9/20/21 Seminar

Hongyuan Cao

Title: Optimal False Discovery Rate Control for Large Scale Multiple Testing with Auxiliary Information

Abstract:
Large-scale multiple testing is a fundamental problem in high dimensional statistical inference. It is increasingly common that various types of auxiliary information, reflecting the structural relationship among the hypotheses, are available. Exploiting such auxiliary information can boost statistical power. To this end, we propose a framework based on a two-group mixture model with varying probabilities of being null for different hypotheses a priori, where a shape-constrained relationship is imposed between the auxiliary information and the prior probabilities of being null. An optimal rejection rule is designed to maximize the expected number of true positives when average false discovery rate is controlled. Focusing on the ordered structure, we develop a robust EM algorithm to estimate the prior probabilities of being null and the distribution of p-values under the alternative hypothesis simultaneously. We show that the proposed method has better power than state-of-the-art competitors while controlling the false discovery rate, both empirically and theoretically. Extensive simulations demonstrate the advantage of the proposed method. Datasets from genome-wide association studies are used to illustrate the new methodology.