20236 Time Series Analysis

Lecture Notes

Stefano Graziosi, Università Bocconi

February 26, 2025

Contents

Introduction

Group Project

The project will be about climate, taking inspiration from the Bocconi summer school (with two famous experts in climate).
Data will be posted fully by week 2.

Part I

Descriptive techniques and forecasting algorithms

Lecture 1

Classical time series decomposition

1.1 What is a time series?

Reference

Chatfield, Ch 2, Sections 2.1-2.6. Subsection 2.5.2: only basic moving average. Section 2.7 will be useful later

A first answer: a sequence of observations taken over time.

$$y_{1:t} = (y_1, \dots, y_t) \tag{1.1}$$

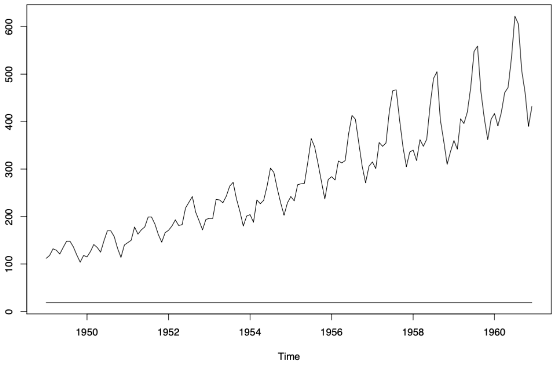


Figure 1.1: Time plot

The time points of measurement can be:

- Regularly spaced[1], with **no** missing values → **discrete time**
- Regularly spaced, with missing values
- Irregularly spaced
- High frequency → **continuous time**

Discrete time

$$y_t, t = 1, 2, 3, \dots, T \tag{1.2}$$

Continuous time

$$y_t, t \in (1, T) \tag{1.3}$$

[1] (monthly, yearly, weekly)

1.1.1 Different Types of Data

- **Types of data by nature**
  - Categorical
  - Discrete
  - Continuous
- **Types of data by dimensionality**
  - Univariate
  - Multivariate

1.1.2 Aims of time series analysis

1. **Exploratory/descriptive analysis**

   Describe trend, seasonality, cycle, etc. Often, in order to de-trend, de-seasonalize, etc. Multivariate/high dimensional time series: summarize; find common patterns, clusters, dimension reduction, etc...

2. **Forecasting**

   There are clever algorithms for forecasting (a first example: simple exponential smoothing). **Issue:** they do not describe *uncertainty* and *risk*, nor give a '*vision*' of the phenomenon (long tern forecasts).

3. **Explain**

   In fact, we may want to explain the phenomenon; understand the relationships among variables, or how one variable of interest Y depends on other variables. These will not be determinist relationships, but 'statistical '. The main tool is regression! But, here, the variables evolve over time.

4. **Modelling and inference**

   In this "explain" task, the info on the phenomenon is formalized through a **probabilistic (statistical) model**. The statistical model will have unknown parameters, that will have to be estimated: **inference**[2].

5. **Forecasting ⇒ Decisions under Risk**

   Still, the ultimate goal is **forecasting**: but now we have probability to express uncertainty and risk.

What is a time series? In this all, a time series is no longer just a sequence of measurements over time, but a stochastic process $Y_t, t \geq 1$.

In the course we will study some of the main probabilistic models for time series analysis, with focus on models for non-stationary processes.

[2] Statistical inference is based on probability.

Remark 1 **Aims of time series analysis**

1. Exploratory/descriptive analysis
2. Explain (modelling and inference)
3. Forecasting
4. Control → One wants to forecast y_t , but also control y_{t+1} through control variables x_t
5. Decisions under risk

1.2 Exploratory analysis

1.2.1 Classical time series decomposition

Classical time series decomposition describes a univariate time series $(Y_t)_{t\geq 1}$ as composed by structural components such as a trend, a seasonal component, a cycle,. . . , and an erratic component

$$Y_t = T_t + S_t + C_t + E_t \quad \text{additive} \tag{1.4}$$

$$Y_t = T_t * S_t * C_t * E_t \quad \text{multiplicative} \tag{1.5}$$

Notice that a *log* transforms a multiplicative decomposition into an additive one, for example **??** becomes:

$$\log Y_t = \log T_t + \log S_t + \log E_t \tag{1.6}$$

1.2.2 Additive seasonality and trend

If in `R` we write:

```r
plot(co2)
```

We get the following output:

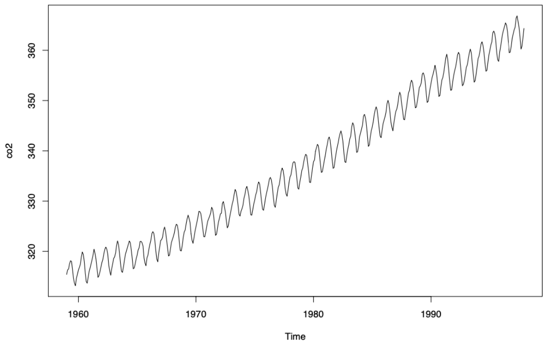


Figure 1.2: Evolution of CO2 concentration levels

If instead we write:

```r
plot(AirPassengers)
```

We get the following output;

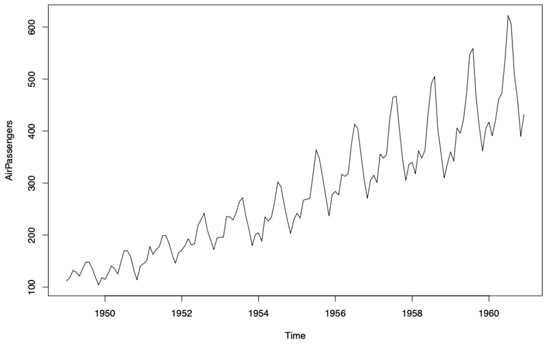


Figure 1.3: Evolution of air passenger travel

1.3 Fitting a trend

Consider a series that only shows a trend (no seasonal component)

$$Y_t = T_t + E_t$$

How can we fit the trend? Main basic tools:

1. **Fit a smooth function of** t

   For example, take a linear trend:

   $$y_t = \beta_0 + \beta_1 t + e_t, \quad t = 1, \dots, n$$

   Estimate β_0 and β_1 by least squares:

   $$\arg\min_{(\beta_0,\beta_1)} \sum_{t=1}^{n} (y_t - (\beta_0 + \beta_1 t))^2 \tag{1.7}$$

   One can use more complex functions, for example a quadratic trend

   $$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t \tag{1.8}$$

   or an exponential trend etc.

Remark 2 But, being too flexible makes it difficult to identify a trend and a cycle

2. **Fitting a moving average**

   For example, a moving average of order $k = 3$:

$$\hat{T}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}, \quad t = 2, 3, \ldots \tag{1.9}$$

   gives a 'smoothed version' ($\hat{T}_t$) of the observed series, interpreted as the trend; the higher k, the smoother the fitted trend.

   If k is even, e.g. $k = 4$, use

$$\hat{T}_t = \frac{0.5y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0.5y_{t+2}}{4} \tag{1.10}$$

   Having fitted the trend $\hat{T}_t$, let

$$y_t = \hat{T}_t + E_t, \quad t = 2, \ldots, n$$

   We can remove the trend, to obtain a *de-trended* time series

$$y_t^{\text{detrended}} = y_t - \hat{T}_t \tag{1.11}$$

1.4 Fitting a seasonal component

Consider a purely seasonal time series (no trend).

$$Y_t = S_t + E_t \tag{1.12}$$

How can we fit the seasonal component? Main basic tools:

1. **Seasonal factors**

   Suppose you have monthly data y_t, with mean $\bar{y}_t = 0$. A simple way to proceed is to introduce *seasonal factors* $\alpha_{Jan}, \alpha_{Feb}, \ldots, \alpha_{Dec}$ and describe

$$y_t = \alpha_{\text{month}(t)} + E_t \tag{1.13}$$

   where, if t corresponds to January, $\alpha_{\text{month}(t)} = \alpha_{Jan}$, and so on.

   For identifiability, we assume that the sum of the seasonal factors is zero. (Remember we are only considering additive decomposition.)

2. **Moving averages**

   For example, with monthly data, use MA(12)

$$y_t^{\text{deseasonalized}} = \frac{0.5y_{t-6} + y_{t-5} + \cdots + y_{t+5} + 0.5y_{t+6}}{12} \tag{1.14}$$

1.5 Fitting trend and seasonal components

Now consider

$$T_t + S_t + E_t \tag{1.15}$$

1. **Detrend, then fit the seasonal component**

   e.g., the R function DECOMPOSE uses MA to fit the trend $\hat{T}_t$ and, on the detrended series

   $$y_t^{\text{detrended}} = y_t - \hat{T}_t \tag{1.16}$$

   estimates the seasonal factors;

2. **Fit the seasonal component, then detrend**

   Fit the seasonal component, and on the deseasonalized series, fit the trend.

```r
plot (co2)
```

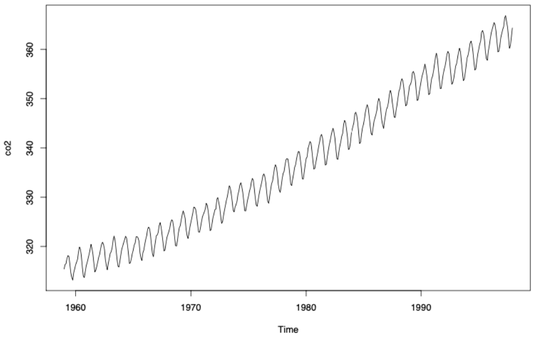


Figure 1.4: Atmospheric concentrations of CO2 are expressed in parts per million (ppm) and reported in the preliminary 1997 SIO manometric mole fraction scale. Monthly data from 1959 to 1997.

Example 1 In Laboratory 1 we'll introduce time series analysis with R and use R functions for time series decomposition.

Lecture 2

Forecasting algorithms: Simple exponential smoothing

References

Chatfield (2004), Ch.5, sect. 5.2.2; Petris, Petrone, Campagnoli (2009), Chapter 3, section 3.3.1

2.1 Forecasting algorithms

Forecasting (and consequent decisions) is often the ultimate goal of time series analysis.

Forecasting algorithms: predict y_{t+1} given $y_{1:t}$

We present **exponential smoothing**, a simple and cleaver algorithm, also called Holt-Winters algorithm, after Holt (1957) and Winters (1960).

2.2 Simple exponential smoothing

Consider a time series that shows no trend and no seasonality. At time t, we want to provide a forecast of y_{t+1} given data $(y_1, \ldots, y_t)$ (that I will denote by $y_{1:t}$ for short).

The idea is to take a weighted average of the past values $y_{1:t}$

$$\hat{y}_{t+1|t} = c_0 y_t + c_1 y_{t-1} + \ldots + c_{t-1} y_1 \tag{2.1}$$

with geometrically decreasing weights:

$$c_j = \alpha(1-\alpha)^j, 0 < \alpha < 1 \tag{2.2}$$

Remembering the geometric series:$\sum_{j=0}^{\infty} \lambda^j = \frac{1}{1-\lambda}$ if $0 < \lambda < 1$, we have

$$\sum_{j=0}^{\infty} \alpha(1-\alpha)^j = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j = \alpha \, \frac{1}{1-(1-\alpha)} = 1 \tag{2.3}$$

Thus, if t is large, the weights' sum in the average above is approximately one.

Thus, the *simple exponential smoothing algorithm* gives the point forecast:

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 \ldots + \alpha(1-\alpha)^{t-1} y_1 \tag{2.4}$$

2.3 Recursive form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + \alpha(1-\alpha)\ldots + \alpha(1-\alpha)^{t-2} y_1\right] \tag{2.5}$$

$$= \alpha y_t + (1-\alpha)\hat{y}_{t|t-1} \tag{2.6}$$

Remark 3

- if $\alpha \approx 1$, $\hat{y}_{t+1|t} \approx y_t$
- if $\alpha \approx 0$, $\hat{y}_{t+1|t} \approx constant$

Definition 1 The parameter α is called **smoothing parameter**

2.3.1 Error correction form

$$\hat{y}_{t+1|t} = \hat{y}_{t|t-1} + \alpha \underbrace{\left(y_t - \hat{y}_{t|t-1}\right)}_{\text{Forecast error}} \tag{2.7}$$

- at time $t-1$, forecast $\hat{y}_{t|t-1}$
- then observe y_t
- and correct the forecast, taking into account the forecast error

2.3.2 Choosing α

Algorithms have parameters that need to be "tuned" (on a training sample). Here, we have to tune the smoothing parameter α:

- consider a (fine) grid of values of α in $(0, 1)$
- for each α, use the algorithm to recursively compute the one-step-ahead forecasts, for $t = 1, \ldots, T$
- "Score" the forecasts and choose the value α that gives the best predictive performance.

2.3.3 Popular measures of predictive performance

$e_t = y_t - \hat{y}_{t|t-1}$ **forecast error**

$$MAE = \frac{\sum_{t=1}^{T} |e_t|}{T} \qquad \text{Mean Absolute Value} \tag{2.8}$$

$$MSE = \frac{\sum_{t=1}^{t} |e_t^2|}{T} \qquad \text{Mean Square Error} \tag{2.9}$$

$$MAPE = \frac{\sum_{t=1}^{T} |\frac{e_t}{y_t}|}{T} \qquad \text{Mean Absolute Percentage Value} \tag{2.10}$$

Note that the latter measure does not depend on the scale.

2.3.4 Extensions

Simple exponential smoothing does not envisage a trend nor a seasonal behaviour. The so-called *Holt-Winters* forecasting algorithm extends simple exponential smoothing to the case of trend and/or seasonality.
Here we briefly present these extensions; our aim is simply to understand how the R functions for exponential smoothing work.
Only simple exponential smoothing is part of the program for the written proof in the final exam; but you may want to use these extension in your data project.

Remark 4 We will give a probabilistic interpretation of these algorithms as Dynamic Linear Models.

2.4 Forecast function

We didn't give a definition of "trend". Let's here think of the trend as the "expected" future behaviour of the time series.

Definition 2 Define the forecast function

$$f_t : h \to f_t(h) \equiv \hat{y}_{t+h|t} \,, \qquad h = 1, 2, \ldots \tag{2.11}$$

In simple exponential smoothing, the forecast function at time t is defined as a constant function of h,

$$f_t(h) \equiv \hat{y}_{t+h|t} = \hat{y}_{t+1|t}, \quad h = 1, 2, \ldots \tag{2.12}$$

Which is a "flat" prediction.

2.4.1 Intuition: y_t only shows a "level"

The idea is that

$$y_t = level + \epsilon_t$$

and we are forecasting the level; thus a "flat" forecast function seems reasonable.

In fact (as we will explain rigorously through DLMs), the idea is more clever than that,

$$y_t = level_t + \epsilon_t$$

and we update the forecast of the level as

$$L_t = \alpha y_t + (1 - \alpha) L_{t-1}.$$

Denoting $\hat{y}_{t+1|t} = L_t$ (level forecast, made at time t), thus $\hat{y}_{t|t-1} = L_{t-1}$ (level forecast, made at time $t - 1$), we see that the rule above corresponds to the forecast rule we had

$$\hat{y}_{t+1|t} = \alpha \, y_t + (1 - \alpha) \hat{y}_{t|t-1}.$$

Extension to a trend

Let us now consider a time series that shows a (possibly time-evolving) trend; no seasonal component.

Forecast function:

$$f_t(h) = L_t + B_t\, h, \qquad h = 1, 2, \ldots,$$

where the forecasts for the level, L_t, and for the growth, B_t, are recursively updated with a similar rule

$$\begin{aligned} L_t &= \alpha y_t + (1-\alpha)(L_{t-1} + B_{t-1}) && (2.13)\\ B_t &= \beta(L_t - L_{t-1}) + (1-\beta)B_{t-1} && (2.14) \end{aligned}$$

where

- $L_{t-1} + B_{t-1} = \hat{y}_{t|t-1}$, so the first update is "the same" as in simple exponential smoothing;
- $(L_t - L_{t-1})$ growth estimate at time t (difference of the levels);
- $0B_{t-1}$ growth estimate at time $t-1$.

Extension to a seasonality

Consider, for example, monthly data

Seasonal factors: (think, e.g., of t=January)

$$S_t \approx S_{t+12} \approx S_{t+2*12} \approx \cdots$$

Similar idea for this forecast function:

$$\hat{y}_{t+h|t} = L_t + B_t\, h + S_{t+m+h} \qquad (2.15)$$

where

$$\begin{aligned} L_t &= \alpha(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + B_{t-1}) && (2.16)\\ B_t &= \beta(L_t - L_{t-1}) + (1-\beta)B_{t-1} && (2.17)\\ S_t &= \gamma(y_t - (L_{t-1} + B_{t-1})) + (1-\gamma)S_{t-12} && (2.18) \end{aligned}$$

For now, enough to remember that there are three smoothing parameters, α, β, γ, referring to level, growth, seasonality

Part II

Probabilistic Models for Time Series Analysis

Lecture 3

Introd. to Probabilistic Models for Time Series Analysis

Further readings:

Chatfield, Chapter 3, Sections 3.1–3.4, and references therein

3.1 Time series as a stochastic process

So far, we have been thinking of a time series as a sequence of measurements taken over time, $y_{1:t} = (y_1, y_2, \ldots, y_t)$. We presented simple exponential smoothing as a simple, but quite clever, forecasting algorithm that recursively provides a point forecast $\hat{y}_{t+1|t}$ given data $y_{1:t}$. But how reliable are the algorithm's forecasts? In fact, this is quite a timely question that we could also address to the current hype of AI algorithms.

As a predictive algorithm, simple exponential smoothing does not provide any measure of the uncertainty associated with the forecasts. But being able to quantify uncertainty, and the implied risk of decisions based on those forecasts, is in fact crucial.

There are many ways to express uncertainty. We could express vague claims such as "we are pretty sure," "I'm not that confident"; but we want to have a more structured methodology. The statistical methodology quantifies uncertainty around estimates and forecasts through *probability.*

This means, to start with, a different notion of "time series."

Definition 3 (Time series) We define a time series as a *stochastic process* $(Y_t)_{t \in T}$, with $Y_t \in \mathcal{Y}$, where the index t refers to time. The data (y_t) are then regarded as a realization of the stochastic process. (I will often use the short notation (Y_t) for $(Y_t)_{t \in T}$.)

3.1.1 What is a stochastic process?

You might have studied this notion in previous courses in Statistics or Probability. To our aims, let us recall that a stochastic process $(Y_t)_{t \in T}$ is a sequence of random vectors indexed by t, with t in a set T. If T is countable, for example $t = 1, 2, \ldots,$ we say that (Y_t) is a discrete-time stochastic process. Or we may have continuous-time stochastic processes.

Classic time series analysis refers to discrete-time processes. The time series is univariate, if $Y_t \in \mathcal{Y} \subseteq \mathbb{R}$, or multivariate, if $Y_t = (Y_{1,t}, \ldots, Y_{m,t})^T$ is a random vector in $\mathcal{Y} \subseteq \mathbb{R}^p$. Let us focus, in this introduction, on the univariate case. Then, for any t, Y_t is a random variable, and y_t denotes a realization of Y_t. See the beautiful Figure **??** for a simple illustration.

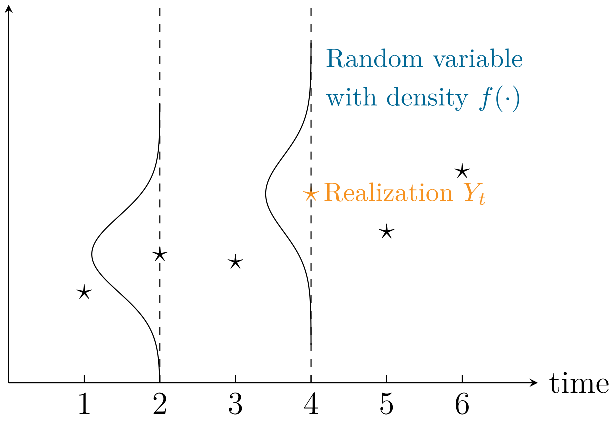


Figure 3.1: A time series is a discrete-time stochastic process. For each t, Y_t is a random variable.

Assuming a statistical model for a time series (Y_t) **means to assume the probability law of the stochastic process** (Y_t). We would write $(Y_t) \sim P$. But here we see a new notion: the probability law P is not assigned on a random variable, or on a random vector—which are familiar notions to us—but on an infinite-dimensional object, the sequence (Y_t). A stochastic process is indeed a new notion: it is a sequence of random variables indexed by t, but it involves more than just single random variables or random vectors.

Intuitively, we may think of proceeding by first assigning the one-dimensional marginal densities: $Y_t \sim f_t(\cdot)$, for all t. For example, suppose that the time series $(Y_t)_{t \geq 1}$ describes the evolution of the Gross Domestic Product (GDP) of a country over time. Before observing the data, the GDP at time t is unknown, and we model it as a random variable Y_t with density f_t.[1] Then, we would specify all the bivariate densities, $(Y_{t_1}, Y_{t_2}) \sim f_{t_1,t_2}(\cdot,\cdot)$ for any t_1, t_2 (the joint density of the GDP in years t_1 and t_2, say), and so on. We would thus assign the joint densities of vectors $(Y_{t_1}, \ldots, Y_{t_k})$ for any $k \geq 1$ and any choice of $t_1, \ldots, t_k$:

$$\begin{pmatrix} Y_{t_1} \\ \vdots \\ Y_{t_k} \end{pmatrix} \sim f_{t_1,\ldots,t_k}(\ldots) \tag{3.1}$$

and so on.

Is this enough to assign the probability law of the sequence (Y_t)? We are not sure yet: assigning the probability law of a stochastic process (Y_t) is something more, and new: we have to also assign the probability of events that involve the infinite-dimensional path of the process (e.g., the probability that $Y_1 < Y_2 < Y_3 < \cdots$, i.e. that the path is increasing). Dealing with infinite-dimensional events is something new.

[1] If Y_t is a continuous random variable, then, more formally, f_t is a probability density with respect to the Lebesgue measure; if Y_t is discrete, f_t is a probability mass function (which is still a probability density, but with respect to the counting measure). For brevity, we will use the term *density* in both cases.

Luckily, a fundamental result in Probability ensures that we can define the probability law of the process (Y_t) through the *finite-dimensional distributions*

$$f_{t_1,\ldots,t_k}, \quad \text{for all } k \geq 1 \text{ and any choices of } (t_1,\ldots,t_k),$$

as in **??**. It is proved that, if the family of densities $\{f_{t_1,\ldots,t_k}\}$ satisfies a consistency condition[2] then they uniquely define the probability law, say P, of the process (Y_t), so that $(Y_t) \sim P$; in other words, there exists a stochastic process (Y_t) such that $(Y_{t_1},\ldots,Y_{t_k}) \sim f_{t_1,\ldots,t_k}$ for any k and $(t_1,\ldots,t_k)$.

We do not enter into details about this fundamental result. The crucial point is that we can specify the probability law of the process (Y_t)—that is, the *statistical model* for the time series (Y_t)—by specifying the joint densities **??**.

Remark 5 This means that, when you define a statistical model for a time series, you have to specify all (and only) the assumptions that allow one to write the finite-dimensional distributions.

This is what you have been doing in your basic courses of Statistics. In the basic setting (random sampling), the Y_i are **independent**:

$$f_{t_1,\ldots,t_k}(y_1,\ldots,y_k) \;=\; \prod_{j=1}^{k} f_{t_j}(y_j),$$

or even **independent and identically distributed (i.i.d.)**:

$$f_{t_1,\ldots,t_k}(y_1,\ldots,y_k) \;=\; \prod_{j=1}^{k} f(y_j).$$

However, for time series, independence would be a trivial assumption: we want to introduce *temporal dependence*!

We will study some main notions of temporal dependence, and statistical models that express these forms of dependence. We will mainly use *parametric models*, which assume that the joint densities have an analytic expression indexed by a vector of parameters θ taking values in a parameter space Θ:

$$f_{t_1,\ldots,t_k}(y_1,\ldots,y_k;\theta), \quad \theta \in \Theta.$$

A problem will then be to *estimate* the parameter θ; note that inference on θ is in fact, in most applications, an intermediate step for making *forecasts* based on the model and expressing the uncertainty about those forecasts through probability.

[2] Informally, Kolmogorov consistency conditions mean that lower-dimensional densities correspond to the densities obtained by marginalizing the joint densities: for example, $\int f_{t_1,t_2}(y_1,y_2)\,dy_2 \;=\; f_{t_1}(y_1)$,and so on for any choice of $(t_1,\ldots,t_k)$ and $k \geq 1$.

Example 2 (White noise) A process $(Y_t)_{t\geq 1}$ such that the Y_t are i.i.d. with common finite mean μ and variance σ^2 is called a *white noise process*. Assuming independence, a white noise is clearly uninteresting as a model for a time series, but it may be an important building block for more complex models with temporal dependence.
With the additional assumption that the Y_t have a common, Gaussian density

$$Y_t^{\text{i.i.d.}} \sim \mathcal{N}(\mu, \sigma^2),$$

the process (Y_t) is called a *Gaussian white noise*. A simulated sample $y_{1:T}$ from a Gaussian white noise with

$$Y_t^{\text{i.i.d.}} \sim \mathcal{N}(0, 1)$$

is plotted in Figure **??**.

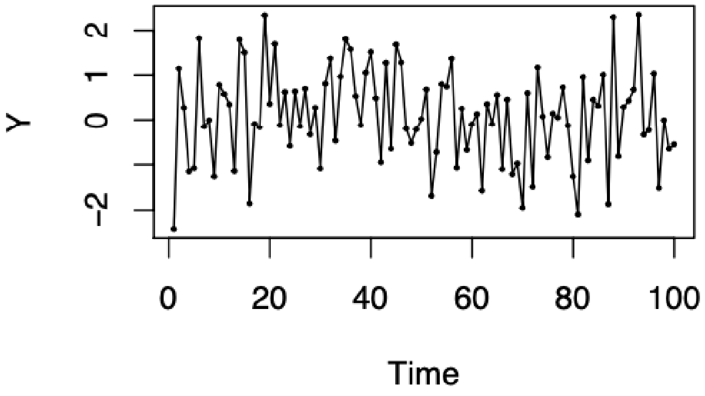


Figure 3.2: A simulated path of a Gaussian white noise process, with $Y \sim \mathcal{N}(0, 1)$

We will present more example in the next lectures, including Markov processes, ARMA models, Hidden Markov processes, dynamic linear models... Let's first see some important general notions.

3.2 Summaries

For a random variable X, we are used to providing summaries such as the mean (i.e. the expected value $E(X)$) and the variance $V(X)$. For a random vector (X, Y), we would provide measures of dependence, such as the covariance

$$\operatorname{Cov}(X,Y) \;=\; E\Big((X-E(X))(Y-E(Y))\Big), \tag{3.2}$$

and the correlation coefficient

$$\rho \;=\; \frac{\operatorname{Cov}(X,Y)}{\sqrt{V(X)\,V(Y)}}. \tag{3.3}$$

We want to provide similar summaries for time series. A main difference is that, for a time series (Y_t), the mean and the variance, as well as other relevant measures, will generally depend on time. More precisely, let us define the following notions.

Definition 4 (Mean function) The mean function of (Y_t) is the function

$$\mu:\, t \;\mapsto\; \mu(t) \;=\; F(Y_t), \quad t = 1, 2, \ldots.$$

In the Gaussian white noise example above, the mean function is constant: $\mu(t) = \mu$ for any t. In general, however, $F(Y_t)$ varies with t and, roughly speaking, the mean function expresses the expected path over time.

Definition 5 (Variance function) The variance function of (Y_t) is the function

$$\sigma^2:\, t \;\mapsto\; \sigma^2(t) \;=\; V(Y_t), \quad t = 1, 2, \ldots.$$

Definition 6 (Autocovariance function) The autocovariance function of (Y_t) is the function

$$\gamma:\, (t_1, t_2) \;\mapsto\; \gamma(t_1, t_2) \;=\; \operatorname{Cov}\big(Y_{t_1}, Y_{t_2}\big), \quad t_1, t_2 = 1, 2, \ldots,$$

with $\gamma(t,t) = \sigma^2(t)$.

Definition 7 (Autocorrelation function) The autocorrelation function of (Y_t) is the function

$$\rho:\, (t_1, t_2) \;\mapsto\; \rho\big(t_1, t_2\big) \;=\; \frac{\operatorname{Cov}\big(Y_{t_1}, Y_{t_2}\big)}{\sqrt{\sigma^2(t_1)\,\sigma^2(t_2)}}, \quad t_1, t_2 = 1, 2, \ldots.$$

The autocovariance function and the autocorrelation function inherit several properties from the properties of the covariance and the correlation coefficient for random vectors. See Chatfield, Chapter 3.

3.3 Stationarity

Many statistical models, such as popular ARMA processes, assume that the time series under study is *stationary*.

Definition 8 (Strict stationarity) A time series (Y_t) is *strictly stationary* if the joint distribution of

$$(Y_t, Y_{t+1}, \ldots, Y_{t+k})$$

does not depend on t, for any $k = 1, 2, \ldots$.

This implies that

$$(Y_{t_1}, \ldots, Y_{t_k}) \overset{d}{=} (Y_{t_1+h}, \ldots, Y_{t_k+h}),$$

for any choice of times $(t_1, \ldots, t_k)$ and any lag h, where $\overset{d}{=}$ means equality in distribution. In particular, the marginal distribution of Y_t does not depend on t, i.e.,

$$Y_1 \overset{d}{=} Y_2 \overset{d}{=} \ldots \overset{d}{=} Y_t \overset{d}{=} \ldots.$$

If $E(Y_t)$ and $E(Y_t Y_{t+h})$ exist and are finite, this implies that the mean function and variance function are constant over time, namely

$$\mu(t) = \mu \quad \text{and} \quad \sigma^2(t) = \sigma^2.$$

Moreover, the bivariate distribution of (Y_t, Y_{t+h}) does not depend on t. thus,

$$(Y_1, Y_{1+h}) \overset{d}{=} (Y_2, Y_{2+h}) \overset{d}{=} \ldots \overset{d}{=} (Y_t, Y_{t+h}) \overset{d}{=} \ldots.$$

Consequently, the autocovariance $\gamma(t, t+h)$, which depends on the distribution of (Y_t, Y_{t+h}), does not depend on t but only on the lag h.

Definition 9 (Stationarity) A time series (Y_t) is *stationary* (or, more precisely, *second-order stationary*) if $E(Y_t)$ and $E(Y_tY_{t+h})$ exist and do not depend on t, for every lag $h = 1, 2, \ldots$.

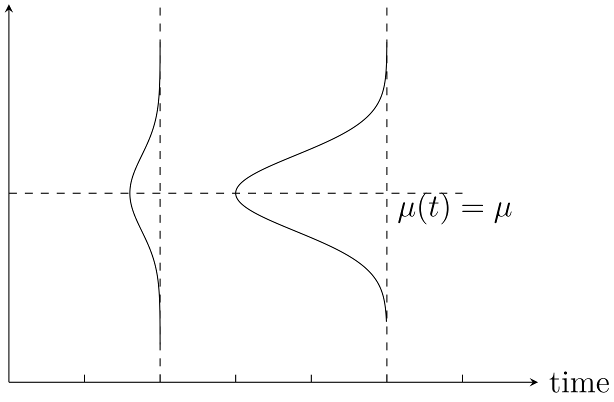


Figure 3.3: Stationarity does not imply strict stationarity. For example, the Y_t may have the same mean, but the marginal distributions (of Y_1 and of Y_t, in the plot) are different.

In other words, this means that

$$\begin{aligned} \mu(t) &= \mu && \text{the mean function is constant,} && (3.4)\\ \sigma^2(t) &= \sigma^2 && \text{the variance function is constant,} && (3.5)\\ \gamma(t,t+h) &= \tilde{\gamma}(h) && \text{the autocovariance function only depends on the lag } h. && (3.6) \end{aligned}$$

Clearly, a strictly stationary time series with finite second moments is also stationary. However, the reverse is not true. See another beautiful hand-made picture in Figure 3.

The two notions are equivalent for a *Gaussian* time series, that is, for any $(t_1,\dots,t_k)$, the joint density is Gaussian

$$\begin{bmatrix} Y_{t_1}\\ \vdots\\ Y_{t_k}\end{bmatrix} \sim \mathcal{N}_k\left(\begin{bmatrix}\mu(t_1)\\ \vdots\\ \mu(t_k)\end{bmatrix}, \begin{bmatrix} \sigma^2(t_1) & \gamma(t_1,t_2) & \cdots & \gamma(t_1,t_k)\\ \gamma(t_2,t_1) & \sigma^2(t_2) & \cdots & \gamma(t_2,t_k)\\ \vdots & \vdots & \ddots & \vdots\\ \gamma(t_k,t_1) & \gamma(t_k,t_2) & \cdots & \sigma^2(t_k)\end{bmatrix}\right)$$

You can find an overview of main properties of multivariate Gaussian distributions in the Appendix of the textbook Petrone, Petris, Campagnoli, *Dynamic Linear Models*, Springer 2009.

Example 3 (Gaussian white noise) If $Y_t^{\text{i.i.d.}} \sim \mathcal{N}(\mu,\sigma^2)$, then

$$\mu(t)=\mu, \quad \gamma(t,t)=\sigma^2, \quad \gamma(t,t')=0 \quad (t\neq t').$$

In the next lecture, we will introduce random walks and Markov processes, that will be very important tools for our developments.

Lecture 4

Random walks

References

lecture notes on BBoard and Feller, "*Introduciton to Probability Theory*" Chapter 14 (page 360 of the PDF in Desktop → Bocconi → ESS Y1S2 → 20236 Books)

From the previous lecture

In statistical time series analysis, we regard a time series as a stochastic process (i.e. a sequence of random variables or random vectors indexed by physical time $(Y_t)_{t\geq 1}$).
Statistical model: $(Y_t)_{t\geq 1} \sim \mathbb{P}$
To choose a statistical model, it is enough to specify the finite-dimensional distribution (i.e. the univariate, bivariate, ..., multivariate distribution) with some consistency conditions

4.1 Simple Random Walk

Statisticians usually have this *explain* step. Suppose we have this data:

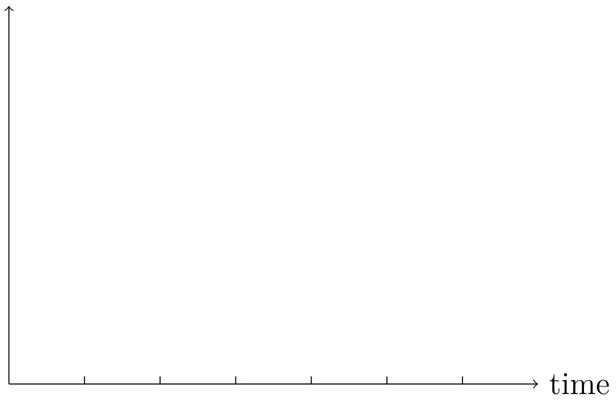


Figure 4.1: Graphical representation of an investment

Leo Breiman, *"The Two Cultures"*, Statistical Science, 2001

$$y_o > 0, \text{ at each step} \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases} \tag{4.1}$$

$$Z \sim \begin{cases} -1 & \text{with probability } 1-p \\ +1 & \text{with probability } p \end{cases}$$

Take as starting condition $Y_0 = y_0$, and define the successive steps as $Y_1 = y_0 + Z_1$ and $Y_2 = Y_1 + Z_2 = y_0 + Z_1 + Z_2$.

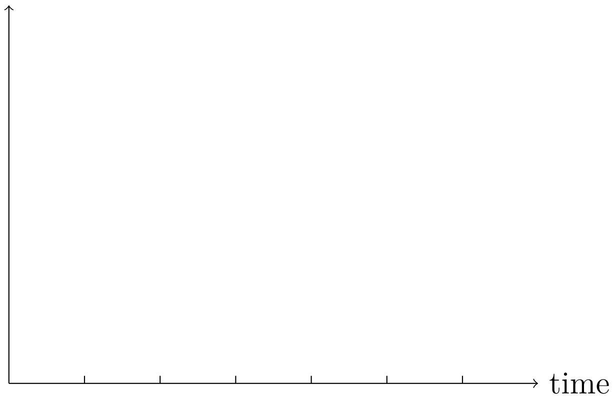


Figure 4.2: Graphical representation of an investment

Definition 10 (Simple random walk) We define (Y_t) as a simple random walk.

Remark 6 Y_t depends on the past only through the very last value.

Now let's consider the case where Z_t is not necessarily binary as in the previous case, but it is instead described as follows:

Definition 11 (General random walk) Take $(Y_t)_{t\geq 1}$ starting at y_0. Now define:

$$Z_t \sim \text{with } E(Z_t) = \mu, \quad V(Z_t) = \sigma^2, \quad \text{usually gaussian} \tag{4.2}$$

We get to the law of motion:

$$Y_t = Y_{t-1} + Z_t = y_0 + \sum_{i=1}^{t} Z_i \tag{4.3}$$

Proposition 1 Is (Y_t) stationary? In order to verify this, we need to check whether

i. $\mu(t) = \mu \quad \Rightarrow \quad \mu(t) = E(Y_t) = E(\underbrace{y_0}_{y_0=0} + \sum_{i=1}^t \overbrace{Z_i}^{E(Z_i)=0}) = \sum_{i=1}^t E(Z_i) = 0$

ii. $\sigma^2(t) = \sigma^2 \quad \Rightarrow$

iii. It is superfluous to check for the 3rd condition

Example 4 Assume a model that predicts a price based on the following law of motion:[a]

$$\underset{\log(\text{price})=\log P_t}{Y_t} = Y_{t-1} + Z_t$$

Now define the returns as

$$\text{returns} : \frac{P_t}{P_{t-1}} \quad \text{log returns} : Y_t - Y_{t-1} = Z_t$$

This shows that returns are i.i.d. distributed in a random fashion, i.e. $(Z_t)_{t\geq 1} \sim \ldots$ stationary

[a] A classical assumption for perfect financial markets

Example 5 (Random walk with noise)

$$\begin{cases} \overset{\text{Ideal log price}_t}{X_t} = X_{t-1} + Z_t \quad \text{latent noise} \\ Y_t = X_t + \varepsilon_t, \quad \varepsilon \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2) \end{cases} \tag{4.4}$$

Lecture 5

Markov processes

References

lecture notes on BBoard

5.1 Markov chains

First of all, some notation:

$$\underset{Y_t\in\{1,\cdots,k\}}{(Y_t)}_{t\geq 1} \sim \mathbb{P}$$

We also ought to define:

$$\text{(i.i.d.)} \Rightarrow \mathbb{P}(y_1,\dots,y_n) = \mathbb{P}(Y_1 = y_1 \cdots Y_n = y_n) = \mathbb{P}(y_1)\cdot\dots\cdot\mathbb{P}(y_n) = \prod_{i=1}^{n} p(y_i) \tag{5.1}$$

$$\text{(general case)} \Rightarrow \mathbb{P}(y_1,\dots,y_n) = \dots = \mathbb{P}(y_1)\cdot\mathbb{P}(y_2|y_1)\cdot\dots\cdot\mathbb{P}(y_n|y_{n-1}) = \prod_{i=1}^{n} p(y_i|y_{i-1}) \tag{5.2}$$

$$\text{(Markov)} \Rightarrow \mathbb{P}(y_1,\dots,y_n) \tag{5.3}$$

Definition 12 (Markov chain)

From the definition it follows that a finite dimensional distribution can be written as in here (see eq.5)
To specify the probability law $\mathbb{P}$ of a Markov chain (Y_t), $Y_t \in \{1,2,\dots,k\}$ it is enough to give:

i. **Initial probabilities**:

$$p_0$$

ii. **Transition probabilities**:

Lecture 6

Inference for Markov chains

References

lecture notes on BBoard

Lecture 7

Stationary time series: ARMA models I

References

A basic reference is Chatfield (2004), Ch.3 (pp 39-49), Ch. 4, Ch 5, section 5.2.4

Lecture 8

Stationary time series: ARMA models II

8.1 Quick overview of ARMA models

Classic models presented in a conurs in time series analysis deal with **stationary processes**.

Autoregressive and

8.1.1 For Stationary Time Series

$$(T_t)_{t\geq 1} \text{ is stationary}$$

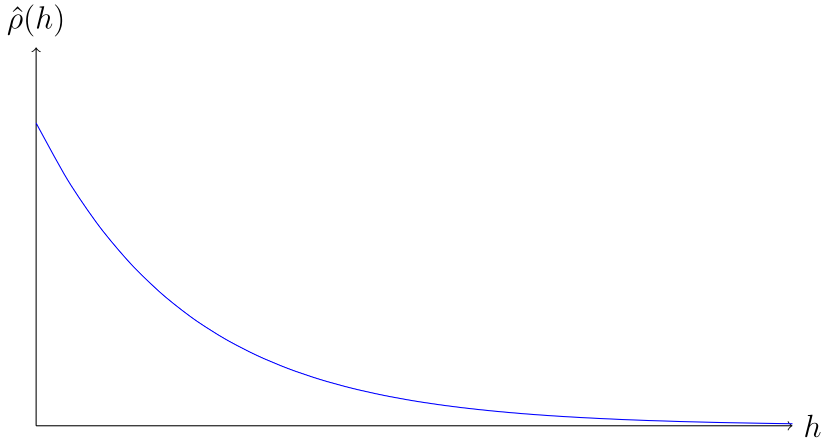
Stationarity is defined and characterized by:

- **Mean** $\mu(t) = \mathbf{E}(\mathrm{T}_t) = \mu$
- **Variance** $\sigma(t) = \mathrm{Var}(T_t) = \sigma^2$
- **Autocovariance** $\gamma(t, t+h) = Cov(Y_t, Y_{t+h}) = \tilde{\gamma}(h)$
- **Autocorrelation** $\rho(t, t+h) = \frac{\gamma(t,t+h)}{\sigma(t)\sigma(t+h)} = \tilde{\rho}(h)$

Both autocovariance and autocorrelation are functions of the lag h.

If the series $(Y_t)_{t\geq 1}$ is stationary, then the mean, variance, autocovariance and autocorrelation are constant over time and they can be calculated. In fact, we can see that:

- Sample mean: $\hat{\mu} = \frac{\sum_{t=}^{T} Y_t}{T} = \bar{Y}$
- Sample variance: $\hat{\sigma}^2 = \frac{\sum_{t=1}^{T}(Y_t-\bar{Y})^2}{T-1}1$
- Sample autocovariance at lag h: $\hat{\gamma}(h) = \frac{\sum_{t=1}^{T-h}(Y_t-\bar{Y})(Y_{t+h}-\bar{Y})}{T-h}$
- Sample autocorrelation at lag h: $\hat{\rho}(h) = \frac{\hat{Cov}(Y_t,Y_{t+h})}{\sqrt{\hat{Var}(Y_t)\hat{Var}(Y_{t+h})}} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$



Plot the estimated **autocorrelogram** to see if the series is stationary.

The estimated autocorrelation function should be approximately close to the autocorrelation sample for large sample size.

Proposition 2 Informally speaking, one can expect that:
Correlogram ≈ True autocorrelation function

8.1.2 For Non-Stationary Time Series

In this case, we cannot use the same properties as the moments we used before now depend on time (i.e.$\mu(t) = E(Y_t) = \mu_t$).

Replicates

In some applications (e.g. a financial market with multiple assets prices as time series) we have "replicates" (a random sample) of time series.

INSERISCI GRAFICO

And we can use

$$\bar{Y_t} = \frac{\sum_{i=1}^{n} Y_{i,t}}{n} \tag{8.1}$$

to estimate the mean of the time series $\mu_t = E(Y_t)$.

However, in many applications we do not have access to replicates.

As we will see, these models introduce a temporal dependence for the μ_t, σ_t and γ_t, so that observations at times different from t can still give "indirect" information on mu_t, and can thus be used for estimating it.

8.2 ARMA(p,q) Models for Stationary Time Series

INSERT THE GRAPH OF THE REALISATION OF A TIME SERIES.

Suppose we have time series data that can be modeled as the realisation of a stationary time series $(Y_t)_{t\geq 1}$.

[1] By dividing by $T-1$ instead of T, we actually get the "corrected" sample variance.

But what model should we use? **ARMA(p,q)** models are a good starting point. One of the reasons why they are attractive is because the order p, q can be estimated from the data, and it can then be understood from the ACF.

8.2.1 AR(1) autoregressive process of order 1

An AR(1) is presented as the (stationary and causal) solution of a finite difference stochastic equation, with $t \in (-\infty, +\infty)$:

$$Y_t = \alpha T_{t-1} + \varepsilon_t \tag{8.2}$$

where ε_t is a white noise process with mean 0 and variance σ^2, i.e. $\varepsilon_t \sim N(0, \sigma^2)$.

A solution of the equation is a proces Y_t that satisfies the equation for all t. In general, the solution is not unique, but we can impose some conditions to make it unique. Namely, if we restrict to stationary solutions, then either there exist no solutions or there exists a unique solution, depending on the value of α.

Example 6 ($\alpha = 1$) To define the AR(1) process, we need to specify:

$$\alpha = 1, \quad Y_t = T_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

In this case, the model just specified is a random walk, and does not hence admit a stationary solution.

Example 7 ($|\alpha| < 1$) If instead we consider a model specified as:

$$\alpha \in (-1, 1), \quad Y_t = \alpha Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

then the model admits a stationary solution, and it is:

$$Y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \quad \text{which is an MA}(\infty) \text{ process}$$

Remark 7 Please note that the above solution is finite if and only if the series converges, which is only possible if $\alpha \in (-1, 1)$.

GUARDA NOTE SUL GRUPPO: Lemma 1.29 AR(p)

Let the absolute value of α be strictly less than 1, then there exists a stationary solution given by $Y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$.

$$\forall \alpha \in (-1, 1) \quad \exists! Y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \tag{8.3}$$

Let us check that this solution is stationary:

- **Mean**:

$$\mu(t) = \mathbf{E}(Y_t) = \mathbf{E}\left(\sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \alpha^j \mathbf{E}(\varepsilon_{t-j}) = 0$$

- **Variance**:

$$\sigma^2(t) = \text{Var}(Y_t) = \text{Var}\left(\sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \alpha^{2j} \text{Var}(\varepsilon_{t-j}) = \sigma^2 \sum_{j=0}^{\infty} \alpha^{2j} = \frac{\sigma^2}{1-\alpha^2} = \sigma^2 \cdot \frac{1}{1-\alpha^2} = \sigma_Y^2$$

- **Autocovariance**:

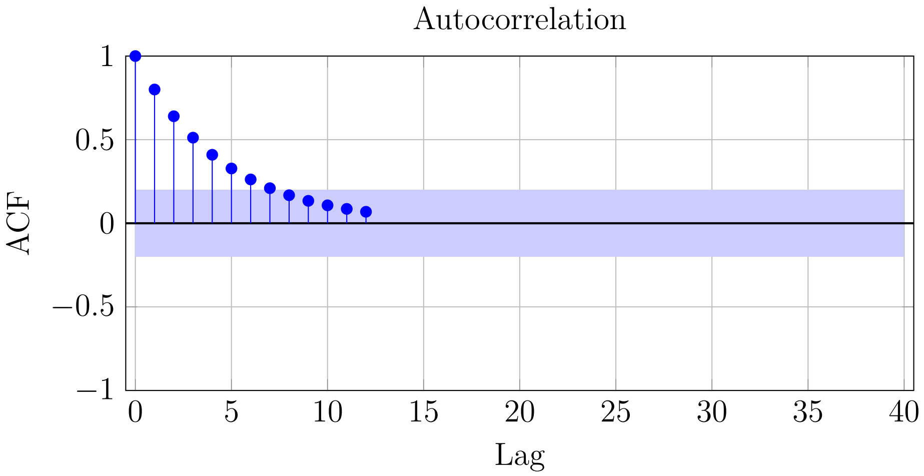
$$\gamma(t,t+h) = (Y_t, Y_{t+h}) = E(Y_t, Y_{t+h}) - \underbrace{E(Y_t)}_{0}\overbrace{E(Y_{t+h})}^{0} = \mathbf{E}\left(\sum_{j=0}^{\infty}\alpha^j\varepsilon_{t-j}, \sum_{j=0}^{\infty}\alpha^j\varepsilon_{t+h-j}\right) = E(\alpha^0\varepsilon_t + \alpha^1\varepsilon$$

Given that all the moments are constant over time (the result on the right do not depend on t), we can say that the solution is stationary.

Example 8 (Do it yourself!) Plot the *acf* of an AR(1) process with:

1. $\alpha = 0.8$
2. $\alpha = -0.8$
3. $\alpha = 2$
4. $\alpha = 0.2$

The *acf* is $\rho(g) = \frac{\gamma(h)}{\gamma(0)} = \frac{\alpha^h \cdot \sigma_Y^2}{\sigma_Y^2} = \alpha^h$. Thus, the functions can be plotted as



Oftentimes, it is not sufficient to look at the autocorrelation function: we need to look at the *partial autocorrelation function.*

Partial Autocorrelation Function

Definition 13 (Partial Autocorrelation Function) For gaussian processes (and ARMA models are indeed Gaussian):

$$\phi : h \mapsto \phi(h) = Corr(Y_t, Y_{t+h}|Y_{t+1}, \dots, Y_{t+h-1}) \tag{8.4}$$

Example 9 For an autoregressive process of order 1, AR(1), we have:

$$\phi(0) = 1 \tag{8.5}$$
$$\phi(1) = Corr(Y_t, Y_{t-1}) \neq 0 \tag{8.6}$$
$$\phi(2) = Corr(Y_t, Y_{t-2}|Y_{t-1}) = 0 \tag{8.7}$$
$$\vdots \tag{8.8}$$
$$\phi(h) = 0 \quad \forall h > 1 \tag{8.9}$$

8.2.2 AR(p) autoregressive process of order p

Partial Autocorrelation Function

Definition 14 (Partial Autocorrelation Function)

8.2.3 MA(q) moving average process of order q

Example 10

8.3 ARMA(p,q) Process

8.3.1 Fitting and ARMA model: the Box-Jenkins approach

Step 0: is this series stationary

Step 1: model specification

Step 2: model estimation

Step 3: model diagnostics

Part III

Inference for Non-Stationary Time Series: State-Space Models

Lecture 9

Hidden Markov Models (HMMs)

References

lecture notes on BB

LAB 2

Part IV

State-Space Models (SSMs) for Time Series Analysis

Lecture 10

State-space models: Definition and main properties

Definition and main properties. Examples: HMMs as state-space models (This is done in class, look at your notes) Dynamic Linear Models (DLMs) Examples of non-linear, non-Gaussian state-space models (stochastic volatility models).

References

Petris, Petrone & Campagnoli (2009), Chapter 2.

Lecture 11

Applying SSMs

Applying DLMs with known parameters: Model specification. Filtering, smoothing and forecasting. –The simplest non-trivial example: random walk plus noise. (a) Static case; it requires notions of Bayesian inference.

References

Petris, Petrone & Campagnoli (2009), Chapter 2.

Lecture 12

Basic notions of Bayesian inference I

References

Petris, Petrone & Campagnoli (2009), Chapter 1. The Gaussian example, with known variance, is on page 7

Lecture 13

Basic notions of Bayesian inference II

Lecture 14

Kalman filter for DLMs with known parameters I

References

Petris, Petrone & Campagnoli (2009), Chapter 2. Behavior for large t (random walk plus noise : reference: textbook on DLMs, Ch 3

Back to the basic example: random walk plus noise. (b) Dynamic case: Filtering (with details on computations)

Kalmam filter for general DLMs

Example: linear growth model.

Lecture 15

Kalman filter for DLMs with known parameters II

LAB 3: DLMs with R

Lecture 16

Smoothing, forecasting, model checking I

References

Petris, Petrone & Campagnoli (2009), Chapter 2.

Smoothing in DLMs: Kalman smoother
Forecasting.
The innovation process. Model checking.

Lecture 17

Smoothing, forecasting, model checking II

Lecture 18

DLMs with unknown parameters: MLE

References

Petris, Petrone & Campagnoli (2009), Chapter 4, section 4.1.

Likelihood function for DLMs. MLE and asymptotic standard errors.

LAB 4: MLE for DLMs with R

Lecture 19

Modeling time series with DLMs: examples I

References

Petris, Petrone & Campagnoli (2009), Chapter 3; only the general idea for local seasonal factors DLMs; no Fourier-based DLMs.

Combining DLMs components
Basic structural models:
DLMs for trend.
DLMs for seasonality (no details)

Lecture 20

Modeling time series with DLMs: examples II

References

Petris, Petrone & Campagnoli (2009), Chapter 3.

Lecture 21

DLMs for spatio-temporal data I

References

No infos

Lecture 22

DLMs for spatio-temporal data II

Lecture 23

Bayesian inference for DLMs with unknown parameters

References

Petris, Petrone & Campagnoli (2009), Chapter 4.

Basic notions.
Markov Chain Monte Carlo (MCMC; only aims and general idea)

References for MCMC

Petris, Petrone & Campagnoli (2009), Chapter 4.