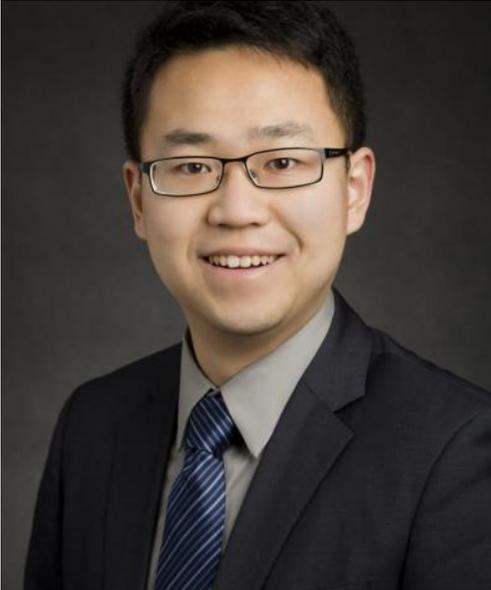


DEPARTMENT OF STATISTICS

**Yihong Wu**

Professor of Statistics and Data Science
Yale University

Self-regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models

Wednesday, February 3rd, 2021

11:45 AM EST

Zoom Meeting: Meeting ID: 977 8097 0143

Password: 801528

Virtual Coffee session before the seminar at 11:30AM EST

Abstract: Introduced by Kiefer and Wolfowitz 1956, the nonparametric maximum likelihood estimator (NPMLE) is a widely used methodology for learning mixture models and empirical Bayes estimation. Sidestepping the non-convexity in mixture likelihood, the NPMLE estimates the mixing distribution by maximizing the total likelihood over the space of probability measures, which can be viewed as an extreme form of overparameterization. In this work we discover a surprising property of the NPMLE solution. Consider, for example, a Gaussian mixture model on the real line with a subgaussian mixing distribution. Leveraging complex-analytic techniques, we show that with high probability the NPMLE based on a sample of size n has $O(\log n)$ atoms (mass points), significantly improving the deterministic upper bound of n due to Lindsay 1983. Notably, any such Gaussian mixture is statistically indistinguishable from a finite one with $O(\log n)$ components (and this is tight for certain mixtures). Thus, absent any explicit form of model selection, NPMLE automatically chooses the right model complexity, a property we term self-regularization. Statistical applications and extensions to other exponential families will be given. Time permitting, we will discuss some recent results on optimal regret in empirical Bayes and the role of NPMLE. This is based on joint work with Yury Polyanskiy (MIT): <https://arxiv.org/abs/2008.08244>

Bio: Yihong Wu is an associate professor in the Department of Statistics and Data Science at Yale University. He received his B.E. degree from Tsinghua University in 2006 and Ph.D. degree from Princeton University in 2011. He is a recipient of NSF CAREER award in 2017 and Sloan fellowship in mathematics in 2018. He is broadly interested in the theoretical and algorithmic aspects of high-dimensional statistics, information theory, and optimization.

