Time Series Analysis

Prof. Lea Petrella

MEMOTEF Department
Sapienza, University of Rome, Italy

Class 3

- Example 2. Process with trend. The following process $X_t = \alpha + \beta t + \epsilon_t$ with $\epsilon_t \sim WN(0, \sigma^2)$ is non- stationary (in mean...although it is stationary in variance!)
- In fact:
- $\mathbb{E}(X_t) = \alpha + \beta t$
- $\gamma(0) = \sigma^2$
- $\gamma(h) = Cov(X_t, X_{t+h}) = \mathbb{E}\left[(X_t - \mathbb{E}(X_t))(X_{t+h} - \mathbb{E}(X_{t+h}))\right]$

$$= \mathbb{E}\left[(\alpha + \beta t + \epsilon_t - (\alpha + \beta t))(\alpha + \beta(t+h) + \epsilon_{t+h} - (\alpha + \beta(t+h)))\right]$$

$$= \mathbb{E}\left[\epsilon_t \epsilon_{t+h}\right] = 0$$

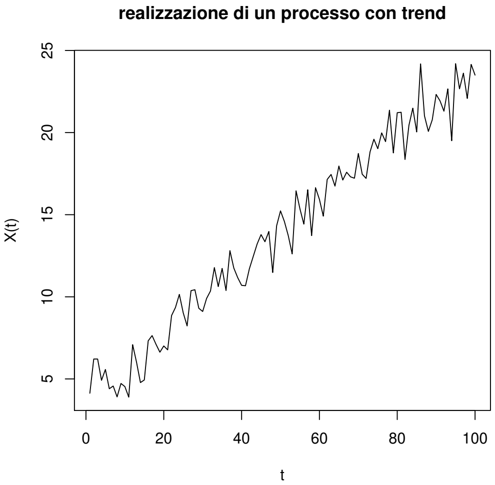

Figure: Model with trend equal to $4 + 0.2t$ and $N(0, 1)$.

- Example 2. Random Walk. A random walk process is the easiest (non stationary) stochastic process one can think of and represents the position of a particle that moves randomly after t steps ($t = 0, 1, 2, \dots$).

$$X_t = \epsilon_t + \epsilon_{t-1} + \ldots + \epsilon_0 = X_{t-1} + \epsilon_t.$$

- If the movement of the particle is such that $P(\epsilon_t = 1) = \frac{1}{2}$ and $P(\epsilon_t = -1) = \frac{1}{2}$, assuming that the process starts form the origin, $\epsilon_0 = 0$, we will obtain the following graphs:

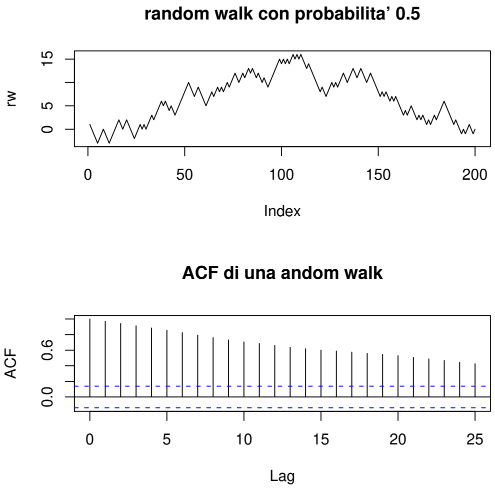

Figure: Representation of a Random Walk and its ACF. Non stationary process.

- If we want to analytically evaluate the stationary property of $X_t \sim WN(0, \sigma^2)$, we have that
- $\mathbb{E}(X_t) = \mathbb{E}(\epsilon_0 + \epsilon_1 + \ldots + \epsilon_{t-1} + \epsilon_t) = 0$,
- $\mathbb{V}ar(X_t) = \gamma(0) = \mathbb{V}ar(\epsilon_0 + \epsilon_1 + \ldots + \epsilon_t) = t\sigma^2$,
- $\gamma(t, t+h) = Cov(X_t, X_{t+h}) = \mathbb{E}(X_t, X_{t+h})$

$$= \mathbb{E}\left[(\epsilon_0 + \epsilon_1 + \ldots + \epsilon_t), (\epsilon_0 + \ldots + \epsilon_t + \epsilon_{t+1} + \ldots + \epsilon_{t+h})\right] =$$

$$= \mathbb{E}(\epsilon_0^2 + \epsilon_1^2 + \ldots + \epsilon_t^2) = t\sigma^2.$$

- The process is non-stationary in that the moments of order 2 depend on time t.
- The Random Walk is also called unit-root. We will discuss later the reason.

Sample ACF

- The ACF is important in order to understand if the process is stationary. It gives information on the existing dependence structure.
- It gives information on which kind of model can be used to study the process.
- It is then important to estimate the ACF generated by the observed data by using samples obtined.
- Let $\{x_1, \ldots, x_n\}$ be the observed time series and

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

be its sample mean.

Stationary property

- Then the sample autocovariance function is given by

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

  and the sample ACF is given by

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

- For example, if the series has length $n = 25$ then the estimate can be $\hat{\rho}(0) = 1, \hat{\rho}(1) = 0.387, \hat{\rho}(2) = 0.164, \ldots, \hat{\rho}(13) = -0.265.$

- $\hat{\rho}(h)$ represents the autocorrelation function observed for the time series for different values of h and gives information about the ACF $\rho(h)$ of the (unknown) process that generated the dataset.
- The graph of $\hat{\rho}(h)$ with respect to different lags h is defined as correlogram of x_t.
- The correlogram gives information about the existing ACFs in the observed sample, that is information about serial dependence.

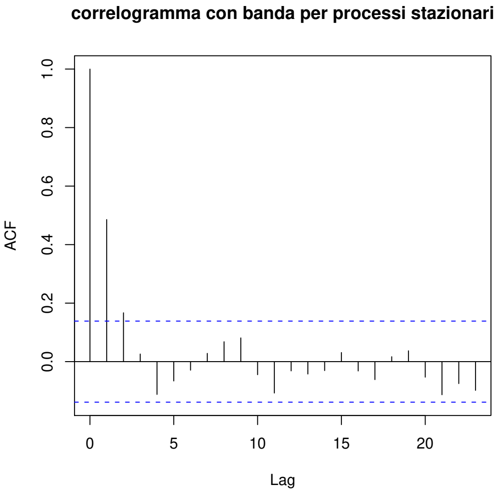

Figure: Correlogram of a stationary series.

- It can be shown that there is no ACF in the generating process (i.e., if $\rho(h) = 0$ when $h \neq 0$), that is, if the process is a white noise then in the observed series and sa n increases we have

$$\hat{\rho}(h) \sim N(0, \frac{1}{n}).$$

- For a series without autocorrelation we expect that the estimated ACF decreses as h increases and that about 95% of the PACF lies in the interval $\pm 1,96/\sqrt{n}$.
- This means that in the plot of the ACF, the bars corresponding to the estimated autocorrelation that are inside the blue lines (the interval $\pm 1,96/\sqrt{n}$) can be considered null.

- The ACF is considered to be significantly different from zero if its value lies outside the interval
$$\left(-\frac{1,96}{\sqrt{n}}, \frac{1,96}{\sqrt{n}}\right).$$
- Values of $\hat{\rho}(h)$ inside the interval, although different for zero, suggest that the estimated autocorrelation can be due to randomness (i.e., not being a property of the process). Think about the correlogram of the white noise.
- Note, however, that even when there is no autocorrelation, we sometimes expect to see $\hat{\rho}(h)$ outside the blue lines.
- That is, when computing the ACF for the first 30 ACF coefficients, we can expect to see one, two or even three ouside the blue lines.

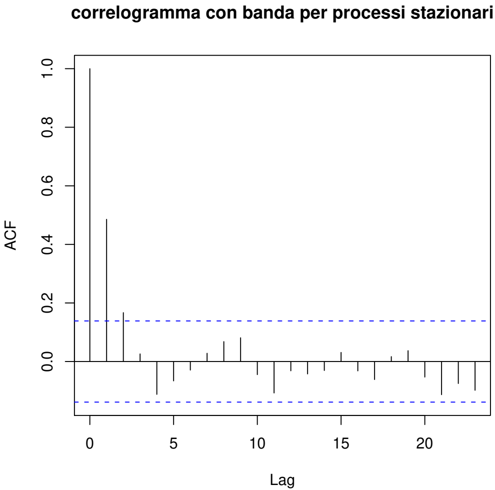

Figure: The blue lines correspond to the interval $\pm 1,96/\sqrt{n}$ per $n = 200$.

Some instruments: lag operator and differences

- lag operator:

$$Bx_t = x_{t-1}$$

- when applied twice to the series we get

$$B^2 x_t = B(Bx_t) = B(x_{t-1}) = x_{t-2}$$

- Morel generally,

$$B^k x_t = x_{t-k}.$$

- difference operator:

$$\nabla_1 x_t = x_t - x_{t-1} = (1 - B)x_t.$$

- More generally,

$$\nabla_k x_t = x_t - x_{t-k} = (1 - B^k)x_t.$$

Modeling

- One of the most important steps in statistics is how to build models that can represent well phenomena of interest, with the goal to describe, interpret and possibly forecast them.
- Generally, a model is characterized by a relation cause-effect among the variables that is the most parsimonious that can be achieved, that is a relation that makes use of the less number of parameters to represent it.

- In statistics, that kind of relation is not expresses in a deterministic form such as

$$X_t = f(X_{t-1}, \ldots, X_{t-k}, \theta).$$

- Rather, it includes a noise component ϵ_t that that accounts for the randomness related with the choice of the model

$$X_t = f(X_{t-1}, \ldots, X_{t-k}, \theta) + \epsilon_t.$$

- For example, we can think of a model

$$X_t = \alpha + \beta X_{t-1} + \epsilon_t.$$

- The goal is to understand the model that generate the observed series.

Building the model

- Choose a class of models that may have generated the observed series (ARMA,ARIMA,ARCH,GARCH).
- Identify the class of models in a parsimonious way, among those that better fit the observed series the model with less explanarory variables (use ACF,PACF, AIC).
- Estimate the parameters (OLS, MLE).
- Diagnostic: evaluate if the selected model fits the observed series and/or if the hypothesis about the distribution of the shock are correct.
- Use the model to forecast.

ARMA models

- We will first study a family of linear models, $ARMA(p, q)$, that evolve in time according to a dynamics that explains present vales depending on past values by means of a linear relation.
- More specifically, within this class of models the observation at time t depends on that at time $t-1, t-2, \ldots, t-p$ and on the shocks at time $t, t-1, t-2, \ldots, t-q$.
- We will consider Moving Average models, $MA(q)$, then Autoregressive, $AR(p)$, and finally the Autoregressive Moving Average $ARMA(p, q)$.