# Sparse recovery by aggregation and langevin monte-carlo

Arnak Dalalyan, Université Paris Est et Ecole des Ponts ParisTech
(joint work with A. Tsybakov)

**Abstract**

The main aim of the present talk is to show that advanced probabilistic tools developed in the theory of continuous-time diffusion processes may be of substantial help in the classical statistical problem of estimation in linear regression with fixed design and i.i.d. noise. More precisely, high-dimensional linear regression model will be considered in the so called $p \gg n$ setup, that is when the number of covariates is possibly much larger than the sample-size. It will be shown that under the sparsity scenario, the aggregate with exponential weights based on a heavy-tailed prior distribution leads to accurate estimates of the unknown parameter-vector. Note that the computation of the exponentially weighted aggregate (EWA), that can be considered as a Bayesian posterior mean, is far from being trivial.

It turns out that a particularly efficient way of computing (approximately) the EWA relies on the use of Langevin diffusion processes, having the posterior as invariant distribution. The efficiency of this approximation will be demonstrated both theoretically and by means of numerical experiments.