Waseda Seminar on Mathematical Statistics

Date: January 10 (Wed.), 2024

Venue: Meeting Room (Dept of Pure & Appl. Math), Building 63-1, Nishi-Waseda Campus, Waseda University

(Access map: https://www.waseda.jp/top/en/access/nishiwaseda-campus)

This seminar is supported by

- JSPS KAKENHI Kiban (C): Grant-in-Aid for Scientific Research No. 23K11018 (Y. Liu)
- Waseda Univ. Tokutei-Kadai: 2023C-094 (Y. Liu)

Program

13:30 ~ 14:30

Optimal Subsampling Bootstrap for Massive Data

Yingying Ma (Beihang Univ.)

14:30 ~ 14:45

Coffee Break

14:45 ~ 15:45

Rerandomization and Regression Adjustment in Stratified Randomized Experiments

Hanzhong Liu (Tsinghua Univ.)

15:45 ~ 16:00

Coffee Break

16:00 ~ 17:30

Integrated Discussion (Master's Students' Presentation on Annual Results)

18:00 ~

Welcome Party

Program

Optimal Subsampling Bootstrap for Massive Data

Yingying Ma (Beihang Univ.)

Abstract: The bootstrap method is a widely used procedure for conducting statistical inference thanks to its simplicity and attractive statistical properties. However, the vanilla version of bootstrap is no longer feasible computationally for many modern datasets that are massive in scale due to the need to repeatedly resample the whole data. As a response, recent years have seen several innovations in advancing the bootstrap method for assessing the quality of estimators by subsampling the full datasets before resampling the subsamples, based on the idea that the size of subsamples is often substantially smaller than the full sample. Naturally, the performance of these modern subsampling methods is influenced by tuning parameters including the size of subsamples, the number of subsamples, and the number of resamples per subsample. In this article, we develop a novel hyperparameter selection methodology for selecting these tuning parameters. Formulated as an optimization problem to find the optimal value of some measure of accuracy of an estimator subject to computational cost, our framework provides closed-form solutions for the optimal hyperparameter values for subsampled bootstrap, subsampled double bootstrap and bag of little bootstraps, at no or little extra time cost. Using the mean square errors as a proxy of the accuracy measure, we apply our methodology to study, compare and improve the performance of these modern versions of bootstrap developed for massive data through simulation study. The results are promising.

Rerandomization and Regression Adjustment in Stratified Randomized Experiments

Hanzhong Liu (Tsinghua Univ.)

Abstract: Stratification and rerandomization are two well-known methods used in randomized experiments for balancing baseline covariates. Renowned scholars in experimental design have recommended combining these two methods; however, limited studies have addressed the statistical properties of this combination. This work proposes two rerandomization methods to be used in stratified randomized experiments, based on the overall and stratum-specific Mahalanobis distances. We obtain the asymptotic distribution of the weighted difference-in-means estimator under the proposed rerandomization schemes and the formulas of variance reduction when compared to stratified randomization. We provide asymptotically conservative variance estimators and confidence intervals for the average treatment effect. Moreover, we discuss the relationship between stratified rerandomization and regression adjustment. Our analysis is randomization-based or design-based and thus does not require strong modeling assumptions on the potential outcome data-generating model. The advantages of the proposed methods are exhibited through simulation study and real-data illustration.