Department of Decision Sciences
Statistics Seminars

Sampling depth trade-off in function estimation under a two-level design

Akira Horiguchi
Duke University

Friday, 7th June 2024
12:00 pm **Room 3-E4-SR03** Via Roentgen 1 Milano

Abstract

Many modern statistical applications involve a two-level sampling scheme that first samples subjects from a population and then samples observations on each subject. These schemes often are designed to learn both the population-level functional structures shared by the subjects and the functional characteristics specific to individual subjects. Common wisdom suggests that learning population-level structures benefits from sampling more subjects whereas learning subject-specific structures benefits from deeper sampling within each subject. Oftentimes these two objectives compete for limited sampling resources, which raises the question of how to optimally sample at the two levels. We quantify such sampling-depth trade-offs by establishing the L2 minimax risk rates for learning the population-level and subject-specific structures under a hierarchical Gaussian process model framework where we consider a Bayesian and a frequentist perspective on the unknown population-level structure. These rates provide general lessons for designing two-level sampling schemes given a fixed sampling budget. Interestingly, they show that subject-specific learning occasionally benefits more by sampling more subjects than by deeper within-subject sampling. We show that the corresponding minimax rates can be readily achieved in practice through simple adaptive estimators without assuming prior knowledge on the underlying variability at the two sampling levels. We validate our theory and illustrate the sampling trade-off in practice through both simulation experiments and two real datasets. While we carry out all the theoretical analysis in the context of Gaussian process models for analytical tractability, the results provide insights on effective two-level sampling designs more broadly.