Lecture 1

A *time series* is a set of observations (x_t), each one being recorded at a specific time t. This type of data appears naturally in everyday life, for example in:

- Medicine: heartbeat of a healthy person;
- Economics and Finance: stock prices at the end of the day;
- Demographics: population per year;
- Transport: international air traffic.

The purpose of the first part of the course is to present techniques for modeling these observations, i.e. the aim is to suggest a probabilistic model to represent the data. We think of x_t as the output of a stochastic process at time t.

Definition 1. *Let $T \subset \mathbb{R}$ be a* time set*, i.e. if $t, s \in T$ then $t+s \in T$. A family $X = (X_t)_{t \in T}$ of random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a* stochastic process*. For $T \in \{\mathbb{N}, \mathbb{Z}\}$, X is called* time series*.*

Remark 1. *A time series is a mathematical sequence, not a series. A time series is an infinite sequence*

$$\ldots, X_{t_1}, \ldots, X_{t_n}, \ldots \quad t_i \in \mathbb{Z} \text{ or } t_i \in \mathbb{N}$$

of random variables (or random vectors).

The aim of the probabilistic modeling behind a time series is

- understanding of the stochastic system;
- predicting the future, i.e. predicting X_{t_n+1} given the observations $(X_{t_1}, \ldots, X_{t_n})$;
- drawning inference.

To achieve these tasks it is necessary to assume some a-priori structure of the time series otherwise any conclusion about the joint distribution of $(X_{t_1}, \ldots, X_{t_n})$ would be possible. A basic type of structure is stationarity.

Definition 2. *$X = (X_t)_{t \in T}$ is* strictly stationary *if*

$$\forall n \in \mathbb{N},\ t_1, \ldots, t_n, t \in T \quad \mathscr{L}\big((X_{t_1}, \ldots, X_{t_n})\big) = \mathscr{L}\big((X_{t_1+t}, \ldots, X_{t_n+t})\big),$$

i.e. $\forall A \in \mathscr{B}(\mathbb{R}^n)$: $\mathbb{P}\big((X_{t_1}, \ldots, X_{t_n}) \in A\big) = \mathbb{P}\big((X_{t_1+t}, \ldots, X_{t_n+t}) \in A\big)$.

Loosely speaking, X is stationary if its statistical properties do not change with time.

Definition 3. *$X = (X_t)_{t\in T}$ is* weakly stationary *(*second order stationary*) if $\mathbb{E}[X_t^2] < \infty$ for all $t \in T$, the expectation function $t \to \mu(t) := \mathbb{E}[X_t]$ is constant and the covariance function satisfies $Cov(X_u, X_s) = Cov(X_{u+t}, X_{s+t}),\ \forall u, s, t \in T$.*

For a weakly stationary series X we use the following vocabulary:

- *Autocovariance at lag h:* $c(h) := \mathrm{Cov}(X_{t+h}, X_t)$.
- *Autocorrelation at lag h:* $\rho(h) := \rho(X_{t+h}, X_t) = \frac{c(h)}{c(0)}$.

Example 1. *If $X = (X_t)_{t\in T}$ are i.i.d., then X is strictly stationary.*

Example 2. *If X is in L_2 (i.e. $\mathbb{E}[X_t^2] < \infty$, for all $t \in T$) and it is strictly stationary, then X is weakly stationary.*

Many time series in real life are not stationary but they can often be transformed in a time series that is (assumed to be) stationary. Indeed, the "classical decomposition" model is

$$X_t = m_t + s_t + Y_t,$$

where

- m_t is a slowly changing function called the *TREND COMPONENT*;
- s_t is a function with known period called the *SEASONAL COMPONENT* (e.g. weaker temperature in winter than in summer);
- Y_t is a stationary time series.

Given $x_t = X_t(\omega)$, a realization of X, there are techniques to extract the deterministic components m_t and s_t so we are only left to study the residual Y_t.

If X is a weakly stationary process then its mean μ and autocovariance function c contribute to characterize it. However, μ and c are generally unknown quantities, they need to be estimated from the data.

Estimation of μ and c

Reminder 1. *Let $\mathscr{P} = (\mathcal{X}, \mathcal{F}, (P_\theta : \theta \in \Theta))$ be a statistical model and $X_1, \dots, X_n$ be random variables defined on $(\mathcal{X}, \mathcal{F})$. An estimator $\widehat{\theta}$ of the unknown parameter $\theta \in \Theta$ from realizations $x_1, \dots, x_n$ of $X_1, \dots, X_n$ is a measurable function of $X_1, \dots, X_n$, i.e.*

$$\widehat{\theta} = \widehat{\theta}(X_1, \dots, X_n) \quad \textit{is a random variable.}$$

Good estimators are estimators that enjoy nice properties:

- *$\widehat{\theta}$ is an* unbiased *estimator of θ if $\mathbb{E}_\theta[\widehat{\theta}] = \theta$, for all $\theta \in \Theta$.*
- *$\widehat{\theta}$ is a* consistent *estimator of θ if $\forall \varepsilon > 0$ $P_\theta(|\widehat{\theta} - \theta| > \varepsilon) \to 0$ as $n \to \infty$, for all $\theta \in \Theta$.*

- *If $\widehat{\theta}$ and θ^* are unbiased estimators of θ we say that $\widehat{\theta}$ is* more effective *than θ^* if $Var_\theta(\widehat{\theta}) \leq Var_\theta(\theta^*)$.*

If nothing is known about the distribution of the time series X, besides that it is stationary, then the natural estimators for μ and c are their sample counterparts:

$$\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \textit{empirical mean (sample mean)}$$

$$\widehat{c}(k) := \frac{1}{n}\sum_{l=1}^{n-k}(X_l - \widehat{\mu}_n)(X_{l+k} - \widehat{\mu}_n), \qquad \textit{empirical autocovariance function at lag } k.$$

Set $\widehat{c}(-k) = \widehat{c}(k)$. The *empirical autocovariance matrix* is

$$\widehat{C}_n = \begin{pmatrix} \widehat{c}(0) & \widehat{c}(1) & \dots & \widehat{c}(n-1) \\ \widehat{c}(1) & \ddots & \ddots & \widehat{c}(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ \widehat{c}(n-1) & \dots & \widehat{c}(1) & \widehat{c}(0) \end{pmatrix}.$$

Observe that $\widehat{C}_n$ is a Toeplitz matrix.

Properties of $\widehat{\mu}_n$

Theorem 1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary process with autocovariance function c and mean μ, then for the sample mean $\widehat{\mu}_n$ the following properties hold:*

1. *(Unbiasedness) $\mathbb{E}[\widehat{\mu}_n] = \mu$;*
2. *(Variance) $Var[\widehat{\mu}_n] = \frac{1}{n}\sum_{k=-n+1}^{n-1}\left(1 - \frac{|k|}{n}\right)c(k)$;*
3. *(Consistency) If $c(k) \xrightarrow[k\to\infty]{} 0$, then $\mathbb{E}[(\widehat{\mu}_n - \mu)^2] \to 0$;*
4. *(Asymptotic variance) If $\sum_{k=-\infty}^{\infty}|c(k)| < \infty$ then $n\,Var(\widehat{\mu}_n) \to \sum_{k=-\infty}^{\infty} c(k)$.*

Proof. 1. $\mathbb{E}[\widehat{\mu}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{n\mu}{n} = \mu$.

2. Using that $(\mathrm{Cov}(X_i, X_j))_{i,j}$ is a Toeplitz matrix we have:

$$\mathrm{Var}(\widehat{\mu}_n) = \frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i, X_j) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} c(i-j)$$
$$= \sum_{k=-(n-1)}^{n-1}\frac{n-|k|}{n^2}c(k) = \frac{1}{n}\sum_{k=-n+1}^{n-1}\left(1 - \frac{|k|}{n}\right)c(k).$$

3. To prove that $\widehat{\mu}_n$ is a consistent estimator of μ it is enough to prove that $\mathbb{E}[(\widehat{\mu}_n - \mu)^2] \to 0$ since convergence in L_2 implies convergence in probability. We have

$$\begin{aligned}\mathbb{E}[(\widehat{\mu}_n - \mu)^2] = \operatorname{Var}(\widehat{\mu}_n) &= \frac{1}{n}\sum_{k=-n+1}^{n-1}\Big(1 - \frac{|k|}{n}\Big)c(k) \le \frac{1}{n}\sum_{k=-n+1}^{n-1}|c(k)| \\ &= \frac{2n-1}{n}\Bigg(\frac{1}{2n-1}\sum_{k=-(n-1)}^{n-1}|c(k)|\Bigg) \to 0\end{aligned}$$

by Cesàro mean theorem using that $c(k) \to 0$.

4. Using the Dominated Convergence Theorem (DCT) we have

$$\lim_{n\to\infty} n\operatorname{Var}(\widehat{\mu}_n) = \lim_{n\to\infty}\sum_{k=-n+1}^{n-1}\Big(1 - \frac{|k|}{n}\Big)c(k) = \sum_{k\in\mathbb{Z}} c(k).$$

□

Remark 2. *Let X be a weakly stationary time series with mean μ and autocovariance function c. Thanks to Theorem 1 we know that if $\sum_{k\in\mathbb{Z}}|c(k)| < \infty$ then $\sqrt{n}(\widehat{\mu}_n - \mu)$ is bounded in L_2 and hence in probability. We can thus conclude that the rate of convergence of $\widehat{\mu}_n$ to μ is $\sqrt{n}$.*

Remark 3. *Under suitable conditions, the sample mean $\widehat{\mu}_n$ of a weakly stationary process X satisfies:*

$$\sqrt{n}(\widehat{\mu}_n - \mu) \to \mathcal{N}\Big(0, \sum_{k\in\mathbb{Z}} c(k)\Big). \tag{1}$$

Provided that one is able to estimate the asymptotic variance $\sum_{k\in\mathbb{Z}} c(k)$, (1) allows to give confidence intervals. If in addition X is Gaussian, then

$$\mathscr{L}\big(\sqrt{n}(\widehat{\mu}_n - \mu)\big) = \mathcal{N}\Bigg(0, \sum_{|k|<n}\Big(1 - \frac{|k|}{n}\Big)c(k)\Bigg), \quad \forall n.$$

Properties of $c(k)$ and $\widehat{c}(k)$

Lemma 1. *The autocovariance function $c : \mathbb{Z} \to \mathbb{R}$ of a weakly stationary process $X = (X_t)_{t\in\mathbb{Z}}$ satisfies*

1. *c is symmetric: $c(k) = c(-k)$ for all $k \in \mathbb{Z}$.*
2. *$c(0) \ge 0$ and $|c(k)| \le c(0)$.*
3. *c is* positive semi-definite:

$$\forall m \in \mathbb{N},\ a_1, \dots, a_m \in \mathbb{R} : \quad \sum_{i,j=1}^{m} a_i a_j c(i-j) \ge 0.$$

Proof. 1. Since X is weakly stationary $\mathrm{Cov}(X_{s+k}, X_s) = \mathrm{Cov}(X_r, X_{r-k})$. Therefore,

$$c(k) = \mathrm{Cov}(X_{s+k}, X_s) = \mathrm{Cov}(X_r, X_{r-k}) = c(-k).$$

2. It is enough to observe that

$$c(0) = \mathrm{Cov}(X_s, X_s) = \mathrm{Var}(X_s) \geq 0, \quad \forall s \in \mathbb{Z}$$

and that, using Cauchy-Schwarz,

$$c(k)^2 = \mathrm{Cov}(X_k, X_0)^2 = \big(\mathbb{E}[(X_k - \mu)(X_0 - \mu)]\big)^2 \leq \mathrm{Var}(X_k)\mathrm{Var}(X_0) = c(0)^2.$$

3. We have

$$\begin{aligned}
\sum_{i,j=1}^{m} a_i a_j c(i-j) &= \sum_{i,j=1}^{m} a_i a_j \mathrm{Cov}(X_i, X_j) = \sum_{i,j=1}^{m} a_i a_j \big(\mathbb{E}[X_i X_j] - \mu^2\big) \\
&= \sum_{i,j=1}^{m} a_i a_j \mathbb{E}[X_i X_j] - \Big(\sum_{i=1}^{m} a_i \mu\Big)^2 \\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^{m} a_i X_i\Big)^2\Big] - \Big(\sum_{i=1}^{m} a_i \mathbb{E}[X_i]\Big)^2 = \mathrm{Var}\Big(\sum_{i=1}^{m} a_i X_i\Big) \geq 0.
\end{aligned}$$

□

Lemma 2. *The empirical autocovariance matrix $\widehat{C}_n$ is positive semi-definite:*

$$\forall n \in \mathbb{N},\ \forall a_1, \ldots, a_n \in \mathbb{R}: \quad \sum_{i,j=1}^{n} a_i a_j \widehat{c}(i-j) \geq 0.$$

Proof. Set $Y_i = (X_i - \widehat{\mu}_n)\mathbf{1}_{(1 \leq i \leq n)}$. Then,

$$\begin{aligned}
\sum_{i,j=1}^{n} a_i a_j \widehat{c}(i-j) &= \frac{1}{n} \sum_{i,j=1}^{n} a_i a_j \sum_{l=1}^{n-|i-j|} (X_l - \widehat{\mu}_n)(X_{l+|i-j|} - \widehat{\mu}_n) \\
&= \frac{1}{n} \sum_{l \in \mathbb{Z}} \sum_{i,j=1}^{n} a_i a_j Y_l Y_{l+|i-j|} = \frac{1}{n} \sum_{l' \in \mathbb{Z}} \sum_{i,j=1}^{n} a_i a_j Y_{l'-i} Y_{l'-j} \\
&= \frac{1}{n} \sum_{l' \in \mathbb{Z}} \Big(\sum_{i=1}^{n} a_i Y_{l'-i}\Big)^2 \geq 0.
\end{aligned}$$

□

Remark 4. *In order to have that $\widehat{C}_n$ is positive semi-definite it is essential that the prefactor before the sum in $\widehat{c}(k)$ does not depend on k.*