SMALL AREA PREDICTION FOR EXPONENTIAL DISPERSION FAMILIES UNDER INFORMATIVE SAMPLING

Abdulhakeem Eideh[1] and Emily Berg[2]

[1] Al-Quds University, Palestine
e-mail: msabdul@staff.alquds.edu

[2] Iowa State University, USA
e-mail: emilyb@iastate.edu

Abstract

In complex surveys, the inclusion probability is often correlated with the response variable after conditioning on model covariates, leading to an informative design. Small area estimates are usually constructed from complex survey data. If the design is informative for the model, then procedures that ignore the design can suffer from important biases. The obvious fact that ignoring the design can render erroneous small area predictions is widely noted in the literature (Pfeffermann & Sverchkov 2007, Eideh 2002, Verret et al. 2015, You & Rao 2002, Parker et al. 2019). Past work on small area estimation under informative sampling has focused heavily on linear models or on prediction of means. We propose to generalize existing small area procedures for an informative sample design.
Concomitantly, response variables often have non-normal distributions and require nonlinear models. We develop a small area procedure that addresses both of these issues simultaneously. We develop small area predictors for the broad class of exponential dispersion families under an informative design. This class of models encompasses linear models as well as nonlinear models. We develop predictors of general parameters that may be nonlinear functions of the model response variable. We study the properties of the procedure under two models for the survey weight. We evaluate the procedures through simulation using a logistic mixed model. We then apply the methods to construct small area estimates of several functions of a wetlands indicator using data from a large-scale survey called the National Resources Inventory. Wetlands are crucial for maintaining ecosystem health. We estimate several functions of a wetlands indicator in New Jersey counties. The National Resources Inventory uses a complex design and the association between the weight and the response variable is significant.

Keywords: informative sampling, mean weight model, small area estimation.

References

Cho, Y., Guadarrama, M., Eideh, A., Molina, I. & Berg, E. (2023) Optimal predictors of nonlinear parameters under informative sampling, Manuscript under review by the Journal of Survey Statistics and Methodology .

Eideh, A. H. (2002) Estimation for longitudinal survey data under informative sampling. Unpublished Ph.D., Hebrew University of Jerusalem.

Kim, J. K. & Wang, H. (2023) A note on weight smoothing in survey sampling, Survey Methodology (accepted)

Parker, P. A., Janicki, R. & Holan, S. H. (2019) Unit level modeling of survey data for small area estimation under informative sampling: A comprehensive overview with extensions, arXiv preprint arXiv:1908.10488 .

Pfeffermann, D. & Sverchkov, M. (2007), 'Small-area estimation under informative probability sampling of areas and within the selected areas', Journal of the American Statistical Association 102(480), 1427–1439.

Rao, J. N. K. & Molina, I. (2015) Small area estimation, John Wiley & Sons.

You, Y. & Rao, J. N. K. (2002) A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights, Canadian Journal of Statistics 30(3), 431–439.

Verret, F., Rao, J. N. K. & Hidiroglou, M. A. (2015) Model-based small area estimation under informative sampling', Survey Methodology 41(2), 333–348.