# Nonparametric Statistics

**Academic Year: ( 2021 / 2022 )** **Review date: 10-06-2021**

**Department assigned to the subject: Statistics Department**

**Coordinating teacher: GARCIA PORTUGUES, EDUARDO**

**Type: Compulsory ECTS Credits : 3.0**

**Year : 1 Semester : 2**

## REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Mathematics for Data Science
Probability
Statistical Inference
Programming in R
Multivariate Analysis
Regression Models
Advanced Programming

## OBJECTIVES

* Basic competences

- CB6: Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
- CB9: That students know how to communicate their conclusions and the knowledge and ultimate reasons that sustain them to specialised and non-specialised audiences in a clear and unambiguous way.
- CB10: That the students have the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

* General competences

- CG1: Ability to apply the techniques of analysis and representation of information, in order to adapt it to real problems.
- CG4: Synthesise the conclusions obtained from data analyses and present them clearly and convincingly in a bilingual environment (Spanish and English) both written and orally.
- CG5: Be able to generate new ideas (creativity) and anticipate new situations, in the contexts of data analysis and decision making.
- CG6: Use social skills for teamwork and to relate to others autonomously.

* Specific competences

- CE1: Apply in the development of methods of analysis of real problems, advanced knowledge of statistical inference.
- CE2: Use free software such as R and Python for the implementation of statistical analysis.
- CE5: Apply the advanced statistical foundations for the development and analysis of real problems, which involve the prediction of a variable response.
- CE6: Apply nonparametric models for the interpretation and prediction of random phenomena.
- CE10: Apply statistical modeling in the treatment of relevant problems in the scientific field.

## DESCRIPTION OF CONTENTS: PROGRAMME

This course is designed to give a panoramic view of several tools available for nonparametric statistics, at an intermediate-advanced level. This view covers in-depth the main concepts in the estimation of the density and regression functions through kernel methods (with their corresponding applications), and the description of several popular nonparametric tests. The focus is placed on providing the main insights on the statistical/mathematical foundations of the methods and on showing the effective implementation of the methods through the use of statistical software. This is achieved by a mixture of theory and reproducible code.

1. Introduction

1.1. Probability review
1.2. Facts about distributions
1.3. Stochastic convergence review
1.4. OP and oP notation
1.5. Review on basic analytical tools
1.6. Why nonparametric statistics?

2. Kernel density estimation I
2.1. Histograms
2.2. Kernel density estimation
2.3. Asymptotic properties
2.4. Bandwidth selection
2.5. Practical issues
2.6. Kernel density estimation with ks

3. Kernel density estimation II
3.1. Multivariate kernel density estimation
3.2. Asymptotic properties
3.3. Bandwidth selection
3.4. Applications of kernel density estimation

4. Kernel regression estimation I
4.1. Kernel regression estimation
4.2. Asymptotic properties
4.3. Bandwidth selection
4.4. Regressogram
4.5. Kernel regression estimation with np

5. Kernel regression estimation II
5.1. Kernel regression with mixed multivariate data
5.2. Bandwidth selection
5.3. Prediction and confidence intervals
5.4. Local likelihood

6. Nonparametric tests
6.1. Goodness-of-fit tests for distributions
6.2. Comparison of distributions
6.3. Independence tests

The program is subject to modifications due to the course development and/or academic calendar.

## LEARNING ACTIVITIES AND METHODOLOGY

The lessons consist of a mixture of theory (methods description) and practice (implementation and practical usage of methods). The implementation of the methods is done with the statistical language R, so good coding abilities on it are fundamental to understand the implementations. Students are expected to bring their own laptops to experience the code during some parts of the lessons.

## ASSESSMENT SYSTEM

The evaluation in the ordinary call is done entirely by continuous evaluation. This is done by a mixture of:

A) three quizzes about key ideas and theoretical concepts;
B) a set of practical exercises;
C) an active following of the course development.

The continuous evaluation grade (in the scale 0-10) is

$$\min(0.5 * A + 0.5 * B + 0.10 * C, 10)$$

where

- A (in the scale 0-10) is the weighted grade of the quizzes.
- B (in the scale 0-10) is the grade of the practical exercises.
- C (in the scale 0-10) is the grade of active following of the course development.

The grade in the extraordinary call is established by a quiz and the delivery of a set of practical

exercises.

Further details are provided in Aula Global. The evaluation is subject to modifications due to the course development and/or academic calendar.

| | |
|---|---|
| **% end-of-term-examination:** | 0 |
| **% of continuous assessment (assigments, laboratory, practicals…):** | 99 |

## BASIC BIBLIOGRAPHY


- Chacón, J. E. and Duong, T. Multivariate Kernel Smoothing and Its Applications, Chapman and Hall/CRC, 2018
- Fan, J. and Gijbels, I. Local polynomial modelling and its applications, 1996, Chapman & Hall
- Li, Q. and Racine, J. S. Nonparametric Econometrics, Princeton University Press, 2007
- Wand, M. P. and Jones, M. C. Kernel Smoothing, Chapman & Hall, 1995
- Wasserman, L. All of Nonparametric Statistics, Springer-Verlag, 2006
- Wasserman, L. All of Statistics, Springer-Verlag, 2004