

Estimation of finite population proportions for small areas -- a statistical data integration approach

Partha Lahiri

University of Maryland College Park, U.S.A.

Abstract

Empirical best prediction (EBP) is a well-known method for producing reliable proportion estimates when the primary data source provides only small or no sample from finite populations. There are potential challenges in implementing existing EBP methodology such as limited auxiliary variables in the frame (not adequate for building a reasonable working predictive model) or unable to accurately link the sample to the finite population frame due to absence of identifiers. In this paper, we propose a new data linkage approach where the finite population frame is replaced by a big probability sample, having a large set of auxiliary variables but not the outcome binary variable of interest. We fit an assumed model on the small probability sample and then impute the outcome variable for all units of the big sample to obtain standard weighted proportions. We develop a new adjusted maximum likelihood (ML) method so that the estimate of model variance doesn't fall on the boundary, which is otherwise encountered in commonly used ML method. We also propose an estimator of the mean squared prediction error using a parametric bootstrap method and address computational issues by developing an efficient Expectation Maximization algorithm. The proposed methodology is illustrated in the context of election projection for small areas.

My talk will be based on my collaborative research with Ms. Aditi Sen.