

Laurea Magistrale in Ingegneria Informatica

Statistics for High Dimensional Data

Module on time series (3 cfu)

aa 2023-2024
Alessandro Fassò

1 / 1

Part I - Stat & R Basics

Content

1. Introduction
2. the AGRIMONIA dataset
3. The R environment
4. Review of statistical inference
 - ▶ The concept of inference
 - ▶ Estimation
 - ▶ Confidence intervals
 - ▶ Test of hypothesis

2 / 1

Introduction

Bibliography

- ▶ Agresti A. & Kateri M. (2022) Foundations of Statistics for Data Scientists: with R and Python. CRC Press.
- ▶ Agresti A. & Kateri M. (2022) Statistica per Data Scientists: con R e Python. Egea.



3 / 1

2. the AGRIMONIA dataset

Data description



<https://doi.org/10.1038/s41597-023-02034-0>

Data source



<https://doi.org/10.5281/zenodo.7956006>

4 / 1

3. The R environment

Installing R

- ▶ Install the "R engine": [CRAN](#)
- ▶ Install the "R dashboard": [Rstudio](#)
- ▶ Install the "package manager": [Rtools4](#)
(while Rstudio is closed!!)

5 / 1

3. The R environment

Beginning with R

- ▶ Panel structure
- ▶ Projects & scripts
- ▶ Calculator
- ▶ Numeric variables and vectors
- ▶ Functions (built-in)
- ▶ Reading/writing data files (csv)
- ▶ Dataframes and tables
- ▶ Non numeric data types: logical, characters, dates, factors

R example with dataset `Agrimonia.Rdata`
available at [zenodo](#)

6 / 1

The R environment

R packages

- ▶ Built-in and optional packages
- ▶ [Rstudio packages](#)
- ▶ Installing an R package for the first time
- ▶ Opening an R package
- ▶ Using the package functions and data

7 / 1

R LAB

R environment & EDA

- ▶ (Start a project named `Stat.RProject`)
- ▶ Plot a function.
For example $y = \sqrt{x}$ in the interval $[0,10]$ and save to file
- ▶ Load the file `Agrimonia.Rdata`
- ▶ screening (many?) single variables
 - ▶ `mean,sd,...` of vectors and tables
 - ▶ handling missing data (NA)
 - ▶ `summary`
 - ▶ basic graphs

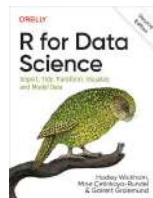
8 / 1

R LAB

Exploratory Data Analysis - part 1

Manipulating data using `tidyverse` and dates using `lubridate`. Exploring multivariate data using `ggplot2` and `psych` packages

- ▶ We navigate `tidyverse` & `ggplot2` capabilities using `Stat.Rproj`
- ▶ [ggplot2: Cheat Sheet](#)



R for Data Science

9 / 1

4. Review of statistical inference

The concept of inference

- ▶ Population
 - ▶ Finite populations
 - ▶ Infinite populations
 - ▶ Real and virtual populations
- ▶ Sample
- ▶ Inference
 - ▶ Point estimation
 - ▶ Interval estimation
 - ▶ Testing
- ▶ Uncertainty



10 / 1

4. Review of statistical inference

distributions and random numbers

Gaussian distribution

- ▶ Density (pdf): `dnorm(x, mean, sd)`
- ▶ Cumulative Distribution (cdf) `pnorm(x, mean, sd)`
- ▶ Quantiles: `qnorm(p, mean, sd)`
- ▶ Random numbers: `rnorm(n, mean, sd)`

Other distributions

- ▶ `unif, t, gamma, chisq, ... binom, pois, ...`

Random object generator

- ▶ `sample`: with/without replacement

11 / 1

4. Review of statistical inference

distributions and random numbers

Gaussian distribution

- ▶ Density (pdf): `dnorm(x, mean, sd)`
- ▶ Cumulative Distribution (cdf) `pnorm(x, mean, sd)`
- ▶ Quantiles: `qnorm(p, mean, sd)`
- ▶ Random numbers: `rnorm(n, mean, sd)`

Other distributions

- ▶ `unif, t, gamma, chisq, ... binom, pois, ...`

Random object generator

- ▶ `sample`: with/without replacement

11 / 1

4. Review of statistical inference

distributions and random numbers

Gaussian distribution

- ▶ Density (pdf): `dnorm(x, mean, sd)`
- ▶ Cumulative Distribution (cdf) `pnorm(x, mean, sd)`
- ▶ Quantiles: `qnorm(p, mean, sd)`
- ▶ Random numbers: `rnorm(n, mean, sd)`

Other distributions

- ▶ `unif, t, gamma, chisq, ... binom, pois, ...`

Random object generator

- ▶ `sample`: with/without replacement

11 / 1

4. Review of statistical inference

Estimation

- ▶ Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$
- ▶ An estimate of θ is a function of (y_1, \dots, y_n) denoted by $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$
- ▶ The estimation uncertainty is characterised by the distribution of $\hat{\theta}$, say $p(\hat{\theta})$
- ▶ Especially by
 - ▶ The mean squared error: $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$
 - ▶ The Variance: $\sigma^2(\hat{\theta})$
 - ▶ The Standard Error $SE(\hat{\theta}) = \hat{\sigma}(\hat{\theta})$

12 / 1

4. Review of statistical inference

Estimation

- ▶ Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$
- ▶ An estimate of θ is a function of (y_1, \dots, y_n) denoted by $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$
- ▶ The estimation uncertainty is characterised by the distribution of $\hat{\theta}$, say $p(\hat{\theta})$
- ▶ Especially by
 - ▶ The mean squared error: $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$
 - ▶ The Variance: $\sigma^2(\hat{\theta})$
 - ▶ The Standard Error $SE(\hat{\theta}) = \hat{\sigma}(\hat{\theta})$

12 / 1

4. Review of statistical inference

Estimation

- ▶ Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$
- ▶ An estimate of θ is a function of (y_1, \dots, y_n) denoted by $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$
- ▶ The estimation uncertainty is characterised by the distribution of $\hat{\theta}$, say $p(\hat{\theta})$
- ▶ Especially by
 - ▶ The mean squared error: $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$
 - ▶ The Variance: $\sigma^2(\hat{\theta})$
 - ▶ The Standard Error $SE(\hat{\theta}) = \hat{\sigma}(\hat{\theta})$

12 / 1

4. Review of statistical inference

Estimation

- ▶ Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$
- ▶ An estimate of θ is a function of (y_1, \dots, y_n) denoted by $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$
- ▶ The estimation uncertainty is characterised by the distribution of $\hat{\theta}$, say $p(\hat{\theta})$
- ▶ Especially by
 - ▶ The mean squared error: $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$
 - ▶ The Variance: $\sigma^2(\hat{\theta})$
 - ▶ The Standard Error $SE(\hat{\theta}) = \hat{\sigma}(\hat{\theta})$

12 / 1

4. Review of statistical inference

Maximum Likelihood Estimation

- ▶ Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$
- ▶ Let $L(\theta) = f_\theta(y_1, \dots, y_n)$ the likelihood of θ for the given data (y_1, \dots, y_n)
- ▶ The maximum likelihood estimate of θ is $\hat{\theta}_{ML} = \operatorname{argmax}_\theta(L(\theta))$
- ▶ Under regularity conditions, for large n MLE is optimal and
- ▶ $\operatorname{var}(\hat{\theta}_{ML}) \cong I(\theta)^{-1}$

13 / 1

Dataset Emissioni CO2

R example with `ferries_panel.Rdata`



Mannarini G., Salinas M.L., Carelli L., Fassò A. (2022) How COVID-19 Affected GHG Emissions of Ferries in Europe. Sustainability. 14(9):5287. 1-19.

14 / 1

4. Review of statistical inference

Maximum Likelihood Estimation

Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$

- ▶ Likelihood function: $L(\theta) = f_\theta(y_1, \dots, y_n)$
- ▶ MLE: $\hat{\theta} = \operatorname{argmax}_\theta(L(\theta))$
- ▶ For large n : $\hat{\theta} \cong N(\theta, I^{-1})$
- ▶ MLE is optimal for large n

R example with `ferries_panel.Rdata`

15 / 1

4. Review of statistical inference

Maximum Likelihood Estimation

Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$

- ▶ Likelihood function: $L(\theta) = f_\theta(y_1, \dots, y_n)$
- ▶ MLE: $\hat{\theta} = \operatorname{argmax}_\theta(L(\theta))$
- ▶ For large n : $\hat{\theta} \cong N(\theta, I^{-1})$
- ▶ MLE is optimal for large n

Reexample with `ferries_panel.Rdata`

15 / 1

Confidence intervals

CI for the mean

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$ with unknown mean μ and standard deviation σ

$$\mu \in \bar{y} \pm \Delta_\alpha$$

where $\Delta_\alpha = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$ and S is the sample standard deviation

Is an interval with $1 - \alpha$ CONFIDENCE LEVEL

Appropriate if

- ▶ f is Gaussian or n is large
- ▶ data y_1, \dots, y_n are independent
- ▶ if n is large then $t_{\alpha/2, n-1} \approx z_{\alpha/2}$

If (y_1, \dots, y_n) is autocorrelated Δ_α IS DIFFERENT!!

16 / 1

Confidence intervals

CI for the mean

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$ with unknown mean μ and standard deviation σ

$$\mu \in \bar{y} \pm \Delta_\alpha$$

where $\Delta_\alpha = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$ and S is the sample standard deviation

Is an interval with $1 - \alpha$ CONFIDENCE LEVEL

Appropriate if

- ▶ f is Gaussian or n is large
- ▶ data y_1, \dots, y_n are independent
- ▶ if n is large then $t_{\alpha/2, n-1} \approx z_{\alpha/2}$

If (y_1, \dots, y_n) is autocorrelated Δ_α IS DIFFERENT!!

16 / 1

Confidence intervals

CI for the mean

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$ with unknown mean μ and standard deviation σ

$$\mu \in \bar{y} \pm \Delta_\alpha$$

where $\Delta_\alpha = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$ and S is the sample standard deviation

Is an interval with $1 - \alpha$ CONFIDENCE LEVEL

Appropriate if

- ▶ f is Gaussian or n is large
- ▶ data y_1, \dots, y_n are independent
- ▶ if n is large then $t_{\alpha/2, n-1} \approx z_{\alpha/2}$

If (y_1, \dots, y_n) is autocorrelated Δ_α IS DIFFERENT!!

16 / 1

Confidence intervals

CI for the mean

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$ with unknown mean μ and standard deviation σ

$$\mu \in \bar{y} \pm \Delta_\alpha$$

where $\Delta_\alpha = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$ and S is the sample standard deviation

Is an interval with $1 - \alpha$ CONFIDENCE LEVEL

Appropriate if

- ▶ f is Gaussian or n is large
- ▶ data y_1, \dots, y_n are independent
- ▶ if n is large then $t_{\alpha/2, n-1} \approx z_{\alpha/2}$

If (y_1, \dots, y_n) is autocorrelated Δ_α IS DIFFERENT!!

16 / 1

Confidence intervals

General case (MLE)

Assume that

- ▶ Our data (y_1, \dots, y_n) is a sample from a distribution $f(y_1, \dots, y_n)$
- ▶ $\hat{\theta}$ is an estimate of θ
- ▶ For large n : $\hat{\theta} \cong N(\theta, I_n^{-1})$

An approximate CI for *theta* is given by

$$\theta \in \hat{\theta} \pm \Delta$$

where $\Delta = z_{\alpha/2} \frac{1}{\sqrt{I_n}}$

17 / 1

4. Review of statistical inference

hypothesis test - Parametric case

Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$

- ▶ Null Hypothesis: $H_0 : \theta = \theta_0$
- ▶ Test statistic: $T(y_1, \dots, y_n)$ with large values against H_0
- ▶ Pvalue: $p = P(T > T_{obs} | H_0)$
Small p: evidence "against" H_0
- ▶ Alternatives to H_0

18 / 1

4. Review of statistical inference

hypothesis test - Parametric case

Assume our data (y_1, \dots, y_n) is a sample from a distribution $f_\theta(y_1, \dots, y_n)$

- ▶ Null Hypothesis: $H_0 : \theta = \theta_0$
- ▶ Test statistic: $T(y_1, \dots, y_n)$ with large values against H_0
- ▶ Pvalue: $p = P(T > T_{obs} | H_0)$
Small p: evidence "against" H_0
- ▶ Alternatives to H_0

18 / 1

4. Review of statistical inference

Example: one-sample t-test

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$

Preliminary observations:

- ▶ $f \equiv N(\mu, \sigma^2)$ or n is large
- ▶ data y_1, \dots, y_n are independent

Test

- ▶ $H_0 : \mu = \mu_0$
- ▶ Alternatives:
 - ▶ $H_1 : \mu \neq \mu_0$
 - ▶ $H_1 : \mu > \mu_0$
 - ▶ $H_1 : \mu < \mu_0$
- ▶ Normality, sample size, nonparametric test

Rexample with `ferries_panel.Rdata`

19 / 1

4. Review of statistical inference

Example: one-sample t-test

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$

Preliminary observations:

- ▶ $f \equiv N(\mu, \sigma^2)$ or n is large
- ▶ data y_1, \dots, y_n are independent

Test

- ▶ $H_0 : \mu = \mu_0$
- ▶ Alternatives:
 - ▶ $H_1 : \mu \neq \mu_0$
 - ▶ $H_1 : \mu > \mu_0$
 - ▶ $H_1 : \mu < \mu_0$
- ▶ Normality, sample size, nonparametric test

Rexample with `ferries_panel.Rdata`

19 / 1

4. Review of statistical inference

Example: two-sample t-test

Assume we have a RANDOM sample (y_1, \dots, y_n) from the Gaussian distribution $N(\mu_Y, \sigma^2)$ and another sample (x_1, \dots, x_n) from $N(\mu_X, \sigma^2)$.

- ▶ Equality of X and Y distributions: $H_0 : \mu_X = \mu_Y$
- ▶ Alternatives:
 - ▶ $H_1 : \mu_X \neq \mu_Y$
 - ▶ $H_1 : \mu_X > \mu_Y$
 - ▶ $H_1 : \mu_X < \mu_Y$
- ▶ Same/different population variances σ^2
- ▶ Normality, sample size, nonparametric test

Rexample with `ferries_panel.Rdata`

20 / 1

4. Review of statistical inference

Example: two-sample t-test

Assume we have a RANDOM sample (y_1, \dots, y_n) from the Gaussian distribution $N(\mu_Y, \sigma^2)$ and another sample (x_1, \dots, x_n) from $N(\mu_X, \sigma^2)$.

- ▶ Equality of X and Y distributions: $H_0 : \mu_X = \mu_Y$
- ▶ Alternatives:
 - ▶ $H_1 : \mu_X \neq \mu_Y$
 - ▶ $H_1 : \mu_X > \mu_Y$
 - ▶ $H_1 : \mu_X < \mu_Y$
- ▶ Same/different population variances σ^2
- ▶ Normality, sample size, nonparametric test

Rexample with `ferries_panel.Rdata`

20 / 1

4. Review of statistical inference

hypothesis test - Functional hypotheses

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f(y)$

- ▶ Null Hypothesis: $H_0 : f = f_0$
- ▶ Special case: Normality tests
- ▶ $H_0 : f \equiv N(\mu, \sigma^2)$ for unknown μ, σ
 - ▶ Jarque Bera test on skewness and kurtosis
 - ▶ Lilliefors test on CDF
 - ▶ χ^2 test of normality on histogram

21 / 1

4. Review of statistical inference

Likelihood ratio tests

Assume our data (y_1, \dots, y_n) is a RANDOM sample from a distribution $f_\theta(y)$ with $\theta \in \Theta \subset R^k$

- ▶ Null Hypothesis: $H_0 : \theta \in \Theta_0$
- ▶ Alternative Hypothesis: $H_1 : \theta \in \Theta \setminus \Theta_0$
- ▶ Reject H_0 for large values of

$$\lambda = -2 \ln \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

- ▶ if $\Theta_0 = \theta_0$

$$\lambda = -2[\ln L(\theta_0) - \ln L(\hat{\theta}_{MLE})]$$

- ▶ Under H_0 , for large samples

$$\lambda \cong \chi_\nu^2$$

where $\nu = \dim(\Theta) - \dim(\Theta_0)$

22 / 1

R LAB

Common parametric tests

Reexample with `ferries_panel.Rdata`

- ▶ Consider year 2020 data
- ▶ Focus on One sided vs Two sided tests
- ▶ Test if "small ferries are less than 30%"
- ▶ Test if "total emissions per ship are above 31'000 kt"
- ▶ Test if "the emissions have a Gaussian distribution"
- ▶ Test if "region A and M have the same avg emissions" (unpaired)
- ▶ Test if "emissions in 2020 are lower than 2019" (paired)
- ▶ Test if "year variations have a Gaussian distribution"

23 / 1

4. Review of statistical inference

Regression/ANOVA

Can we test the equality of the means of three or more distributions?

Reexample with `ferries_panel.Rdata`

24 / 1

Laurea Magistrale in Ingegneria Informatica

Statistics for High Dimensional Data

Topic: Regression and Splines

Module on time series (3 cfu)
aa 2022-2023

Alessandro Fassò

1 / 31

TOC

- ▶ Review of classical linear regression
- ▶ CO2-ferries data set
- ▶ Categorical predictors
- ▶ Cross-validation
- ▶ Variable selection
 - ▶ Stepwise regression
 - ▶ LASSO
- ▶ Mixed effects regression
- ▶ Local linear regression
- ▶ Splines
 - ▶ Regression splines
 - ▶ Smoothing splines

2 / 31

The linear regression model

Consider

- ▶ n observations $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, where
- ▶ $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,k})$, $i = 1, \dots, n$ denotes a k -dim predictor vector
- ▶ Each predictor $x_{i,j}$ may be numerical or categorical (later)
- ▶ y_i is the response variable

The linear regression model is given by:

$$\begin{aligned} y_i &= \beta_0 + \sum_{j=1}^k x_{i,j} \beta_j + \varepsilon_i \\ &= (\mathbf{1}, \mathbf{x}_i') \boldsymbol{\beta} + \varepsilon_i \end{aligned} \quad (1)$$

3 / 31

The linear regression model

Consider

- ▶ n observations $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, where
- ▶ $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,k})$, $i = 1, \dots, n$ denotes a k -dim predictor vector
- ▶ Each predictor $x_{i,j}$ may be numerical or categorical (later)
- ▶ y_i is the response variable

The linear regression model is given by:

$$\begin{aligned} y_i &= \beta_0 + \sum_{j=1}^k x_{i,j} \beta_j + \varepsilon_i \\ &= (\mathbf{1}, \mathbf{x}_i') \boldsymbol{\beta} + \varepsilon_i \end{aligned} \quad (1)$$

3 / 31

Error assumptions

$$\varepsilon \equiv NID(0, \sigma^2)$$

- ▶ zero mean
- ▶ constant variance
- ▶ independent
- ▶ Gaussian distributed

We will discuss their violation soon.

4 / 31

Error assumptions

$$\varepsilon \equiv NID(0, \sigma^2)$$

- ▶ zero mean
- ▶ constant variance
- ▶ independent
- ▶ Gaussian distributed

We will discuss their violation soon.

4 / 31

Matrix representation

$$Y = X\beta + \epsilon$$

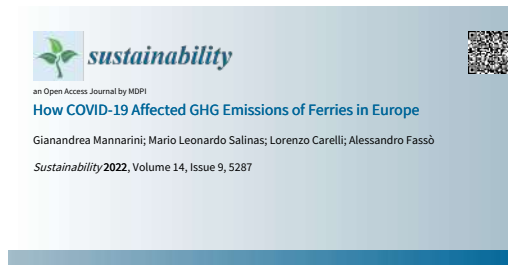
- ▶ Y is a n -dim response vector
- ▶ X is a $n \times k + 1$ matrix, with unit first column
- ▶ X is of full rank
- ▶ β is the $k + 1$ -dim parameter vector
- ▶ $\epsilon = (\epsilon_1, \dots, \epsilon_n)$

5 / 31

R LAB: data from COVID-CO2-ferries

R example using `co2_ferries_panel.csv`
and `co2_ferries_example.R`

For more details see [Zenodo Dataset](#) and [Paper](#)

The image is a screenshot of a Zenodo dataset page. The header shows the Zenodo logo and a search bar. Below the header, the dataset title 'COVID-CO2-ferries' is displayed. Underneath the title, there is a brief description of the dataset. A table is shown with columns for 'name', 'version', 'format', and 'size'. The table lists several files, including 'co2_ferries_panel.csv' and 'co2_ferries_example.R'. The 'size' column shows the file sizes in MB.

6 / 31

Wilkinson notation

- ▶ Simple regression $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$\begin{aligned} y &\sim x \\ y &\sim 1 + x \end{aligned}$$

- ▶ No intercept $y_i = \beta_1 x_i + \varepsilon_i$

$$y \sim -1 + x$$

- ▶ Multiple regression (e.g. $k=3$)

$$\text{co2} \sim \text{covid} + \text{region} + \text{speed}$$

- ▶ Interactions

$$\begin{aligned} \text{co2} &\sim \text{covid} + \text{region} + \text{covid} : \text{region} \\ \text{co2} &\sim \text{covid} * \text{region} \end{aligned}$$

7 / 31

Categorical variables

- ▶ A categorical predictor x has a number p of possible "categories" or "levels".
For example `region = "B", "M", or "N"`
- ▶ In the simplest case $p = 2$, the predictor is dichotomous and may be represented by a *dummy variable*.
- ▶ For example, consider `cbuild`. We put:
 - $x_i = 0$ if `cbuild[i]=="old"`
 - $x_i = 1$ if `cbuild[i]=="new"`.This gives the model:

$$\begin{aligned} \text{co2} &\sim \text{cbuild} \\ \text{co2}_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \end{aligned}$$

- ▶ Note that β_0 is the mean of `co2` for old ferries and β_1 is the variation due to new technology.

8 / 31

Categorical variables

- ▶ For larger p , we may use p dummy variables
- ▶ Considering `region`, we define
 - ▶ x_B dummy for `region="B"`
 - ▶ x_M dummy for `region="M"`
 - ▶ x_N dummy for `region="N"`

This gives the model:

$$\begin{aligned} \text{co2} &\sim \text{region} \\ \text{co2} &= \beta_0 + \beta_1 x_M + \beta_2 x_N + \varepsilon \end{aligned}$$

- ▶ Note that x_B is omitted to avoid a singular matrix X .
- ▶ In this case `region="B"` is the reference level.

9 / 31

R LAB

Reexample with `C02ferries.Rproj`

- ▶ Select year 2018
- ▶ Consider

$$\text{co2} \sim \text{cbuild} * \text{speed} + \text{cbuild} * \text{region}$$

and submodels

10 / 31

Ordinary least squares

OLS provides the β estimate minimizing the sum of squared errors:

$$\hat{\beta} = \operatorname{argmin} \left(\sum (y_t - \beta' x_t)^2 \right)$$

We have a closed form solution:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

The estimate is (approximately) normal under above error assumptions:

$$\hat{\beta} \cong N\left(\beta, \frac{\Sigma_{\beta}}{n}\right)$$

⇒ CIs and tests

11 / 31

Ordinary least squares

OLS provides the β estimate minimizing the sum of squared errors:

$$\hat{\beta} = \operatorname{argmin} \left(\sum (y_t - \beta' x_t)^2 \right)$$

We have a closed form solution:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

The estimate is (approximately) normal under above error assumptions:

$$\hat{\beta} \cong N\left(\beta, \frac{\Sigma_{\beta}}{n}\right)$$

⇒ CIs and tests

11 / 31

Ordinary least squares

OLS provides the β estimate minimizing the sum of squared errors:

$$\hat{\beta} = \operatorname{argmin} \left(\sum (y_t - \beta' x_t)^2 \right)$$

We have a closed form solution:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

The estimate is (approximately) normal under above error assumptions:

$$\hat{\beta} \cong N\left(\beta, \frac{\Sigma_{\beta}}{n}\right)$$

⇒ CIs and tests

11 / 31

Inference on regression model

Degrees of freedom and their decomposition:

Error	$df_e = n - k - 1$
Model	$df_{mod} = k$
Data	$df_y = n - 1 = df_e + df_{mod}$

Fitting:

- ▶ $SSE = \sum (\hat{y}_i - y_i)^2 = \text{Residual SS}$
- ▶ $SST = \sum (y_i - \bar{y})^2 = \text{Total SS}$
- ▶ $SSR = SST - SSE =$
- ▶ $MSE = SSE / df_e$
- ▶ $RMSE = \sqrt{MSE}$

12 / 31

Inference on regression model

Degrees of freedom and their decomposition:

$$\begin{array}{ll}\text{Error} & df_e = n - k - 1 \\ \text{Model} & df_{mod} = k \\ \text{Data} & df_y = n - 1 = df_e + df_{mod}\end{array}$$

Fitting:

- ▶ $SSE = \sum (\hat{y}_i - y_i)^2 = \text{Residual SS}$
- ▶ $SST = \sum (y_i - \bar{y})^2 = \text{Total SS}$
- ▶ $SSR = SST - SSE =$
- ▶ $MSE = SSE / df_e$
- ▶ $RMSE = \sqrt{MSE}$

12 / 31

Inference on regression model

Degrees of freedom and their decomposition:

$$\begin{array}{ll}\text{Error} & df_e = n - k - 1 \\ \text{Model} & df_{mod} = k \\ \text{Data} & df_y = n - 1 = df_e + df_{mod}\end{array}$$

Fitting:

- ▶ $SSE = \sum (\hat{y}_i - y_i)^2 = \text{Residual SS}$
- ▶ $SST = \sum (y_i - \bar{y})^2 = \text{Total SS}$
- ▶ $SSR = SST - SSE =$
- ▶ $MSE = SSE / df_e$
- ▶ $RMSE = \sqrt{MSE}$

12 / 31

Inference)

Fitting:

- ▶ $R^2 = 1 - \frac{SSE}{SST} = r(y, \hat{y})^2$
- ▶ $R_{adj}^2 = 1 - \frac{MSE}{S_y^2} = 1 - \frac{SSE/df_e}{SST/df_y}$

Interpretation:

$$\begin{array}{ll}R^2 & = \text{SS fraction captured by the model} \\ R_{adj}^2 & = \text{Variance fraction captured by the model}\end{array}$$

13 / 31

Testing

F-test - overall model:

$$H_0 : \beta_j = 0 \quad j = 2, \dots, k + 1$$

t-test - single coefficients:

$$H_{0j} : \beta_j = 0$$

$$j = 1, \dots, k + 1$$

14 / 31

Testing

F-test - overall model:

$$H_0 : \beta_j = 0 \quad j = 2, \dots, k + 1$$

t-test - single coefficients:

$$H_{0j} : \beta_j = 0$$

$$j = 1, \dots, k + 1$$

14 / 31

R LAB

Reexample with `CO2ferries.Rproj`

- ▶ Use full panel
- ▶ Consider

$$\text{co2} \sim \text{covid} * \text{speed} + \text{covid} * \text{region}$$

and submodels

15 / 31

cross-validation

The model is assessed on data not used for estimation.
Data is splitted: Learning and Testing.

$$Y = \begin{Bmatrix} Y_L \\ Y_T \end{Bmatrix} \quad X = \begin{Bmatrix} X_L \\ X_T \end{Bmatrix}$$

The Learning set is used for estimation:

$$\hat{\beta}_L = (X_L' X_L)^{-1} X_L' Y_L$$

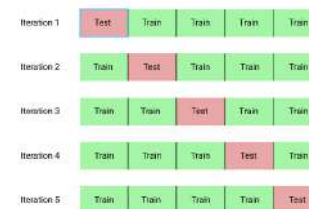
The testing set is used to assess the performance (validation):

$$e_T = \hat{\beta}_L' X_T - Y_T \Rightarrow \text{MSE}_T = e_T' e_T / n_T$$

16 / 31

k-fold cross-validation

A typical cross-validation setup is based on k subsets and k iterations:



- ▶ Each iteration handle one row as in the previous slide.
- ▶ Rows of $[Y, X]$ are randomly shuffled (not if data have spatial or temporal or other structures).

17 / 31

Variable selection

Consider the regression model:

$$Y = X\beta + \varepsilon$$

and the OLS estimate $\hat{\beta}$.

If X is an $n \times k$ matrix with $k \ll n$ and $\text{rank}(X) = k$: How many submodels do we have?

The best model is one out of 2^k with best performance in cross-validation.

The performance may be assessed using MSE, MAE or others.

18 / 31

Variable selection

Consider the regression model:

$$Y = X\beta + \varepsilon$$

and the OLS estimate $\hat{\beta}$.

If X is an $n \times k$ matrix with $k \ll n$ and $\text{rank}(X) = k$: How many submodels do we have?

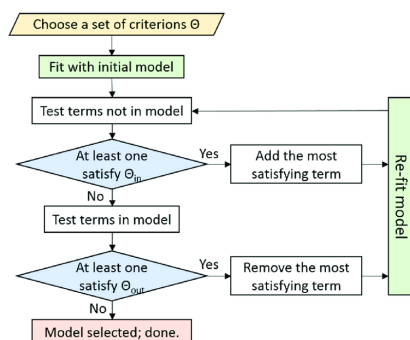
The best model is one out of 2^k with best performance in cross-validation.

The performance may be assessed using MSE, MAE or others.

18 / 31

Stepwise

To avoid testing all 2^k submodels, the stepwise regression method is an iterative forward-backward algorithm that, at each step, includes a new variable if it "adds" something and removes a previously included variable if this is no more needed.



Typical criteria are: the p - value or the $AIC = MSE + 2k$.

19 / 31

LASSO

If k is large, $k \leq n$ and/or $\text{rank}(X) < k$, OLS and stepwise do not work well.

LASSO performs estimation and variable selection at the same time.

It works also for large k and $k > n$.

It is based on penalised LS:

$$\hat{\beta}_\lambda = \arg \min_{\beta} \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \lambda \sum_{j=1}^k |\beta_j|$$

where λ is optimised in cross-validation.

20 / 31

LASSO

If k is large, $k \leq n$ and/or $\text{rank}(X) < k$, OLS and stepwise do not work well.

LASSO performs estimation and variable selection at the same time.

It works also for large k and $k > n$.

It is based on penalised LS:

$$\hat{\beta}_\lambda = \arg \min_{\beta} \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \lambda \sum_{j=1}^k |\beta_j|$$

where λ is optimised in cross-validation.

20 / 31

Mixed Effect Linear Models

Remember the correlation among years in the Rexample with `C02ferries.Rproj`

In general, modelling panel data

$$y_{t,i} = \beta_0 + \beta_1 x_{t,i,1} + \dots + \beta_k x_{t,i,k} + \varepsilon_{t,i}, t = 1, \dots, n_i$$

implies that errors are not independent and OLS does not provide proper inference.

The Random intercept model is given by

$$y_{t,i} = a_i + \beta_0 + \beta_1 x_{t,i,1} + \dots + \beta_k x_{t,i,k} + \varepsilon_{t,i}$$

where $a_i \equiv N(0, \sigma_a^2)$ is the so-called "subject specific" intercept.

21 / 31

Local linear regression

The loess regression is based on the model

$$y = \beta_0(x) + \beta_1(x)x + \varepsilon$$

where $\beta(x)$ varies "slowly" with x .

Such a "functional" β may be estimated for each single x using the data near x

$$\hat{\beta}_\lambda(x) = \argmin \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \times K\left(\frac{x - x_i}{\lambda}\right)$$

where

- ▶ K is the kernel providing the weights of x_i , decreasing with $|x - x_i|$,
- ▶ λ is the band-width factor defining the size of the same neighborhood.

22 / 31

Local linear regression

The loess regression is based on the model

$$y = \beta_0(x) + \beta_1(x)x + \varepsilon$$

where $\beta(x)$ varies "slowly" with x .

Such a "functional" β may be estimated for each single x using the data near x

$$\hat{\beta}_\lambda(x) = \argmin \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \times K\left(\frac{x - x_i}{\lambda}\right)$$

where

- ▶ K is the kernel providing the weights of x_i , decreasing with $|x - x_i|$,
- ▶ λ is the band-width factor defining the size of the same neighborhood.

22 / 31

Basis functions in general

$$y = f(x) + \varepsilon \quad \text{for } a \leq x \leq b$$

where

$$f(x) = \sum_{j=1}^p \beta_j B_j(x)$$

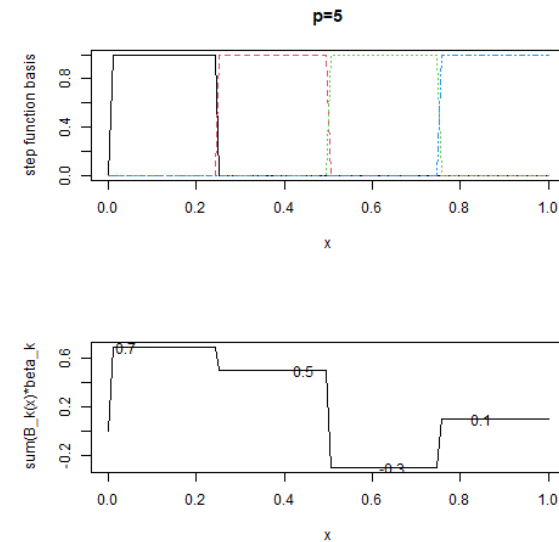
and B_1, \dots, B_p is a set of basis functions.

Examples of basis functions

1. Global polynomials: $f(x) = \beta_0 + \beta_1 x + \dots + \beta_p x^p$
2. Piecewise constant basis functions:
 $B_j(x) = I_{(\tau_j, \tau_{j+1}]}(x)$, $a = \tau_1 < \dots < \tau_K = b$
3. Piecewise linear: $B_{2j-r}(x) = I_{(\tau_{2j}, \tau_{2j+1}]}(x) x^r$, $r = 0, 1$
4. Do 2 and 3 define continuous functions?

23 / 31

Example of piecewise constant basis functions

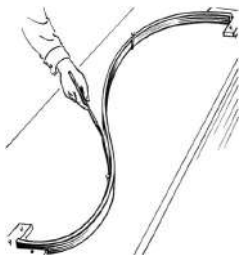


24 / 31

Splines

Historically the term "spline" comes from a flexible thin strip of wood or metal, used by craftsman's to draft smooth curves. For example, the cross-section of an airplane wing.

Wood splines

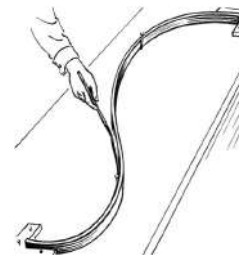


25 / 31

Splines

Historically the term "spline" comes from a flexible thin strip of wood or metal, used by craftsman's to draft smooth curves. For example, the cross-section of an airplane wing.

Wood splines



Powder snow ski tracks:
approximated by spline
bases?



25 / 31

Cubic spline basis

- ▶ Let $x_+ = \max(0, x)$
- ▶ Consider K "knots", $\tau_1 < \dots < \tau_K$
- ▶ Cubic spline basis functions: are defined by

$$B_1(x) = 1, B_2(x) = x, B_3(x) = x^2, B_4(x) = x^3$$

$$B_{4+1}(x) = (x - \tau_1)_+^3, \dots, B_{4+K}(x) = (x - \tau_K)_+^3$$

- ▶ Equivalently:

$$f_\beta(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \\ + \beta_5 (x - \tau_1)_+^3 + \dots + \beta_{4+K} (x - \tau_K)_+^3$$

- ▶ $f()$ is very smooth: it has two continuous derivatives

26 / 31

Cubic spline basis

- ▶ Let $x_+ = \max(0, x)$
- ▶ Consider K "knots", $\tau_1 < \dots < \tau_K$
- ▶ Cubic spline basis functions: are defined by

$$B_1(x) = 1, B_2(x) = x, B_3(x) = x^2, B_4(x) = x^3$$

$$B_{4+1}(x) = (x - \tau_1)_+^3, \dots, B_{4+K}(x) = (x - \tau_K)_+^3$$

- ▶ Equivalently:

$$f_\beta(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \\ + \beta_5 (x - \tau_1)_+^3 + \dots + \beta_{4+K} (x - \tau_K)_+^3$$

- ▶ $f()$ is very smooth: it has two continuous derivatives

26 / 31

Cubic spline basis

- ▶ Let $x_+ = \max(0, x)$
- ▶ Consider K "knots", $\tau_1 < \dots < \tau_K$
- ▶ Cubic spline basis functions: are defined by

$$B_1(x) = 1, B_2(x) = x, B_3(x) = x^2, B_4(x) = x^3$$

$$B_{4+1}(x) = (x - \tau_1)_+^3, \dots, B_{4+K}(x) = (x - \tau_K)_+^3$$

- ▶ Equivalently:

$$f_\beta(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \\ + \beta_5 (x - \tau_1)_+^3 + \dots + \beta_{4+K} (x - \tau_K)_+^3$$

- ▶ $f()$ is very smooth: it has two continuous derivatives

26 / 31

Cubic spline basis

- ▶ Let $x_+ = \max(0, x)$
- ▶ Consider K "knots", $\tau_1 < \dots < \tau_K$
- ▶ Cubic spline basis functions: are defined by

$$B_1(x) = 1, B_2(x) = x, B_3(x) = x^2, B_4(x) = x^3$$

$$B_{4+1}(x) = (x - \tau_1)_+^3, \dots, B_{4+K}(x) = (x - \tau_K)_+^3$$

- ▶ Equivalently:

$$f_\beta(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \\ + \beta_5 (x - \tau_1)_+^3 + \dots + \beta_{4+K} (x - \tau_K)_+^3$$

- ▶ $f()$ is very smooth: it has two continuous derivatives

26 / 31

Cubic spline basis

- ▶ Let $x_+ = \max(0, x)$
- ▶ Consider K "knots", $\tau_1 < \dots < \tau_K$
- ▶ Cubic spline basis functions: are defined by

$$B_1(x) = 1, B_2(x) = x, B_3(x) = x^2, B_4(x) = x^3$$

$$B_{4+1}(x) = (x - \tau_1)_+^3, \dots, B_{4+K}(x) = (x - \tau_K)_+^3$$

- ▶ Equivalently:

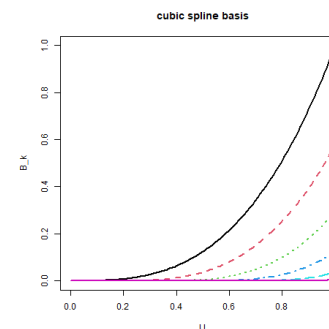
$$f_\beta(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \beta_5 (x - \tau_1)_+^3 + \dots + \beta_{4+K} (x - \tau_K)_+^3$$

- ▶ $f(\cdot)$ is very smooth: it has two continuous derivatives

26 / 31

Example of Polynomial basis functions

Cubic spline

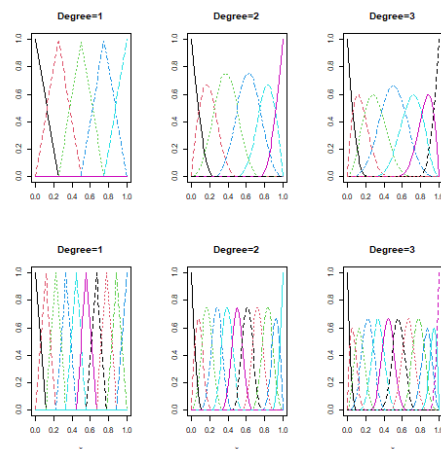


$K = 5$
knots

27 / 31

Example of B-spline basis functions

reparametrization of polynomial splines



\Rightarrow Sparse $X'X \Rightarrow$ efficient computation

28 / 31

Statistical modelling: Regression splines

Consider n bivariate data:

$$(y_1, x_1), \dots, (y_n, x_n)$$

(cubic B-) spline basis and a set of knots, $\tau_1 < \dots < \tau_K$
then a regression spline model is fitted by OLS:

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_{4+K}) = \underset{\beta}{\operatorname{argmin}} \sum (f_\beta(x_i) - y_i)^2$$

In practice one has also to consider:

- ▶ risk of overfitting
- ▶ which and how many knots?
- ▶ endpoint behaviour: for example Natural splines ensure linearity at the endpoints.

29 / 31

Statistical modelling: Smoothing splines

An alternative to regression splines is to have a large number of knots and to penalise for complexity:

$$\hat{\beta}_\lambda = \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \int (f(x''))^2 dx$$

$$\begin{aligned}\hat{\beta}_\lambda &= \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \beta' \Omega \beta \\ &= (B' B + \lambda \Omega)^{-1} B' y\end{aligned}$$

► $\lambda = 0 \Rightarrow \hat{f} = \text{regression spline}$

► $\lambda \gg 0 \Rightarrow \hat{f} \text{ linear}$

30 / 31

Statistical modelling: Smoothing splines

An alternative to regression splines is to have a large number of knots and to penalise for complexity:

$$\hat{\beta}_\lambda = \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \int (f(x''))^2 dx$$

$$\begin{aligned}\hat{\beta}_\lambda &= \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \beta' \Omega \beta \\ &= (B' B + \lambda \Omega)^{-1} B' y\end{aligned}$$

► $\lambda = 0 \Rightarrow \hat{f} = \text{regression spline}$

► $\lambda \gg 0 \Rightarrow \hat{f} \text{ linear}$

30 / 31

Statistical modelling: Smoothing splines

An alternative to regression splines is to have a large number of knots and to penalise for complexity:

$$\hat{\beta}_\lambda = \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \int (f(x''))^2 dx$$

$$\begin{aligned}\hat{\beta}_\lambda &= \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \beta' \Omega \beta \\ &= (B' B + \lambda \Omega)^{-1} B' y\end{aligned}$$

► $\lambda = 0 \Rightarrow \hat{f} = \text{regression spline}$

► $\lambda \gg 0 \Rightarrow \hat{f} \text{ linear}$

30 / 31

Statistical modelling: Smoothing splines

An alternative to regression splines is to have a large number of knots and to penalise for complexity:

$$\hat{\beta}_\lambda = \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \int (f(x''))^2 dx$$

$$\begin{aligned}\hat{\beta}_\lambda &= \operatorname{argmin}_\beta \sum (f_\beta(x_i) - y_i)^2 + \lambda \beta' \Omega \beta \\ &= (B' B + \lambda \Omega)^{-1} B' y\end{aligned}$$

► $\lambda = 0 \Rightarrow \hat{f} = \text{regression spline}$

► $\lambda \gg 0 \Rightarrow \hat{f} \text{ linear}$

30 / 31

Degrees of freedom

Equivalent number of parameters

In full rank linear regression

$$Y = X\beta + \varepsilon$$

we have

$$df = k + 1 = \sqrt{\text{trace}((X'(X'X)^{-1}X))}$$

in smoothing splines we have

$$df = \sqrt{\text{trace}(B'(B'B + \lambda\Omega)^{-1}B)}$$

31 / 31

Degrees of freedom

Equivalent number of parameters

In full rank linear regression

$$Y = X\beta + \varepsilon$$

we have

$$df = k + 1 = \sqrt{\text{trace}((X'(X'X)^{-1}X))}$$

in smoothing splines we have

$$df = \sqrt{\text{trace}(B'(B'B + \lambda\Omega)^{-1}B)}$$

31 / 31

Laurea Magistrale in Ingegneria Gestionale

Statistica II

L06: Introduzione alle serie storiche

aa 2021-2022
Alessandro Fassò

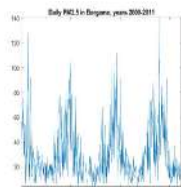
1 / 1

Contenuti

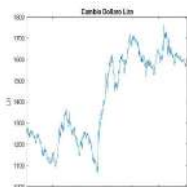
2 / 1

Serie storiche - esempi

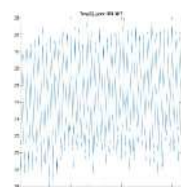
Ricordiamo gli esempi dell'introduzione (L01):



Daily PM2.5
in Bergamo



Daily USD



IGRA Monthly
Temperature (EU)

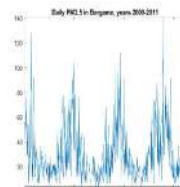
Una serie storica è data da una serie di osservazioni indicizzate nel tempo:

$$y_t, \quad t = 1, 2, \dots, n$$

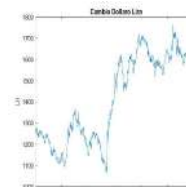
3 / 1

Serie storiche - esempi

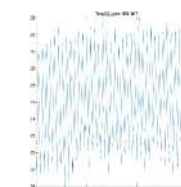
Ricordiamo gli esempi dell'introduzione (L01):



Daily PM2.5
in Bergamo



Daily USD



IGRA Monthly
Temperature (EU)

Una serie storica è data da una serie di osservazioni indicizzate nel tempo:

$$y_t, \quad t = 1, 2, \dots, n$$

3 / 1

Processi Stocastici

I modelli probabilistici per le serie storiche sono dei processi stocastici.

Un PS è dato da una famiglia di variabili casuali

$$y_t, \quad \text{con } t \in \mathcal{T} = \{0, \pm 1, \pm 2, \dots\}$$

che è caratterizzata dall'insieme delle sue distribuzioni congiunte

$$f_{y_{t_1}, y_{t_2}, \dots, y_{t_k}}(z_1, z_2, \dots, z_k)$$

per ogni $t_j \in \mathcal{T}, j = 1, \dots, k$ e $k = 1, 2, \dots$

4 / 1

Processi Stocastici

Processi stocastici a tempo discreto:

$$\mathcal{T} = \{0, \pm 1, \pm 2, \dots\} \text{ oppure } \mathcal{T} = \{1, 2, \dots\}$$

Processi stocastici a tempo continuo:

$$\mathcal{T} \in \mathbb{R} \text{ oppure } \mathcal{T} \in \mathbb{R}^+$$

Processi stocastici spaziali (campi stocastici):

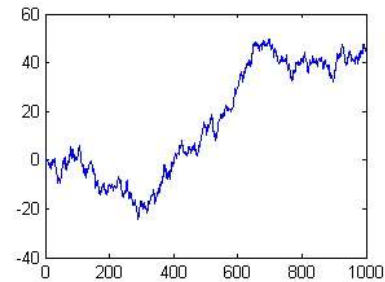
$$\mathcal{T} \in \mathbb{R}^d, \text{ con } d \geq 1$$

5 / 1

Dati simulati da Modello Random Walk

$$y_t = y_{t-1} + \varepsilon_t$$

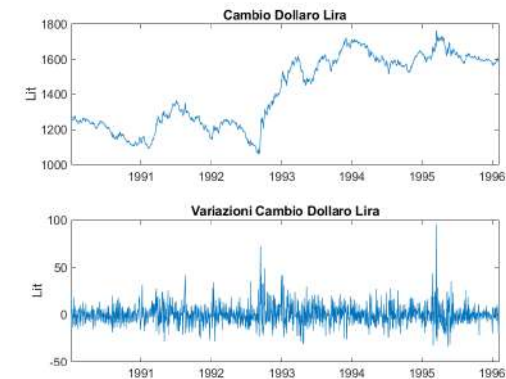
$$\varepsilon_t \equiv NID(0, \sigma^2)$$



6 / 1

Cambio Dollaro Lira

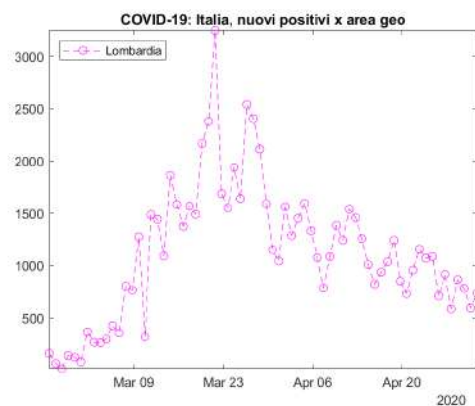
Qual'è il modello per una serie osservata?



7 / 1

COVID-19

Lombardia, Nuovi positivi



8 / 1

Distribuzioni di un PS

► Distribuzioni Marginali

$$f_{y_t}(y) = f_t(y) \quad t \in \mathcal{T}$$

► Distribuzioni k-variate

$$f_{y_{t_1}, \dots, y_{t_k}}(\mathbf{x}) \quad t_1, \dots, t_k \in \mathcal{T}$$

9 / 1

Distribuzioni di un PS

► Distribuzioni Marginali

$$f_{y_t}(y) = f_t(y) \quad t \in \mathcal{T}$$

► Distribuzioni k-variate

$$f_{y_{t_1}, \dots, y_{t_k}}(\mathbf{x}) \quad t_1, \dots, t_k \in \mathcal{T}$$

9 / 1

Esempio IID

$$\begin{aligned} y_t & \text{IID } N(0, \sigma^2) \\ \Downarrow \\ f_{y_{t_1}, y_{t_2}, \dots, y_{t_k}}(z_1, z_2, \dots, z_k) &= \prod_{j=1}^k \frac{1}{\sigma} \phi(z_j / \sigma) \end{aligned}$$

NB: è il caso di processo "banale" della statistica classica in cui non vi è nè dinamica temporale nè prevedibilità.

10 / 1

Esempio trend lineare + IID

$$\begin{aligned} y_t &= \beta t + \varepsilon_t, \quad \varepsilon_t \text{ IID } N(0, \sigma^2) \\ \Downarrow \\ f_{y_{t_1}, \dots, y_{t_k}}(z_1, \dots, z_k) &= \prod_{j=1}^k \frac{1}{\sigma} \phi((z_j - \beta t_j) / \sigma) \end{aligned}$$

11 / 1

Esercizio AR(1)

Il modello $AR(1)$ è dato da:

$$y_t = \alpha y_{t-1} + \varepsilon_t \quad t = 1, 2, \dots$$

$$\varepsilon_t \equiv \text{IID } (0, \sigma^2)$$

$$|\alpha| < 1$$

con valore iniziale $y_0 = 0$ e ha distribuzione

$$f_{y_{t_1}, \dots, y_{t_k}}(z_1, \dots, z_k) = \prod_{j=1}^k \frac{1}{\sigma} \phi((z_j - \alpha z_{t-j}) / \sigma).$$

12 / 1

Processi Gaussiani

Un PS si dice gaussiano se tutte le sue distribuzioni sono gaussiane. Posto $\mathbf{t} = (t_1, \dots, t_k)$ vale:

$$(y_{t_1}, \dots, y_{t_k}) \equiv N_k(\mu_{\mathbf{t}}, \Sigma_{\mathbf{t}})$$

$$E(y_{t_k} | y_{t_1}, \dots, y_{t_{k-1}}) \text{ è lineare in } y_{t_1}, \dots, y_{t_{k-1}}$$

$$\text{Var}(y_{t_k} | y_{t_1}, \dots, y_{t_{k-1}}) = \sigma_{\mathbf{t}}^2 \text{ è indep. da } y_{t_1}, \dots, y_{t_{k-1}}$$

dove $\Sigma_{\mathbf{t}}$ è la matrice di varianze-covarianze di $(y_{t_1}, \dots, y_{t_k})$

13 / 1

Var-cov di AR(1)

Se y_t è generato da n modello AR(1), e $Y = (y_1, \dots, y_n)$ allora

$$\Sigma_Y = \text{VarCov}(Y) = \frac{\sigma^2}{1 - \alpha^2} (\alpha^{|i-j|})_{i,j=1,\dots,n}$$

NB:

$$\text{cov}(y_t, y_{t-h}) = \frac{\sigma^2}{1 - \alpha^2} \alpha^{|h|}$$

$$\text{corr}(y_t, y_{t-h}) = \alpha^{|h|}$$

14 / 1

Processi condizionatamente Gaussiani

Sono PS in cui "l'equazione di misura" è gaussiana ma la dinamica può essere non lineare.

Sia

$$\mathbf{S}_{t-1} = y_1, \dots, y_{t-1}$$

la storia antecedente il tempo t .

$$(y_t | \mathbf{S}_{t-1}) \equiv N(\mu(\mathbf{S}_{t-1}), \sigma^2(\mathbf{S}_{t-1}))$$

- ▶ $\mu()$: Dinamica Lineare/Nonlineare
 - ▶ regressione lineare con errori correlati
 - ▶ Spline, polinomi locali, reti neurali, ...
- ▶ $\sigma^2()$: Omoschedasticità/Eteroschedasticità

15 / 1

Processi condizionatamente Gaussiani

Sono PS in cui "l'equazione di misura" è gaussiana ma la dinamica può essere non lineare.

Sia

$$\mathbf{S}_{t-1} = y_1, \dots, y_{t-1}$$

la storia antecedente il tempo t .

$$(y_t | \mathbf{S}_{t-1}) \equiv N(\mu(\mathbf{S}_{t-1}), \sigma^2(\mathbf{S}_{t-1}))$$

- ▶ $\mu()$: Dinamica Lineare/Nonlineare
 - ▶ regressione lineare con errori correlati
 - ▶ Spline, polinomi locali, reti neurali, ...
- ▶ $\sigma^2()$: Omoschedasticità/Eteroschedasticità

15 / 1

Teoria della Previsione

Sia \mathbf{S}_t l'insieme delle informazioni (storia) fino a t .

Un previsore a k passi di y_{t+k} è una funzione di \mathbf{S}_t :

$$\hat{y}_{t+k}(k) = f_k(y_t, y_{t-1}, \dots) = f_k(\mathbf{S}_t)$$

Nel ns. caso:

$$\mathbf{S}_t = (y_t, y_{t-1}, \dots).$$

16 / 1

17 / 1

Teoria della Previsione

Sia \mathbf{S}_t l'insieme delle informazioni (storia) fino a t .

Un previsore a k passi di y_{t+k} è una funzione di \mathbf{S}_t :

$$\hat{y}_{t+k}(k) = f_k(y_t, y_{t-1}, \dots) = f_k(\mathbf{S}_t)$$

Nel ns. caso:

$$\mathbf{S}_t = (y_t, y_{t-1}, \dots).$$

17 / 1

Teoria della Previsione

Se $E y_t^2 < \infty$ allora la previsione ottima (a minima varianza) è

$$\hat{y}_{t+k}(k) = E(y_{t+k} | \mathbf{S}_t)$$

In particolare è importante la **previsione ad un passo**

$$\hat{y}_{t+1} = \hat{y}_{t+1}(1) = E(y_{t+1} | \mathbf{S}_t)$$

Linearità

Un previsore si dice lineare se è una funzione lineare del passato:

$$\hat{y}_{t+1} = E(y_{t+1} | \mathbf{S}_t) = \sum_{j=1}^{\infty} \xi_j y_{t-j}$$

18 / 1

Teoria della Previsione

Se $Ey_t^2 < \infty$ allora la previsione ottima (a minima varianza) è

$$\hat{y}_{t+k}(k) = E(y_{t+k} | \mathbf{S}_t)$$

In particolare è importante la **previsione ad un passo**

$$\hat{y}_{t+1} = \hat{y}_{t+1}(1) = E(y_{t+1} | \mathbf{S}_t)$$

Linearità

Un previsore si dice lineare se è una funzione lineare del passato:

$$\hat{y}_{t+1} = E(y_{t+1} | \mathbf{S}_t) = \sum_{j=1}^{\infty} \xi_j y_{t-j}$$

18 / 1

Teoria della Previsione

Se $Ey_t^2 < \infty$ allora la previsione ottima (a minima varianza) è

$$\hat{y}_{t+k}(k) = E(y_{t+k} | \mathbf{S}_t)$$

In particolare è importante la **previsione ad un passo**

$$\hat{y}_{t+1} = \hat{y}_{t+1}(1) = E(y_{t+1} | \mathbf{S}_t)$$

Linearità

Un previsore si dice lineare se è una funzione lineare del passato:

$$\hat{y}_{t+1} = E(y_{t+1} | \mathbf{S}_t) = \sum_{j=1}^{\infty} \xi_j y_{t-j}$$

18 / 1

Esempio AR(1)

- Previsione a un passo

$$\hat{y}_{t+1} = \alpha y_t$$

- Previsione a k passi

$$\hat{y}_{t+1}(k) = \alpha^k y_t$$

- Memoria breve:

$$|\alpha| < 1 \Rightarrow \hat{y}_t(k) \rightarrow 0 \text{ per } k \rightarrow \infty$$

19 / 1

Esempio AR(1)

- Previsione a un passo

$$\hat{y}_{t+1} = \alpha y_t$$

- Previsione a k passi

$$\hat{y}_{t+1}(k) = \alpha^k y_t$$

- Memoria breve:

$$|\alpha| < 1 \Rightarrow \hat{y}_t(k) \rightarrow 0 \text{ per } k \rightarrow \infty$$

19 / 1

Esempio AR(1)

- Previsione a un passo

$$\hat{y}_{t+1} = \alpha y_t$$

- Previsione a k passi

$$\hat{y}_{t+1}(k) = \alpha^k y_t$$

- Memoria breve:

$$|\alpha| < 1 \Rightarrow \hat{y}_t(k) \rightarrow 0 \text{ per } k \rightarrow \infty$$

19 / 1

L'errore di previsione

$$e_t = \hat{y}_t - y_t$$

- ha media nulla: $E(e_t) = 0$
- è incorrelato (ortogonale) col passato

$$E(e_t y_{t-j}) = 0, \quad j = 1, 2, \dots$$

- La sua varianza determina la precisione della previsione:

$$\sigma_{e_t}^2 = E(e_t^2)$$

- Supporremo nel seguito omoschedasticità:

$$\sigma^2 = \sigma_{e_t}^2 = E(e_t^2 | \mathbf{S}_{t-1})$$

20 / 1

L'errore di previsione

$$e_t = \hat{y}_t - y_t$$

- ha media nulla: $E(e_t) = 0$
- è incorrelato (ortogonale) col passato

$$E(e_t y_{t-j}) = 0, \quad j = 1, 2, \dots$$

- La sua varianza determina la precisione della previsione:

$$\sigma_{e_t}^2 = E(e_t^2)$$

- Supporremo nel seguito omoschedasticità:

$$\sigma^2 = \sigma_{e_t}^2 = E(e_t^2 | \mathbf{S}_{t-1})$$

20 / 1

L'errore di previsione

$$e_t = \hat{y}_t - y_t$$

- ha media nulla: $E(e_t) = 0$
- è incorrelato (ortogonale) col passato

$$E(e_t y_{t-j}) = 0, \quad j = 1, 2, \dots$$

- La sua varianza determina la precisione della previsione:

$$\sigma_{e_t}^2 = E(e_t^2)$$

- Supporremo nel seguito omoschedasticità:

$$\sigma^2 = \sigma_{e_t}^2 = E(e_t^2 | \mathbf{S}_{t-1})$$

20 / 1

L'errore di previsione

$$e_t = \hat{y}_t - y_t$$

- ▶ ha media nulla: $E(e_t) = 0$
- ▶ è incorrelato (ortogonale) col passato

$$E(e_t y_{t-j}) = 0, \quad j = 1, 2, \dots$$

- ▶ La sua varianza determina la precisione della previsione:

$$\sigma_{e_t}^2 = E(e_t^2)$$

- ▶ Supporremo nel seguito omoschedasticità:

$$\sigma^2 = \sigma_{e_t}^2 = E(e_t^2 | \mathbf{S}_{t-1})$$

20 / 1

L'errore di previsione

$$e_t = \hat{y}_t - y_t$$

- ▶ ha media nulla: $E(e_t) = 0$
- ▶ è incorrelato (ortogonale) col passato

$$E(e_t y_{t-j}) = 0, \quad j = 1, 2, \dots$$

- ▶ La sua varianza determina la precisione della previsione:

$$\sigma_{e_t}^2 = E(e_t^2)$$

- ▶ Supporremo nel seguito omoschedasticità:

$$\sigma^2 = \sigma_{e_t}^2 = E(e_t^2 | \mathbf{S}_{t-1})$$

20 / 1

Processi Stazionari

Un processo stocastico si dice **stazionario in senso forte** se le sue distribuzioni non dipendono dall'origine dei tempi:

$$\begin{aligned} f_{y_{t_1}, y_{t_2}, \dots, y_{t_k}}(z_1, z_2, \dots, z_k) = \\ f_{y_{t_1+h}, y_{t_2+h}, \dots, y_{t_k+h}}(z_1, z_2, \dots, z_k) \quad \text{per ogni } h \end{aligned}$$

Si dice **stazionario in senso debole** se i momenti del primo e del second'ordine non dipendono da t

$$\begin{aligned} E(y_t) &= \mu_y && \text{per ogni } t \\ \text{Var}(y_t) &= \sigma_y^2 && \text{per ogni } t \\ E(y_t y_{t+h}) &= \gamma(h) && \text{per ogni } t \end{aligned}$$

21 / 1

22 / 1

Processi Stazionari

Un processo stocastico si dice **stazionario in senso forte** se le sue distribuzioni non dipendono dall'origine dei tempi:

$$\begin{aligned} f_{y_{t_1}, y_{t_2}, \dots, y_{t_k}}(z_1, z_2, \dots, z_k) &= \\ f_{y_{t_1+h}, y_{t_2+h}, \dots, y_{t_k+h}}(z_1, z_2, \dots, z_k) &\text{ per ogni } h \end{aligned}$$

Si dice **stazionario in senso debole** se i momenti del primo e del second'ordine non dipendono da t

$$\begin{aligned} E(y_t) &= \mu_y && \text{per ogni } t \\ \text{Var}(y_t) &= \sigma_y^2 && \text{per ogni } t \\ E(y_t y_{t+h}) &= \gamma(h) && \text{per ogni } t \end{aligned}$$

22 / 1

Violazioni tipiche della stazionarietà

Vediamo alcuni esempi di non-stazionarietà.

- ▶ La media $\mu_t = E(y_t)$ non è stazionaria:
 1. trend lineari $\mu_t = \alpha + \beta t$
 2. trend non-lineari $\mu_t = \mu(t) =$
trigonometrica/esponenziale/polinomiale/spline ...
 3. regressione $\mu_t = x_t' \beta$
- ▶ Se $y_t - \mu_t$ è stazionario, posso usare $y_t - \hat{\mu}_t$
- ▶ Non stazionarietà in varianza (eteroschedasticità)
- ▶ Non stazionarietà in covarianza

23 / 1

Autocovarianza

L'autocovarianza di un PS che ha varianza finita è data da:

$$\gamma_t(h) = \text{Cov}(y_t, y_{t+h}) = E(y_t - \mu_t)(y_{t+h} - \mu_{t+h})$$

Se il PS è stazionario allora l'autocovarianza dipende solo dalla distanza $|h|$:

$$\gamma_t(h) = \gamma(|h|) = E(y_t y_{t+h}) - \mu_y^2$$

24 / 1

Autocovarianza

L'autocovarianza di un PS che ha varianza finita è data da:

$$\gamma_t(h) = \text{Cov}(y_t, y_{t+h}) = E(y_t - \mu_t)(y_{t+h} - \mu_{t+h})$$

Se il PS è stazionario allora l'autocovarianza dipende solo dalla distanza $|h|$:

$$\gamma_t(h) = \gamma(|h|) = E(y_t y_{t+h}) - \mu_y^2$$

24 / 1

Autocorrelazione

L'autocorrelazione del PS y_t è data da

$$\rho_t(h) = \rho(y_t, y_{t+h}) = \frac{\text{Cov}(y_t, y_{t+h})}{\sigma_{y_t} \sigma_{y_{t+h}}}, \quad h = 0, \pm 1, \pm 2, \dots$$

Se il PS è stazionario allora

$$\rho_t(h) = \rho(|h|) = \frac{\gamma(h)}{\gamma(0)}$$

25 / 1

Autocorrelazione campionaria

L'autocovarianza campionaria è data da

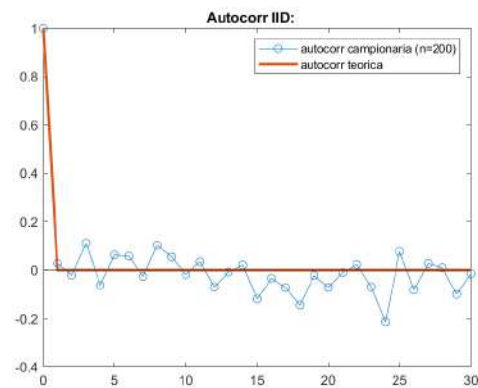
$$c(h) = \frac{1}{n} \sum_{t=1}^{n-h} (y_t - \bar{y})(y_{t+h} - \bar{y})$$

L'autocorrelazione campionaria è data da

$$r(h) = \frac{c(h)}{c(0)}$$

26 / 1

Autocorrelazione IID



27 / 1

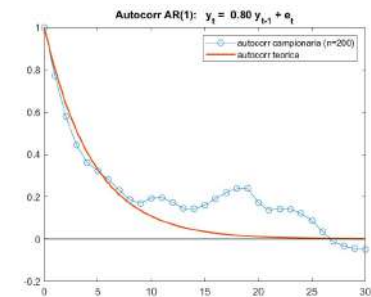
Autocorrelazione AR(1)

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

$$|\alpha| < 1$$

\Downarrow

$$\rho(h) = \alpha^h$$



28 / 1

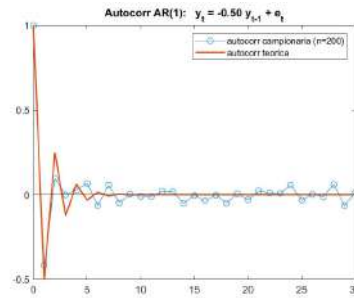
Autocorrelazione AR(1)

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

$$|\alpha| < 1$$

↓

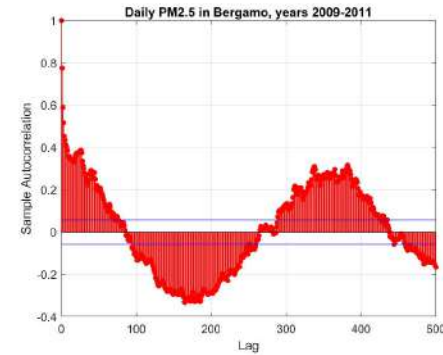
$$\rho(h) = \alpha^h$$



29 / 1

Dati giornalieri

PM2.5 a Bergamo

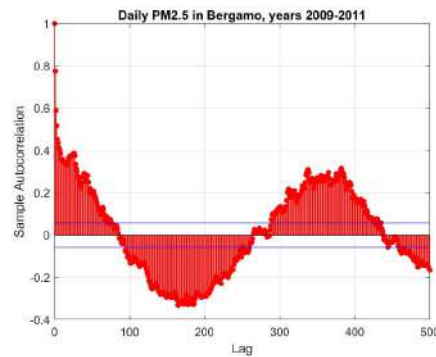


Cosa sono le due rette orizzontali blu?

30 / 1

Dati giornalieri

PM2.5 a Bergamo



Cosa sono le due rette orizzontali blu?

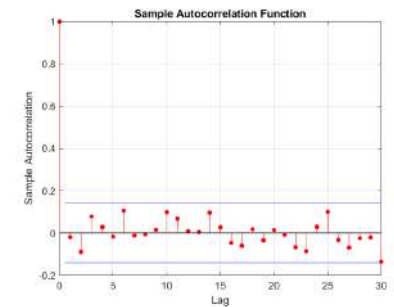
30 / 1

Distribuzione dell'autocorrelazione campionaria

$$y_t \text{ IID}$$

↓

$$\sqrt{nr}(h) \cong N(0, 1)$$

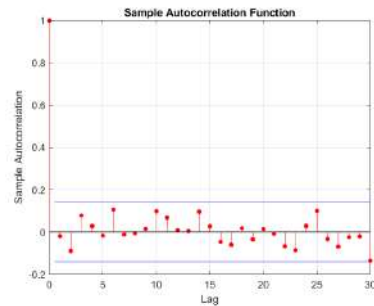


Correlogramma di n=200 osservazioni
NID

31 / 1

Distribuzione dell'autocorrelazione campionaria

$$y_t \text{ IID}$$
$$\Downarrow$$
$$\sqrt{nr}(h) \cong N(0, 1)$$



Correlogramma di n=200 osservazioni
NID

31 / 1

32 / 1

fare autocorr residui
regressione dati QA

33 / 1

Modelli ARMA

I modelli AutoRegressivi (AR) a Media Mobile (MA) sono dei modelli lineari che descrivono un'ampia casistica di funzioni di autocorrelazione caratterizzate da memoria breve.

Per memoria breve si intende che la correlazione tende a zero velocemente:

$$\rho(h) \approx \lambda^h$$

con $|\lambda| < 1$ Iniziamo ora lo studio della parte AR.

34 / 1

Modelli ARMA

I modelli AutoRegressivi (AR) a Media Mobile (MA) sono dei modelli lineari che descrivono un'ampia casistica di funzioni di autocorrelazione caratterizzate da memoria breve.

Per memoria breve si intende che la correlazione tende a zero velocemente:

$$\rho(h) \approx \lambda^h$$

con $|\lambda| < 1$ Iniziamo ora lo studio della parte AR.

Modelli ARMA

I modelli AutoRegressivi (AR) a Media Mobile (MA) sono dei modelli lineari che descrivono un'ampia casistica di funzioni di autocorrelazione caratterizzate da memoria breve.

Per memoria breve si intende che la correlazione tende a zero velocemente:

$$\rho(h) \approx \lambda^h$$

con $|\lambda| < 1$ Iniziamo ora lo studio della parte AR.

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

35 / 1

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

35 / 1

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

35 / 1

Studio dettagliato di AR(1)

Caratteristiche del modello autoregressivo di ordine uno:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \\ \varepsilon_t &\equiv NID(0, \sigma^2) \\ |\alpha| &< 1 \end{aligned}$$

1. Rappresentazione MA(∞): $y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$
2. Media: $E(y_t) = 0$
3. Varianza: $Var(y_t) = \sigma^2 / (1 - \alpha^2)$
4. Autocovarianza: $\gamma(h) = \alpha^h \times \sigma^2 / (1 - \alpha^2)$
5. Autocorrelazione : $\rho(h) = \alpha^h$
6. Discutere il ruolo dell'assunzione $|\alpha| < 1$.

35 / 1

Random Walk

Modello di persistenza

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t, & t = 1, 2, \dots & \quad \varepsilon_t \text{ IID}(0, \sigma^2) \\ y_0 &= 0 \end{aligned}$$

segue

$$\begin{aligned} E(y_t) &= 0 \\ \text{Var}(y_t) &= t\sigma^2 \rightarrow \infty, \quad t \rightarrow \infty \end{aligned}$$

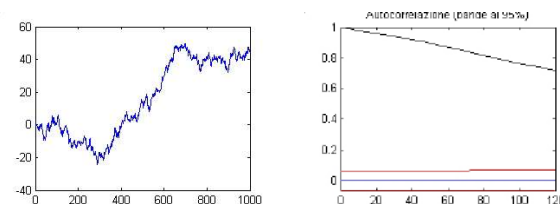
inoltre

$$\rho(y_t, y_{t+h}) = \sqrt{\frac{t}{t+h}} \rightarrow 1, \quad t \rightarrow \infty$$

36 / 1

Esempio

$n = 1000$ simulazioni di un RW con $\varepsilon_t \equiv \text{NID}(0, 1)$



Differenze prime

$$\Delta y_t = y_t - y_{t-1} \equiv \varepsilon_t$$

37 / 1

Markovianità

Un processo è markoviano se il futuro è indipendente dal passato, condizionatamente al presente:

$$f(y_{t+j} | y_t, y_{t-1}, \dots) = f(y_{t+j} | y_t)$$

Esempi: Random Walk, AR(1)

38 / 1

Modelli AR(p)

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \\ \varepsilon_t &\text{ iid } (0, \sigma^2) \\ t &= 0, \pm 1, \pm 2, \dots \end{aligned}$$

Problema: è markoviano ?

Se $h > p$

$$E(y_t y_{t-h} | y_{t-1}, \dots, y_{t-p}) = 0$$

Viene detto Markoviano di ordine p

39 / 1

Modelli AR(p)

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \\ \varepsilon_t &\text{ iid } (0, \sigma^2) \\ t &= 0, \pm 1, \pm 2, \dots \end{aligned}$$

Problema: è markoviano ?

Se $h > p$

$$E(y_t y_{t-h} | y_{t-1}, \dots, y_{t-p}) = 0$$

Viene detto Markoviano di ordine p

39 / 1

Autocorrelazione parziale

L'autocorrelazione parziale del PS stazionario y_t è data da

$$\bar{\rho}(h) = \rho(y_t, y_{t+h} | y_{t+1}, \dots, y_{t+h-1}) \quad h = 0, \pm 1, \pm 2, \dots$$

Se y_t è un modello AR(p) allora, per $h > p$, si ha

$$\bar{\rho}(h) = 0.$$

40 / 1

Autocorrelazione parziale

L'autocorrelazione parziale del PS stazionario y_t è data da

$$\bar{\rho}(h) = \rho(y_t, y_{t+h} | y_{t+1}, \dots, y_{t+h-1}) \quad h = 0, \pm 1, \pm 2, \dots$$

Se y_t è un modello AR(p) allora, per $h > p$, si ha

$$\bar{\rho}(h) = 0.$$

40 / 1

Forma polinomiale di un AR

Operatore ritardo

Consideriamo dapprima B l'operatore di ritardo:

$$B y_t = y_{t-1},$$

allora

$$B^0 y_t = y_t$$

$$B(B y_t) = B^2 y_t = y_{t-2}$$

$$B^k y_t = y_{t-k}$$

41 / 1

Forma polinomiale di un AR

Operatore ritardo

Consideriamo dapprima B l'operatore di ritardo:

$$By_t = y_{t-1},$$

allora

$$B^0 y_t = y_t$$

$$B(B y_t) = B^2 y_t = y_{t-2}$$

$$B^k y_t = y_{t-k}$$

41 / 1

Forma polinomiale di un AR

Operatore AR

Riscriviamo il modello AR(p):

$$1y_t - a_1y_{t-1} - a_2y_{t-2} - \dots - a_p y_{t-p} = \varepsilon_t$$

allora

$$B^0 y_t - B^1 a_1 y_t - \dots - B^p a_p y_t$$

$$a(B) y_t = \varepsilon_t$$

dove l'operatore AR è dato da:

$$a(B) = 1 - a_1 B - \dots - a_p B^p$$

42 / 1

Esempio AR(1)

Per il modello AR(1) abbiamo le seguenti identità

$$y_t - \alpha y_{t-1} = \varepsilon_t$$

$$a(B) y_t = \varepsilon_t$$

$$(1 - \alpha B) y_t = \varepsilon_t$$

In particolare per $\alpha = 1$ abbiamo il Random Walk in cui

$$y_t - y_{t-1} = (1 - B) y_t = \varepsilon_t$$

43 / 1

Esempio AR(1)

Per il modello AR(1) abbiamo le seguenti identità

$$y_t - \alpha y_{t-1} = \varepsilon_t$$

$$a(B) y_t = \varepsilon_t$$

$$(1 - \alpha B) y_t = \varepsilon_t$$

In particolare per $\alpha = 1$ abbiamo il Random Walk in cui

$$y_t - y_{t-1} = (1 - B) y_t = \varepsilon_t$$

43 / 1

Stazionarietà di un AR(p)

Sia $z \in \mathbb{C}$ la variabile complessa. Il PS y_t dato da un modello AR(p) è stazionario sse le p soluzioni¹ dell'equazione caratteristica

$$a\left(\frac{1}{z}\right) = 0$$

sono interne al cerchio unitario $|z| < 1$.

Ovvero sse

$$a\left(\frac{1}{z}\right) = 1 - a_1 \frac{1}{z} - \dots - a_p \frac{1}{z^p} \neq 0 \text{ per } |z| \geq 1$$

¹Teorema fondamentale dell'algebra e dintorni

Stazionarietà di un AR(p)

Sia $z \in \mathbb{C}$ la variabile complessa. Il PS y_t dato da un modello AR(p) è stazionario sse le p soluzioni¹ dell'equazione caratteristica

$$a\left(\frac{1}{z}\right) = 0$$

sono interne al cerchio unitario $|z| < 1$.

Ovvero sse

$$a\left(\frac{1}{z}\right) = 1 - a_1 \frac{1}{z} - \dots - a_p \frac{1}{z^p} \neq 0 \text{ per } |z| \geq 1$$

¹Teorema fondamentale dell'algebra e dintorni

Esempio AR(1)

Per il modello AR(1) abbiamo

$$a\left(\frac{1}{z}\right) = 1 - \frac{\alpha}{z}$$

e la soluzione dell'eq^{ne} caratteristica è

$$z = \alpha$$

Collegamento AR(1) - AR(p)

Se $\lambda_1, \dots, \lambda_p$ sono le soluzioni dell'eq^{ne} caratteristica, allora l'operatore AR(p) si può riscrivere² come prodotto di operatori AR(1)

$$a\left(\frac{1}{z}\right) = \prod_{j=1}^p \left(1 - \frac{\lambda_j}{z}\right)$$

così i ragionamenti che abbiamo fatto sugli AR(1) si estendono al caso generale.

²vedi ancora Teorema fondamentale dell'algebra e dintorni

Forma MA(∞)

Se y_t è un $AR(p)$ stazionario allora esistono i coefficienti $\omega_j, j = 1, \dots, \infty$, tali che:

$$y_t = \frac{1}{a(B)}\varepsilon_t = \sum_{j=1}^{\infty} \omega_j \varepsilon_{t-j} + \varepsilon_t$$

Caso AR(1): $\omega_j = \alpha^j$

$$y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

Forma MA(∞)

Se y_t è un $AR(p)$ stazionario allora esistono i coefficienti $\omega_j, j = 1, \dots, \infty$, tali che:

$$y_t = \frac{1}{a(B)}\varepsilon_t = \sum_{j=1}^{\infty} \omega_j \varepsilon_{t-j} + \varepsilon_t$$

Caso AR(1): $\omega_j = \alpha^j$

$$y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

Esempio AR(2)

Il modello

$$y_t = 1.7y_{t-1} - 0.72y_{t-2} + \varepsilon_t, \quad \varepsilon_t \equiv NID(0, \sigma^2)$$

ha operatore polinomiale

$$a(B) = 1 - 1.7B + 0.72B^2.$$

Le soluzioni dell'equazione caratteristica $a(\frac{1}{z}) = 0$ sono

$$z_1 = 0.9 \text{ e } z_2 = 0.8.$$

Poichè $|z_{1,2}| < 1$ il processo generato dal modello di cui sopra è **stazionario**.

Esempio AR(2)

Il modello

$$y_t = 1.7y_{t-1} - 0.72y_{t-2} + \varepsilon_t, \quad \varepsilon_t \equiv NID(0, \sigma^2)$$

ha operatore polinomiale

$$a(B) = 1 - 1.7B + 0.72B^2.$$

Le soluzioni dell'equazione caratteristica $a(\frac{1}{z}) = 0$ sono

$$z_1 = 0.9 \text{ e } z_2 = 0.8.$$

Poichè $|z_{1,2}| < 1$ il processo generato dal modello di cui sopra è **stazionario**.

... segue Esempio AR(2) - matlab

Il modello di cui sopra con $\sigma^2 = 0.5$ è implementato da:

```
>>m_AR2 = regARIMA('AR',[1.7, -.72],'Variance',.5,  
'Intercept',0)3
```

e le sue radici sono date da

```
>>r = roots([1, -1.7, .72])4
```

con modulo

```
>>abs(r)
```

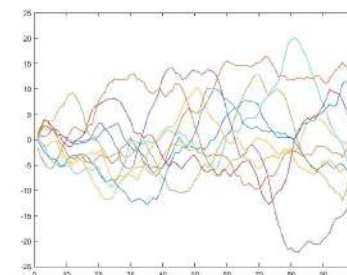
³Ovvero: `arima(...)`

⁴Ovvero: `roots([1, -cell2mat(m_AR2.AR)])`

... segue Esempio AR(2) - simulazione

Simulazione di $m=10$ repliche di una serie storica lunga $n=100$ dell'AR(2) Gaussiano di cui sopra

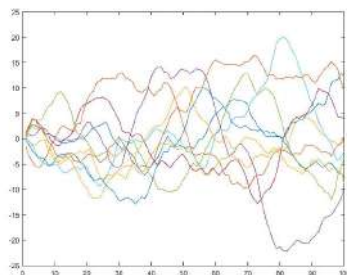
```
>> y_star = simulate(m_AR2,n,'NumPaths',m);
```



... segue Esempio AR(2) - simulazione

Simulazione di $m=10$ repliche di una serie storica lunga $n=100$ dell'AR(2) Gaussiano di cui sopra

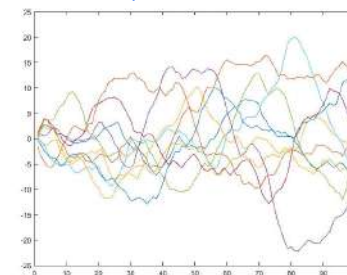
```
>> y_star = simulate(m_AR2,n,'NumPaths',m);
```



... segue Esempio AR(2) - simulazione

Simulazione di $m=10$ repliche di una serie storica lunga $n=100$ dell'AR(2) Gaussiano di cui sopra

```
>> y_star = simulate(m_AR2,n,'NumPaths',m);
```



... segue: altri modelli AR(2)

Esaminare secondo la logica vista sopra i modelli definiti dai seguenti operatori AR:

$$a(B) = 1 - 1B + 0.34B^2$$

$$a(B) = 1 - 0.4B + 0.9425B^2$$

$$a(B) = 1 - 1.2B + 1B^2$$

Esaminare anche autocorrelazione e autocorrelazione parziale!!!

51 / 1

Esempio MA(1)

Il modello a media mobile di ordine uno è dato da:

$$\begin{aligned} y_t &= \varepsilon_t + c\varepsilon_{t-1} \\ \varepsilon_t &\text{ iid } (0, \sigma^2) \end{aligned}$$

NB: non è markoviano, infatti, per $|c| < 1$, sostituendo ricorsivamente $\varepsilon_t = y_t - c\varepsilon_{t-1}$ si ottiene,

$$y_t = - \sum_{j=1}^{\infty} (-c)^j y_{t-j} + \varepsilon_t$$

o, equivalentemente,

$$\varepsilon_t = \sum_{j=0}^{\infty} (-c)^j y_{t-j}.$$

52 / 1

Momenti di MA(1)

I momenti del modello MA(1)

$$\begin{aligned} E(y_t) &= 0 \\ \gamma_{t,t} &= \text{Var}(y_t) = (1 + c^2) \sigma_\varepsilon^2 \\ \gamma_{t,t\pm 1} &= E(\varepsilon_t + c\varepsilon_{t-1})(\varepsilon_{t-1} + c\varepsilon_{t-2}) = c\sigma_\varepsilon^2 \\ \gamma_{t,t\pm h} &= 0 \quad \forall h > 1 \end{aligned}$$

Autocorrelazione:

$$\begin{aligned} \rho(\pm 1) &= \frac{c}{1+c^2} \\ \rho(h) &= 0 \quad |h| > 1 \end{aligned}$$

NB: MA è un modello con memoria finita

53 / 1

Momenti di MA(1)

I momenti del modello MA(1)

$$\begin{aligned} E(y_t) &= 0 \\ \gamma_{t,t} &= \text{Var}(y_t) = (1 + c^2) \sigma_\varepsilon^2 \\ \gamma_{t,t\pm 1} &= E(\varepsilon_t + c\varepsilon_{t-1})(\varepsilon_{t-1} + c\varepsilon_{t-2}) = c\sigma_\varepsilon^2 \\ \gamma_{t,t\pm h} &= 0 \quad \forall h > 1 \end{aligned}$$

Autocorrelazione:

$$\begin{aligned} \rho(\pm 1) &= \frac{c}{1+c^2} \\ \rho(h) &= 0 \quad |h| > 1 \end{aligned}$$

NB: MA è un modello con memoria finita

53 / 1

Modelli Autoregressivi a media mobile

Un processo ARMA(p,q) è dato dall'equazione

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + c_1 \varepsilon_{t-1} + \dots + c_q \varepsilon_{t-q} + \varepsilon_t, \\ t = 0, \pm 1, \pm 2, \dots$$

dove, $\varepsilon_t \equiv NID(0, \sigma^2)$ è l'**innovazione** di y_t .

Per semplicità di notazione assumiamo $Ey_t = 0$.

Nelle applicazioni, ove normalmente $Ey_t \neq 0$, si usa l'intercetta a_0 ⁵.

⁵Equivalentemente si possono usare gli scarti da \bar{y} . Le due strade sono equivalenti se il modello è a memoria breve. Il caso della memoria lunga (opzionale) è discusso nel Capitolo 06.

Modelli Autoregressivi a media mobile

Un processo ARMA(p,q) è dato dall'equazione

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + c_1 \varepsilon_{t-1} + \dots + c_q \varepsilon_{t-q} + \varepsilon_t, \\ t = 0, \pm 1, \pm 2, \dots$$

dove, $\varepsilon_t \equiv NID(0, \sigma^2)$ è l'**innovazione** di y_t .

Per semplicità di notazione assumiamo $Ey_t = 0$.

Nelle applicazioni, ove normalmente $Ey_t \neq 0$, si usa l'intercetta a_0 ⁵.

⁵Equivalentemente si possono usare gli scarti da \bar{y} . Le due strade sono equivalenti se il modello è a memoria breve. Il caso della memoria lunga (opzionale) è discusso nel Capitolo 06.

Forma polinomiale di un ARMA

Il modello ARMA appena introdotto può esser scritto come

$$a(B)y_t = c(B)\varepsilon_t$$

con

$$a(B) = 1 - a_1 B - \dots - a_p B^p$$

e

$$c(B) = 1 + c_1 B + \dots + c_q B^q$$

dove B è l'operatore di ritardo: $By_t = y_{t-1}$.

Stazionarietà

La stazionarietà di un ARMA è determinata dalla componente AR come già visto per i modelli AR(p): le soluzioni dell'equazione caratteristica dell'operatore AR devono essere interne al cerchio unitario:

$$a\left(\frac{1}{z}\right) = 0 \Rightarrow |z| < 1.$$

In questo caso vale la forma MA(∞):

$$y_t = \frac{c(B)}{a(B)} \varepsilon_t = \sum_{j=1}^{\infty} \omega_j \varepsilon_{t-j} + \varepsilon_t$$

Stazionarietà

La stazionarietà di un ARMA è determinata dalla componente AR come già visto per i modelli AR(p): le soluzioni dell'equazione caratteristica dell'operatore AR devono essere interne al cerchio unitario:

$$a\left(\frac{1}{z}\right) = 0 \Rightarrow |z| < 1.$$

In questo caso vale la forma MA(∞):

$$y_t = \frac{c(B)}{a(B)} \varepsilon_t = \sum_{j=1}^{\infty} \omega_j \varepsilon_{t-j} + \varepsilon_t$$

56 / 1

Invertibilità

Un modello ARMA si dice invertibile sse le soluzioni dell'equazione caratteristica dell'operatore MA sono interne al cerchio unitario:

$$c\left(\frac{1}{z}\right) = 0 \Rightarrow |z| < 1.$$

In questo caso vale la forma AR(∞):

$$\varepsilon_t = \frac{a(B)}{c(B)} y_t = y_t - \sum_{j=1}^{\infty} \psi_j y_{t-j}$$

57 / 1

Invertibilità

Un modello ARMA si dice invertibile sse le soluzioni dell'equazione caratteristica dell'operatore MA sono interne al cerchio unitario:

$$c\left(\frac{1}{z}\right) = 0 \Rightarrow |z| < 1.$$

In questo caso vale la forma AR(∞):

$$\varepsilon_t = \frac{a(B)}{c(B)} y_t = y_t - \sum_{j=1}^{\infty} \psi_j y_{t-j}$$

57 / 1

Previsione AR

La previsione a un passo di un modello AR(p) è data da:

$$\hat{y}_t = \hat{y}_t(1) = E(y_t | \mathbf{S}_{t-1}) = a_1 y_{t-1} + \dots + a_p y_{t-p}$$

Proprietà

1. L'errore di previsione

$$e_t = y_t - \hat{y}_t = \varepsilon_t$$

è anche l'**Innovazione** del PS.

2. Linearità del previsore

3. Omoschedasticità: la previsione ha sempre la stessa incertezza

58 / 1

Previsione AR

La previsione a un passo di un modello AR(p) è data da:

$$\hat{y}_t = \hat{y}_t(1) = E(y_t | \mathbf{S}_{t-1}) = a_1 y_{t-1} + \dots + a_p y_{t-p}$$

Proprietà

1. L'errore di previsione

$$e_t = y_t - \hat{y}_t = \varepsilon_t$$

è anche l'**Innovazione** del PS.

2. Linearità del previsore
3. Omoschedasticità: la previsione ha sempre la stessa incertezza

58 / 1

Previsione AR a più passi

Per i modelli lineari in generale, e in particolare per gli ARMA, la previsione a più passi è ricorsiva.

Se l'orizzonte previsivo è $k = 2$, abbiamo:

$$\hat{y}_t(2) = a_1 \hat{y}_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p}$$

Più in generale

$$\hat{y}_t(k) = a_1 \hat{y}_{t-1}(k-2) + a_2 \hat{y}_{t-2}(k-1) + \dots + a_p \hat{y}_{t-p}(k-p).$$

NB: Se il PS è non lineare questo approccio può divergere.

59 / 1

Previsione AR a più passi

Per i modelli lineari in generale, e in particolare per gli ARMA, la previsione a più passi è ricorsiva.

Se l'orizzonte previsivo è $k = 2$, abbiamo:

$$\hat{y}_t(2) = a_1 \hat{y}_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p}$$

Più in generale

$$\hat{y}_t(k) = a_1 \hat{y}_{t-1}(k-2) + a_2 \hat{y}_{t-2}(k-1) + \dots + a_p \hat{y}_{t-p}(k-p).$$

NB: Se il PS è non lineare questo approccio può divergere.

59 / 1

Previsione AR a più passi

Per i modelli lineari in generale, e in particolare per gli ARMA, la previsione a più passi è ricorsiva.

Se l'orizzonte previsivo è $k = 2$, abbiamo:

$$\hat{y}_t(2) = a_1 \hat{y}_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p}$$

Più in generale

$$\hat{y}_t(k) = a_1 \hat{y}_{t-1}(k-2) + a_2 \hat{y}_{t-2}(k-1) + \dots + a_p \hat{y}_{t-p}(k-p).$$

NB: Se il PS è non lineare questo approccio può divergere.

59 / 1

Previsione ARMA

Nel caso di ARMA invertibile la previsione a un passo è data dalla seguente ricorsione

$$\hat{y}_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + c_1 e_{t-1} + \dots + c_q e_{t-q}$$

$$e_t = y_t - \hat{y}_t$$

con valori iniziali

$$e_0, \dots, e_{1-q} = 0.$$

Affinché la ricorsione converga per $t \rightarrow \infty$, occorre che l'operatore MA sia invertibile. Allora

$$e_t = y_t - \hat{y}_t \cong y_t - \sum_{j=1}^{\infty} \psi_j y_{t-j} = \varepsilon_t$$

60 / 1

Radici Unitarie e modelli ARIMA

Se l'equazione caratteristica dell'operatore AR si annulla per $z = 1$:

$$a(1) = 0,$$

allora il modello ARMA è non stazionario.

In questi casi si fattorizza il termine con la radice unitaria e si ottiene un modello ARIMA che ha la forma:

$$a(B)(1 - B)y_t = c(B)\varepsilon_t.$$

In sostanza si usano le differenze prime:

$$(1 - B)y_t = y_t - y_{t-1} \equiv \text{ARMA stazionario}$$

61 / 1

Test per radici unitarie

Test di Dickey-Fuller

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

$$H_0 : \alpha = 1, H_1 : \alpha < 1$$

per $n \rightarrow \infty$: $n(\hat{\alpha}_{OLS} - 1) \rightarrow v.c. \text{ nota sotto } H_0$

dove

$$\hat{\alpha}_{OLS} = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2}$$

62 / 1

Stagionalità deterministica

La serie storica ha una periodicità regolare:

$$y_t = m(t) + \varepsilon_t$$

con $m(t)$ periodica di periodo T

Esempio

$$y_t = \alpha \cos(2\pi t/P) + \beta \sin(2\pi t/P) + \varepsilon_t$$

IGRA_Media_clu14_Mensili.mat IGRA_periodicita.m

63 / 1

Stagionalità deterministica

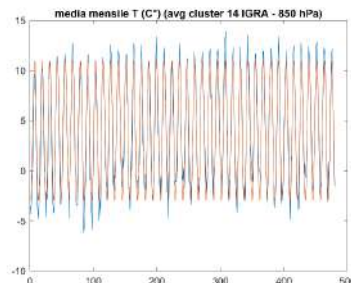
La serie storica ha una periodicità regolare:

$$y_t = m(t) + \varepsilon_t$$

con $m(t)$ periodica di periodo T

Esempio

$$y_t = \alpha \cos(2\pi t/P) + \beta \sin(2\pi t/P) + \varepsilon_t$$



IGRA_Media_clu14_Mensili.mat IGRA_periodicita.m

63 / 1

Modelli ARMA stagionali

$$y_t = \frac{a(B)}{c(B)} \frac{A(B)}{C(B)} \varepsilon_t$$

dove $A(B)$ e $C(B)$ sono la componente stagionale.

Per esempio una componente SAR(1) è data:

$$A(B) = 1 - A_1 B^P$$

e un modello SAR(1) è dato da:

$$y_t = A_1 y_{t-P} + \varepsilon_t.$$

Inoltre un modello AR(1)×SAR(1) è dato:

$$y_t = a_1 y_{t-1} + A_1 y_{t-P} + a_1 A_1 y_{t-P-1} + \varepsilon_t.$$

64 / 1

Modelli ARMA stagionali

$$y_t = \frac{a(B)}{c(B)} \frac{A(B)}{C(B)} \varepsilon_t$$

dove $A(B)$ e $C(B)$ sono la componente stagionale.

Per esempio una componente SAR(1) è data:

$$A(B) = 1 - A_1 B^P$$

e un modello SAR(1) è dato da:

$$y_t = A_1 y_{t-P} + \varepsilon_t.$$

Inoltre un modello AR(1)×SAR(1) è dato:

$$y_t = a_1 y_{t-1} + A_1 y_{t-P} + a_1 A_1 y_{t-P-1} + \varepsilon_t.$$

64 / 1

Modelli ARMA stagionali

$$y_t = \frac{a(B)}{c(B)} \frac{A(B)}{C(B)} \varepsilon_t$$

dove $A(B)$ e $C(B)$ sono la componente stagionale.

Per esempio una componente SAR(1) è data:

$$A(B) = 1 - A_1 B^P$$

e un modello SAR(1) è dato da:

$$y_t = A_1 y_{t-P} + \varepsilon_t.$$

Inoltre un modello AR(1)×SAR(1) è dato:

$$y_t = a_1 y_{t-1} + A_1 y_{t-P} + a_1 A_1 y_{t-P-1} + \varepsilon_t.$$

64 / 1

Esempio COVID

- ▶ Nuovi casi COVID in Italia: Periodicità settimanale.
- ▶ File: COVID_Italia.mat
- ▶ Script: ese_SARMA_COVID.m

65 / 1

Stima MLE (estimate)

La stima di massima verosimiglianza dei modelli ARIMA riguarda il vettore di parametri

$$\theta = (a_1, \dots, a_p, c_1, \dots, c_q, \sigma_\varepsilon^2)^t$$

allora la stima del modello è data da:

$$\hat{\theta}_{MLE} = \arg \max(\ln L_n(\theta))$$

Proprietà asintotiche: per n grande abbiamo:

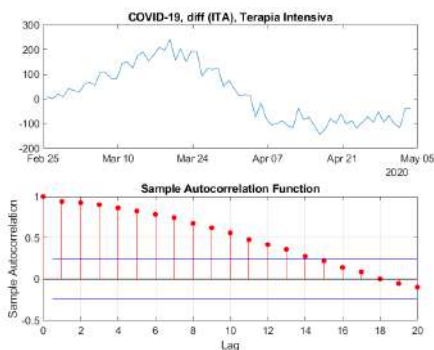
$$(\hat{\theta} - \theta_0) \approx N(0, V_n)$$

dove

$$V_n = \left(\frac{\partial^2}{\partial \theta \partial \theta'} \ln L_n(\theta) \right)^{-1}$$

66 / 1

COVID-19: differenze prime



- ▶ Differenze prime con forte autocorrelazione

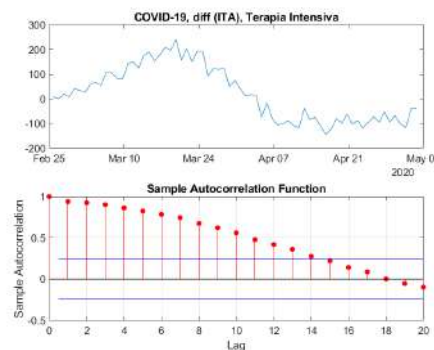
▶ $\Rightarrow \text{ARIMA}(2,1,2)^a$

▶ Matlab:
`>>m0=arima(2,1,2);`

^aModello non ottimizzato

67 / 1

COVID-19: differenze prime



- ▶ Differenze prime con forte autocorrelazione

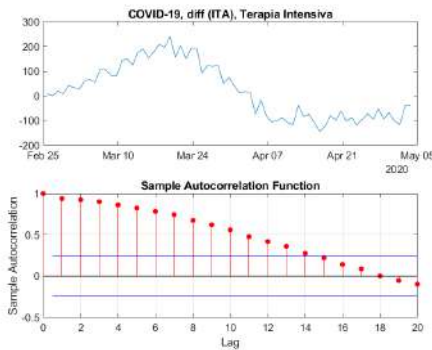
▶ $\Rightarrow \text{ARIMA}(2,1,2)^a$

▶ Matlab:
`>>m0=arima(2,1,2);`

^aModello non ottimizzato

67 / 1

COVID-19: differenze prime

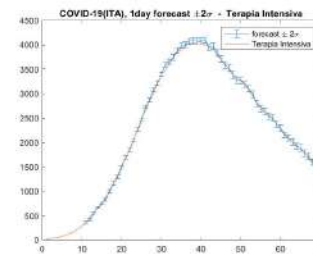


- Differenze prime con forte autocorrelazione
- $\Rightarrow \text{ARIMA}(2,1,2)^a$
- Matlab:
`>>m0=arima(2,1,2);`

^aModello non ottimizzato

67 / 1

COVID-19: Previsione



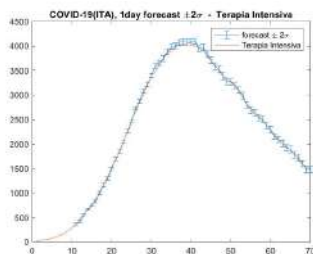
Stima e previsione iterativa ($t=11:n$)

```
>>mhat{t-1}=estimate(m0,y(1:t-1));
>>[yhat(t),ymse(t)]=...
    forecast(mhat{t-1},y(1:t-1));
>>errorbar(yhat,2*sqrt(ymse))
```

Forecast performance: $\text{rmse}=38.6$, $\text{rrmse}=2\%$

68 / 1

COVID-19: Previsione



Stima e previsione iterativa ($t=11:n$)

```
>>mhat{t-1}=estimate(m0,y(1:t-1));
>>[yhat(t),ymse(t)]=...
    forecast(mhat{t-1},y(1:t-1));
>>errorbar(yhat,2*sqrt(ymse))
```

Forecast performance: $\text{rmse}=38.6$, $\text{rrmse}=2\%$

68 / 1

69 / 1

Regressione con errori ARIMA

Consideriamo il modello

$$y_t = m(x_t) + e_t$$
$$a(B)e_t = c(B)\varepsilon_t.$$

Casi importanti sono:

- ▶ Regressione lineare con errori correlati:

$$m(x) = \beta'x$$

- ▶ Modelli nonlineari con errori correlati. Per esempio

$$m(x) = \text{spline}(x)$$

70 / 1

OSSERVAZIONE

Se $a(\frac{1}{z}) = 0$ ha solo radici interne al cerchio unitario, il modello di regressione lineare a errori ARMA può essere scritto come:

$$y_t = \beta'x_t + \frac{c(B)}{a(B)}\varepsilon_t$$

71 / 1

Esempio: trend lineare su dati giornalieri

$$y_t = 2 + \frac{1.5}{365}t + \frac{\varepsilon_t}{1 - .75B + .5B^2}, \quad \varepsilon_t \text{ iid } t(5)$$

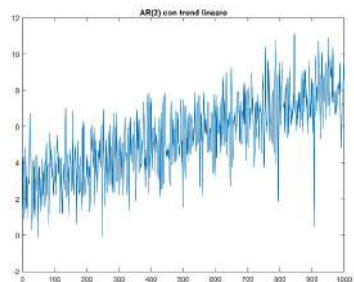


Figura: Trend lineare con errori AR(2)

Esercizio: Valutare se l'errore è stazionario

72 / 1

Esempio: trend lineare su dati giornalieri

In Matlab il modello

$$y_t = 2 + \frac{1.5}{365}t + \frac{\varepsilon_t}{1 - .75B + .5B^2}, \quad \varepsilon_t \text{ iid } t(5)$$

si implementa in classe regARIMA:

```
>>m=regARIMA('Intercept',2,'Beta',1./365,'AR',[.75 -.5]);  
>>e_distr.Name='t';  
>>e_distr.DoF=5;  
>>m.Distribution=e_distr;
```

73 / 1

ARMAX e regARIMA

La classe `arima()` di Matlab implementa i modelli ARMAX che hanno la forma:

$$a(B)y_t = \beta'x_t + (B)\varepsilon_t$$

Questi corrispondono ai modelli regARIMA tramite l'equazione:

$$y_t = \beta' \frac{1}{a(B)} x_t + \frac{c(B)}{a(B)} \varepsilon_t$$

74 / 1

ESEMPIO

Il modello AR(1) con intercetta:

$$(1 - aB)y_t = k + \varepsilon_t$$

corrisponde al seguente modello regARIMA:

$$y_t = \frac{k}{1 - a} + \frac{1}{a(B)} \varepsilon_t$$

Si noti che la media è

$$E(y_t) = \frac{k}{1 - a}$$

75 / 1

Determinazione di (p,q)

Penalizzazione per la complessità:

► Criterio BIC $-2\log L + n_{par} \log(n)$

► Criterio AIC $-2\log L + 2n_{par}$

Matlab:

```
[m_hat,V_hat,logL,info] = estimate(m0,y);  
n = length(y);  
npar = sum(any(V_hat))-1  
[aic,bic] = aicbic(logL,npar,n)
```

76 / 1

77 / 1

Determinazione di (p,q)

Penalizzazione per la complessità:

- Criterio BIC $-2\log L + n_{par}\log(n)$
- Criterio AIC $-2\log L + 2n_{par}$

Matlab:

```
[m_hat,V_hat,logL,info] = estimate(m0,y);  
n = length(y);  
npar = sum(any(V_hat))-1  
[aic,bic] = aicbic(logL,npar,n)
```

77 / 1

Determinazione di (p,q)

Penalizzazione per la complessità:

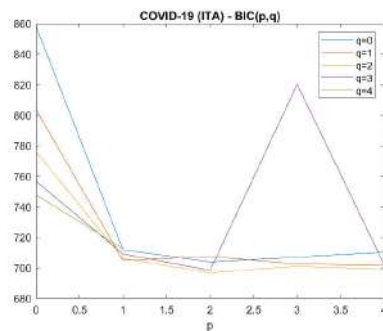
- Criterio BIC $-2\log L + n_{par}\log(n)$
- Criterio AIC $-2\log L + 2n_{par}$

Matlab:

```
[m_hat,V_hat,logL,info] = estimate(m0,y);  
n = length(y);  
npar = sum(any(V_hat))-1  
[aic,bic] = aicbic(logL,npar,n)
```

77 / 1

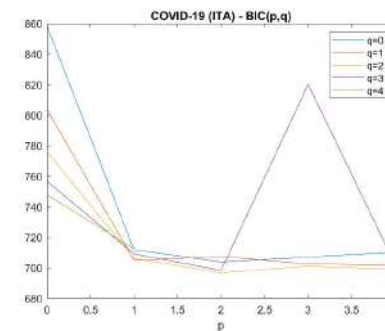
COVID-19: scelta modello



- Minimizzazione BIC:
ARIMA(2,1,2)
- Altri modelli simili
- Validazione esterna: ricerca modelli che minimizza RMSE di previsione

78 / 1

COVID-19: scelta modello



- Minimizzazione BIC:
ARIMA(2,1,2)
- Altri modelli simili
- Validazione esterna: ricerca modelli che minimizza RMSE di previsione

78 / 1

Esercizio aic-bic per AR

1. Simulare $n = 750$ osservazioni di un modello AR(2) con

$$a(B) = 1 - .75B + .5B^2$$

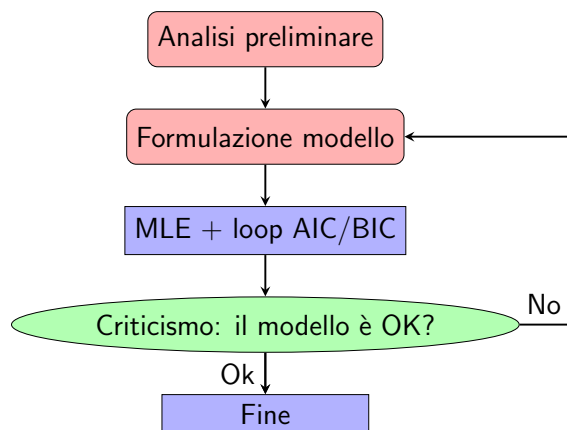
e $\sigma^2 = 0.001$;

2. calcolare e plottare aic e bic per $p = 0, \dots, 4$;
3. individuare l'ordine ottimale secondo i due criteri aic e bic .

79 / 1

Diagnostica e validazione

Flowchart modellazione



81 / 1

Diagnostica e validazione

La fase del "criticismo" si applica al modello individuato dalla procedura automatica e consiste nell'analizzare:

- ▶ la significatività statistica e sostanziale dei coefficienti individuati;
- ▶ le proprietà del modello (stazionarietà, linearità etc)
- ▶ simulazione di traiettorie e bootstrap
- ▶ le proprietà dei residui: **Diagnostica**;
- ▶ la capacità previsiva: **Validazione**.

Trattiamo nel seguito la **Diagnostica** dei residui, la **Validazione** del modello ed il **Bootstrap**.

82 / 1

Diagnostica e validazione

La fase del "criticismo" si applica al modello individuato dalla procedura automatica e consiste nell'analizzare:

- ▶ la significatività statistica e sostanziale dei coefficienti individuati;
- ▶ le proprietà del modello (stazionarietà, linearità etc)
- ▶ simulazione di traiettorie e bootstrap
- ▶ le proprietà dei residui: **Diagnostica**;
- ▶ la capacità previsiva: **Validazione**.

Trattiamo nel seguito la **Diagnostica** dei residui, la **Validazione** del modello ed il **Bootstrap**.

82 / 1

Diagnostica dei residui

La diagnostica dei residui valuta se le assunzioni fatte sui residui (e sul modello) siano soddisfatte e, in caso contrario, dà indicazioni sulle modifiche da apportare al modello.

Ci si chiede se i residui si comportino come una realizzazione di VC IID, stanti i vincoli legati alla procedura di ottimizzazione:

- ▶ L'autocorrelazione è zero?
- ▶ La varianza è costante?
- ▶ I residui sono IID?
- ▶ I residui sono gaussiani?
- ▶ Ci sono outliers?

La diagnostica dei residui in genere si svolge nel set di stima (learning) ma può essere anche fatta nel set di validazione.

83 / 1

Diagnostica dei residui

La diagnostica dei residui valuta se le assunzioni fatte sui residui (e sul modello) siano soddisfatte e, in caso contrario, dà indicazioni sulle modifiche da apportare al modello.

Ci si chiede se i residui si comportino come una realizzazione di VC IID, stanti i vincoli legati alla procedura di ottimizzazione:

- ▶ L'autocorrelazione è zero?
- ▶ La varianza è costante?
- ▶ I residui sono IID?
- ▶ I residui sono gaussiani?
- ▶ Ci sono outliers?

La diagnostica dei residui in genere si svolge nel set di stima (learning) ma può essere anche fatta nel set di validazione.

83 / 1

Diagnostica dei residui

La diagnostica dei residui valuta se le assunzioni fatte sui residui (e sul modello) siano soddisfatte e, in caso contrario, dà indicazioni sulle modifiche da apportare al modello.

Ci si chiede se i residui si comportino come una realizzazione di VC IID, stanti i vincoli legati alla procedura di ottimizzazione:

- ▶ L'autocorrelazione è zero?
- ▶ La varianza è costante?
- ▶ I residui sono IID?
- ▶ I residui sono gaussiani?
- ▶ Ci sono outliers?

La diagnostica dei residui in genere si svolge nel set di stima (learning) ma può essere anche fatta nel set di validazione.

83 / 1

Test di Ljung-Box

Il test di Box-Pierce/Ljung-Box verifica l'ipotesi nulla che il modello sia correttamente specificato e, quindi, che i residui corrispondano a delle innovazioni IID

$$Q_h^2(e) = n \sum_{j=1}^h r_e(j)^2 \frac{n+2}{n-j} \approx \begin{cases} \chi_{h-(p^*+q^*)}^2 & \text{Learning set} \\ \chi_h^2 & \text{Testing set} \end{cases}$$

Dove p^* e q^* sono il numero di coefficienti stimati della parte AR ed MA.

Nell'analisi preliminare il test LB può essere applicato ai dati stessi, y_t , per testare l'ipotesi che siano IID. In questo caso non vi è perdita di gradi di libertà.

84 / 1

Test di Ljung-Box

Il test di Box-Pierce/Ljung-Box verifica l'ipotesi nulla che il modello sia correttamente specificato e, quindi, che i residui corrispondano a delle innovazioni IID

$$Q_h^2(e) = n \sum_{j=1}^h r_e(j)^2 \frac{n+2}{n-j} \approx \begin{cases} \chi_{h-(p^*+q^*)}^2 & \text{Learning set} \\ \chi_h^2 & \text{Testing set} \end{cases}$$

Dove p^* e q^* sono il numero di coefficienti stimati della parte AR ed MA.

Nell'analisi preliminare il test LB può essere applicato ai dati stessi, y_t , per testare l'ipotesi che siano IID. In questo caso non vi è perdita di gradi di libertà.

84 / 1

Omoschedasticità e outliers

Si vuole capire se

$$\text{var}(\varepsilon_t) = \sigma^2 = \text{cost}$$

oppure:

- dipende da t
- dipende dal passato.

Nel seguito consideriamo solo il primo punto.

L'identificazione degli outliers, nel caso più semplice, è basata sui valori estremi dei residui.

85 / 1

Omoschedasticità e outliers

Si vuole capire se

$$\text{var}(\varepsilon_t) = \sigma^2 = \text{cost}$$

oppure:

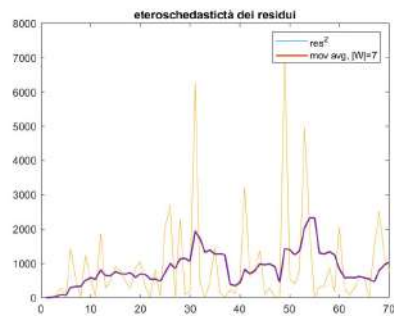
- dipende da t
- dipende dal passato.

Nel seguito consideriamo solo il primo punto.

L'identificazione degli outliers, nel caso più semplice, è basata sui valori estremi dei residui.

85 / 1

COVID-19: diagnostica residui 2



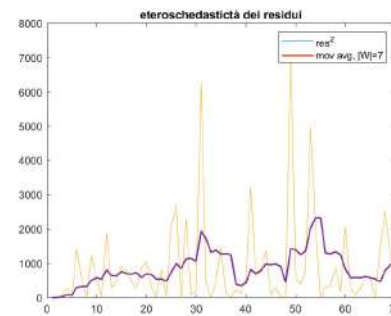
- ▶ Eteroschedasticità moderata:
 - fase iniziale a variabilità ridotta (conteggi)
 - effetto mitigato da previsione iterativa

▶ Matlab:
`>>plot(movavg(res2,'a'))`

`'a','simple',7`

86 / 1

COVID-19: diagnostica residui 2



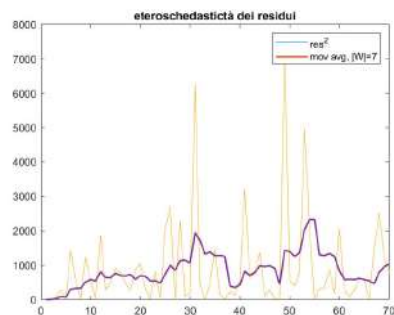
- ▶ Eteroschedasticità moderata:
 - fase iniziale a variabilità ridotta (conteggi)
 - effetto mitigato da previsione iterativa

▶ Matlab:
`>>plot(movavg(res2,'a'))`

`'a','simple',7`

86 / 1

COVID-19: diagnostica residui 2



- ▶ Eteroschedasticità moderata:
 - fase iniziale a variabilità ridotta (conteggi)
 - effetto mitigato da previsione iterativa

▶ Matlab:
`>>plot(movavg(res2,'a'))`

`'a','simple',7`

86 / 1

87 / 1

Validazione e capacità previsiva

La validazione è prevalentemente orientata a valutare la capacità previsiva del modello individuato. Il significato assume sfumature diverse in relazione all'obiettivo del modello stesso:

1. Descrittivo
2. Previsivo
3. Qualità del dato
4. Monitoraggio e/o sorveglianza

Poiché individueremo più di un criterio, l'eventuale ottimizzazione dovrà, appunto, essere multi-criterio.

88 / 1

Validazione e capacità previsiva

La validazione è prevalentemente orientata a valutare la capacità previsiva del modello individuato. Il significato assume sfumature diverse in relazione all'obiettivo del modello stesso:

1. Descrittivo
2. Previsivo
3. Qualità del dato
4. Monitoraggio e/o sorveglianza

Poiché individueremo più di un criterio, l'eventuale ottimizzazione dovrà, appunto, essere multi-criterio.

88 / 1

Validazione interna ed esterna

Nella **validazione interna** la capacità previsiva viene valutata sugli stessi dati usati per la stima. Questo approccio può essere esposto al rischio di sovra-ottimismo e overfitting.

Nella **validazione esterna** chiamata cross-validazione (CV) in senso lato si usano dati diversi per stima e valutazione della capacità previsiva:

- ▶ CV classica (k-fold): la randomizzazione crea problemi per dati equispaziati nel tempo.
- ▶ CV Leave-one-out
- ▶ CV bootstrap
- ▶ CV incrementale (cfr previsione COVID-19)

89 / 1

Validazione interna ed esterna

Nella **validazione interna** la capacità previsiva viene valutata sugli stessi dati usati per la stima. Questo approccio può essere esposto al rischio di sovra-ottimismo e overfitting.

Nella **validazione esterna** chiamata cross-validazione (CV) in senso lato si usano dati diversi per stima e valutazione della capacità previsiva:

- ▶ CV classica (k-fold): la randomizzazione crea problemi per dati equispaziati nel tempo.
- ▶ CV Leave-one-out
- ▶ CV bootstrap
- ▶ CV incrementale (cfr previsione COVID-19)

89 / 1

Capacità previsiva

Indici di capacità previsiva per $y = \hat{y} + e$

	Assoluti	Relativi	(Pseudo) Normalizzati
Bias	$\frac{-1}{n} \sum e$	$\frac{-1}{n} \sum \frac{e}{y}$	
Corr			$r(y, \hat{y})$
MSE	$\frac{1}{n} \sum e^2$		$R^2 = 1 - \frac{MSE}{MSE_0}$
RMSE	$\sqrt{\frac{1}{n} \sum e^2}$	$\sqrt{\frac{1}{n} \sum (\frac{e}{y})^2}$	
MAE	$\frac{1}{n} \sum e $	$\frac{1}{n} \sum \frac{e}{y} $	$1 - MAE/MAE_0$

$$MSE_0 \leq \sigma_y^2, MAE_0 \leq MAD$$

90 / 1

COVID-19 Performance analysis

Valutazione delle prestazioni del modello rispetto al modello "oggi come ieri" (RW).

	bias	rmse	rrmse
ARIMA	8.73	38.6	0.02
ARIMA-Ante picco	12.91		
ARIMA-Post picco	5.07		
RW	-21.38	110.7	0.12
RW-Ante picco	-108.3		
RW-Post picco	82.32		

91 / 1

Change detection

Consideriamo il caso più semplice dove il cambiamento riguarda il livello di un ARMA:

$$a(B)y_t = \theta_t + x_t' \beta + c(B)\varepsilon_t$$

$$\text{In generale abbiamo: } \theta_t = \begin{cases} \mu & t \in T_1 \\ \mu + \delta & t \in T_2 \end{cases}$$

- ▶ cambiamenti a gradino: $T_2 = (t^*, \dots, n)$, $T_1 = N - T_2$
- ▶ cambiamenti a impulso: $T_2 = t^*$, $T_1 = N - T_2$
- ▶ cambiamenti temporanei: $T_2 = (t_a, \dots, t_b)$, $T_1 = N - T_2$
- ▶ cambiamenti osservazionali: $y_t = \theta_t + x_t' \beta + \frac{a(B)}{c(B)} \varepsilon_t$

92 / 1

Change detection

Altri modelli di cambiamento:

- ▶ Funzione di regressione: $\theta_t = \beta_t$
- ▶ Varianza: $\theta_t = \text{Var}(\varepsilon_t)$
- ▶ Autocorrelazione: $\theta_t = (a_1, \dots, a_p, c_1, \dots, c_q)_t$

93 / 1

Statistica II

L09: ARCH/GARCH

aa 2020-2021
Alessandro Fassò

1/4

Introduzione

$$y_t = \mu_t + \varepsilon_t$$

Trattiamo in questo capitolo i modelli per la varianza condizionata:

$$\sigma_t^2 = \text{Var}(\varepsilon_t | S_{t-1})$$

dove ε_t è incorrelato.

Tipicamente questo è il caso dei residui di un modello per la media (ARMA o simili).

Questi modelli sono stati introdotti dal premio Nobel Engle per l'analisi di dati finanziari ma si applicano a dati - in genere in alta frequenza - anche di altra natura.

HET σ_t^2 ? \swarrow \searrow $\text{funct. of } t, x_t$ "DETERM"
 stochastic

2/4

Modelli ARCH(q)

I modelli ARCH (AutoRegressive Conditionally Heteroskedastic) descrivono la varianza condizionata tramite un modello AR per i quadrati degli errori:

$$y_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \kappa + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

dove $z_t \text{ iid } N(0, 1)$ oppure t_ν standardizzata.

3/4

Modelli GARCH(p,q)

I modelli ARCH generalizzati (GARCH) descrivono la varianza condizionata tramite un modello di tipo ARMA:

$$\sigma_t^2 = \kappa + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

Dominio

$$\kappa, \alpha_j, \beta_j > 0$$

Stazionarietà

$$\sum_{j=1}^p \beta_j + \sum_{j=1}^q \alpha_j < 1$$

NB: la ricorsione su σ_t consente di modellare la persistenza della varianza analogamente a quanto fatto dalla parte AR di un modello ARMA.

4/4

Modelli GARCH(p,q)

I modelli ARCH generalizzati (GARCH) descrivono la varianza condizionata tramite un modello di tipo ARMA:

$$\sigma_t^2 = \kappa + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

Dominio

$$\kappa, \alpha_j, \beta_j > 0$$

Stazionarietà

$$\sum_{j=1}^p \beta_j + \sum_{j=1}^q \alpha_j < 1$$

NB: la ricorsione su σ_t consente di modellare la persistenza della varianza analogamente a quanto fatto dalla parte AR di un modello ARMA.

Modelli GARCH(p,q)

I modelli ARCH generalizzati (GARCH) descrivono la varianza condizionata tramite un modello di tipo ARMA:

$$\sigma_t^2 = \kappa + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

Dominio

$$\kappa, \alpha_j, \beta_j > 0$$

Stazionarietà

$$\sum_{j=1}^p \beta_j + \sum_{j=1}^q \alpha_j < 1$$

NB: la ricorsione su σ_t consente di modellare la persistenza della varianza analogamente a quanto fatto dalla parte AR di un modello ARMA.