Data integration with nonprobability sample: semiparametric model-assisted approach

Sixia Chen

University of Oklahoma

This paper introduces a novel semiparametric model-assisted estimation method that integrates data from both probability and nonprobability samples, thereby facilitating robust and efficient inferences regarding finite population parameters. To mitigate selection bias, whether ignorable or nonignorable, associated with the nonprobability sample, we propose a flexible semiparametric propensity score model that extends beyond the missing at random assumption. Our approach employs a pseudo-profile-likelihood method to estimate the propensity score model. Subsequently, a difference estimator is constructed utilizing the probability sample as a foundation, where the proxy values of the study variable for the finite population are derived from the nonprobability sample using the estimated propensity score model. We present the asymptotic properties of the proposed estimators and provide formulae for variance estimation. Through a series of simulations and a real data application, we validate our proposed estimation procedure and demonstrate its superiority over some existing estimators.