Statistica Sinica **34** (2024), 1-20
doi:https://doi.org/10.5705/ss.202022.0181

A COMPARISON OF ESTIMATORS OF MEAN AND ITS FUNCTIONS IN FINITE POPULATIONS

Anurag Dey* and Probal Chaudhuri

Indian Statistical Institute

Abstract: We investigate several well-known estimators of finite population means and the functions of these means under standard sampling designs. Such functions include the variance, correlation coefficient, and regression coefficient in the population as special cases. We compare the performance of these estimators under different sampling designs, based on their asymptotic distributions. We construct equivalence classes of estimators under different sampling designs so that estimators in the same class have equivalent performance in terms of the asymptotic mean squared error (MSE). We then compare estimators from different equivalence classes under superpopulations that satisfy linear models. We show that the pseudo empirical likelihood (PEML) estimator of the population mean under simple random sampling without replacement (SRSWOR) has the lowest asymptotic MSE of the estimators considered here. In addition, for the variance, correlation coefficient, and regression coefficient of the population, the plug-in estimators based on the PEML estimator have the lowest asymptotic MSEs under SRSWOR. However, for any high entropy πPS sampling design, which uses auxiliary information, the plug-in estimators based on the Hájek estimator have the lowest asymptotic MSEs.

Key words and phrases: Asymptotic normality, equivalence classes of estimators, high entropy sampling designs, inclusion probability, linear regression model, rejective sampling design, relative efficiency, superpopulation models.

1. Introduction

Suppose that $\mathcal{P} = \{1, 2, \ldots, N\}$ is a finite population of size N, s is a sample of size n ($< N$) from $\mathcal{P}$, and $\mathcal{S}$ is the collection of all possible samples of size n. Then, a sampling design $P(s)$ is a probability distribution on $\mathcal{S}$ such that $0 \leq P(s) \leq 1$ for all $s \in \mathcal{S}$ and $\sum_{s\in\mathcal{S}} P(s) = 1$. In this study, we consider the following designs: simple random sampling without replacement (SRSWOR), the Lahiri-Midzuno-Sen (LMS) sampling design (see Lahiri (1951); Midzuno (1952); Sen (1953)), the Rao-Hartley-Cochran (RHC) sampling design (see Rao, Hartley and Cochran (1962)), and high entropy πPS (HEπPS) sampling designs (see Section 2). Note that all of the above sampling designs other than SRSWOR use some auxiliary variable.

Let (Y_i, X_i) be the value of (y, x) for the ith population unit, for $i = 1, \ldots, N$, where y is a univariate or multivariate study variable, and x is a positive

*Corresponding author.