Statistica Sinica **33** (2023), 1-23
doi:https://doi.org/10.5705/ss.202020.0446

ARC-SIN TRANSFORMATION FOR BINOMIAL SAMPLE PROPORTIONS IN SMALL AREA ESTIMATION

Masayo Y. Hirose, Malay Ghosh and Tamal Ghosh

Kyushu University, University of Florida and Citibank, N.A.

Abstract: The arc-sin transformation has long been used as a variance stabilizer for the binomial sample proportion arising out of binary data. The natural back-transformed function is useful for returning an estimate to the original scale of the parameter of interest. However, it is known that such a transformation leads to bias when estimating the original parameter of interest. In this study, we find explicit asymptotic bias-adjusted empirical Bayes (EB) estimators for binomial sample proportions in the context of small area estimation. We obtain an explicit second-order correct approximation of the mean squared errors (MSEs) of such estimators, as well as second-order correct estimators of these MSEs. Moreover, the proposed EB estimators and corresponding MSE estimators outperform their competitors in terms of the bias and variance, as demonstrated in a simulation study. We apply our methodology to real data associated with Coronavirus Disease 2019 (COVID-19) for each prefecture in Japan.

Key words and phrases: Area level model, COVID-19, linear mixed model, mean squared error estimation.

1. Introduction

Small area estimation is receiving increasing attention from both from the public and the private sectors. An important example is the small area estimation of poverty and income undertaken by the United States Bureau of the Census. Federal agencies are often mandated to produce reliable estimates for small areas, such as counties, census tracts, and school districts. Of equal importance is to provide reliable small domain estimates cross-classified by age, sex, race, and ethnicity. For more details on small area estimation, refer to Ghosh and Rao (1994), Pfeffermann (2002, 2013), and Rao and Molina (2015).

Small area estimation can either be at the area level or at the unit level. The former is more popular, because in most instances, unit level data are not available to secondary users of survey data. The classic area level model is attributed to Fay and Herriot (1979), and is essentially a mixed effects normal linear model, with

Corresponding author: Masayo Y. Hirose, Institute of Mathematics for Industry, Kyushu University, Nishi-ku, Fukuoka, Japan. E-mail: masayo@imi.kyushu-u.ac.jp.