Name: Statistics, Monte-Carlo methods and data processing - 34056

Type: core **Semester:** 1st

ECTS: 5

Periodicity: annual

Departments involved: Department of Structure and Constituents of Matter (UB)
Department of Astronomy and Meteorology (UB)

Coordinator: Francesc Xavier Luri Carrascoso

Professors: LLuís Garrido Beltran, Francesc Xavier Luri Carrascoso

Language: English/Catalan (depending on the audience)

Prerequisite:

Aims:

The basic objective of this course is to inform the students on the concepts of the theory of probability and statistics, and on the Monte Carlo generation and data processing techniques that come from these concepts. Thanks to the combination of theoretical and practical lessons, the students will acquire the necessary skills that will allow them to apply the knowledge to real problems in any scientific field. At the end of the course the students will have a set of tools (the already available software) necessary to apply the techniques that are now being used in scientific research.

Syllabus:

1. Revision of probability and statistics

- Basic concepts. Random Variable.
- Bayes' theorem and its use
- Probability functions of one single random variable
- Examples: Bernouilli, Binomial, Uniforme, Poisson, Normal and chi**2
- Tchebycheff's theorem. Law of large numbers
- ROOT tutorial (an analysis package in C++)
- Probability functions of with several random variables
- Examples: Multinomial and normal in 2-dimensions
- Independence of variable. Propagation of errors. Reproductivity of the Poisson and Normal distributions

2. Monte Carlo

- Generation of random uniform numbers
- Generation of Gaussian numbers
- Methods for generating one-dimensional distributions
- Inverse transforms method
- Composition method
- Von Neumann method
- Importance sampling
- Generation of multidimensional distributions
- MC Integration

3. Statistical inference

- Parametric and non-parametric inference. Statistics.

- Confidence intervals
- Fisher information
- Sufficient statistics
- Relationship between sufficiency and information
- Estimators theory
- Maximum likelihood method
- Chi-squared method

4. Hypothesis testing and decision theory

- Comments about hypothesis testing and decision theory
- Goodness of fit
- TMVA tutorial (Toolkit for MultiVAriate analysis in ROOT)
- Kullback and Shannon information
- Examples of application

5- Multivariate analysis

- Analysis and representation of data. Statistical distances
- Analysis of principal components
- Hierarchical classification
- Discriminant analysis
- Bayesian discrimination

6- Techniques of statistical treatment

- Introduction to classification by means of neural networks
- Supervised and non-supervised learning algorithms
- Non-parametric estimation methods of the density function
- Histograms, simple estimator, Kernel estimator
- Introduction to wavelets
- Free software of statistical classification available on the internet

7- Database and data mining

- Database: essential concepts
- Study of some cases: the genome project
- Introduction to data mining
- Future tendencies: the project "Virtual Observatory"

Method:

The teaching of each subject in the program is structured in three parts:

1. Theory: introduction of the theoretical foundations in the classroom by the professor

2. Application to practical cases: using the computer the professor applies the theory to several practical cases

3. Student work: the students have to apply themselves the methods they have learnt to several practical cases proposed by the professor, using the computers in the computer room or their own computer at home.

Evaluation:

Continuous evaluation. The evaluation will be based on the solution of practical problems proposed by the professor; eight practical cases will be proposed along the course that the students will have to solve in two weeks, delivering a written report summarizing the work done and the results. Additionally, the students will have to do a public presentation of one of their reports. The reports and the presentation will be evaluated and the average of the marks will be the global course mark; all the reports and the presentation are required to pass the course.

For those students willing to improve their marks additional work on more advanced and specific subjects will be offered.

Bibliography:

Eadie, W.T.; et al. Statistical Methods in experimental Physics, NORTH-HOLLAND
Frodesen, A.G.; et al Probability and Statistics in particle Physics, Columbia University Press
Hermann, D.W. Computer simulation methods in Theoretical Physics, Springer- Verlag
Murtagh, F., Heck, A Multivariate data analysis. D. Reidel Publ. Company. 1985.
Silverman, B.W. Density estimation for statistics and data analysis. Chapman and Hall. 1986.
Numerical Recipes: The art of Scientific computing. 2on edition". W.H. Press. Cambridge Univ. Press. 1992.
Fruhwirth, R.; et al. Data Analysis Techniques for High-Energy PhysicsCambridge University Press
Babu, G.J.; Feigelson; E.D. Astrostatistics. Chapman and Hall. 1996.