Topics in Statistical Theory (M16)

Richard J. Samworth

This course will provide an introduction to the theory behind a selection of statistical problems that play a key role in modern statistics. Most undergraduate statistics courses are restricted to the study of parametric models; here we will no longer assume that our distributions belong to finite dimensional classes and will instead study fundamental nonparametric problems such the estimation of a distribution function, a density function or a regression function. We will also study minimax lower bounds, which characterise the intrinsic difficulty of a statistical problem, and provide benchmarks against which statistical procedures can be compared.

An outline of the course is as follows:

- An introduction to nonparametric statistics: the basics of empirical process theory, Glivenko–Cantelli theorem, Dvoretzky–Kiefer–Wolfowitz inequality, order statistics, quantile estimation and associated asymptotic distribution theory. Concentation inequalities, including Hoeffding, Bennett and Bernstein inequalities.

- Kernel density estimation: Definition, bounds on bias and variance, uniform nonasymptotic bounds on MSE and MISE. Bandwidth selection via least squares cross validation and Lepski's method, choice of kernel, multivariate density estimation.

- Nonparametric regression: Local poynomial estimation, bounds on weights, bias, variance and MSE. Cubic splines, natural cubic smoothing splines, choice of smoothing parameter.

- Minimax lower bounds: Reduction to testing, f-divergences, Le Cam's two point lemma, Assoaud's lemma, the data processing inequality, Fano's lemma; examples.

Pre-requisites

A good background in undergraduate probability theory, elements of linear algebra and real analysis. Measure theory is not necessary but may be helpful; similarly for a preliminary course in mathematical statistics. Though the material in the Modern Statistical Methods course will not be needed here, the two courses complement each other well.

Literature

The lecturer is currently writing a book based on the course, and this should be available (if not published) in time for the course. Some of the material is covered in:

1. S. Boucheron, Lugosi, G. and Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

2. A. Tsybakov, *Introduction to Nonparametric Estimation*. Springer, 2009.

3. M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

Additional support

Three examples sheets will be provided and associated examples classes will be given. There will be weekly office hours during Michaelmas and a revision class in the Easter Term.