

Probability and Statistics for Data Science

Course title – Intitulé du cours	Probability and Statistics for Data Science
Level / Semester – Niveau /semestre	M1/S1
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	Gadat Sébastien
Other teacher(s) – Autre(s) enseignant(s)	Ibrahim Jean-Paul
Other teacher(s) – Autre(s) enseignant(s)	Gruen Christine
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	30
TA Hours – Volume horaire TD	15
TP Hours – Volume horaire TP	
Course Language – Langue du cours	Anglais
TA and/or TP Language – Langue des TD et/ou TP	Anglais

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

Jean-Paul Ibrahim - jean-paul.ibrahim@ut-capitole.fr - office TJ16

Sébastien Gadat – sebastien.gadat@tse-fr.eu – office T586

Preferred means of interaction: at the end of class, by appointment, by e-mail

For tutorials:

Gruen Christine Preferred means of interaction: at the end of class, by appointment, by e-mail

Course Objectives – Objectifs du cours :

The first part of the course is dedicated to review and complements of probability tools that will be used in the second part. The second part of the course is about sampling theory in a statistical model and statistical decision theory. The course highlights the general principles used to support the choice of an optimal statistical technique for a given problem. It is meant to provide a general framework for statistical decision theory.

The probability review part involves the following topics: cumulative distribution function for continuous and discrete probability laws, quantiles. Classical laws. Random vectors (joint, marginal and conditional distribution). Bivariate and multivariate Gaussian distribution. Conditional expectation. Asymptotic behavior (law of large numbers, central limit theorem, Cramer's theorem or delta method, asymptotic distribution of empirical moments and quantiles).

In the statistical part, we will answer the question: what is mathematical statistics about ? We introduce a general framework for statistical models, in particular exponential models, and sampling theory. We then discuss the principles of statistical decision theory. The student is supposed to

acquire the ability to advocate for his choices between several models and/or techniques in a statistical problem.

Skills that will be developed during the course: Compute the expectation and the variance of a discrete or an absolutely continuous random variable. Find the distribution of a transformed random variable. Compute joint and marginal distributions (discrete, continuous and mixed cases). Compute a conditional density and a conditional expectation. Deal with Gaussian vectors (joint density, affine transformation and conditional expectation). Show the convergence in distribution and the convergence in probability. Apply the central limit theorem and the Delta-method. Write a statistical model corresponding to a given experiment. Use quantiles for descriptive or modeling purposes. Compare two estimators in terms of risk. Construct new estimators using the bayesian principle.

Prerequisites - Pré requis :

A probability course introducing the tools to study the distribution of random vectors, and an introduction to statistical estimation theory.

Set theory, series and integrals calculus Cours de probabilités de base (semestre 3)

Practical information about the sessions - Modalités pratiques de gestion du cours :

Personal computers allowed.

Grading system - Modalités d'évaluation :

The grade will be a weighted average of the midterm exams (50%), the final exam (50%).

Bibliography/references - Bibliographie/références :

Mathematical Statistics, Jun Shao, Springer texts in Statistics, 1999.

Theory of Statistics, Mark Schervish, Springer series in Statistics,1995.

A course in mathematical statistics, G. Roussas, Academic Press, second edition, 1997(available on internet)

Initiation à la statistique avec R, Bertrand et Maumy-Bertrand, Dunod, 2010.

Basic Probability Theory, Robert Ash, Dover Publications, 2008 (available on internet).

Probabilités, Philippes Barbe et Michel Ledoux, EDP Sciences, 2007.

Lecture Notes for Introductory Probability, Janko Gravner, 2014 (available on internet).

Session planning - Planification des séances :

September to December.