G35 – Research Fellows Meeting
Presenter's Abstract
21 November 2018

Raymond Chambers

NIASRA, University of Wollongong

A Linked Double-Header: Domain Estimation Under Informative Linkage and Outlier Robust Inference Using Probabilistically Linked Data

In the first half of this presentation. I will look at the situation where a finite population is partitioned into a set of domains and the aim is to estimate the average value of a target variable Y in each domain. The available auxiliary information includes aggregated domain averages, corresponding to a smaller set of strata that cross the domains of interest. Both domain and stratum affiliation are available on a sampling frame that we denote by RX. The values of Y, along with stratum affiliations, are kept on a separate unavailable register RY. An independent sample is taken from each stratum in RX. These sampled records are then probabilistically linked to records on RY (e.g. via Trusted Third Party Linkage), and the linked values Y* for these sampled records, along with their domain affiliations, are then made available for estimation of the small area means of Y. When modelling such linked data it is standard to assume that the stochastic properties of the linking process and those of the response variable Y are independent of one another given population covariate information ("non-informative linkage"). But what happens when this assumption fails? Here I investigate the robustness of different domain estimation methods where the sampling method within each stratum is biased towards more easily linked records.

In the second half of my presentation I switch to a different type of robustness, in this case robustness to outliers in linked data. In particular, I investigate an estimating equations approach to bias corrected secondary analysis of probabilistically linked data, based on a realistic scenario of dependent linkage errors in a linear regression setting. Outlier robust solutions when population auxiliary information in the form of population summary statistics is available will be presented. Simulation results show that standard outlier robust methods under an incorrect assumption of independent linkage errors can lead to insufficient linkage error bias correction, while an outlier robust approach that allows for correlated linkage errors appears to substantially correct this bias.

Both parts of my presentation are based on ongoing work with Nicola Salvati at the University of Pisa and Enrico Fabrizi at the Catholic University of the Sacred Heart of Milan.