

Combining Probability and Non-probability Samples Using Semi-parametric Quantile Regression and a Non-parametric Estimator of the Participation Probability

Emily Berg

Iowa State University

Non-probability samples are prevalent in various fields, such as biomedical studies, educational research, and business investigations, owing to the escalating challenges associated with declining response rates and the cost-effectiveness and convenience of utilizing such samples. However, relying on naive estimates derived from non-probability samples, without adequate adjustments, may introduce bias into study outcomes. Addressing this concern, data integration methodologies, which amalgamate information from both probability and non-probability samples, have demonstrated effectiveness in mitigating selection bias. Nonetheless, the efficacy of these methods hinges upon the assumptions underlying the models. This paper introduces innovative and robust data integration approaches, notably a semi-parametric quantile regression-based mass imputation approach and a doubly robust approach that integrates a non-parametric estimator of the participation probability for non-probability samples. Our proposed methodologies exhibit greater robustness compared to existing parametric approaches, particularly concerning model misspecification and outliers. Theoretical results are established, including variance estimators for our proposed estimators. Through comprehensive simulation studies and real-world applications, our findings demonstrate the promising performance of the proposed estimators in facilitating valid statistical inference. This research contributes to the advancement of robust methodologies for handling non-probability samples, thereby enhancing the reliability and validity of research outcomes across diverse domains.

Co-authors: Sixia Chen and Cindy Yu