Topics in Mathematics and Economic Decision graduate course

For Masters, MRes, PhD students and researchers

PAC-Bayes and Information bounds, with Applications to Variational Inference

by **Pierre Alquier** (Riken AIP Tokyo, in office T523 while at TSE)

In Room 1 RDJ floor of the TSE building in November 2022

- Tuesday 8 from 5 to 6:30 PM
- Tuesday 15 from 3:40 to 5:10 PM
- Monday 21 from 3:40 to 5:10 PM
- Tuesday 22 from 5 to 6:30 PM
- Monday 28 from 3:40 to 5:10 PM

Syllabus:
Aggregated predictors are obtained by making a set of basic predictors vote according to some weights, that is, to some probability distribution. Randomized predictors are obtained by sampling in a set of basic predictors, according to some prescribed probability distribution. Thus, aggregated and randomized predictors have in common that they are not defined by a minimization problem, but by a probability distribution on the set of predictors. In statistical learning theory, there is a set of tools designed to understand the generalization ability of such procedures: PAC-Bayes bounds. The original PAC-Bayes bounds of McAllester essentially provided numerical certificates on the generalization ability of randomized predictors. Catoni and Zhang proved that they could also lead to prove strong statistical properties of these algorithms: oracle inequalities, and rates of convergence (the oracle version of PAC-Bayes bounds are often referred to as "Information bounds").

In the past 5 years, PAC-Bayes bounds led to many advances in statistics and machine learning. Notably, Dziugaite and Roy obtained tight generalization certificates for deep neural networks for the first time, via such techniques. With James Ridgway and Nicolas Chopin, we proved consistency of a popular algorithm in Bayesian statistics, variational inference, thanks to Information bounds.

In these lectures, we will introduce the basic theory of PAC-Bayes bounds and Information bounds. We will show how they can be used both to design efficient learning algorithms, and to derive their statistical properties. The lectures will be organized as follows:
1) statistical learning, randomized and aggregated predictors, probabilistic tools (exponential inequalities).
2) PAC-Bayes bounds, and algorithms to minimizing such bounds.
3) PAC-Bayes bounds in expectation, Mutual Information bounds, derivation of rates of convergence.
4) comparison between PAC-Bayes bounds and asymptotic results on the contraction of the posterior in statistics.

5) variational inference: definition, examples of applications, convergence via PAC-Bayes bounds.

References:
1) general introduction to PAC-Bayes bounds:
- Alquier, User-friendly introduction to PAC-Bayes bounds, preprint, 2021. https://arxiv.org/abs/2110.11216
2) PAC-Bayes bounds for deep neural networks:
- Dziugaite and Roy, Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data, UAI 2017. http://auai.org/uai2017/proceedings/papers/173.pdf
- Pérez-Ortiz, Rivasplata, Shawe-Taylor and Szepesvári, tighter risk certificates for neural networks, Journal of Machine Learning Research, 2021. https://www.jmlr.org/papers/v22/20-879.html
3) PAC-Bayes bounds and variational inference:
- Alquier and Ridgway, Concentration of tempered posteriors and of their variational approximations, Annals of Statistics, 2020. https://arxiv.org/abs/1706.09293
4) advanced topics on PAC-Bayes/Information bounds:
- Catoni, PAC-Bayesian supervised classification: the thermodynamics of statistical learning, IMS Lecture Notes, 2007. https://arxiv.org/abs/0712.0248