UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistics and Data Analysis

2627-1-F1703Q051

Aims

Here are the main goals of this class:

1. Knowledge and understanding

To provide students with a solid grounding in statistical methods and data analysis techniques commonly used in physics, including the conceptual foundations necessary to critically interpret experimental data.

2. Applying knowledge and understanding

To develop the ability to apply statistical tools and data analysis methods to real-world physics problems, through hands-on exercises and interactive sessions aimed at reinforcing practical understanding.

3. Making judgements

To cultivate critical thinking in the evaluation of data quality, in the choice of appropriate statistical models, and in the interpretation of outcomes within the broader context of physical theories and experimental limitations.

4. Communication skills

To strengthen students' capacity to clearly and effectively present their data analysis process and results, using written reports, oral discussions, and visual tools tailored to both scientific and interdisciplinary audiences.

5. Learning skills

To encourage autonomous learning and the continuous development of data analysis skills, enabling students to confidently approach increasingly complex datasets and keep pace with evolving techniques in physics research.

Contents

- Recap of fundamental concepts of probability and statistics
- Recap of parameters estimation
- Hypothesis Testing
- Confidence interval estimation
- Intro to MC methods
- Intro to Machine learning and Deep Learning
- Unfolding
- Analysis workflow Walk-Through

Detailed program

Probability & Monte Carlo Foundations

- Foundations, interpretations, PDFs, CDFs, and Quantiles
- Joint PDFs, moments, transformations, and error propagation
- MC methods: Hit-or-miss, rejection sampling, transformation method

Frequentist Inference

- Hypothesis testing (Neyman-Pearson lemma, p-values)
- Estimators and Maximum Likelihood (MLE)
- Covariance, confidence intervals/regions, and Wilks theorem
- Nuisance parameters (Profile likelihood)

Bayesian Inference

- Parameter estimation (Priors, sensitivity, credible intervals)
- Nuisance parameters (Marginalization)
- Markov Chain Monte Carlo (MCMC)
- Hypothesis testing (Bayes factors)

Intro to machine learning

- Fundamental concepts (loss, generalization)
- Supervised learning (regression, classification)
- Unsupervised and self-supervised learning (clustering, auto-encoders)
- Dimensionality reduction
- MLP, BDT, PCA, …

Deep Learning

- Intro to Convolutional Networks, Graph Neural Networks, Recurrent Neural Networks

Uncertainty in ML

- Incorporating uncertainties in ML, Bayesian Neural Networks

Unfolding

- Inverse Problem

- Regularisation Methods
- ML based unfolding

Analysis example end-to-end

- examples of analysis flows and presentation of results (examples from multiple fields)
- systematic uncertainty (estimation and treatment)

Prerequisites

A basic knowledge of programming (preferably Python) is preferred. Material will be provided before the beginning of the course for those who don’t have programming experience.

Teaching form

Frontal lectures and hands on exercises. Students are required to bring their own laptop.

Textbook and teaching resource

Semester

Second semester

Assessment method

Each student is required to develop a data analysis project (subject to be agreed with the teacher) and that will serve as a starting point for the oral exam. The oral exam will also include questions on all the topics seen during the lectures.

Office hours

Contact us via email to book an appointment.

Sustainable Development Goals

QUALITY EDUCATION