CREST - GENES
Cours doctoraux 2018 – 2019

Polynomial method in statistical estimation: from large domain to mixture models

Wu YIHONG
Yale University

Approximation-theoretic methods have been among the key techniques in nonparametric statistics. In recent years, these ideas, especially those involving polynomials, have been found useful in various statistical inference problems in discrete settings. At a high level, on one hand, the apparatus of polynomial approximation and interpolation provides powerful tools for constructing computationally efficient estimators with provable statistical guarantees; on the other hand, through the connection to the theory of moments, these procedures come with natural minimax lower bounds that certify their optimality.

The focus of this short course is to introduce the necessary background on the polynomial method and discuss two of its main applications: (a) statistical estimation on large domains; (b) method of moments in mixture models. The tentative plan is:

- Preliminaries on the theory of polynomials and moments:
  - Uniform approximation and linear programming duality
  - Moment space and characterization, Gauss quadrature
  - Polynomial interpolation: Lagrange, Newton and Hermite
- Entropy estimation
  - Estimator via best polynomial approximation
  - Minimax lower bound
  - Nonparametric setting
- Estimating the unseens
  - Support size of distributions
  - Distinct element problem and counting cliques
  - Species discovery curve
- Learning Gaussian mixtures
  - Method of moments and variants
  - Optimal transport and moment comparison theorems

Along the way, I will discuss extensions, such as applications to sampling large graphs, and various open problems.

References:

- Oleg Lepski, Arkady Nemirovski, and Vladimir Spokoiny. On estimation of the L_r norm of a regression function. Probability theory and related fields, 113(2):221–253, 1999.
- T.T. Cai and M. G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. The Annals of Statistics, 39(2):1012–1041, 2011.
- Y. Wu and P. Yang, "Chebyshev polynomials, moment matching, and optimal estimation of the unseen", https://arxiv.org/abs/1504.01227, Mar 2015.
- Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. IEEE Transactions on Information Theory, 61(5):2835–2885, 2015.
- Yihong Wu and Pengkun Yang, "Minimax rates of entropy estimation on large alphabets via best polynomial approximation", IEEE Transactions on Information Theory, vol. 62, no. 7, June 2016.
- Alon Orlitsky, Ananda Theertha Suresh, and Yihong Wu, "Optimal prediction of the number of unseen species", Proceedings of the National Academy of Sciences, vol. 113, no. 47, pp. 13283–13288, Nov 2016.
- Y. Wu, "Information-theoretic Methods for High-dimensional Statistics", Lecture notes, Aug 2017.

Cours	Lundis	19 mars 2018 26 mars 2018	de 10h30 à 12h30 de 10h30 à 12h30	salle 2002 **salle 2001**
	Jeudis	22 mars 2018 29 mars 2018	de 14h00 à 17h00 de 14h00 à 17h00	salle 2002 salle 2002

à l'ENSAE, 5 Av. Henry Le Chatelier - Palaiseau (REB B Massy Palaiseau & bus 9106 C ou B)

Ces cours sont proposés aux étudiants de 3ème année de l'ENSAE, de l'ENSAI, ouverts aux étudiants de M2 ou inscrits en *thèse.*

Une inscription préalable est demandée impérativement *pour tous les étudiants de l'ENSAE, de l'ENSAI, ou* extérieurs, à Lyza RACON : lyza.racon@ensae.fr ou par téléphone **au 01.70.26.69.26 afin de pouvoir être admis dans les locaux de l'ENSAE et pouvoir être joints en cas de nécessité par les organisateurs du cours.**