol{\Gamma}_3$ *is singular.*

We can compute the 1-step predictor $\tilde{X}_{h+1}$ iteratively by starting from $h = 1$ up to the fixed $n$ without requiring any matrix inversion. We will see the **Durbin-Levinson** algorithm:

- $$h = 1 \quad v_0^2 = \gamma(0) \quad \alpha_{1,1} = \frac{\gamma(1)}{\gamma(0)} = \rho(1) \Leftarrow \begin{cases} v_0^2 = MSE(\tilde{X}_2) \\ \tilde{X}_2 = \alpha_{1,1}X_1 \qquad \alpha_{1,1} = \pi_{11} \end{cases}$$

- $$h = 2 \quad v_1^2 = (1-\alpha_{1,1}^2)v_0^2 \quad \alpha_{2,2} = \frac{\gamma(2) - \alpha_{1,1}\gamma(1)}{v_1^2} \Leftarrow \begin{cases} v_1^2 = MSE(\tilde{X}_3) \\ \tilde{X}_3 = \alpha_{2,1}X_2 + \alpha_{2,2}X_1 \quad \alpha_{2,2} = \pi_{22} \end{cases}$$

- $$h = 3 \quad v_2^2 = (1-\alpha_{2,2}^2)v_1^2 \quad \alpha_{3,3} = \frac{\gamma(3) - \sum_{j=1}^{2}\alpha_{2,j}\gamma(3-j)}{v_2^2} = \Leftarrow \begin{cases} v_2^2 = MSE(\tilde{X}_4) \\ \tilde{X}_4 = \alpha_{3,1}X_3 + \alpha_{3,2}X_2 + \alpha_{3,3}X_1 \quad \alpha_{3,3} \end{cases}$$

- ...

up to $h = n$. This method as the advantages that we know the values of $\{\pi_{h,h}\}$ and we have the BLP $\left\{\tilde{X}_{h+1}\right\}$ for $h = 1, ..., n$.

*Proof.* The main idea of this algorithm is that

$$\bar{sp}\left\{X_1, ..., X_h\right\} = \bar{sp}\left\{X_2, ..., x_h\right\} \oplus \bar{sp}\left\{X_1 - \tilde{X}_1\right\}$$

where $\tilde{X}_1 = P_{\bar{sp}\{X_2,...,X_h\}}(X_1)$, beacuse

$$\bar{sp}\left\{X_2, ..., X_h\right\} \perp \bar{sp}\left\{X_1 - \tilde{X}_1\right\}$$

implying that

$$\begin{aligned}\tilde{X}_{h+1} &= P_{\bar{sp}\{X_1,...,X_h\}}(X_{h+1}) \\ &= P_{\bar{sp}\{X_2,...,X_h\}}(X_{h+1}) + a(X_1 - P_{\bar{sp}\{X_2,...,X_h\}}(X_1))\end{aligned}$$

where $P_{\bar{sp}\{X_2,...,X_h\}}(X_{h+1})$ and $P_{\bar{sp}\{X_2,...,X_h\}}(X_1)$ are the BLP in $\pi_{hh}$. □

How to compute $\tilde{X}_{n+k}$ iteratively?

1. If $X \in \bar{sp}\left\{X_1, ..., X_n\right\}$ then $P_{\bar{sp}\{X_1,...,X_n\}}(X) = X$
2. If $M_1, M_{\subseteq}\mathcal{L}^2$ such that $M_1 \subseteq M_2$ then $P_{M_1}(X) = P_{M_1}(P_{M_2}(X))$

Suppose $k = 2$ and to want to compute $\tilde{X}_{n+2} = P_{\bar{sp}\{X_1,...,X_n\}}(X_{n+2})$. As $\bar{sp}\left\{X_1, ..., X_n\right\} \subseteq \bar{sp}\left\{X_1, ..., X_n, X_{n+1}\right\}$ we have that

$$\begin{aligned}\tilde{X}_{n+2} &= P_{\bar{sp}\{X_1,...,X_n\}}(X_{n+2}) \\ &= P_{\bar{sp}\{X_1,...,X_n\}}(P_{\bar{sp}\{X_1,...,X_{n+1}\}}(X_{n+2})) \quad \textit{for condition } 2. \\ &= P_{\bar{sp}\{X_1,...,X_n\}}(\alpha_{n+1,1}X_{n+1} + \alpha_{n+1,2}X_n + ... + \alpha_{n+1,n+1}X_1) \quad \textit{for linearity} \\ &= \alpha_{n+1,1}P_{\bar{sp}\{X_1,...,X_n\}}(X_{n+1}) + \alpha_{n+1,2}P_{\bar{sp}\{X_1,...,X_n\}}(X_n) + ... + \alpha_{n+1,n+1}P_{\bar{sp}\{X_1,...,X_n\}}(X_1) \\ &= \alpha_{n+1,1}\tilde{X}_{n+1} + \alpha_{n+1,2}X_n + ... + \alpha_{n+1,n+1}X_1\end{aligned}$$

So the final structure of the algorithm is

```
compute $\hat{\gamma}(0), ..., \hat{\gamma}(n)$
for $h \in 1, ..., n$ do
    compute $\hat{v}^2_{h-1}$
    compute $\alpha_{\hat{h},1}; ...; \alpha_{\hat{h},h}$
    compute $\tilde{x}_{h+1}$
end for
if $k \geq 2$ then
    for $h = n, ..., n + k - 2$ do
        repeat 1. and 2. adding to the sample $\tilde{x}_{h+1}$ to obtain $\tilde{x}_{h+2}$
    end for
end if
```

# 15 Lecture 15

**The innovations algorithm. Properties of innovations. Recursions for the mean square errors. Applications to MA(q) and ARMA(p,q) models: the fitted innovations MA(q) and ARMA(p,q) models. Maximum likelihood estimators for Gaussian time series. ARIMA models.**

There is a second method allowing us to compute the BLP, the **innovations algorithm**. This method can be applied to any series with finite second moment (stationary or not).

**Proposition 15.1.** *Suppose:*

- $(X_t) \in \mathcal{L}^2$
- $\mathbb{E}\left[X_t\right] = 0 \quad \forall t \in \mathbb{Z}$
- $\tilde{X}_h = \begin{cases} 0 & h = 1 \\ P_{\overline{sp}\{X_1,...,X_{h-1}\}}(X_h) = \sum_{j=1}^{h-1} \alpha_{h-1,j} X_{h-j} & h = 2,3,... \end{cases}$
- $U_h = X_h - \tilde{X}_h \quad h = 1,2,...$

*then*

$$\tilde{X}_{h+1} = \begin{cases} 0 & h = 0 \\ \sum_{j=1}^{h} b_{h,j} U_{h+1-j} & h = 1,2,... \end{cases}$$

*where*

$$b_{h,j} = \left(\boldsymbol{A}_h^{-1} - \boldsymbol{I}_h\right)_{h,j}$$

*and*

$$\left(\boldsymbol{A}_h\right)_{ij} = \begin{cases} 0 & i < j \\ 1 & i = j \\ -\alpha_{i-1,i-j} & i > j \end{cases}$$

**Remark 15.1.** *The sequence $U_h$ is the sequence of innovations, that is the 1-step predicion errors sequence.*

*Proof.* Consider

$$\begin{aligned} U_1 &= X_1 - \tilde{X}_1 = X_1 \\ U_2 &= X_2 - \tilde{X}_2 = X_2 - \alpha_{1,1} X_1 \\ U_3 &= X_3 - \tilde{X}_3 = X_3 - \alpha_{2,1} X_2 - \alpha_{2,2} X_1 \\ &... \end{aligned}$$

that is

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ ... \\ U_h \end{pmatrix} = \begin{pmatrix} 1 & 0 & ... & 0 \\ -\alpha_{1,1} & 1 & ... & 0 \\ -\alpha_{2,1} & -\alpha_{2,2} & ... & 0 \\ ... & ... & ... & ... \\ -\alpha_{h-1,h-1} & \alpha_{h-1,h-2} & ... & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ ... \\ X_h \end{pmatrix}$$

in matrix notation

$$\boldsymbol{U}_{(h)} = \boldsymbol{A}_h \boldsymbol{X}_{(h)}$$

The determinant of $\boldsymbol{A}_h = 1$ because it is a lower triangular matrix, so its inverse exists; then $\boldsymbol{X}_{(h)} = \boldsymbol{A}_h^{-1}\boldsymbol{U}_{(h)}$. Now consider

$$\boldsymbol{U}_{(h)} = \boldsymbol{X}_{(h)} - \tilde{\boldsymbol{X}}_{(h)}$$

this implies that

$$\begin{aligned}\tilde{\boldsymbol{X}}_{(h)} &= \boldsymbol{X}_{(h)} - \boldsymbol{U}_{(h)} \\ &= \boldsymbol{A}_h^{-1}\boldsymbol{U}_{(h)} - \boldsymbol{U}_{(h)} \\ &= \left(\boldsymbol{A}_h^{-1} - \boldsymbol{I}_h\right)\boldsymbol{U}_{(h)} \\ &= \boldsymbol{B}_h\boldsymbol{U}_{(h)}\end{aligned}$$

then

$$\boldsymbol{B}_h = \begin{pmatrix} 0 & 0 & 0 & ... & 0 \\ b_{1,1} & 0 & 0 & ... & 0 \\ b_{2,1} & b_{2,2} & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ b_{h-1,h-1} & b_{h-1,h-2} & b_{h-1,h-3} & ... & 0 \end{pmatrix}$$

Expanding the previous product gives the final result and completes the proof.

□

**Example 15.1.** $\tilde{X}_1 = 0,\ \tilde{X}_2 = b_{1,1}(X_1 - \tilde{X}_1) = b_{1,1}X_1,\ \tilde{X}_3 = b_{2,2}(X_1 - \tilde{X}_1) + b_{2,1}(X_2 - \tilde{X}_2), ...$

The innovations have interesting properties:

1. $sp\left\{U_1, ..., U_h\right\} = sp\left\{X_1..., X_h\right\}$ since in $U_j$, with $J = 1, ..., h$, these is the contribution of $X_j$ and $\tilde{X}_j$, which in turn contains the contributions of $X_1, ..., X_{j-1}$.

2. $(U_h)$ are uncorrelated random variables, that is $\mathbb{E}\left[U_iU_j\right] = 0$ for $i \neq j$. It can be prove in many ways; this is one:

   *Proof.* Consider

   $$\begin{aligned}\mathbb{E}\left[U_iU_j\right] &= \mathbb{E}\left[(X_i - \tilde{X}_i)(X_j - \tilde{X}_j)\right] \quad fix\ j \geq 2,\ i < j \\ &= \mathbb{E}\left[X_i(X_j - \tilde{X}_j)\right] - \mathbb{E}\left[\tilde{X}_i(X_j - \tilde{X}_j)\right]\end{aligned}$$

   Note that $\mathbb{E}\left[X_i(X_j - \tilde{X}_j)\right] = 0$, since

   $$X_i \in \bar{sp}\left\{X_1, ..., X_{j-1}\right\} \quad and \quad X_j - \tilde{X}_j \in \bar{sp}\left\{X_1, ..., X_{j-1}\right\}^{\perp}$$

   als also $\mathbb{E}\left[\tilde{X}_i(X_j - \tilde{X}_j)\right] = 0$, since

   $$\tilde{X}_i \in \bar{sp}\left\{X_1, ..., X_{i-1}\right\} \subseteq \bar{sp}\left\{X_1, ..., X_{j-1}\right\} \quad and \quad X_j - \tilde{X}_j \in \bar{sp}\left\{X_1, ..., X_{j-1}\right\}^{\perp}$$

   □

Now we will see how to recursively compute $\{b_{h,j}\}$:

$$\tilde{X}_{h+1} = \sum_{j=1}^{h} b_{h,j}(X_{h+1-j} - \tilde{X}_{h+1-j})$$
$$\mathbb{E}\left[\tilde{X}_{h+1}(X_k - \tilde{X}_k)\right] = \sum_{j=1}^{h} b_{h,j}\mathbb{E}\left[(X_{h+1-j} - \tilde{X}_{h+1-j})(X_k - \tilde{X}_k)\right]$$

Note that

$$\mathbb{E}\left[(X_{h+1-j} - \tilde{X}_{h+1-j})(X_k - \tilde{X}_k)\right] = \mathbb{E}\left[U_{h+1-j}U_k\right] \neq 0 \iff h+1-j = k \iff j = h+1-k$$

then

$$\mathbb{E}\left[\tilde{X}_{h+1}(X_k - \tilde{X}_k)\right] = b_{h,h+1-k}\mathbb{E}\left[X_k - \tilde{X}_k\right]^2$$

observe that $\tilde{X}_{h+1} = X_{h+1} + (\tilde{X}_{h+1} - X_{h+1})$ and $\mathbb{E}\left[X_k - \tilde{X}_k\right]^2 = MSE(\tilde{X}_k) = v_k^2$. Then

$$\mathbb{E}\left[X_{h+1}(X_k - \tilde{X}_k)\right] + \mathbb{E}\left[(\tilde{X}_{h+1} - X_{h+1})(X_k - \tilde{X}_k)\right] = \mathbb{E}\left[X_{h+1}(X_k - \tilde{X}_k)\right] = b_{h,h+1-k}v_k^2$$

For $k = 1$ we have that $\mathbb{E}\left[X_{h+1}X_1\right] = b_{h,h}v_1^2$, implying $b_{h,h} = \frac{cov(X_{h+1},X_1)}{v_1^2}$. For $k = 2, ..., h$ we can rewrite the last equation substituting $\tilde{X}_k$ with $\sum_{j=1}^{k-1} b_{k-1,j}(X_{k-j} - \tilde{X}_{k-j})$:

$$\begin{aligned} b_{h,h+1-k}v_k^2 &= \mathbb{E}\left[X_{h+1}X_k\right] - \sum_{j=1}^{k-1} b_{k-1,j}\mathbb{E}\left[X_{h+1}(X_{k-j}\tilde{X}_{k-j})\right] \quad k-j \in \{1, ..., k-1\} \\ &= cov(X_{h+1}X_k) - \sum_{j=1}^{k-1} b_{k-1,j}b_{h,h+1-(k-j)}v_{k-j}^2 \quad set\ k-j = 1 \\ &= cov(X_{h+1}X_k) - \sum_{i=1}^{k-1} b_{k-1,k-i}b_{h,h+1-i}v_i^2 \end{aligned}$$

Then we have that

$$b_{h,h+1-k} = \frac{cov(X_{h+1}, X_k) - \sum_{i=1}^{k-1} b_{k-1,k-i}b_{h,h+1-i}v_i^2}{v_k^2}$$

for $k = 2, 3, ..., h$.

**Example 15.2.** *Suppose $(X_t)$ stationary, then:*

- ***h=1**: compute $v_1^2$*
  - $k = 1 \to b_{1,1} = \frac{\gamma_X(1)}{v_1^2}$
- ***h=2**: compute $v_2^2$*
  - $k = 1 \to b_{2,2} = \frac{\gamma_X(2)}{v_2^2}$

- $k = 2 \to b_{2,1} = \frac{\gamma_X(1) - b_{1,1}b_{2,2}v_1^2}{v_2^2}$

- ***h=3**: compute $v_3^2$*
  - $k = 1 \to b_{3,3} = \frac{\gamma_X(3)}{v_3^2}$
  - $k = 2 \to b_{3,2} = \frac{\gamma_X(2) - b_{1,1}b_{3,3}v_1^2}{v_3^2}$
  - $k = 3 \to b_{3,1} = \frac{\gamma_X(1) - b_{2,2}b_{3,3}v_1^2 - b_{2,1}b_{3,2}v_2^2}{v_3^2}$
- ***h=4**: compute $v_4^2$*
  - $k = 1 \to b_{4,4} = \frac{\gamma_X(4)}{v_4^2}$
  - $k = 2 \to b_{4,3} = \frac{\gamma_X(3) - b_{1,1}b_{4,4}v_1^2}{v_4^2}$
  - $k = 3 \to b_{4,2} = \frac{\gamma_X(2) - b_{2,2}b_{4,4}v_1^2 - b_{2,1}b_{4,3}v_2^2}{v_4^2}$
  - $k = 4 \to b_{4,1} = \frac{\gamma_X(1) - b_{3,3}b_{4,4}v_1^2 - b_{3,2}b_{4,3}v_2^2 - b_{3,1}b_{4,2}v_3^2}{v_4^2}$

*Note that this computation simplify for $(X_t) \sim MA(q)$: $\gamma_X(h) = 0$ for $h > q$.*

Now we will see the computation of $\left\{v_h^2\right\}$: consider

$$v_h^2 = \mathbb{E}\left[(X_h - \tilde{X}_h)^2\right] = \mathbb{E}\left[X_h^2\right] + \mathbb{E}\left[\tilde{X}_h^2\right] - 2\mathbb{E}\left[X_h\tilde{X}_h\right]$$

note that $\mathbb{E}\left[(X_h - \tilde{X}_h)\tilde{X}_h\right] = 0$ since $(X_h - \tilde{X}_h) \in \bar{sp}\left\{X_1, ..., X_{h-1}\right\}^\perp$ and $\tilde{X}_h \in \bar{sp}\left\{X_1, ..., X_{h-1}\right\}$; so $\mathbb{E}\left[X_h\tilde{X}_h\right] = \mathbb{E}\left[\tilde{X}_h^2\right]$. We than have

$$v_h^2 = \mathbb{E}\left[X_h^2\right] - \mathbb{E}\left[\tilde{X}_h^2\right]$$

for $h = 1, 2, ....$ In particular, for $h = 1$ we have $v_1^2 = \mathbb{E}\left[X_1^2\right] - \mathbb{E}\left[\tilde{X}_1^2\right] = \mathbb{E}\left[X_1^2\right] = Var(X_1)$. For $h = 2, 3, ...$

$$\begin{aligned}
v_h^2 &= Var(X_h) - \mathbb{E}\left[\left(\sum_{j=1}^{h-1} b_{h-1,j}(X_{h-j} - \tilde{X}_{h-j})\right)^2\right] \\
&= Var(X_h) - \sum_{j=1}^{h-1} b_{h-1,j}^2 \mathbb{E}\left[\left(X_{h-j} - \tilde{X}_{h-j}\right)^2\right] \\
&= Var(X_h) - \sum_{j=1}^{h-1} b_{h-1,j}^2 MSE(\tilde{X}_{h-j}) \\
&= Var(X_h) - \sum_{j=1}^{h-1} b_{h-1,j}^2 v_{h-j}^2
\end{aligned}$$

That means that in the previous example we can replace *compute $v_1^2$* with $\gamma_X(0)$, *compute $v_2^2$* with $\gamma_X(0) - b_{1,1}^2 v_1^2$, *compute $v_3^2$* with $\gamma_X(0) - b_{2,1}^2 v_2^2 - b_{2,2}^2 v_1^2$ and so on.

**Exercise 15.1.** *Find the 1-step predictor of $MA(1)$.*

It is also possible to apply the innovations algorithm to $ARMA$ model, and it simplifies drastically. The idea is to apply it tp a suitable transformation of $(X_t)$.

**Proposition 15.2.** *If $(X_t) \sim ARMA(p, q)$ and*

$$Y_t = \begin{cases} \frac{X_t}{\sigma} & t = 1, 2, ..., m \\ \theta(B)\frac{X_t}{\sigma} & t > m \end{cases}$$

*with $\sigma^2 = \mathbb{E}\left[W_t^2\right]$ the variance of the white noise and $m = \max\{p, q\}$ with $p, q \geq 1$ then*

1. $\bar{sp}\{Y_1, ..., Y_h\} = \bar{sp}\{X_1, ..., X_h\} \quad \forall h \geq 1$
2. $Y_h - \tilde{Y}_h = \frac{1}{\sigma}(X_h - \tilde{X}_h) \quad \forall h \geq 1$

*Proof.* The first point i straightforward to prove, since $Y_t$ is a linear combination of $X_t$. We will now prove the second point. For $h \leq m$ we have that

$$\tilde{Y}_h = P_{\bar{sp}\{Y_1,...,Y_{h-1}\}}(Y_h) = P_{\bar{sp}\{X_1,...,X_{h-1}\}}\left(\frac{X_h}{\sigma}\right) = \tilde{X}_{\frac{h}{\sigma}}$$

For $h > m$ instead

$$\begin{aligned} \tilde{Y}_h &= P_{\bar{sp}\{Y_1,...,Y_{h-1}\}}(Y_h) \\ &= \sigma^{-1}P_{\bar{sp}\{X_1,...,X_{h-1}\}}(X_h - \theta_1 X_{h-1} + ... - \theta_p X_{h-p}) \\ &= \sigma^{-1}\left[P_{\bar{sp}\{X_1,...,X_{h-1}\}}(X_h) - \theta_1 P_{\bar{sp}\{X_1,...,X_{h-1}\}}(X_{h-1}) + ... - \theta_p P_{\bar{sp}\{X_1,...,X_{h-1}\}}(X_{h-p})\right] \\ &= \sigma^{-1}\left[\tilde{X}_h - \theta_1 X_{h-1} + ... - \theta_p X_{h-p}\right] \end{aligned}$$

therefore

$$Y_h - \tilde{Y}_h = \frac{\theta(B)X_h}{\sigma} - \frac{\tilde{X}_h - \theta_1 X_{h-1} + ... - \theta_p X_{h-p}}{\sigma} = \frac{X_h - \tilde{X}_h}{\sigma}$$

completing the proof. □

**Proposition 15.3.** *If $(X_t) \sim ARMA(p, q)$ and*

$$Y_t = \begin{cases} \frac{X_t}{\sigma} & t = 1, 2, ..., m \\ \theta(B)\frac{X_t}{\sigma} & t > m \end{cases}$$

*with $\sigma^2 = \mathbb{E}\left[W_t^2\right]$ the variance of the white noise and $m = \max\{p, q\}$ with $p, q \geq 1$ then*

1. $\tilde{Y}_{h+1} = \begin{cases} \sum_{j=1}^{h} b_{h,j}(Y_{h+1-j} - \tilde{Y}_{h+1-j}) & 1 \leq h < m \\ \sum_{j=1}^{q} b_{h,j}(Y_{h+1-j} - \tilde{Y}_{h+1-j}) & h \geq m \end{cases}$
2. $\tilde{X}_{h+1} = \begin{cases} \sum_{j=1}^{h} b_{h,j}(X_{h+1-j} - \tilde{X}_{h+1-j}) & h < m \\ \theta_1 X_h + ... + \theta_p X_{h+1-p} + \sum_{j=1}^{q} b_{h,j}(X_{h+1-j} - \tilde{X}_{h+1-j}) & h \geq m \end{cases}$

*Proof.* For the first point there is little to say, since for $h < m$ the formula is just the application of the innovations algorithm to $Y_t$, while for $h > m$ we know that $(Y_t) \sim MA(q)$ and there there is a cutoff in the coefficient of the innovations algorithm, and for $h = m$ the BLP $\tilde{Y}_{m+1} \to Y_{m+1}$, which is a $MA(q)$ and then as the same cutoff.

About the second point we have that the first equation follows from the first point, since for $h < m$ we know that $h + 1 \leq m$, and so $\tilde{Y}_{h+1} = \frac{\tilde{X}_{h+1}}{\sigma}$. For the second equation note that

$$\tilde{Y}_{h+1} = \sum_{j=1}^{q} b_{h,j}(Y_{h+1-j} - \tilde{Y}_{h+1-j})$$

$$\sigma^{-1}(\tilde{X}_{h+1} - \theta_1 X_h + ... - \theta_p X_{h+1-p}) = \sum_{j=1}^{q} b_{h,j}(X_{h+1-j} - \tilde{X}_{h+1-j})\sigma^{-1}$$

and recovering $\tilde{X}_{h+1}$ completes the proof. □

**Remark 15.2.**

$$MSE(\tilde{X}_{h+1}) = \mathbb{E}\left[(X_{h+1} - \tilde{X}_{h+1})^2\right] = \sigma^2 \mathbb{E}\left[(Y_{h+1} * \tilde{Y}_{h+1})^2\right] = \sigma^2 MSE(\tilde{X}_{h+1})$$

*Moreover, if* $(X_t)$ *is invertible then* $MSE(\tilde{Y}_{h+1}) \to 1$ *when* $h \to \infty$ *an also* $b_{h,j} \to \phi_j$ *for* $j = 1, 2, ..., q$ *when* $h \to \infty$*.*

It is also possible to fit a $MA(q)$ model with the innovations algorithm:

$$X_t = W_t + \hat{b}_{q1} W_{t-1} + ... + \hat{b}_{qq} W_{t-q}$$

with $(W_t) \sim \mathcal{WN}(0, \hat{v}_q^2)$ where the $\left\{\hat{b}_{q,j}\right\}_{j=1}^{q}$ and $\hat{v}_q^2$ are the **innovation estimates** obtained by the innovations algorithm with $\gamma_X(h)$ instead of $\hat{\gamma}_X(h)$. This is the **fitted innovations MA(q) model**. An other motivation to study this model is that $(X_t)$ is causal

$$X_t = \sum_{j \geq 0} \psi_j W_{t-j}$$

where $\psi_0 = 1$ and $\psi_j = \phi_j + \sum_{k=1}^{j} \theta_k \psi_{j-k}$ for $j = 1, 2, ...$ with $\theta_k = 0$ for $k > p$ and $\phi_j = 0$ for $j > q$ then we can estimate $\hat{\psi}_1, ..., \hat{\psi}_{p+q}$ by using the innovation estimates by solving

$$\hat{\psi}_j = \sum_{k+1}^{j} \theta_k \hat{\psi}_{j-k}$$

in $\theta_k$ for $j = q + 1, ..., q + p$, denoting the solutions with $\hat{\theta}_1, ..., \hat{\theta}_p$ and then solving

$$\hat{\psi}_j = \phi_j + \sum_{k=1}^{j} \hat{\theta}_k \hat{\psi}_{j-h}$$

in $\phi_j$ for $j = 1, 2, ..., q$ and denoting the results as $\hat{\phi}_1, ..., \hat{\phi}_1$.

An other method that can be used in order to fit $ARMA(p, q)$ models is the **maximum likelihood estimator**. It has often a lower variance than other methods, but it requires the time series to be Gaussian, and is usually asymptotically robust; it also has optimization probles and requires a good starting point to converge to a good solution. The methods that we have just seen can provide a good starting point.

Suppose $(X_t)$ a Gaussian $ARMA(p,q)$ model and $n$ as sample size. Then the likelihood function is

$$\mathcal{L}(\boldsymbol{x}_n|\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)=\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)$$

where $\boldsymbol{x}_n=(x_1,...,x_n)^\intercal$, $\boldsymbol{\phi}=(\phi_1,...,\phi_q)^\intercal$, $\boldsymbol{\theta}=(\theta_1,...,\theta_p)$ and $\sigma^2=\mathbb{E}\left[W_t^2\right]$. Note that this likelihood function is equal to

$$\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}(det\boldsymbol{\Gamma}_n)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{x}_{(n)}^\intercal\boldsymbol{\Gamma}_n^{-1}\boldsymbol{x}_{(n)}\right\}$$

where $\boldsymbol{\Gamma}_n=\mathbb{E}\left[\boldsymbol{X}_{(n)}\boldsymbol{X}_{(n)}^\intercal\right]$ as $\mathbb{E}\left[\boldsymbol{X}_{(n)}\right]=0$ and $\boldsymbol{X}_{(n)}=(X_1,...,X_n)^\intercal$. Remember also that

$$\boldsymbol{X}_{(n)}=\boldsymbol{A}_n^{-1}\boldsymbol{U}_{(n)}$$

where $\boldsymbol{A}_n=\boldsymbol{B}_n+\boldsymbol{I}_n$. Denote with $\boldsymbol{C}_n=\boldsymbol{A}_n^{-1}$ and note that $det\boldsymbol{C}_n=1$. We have that

$$\begin{aligned}\boldsymbol{\Gamma}_n&=\mathbb{E}\left[\boldsymbol{X}_{(n)}\boldsymbol{X}_{(n)}^\intercal\right]\\&=\mathbb{E}\left[\boldsymbol{C}_n\boldsymbol{U}_{(n)}\boldsymbol{U}_{(n)}^\intercal\boldsymbol{C}_n^\intercal\right]\\&=\boldsymbol{C}_n\mathbb{E}\left[\boldsymbol{U}_{(n)}\boldsymbol{U}_{(n)}^\intercal\right]\boldsymbol{C}_n^\intercal\\&=\boldsymbol{C}_n diag(\sigma^2v_1^2,...,\sigma^2v_n^2)\boldsymbol{C}_n^\intercal\\&=\boldsymbol{C}_n MSE(\tilde{Y})\boldsymbol{C}_n^\intercal\\&=\boldsymbol{C}_n\boldsymbol{D}_n\boldsymbol{C}_n^\intercal\end{aligned}$$

then

$$det(\boldsymbol{\Gamma}_n)=det(\boldsymbol{C}_n)det(\boldsymbol{D}_n)det(\boldsymbol{C}_n^\intercal)=(\sigma^2)^n v_1^2...v_n^2$$

We now hava that

$$\boldsymbol{x}_{(n)}^\intercal\boldsymbol{\Gamma}_n^{-1}\boldsymbol{x}_{(n)}=\boldsymbol{v}_{(n)}^\intercal\boldsymbol{C}_n^\intercal(\boldsymbol{C}_n^\intercal)^{-1}\boldsymbol{D}_n^{-1}\boldsymbol{C}_n^{-1}(\boldsymbol{C}_n)\boldsymbol{v}_{(n)}=\frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}$$

Putting all together:

$$\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\frac{1}{\sqrt{v_1^2...v_n^2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}\right\}$$

considering the log:

$$\log\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2}\sum_{j=1}^{n}\log v_j^2--\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}$$

and taking the partial derivative:

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\log\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2) &= -\frac{n}{2}\log\sigma^2+\frac{1}{2}\frac{1}{\sigma^4}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}\\ &= -\frac{1}{2\sigma^2}\left(n-\frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}\right)\end{aligned}$$

This quantity is equal to zero if and only if

$$\frac{1}{n}\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}=\sigma^2$$

denote $S(\boldsymbol{\phi},\boldsymbol{\theta})=\sum_{j=1}^{n}\frac{(x_j-\tilde{x}_j)^2}{v_j^2}$. We now have that

$$MLE(\sigma^2)=\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{n}$$

Now we will plug the estimated parameters in the likelihood function:

$$\begin{aligned}\log\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2) &= -\frac{n}{2}\log 2\pi-\frac{n}{2}\log\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{n}-\frac{1}{2}\sum_{j=1}^{n}\log v_j^2-\frac{n}{2}\\ &= -\frac{n}{2}\left[\log\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{n}+\frac{1}{n}\sum_{j=1}^{n}\log v_j^2+(1+\log 2\pi)\right]\\ &= -\frac{n}{2}\left[l(\boldsymbol{\phi},\boldsymbol{\theta})+(1+\log 2\pi)\right]\end{aligned}$$

where $l(\boldsymbol{\phi},\boldsymbol{\theta})$ is called the **reduced log-likelihood function**. Now we have a method for estimating our parameters, since the $MLE(\boldsymbol{\phi})$ and the $MLE(\boldsymbol{\theta})$ are found by minimizing $l(\boldsymbol{\phi},\boldsymbol{\theta})$ when $(X_t)$ is causal. If $(X_t)$ is invertible then there is an alternative method: finding $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ least square estimates by minimizing $S(\boldsymbol{\phi},\boldsymbol{\theta})$ and then setting the estimate of $\sigma^2$ as

$$\hat{\sigma}^2=\frac{S(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})}{(n-p-q)}$$

Now we will see an other type of model: **ARIMA (autoregressive-integrated moving average)**.

**Definition 15.1.** *A time series is $(X_t)\sim ARIMA(p,d,q)$ if $\big((1-B)^dX_t\big)\sim ARMA(p,q)$ causal.*

**Example 15.3.** *If $(1-\theta B)(1-B)X_t=W_t$ with $t\in\mathbb{Z}$ and $|\theta|<1$ then $(1-\theta B)Y_t=W_t$ with $Y_t=(1-B)X_t$ is an $ARMA$ model and $(X_t)$ is an $ARIMA(1,1,0)$ model.*

**Remark 15.3.** *$(X_t)$ is the solution of $\theta^*(B)X_t=\phi(B)W_t$ with $\theta^*(B)=\theta(Z)(1-Z)^d$ which has a zero of order $d$ at $Z=1$.*

# 16 Lecture 16

**How to chose the order p and q for ARMA(p,q) in fitting: Akaike index. The parsimony principle. Evaluation of the accuracy. Case studies in R: decomposition in an additive model, analysis, fitting, and forecasting of the residual terms. SARIMA models.**

**Example 16.1.**

```
##############################
# fitting and forecasting
##############################

#load the dataset
library(TSA)
data(tempdub)
mean(tempdub)

#plot the data
windows()
plot(tempdub,ylab='temperature',type='o')

# plot ACF and PACF
library(astsa)
windows()
acf2(tempdub,max.lag=80)

# test stationarity
library(tseries)
adf.test(tempdub)
kpss.test(tempdub)

# Guess AR(6): order=c(p,d,q) -> AR(p) & MA(q)
(out1=arima(tempdub,order=c(6,0,0))
# The forecasted values of time series with the errors
rec.pr=predict(out1,n.ahead=25)

# To print the values:
# - the first vector gives the predicted values
# - the second vector gives the standard errors
str(rec.pr)

# To plot the forecasting: n.ahead=lead time
windows()
plot(out1,n.ahead=25,type='b')
```

**Proposition 16.1.** *Consider* $f(\boldsymbol{x}, \boldsymbol{\theta}_{(k)})$ *a pdf such that*

$$\boldsymbol{\theta}_{(k)} = (\theta_1, ..., \theta_k) \in \Theta_k \subseteq \mathbb{R}^k$$

*and $g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})$ an approximation of $f$ such that*

$$G(k) = \left\{g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})/\boldsymbol{\theta}_{(k)} \in \Theta_k\right\}$$

*is a k-dimensional parametric class of the pdf. Consider also*

$$G = \{G(k_1), ..., G(k_l)\}$$

*Which $k \in \{k_1, ..., k_l\}$ gives the best approximation $g$ to $f$? Firstly we need a measure of 'disparity' between $g$ and $f$, and we also need the estimates of the parameters $\boldsymbol{\theta}_{(k)}$, which can be obtained by the ML estimation $\hat{\boldsymbol{\theta}}_{(k)} = MLE(\boldsymbol{\theta}_{(k)})$ with $\boldsymbol{\theta}_{(k)} \in \Theta_k$; in this way we identify a set $g(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_{(k_1)}), ..., g(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_{(k_l)})$. About the measurement of disparity all the information we need is in the* ***likeliyhood ratio****: given $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ the the critical region is defined by the likelyhood ratio*

$$\frac{L(\boldsymbol{x}|\theta_0)}{L(\boldsymbol{x}|\theta_1)}$$

*So, as a measuring tool for our case, we can use*

$$\int_{\mathbb{R}^n} \Phi\left(\frac{g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})}{f(\boldsymbol{x})}\right) f(\boldsymbol{x}) d\boldsymbol{x}$$

*Akaike, in his seminal paper, proposes to use $\Phi(t) = -2\log t$, which implies that*

$$\begin{aligned} W(f,g) &= -2\int_{\mathbb{R}^n} \log \frac{g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})}{f(\boldsymbol{x})} \\ &= 2\int_{\mathbb{R}^n} \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})} \end{aligned}$$

*which is the well-known Kullback-Leibler information (or divergence), which measures the difference between two distributions.*

**Definition 16.1.** *The* ***Kullback-Leibler information (or divergence)*** *is defined as*

$$KL(f,g) = \mathbb{E}_f\left[\ln \frac{f}{g}\right]$$

*where $f, g$ are pdf with the same support.*

**Remark 16.1.** *Some remarks:*

- *$KL(f,g) \geq 0$ and $KL(f,g) = 0 \iff f = g$*
- *$KL(f,g)$ is not a distance: $KL(f,g) \neq KL(g,f)$*

In our case, $f$ is the true model and $g$ the approximating model. Therefore

$$\begin{aligned} KL(f,g) &= \int_{\mathbb{R}^n} \ln \frac{f(\boldsymbol{x})}{g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})} f(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} \ln f(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - \int_{\mathbb{R}^n} \ln g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)} f(\boldsymbol{x}) d\boldsymbol{x} \end{aligned}$$

gives the information lost in replacing $f(\boldsymbol{x})$ with $g(\boldsymbol{x}|\boldsymbol{\theta})$. Taking into account the proposal of Akaike, set $d(\boldsymbol{\theta}_{(k)}) = \mathbb{E}_f(-2\ln g(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}))$ so that we have

$$W(f, g) = 2KL(f, g) = d(\boldsymbol{\theta}_{(k)}) - \mathbb{E}_f[-2\ln f(\boldsymbol{x})]$$

note that this last expectation does not depend on $\boldsymbol{\theta}_{(k)}$, so any ranking of a set of candidate models based on the KL information would be equivalent to a ranking based on $d(\boldsymbol{\theta}_{(k)})$. So in order to select among a set of approximating models $d(\boldsymbol{\theta}_{(k)})$ is used instead of the KL information. This quantity takes the name of **Kullback discrepancy**. Unfortunately, it is not possible to calculate this quantity since $f(\boldsymbol{x})$ is not known; but we can use an estimator of it; in his work Akaike proves that

$$-2\ln g(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_{(k)})$$

is a biased estimator of $d(\boldsymbol{\theta}_{(k)})$ and that

$$\mathbb{E}\left[-2\ln g(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_{(k)})\right] + 2k \xrightarrow{n} \mathbb{E}\left[d\left(\hat{\boldsymbol{\theta}}_{(k)}\right)\right]$$

An empirical rule to choose an approximating model is then to choose a $k$ that minimizes this **Akaike Information Criterion**.

Another fundamental topic for choosing a model is performance measurement. One simple of such metrics is the **forecast error**:

$$e_t = x_t - \tilde{x}_{t|T}$$

where $x_t$ are the original observations and $\tilde{x}_{t|T}$ are the forecasted values obtained by the fitted model on $\{x_1, ..., x_t\}$ (the training set) with $t = T + 1, ..., n$. This function measures the forecast accuracy in different ways by using different indexes, which are

- **Mean Error (ME)** $= mean(e_t)$, also called **forecast bias**
- **Mean Absolute Error (MAE)** $= mean(|e_t|)$
- **Root Mean Squared Error (RMSE)** $= \sqrt{mean(e_t^2)}$
- **Mean Percentage Error (MPE)** $= mean(p_t)$ with $p_t = 100\frac{e_t}{x_t}$
- **Mean Absolute Percentage Error (MAPE)** $= mean(|p_t|)$
- **Mean Absolute Scaled Error (MASE)** $= mean(q_t)$ with
  - for non seasonal time series $q_t = \frac{e_t}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}$
  - for seasonal time series $q_t = \frac{e_t}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}$

  where $y_t$ is the naive forecast at time $t$, $T$ is the trainig set size and $m$ is the period of the seasonal component.
- **ACF1**, which is the correlation function at lag 1 of the forecast error.

All these metrics can also be computed on the training set:

$$e_t = x_t - \tilde{x}_t$$

with $t = 1, ..., T$.

**Example 16.2.**

```
#
# Exercise: by using the output of PREDICT, replicate the previous plot.
#
# Changing the order p? The AIC index

# Guess AR(8): check AIC index
(out2=arima(tempdub,order=c(8,0,0)))

# Guess AR(6): check AIC index
(out1=arima(tempdub,order=c(6,0,0)))

# Similar AIC: parcimony principle, choose p=6
# evaluating the accuracy
library(forecast)
(train.rem = window(tempdub, start=start(tempdub), end=c(1973,12)))
(test.rem = window(tempdub, start=c(1974,1), end=end(tempdub)))
(out1train=arima(train.rem,order=c(6,0,0)))
rec.pr.train=predict(out1train,n.ahead=24)

# plot train data + test data + forecasted data
windows()
plot(test.rem,col="red",main='Forecasted vs test data')
lines(rec.pr.train$pred,type='o',col="blue")
legend("bottomright", legend = c("test set", "forecasted"),
lty = 1, col=c('red','blue'),
title = "Line colors", cex = 1.0)

# accuracy
class(rec.pr.train$pred)
accuracy(rec.pr.train$pred,test.rem)

# different ways to extract the the training data and the test data
(train.rem=head(tempdub,12*10))
(test.rem=tail(tempdub, 12*2))


# Solution of the exercise
# set the upper and lower bounds
U=rec.pr$pred+qnorm(0.975)*rec.pr$se
L=rec.pr$pred-qnorm(0.975)*rec.pr$se

# set the yaxis
minx=min(tempdub,L)
maxx=max(tempdub,U)

# produce the plot
windows()
ts.plot(tempdub,rec.pr$pred,ylim=c(minx,maxx),main='Forecasting')
```

```
# add a red line for the forecasted values
lines(rec.pr$pred,col='red',type='o')

# add blue lines for upper and lower bounds
lines(U,col='blue',lty='dashed')
lines(L,col='blue',lty='dashed')
```

**Example 16.3.**

```
####################################
# the forecast package: jj
####################################

# load the libraries
library(astsa)
library(forecast)
library(tseries)

#load the data
data(jj)

# ndiffs to get a suggestion for the iterated difference operator
(ndiffs(log(jj)))

# consider diff(log(jj))
workjj=diff(log(jj))

# be aware: the starting point is now changed
start(workjj)
start(jj)

# decomposition
windows()
comp=stl(workjj, 'per')
plot(comp)

# looking for a model for the reminder term
(reminder=comp$time.series[,3])

# check stationarity
mean(reminder)kpss.test(reminder)
adf.test(reminder)

# ACF and PACF
windows()
acf2(reminder,max.lag=40)

# a different way to choose the parameters using
# the package "forecast"
(fit=auto.arima(reminder))
```

```
# check causality and invertibility
windows()
plot(fit)

# plot Arima model versus original time series
windows()
plot(reminder,col="red")
lines(fitted(fit),type='o',col="blue")
legend("bottomright", legend = c("dataset", "fitting"),
lty = 1, col=c('red','blue'),
title = "Line colors", cex = 1.0)

# check residuals of the fitted ARMA
windows()
checkresiduals(fit,plot=TRUE)
jarque.bera.test(fit$residuals)

# forecast: lead time = 20 steps
ffit=forecast(fit,h=20)
windows()
plot(ffit)

# print the predicted values with bounds
ffit$mean
head(ffit$lower,10)
head(ffit$upper,10)

# evaluate the accuracy of fitting/forecasting
# split the reminder term in training set and test set
(train.rem = head(reminder, 67))
(test.rem = tail(reminder, 4*4))
(fit.train=auto.arima(train.rem))
ffit.test=forecast(fit.train,h=16)

# plot forecasted vs test data
windows()
plot(test.rem,col="red",main='Forecasted vs test data')
lines(ffit.test$mean,type='o',col="blue")
legend("topleft", legend = c("test set", "forecast"),
lty = 1, col=c('red','blue'),
title = "Line colors", cex = 1.0)

# accuracy ARMA(3,0)
class(ffit.test)
accuracy(ffit.test,test.rem)

# accuracy ARMA(3,0,1)
fit.train1=arima(train.rem,order=c(3,0,1))
ffit.test1=predict(fit.train1,n.ahead=16)accuracy(ffit.test1$pred,test.rem)
```

```
# plot forecasted vs test data
windows()
plot(test.rem,col="red",main='Forecasted vs test data')
lines(ffit.test$mean,type='o',col="blue")
lines(ffit.test1$pred,type='o',col="green")
legend("topleft", legend = c("test set", "forecast 3,0", "forecast 3,1"),
lty = 1, col=c('red','blue','green'),
title = "Line colors", cex = 1.0)

# what happens with auto.arima(workjj)?
(fit.workjj=auto.arima(workjj))
```

Up to now, we have worked with additive models, but sometimes a deterministic seasonal component is not suitable (es. economics). In such a case, it is helpful to introduce **autoregressive moving average seasonal operators**.

**Definition 16.2.** *$(X_t) \sim ARMA(P, Q)_s$ if it is the stationary solution of*

$$\Theta_P(B^s)X_t = \Phi_Q(B^s)W_t$$

*where $W_t \sim WN(0, \sigma^2)$ and*

$$\Theta_p(B^s) = 1 - \tilde{\theta}_1 B^s + ... - \tilde{\theta}_P B^{Ps}$$

*is the seasonal autoregressive operator and*

$$\Phi_Q(B^s) = 1 + \tilde{\phi}_1 B^s + ... + \tilde{\phi}_Q B^{Qs}$$

Now we can combine the seasonal operators $(\Theta_P(B^S), \Phi_Q(B^S))$ and the non-seasonal operators $(\theta(B), \phi(B))$ in order to obtain a **multiplicative seasonal ARMA model**, which is the solution to

$$\theta(B)\Theta_P(B^S)X_T = \phi(B)\Phi_Q(B^S)W_t$$

and we can also add non-stationarity by considering that

$$X_t = (1 - B)^d \tilde{X}_t$$

but we can go one step further considering that the non-stationarity can be in the seasonal component.

**Example 16.4.** *Consider*

$$X_t = S_t + W_t$$

*with $S_t$ is a random walk such that $S_t = S_{t-12} + W'_t$ with $(W_t), (W'_t) \sim WN$ uncorrelated. Observe that*

$$\begin{aligned}\nabla_{12} X_t &= (1 - B^{12})X_t \\ &= X_t - X_{t-12} \\ &= S_t + W_t - S_{t-12} - W_{t-12} \\ &= W'_t + W_t - W_{t-12}\end{aligned}$$

*so this is a stationary time series. So it is possible to recover a stationary time series by applying the difference operator taking into account the period of the non stationary component.*

**Definition 16.3.** *If $d$ and $D$ are non-negative integes then $(X_t) \sim$ **ARIMA**$(p, d, q) \times (P, D, Q)_s$ **(seasonal ARIMA)** with period $s$ if*

$$Y_t = (1 - B)^d (1 - B^s)^D X_t$$

*is a multiplicative seasonal ARMA model.*

**Example 16.5.** *Suppose $(X_t) \sim SARIMA(1, 1, 0) \times (1, 1, 2)_{12}$. $(X_t)$ is such that $Y_t = (1 - B)(1 - B^{12})$ is a solution of*

$$\theta(B)\Theta_1(B^{12})Y_t = \phi(B)\Phi_2(B^12)W_t$$

*where*

- $\theta(B) = 1 - \theta_1 B, \ \phi(B) = 1$
- $\Theta_1(B^{12}) = 1 - \tilde{\theta}_1 B^{12}$
- $\Phi_2(B^{12}) = 1 + \tilde{\phi}_1 B^{12} + \tilde{\phi}_2 B^{24}$

**Example 16.6.** *Consider $(X_t) \sim SARIMA(0, 0, 1) \times (1, 1, 0)_4$ so $(X_t)$ is such that $Y_t = (1 - B)^0 (1 - B^4)^1 X_t$ is a solution of*

$$\theta(B)\Theta_1(B^4)Y_t = \phi(B)\Phi_0(B)^4 W_t$$

*where*

- $\theta(B) = 1$
- $\Theta_1(B^4) = 1 - sar1B^4$
- $\phi(B) = 1 + ma1B$
- $\Phi_0(B^4) = 1$

*according to the results of the R functions.*

**Remark 16.2.** *The parsimony principle should always guide the choice of the parameters. An empirical rule suggest that $p + d + q + P + D + Q \leq 6$.*