# Computational Statistics

## Prof. Francesco Denti; Prof. Reza Mohammadi

***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 create a practical understanding of how and why existing methods work, enabling the students to use modern statistical methods effectively. Furthermore, since many modern computational approaches in Statistics and Data Analytics are built upon 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 ML 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 an 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*Reza Mohammadi*

## *- Introduction to advanced R programming:* Recap of R basics - Efficient programming - R markdown for producing reports - The Tidyverse package

* *Methods for generating random variables:* Transformation methods – The inverse transform method – The acceptance-rejection method
* *Monte Carlo integration:* Numerical integration – Basic Monte Carlo – Variance reduction techniques – Importance sampling
* *Monte Carlo Markov Chains:* Introduction to Markov Chains – Metropolis-Hastings algorithm – Gibbs sampler

Module 2 *Francesco Denti*

* *Bootstrapping:* Bootstrap inference – Bootstrapping regression – Parametric bootstrap for generalized linear models.
* *Optimization:* Univariate and multivariate optimization problems for Statistical Inference
* *Multiple Hypothesis Testing:* issues with standard approaches – Beyond the control of the first type error – Benjamini-Hochberg method – The false discovery rate and the two-group model
* *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***

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:

* Efron, B., Hastie, T., Computer Age statistical inference
  + https://hastie.su.domains/CASI\_files/PDF/casi.pdf
* 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 with R (20 hours). Attending the classes, active participation, and ongoing personal study are strongly recommended.

***ASSESSMENT METHOD AND CRITERIA***

The assessment method for this course articulates in two parts: weekly homework and a final group project. The final mark will be computed as a weighted average of the scores in the two parts. In detail,

* Students will be required to complete weekly homework, for which dedicated lab sessions are devised. The homework is intended to assess the student’s understanding and participation. (1/3 of the mark, up to 10 points)
* The final group project focuses on the independent implementation and understanding of state-of-the-art techniques in the field of Computational Statistics. The groups are intended to be comprised of 3-4 students, according to the total number of students in the course.
  + Each group will be asked to produce a report that will obtain a score shared across the group components (1/3 of the mark, up to 10 points)
  + There will also be an individual oral examination, where each student will be asked about their presented project and be assessed on their understanding of the concepts presented in class (1/3 of the mark, 5 points on the understanding/participation of the group project + 5 points on the understanding of the course topics)

***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”. Students should also have a basic knowledge of the R language.