Optimal convex M-estimation via score matching

Richard Samworth

(University of Cambridge)

15/05/2024, 5.00pm

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 683 0699 4223; Password: StatsCol23)

In the context of linear regression, we construct a data-driven convex loss function with respect to which empirical risk minimisation yields optimal asymptotic variance in the downstream estimation of the regression coefficients. Our semiparametric approach targets the best decreasing approximation of the derivative of the log-density of the noise distribution. At the population level, this fitting process is a nonparametric extension of score matching, corresponding to a log-concave projection of the noise distribution with respect to the Fisher divergence. The procedure is computationally efficient, and we prove that our procedure attains the minimal asymptotic covariance among all convex M-estimators. As an example of a non-log-concave setting, for Cauchy errors, the optimal convex loss function is Huber-like, and our procedure yields an asymptotic efficiency greater than 0.87 relative to the oracle maximum likelihood estimator of the regression coefficients that uses knowledge of this error distribution; in this sense, we obtain robustness without sacrificing much efficiency.

Biography:
Richard J. Samworth is Professor of Statistical Science and Director of the Statistical Laboratory at the University of Cambridge. His main research interests are in high-dimensional and nonparametric statistics: he has developed methods and theory for shape-constrained inference, data perturbation techniques (subsampling, the bootstrap, random projections, knockoffs), changepoint estimation and classification.