PROBABILITY AND STATISTICS
MA2 2021–2022

Contents

1. Probability and statistics

This program is made of two closely intertwined lecture series in probability (P) and statistics (S), that will provide a strong and interdisciplinary training in the modern mathematics of random phenomena.

Via their choice of lectures (3 out of 4 in the first semester and 4 out of 6 in the second), students may choose to give a stronger focus on the probabilistic or statistical aspects of this program. It is also possible to follow lectures from other programs, especially in PDEs. It is strongly advised to follow the corresponding refresher courses.

2. 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.

3. Courses

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+8h). 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: Non-parametrics (Irène Gannaz, Clément Marteau, Franck Picard, 24h). In this course we will focus on recent developments in non parametric statistics, with a special focus on non-parametric model selection and high dimensional statistics. Variable selection through the LASSO (Least Absolute Shrinkage and Selection Operator) has revolutionized high dimensional statistics thanks to the use of a L1-constrained optimization problem that ensures powerful statistical properties. These connections between non-convex optimization and Statistics has been very fruitful from both the applied and theoretical aspects of Machine Learning. A non-negligible part of this course will focus on the theoretical properties (model selection, convergence ...) of penalized estimators, with the use of oracle inequalities as a building block. These penalized methods will be put into perspective with kernel-based estimation and regression, another popular non-parametric strategy that requires fine theoretical and computational calibration (bandwith choice). Some attention may also be payed to related topics such as wavelet transform, multiple testing issues or signal processing on graphs.

P3 : 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.

P4 : Random Graphs (Dieter Mitsche, 18h). In the last years, complex networks have become central elements in many areas (telecommunication networks, internet, neural networks, social networks, propagation of infectious diseases, propagation of rumors, ...). It is a booming area, and it is crucial to develop mathematical models to represent these networks.

A network is often modelled by a random graph, and this course proposes the study of different random graph models, in particular the Erdö-Rényi random graph model, the configuration model and random geometric graphs. A special focus of this course will be given on the threshold of the giant component in different graph models. Contents:

(1) Erdös-Rényi model
   * Introduction, subgraph count
   * Local weak convergence
   * Phase transition, appearance of a giant component
   * Hamiltonicity

(2) Configuration model
   * Differential equation method
   * Emergence of a giant component 3. Random geometric graphs
   * Euclidean random geometric graphs - emergence of the giant component
   * Introduction to random hyperbolic graphs

References

[1] N. Alon, J. Spencer, The probabilistic method, 3rd ed., John Wiley & Sons, 2008.

[2] C. Bordenave, Lecture notes on random graphs and combinatorial optimization, https://www.math.univ-toulouse.fr/~bordenave/coursRG.pdf

[3] A. Frieze, M. Karonski, Introduction to random graphs, CUP, 2015.

[4] M. Penrose, Random geometric graphs, Oxford Univ. Press, 2003.

[5] R. van der Hofstad, Random graphs and complex networks,
Vol. 1, Cambridge Series in Statistical and Probabilistic Mathematics, 2017.
Volume 2, https://www.win.tue.nl/~rhofstad/ NotesRGCN.html
https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf

P5 : Determinantal processes (Adrien Kassel, 18h). A determinantal process on a nice topological measured space S is a random discrete collection of points such that the correlation functions – loosely speaking, the density of probability of seeing a finite subcollection of points at given locations – exist, and may be written in the form

$$\rho_m(x_1, \dots, x_m) = \det[(K(x_i, x_j))_{1 \le i,j \le m}] \,,$$

where $K : S^2 \to \mathbb{C}$ is a two-point function, also called a kernel.

There are many examples of stochastic models which give rise to interesting determinantal processes, many of which find their origin in mathematical physics. The corresponding kernel can sometimes be computed rather explicitly, which enables the study of fine properties of the model. We may broadly distinguish two classes of processes according to the topology of S: discrete ones (e.g. $S = \mathbb{Z}^d$) and continuous ones (e.g. $S = \mathbb{R}^d$). Examples of discrete processes include random spanning forests on finite and infinite graphs; examples of continuous processes include eigenvalues of certain random matrices of finite or infinite size.

The goal of this course will be to present a theory of determinantal processes, namely to present what is common to these examples beyond their particularities. For that matter, we will focus on the better understood case where K is self-dual, namely when the symmetry $K(x, y) = \overline{K(y, x)}$ holds.

We will start with the case where S is finite, for which a very complete understanding is available. The kernel K is then a Hermitian matrix, and the process is completely described in terms of linear algebra in $\mathbb{C}^S$, and its Euclidean geometry. This allows to explain the link to theoretical physics in quite a transparent way. This part of the theory may easily be extended to the case where S is countable, where now $\ell^2(S)$ is the relevant geometry. To move on to the case of uncountable S requires extra caution, and the dictionary between kernels and Euclidean geometry now needs to be enhanced to the setup of bounded integral operators on the Hilbert space $L^2(S)$.

Examples we will present, at least superficially, include: uniform spanning forests of infinite lattices; zeros of the Gaussian analytic function on the unit disc; eigenvalues of a random Hermitian matrix distributed according to the Gaussian unitary ensemble.

References

[1] A. Borodin. Determinantal point processes. *Oxford handbook of random matrix theory*, pp 231–249, Oxford Univ. Press, 2011.

[2] R. Lyons. Determinantal probability measures. *Publ. Math. Inst. Hautes Études Sci.*, (98):167–212, 2003.

[3] A. Soshnikov. Determinantal random point fields. *Uspekhi Mat. Nauk*, 55(5(335)):107–160, 2000.

S3: Mathematical foundations of deep neural networks (Rémi Gribonval, Aurélien Garivier, 18h). This course will 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, the course will then show 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 described, both in the classical underfitting scenario and in the overfitting scenario leading to the so-called "double descent" phenomenon.

S4: Inverse problems and high dimension (Yohann de Castro, Rémi Gribonval, 18h). 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 turned to be Semi-Definite representable (SDP) and hence tractable in practice. The theoretical part of the course will focus on the guarantees of these algorithms under the sparsity assumption. The practical part of this cours will present the standard SDP solvers of these learning problems.

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

S5: Advanced machine learning theory (Laurent Jacob, Joseph Salmon, 18h). Choosing an appropriate data representation is a key element in modern machine learning. While most existing learning algorithms manipulate vectors, important data types including webpages, sequences or graphs do not admit a straightforward vectorial representation. Other data types admit a natural vectorial encoding, but the resulting representation is not necessarily appropriate for learning (this is the case for images).

This course will introduce positive definite kernels, a powerful mathematical framework to create, analyze and manipulate data representation. Kernels are functions measuring the similarity between pairs of objects. We will show how they implicitly define a mapping of the data to a Hilbert space, and how this fact can be used to manipulate large or even infinite sets of descriptors. We will specifically discuss how these kernels can be used in supervised and unsupervised learning algorithms, and provide examples of kernels defined on sequences and graphs. Finally, we will consider how this framework can shed some light on convolutional neural networks.

Keywords: positive definite kernel, RKHS, kernel methods, sequences, graphs, libsvm, large scale learning, deep kernel machines.