# NDA: Time Series Analysis

Lionel Tabourier

LIP6 – CNRS and Sobonne University

first_name.last_name@lip6.fr

January 2020

# Bibliography

Formal content:

- Peter Brockwell and Richard Davis
  *Introduction to Time Series and Forecasting*
- William Thistleton and Tural Sadigov
  MOOC Coursera: *Practical Time Series Analysis*

Informal guide in python:

- `www.machinelearningplus.com/time-series/`

Illustrative datasets:

- `data.world/datasets/time-series`
- `www.kaggle.com/tags/time-series`

# Outline

# Outline

# What is time series analysis

**Definition**

Set of observations $\{x_t\}$, recorded at time $t \in T_0$

Think of each $x_t$ as a realization from a distribution

**Specificities of the problem**

A unique realization of the process
$\Rightarrow$ necessary to make assumptions

- observe time series, identify particularities
- choose a family of models $X_t$ to represent data
- check the goodness of the model

# What is time series analysis

## Definition

Set of observations $\{x_t\}$, recorded at time $t \in T_0$

Think of each $x_t$ as a realization from a distribution

## Specificities of the problem

A unique realization of the process

$\Rightarrow$ necessary to make assumptions

- observe time series, identify particularities
- choose a family of models $X_t$ to represent data
- check the goodness of the model

# Assumptions for this course

## Restrictions to a subfamily of problems

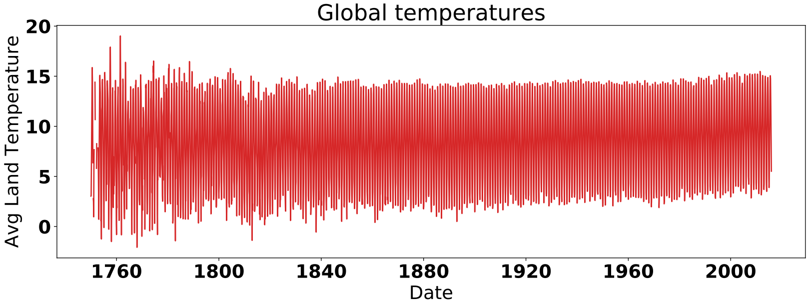
- discrete time series (discrete time set)
- fixed time steps (time resolution)
- univariate (one single variable over time)
  $\rightarrow$ processes have values in $\mathbb{R}$

## And only a few approaches

- e.g. no Fourier analysis

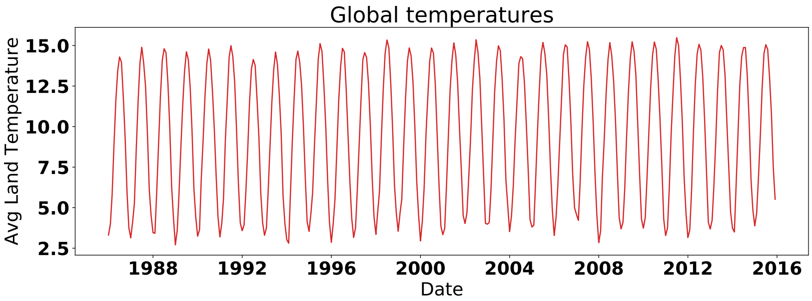
# A few examples

Average global land temperature (per month)



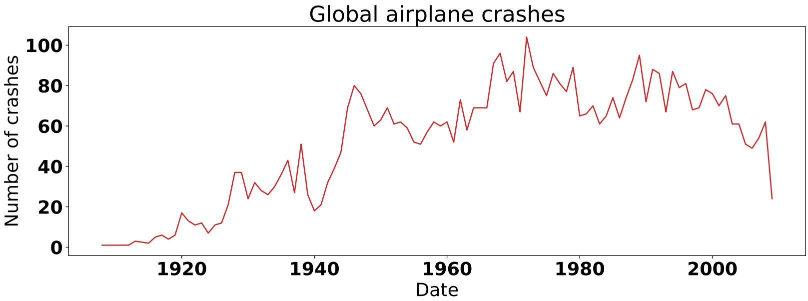
# A few examples

Average global land temperature (per month)



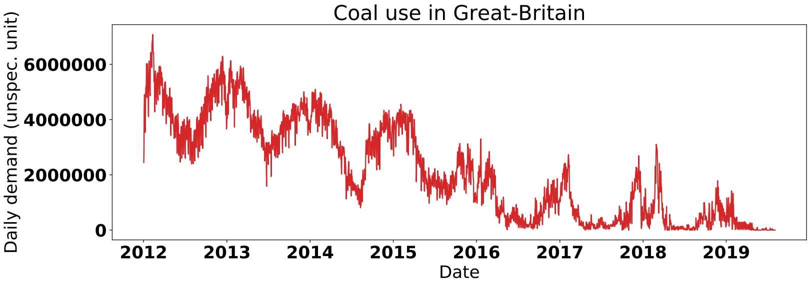
# A few examples

Number of airplane crashes (per year)

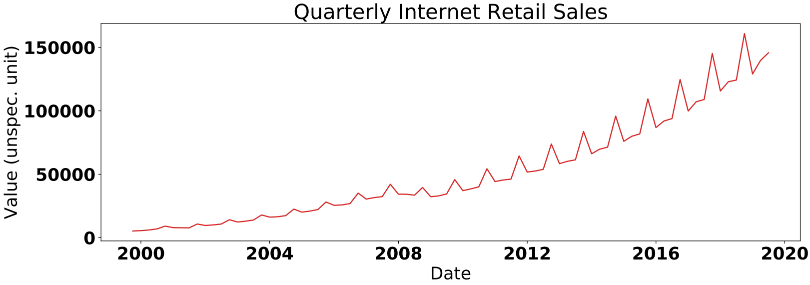


# A few examples

Daily demand of power obtained with coal in GB (per year)



# A few examples



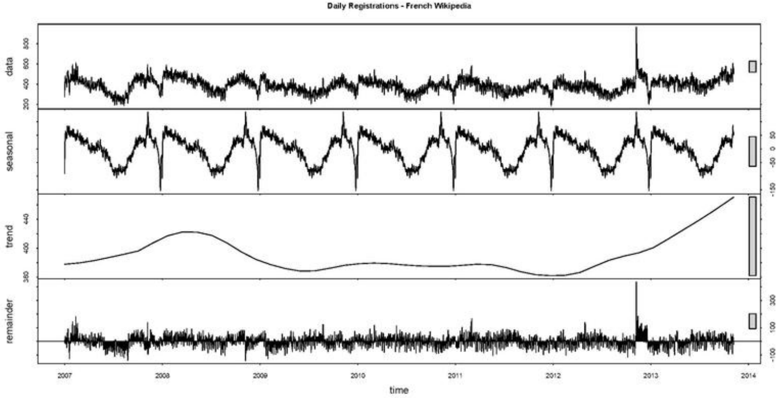
# Goals of time series analysis

- Have a simplified description of the data
  $\rightarrow$ improve our understanding (*ex: climate data*)
- Test an assumption
  *ex: is there a significant measurable global warming?*
- Filter: separate signal from noise
  *ex: known physical signal broadcast* $\rightarrow$ *filter noise*
- Predict future values
  *ex: predict the future demand for a product*
- Simulate a process in a complex model
  *ex: expectation for the GDP to predict economic activity*

# How to analyze a time series? (1)

Analyse from Greek *análusis* $\sim$ unravel $\Rightarrow$ decompose

Decompose the time series into parts, for example:

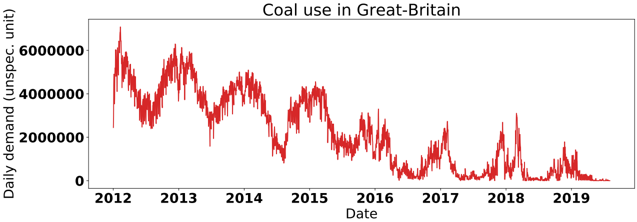


# How to analyze a time series? (1)

## First step

Plot the time series to:

- identify the existence of a trend *(tendance)*
- uncover seasonal variations *(variations saisonnières)*
- detect changes of behavior
- spot outliers *(valeurs aberrantes)*

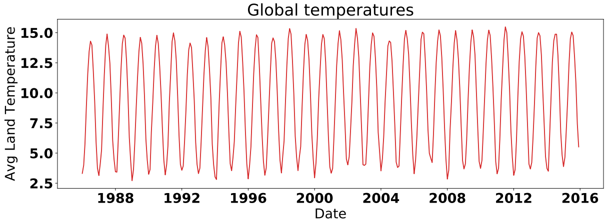


# How to analyze a time series? (1)

## First step

Plot the time series to:

- identify the existence of a trend *(tendance)*
- uncover seasonal variations *(variations saisonnières)*
- detect changes of behavior
- spot outliers *(valeurs aberrantes)*

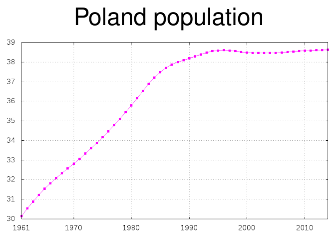


# How to analyze a time series? (1)

## First step

Plot the time series to:

- identify the existence of a trend *(tendance)*
- uncover seasonal variations *(variations saisonnières)*
- detect changes of behavior
- spot outliers *(valeurs aberrantes)*

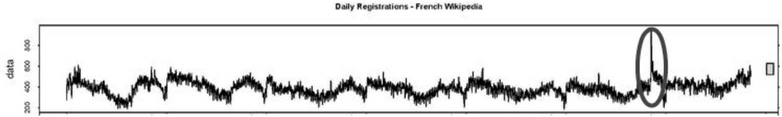


# How to analyze a time series? (1)

## First step

Plot the time series to:

- identify the existence of a trend *(tendance)*
- uncover seasonal variations *(variations saisonnières)*
- detect changes of behavior
- spot outliers *(valeurs aberrantes)*



# How to analyze a time series? (1)

**First step**

Plot the time series to:

- identify the existence of a trend *(tendance)*
- uncover seasonal variations *(variations saisonnières)*
- detect changes of behavior
- spot outliers *(valeurs aberrantes)*

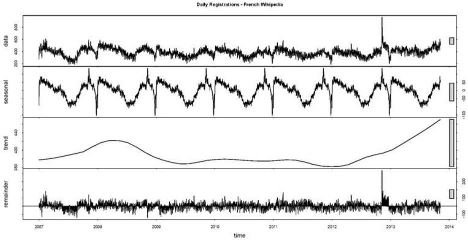
$\rightarrow$ subjective components in this analysis

# The classical decomposition

## Classical decomposition of the time series

$$X_t = s_t + m_t + r_t$$

- seasonality $s_t$
- trend $m_t$
- remainder $r_t$



# Outline

# Mean and covariance of a time series

## Two fundamental definitions

Let $\{X_t\}$ a time series with $\mathbb{E}[X_t^2] < \infty$ (finite variance)

*rk: here we consider $X_t$ as a model*

- **mean function** of $X_t$, defined for all $t$:

$$\mu_X(t) = \mathbb{E}[X_t]$$

- **covariance function** of $X_t$, defined for all $r, s$:

$$\gamma_X(r, s) = Cov(X_r, X_s) = \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

# Stationarity

**Informal definition**

A process is **stationary** *(stationnaire)* if its statistical properties are similar when shifted in time

Remarks:

- stationarity is a property of a model (not of data)
- stationary processes are simpler to investigate
  $\Rightarrow$ usual to transform a TS to obtain a stationary process

# Stationarity

### Informal definition

A process is **stationary** *(stationnaire)* if its statistical properties are similar when shifted in time

### Formal definition

A process is said to be **weakly stationary** if

- the mean function $\mu_X(t)$ is independent of $t$
- $\gamma_X(t+h, t)$ is independent of $t$ for any $h$ (including $h = 0$)

  $h$ is called the *lag* *(décalage)*

# Stationarity

## Informal definition

A process is **stationary** *(stationnaire)* if its statistical properties are similar when shifted in time

## Formal definition

A process is said to be **strictly (or strongly) stationary** if

- $\forall n$ and $\forall h$

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_{1+h} = x_1, \dots, X_{n+h} = x_n)$$

*Unless specified otherwise, we talk about weak stationarity in the following*

# Stationarity

## Informal definition

A process is **stationary** *(stationnaire)* if its statistical properties are similar when shifted in time

## Formal definition

A process is said to be **strictly (or strongly) stationary** if

- $\forall n$ and $\forall h$

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_{1+h} = x_1, \ldots, X_{n+h} = x_n)$$

*Unless specified otherwise, we talk about weak stationarity in the following*

# Autocorrelation function

Notice that for a stationary time series: $\gamma_X(t+h,t) = \gamma_X(h)$
$\Rightarrow$ the covariance function $\gamma_X$ has one variable (the lag)

## Definition

For a stationary time series:

- the **autocovariance function** at lag $h$ is:

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

- the **autocorrelation function** (ACF) at lag $h$ is:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

# Equivalent on real data

Concepts well defined on models, but what about real data?
Let $\{x_1, \ldots, x_n\}$ be a series of observations

## Sample mean

- the **sample mean** estimator is

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

# Equivalent on real data

Concepts well defined on models, but what about real data?
Let $\{x_1, \ldots, x_n\}$ be a series of observations

## Sample autocovariance function

- the **sample autocovariance function** estimator is

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

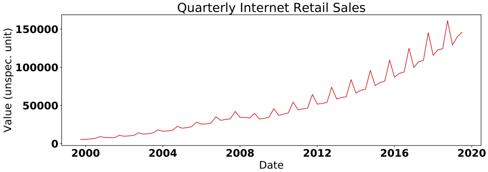
*remark: notice the denominator (because of mathematical properties)*

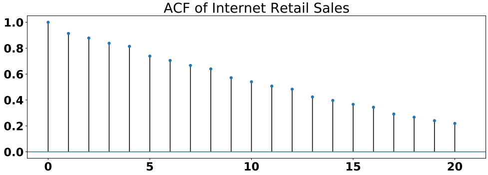
- the **sample autocorrelation function** estimator is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \ , \ -n < h < n. \text{ Note that } \hat{\rho}(h) \in [-1; 1]$$

# Equivalent on real data

Data with strong trend:

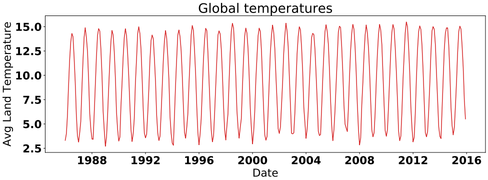


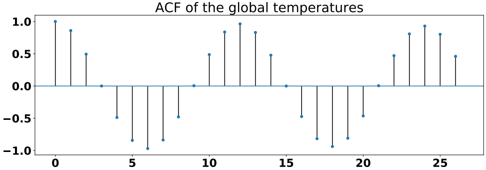




slow decay of correlations with $h$

# Equivalent on real data

Data with strong seasonality:




periodicity on the ACF (here monthly measures $\Rightarrow$ period = 12)

# Outline

# What is a time series model?

### Definition

**Time series model**: specification of the joint distributions of a sequence of random variables $X_t$ of which the observed data is supposed to be the realization

Remarks:

- suppose to know $\forall n$ the distribs $P(X_1 = x_1, \ldots, X_n = x_n)$
  $\Rightarrow$ in most case too many parameters...
- in practice, we focus on first and second order moments:
  - expected values $\mathbb{E}[X_t]$
  - and expected products $\mathbb{E}[X_{t+h}X_t]$, $h = 1, 2, \ldots$

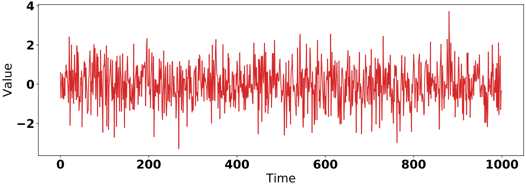
# Independent Identically Distributed noise model

## IID noise

- independant:
  $P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$
- identically distributed: $P(X_t = x) = P(X_{t'} = x)$

IID noise is obviously stationary

*ex: repeated coin flipping with heads=1, tails=-1 should be IID noise*



# Independent Identically Distributed noise model

## IID noise

- independant:
  $P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$
- identically distributed: $P(X_t = x) = P(X_{t'} = x)$

IID noise is obviously stationary

*ex: repeated coin flipping with heads=1, tails=-1 should be IID noise*

## White noise *(bruit blanc)*

Special case IID noise with

- 0 mean: $E[X_t] = 0$
- autocovariance function:
  $\gamma_X(h) = \sigma^2$ if $h = 0$ and $\gamma_X(h) = 0$ if $h \neq 0$

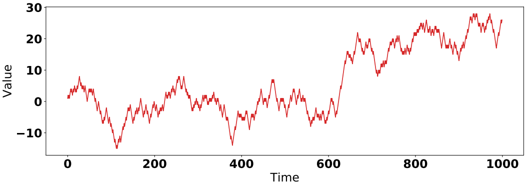
# Random Walk model

**How to build a random walk?** *(marche aléatoire)*

Suppose $\{X_t\}$ is IID noise, then $\{S_t\}$ defined as:

$$S_t = X_1 + \ldots + X_t$$

is a random walk



# Random Walk model

**How to build a random walk?** *(marche aléatoire)*

Suppose $\{X_t\}$ is IID noise, then $\{S_t\}$ defined as:

$$S_t = X_1 + \ldots + X_t$$

is a random walk

Remarks:

- is a random walk stationary?
- it's a summation of an IID process
- and conversely $X_t = S_t - S_{t-1}$

# Random Walk model

**How to build a random walk?** *(marche aléatoire)*

Suppose $\{X_t\}$ is IID noise, then $\{S_t\}$ defined as:

$$S_t = X_1 + \ldots + X_t$$

is a random walk

Remarks:

- is a random walk stationary? No
- it's a summation of an IID process
- and conversely $X_t = S_t - S_{t-1}$

# Outline

# How to analyze a time series? (2)

**Second step**:

- (if necessary) transform data
- remove the trend and seasonal components to get stationary residuals *(résidus)*

Residual time series obtained (remainder) should be stationary, but not necessarily IID noise…

# Back to the classical decomposition

## Classical decomposition of the time series

$$X_t = s_t + m_t + r_t$$

- seasonality $s_t$
- trend $m_t$
- remainder $r_t$

What is the difference between seasonality and trend?

$$s_{t+d} = s_t \text{ and } \sum_{j=1}^{d} s_j = 0$$

# Back to the classical decomposition

## Classical decomposition of the time series

$$X_t = s_t + m_t + r_t$$

- seasonality $s_t$
- trend $m_t$
- remainder $r_t$

What is the difference between seasonality and trend?

$$s_{t+d} = s_t \text{ and } \sum_{j=1}^{d} s_j = 0$$

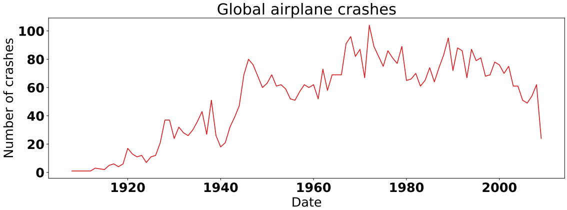
# Isolate the trend component

## Method 1: model and regression

→ cf. course *Regression*

E.g.: $2^{nd}$ order polynomial model with least squares regression

Minimize $\sum_{t=1}^{n}(x_t - m_t)^2$, with $m_t = a_0 + a_1 t + a_2 t^2$



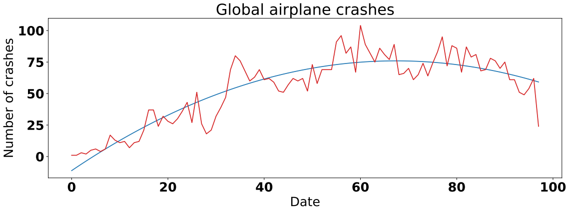
# Isolate the trend component

## Method 1: model and regression

$\rightarrow$ cf. course *Regression*

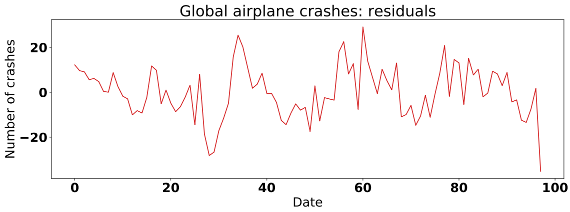
E.g.: $2^{nd}$ order polynomial model with least squares regression

$$\text{Minimize } \sum_{t=1}^{n}(x_t - m_t)^2, \text{ with } m_t = a_0 + a_1 t + a_2 t^2$$



# Isolate the trend component

Then we plot the residuals $\{x_t - m_t\}$




Questions to ask oneself:

- Does it look stationary? Perceptible trend?
- Does it look like noise? Is it smooth? Do we see stretch of values of the same sign?

# Moving Average models: MA(1)

**What is a moving average?** *(moyenne mobile)*

To smooth *(rendre lisse)* a signal $x_t$, one possibility:

$$x'_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \; , \; q < t < n-q$$

tool for signal processing *(low pass filter, filtre passe-bas)*

# Moving Average models: MA(1)

**What is a moving average?** *(moyenne mobile)*

To smooth *(rendre lisse)* a signal $x_t$, one possibility:

$$x'_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \,,\; q < t < n - q$$

tool for signal processing *(low pass filter, filtre passe-bas)*

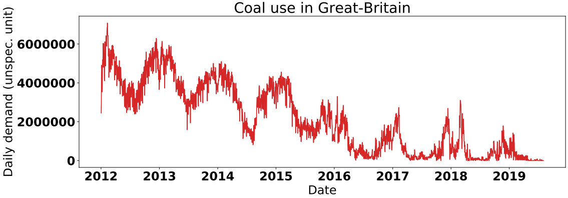
# Moving Average models: MA(1)

## What is a moving average? *(moyenne mobile)*

To smooth *(rendre lisse)* a signal $x_t$, one possibility:

$$x'_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \ , \ q < t < n - q$$

tool for signal processing *(low pass filter, filtre passe-bas)*



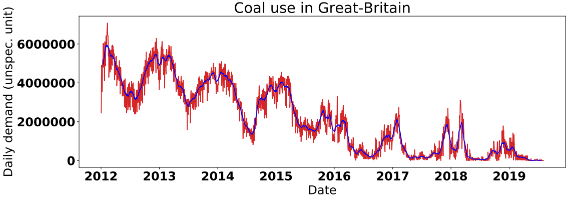
# Moving Average models: MA(1)

## What is a moving average? *(moyenne mobile)*

To smooth *(rendre lisse)* a signal $x_t$, one possibility:

$$x'_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \;,\; q < t < n - q$$

tool for signal processing *(low pass filter, filtre passe-bas)*



# Isolate the trend component (2)

can also be seen as a method to isolate the trend

**Method 2: moving average**

consider that $m_t$ can be computed as a MA:

$$m_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \ , \ q < t < n - q$$

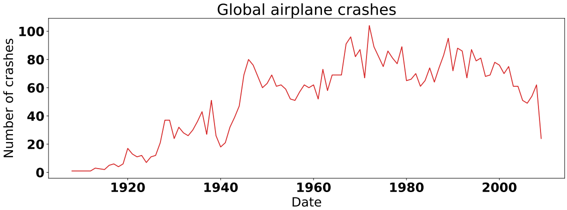
# Isolate the trend component (2)

can also be seen as a method to isolate the trend

## Method 2: moving average

consider that $m_t$ can be computed as a MA:

$$m_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \ , \ q < t < n - q$$



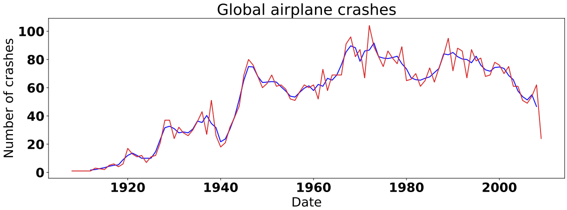
# Isolate the trend component (2)

can also be seen as a method to isolate the trend

## Method 2: moving average

consider that $m_t$ can be computed as a MA:

$$m_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \; , \; q < t < n - q$$



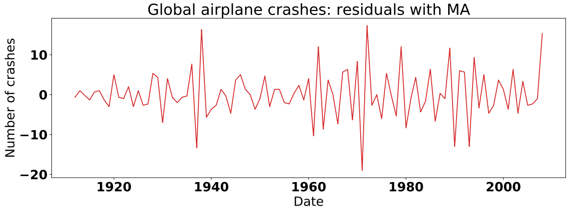
# Isolate the trend component (2)

can also be seen as a method to isolate the trend

## Method 2: moving average

consider that $m_t$ can be computed as a MA:

$$m_t = \frac{1}{2q+1} \sum_{h=-q}^{h=+q} x_{t-h} \ , \ q < t < n - q$$



# Isolate seasonal component

## Regression

Which model? Harmonic regression

$$s_t = a_0 + \sum_{j=1}^{k} a_j cos\left(\frac{2\pi t}{T_j}\right) + b_j sin\left(\frac{2\pi t}{T_j}\right)$$

where $T_j$ are the expected periods of the process

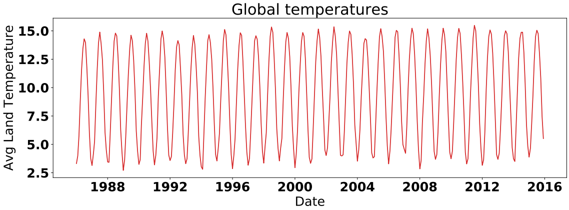
# Isolate seasonal component

## Regression

Which model? Harmonic regression

$$s_t = a_0 + \sum_{j=1}^{k} a_j cos\left(\frac{2\pi t}{T_j}\right) + b_j sin\left(\frac{2\pi t}{T_j}\right)$$

where $T_j$ are the expected periods of the process



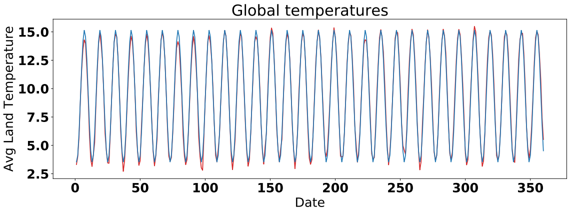
# Isolate seasonal component

## Regression

Which model? Harmonic regression

$$s_t = a_0 + \sum_{j=1}^{k} a_j cos\left(\frac{2\pi t}{T_j}\right) + b_j sin\left(\frac{2\pi t}{T_j}\right)$$

where $T_j$ are the expected periods of the process



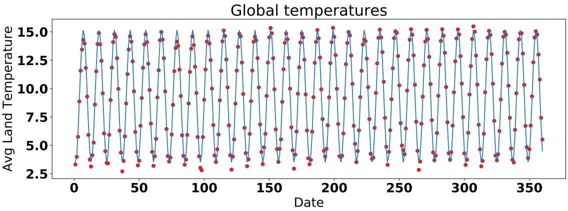
# Isolate seasonal component

## Regression

Which model? Harmonic regression

$$s_t = a_0 + \sum_{j=1}^{k} a_j cos\left(\frac{2\pi t}{T_j}\right) + b_j sin\left(\frac{2\pi t}{T_j}\right)$$

where $T_j$ are the expected periods of the process



# About pre-processing

**Second step**:

- (if necessary) transform data
- remove the trend and seasonal components to get stationary residuals

When is it necessary to transform data ?

**Some cases**

- if outliers $\rightarrow$ discard them if justified
  *ex: external stimulus, mistake in data acquisition, ...*
- if obvious different regimes
  $\rightarrow$ break data into homogeneous segments
- if noise or seasonality component increases with level
  $\rightarrow$ logarithmic transformation of the data

# About pre-processing

**Second step**:

- (if necessary) transform data
- remove the trend and seasonal components to get stationary residuals

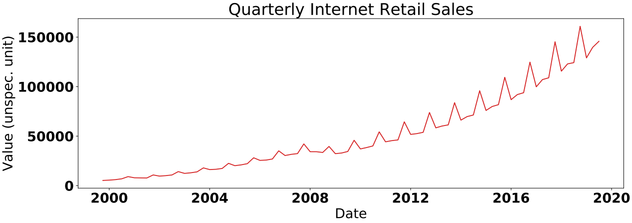
When is it necessary to transform data ?

## Some cases

- if outliers → discard them if justified
  *ex: external stimulus, mistake in data acquisition, ...*
- if obvious different regimes
  → break data into homogeneous segments
- if noise or seasonality component increases with level
  → logarithmic transformation of the data

# Logarithmic transformation

If fluctuations (seasonality, noise) grow with magnitude…



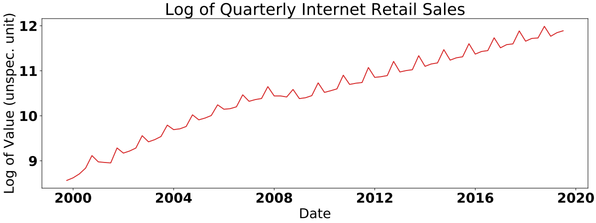

…after logarithmic transform

→ c.f. course *Regression (heteroscedasticity)*

Conduct similar analysis on the transformed time series and reverse the transformations in the end to make predictions etc.

# Logarithmic transformation

If fluctuations (seasonality, noise) grow with magnitude…



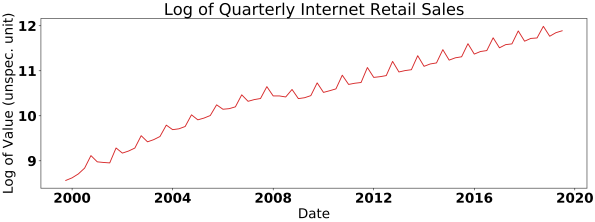

…after logarithmic transform

→ c.f. course *Regression (heteroscedasticity)*

Conduct similar analysis on the transformed time series and reverse the transformations in the end to make predictions etc.

# Logarithmic transformation

If fluctuations (seasonality, noise) grow with magnitude…




…after logarithmic transform

→ c.f. course *Regression (heteroscedasticity)*

Conduct similar analysis on the transformed time series and reverse the transformations in the end to make predictions etc.

# Outline

# How to analyze a time series? (3)

**Third step**:

- fit the residuals

For this purpose, we introduce new families of models

# Auto-Regressive models: AR(1)

### What is autoregression?

*auto* means self $\Rightarrow$ regression from itself

### The most basic AR model: $1^{st}$ order regression or AR(1)

$\{X_t\}$ is a series satisfying:

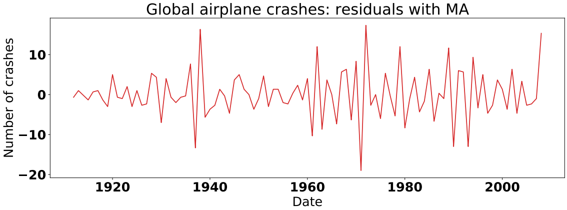
$$X_t = \phi X_{t-1} + W_t$$

where $W_t$ is a white noise (0 means, $\sigma^2$ variance)

**if stationary**, we can check that $\mathbb{E}[X_t] = 0$ and $\gamma_X(h) = \phi^{|h|}\gamma_X(0) = \phi^{|h|}\frac{\sigma^2}{1-\phi^2}$

Note: random walk is AR(1) with $\phi = 1$, in general assume $|\phi| < 1$ w. AR(1)

# Auto-Regressive models: AR(1)



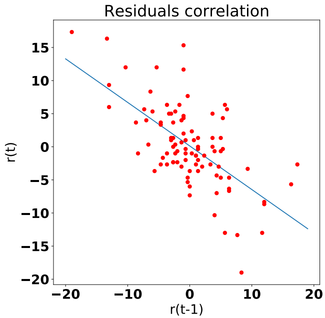

Suppose AR(1) model for residuals $r_t$, how to compute $\phi$?

- plot $r_t$ as a function of $r_{t-1}$ *(lag-1 plot)*
- linear fit, slope is $\phi$

# Auto-Regressive models: AR(1)

Suppose AR(1) model for residuals $r_t$, how to compute $\phi$?

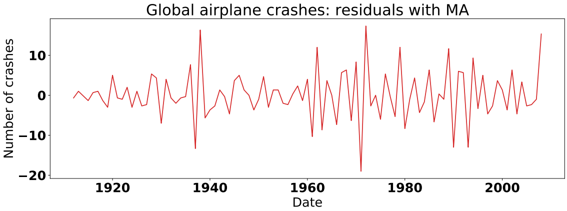
- plot $r_t$ as a function of $r_{t-1}$ *(lag-1 plot)*
- linear fit, slope is $\phi$



# Auto-Regressive models: AR(1)

Now, compute we compute the “residuals of residuals”:

$$r_t - \phi r_{t-1}$$

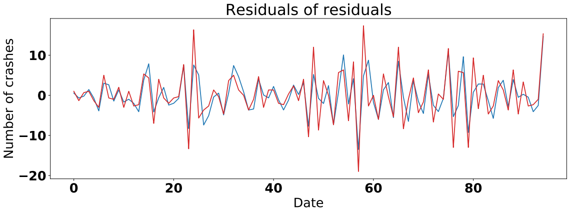



Is it much better?

# Auto-Regressive models: AR(1)

Now, compute we compute the “residuals of residuals”:

$$r_t - \phi r_{t-1}$$




Is it much better?

# Testing if the residuals time series is IID: method 1

**General idea:** suppose IID random variables
What should we observe? Is it the case?

# Testing if the residuals time series is IID: method 1

**General idea:** suppose IID random variables
What should we observe? Is it the case?

## Sample ACF criterion

**Theorem** (admitted):

- suppose $x_t$ IID with mean 0 and variance 1 (white noise)
- if $n$ large enough, $\hat{\rho}_x(h)$ is approx. distributed as $\mathcal{N}(0, \frac{1}{\sqrt{n}})$

In practice, consider the 95% confidence interval:
we measure how many values fall out of $\left[\frac{-1.96}{\sqrt{n}}, \frac{+1.96}{\sqrt{n}}\right]$

$\rightarrow$ c.f. course *Hypothesis testing*

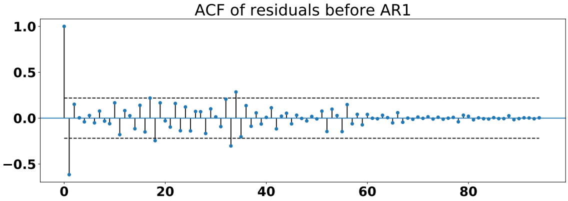
# Testing if the residuals time series is IID: method 1

**General idea:** suppose IID random variables
What should we observe? Is it the case?

## Sample ACF criterion

**Theorem** (admitted):

- suppose $x_t$ IID with mean 0 and variance 1 (white noise)
- if $n$ large enough, $\hat{\rho}_x(h)$ is approx. distributed as $\mathcal{N}(0, \frac{1}{\sqrt{n}})$



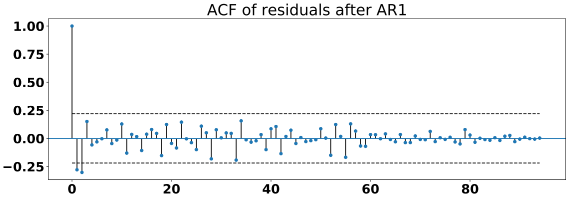
# Testing if the residuals time series is IID: method 1

**General idea:** suppose IID random variables
What should we observe? Is it the case?

## Sample ACF criterion

**Theorem** (admitted):

- suppose $x_t$ IID with mean 0 and variance 1 (white noise)
- if $n$ large enough, $\hat{\rho}_X(h)$ is approx. distributed as $\mathcal{N}(0, \frac{1}{\sqrt{n}})$



# Testing if the residuals time series is IID: method 2

**General idea:** suppose IID random variables
What should we observe? Is it the case?

## Turning point test

Turning point $x_i$ (only defined for $1 < i < n$):
if $x_i \geq x_{i-1}$ and $x_i \geq x_{i+1}$ or $x_i$ if $x_i \leq x_{i-1}$ and $x_i \leq x_{i+1}$

- Probability that a point is a turning point if IID? $\frac{2}{3}$
- $\Rightarrow \mu_T = \mathbb{E}[T_n] = \frac{2(n-2)}{3}$, with $T_n$ number of turning points
- $Var(T_n) = \mathbb{E}[T_n^2] - \mathbb{E}[T_n]^2 \Rightarrow \sigma_T^2 = Var(T_n) = \frac{16n-29}{90}$

If $x_i$ is IID, $T_n$ is approximately $\mathcal{N}(\mu_T, \sigma_T^2)$

- test if $1.96 > \frac{T_n - \mu_T}{\sigma_T} > -1.96$ for the 95% CI

# Testing if the residuals time series is IID: method 2

**General idea:** suppose IID random variables
What should we observe? Is it the case?

### Turning point test

Turning point $x_i$ (only defined for $1 < i < n$):
if $x_i \geq x_{i-1}$ and $x_i \geq x_{i+1}$ or $x_i$ if $x_i \leq x_{i-1}$ and $x_i \leq x_{i+1}$

- Probability that a point is a turning point if IID? $\frac{2}{3}$
- $\Rightarrow \mu_T = \mathbb{E}[T_n] = \frac{2(n-2)}{3}$, with $T_n$ number of turning points
- $Var(T_n) = \mathbb{E}[T_n^2] - \mathbb{E}[T_n]^2 \Rightarrow \sigma_T^2 = Var(T_n) = \frac{16n-29}{90}$

If $x_i$ is IID, $T_n$ is approximately $\mathcal{N}(\mu_T, \sigma_T^2)$

- test if $1.96 > \frac{T_n-\mu_T}{\sigma_T} > -1.96$ for the 95% CI

# Testing if the residuals time series is IID: method 2

**General idea:** suppose IID random variables
What should we observe? Is it the case?

## Turning point test

Turning point $x_i$ (only defined for $1 < i < n$):
if $x_i \geq x_{i-1}$ and $x_i \geq x_{i+1}$ or $x_i$ if $x_i \leq x_{i-1}$ and $x_i \leq x_{i+1}$

- Probability that a point is a turning point if IID? $\frac{2}{3}$
- $\Rightarrow \mu_T = \mathbb{E}[T_n] = \frac{2(n-2)}{3}$, with $T_n$ number of turning points
- $Var(T_n) = \mathbb{E}[T_n^2] - \mathbb{E}[T_n]^2 \Rightarrow \sigma_T^2 = Var(T_n) = \frac{16n-29}{90}$

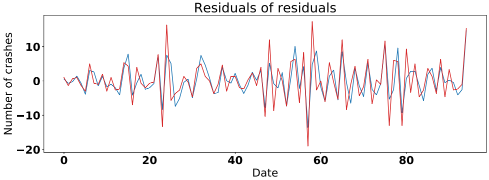
If $x_i$ is IID, $T_n$ is approximately $\mathcal{N}(\mu_T, \sigma_T^2)$

- test if $1.96 > \frac{T_n - \mu_T}{\sigma_T} > -1.96$ for the 95% CI

# Testing if the residuals time series is IID: method 2

**General idea:** suppose IID random variables
What should we observe? Is it the case?

In our example (airplane crashes):




- $T_n^{res} = 64$ on the residuals
- $T_n^{resAR1} = 59$ on the residuals of the residuals (after AR1)

$\Rightarrow$ both pass this test

# MA models

Now let's think of the Moving Average process as a model

**MA(1) model**

$W_t$ is white noise
signal = weighted average of noise at $t$ and of noise at $t-1$

- $X_t = \beta_0 W_t + \beta_1 W_{t-1}$
- $X_t = \beta_0 W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$

# MA models

Now let's think of the Moving Average process as a model

**MA(q) model**

$W_t$ is white noise
signal = weighted average of noise at $t$ and $q$ previous steps

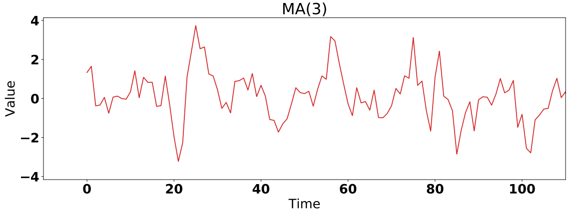
- $X_t = \beta_0 W_t + \beta_1 W_{t-1}$
- $X_t = \beta_0 W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$

# MA models

$$X_t = \beta_0 W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$$

## Some characteristics

- Stationary process
  mean and autocovariance at lag $h$ do not depend on time
  *Ex: prove if $h \leq q$, $Cov(X_t, X_{t+h}) = \sigma^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h}$*
- ACF cuts off at lag $q$

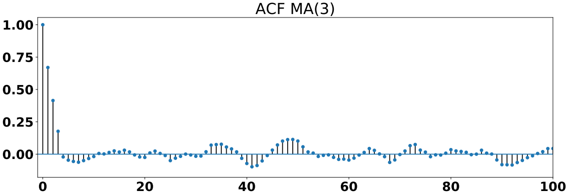


# MA models

$$X_t = \beta_0 W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$$

## Some characteristics

- Stationary process
  mean and autocovariance at lag $h$ do not depend on time
  *Ex: prove if* $h \leq q$, $Cov(X_t, X_{t+h}) = \sigma^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h}$
- ACF cuts off at lag $q$



# Generalized AR model

## AR(1) model

$W_t$ is white noise
signal = noise and (weighted) influence of the signal at $t-1$

- $X_t = \phi X_{t-1} + W_t$
- $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + W_t$

# Generalized AR model

**AR(p) model**

$W_t$ is white noise
signal = noise and influence of the signal at $p$ previous steps

- $X_t = \phi X_{t-1} + W_t$
- $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + W_t$

# Generalized AR model

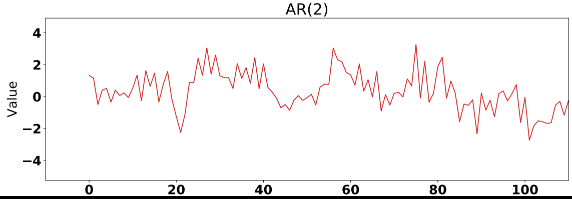
$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = W_t$$

## Some characteristics

- Stationary process? $\rightarrow P(x) = 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p$

  if all its roots are out of the unit circle, then AR(p) stationary

  *example: AR(1),* $P(x) = 1 - \phi x \Rightarrow$ *root is* $\frac{1}{\phi} \Rightarrow |\phi| < 1$
- smoother decay, no cut-off (as with MA models)



# Generalized AR model

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = W_t$$

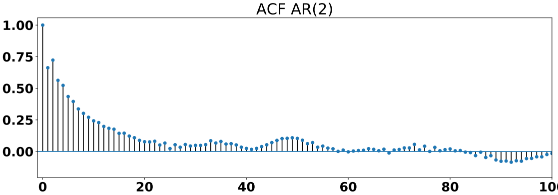
## Some characteristics

- Stationary process? $\rightarrow P(x) = 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p$

  if all its roots are out of the unit circle, then AR(p) stationary

  *example: AR(1),* $P(x) = 1 - \phi x \Rightarrow$ *root is* $\frac{1}{\phi} \Rightarrow |\phi| < 1$

- smoother decay, no cut-off (as with MA models)



# How to find AR(p) ACF coefficients?

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + W_t$$

## Yule-Walker equations

We have seen that for AR(1):

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

*Note that it would diverge with h if* $|\phi| > 1$*, AR(1) stationary* $\Leftrightarrow |\phi| < 1$

What about the general case? If stationary,

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \ldots$$
$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \ldots$$

$\rightarrow$ these are called the Yule-Walker equations

# How to find AR(p) ACF coefficients?

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + W_t$$

## Yule-Walker equations

We have seen that for AR(1):

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

*Note that it would diverge with h if* $|\phi| > 1$*, AR(1) stationary* $\Leftrightarrow |\phi| < 1$

What about the general case? If stationary,

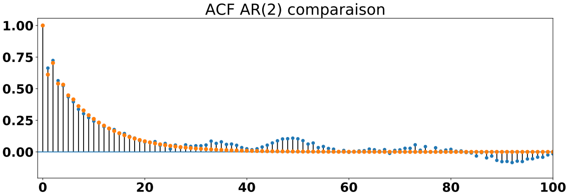
$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \ldots$$
$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \ldots$$

$\rightarrow$ these are called the Yule-Walker equations

# How to find AR(p) ACF coefficients?

## Illustration on a practical case

$$X_t = \frac{1}{3}X_{t-1} + \frac{1}{2}X_{t-2} + W_t$$



# Connection between MA and AR

## From MA to AR

Considering the MA(1) process:

$$
\begin{aligned}
X_t &= W_t + \beta W_{t-1} \\
\Rightarrow W_t &= X_t - \beta W_{t-1} \\
\Rightarrow W_t &= X_t - \beta(X_{t-1} - \beta W_{t-2}) \\
\Rightarrow W_t &= X_t - \beta X_{t-1} + \beta^2 X_{t-2} - \beta^3 X_{t-3} + \dots \\
\Rightarrow X_t &= W_t + \beta X_{t-1} - \beta^2 X_{t-2} + \beta^3 X_{t-3} - \dots
\end{aligned}
$$

In other words, MA(1) is an AR($\infty$) process

More generally, any MA(q) can be seen as an AR($\infty$) process (if MA(q) is invertible)

# Connection between MA and AR

## From MA to AR

Considering the MA(1) process:

$$\begin{array}{rcl} X_t & = & W_t + \beta W_{t-1} \\ \Rightarrow W_t & = & X_t - \beta W_{t-1} \\ \Rightarrow W_t & = & X_t - \beta(X_{t-1} - \beta W_{t-2}) \\ \Rightarrow W_t & = & X_t - \beta X_{t-1} + \beta^2 X_{t-2} - \beta^3 X_{t-3} + \dots \\ \Rightarrow X_t & = & W_t + \beta X_{t-1} - \beta^2 X_{t-2} + \beta^3 X_{t-3} - \dots \end{array}$$

In other words, MA(1) is an AR($\infty$) process

More generally, any MA(q) can be seen as an AR($\infty$) process (if MA(q) is invertible)

# Connection between MA and AR

## From AR to MA

$$
\begin{aligned}
X_t &= W_t + \phi X_{t-1} \\
\Rightarrow X_t &= W_t + \phi(W_{t-1} + \phi X_{t-2}) \\
\Rightarrow X_t &= W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \phi^3 X_{t-3} + \dots
\end{aligned}
$$

In other words, AR(1) is an MA($\infty$) process

More generally, any AR(p) can be seen as an MA($\infty$) process (if AR(p) is invertible)

# Connection between MA and AR

## From AR to MA

$$
\begin{aligned}
X_t &= W_t + \phi X_{t-1} \\
\Rightarrow X_t &= W_t + \phi(W_{t-1} + \phi X_{t-2}) \\
\Rightarrow X_t &= W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \phi^3 X_{t-3} + \dots
\end{aligned}
$$

In other words, AR(1) is an MA($\infty$) process

More generally, any AR(p) can be seen as an MA($\infty$) process (if AR(p) is invertible)

# The ARMA models

## ARMA(p,q) model

ARMA(p,q) model is a combination of AR(p) and MA(q) model:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$$

As for AR(p) and MA(q) parameters can be found from the ACF

## In practice

- fit the residuals with several (low) values of $p$ and $q$
- select what is the *best* model
- $\Rightarrow$ complete model:

  trend + seasonality + ARMA(p,q) residuals

# The ARMA models

## ARMA(p,q) model

ARMA(p,q) model is a combination of AR(p) and MA(q) model:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + W_t + \beta_1 W_{t-1} + \ldots + \beta_q W_{t-q}$$

As for AR(p) and MA(q) parameters can be found from the ACF

## In practice

- fit the residuals with several (low) values of *p* and *q*
- select what is the *best* model
- $\Rightarrow$ complete model:

  trend + seasonality + ARMA(p,q) residuals

# A few questions that we have left aside ...

- How to choose $p$ and $q$ of an ARMA process?
  $\rightarrow$ *useful to define partial ACF*
- How to find the coefficients of an ARMA process?
  $\rightarrow$ *transform it in a MA($\infty$) or AR($\infty$) process*
- How do we select the best models?
  $\rightarrow$ *complexity criteria and the overfitting problem*
- How to deal more systematically with the trend?
  $\rightarrow$ *differentiation method*
- How do we make a prediction with our model?
  $\rightarrow$ *extrapolate complete model to future values*

# A few questions that we have left aside ...

- How to choose $p$ and $q$ of an ARMA process?
  $\rightarrow$ *useful to define partial ACF*
- How to find the coefficients of an ARMA process?
  $\rightarrow$ *transform it in a MA($\infty$) or AR($\infty$) process*
- How do we select the best models?
  $\rightarrow$ *complexity criteria and the overfitting problem*
- How to deal more systematically with the trend?
  $\rightarrow$ *differentiation method*
- How do we make a prediction with our model?
  $\rightarrow$ *extrapolate complete model to future values*

# A few questions that we have left aside ...

- How to choose $p$ and $q$ of an ARMA process?
  $\rightarrow$ *useful to define partial ACF*
- How to find the coefficients of an ARMA process?
  $\rightarrow$ *transform it in a MA($\infty$) or AR($\infty$) process*
- How do we select the best models?
  $\rightarrow$ *complexity criteria and the overfitting problem*
- How to deal more systematically with the trend?
  $\rightarrow$ *differentiation method*
- How do we make a prediction with our model?
  $\rightarrow$ *extrapolate complete model to future values*

# A few questions that we have left aside . . .

- How to choose $p$ and $q$ of an ARMA process?
  $\rightarrow$ *useful to define partial ACF*
- How to find the coefficients of an ARMA process?
  $\rightarrow$ *transform it in a MA($\infty$) or AR($\infty$) process*
- How do we select the best models?
  $\rightarrow$ *complexity criteria and the overfitting problem*
- How to deal more systematically with the trend?
  $\rightarrow$ *differentiation method*
- How do we make a prediction with our model?
  $\rightarrow$ *extrapolate complete model to future values*

# A few questions that we have left aside ...

- How to choose $p$ and $q$ of an ARMA process?
  $\rightarrow$ *useful to define partial ACF*
- How to find the coefficients of an ARMA process?
  $\rightarrow$ *transform it in a MA($\infty$) or AR($\infty$) process*
- How do we select the best models?
  $\rightarrow$ *complexity criteria and the overfitting problem*
- How to deal more systematically with the trend?
  $\rightarrow$ *differentiation method*
- How do we make a prediction with our model?
  $\rightarrow$ *extrapolate complete model to future values*

# Longer term perspectives

- Traditional Box-Jenkins decomposition improvements: ARIMA, SARIMA
- Spectral methods using Fourier transform
- Learning methods: neural networks *(seq2seq)*

# Studying time series in python

Among several options, `pandas` library

A few useful functions:

- Load data as dataframe:
  `read_csv` from `pandas` library
- Moving average:
  `rolling` from `pandas` library
- Fitting:
  `curve_fit` in `scipy.optimize` library
- Autocorrelation function:
  `plot_ACF` in `statsmodels` library
- ARMA model fit:
  `ARMA.fit` in `statsmodels` library