Survey Methodology

Chair: Noel Cadigan (Memorial University)

J. N. K. RAO, Carleton University
Confidence Intervals for Mean and Distribution Function Under Imputation for Missing Data

Missing observations are commonly encountered in data from sample surveys due to item non-response. Imputation is used to compensate for non-response. In particular, single or fractional imputation is often used. This paper develops asymptotically valid bootstrap confidence intervals for the mean and the distribution function, using the bootstrap percentile and the bootstrap empirical likelihood methods, under fractional imputation. Imputation is done independently within classes formed on the basis of known auxiliary variables. Performance of the proposed confidence intervals is studied through simulations.

CAREN HASLER, Université de Neuchâtel
Nonparametric Imputation Method for Nonresponse in Surveys

Many imputation methods are based on statistical models that assume that the variable of interest is a noisy observation of a function of the auxiliary variables. Misspecification of this model may lead to severe errors in estimates and to misleading conclusions. A new imputation method for item nonresponse in surveys is proposed based on a nonparametric estimation of the functional dependence between the variable of interest and the auxiliary variables. We consider the use of smoothing spline estimation within an additive model framework to flexibly build an imputation model in the case of multiple auxiliary variables.

KYLE VINCENT, Bank of Canada
Estimating the Size and Distribution of Networked Populations with Snowball Sampling

A new strategy is introduced for estimating networked population characteristics. Sample selection is based on the one-wave snowball sampling design. A generalized stochastic block model is posited for the population's network topology. Inference is based on a Bayesian data augmentation procedure. This procedure has the advantage over existing methods in that it can be applied to a networked population of unknown size. An application is provided to a study of an empirical population at risk for HIV/AIDS. The results demonstrate that efficient estimates of the size and distribution of the population can be achieved with this novel strategy.

VALÉRY DONGMO JIONGO, Statistique Canada
Bootstrapping Mean Squared Errors of Robust Small Area Estimators

Robust small-area estimation has received considerable attention in recent years, and the mean squared error has been the main way in which the estimators performance is measured. This paper proposes a new bootstrap procedure for mean squared errors of robust small area estimators. We formally prove the asymptotic validity of the proposed bootstrap method and examine its finite sample performance through Monte Carlo simulations. The results show that our procedure performs reasonably well and outperforms existing ones. We also provide a real data example to illustrate the usefulness of the proposed bootstrap method in practice.

XICHEN SHE, University of Waterloo
Analysis of Ordinal Survey Responses with "Don't Know"

Ordinal responses are frequently involved in social and health survey researches to evaluate performance, attitude, severity of diseases, etc... It is also a common practice to list "Don't Know" as an option in the responses, especially for questions with sensitive nature. In this talk, we first briefly introduce approaches dealing with regular ordinal data, then explore methods

for analyzing ordinal responses with "Don't Know" as part of the response. Consistency and efficiency are compared among alternative estimators and results from a limited simulation study will be discussed.

CHRISTIANE LAPERRIERE, Statistics Canada

Data Collection Using a Diary: The Experience of the Survey of Household Spending

The Survey of Household Spending collects household expenditure data using a personal interview as well as an expenditure diary. While the diary gives us detailed expenditure information, it comes with a set of challenges that are unique to this collection mode. For example, respondents' fatigue increases with the use of diaries, which may lead to under-reporting of certain key expenditures, thus requiring further adjustments. In this talk, we discuss challenges and present the strategies we have implemented so far as well as potential strategies we seek to pursue in the future.