Probability and Statistics course, MA2 2022-2023

Contents

Probability and Statistics

This program consists of two series of intertwined courses in probability (P) and statistics (S), the aim of which is to provide a solid and interdisciplinary education in modern theoretical aspects of the mathematics of randomness. We have tried to offer frontier subjects (probability and statistics, PDE, mathematical physics, neural networks).

By their choice of lecture series (3 out of 4 in the first semester and 4 out of 6 in the second), students will be able to choose a “ probabilistic ” or “ statistician ” coloration for their course. It is also possible to follow the lectures of other courses, in particular in PDEs.

Note that four of the “ statistics ” oriented lecture series (one in the first semester and three in the second) are offered by the M2 course em Maths en action.

Finally, it is strongly recommended to take the refresher courses.

Refresher course

Stochastic tools (Grégory Miermont, 15h)

1. Discrete time martingales: stopping theorems and convergence. Extensions for continuous time martingales.

2. Construction of Brownian motion. Regularity of trajectories.

3. Some properties of Brownian trajectories. Connection with the heat equation.

Semester One

P1: Statistical physics (Christophe Garban, 24h)

In statistical physics, one is interested in physical models made of a large number of microscopic elements which interact together in a simple fashion. The goal is then to understand how come such simple microscopic mechanisms can generate interesting (and surprising!) macroscopic phenomena such as phase transitions or symmetry breaking. This program has lead to the development of an important branch of probability theory and the aim of this course is to give a panorama of the field together with tools and techniques that are used in statistical mechanics. We shall focus on three fundamental models: percolation, Ising model and O(n) spherical spin model. Program of the course:

1. Percolation

   * Definition, phase transition

   * FKG inequality, $p_c = 1/2$

   * Exponential decay in the sub-critical regime

   * Russo-Seynour-Welsh theorem for critical percolation

2. Ising model

   * Definition, correlation inequalities

   * Infinite volume limit, Free energy, phase transition

   * Low temperature and Peierls argument

   * Uniqueness at high temperature

3. Phase transition KT (Kosterlitz-Thouless) Nobel prize in phyics 2016

   * Spin models with continuous symmetry ($\sigma_x \in S^d, d \geq 1$)

   * No symmetry breaking in dimension 2 (Mermin-Wagner theorem)

   * Gaussian Free Field

   * Vortices and Coulomb gas

   * A glimpse of Frölich-Spencer Theorem on the KT transition for the XY model $[\sigma = (\sigma_x)_{x \in Z^2} \in (S^1)^{Z^2}]$

References
[1] W. Werner, Percolation et mod* èle d'Ising, Soc. Math. France, 2009.
[2] Y. Velenik, Introduction aux champs aléatoires markoviens et gibbsiens,
http://www.unige.ch/math/folks/velenik/Cours/2006-2007/Gibbs/gibbs.pdf.

P2: Stochastic calculus (Grégory Miermont, 24h)

This lecture series will present some of the most important tools allowing to build and study continuous-time stochastic processes, the central example of which is of course Brownian motion. To this end, we will have to introduce and study semimartingales, a rich class of processes for which one can develop a differential and integral calculus, and set and solve certain type of stochastic differential equations.

Just as for the familiar ordinary differential equations (or PDEs), the motivation to study such stochastic differential equations comes from the goal of understanding the global behaviour of random processes by equations describing their infinitesimal behaviour. But since we are dealing with random processes, these equations contain a random "noise", which informally is an infinitesimal increment of Brownian motion. The main problem of their study comes from the fact that Brownian motion (and therefore the other processes of interest) have too rough trajectories (nowhere differentiable, for instance) for the usual differential and integral calculus to make sense.

In front of this obstacle, we will develop a notion of stochastic integral, due to Itô. It will give rise to a particular integral calculus, in which Itô's formula acts as an integration by parts (or a fundamental theorem of analysis) of a new kind. This integral calculus will allow us to study the stochastic differential equations for continuous semimartingales, and will shed a new light on these processes, for instance via Lévy's characterization of Brownian motion, or the Dubins-Schwarz theorem according to which continuous martingales are appropriate time-changes of Brownian motion. Contents:

1. Generalities on continuous-time processes
2. Continuous-time martingales, regularization. Local martingales, semimartingales. Bracket of a continuous semimartingale.
3. Stochastic integral with respect to a continuous semimartingale.
4. Itô's formula and applications. The theorems of Lévy, Dubins-Schwarz, Girsanov. Burkholder-Davis-Gundy inequalities.
5. Stochastic differential equations. The Lipschitz case.
6. (Time allowing) Continuous-time Markov processes. Generators. Diffusions.

References

[1] Karatzas-Shreve: Brownian motion and stochastic calculus
[2] Le Gall: Brownian motion and stochastic calculus
[3] Mörters-Peres: Brownian motion
[4] Revuz-Yor: Continuous martingales and Brownian motion
[5] Varadhan: Stochastic processes

S1: Concentration of measure in probability and high-dimensional statistical learning (Guillaume Aubrun, Aurélien Garivier, Rémi Gribonval, 24h)

This course will introduce the notion of concentration of measure and highlight its applications, notably in high dimensional data processing and machine learning. The course will start from deviations inequalities for averages of independent variables, and illustrate their interest for the analysis of random graphs and random projections for dimension reduction. It will then be shown how other high-dimensional random functions concentrate, and what guarantees this concentration yields for randomized algorithms and machine learning procedures to learn from large training collections. This course will be based on a sample of the classical textbook "Concentration Inequalities" by Boucheron, Massart, Lugosi, and on "High-Dimensional Probability" by Roman Vershynin. Applications to machine learning will rely on "Understanding Machine Learning", by Shalev-Schwartz and Ben-David.

S2: Stochastic modeling and statistical learning (Aurélien Garivier and Clément Marteau, 24h)

[course also offered in the master maths in action]

- Large-dimensional regression: examples and modeling, reminders and development around the linear model (modeling, hypotheses, least squares and likelihood, Fisher / Student test, etc.).
  - Introduction to the selection of models (construction of the criteria Cp/AIC/BIC, inequalities à la Birgé & Massart, behavior in large dimensions).
  - Ridge method (heuristic, link with Tikhonov, property of risk, a few words on the choice of lambda by estimation of the risk.
  - Introduction to LASSO (construction and heuristics, link with the compressed sensing, theoretical properties / oracle inequalities, compatibility conditions).
- Supervised classification: examples and modeling, overview of some algorithms (kNN, SVM, neural networks, logistic regression, ...), theoretical aspects (inequalities of concentrations, Vapnik theory, kernels, etc.).
- Unsupervised classification: ACP, Clustering (kmeans, hierarchical methods, ...), Gaussian mixing models, Spectral clustering.
- Other risks, extremes, introductions to research issues.

Semester 2

P3: Scaling limits of interacting particles systems (Oriane Blondel and Christophe Poquet, 18h)

This course aims to present how probabilistic particle models can be described by PDEs in the large population limit. The studied systems come from modeling of physical or biological phenomena. The course will use notions of stochastic calculus, continuous-time Markov chains and will introduce some notions of PDE.

This course will consist in two parts, the first on mean-field interactions and the other on network systems with short-range interactions. In the first part, we will present the classical results of convergence of diffusion models in mean field, the notion of propagation of chaos as well as more recent results for dense but incomplete interaction graphs. In the second part, we will focus on exclusion processes, which are Markov processes on $\{0,1\}^N$. We will study their hydrodynamic limit and describe their fluctuations.

References

[1] L. Bertini, G. Giacomin, Stochastic Burgers and KPZ equations from particle systems. Comm. Math. Phys. 183(3): 571-607 (1997).

[2] F. Coppini, H. Dietert and G. Giacomin, A law of large numbers and large deviations for interacting diffusions on Erdős–Rényi graphs. Stochastics and Dynamics 20 (2020).

[3] S. Delattre, G. Giacomin and E. Luçon, A Note on Dynamical Models on Random Graphs and Fokker–Planck Equations. Journal of Statistical Physics 165 (2016).

[4] J. Gärtner, On the McKean–Vlasov limit for interacting diffusions. Mathematische Nachrichten 137 (1988).

[5] P. Gonçalves, Hydrodynamics for symmetric exclusion in contact with reservoirs. Stochastic Dynamics Out of Equilibrium, Institut Henri Poincaré, Paris, France, 2017, Springer Proceedings in Mathematics and Statistics book series, 137-205 (2019).

[6] P. Gonçalves, M. Jara, M. Simon. Second order Boltzmann-Gibbs Principle for polynomial functions and applications, Journal of Statistical Physics, Volume 166, Issue 1, 90–113 (2017).

[7] C. Kipnis and C. Landim, Scaling limits of interacting particle systems. Grundlehren der Mathematischen Wissenschaften. 320. Berlin: Springer. xvi, 442 p. (1999).

[8] A.-S. Sznitman, Topics in propagation of chaos. Ecole d'été de probabilités de Saint-Flour XIX–1989. Springer (1991).

P4 : Large random matrices and applications (Alice Guionnet, 18h)

Large random matrices have appeared in the work of the statistician Wishart, then in those of the physicist Wigner. Their theory has developed at a tremendous speed since the years 1990's. This lecture series will be opportunity to explore this theory, in particular the questions of almost-sure convergence of the spectrum, the study of their local and global fluctuations, and their large deviations properties. Finally, we will study some applications in statistics.

P5 : An approach of disordered systems via partial differential equations (Jean-Christophe Mourrat, 18h)

The aim of statistical mechanics is to describe the behavior at large-scale of collections of simple elements, often called spins, which interact locally according to elementary rules and are subject to disorder. We will focus on the situation where local interactions are chosen at random, in which case the models are generally called "spin glasses". Such models are already surprisingly difficult to analyze when all the spins interact between them. In this course, we will revisit this analysis using the tools of the theory of Hamilton-Jacobi equations.

After giving an overview of the course, we will start with us focus on the very simple Curie-Weiss model. This will be the occasion to introduce the analytical tools linked to the study of the equations of Hamilton-Jacobi, which we will use to identify the limiting free energy of the model. We will then transpose this strategy to a first class of disordered models, derived from statistical inference. In terms of difficulty, these models constitute a useful bridge between the Curie-Weiss model and the spin glasses. Finally, we will focus on these latter models, in which Hamilton-Jacobi equations of infinite dimension arise.

References:

[1] Friedli, S. and Velenik, Y. Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction Cambridge: Cambridge University Press, 2017.

[2] Evans, Lawrence C. Partial differential equations. Second edition. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010. xxii+749

S3: Neural networks (Aurélien Garivier, 18h)

[Also offered in the course *Maths en action*]

The goal of this course is twofold:

- To present the principles of modern deep neural networks, as well as the technical ways to implement them for solving classification and regression problems.
- To provide a detailed overview of the mathematical foundations of modern learning techniques based on deep neural networks.

Starting with the universal approximation property of neural networks, we will then see why depth improves the capacity of networks to provide accurate function approximations for a given computational budget. Tools to address the optimization problems appearing when training networks on large collections will then be covered, and their convergence properties will be reviewed. Finally, statistical results on the generalization guarantees of deep neural networks will be presented, both in the classical underfitting scenario and in the overfitting scenario leading to the so-called "double descent" phenomenon.

S4: Optimal transport and learning (Philippo Santambrigio, Ivan Gentil, Yohann de Castro, Ievgen Reedko, Julie Digne, Nicolas Bonnel, 6 p.m.)

[Also offered in the course *Maths en action*]

The aim of the course is to present the main lines of the theory of optimal transport and some of its applications in data science.

The first part of the course will detail the Monge-Kantorovich problem, its formulation as a linear programming problem and the use of convex duality, as well as distances (known as Wasserstein distances) that optimal transport allows to define on the space of probability measures. Geodesics and barycenters in Wasserstein space, of great importance in interpolation and comparison of data, will also be introduced.

A second part of the course will focus on numerical methods for solving optimal transport problems, with particular attention to the methods best suited to large dimension and unstructured data, in particular the Sinkhorn algorithm.

Finally, the third part of the course will present a choice of applications of transport and Wasserstein distances in statistical learning, of which we cite as examples the Wasserstein GANs, transfer learning, data generation models,...

S5: Inverse problems and parcimony (Yohann de Castro et Rémi Gribonval, 18h)

[Also offered in the course *Maths en action*]

Sparsity and convexity are ubiquitous notions in Machine Learning and Statistics. In this course, we study the mathematical foundations of some powerful methods based on convex relaxation: L1-regularisation techniques in Statistics and Signal Processing; Nuclear Norm minimization in Matrix Completion; K-means and Graph Clustering.

These approaches turn out to be Semi-Definite representable (SDP) and hence tractable in practice. The theoretical part of the course will focus on the performance guarantees for theses approaches and for the corresponding algorithms under the sparsity assumption. The practical part of this course will present the standard SDP solvers for these learning problems.

Keywords: L1-regularisation; Matrix Completion; K-Means; Graph Clustering; Semi-Definite Programming.