Annals of the Institute of Statistical Mathematics (2023) 75:463–492
https://doi.org/10.1007/s10463-022-00851-5

Matrix completion under complex survey sampling

Xiaojun Mao[1] · Zhonglei Wang[2] · Shu Yang[3]

Received: 16 November 2021 / Revised: 12 August 2022 / Accepted: 17 August 2022 /
Published online: 19 September 2022

Abstract

Multivariate nonresponse is often encountered in complex survey sampling, and simply ignoring it leads to erroneous inference. In this paper, we propose a new matrix completion method for complex survey sampling. Different from existing works either conducting row-wise or column-wise imputation, the data matrix is treated as a whole which allows for exploiting both row and column patterns simultaneously. A column-space-decomposition model is adopted incorporating a low-rank structured matrix for the finite population with easy-to-obtain demographic information as covariates. Besides, we propose a computationally efficient projection strategy to identify the model parameters under complex survey sampling. Then, an augmented inverse probability weighting estimator is used to estimate the parameter of interest, and the corresponding asymptotic upper bound of the estimation error is derived. Simulation studies show that the proposed estimator has a smaller mean squared error than other competitors, and the corresponding variance estimator performs well. The proposed method is applied to assess the health status of the U.S. population.

Keywords Asymptotic upper bound · Augmented inverse probability weighting estimator · Low-rank structure · Missingness at random

✉ Zhonglei Wang
wangzl@xmu.edu.cn

Xiaojun Mao
maoxj@sjtu.edu.cn

Shu Yang
syang24@ncsu.edu

[1] School of Mathematical Sciences, Ministry of Education Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

[2] Wang Yanan Institute for Studies in Economics and School of Economics, Xiamen University, Xiamen 361005, Fujian, People's Republic of China

[3] Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA