# Computational Statistics

## Prof. Raffaele Argiento; Prof. Ioannis Ntzoufras

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

This courses covers most topics needed to develop a broad and thorough working knowledge of modern computational statistics. The aim is to develop a practical understanding of how and why existing methods work, enabling the students to use modern statistical methods effectively. Since many modern computational methods in Statistics and Data Analytics are built from components of existing techniques, the ultimate goal of the course is to provide students with the tools they will need to learn and/or develop new computational methods in their educational path as well as professional career.

By the end of the course, students are expected to:

1. have acquired advanced knowledge of the computational techniques used to make inference under modern statistical models and be able to discern which computational technique they must use when dealing with a given statistical problem/model. (*knowledge and understanding*);
2. be able to implement a computational algorithm for a broad range of statistical models. In particular, they will be able to implement an ad hoc optimization method for ML estimation or an Expectation Maximization algorithm. They will know the fundamentals of Monte Carlo Integration as well as Bootstrap Inference and Cross-Validation. Students will be able to apply this quantitative tools to problems arising in Economics and Business (A*pplying knowledge and understanding*);
3. be able to check convergence as well as assess the performances of the studied algorithms. Moreover, they will be able to compare different computational approaches applied to the same statistical problem/model (M*aking judgements*);
4. be able to describe with an appropriate language a computational methods as well as the mathematical and statistical assumptions behind it. Moreover, they are expected to be capable of communicating the result of a computational algorithm to general audience (C*ommunication skills*);
5. have acquired the main concepts and tools to independently learn and/or develop new computational methods for Business Analytics (*Lifelong learning skills*).

***COURSE CONTENT***

Module 1*Iannis Ntzoufras*

* *Methods for generating random variables:* Transformation methods – The inverse transform method – The acceptance-rejection method
* *Monte Carlo integration:* Variance reduction – Control variates – Importance sampling – Monte Carlo based inference
* *Bootstrapping:* The Bootstrap principle – Basic methods nonparametric and parametric bootstrap – Bootstrap inference, confidence intervals

Module 2 *Raffaele Argiento*

* *Bootstrapping:* Bootstrap inference hypothesis tests, Bootstrapping regression – Bootstrap bias correction*.*
* *Optimization:* A gentle introduction to univariate and multivariate optimization problems for Statistical Inference – The Newton-Raphson method to compute the Maximum Likelihood estimator.
* *Em optimization methods:* Missing Data and marginalization – The EM Algorithm: convergence and the usage of EM for inference in the exponential families – The EM for mixture models.

***READING LIST[[1]](#footnote-1)***

Class notes, slides, papers, coding and further material will be posted on the University platform Blackboard.

Useful readings are:

* Geof H. Givens Jennifer A. Hoeting (2012) Computational Statistics, Second Edition
* Rizzo, M. L. (2019). Statistical computing with R. CRC Press
* Jochen Voss (2013) An Introduction to Statistical Computing – A Simulation-based Approach-Wiley

***TEACHING METHOD***

A blend of lectures and coding (60 hours), exercise sessions and lab-sessions on R (20 hours). Attending the lectures, active participation and ongoing personal study are strongly recommended.

***ASSESSMENT METHOD AND CRITERIA***

The assessment is based on two parts:

1. Two assignments: consisting in two work-home data analysis project implemented in R and carried out individually by the student;
2. Written examination with open-ended questions on methods, exercises and R-language.

The final grade is equally based on the two parts.

***NOTES AND PREREQUISITES***

Students enrolling in this course are expected to know data analysis, probability and frequentist inference, at the level of Statistics courses usually taught in a bachelor degree in Economics; see for instance the topics covered in ‘Statistica (analisi dei dati e probabilità) and ‘Statistica applicata’ (or ‘Statistics’ and ‘Applied Statistics’) at this University. They should also have a fair knowledge of the R-language. It is also recommended that they have attended the following courses: Mathematical methods and probability, Statistical inference and Applied linear models taught within this MSc programme.

If the COVID-19 public health emergency would not allow face-to-face teaching, distance learning will be guaranteed. In this eventuality, teaching timing and modalities will be promptly communicated to students.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)