Indirect Sampling using Dual Frame Surveys

Manuela Maia*
School of Economics and Management
Catholic University of Portugal, mmaia@porto.ucp.pt

In any survey, the random selection of the sample requires that a sampling frame is available. There are many populations, though, for which a sampling frame that lists the individual elements are not readily available.

When available, one central statistical concern for the survey researcher is how well the sampling frame actually covers the target population. In practice, perfect frames seldom exist; there are always problems that can occur, namely: (a) undercoverage, (b) duplication and (c) overcoverage.

Selecting a sample from a sampling frame that suffers from undercoverage can cause biases on survey statistics. One of the strategies to reduce coverage errors is to use multiple frames. Selecting a sample from a sampling frame that suffers from either overcoverage or duplication poses several difficulties to estimation, namely in what concerns sample weights computation.

Indirect Sampling (Lavallée, 1995) is an alternative approach to the classical sampling theory dealing with the problem of overlapping sampling frames on survey estimates.

In this paper, a new class of estimators merging multiple frames estimators (dual frames) with indirect sampling estimators to obtain a single estimator that reflects the effect of several frames on survey estimates is presented. The optimal estimator of Deville and Lavallée (2006) is compared with two classes of estimators —Domain Membership estimator and Unit Multiplicity estimator — after being translated into context of Indirect Sampling. The different kinds of cases that can occur when having two frames are presented, and a comparison is made regarding these three classes of estimators.

We also are presenting the variance of the optimal estimator of Deville and Lavallée (2006) when we consider Poisson sampling under equal and unequal inclusion probabilities. Then results of a simulation study aimed at comparing the variance of the Optimal, the Domain and the Multiplicity estimators are presented and discussed. Finally, we present the results of the non-response in links identifications and will find if the actual proposal corrections by Xu and Lavallée (2006)are enough to correct the problem in context of dual frames.