UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Probability and Statistics for Computer Science

2627-2-E3102Q114

Aims

Learning Objectives (Dublin Descriptors)

At the end of the course, students will have achieved the following learning outcomes:

1. Knowledge and Understanding. Students will acquire basic knowledge in probability theory and mathematical statistics, including fundamental concepts, definitions, and key results. Particular emphasis will be placed on the most relevant probability distributions and on the foundations of descriptive and inferential statistics.

2. Applying Knowledge and Understanding. Students will be able to apply theoretical concepts to analyze random phenomena, reason under uncertainty, and effectively summarize data sets. They will also be able to perform parameter estimation and hypothesis testing in practical contexts.

3. Making Judgements. Students will develop the ability to critically assess probabilistic and statistical models, selecting appropriate methods for data analysis and interpretation.

4. Communication Skills. Students will be able to communicate concepts and results related to probability and statistics in a clear and rigorous manner, both in written and oral form, using terminology appropriate to the field of computer science.

5. Learning Skills. Students will develop the necessary skills to independently deepen their knowledge of probability and statistics, including through the use of software tools, with particular reference to the R programming language.

Contents

The first part of the course, devoted to **probability theory**, starts with an introduction to the so-called **descriptive statistics** which allows to summarize the salient features of some data set. Subsequently we present **probability spaces**, a key concept in the modeling of random phenomena, along with **random variables** which form its "operating language".

The second part of the course, devoted to **mathematical statistics**, starts with the main **limit theorems** (law of large numbers and central limit theorem) which describe the behavior of random phenomena with a large number of components. Then we discuss **hypothesis testing** and the corresponding statistical tests, concerning both unknown parameters in a random model (*parametric tests*) and the comparison between distributions (*non-parametric tests*). The last part of the course is devoted to the **linear regression**, a powerful technique to investigate the link that might exist between some variable (output) and a given set of variables (input) under uncertainty.

Detailed program

1. Descriptive Statistics

- Introduction to data analysis
- Sample statistics (mean, median, quantiles, variance, correlation)
- Graphical representations

2. Probability spaces

- Random phenomena, probability spaces and events
- Basic properties of probability
- Conditional probability
- Elements of combinatorial calculus
- Independence of events

3. Random variables

- Discrete random variables
- Expectation, moments, variance and covariance
- Absolutely continuous random variables
- Important discrete and absolutely continuous distributions
- Normal random variables

4. Convergence Theorems

- Convergence of random variables and distribuitons (hints)
- Law of Large Numbers
- Central Limit Theorem

5. Parameters Estimation

- Samples and statistics
- Estimators (sample mean and variance)
- Confidence intervals

6. Hypothesis Testing

- Hypothesis testing, I type and II type errors

- Parametric hypothesis testing for mean and variance
- Non parametric hypothesis testing for goodness of fit and independence

7. Linear Regression

- Introduction to linear regression
- Statistical inference for parameters
- Residual analysis

Prerequisites

The knowledge, competences and skills taught in previous mathematical courses, in particular concerning Mathematical Analysis.

Teaching form

The course is given in Italian in *blended-learning*.

Lectures in the classroom are organized as follows:

- 16 x 2 hours of in-person, lecture-based teaching (theoretical lectures, delivered mode) focused on the knowledge of definitions, results, proofs and relevant examples;
- 10 x 2 hours of in-person, lecture-based exercise classes (recitations, delivered mode) focused on the skills necessary to apply the theoretical knowledge and competencies to the solution of exercises.

The blended part of the course (computer workshops with R) consists of on-line practical sessions with the software R in interactive mode, focused on the skills required to do statistical analysis with R. This activity is organized into seven distinct workshops (one for each topic of the course) which require a commitment of approximately 3 hours 30 minutes per session.

Textbook and teaching resource

Reference book:

- S. M. Ross (2014), Introduzione alla Statistica, seconda edicione, Apogeo Editore

Moreover, the following teaching material is made avaliable:

- Theacher's notes
- Exercise sheets

Computer workshops with R

- video lectures and slides

Semester

Spring term (Second semester)

Assessment method

The exam consists of a written part and of a possible oral part. It receives a mark out of 30.

The written part consists of two parts:

- a first part of closed-ended questions, which contributes one third to the final mark, where theoretical skills are tested;
- a second part of open questions, which contributes two thirds to the final mark, where practical skills are tested.

A midterm exam is scheduled halfway through the course. The first exam session coincides with the second midterm and is reserved for students who passed the first one.

The oral part is optional (or at the request of the student and/or of the teacher) and can increase or decrease the final mark.

A project using R software is available, to be carried out according to the guidelines provided at the beginning of the course. It is worth 2 points.

Office hours

By appointment

Sustainable Development Goals

QUALITY EDUCATION