

# Bayesian Inference for Sample Surveys in the Presence of High-Dimensional Auxiliary Information

Qixuan Chen

*Columbia University*

*E-mail: qc2138@cumc.columbia.edu*

**Abstract:** Survey inference can be challenged by non-representativeness of survey samples, either imperfect probability samples or non-probability samples without a probability sampling design. We consider improving survey inference with a non-representative survey sample in the presence of high-dimensional auxiliary information, which are measured in the survey sample and also available about the population via such as census data or administrative records. We propose Bayesian model-based predictive methods for estimating finite population totals by modeling the conditional distribution of the survey outcome using Bayesian additive regression trees (BARTs), which naturally handles high-dimensional auxiliary variables allowing possible interactions and nonlinear associations. Besides the auxiliary variables, inspired by Little and An (2004), we estimate the propensity score for a unit to be included in the sample using another BART. We include both the propensity score and key predictors of the survey outcome as covariates in the model to improve the estimation of population totals. We show through simulations studies and a real survey that the Bayesian model-based methods using BARTs improve survey inference.