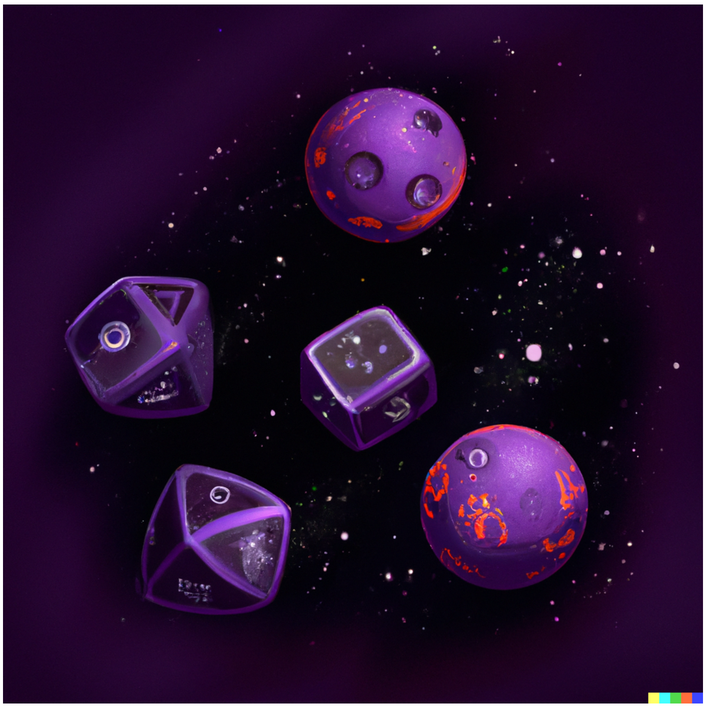
Mathematical and Statistical Methods



Supervisor: **D. Mary**

Contact

✉ David.Mary@oca.eu
☎ +33 4 92 07 63 87

Objectives

During this course, the students will substantially improve their skills in:

1. **Autonomy**:
   (a) The students are able to study without the supervisor, alone or in groups, by means of supporting documents (the pdf of the lecture notes and other online ressources).
   (b) They are able to synthesize, present, explain and illustrate the main points of what they have learned.
   (c) They are able to identify their difficulties and to formulate questions allowing to overcome these difficulties.
   (d) They are able to organize their time in order to progress on specific tasks in parallel to other tasks.
2. **Theory and application of statistical techniques**:
   (a) The students are able to use their statistical background to model data and instruments involved in astrophysical applications.
   (b) The students are able to define a statistical data processing procedure that is relevant for a particular estimation or detection problem involving astrophysical data, along with an estimation of the performance of this procedure.
   (c) The students are able to develop numerical codes to implement classical statistical inference methods.
   (d) The students have clear notions about the contours of statistical estimation, statistical detection, frequentist and Bayesian inference, machine learning.
   (e) The students have specific skills and experience in some important techniques used in theses fields.

Evaluation

The objectives and skills above are evaluated as follows:

1. On each chapter, a 'friendly quiz' is provided, along with a correction, plus anaytical and numerical exercises (python). On the Discord server of the course, each student has a personal channel on which the solution of the exercises and questions can be posted. A noted quiz evaluates the students' level on the current chapter. The dates of the quizs are provided to the student in the beginning of the lecture (along with a detailed lecture by lecture planning of the program). The average of the four quizs' grades makes a grade 'A'.
2. A written exam on all chapters provides a grade 'B'. This exam is typically based on the numerous exercises of the lecture to reward hard and steady working students.

The final grade is $(A + B)/2$.

Contents & Resources

- Chapter 1 provides a short history of statistical inference. It makes an important difference between algorithmic developments (that are driven by or focused on applications) vs inferential arguments (that tend to support the methodology). Statistical Inference has developped in three stages that will be studied in this lecture:
  1. Great themes of classical inference : Bayesian, frequentist, and Fisherian (Maximum Likelihood Estimation)
  2. Early computer-age developments, from the 1950s through the 1990s. Many effects and progresses have been brought by technology (bootstrap, jackknife, cross-validation, empirical Bayes, MCMC - to be seen later in this lecture). This was the early rise of machine learning.
  3. "XX1 century topics": the area of ambitious algorithms of machine learning, data mining and AI (Artifical Intelligence). These algorithms always involve and combine classical algorithms and concepts. The justification and understanding of such algorithms is the ongoing task of modern inference. Hence, when dealing with AI and machine learning, the main concepts and tools of classical inference must be known.

  Important concepts are introduced through three application examples on regression, bootstrap confidence intervals and hypothesis testing.

  The chapter ends with an introduction to Data Mining and Artificial Intelligence, with an analytical and numerical homework on a simplistic Neural Network training. (The picture of this Syllabus was generated by DALL·E by the way).

- Chapter 2 deals with statistical estimation, which is the process of inferring the values of parameters of interest from data, and providing confidence levels for the result.

  Fundamentals quantities and properties of estimators are studied : bias, variance, Mean Square Error, efficiency, and optimality of estimators. The Cramer-Rao lower Bound, that characterizes the minimum variance of any unbiased estimator, is studied as well as the Fisher information.

  The last part of the chapter deals with one of the most used estimator, the Maximum Likelihood estimator, and its performance (asymptotic efficiency and distribution).

- Chapter 3 deals with statistical detection, which is the process of deciding whether some specific signal is hidden or not in the data, and providing estimation on the significance level of the result.

  This Chapter first introduces general concepts related to detection theory : statistical test, test statistic, statistical hypotheses, probability of false alarm and of detection, p-values and ROC (Receiver Operating Characteristics) curves.

  The first important (and most powerful) testing procedure that is studied is the Likelihood Ratio (or Neyman-Pearson) test, which requires all parameters of the statistical model to be known. This test serves as a very useful benchmark to any practical testing approach.

  In real applications, some parameters are unknown. In this case, a useful and often powerful testing approach is the Generalized Likelihood Ratio, where the Maximum Likelihood Estimates of the parameters are plugged in the Likelihood Ratio to mimic the Neyman-Pearson test.

- Chapter 4 turns to Bayesian estimation. Bayes' rule is first introduced along with historical aspects.

  The central idea of Bayes estimation is to incorporate prior knowledge in the inference process through a prior distribution on the unknown parameters. If we want to measure the mass of a planet, or an apple, or an elementary particle, this mass is indeed not a random number; it is what is and cannot have a distribution. But when using the Bayesian formalism, we assign such a probability to the parameters. The prior probability corresponds to the state of our knowledge (i.e., belief) about the parameter. This change of interpretation of the symbols regarding probabilities (note that the mathematical axioms of probability do not change under Bayesianism, however !) introduces the notion of posterior probability distribution for parameters.

  Examples of the prior's influence on the posterior distribution are first provided to develop intuition on Bayes' estimation.

  The second part turns to 'uninformative' priors (Laplace, Jeffreys), to the most commonly used Bayesian point estimators (posterior mean, maximum and median) and to credible intervals.

  The chapter ends with a detailed comparison list of frequentist versus Bayesian inference. According to B. Efron, a famous contemporary statistician, Statistical inference at its most successful combines elements of the two philosophies, as for instance in the empirical Bayes methods.

Bibliography & Resources

- Lecture notes, slides, python codes and data available online.

- "Computer Age Large Scale Inference", B. Efron, Cambridge University Press, 2019

- "Statistics, Data Mining & Machine Learning in Astronomy", Z. Ivezic, Princeton Series in Modern Observational Astronomy, Second Edition, 2020

- "Fundamentals of Statistical Signal Processing, Volume I: Detection Theory", S. Kay, Prentice Hall, 1993

- "Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory", S. Kay, Prentice Hall, 1993