

SCMS & SDS Joint Seminar

JOIN STATISTICS SEMINAR OF SCMS AND SDS

COMBINING MULTIPLE OBSERVATIONAL DATA SOURCES TO ESTIMATE CAUSAL EFFECTS

Dr. Peng Ding

University of California, Berkeley

Lecture

Time: 2:30-3:30 pm., Thursday, Dec. 21, 2017

Venue: Room 205, Shidai Building, Handan Campus

Abstract: The era of big data has witnessed the increasing availability of multiple data sources for statistical analyses. As an important example in causal inference, we consider estimation of causal effects combining big main data with unmeasured confounders and smaller validation data with supplementary information on these confounders. Under the unconfoundedness assumption with completely observed confounders, the smaller validation data allow for constructing consistent estimators for causal effects, but the big main data can only give error-prone estimators in general. However, by leveraging the information in the big main data in a principled way, we can improve the estimation efficiencies while still preserve the consistencies of the initial estimators based solely on the validation data. The proposed framework incorporates asymptotically normal initial estimators, including the commonly-used regression imputation, weighting, and matching estimators, and does not require a correct specification of the model relating the unmeasured confounders with the observed variables. Coupled with appropriate bootstrap procedures, our method is straightforward to implement requiring only software routines for existing estimators.