Department of Statistics

STATISTICS COLLOQUIUM

RAYMOND CARROLL

Department of Statistics
Texas A&M University

Constrained Maximum Likelihood Estimation for Model Calibration Using Summary-Level Information from External Big Data Sources

MONDAY, May 9, 2016, at 4:00 PM
Eckhart 133, 5734 S. University Avenue

ABSTRACT

Information from various public and private data sources of extremely large sample sizes is now increasingly available for research purposes. Statistical methods are needed for utilizing information from such big data sources while analyzing data from individual studies that may collect more detailed information required for addressing specific hypotheses of interest. We consider the problem of building regression models based on individual-level data from an ``internal" study while utilizing summary-level information, such as information on parameters for reduced models, from an ``external" big-data source. We identify a set of very constraints that link internal and external models. These constraints are used to develop a framework for semiparametric maximum likelihood inference that allows the distribution of the covariates to be estimated using either the internal sample or an external reference sample. We develop extensions for handling complex stratified sampling designs, such as case-control sampling, for the internal study. Asymptotic theory and variance estimators are developed for each case. We use simulation studies and a real data application to assess the performance of the proposed methods.

This is joint work with Nilanjan Chatterjee (Johns Hopkins), Yi-Hau Chen (Academia Sinica, Taipei) and Paige Maas (National Cancer Institute)

For further information and about building access for persons with disabilities, please contact Courtney Tillman at 773.702.8333 or send email (cmtillman@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/arc/statseminars.