Time Series

Introduction

A time series is a series of data points indexed in time order. Most often, a time series is a sequence taken at successive spaced points in time $\Rightarrow$ it is a sequence of discrete-time data denoted $(X_t)_{t \in \mathbb{N}}$ or $(X_t)_{t \in \mathbb{Z}}$ or $(X_t)_{t \in \mathbb{N}^*}$.

Time series are used in a lot of domains to describe a physical phenomenon. Here is some examples: statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, ...
Concrete examples of time series: heights of ocean tides, concentration in ozone, gas ~~con~~ consumption, ...

The analysis of time series has several aims:
- a better understanding of the physical phenomenon
- a simplified representation by a stochastic model
- forecasting

Trend, seasonality and noise

The first idea we may have is to draw the graph of the time series, ie $t \mapsto y_t$
By this, we can try to determine a family of functions that have the same shape, and then to ~~det~~ select the 'best' function by minimizing a goodness-of-fit criterion as least squared criterion, and potentially by considering a penalized version to take into account the complexity of the function.

RB: the family of functions is the model!

As the models are false but often usefull, we add some noise.
By this way, we obtain time series!

In this lecture, we consider models that can be decomposed as:

$$y_t = d_t + z_t$$

with $t \mapsto d_t$ a deterministic function
$t \mapsto z_t$ a random noise

Most often, $t \Rightarrow d_t$ depends on a small number of parameters that we want to estimate, whereas the random noise is assumed to be stationary. We are going to see the definition of this notion further but, in some sense, it means that its expectation and its covariance do not vary with the time.

More precisely:
$(x_t)_{t \in T}$ is modelised by one trajectory of a stochastic process $(X_t)_{t \in T}$ where:

$$X_t = \underbrace{(m_t + s_t)}_{= d_t} + z_t$$

- $t \Rightarrow m_t$ is <u>the trend</u>. In general, this is a function that varies slowly.
  Ex: a polynomial function $m_t = a_0 + a_1 t + \dots + a_d t^d$

In practice: we get an idea of d by the graph
after if we let $a = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix}$, we estimate it by least squared.

Recall : least squared method for $d=2$

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \quad A = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}$$

x_i being observed at time t_i

$$\hat{a}_n \in \underset{a}{\operatorname{argmin}} \sum_{i=1}^{n} \left(x_i - (a_0 + a_1 t_i + a_2 t_i^2) \right)^2$$

then $\hat{a}_n = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} \in \underset{a}{\operatorname{argmin}} \| X - Aa \|_2^2$

and we find $\hat{a}_n = ({}^tAA)^{-1} \, {}^tAX$

the residuals are : $x - A\hat{a} = x - A({}^tAA)^{-1} \, {}^tAX$

. $t \to s_t$ is <u>the seasonality</u>. This is a periodic function.

an example is $s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos(\lambda_t t) + b_j \sin(\lambda_j t) \right)$

with a_j, b_j to estimate and $\lambda_j \in \left\{ k \frac{2\pi}{d}, k \in \mathbb{Z} \right\}$

we identify the period d et the λ_j and the we estimate the a_j, b_j by least squared method.

. $(Z_t)_{t \in \mathbb{Z}}$ is <u>the noise</u>. We hope it is stationary.

<u>Stationary process</u>

A stochastic process a a process is a family $(X_t)_{t \in \mathbb{Z}}$ of random variables with values in $\mathbb{R}$.

The time is the index, and the

$$\begin{aligned} \Omega \times \mathbb{Z} &\longrightarrow \mathbb{R} \\ (\omega, t) &\longrightarrow X_t(\omega) \end{aligned}$$

Rk: for all $t \in T$, X_t is a random variable !
for all $\omega \in \Omega$, $t \to X_t(\omega)$ is a trajectory of the process !

definition:
The process $(X_t)_{t \in \mathbb{Z}}$ is strongly stationary if for all $h \in \mathbb{Z}$, and for all finite sequence $t_1, \dots, t_n \in \mathbb{Z}$ whose length is n with $n \geq 1$, the random vectors $(X_{t_1}, \dots, X_{t_n})$ and $(X_{t_1+h}, \dots, X_{t_n+h})$ have the same distribution.

Rk: This property of strongly stationary is a property of time translation invariance. Thus parameters as expectation and variance do not change over time !

definition: mean and autocovariance
Let $(X_t)_{t \in \mathbb{Z}}$ a process of order two (it means that $\mathbb{E}[|X_t|^2] < +\infty$ for all $t \in \mathbb{Z}$).
For such a process, we define
. the mean function μ_X
$$\begin{aligned} \mathbb{Z} &\to \mathbb{R} \\ t &\to \mathbb{E}[X_t] \end{aligned}$$

. the autocovariance function γ_X
$$\begin{aligned} \mathbb{Z} \times \mathbb{Z} &\to \mathbb{R} \\ (t, s) &\to \text{cov}(X_t, X_s) \end{aligned}$$

Rk: . $\gamma_X(s, t) = \text{cov}(X_t, X_s)$
$= \mathbb{E}[X_t X_s] - \mu_X(t) \times \mu_X(s)$

. $\gamma_X(t, t) = \mathbb{V}[X_t]$

definition:
Let $(X_t)_{t \in \mathbb{Z}}$ a process of order two.
We say that $(X_t)_{t \in \mathbb{Z}}$ is stationary if for all $h \in \mathbb{Z}$, for all finite sequence ~~(...)~~ $t_1, \dots, t_n \in \mathbb{Z}$ with $n \geq 1$,
$(X_{t_1}, \dots, X_{t_n})$ and $(X_{t_1+h}, \dots, X_{t_n+h})$ have the same expectation and

the same autocovariance function.
It means that:
$\forall s,t,h \in \mathbb{Z}, \quad \mu_X(t) = \mu_X(t+h)$ and $\gamma_X(t,s) = \gamma_X(t+h, s+h)$

Rk: If $(X_t)_{t\in\mathbb{Z}}$ is stationary then $\forall s,t\in\mathbb{Z}, \ \gamma_X(s,t) = \gamma_X(0, t-s)$.

Thus the autocovariance function $\gamma_X(s,t)$ just depends on the delay $t-s$.
In this case, γ_X can be replaced by the function that just depends on one input γ_X given by the following definition:

definition:
If $(X_t)_{t\in\mathbb{Z}}$ is stationary, its autocovariance function is defined by:

$$\gamma_X : \mathbb{Z} \longrightarrow \mathbb{R}$$
$$h \longrightarrow \gamma_X(h) = \gamma_X(0,h) = \operatorname{cov}(X_t, X_{t+h}) \ \forall t\in\mathbb{Z}$$

So $\forall t,s\in\mathbb{Z} \quad \gamma_X(s,t) = \gamma_X(t-s)$

Rk: If $(X_t)_{t\in\mathbb{Z}}$ is stationary, then:

- $t \Rightarrow \mu_X(t)$ is constant
- $t \rightarrow V[X_t]$ is also constant

definition
Let $(X_t)_{t\in\mathbb{Z}}$ a stationary process.
Its autocorrelation function is given by:

$$\rho_X : \mathbb{Z} \longrightarrow [-1,1]$$
$$h \longrightarrow \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\operatorname{cov}(X_t, X_{t+h})}{\sqrt{V[X_t]\,V[X_{t+h}]}} \quad \forall t\in\mathbb{Z}$$