

Multiple bias calibration for valid statistical inference under nonignorable nonresponse

Z. Wang¹ and J.K. Kim² and S. Yang³

¹ Xiamen University, China
e-mail: wangzl@xmu.edu.cn

² Iowa State University, U.S.A.
e-mail: jkim@iastate.edu

³ North Carolina State University, U.S.A.
e-mail: syang24@ncsu.edu

Abstract

Valid statistical inference is challenging when the sample is subject to unknown selection bias. Data integration can be used to correct for selection bias when we have a probability sample from the same population with some common measurements. How to model and estimate the selection probability of a non-probability sample using an independent probability sample is the challenging part of the data integration. We approach this difficult problem by employing multiple candidate models for the propensity score (PS) function combined with empirical likelihood. By incorporating multiple propensity score models into the internal bias calibration constraint in the empirical likelihood setup, the selection bias can be safely eliminated so long as the multiple candidate models contain the true PS model. The bias calibration constraint for the multiple PS models in the empirical likelihood is called the multiple bias calibration. The multiple PS models can include both missing-at-random and missing-not-at-random models. Asymptotic properties are discussed and some limited simulation studies are presented to compare with the existing methods. The proposed method is applied to a real-data-based simulation platform using the Culture & Community in a Time. of Crisis (CCTC) dataset.

Keywords: Empirical likelihood; Propensity score; Selection bias.