Classical Analysis of Time Series

Josep Vives

February 15, 2024

1 What is a time series? What is its importance?

A time series is a collection of data indexed by time. Usually is obtained from a series of measures or quantitative observations collected over time, in equally spaced time intervals, that is, in a regular basis. We talk about yearly, monthly, daily of hourly data. We will write

$$\{x_i,\ i = 0, 1, \ldots, n, \ldots\}.$$

Examples of time series are closing daily prices of any financial asset, monthly volume of sales of a good in a market, monthly victims of accidents in a place, the yearly mean level of water in a lake, the amount of people every ten years in a region, and so on.

Data x_i can be one dimensional (scalar data) or multidimensional (vector data). And can be discrete observations of continuous (or almost continuous) phenomena as price and temperature, or discrete observations of discrete data as a volume of sales or a number of accidents.

As we can see reading scientific journals or newspapers, time series appear in all domains of knowledge where time passing plays a role.

The central idea of time series analysis is the idea of stochastic dependence, in front of the idea of stochastic independence.

If $x_1, \ldots, x_n$ are independent data, the order doesn't play any role and we can analyze these data by classical techniques of Statistics. On the contrary, if data are not independent, to take into account the order is fundamental. Usual techniques of Mathematical Statistics have no sense and we need a different Probability Theory: the Theory of Stochastic Processes. We insist in the key fact that in a time series, after x_i comes x_{i+1}, and to change this order implies to loose information.

Assume as usual in Probability Theory that our observations $x_1, \ldots, x_n$ are observations of random variables $X_1, \ldots, X_n$. From now on, denote by $\mathbb{E}$, $\mathbb{V}$ and $\mathbb{C}$ the expectation, the variance and the covariance of these random variables, respectively.

In a first step, usually, Mathematical Statistics assumes $X_1, \ldots, X_n$ are independent and identically distributed (iid) random variables and we say that these random variables constitute an iid sample. This will be not true any more in this course, because our random variables will not be necessarily independent nor identically distributed.

If we assume that we have an iid sample when in fact it is not true, we can commit errors as the following example shows.

Assume for the moment that $X_1, \ldots, X_n$ is an iid sample, that is, a collection of iid random variables. Assume $\mathbb{E}(X_i) = \mu \in \mathbb{R}$ and $\mathbb{V}(X_i) = \sigma^2 > 0$, for any $i \geq 1$.

Consider the typical mean estimator

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

It is well known that this is a random variable with expectation μ and variance $\frac{\sigma^2}{n}$. But what happens if the random variables $X_1, \ldots, X_n$ are not iid? In this case we have

$$\mathbb{V}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}(X_i) + \frac{1}{n^2} \sum_{i \neq j} \mathbb{C}(X_i, X_j)$$

where now $\mathbb{C}(X_i, X_j) \neq 0$ for any i, j.

Assume for example that all variables are identically distributed with variance σ^2 and correlations ρ_{ij}. Recall that this means $\mathbb{C}(X_i, X_j) = \sigma^2 \rho_{ij}$. Then,

$$\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n}(1 + \frac{2}{n} \sum_{i<j} \rho_{ij})$$

and this quantity can be quite different from $\frac{\sigma^2}{n}$.

For example under the iid hypothesis, the central limit theorem says that $\bar{X}_n$ is approximately normal and we write

K.O.θ

$$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$

.

This allows to establish the following asymptotic confidence interval

$$\mathbb{P}\left\{\mu \in \left[\bar{X}_n - 1.96\,\frac{\sigma}{\sqrt{n}}, \quad \bar{X}_n + 1.96\,\frac{\sigma}{\sqrt{n}}\right]\right\} = 0.95.$$

Recall that if σ is unknown, Mathematical Statistics allows also to prove that

$$\mathbb{P}\left\{\mu \in \left[\bar{X}_n - 1.96\,\frac{S}{\sqrt{n}}, \quad \bar{X}_n + 1.96\,\frac{S}{\sqrt{n}}\right]\right\} \simeq 0.95.$$

where

$$\boxed{S := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}}$$

and so, we can say that

$$\left[\bar{X}_n - 1.96\,\frac{S}{\sqrt{n}}, \quad \bar{X}_n + 1.96\,\frac{S}{\sqrt{n}}\right]$$

is an approximated confidence interval with confidence level of 95% for μ.

But if we have no independence and $\frac{2}{n}\sum_{i<j}\rho_{ij}$ is a positive quantity much greater than 0, $\mathbb{V}(\bar{X}_n)$ is much greater than $\frac{\sigma^2}{n}$ and so, a confidence interval not taking this into account doesn't attain the desired confidence level.

In a time series, the correlation between the random variables in different time instants, called autocorrelation, is the key element to analyze.

2 Classical or macroscopic analysis of time series

The first step in analyzing a time series is to observe its graphical representation in an adequate scale. We take into account immediately some phenomena: trend, periodicities, outliers, etc.

The classical point of view, the point of view before the seventies, assumes that a time series is composed by different components: a trend (T), some cycles ($C^{(j)}$), and a stationary residual (R). We write

$$X_i = f(T_i, C_i^{(1)}, \dots, C_i^{(k)}, R_i).$$

For example, if X_i represents daily sales of a big store we can observe at minimum an increasing or decreasing trend, a yearly cycle (September if different from August), a monthly cycle (it is not the same the beginning than the end), a weekly cycle (Saturday is not the same as Monday) and a residual component.

In this course we will always assume that our model is additive, that is

$$X_i = T_i + C_i^{(1)} + \dots + C_i^{(k)} + R_i.$$

Recall that frequently, a non additive model can be transformed to an additive one applying a logarithmic transformation to data. In this sense, to assume additivity is not so much restrictive.

Classical methodology establishes several methods to estimate and isolate the different components of a times series.

Note that if we assume for example only one cyclic component, say a yearly one, we can write

$$X_i = T_i + C_i + R_i$$

but we could also write

$$X_{i,j,k} = T_i + C_j + R_{i,j,k}$$

where i denotes the year, j the month and k the day in the month, for example.

Given a time series, we want to decompose it in a way that trend and cycles be clearly distinguished and the residual component seems to be stationary, in the sense that it moves around zero, with no seasonal or cyclical patterns and with no increasing or decreasing variability. In the following subsections techniques to do so will be given. These are descriptive techniques, so, any probabilistic model is assumed. We simply manipulate data.

The ideas presented in this chapter are quite simple and model free, and classical, but extremely useful in a first attempt, to understand what happens in a concrete time series.

2.1 Trend identification

Essentially there are three ways to identify the trend of a time series. One is functional fitting, that is, to fit a certain type of function to data, usually a polynomial. Another is filtering the series to obtain a smooth function that could be interpreted as a trend. The third one is differencing, a way to erase the trend from a series.

2.1.1 Functional Fitting

From the graphics of the series we can identify a type of function that can describe well the trend of data, usually a polynomial, and adjust it using the well-known least squares method. That is, we can assume

$$x_i = f(i) + e_i$$

with

$$f(i) = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m$$

and choose $a_0, \ldots, a_m$ minimizing

$$\boxed{\varphi(a_0, a_1, \ldots, a_m) := \sum_{i=1}^{n} (x_i - f(i))^2.}$$

Usual cases are to fit a straight line $a_0 + a_1 i$ or a parabola $a_0 + a_1 i + a_2 i^2$. Of course the method can be generalized to other families of functions like splines, exponential functions or others.

2.1.2 Filtering

A filter is a linear combination of variables of type

$$\boxed{F_q(x_n) = \sum_{r=-q}^{q} a_r x_{n+r}.}$$

The filter F_q is usually assumed symmetric, that is $a_r = a_{-r}$, and such that $\sum_{r=-q}^{q} a_r = 1$.

The simplest case is the case of moving averages, that is, the case $a_r = \frac{1}{2q+1}$ for any r. We write

$$\boxed{A_q(x_n) := \frac{1}{2q+1}\sum_{r=-q}^{q} x_{n+r}}$$

Note that in this case $A_q(x_n)$ is simply the mean of x_n and the q previous and posterior data. If $y_n := A_q(x_n)$, the data of process y runs in the set $q+1 \leq n \leq N-q$, that is, we loose $2q$ data, q on every side. Of course, the series $\{y_k,\ k = q+1, \ldots, n-q\}$ is smoother than the series $\{x_k,\ k = 1, \ldots, n\}$.

Note that given a functional trend we can construct a filter that maintains the trend invariant. Assume

$$x_i = a_0 + a_1 i + \varepsilon_i$$

where $\{\varepsilon_i,\ i = 1, \ldots, n\}$ is a centered series.

Consider $q = 1$ and

$$A_1(x_n) = \frac{1}{3}(x_{n-1} + x_n + x_{n+1}).$$

Then,

$$\begin{aligned}
A_1(x_n) &= \frac{1}{3}\sum_{r=-1}^{1}(a_0 + a_1(n+r) + \varepsilon_{n+r}) \\
&= a_0 + a_1 n + a_1\frac{1}{3}\sum_{r=-1}^{1} r + \frac{1}{3}\sum_{r=-1}^{1}\varepsilon_{n+r} \\
&= a_0 + a_1 n + \frac{1}{3}\sum_{r=-1}^{1}\varepsilon_{n+r} \\
&= a_0 + a_1 n + A_1(\varepsilon_n)
\end{aligned}$$

because $\frac{1}{3}(-1+0+1) = 0$.

Note that for any q, the moving average series

$$\{A_q(\varepsilon_n),\ n = q+1, \ldots, N-q\}$$

is smoother than the initial series of residuals ε. So, moving averages, used appropriately, are a good method to almost isolate trends. But this method is not able to predict; on the contrary, the fitting method it is.

A particular case of filtering that is used to predict is the so-called exponential smoothing, a method related with the well-know Holt-Winter prediction algorithm that will see later in this course. It is similar to moving average but the smoothing is done only on the left hand side.

Imagine we have a times series $x_0, x_1, \dots, x_n$ with no cyclic components.

We define the filtered times series $y_0, \dots, y_n$ recursively as $y_0 := x_0$ and

$$\boxed{y_n = E_a(x_n) := a x_n + (1-a) y_{n-1},}$$

with $a \in (0,1)$.

Note that this implies

$$\boxed{y_n := \sum_{j=0}^{n-1} a(1-a)^j x_{n-j} + (1-a)^n x_0}$$

Note that if a is close to 1, the filtered series is almost equal to the original one. On the contrary, if a is close to 0 the filtered series is much more smooth and changes very slowly from x_0. The closer is a to 0 the smoother is the filtered series. A typical value is $a = 0.2$.

2.1.3 Differencing

This is essentially a method to erase the trend. But it is a reversible method, so we can recuperate the trended series from the untrended one.

Consider the operators

$$Bx_i := x_{i-1}.$$

and

$$\nabla x_i := (\mathrm{Id} - B)x_i = x_i - x_{i-1},$$

Operator B is called the backward shift operator and operator ∇ is called the first difference operator.

We denote the series $\{y_i,\ \ i \geq 1\}$ with $y_i = \nabla x_i$, as the series of first differences of the series $\{x_i, i \geq 0\}$.

Of course we can apply operator ∇ recursively and then

$$\nabla^j x_i = (\mathrm{Id} - B)^j x_i, \quad i \geq 1.$$

Note also that

$$x_i = x_i - x_{i-1} + x_{i-1} - x_{i-2} + \cdots + x_1 - x_0 + x_0,$$

that is,

$$\boxed{x_i = x_0 + \sum_{l=1}^{i} y_l}$$

ans so, series x can be recuperated from series y.

Note that to transform series x in series y is a way to erase the trend. If we assume for example a linear trend

$$x_i = a_0 + a_1 i + \epsilon_i$$

with ϵ a centered series, we have

$$y_i = a_1 + \epsilon_i - \epsilon_{i-1}$$

that of course is a series with a constant trend a_1.

2.2 Identification of cycles

To fix ideas we analyze a particular case of cyclic behavior, the seasonal one. Its generalization is straightforward. As before, we have three main ways to identify cycles.

2.2.1 Direct identification

Consider a series $\{x_i,\ i = 1, \ldots, n\}$. Assume it has a seasonal component $\{s_i,\ i = 1, \ldots, n\}$ with $s_i = s_{i+p}$, $p = 12$ and $\sum_{i=1}^{p} s_i = 0$. We want to identify it. Assume we have data of m years, that is $n = mp$ where n is the number of data.

We can find the monthly means and to write

$$e_i := \frac{1}{m} \sum_{j=0}^{m-1} x_{i+pj},\ i = 1, \ldots, p.$$

In order to make the component centered we have to compute

$$\bar{e} := \frac{1}{p} \sum_{i=1}^{p} e_i.$$

Finally,

$$s_i := e_i - \bar{e}$$

is the seasonal component.

2.2.2 Filtering

Filtering can be used to erase cycles from a series and identify it. Assume e is a cycle (a cyclic series) with an odd period $p := 2q + 1$, that is, assume $e_i = e_{i+p}$ for any $i \geq 0$. Any symmetric filter with p points eliminates the cycle. Note that the filter

$$A_q(x_i) := \frac{1}{2q+1} \sum_{j=-q}^{q} x_{i+j}$$

satisfies

$$A_q e_i = A_q e_{i+1}, \ \forall i.$$

But usual periods are even; the year has twelve month, or four seasons. In this case, the suitable filter is

$$\tilde{A}_q(x_i) := \frac{1}{2q} \left\{ \frac{1}{2} x_{i-q} + x_{i-q+1} + \cdots + x_{i+q-1} + \frac{1}{2} x_{i+q} \right\}$$

with $p = 2q$ and $i \in \{q+1, \ldots, N-q\}$.

In any case, using the adequate filter we can eliminate the seasonal effect. Of course, then, subtracting the new unseasonalized series from the original one we identify the seasonal component.

2.2.3 Differencing

Differencing is also useful to eliminate the seasonal component. For example, if we apply the operator

$$\nabla_p x_i := (\mathrm{Id} - B^p) x_i = x_i - x_{i-p}$$

where p is the period of the seasonal component, we erase it.

Note that in the case of yearly periodicity, variables $y_i := \nabla_{12} x_i$ represent the increments between a month and the same month of the previous year.

2.3 Transforming

Frequently it is convenient to transform initial data in order to treat them in a more significant way. The most important transformation is the logarithmic one. That is

$$y_i := \log x_i, \quad i \geq 1.$$

This transformation has several utilities. The first one, to transform products in sums. The second, to transform positive data in real data. The third, to neutralize exponential trends. The fourth one is that to take logarithms stabilize the series in the sense changes on variability reduces. Of course, a logarithmic transformation is reversal.

In Finance, for example, the basic objects of interest are not prices but the so called log-prices.

3 Exercises

1. Show that a linear filter of coefficients $\{a_j\}$ maintains a polynomial of degree k, $P(j) = c_0 + c_1 j + \cdots c_k j^k$ without distortion if and only if $\sum_j a_j = 1$ and $\sum_j j^r a_j = 0$, $\forall r = 1, \ldots, k$.

2. If $P(j) = \sum_{k=0}^{p} c_k j^k$ for any $j \in \mathbb{Z}$, prove that ∇P, where $\nabla := Id - B$, is a polynomial of degree $p-1$ in j and so, $\nabla^{p+1} P = 0$.