# Computational Statistics

## Prof. Augusto Fasano

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

This course 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. Furthermore, since many modern computational methods in Statistics and Data Analytics are built from components of existing techniques, the goal of the course is to provide students with the tools they will need to learn and design new computational methods in their educational path and 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. They will be able to implement an ad hoc optimization method for Maximum Likelihood estimation or an Expectation-Maximization algorithm. They will know the fundamentals of Monte Carlo Integration and Simulation and Bootstrap Inference. Students will be able to apply these quantitative tools to problems arising in Economics and Business (Applying 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 (Making judgments);

4. be able to describe with appropriate language a computational method 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 a general audience, develop the ability to work in teams to complete a project, address complex tasks, and improve their problem-solving skills (Communication 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*(Augusto Fasano)*

* **The concept of simulation**: formal introduction to the problem - how to use the computer to approximate probabilities of interest
* **Methods for generating random variables**: Uniform simulation - Transformation methods – The inverse transform method – The acceptance-rejection method
* **Monte Carlo integration**: MC crude method - Variance reduction techniques

## Module 2*(Augusto Fasano)*

* **Bootstrapping**: Bootstrap inference -- Bootstrapping regression – Parametric bootstrap for generalized linear models – Bootstrapping for dependent data and time series – Jack-knife and permutation test
* **Optimization**: Univariate and multivariate optimization problems for Statistical Inference
* **EM optimization methods**: the concept of auxiliary variable – the EM Algorithm and its convergence – 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. Ground-breaking scientific and review papers in the field of Computational Statistics will be proposed as well.

Useful readings are:

* Geof H. Givens Jennifer A. Hoeting (2012) Computational Statistics, Second Edition
* Rizzo, M. L. (2019). Statistical computing with R. CRC Press
* Robert and Casella (2010) Introducing Monte Carlo methods with R. Springer

The reading list may change. Please monitor the course websites for up-to-date information.

***Note: some of the proposed books are freely available online or via the University Library***

***TEACHING METHOD***

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

***ASSESSMENT METHOD AND CRITERIA***

The final assessment will be carried out via an open-ended exam, composed of two parts. The first part (closed-book) consists of theoretical questions and exercises, while in the second part (open-book) students are asked to solve a problem with the support of statistical software by developing R code.

***NOTES AND PREREQUISITES***

Students enrolling in this course are expected to know mathematics, data analysis, probability, and frequentist inference, at the level of “Mathematical methods and probability”, “Statistical inference”, and “Applied linear models”, which are courses taught -- on the first and the second term – within this MSc program. Students should also have a good knowledge of the R language.

*Office hours*

Students can ask for a meeting via email. The meeting can either take place online (via Microsoft Teams) or in the office of the teacher (313, Lanzone 18)

1. The texts listed in the bibliography can be purchased from the University bookstores; they can also be purchased from other retailers. [↑](#footnote-ref-1)