PLOS ONE

RESEARCH ARTICLE

Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling

Sardar Hussain[1], Sohaib Ahmad[1], Mariyam Saleem[2], Sohail Akhtar[2]*

1 Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan, **2** Department of Statistics, Government College Univerisity, Lahore, Pakistan

* s.akhtar@gcu.edu.pk

Abstract

The main purpose of this paper is to propose two new estimators for estimating the finite population distribution function under simple and stratified random sampling schemes using supplementary information on the distribution function, mean and ranks of the auxiliary variable. The mathematical expressions for the bias and mean squared error of the proposed estimators are derived under the first order of approximation. The theoretical and empirical studies showed that the proposed estimators uniformly perform better than the existing estimators in terms of the percentage relative efficiency.

OPEN ACCESS

Citation: Hussain S, Ahmad S, Saleem M, Akhtar S (2020) Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. PLoS ONE 15(9): e0239098. https://doi.org/10.1371/journal.pone.0239098

Editor: Feng Chen, Tongii University, CHINA

Received: June 30, 2020

Accepted: August 29, 2020

Published: September 28, 2020

Data Availability Statement: The data are all contained within the manuscript.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

In survey sampling, the use of suitable auxiliary information improves the precision of estimators of the unknown population parameter(s). Several estimators of population parameters, including the population mean, median, total, distribution function, quantiles, etc., exist in the literature, and requires, supplementary information on one or more auxiliary variables along with the information on the study variable. A number of studies have been published on the estimation of the population mean. Some important references to the population mean estimation using auxiliary information include Murthy [1], Sisodia and Dwivedi [2], Srivastava and Jhajj [3], Rao [4], Upadhyaya and Singh [5], Singh [6], Kadilar and Cingi [7], Kadilar and Cingi [8], Gupta and Shabbir [9], Grover and Kaur [10], Grover and Kaur [11], Lu [12], Muneer et al. [13], Shabbir and Gupta [14], and Gupta and Yadav [15]. In these studies, the authors have proposed improved ratio, product, and regression type estimators for estimating the finite population mean. These authors have used a single auxiliary variable for the estimation procedure. In a recent study, Haq et al. [16] suggested using ranks of the auxiliary varible as an additional auxiliary variable to increase the precision of the estimator of the population mean in simple random sampling. However, to the best of our knowledge, there is no study concerning the use of two auxiliary information variables for the estimation of the finite population distribution function.

The problem of estimating the finite population cumulative distribution function (CDF) arises when the interest lies in finding out the proportion of the values of the study variable that are less or equal to a certain value. There are situations where estimating the CDF is deemed necessary. For example, for a nutritionist, it is interesting to know the proportion of the population that consumes 25% or more of the calorie intake from saturated fat. In the literature, many authors have estimated the CDF using information about one or more auxiliary variables. Chambers and Dunstan [17] suggested an estimator for estimating the CDF that requires information both on the study and auxiliary variables. Similarly, Rao et al. [18] and Rao [19] proposed ratio and difference/regression estimators for estimating the CDF under a general sampling design. Kuk [20] suggested a kernel method for estimating the CDF using the auxiliary information. Ahmed and Abu-Dayyeh [21] estimated the CDF using the information on multiple auxiliary variables. Rueda et al. [22] used a calibration approach to develop an estimator for estimating the CDF. Singh et al. [23] considered the problem of estimating the CDF and quantiles with the use of at the estimation stage of a survey. Moreover, Yaqub and Shabbir [24] considered a generalised class of estimators for estimating the CDF in the presence of non-response. Chen and Chen [25] investigated the injury severities of truck drivers in single-and multi-vehicle accidents on rural highways while Zeng et al. [26] worked on a multivariate random-parameter Tobit model for analysing highway crash rates by injury severity, and Yaqub and Shabbir [24] considered a generalised class of estimators for estimating the CDF in the presence of non-response. Dong et al. [27] investigated the differences of single-vehicle and multi-vehicle accident probability using a mixed logit model, Chen et al. [28] worked on an analysis of hourly crash likelihood using an unbalanced panel data mixed logit model and real-time driving environmental big data, Zeng et al. [29] suggested jointly modelling area-level crash rates by severity, and Zeng et al. [30] used spatial joint analysis for zonal daytime and night-time crash frequencies using a Bayesian bivariate conditional autoregressive model. However, these estimators only used one auxiliary variate.

In this paper, we propose two new families of estimators for estimating the CDF using the information on the distribution function, ranks, and mean of the auxiliary variable under simple random sampling and stratified random sampling. The bias and mean squared errors (MSEs) of the existing and proposed estimators of the CDF are derived under the first order of approximation. The theoretical and numerical comparisons showed that the proposed estimators are more precise than the existing adapted estimators when estimating the CDF of a finite population.

2 Notation in simple random sampling

Consider a finite population $\Omega = \{1,2,\ldots,N\}$ of N distinct units. In order to estimate the finite population distribution function, a sample of size n units is drawn from Ω using simple random sampling without replacement. Suppose Y and X are the study and auxiliary variables, respectively. Let Z denote the ranks of X, $I(Y\leq y)$ is indicator variable based on Y and $I(X\leq x)$ is indicator variable based on X. Similarly, $F(y) = \sum_{i=1}^{N} I(Y_i \leq y)/N$ and $\hat{F}(y) = \sum_{i=1}^{n} I(Y_i \leq y)/n$ ($F(x) = \sum_{i=1}^{N} I(X_i \leq x)/N$ and $\hat{F}(x) = \sum_{i=1}^{n} I(X_i \leq x)/n$) are the population and sample distribution functions of Y (X), respectively. Let $\bar{X} = \sum_{i=1}^{N} X_i/N$ and $\hat{\bar{X}} = \sum_{i=1}^{n} X_i/n$ ($\bar{Z} = \sum_{i=1}^{N} Z_i/N$ and $\hat{\bar{Z}} = \sum^{i=1n} Z_i/n$) are the population and sample means of X (Z), respectively.

In order to obtain the biases and mean squared errors (MSEs) of the adapted and proposed estimators of F(y), we consider the following relative error terms. Let

$$e_1 = \frac{\hat{F}(y) - F(y)}{F(y)}, e_2 = \frac{\hat{F}(x) - F(x)}{F(x)}, e_3 = \frac{\hat{\bar{X}} - \bar{X}}{\bar{X}} \text{ and } e_4 = \frac{\hat{\bar{Z}} - \bar{Z}}{\bar{Z}},$$

such that $e_i = 0$ for $i = 0,1,2,3$, where $E(\cdot)$ is the mathematical expectation of $(\cdot)$. Let

$$V_{rstu} = E[e_1^r\ e_2^s\ e_3^t\ e_4^u],$$

$$E(e_1^2) = \lambda C_1^2 = V_{2000}, E(e_2^2) = \lambda C_2^2 = V_{0200}, E(e_3^2) = \lambda C_3^2 = V_{0020}, E(e_4^2) = \lambda C_4^2 = V_{0002},$$

$$E(e_1 e_2) = \lambda R_{12} C_1 C_2 = V_{1100}, E(e_1 e_3) = \lambda R_{13} C_1 C_3 = V_{1010}, E(e_1 e_4) = \lambda R_{14} C_1 C_4 = V_{1001},$$

$$E(e_2 e_3) = \lambda R_{23} C_2 C_3 = V_{0110}, E(e_2 e_4) = \lambda R_{24} C_2 C_4 = V_{0101},$$

where $\lambda = (N-n)/(nN)$, $S_1^2 = \sum_{i=1}^N (I(Y_i \leq y) - F(y))^2/(N-1)$,
$S_2^2 = \sum_{i=1}^N (I(X_i \leq x) - F(x))^2/(N-1)$,
$S_3^2 = \sum_{i=1}^N (X_i - \bar{X})^2/(N-1)$,
$S_4^2 = \sum_{i=1}^N (Z_i - \bar{Z})^2/(N-1)$,
$C_1 = S_1/F(y)$, $C_2 = S_2/F(x)$,
$C_3 = S_3/\bar{X}$, $C_4 = S_4/\bar{Z}$, $R_{12} = S_{12}/(F(y)F(x))$,
$R_{13} = S_{13}/(F(y)\bar{X})$, $R_{23} = S_{23}/(F(y)\bar{Z})$, $R_{14} = S_{14}/(F(x)\bar{X})$, $R_{24} = S_{24}/(F(x)\bar{Z})$,
$S_{12} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(I(X_i \leq x) - F(x))\}/(N-1)$,
$S_{13} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(X_i - \bar{X})\}/(N-1)$,
$S_{23} = \sum_{i=1}^N \{(I(X_i \leq x) - F(x))(X_i - \bar{X})\}/(N-1)$,
$S_{14} = \sum_{i=1}^N \{(I(Y_i \leq y) - F(y))(Z_i - \bar{Z})\}/(N-1)$,
$S_{24} = \sum_{i=1}^N \{(I(X_i \leq x) - F(x))(Z_i - \bar{Z})\}/(N-1)$.

In addition, let $R_{1.23}^2 = (R_{12}^2 + R_{13}^2 - 2R_{12}R_{13}R_{23})/(1 - R_{23}^2)$ be the multiple correlation coefficient of $I(Y \leq y)$ on $I(X \leq x)$ and X, and let $R_{1.24}^2 = (R_{12}^2 + R_{14}^2 - 2R_{12}R_{14}R_{24})/(1 - R_{24}^2)$ be the multiple correlation coefficient of $I(Y \leq y)$ on $I(X \leq x)$ and Z, under simple random sampling.

3 Adapted estimators in simple random sampling

In this section, some estimators of finite population mean are adapted for estimating the finite CDF under simple random sampling. The biases and MSEs of these adapted estimators are derived under the first order of approximation.

1. The traditional unbiased estimator of $F(y)$ is

$$\hat{F}_1(y) = \frac{1}{n}\sum_{i=1}^n I(Y_i \leq y). \tag{1}$$

The variance of $\hat{F}_1(y)$ is

$$\mathrm{Var}(\hat{F}_1(y)) = F^2(y)V_{2000}. \tag{2}$$

2. Cochran [31] adapted ratio estimator of $F(y)$ is

$$\hat{F}_2(y) = \hat{F}(y)\left(\frac{F(x)}{\hat{F}(x)}\right). \tag{3}$$

The bias and MSE of $\hat{F}_2(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_2(y)) \cong F(y)(V_{0200} - V_{1100}),$$

$$\text{MSE}(\hat{F}_2(y)) \cong F^2(y)(V_{2000} + V_{0200} - 2V_{1100}). \tag{4}$$

If $R_{12} > C_2/(2C_1)$, then $\hat{F}_2(y)$ is better than $\hat{F}_1(y)$ in terms of MSE.

3. Murthy [32] adapted product estimator of $F(y)$ is

$$\hat{F}_3(y) = \hat{F}(y)\left(\frac{\hat{F}(x)}{F(x)}\right). \tag{5}$$

The bias and MSE of $\hat{F}_3(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_3(y)) = F(y)V_{1100},$$

$$\text{MSE}(\hat{F}_3(y)) \cong F^2(y)(V_{2000} + V_{0200} + 2V_{1100}). \tag{6}$$

If $-C_2/(2C_1) > R_{12}$, then $\hat{F}_3(y)$ is better than $\hat{F}_1(y)$ in terms of MSE.

4. The adapted difference estimator of $F(y)$ is

$$\hat{F}_4(y) = \hat{F}(y) + k(F(x) - \hat{F}(x)). \tag{7}$$

where k is an unknown constant. Here, $\hat{F}_4(y)$ is an unbiased estimator of $F(y)$. The minimum variance of $\hat{F}_4(y)$ at the optimum value $k_{(opt)} = (F(y)V_{1100})/(F(x)V_{0200})$ is

$$\text{Var}_{\min}(\hat{F}_4(y)) = \frac{F^2(y)(V_{2000}V_{0200} - V_{1100}^2)}{V_{0200}}. \tag{8}$$

Here, (8) may be written as

$$\text{Var}_{\min}(\hat{F}_4(y)) = F^2(y)V_{2000}(1 - R_{12}^2). \tag{9}$$

5. Rao [4] adapted difference-type estimator of $F(y)$ is

$$\hat{F}_5(y) = k_1\hat{F}(y) + k_2(F(x) - \hat{F}(x)), \tag{10}$$

where k_1 and k_2 are unknown constants. The bias and MSE of $\hat{F}_5(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_5(y)) = F(y)(k_1 - 1),$$

$$\text{MSE}(\hat{F}_5(y)) \cong F^2(y) - 2k_1F^2(y) + k_1^2F^2(y) + k_1^2F^2(y)V_{2000}$$

$$-2k_1k_2F(y)F(x)V_{1100} + k_2^2F^2(x)V_{0200}. \tag{11}$$

The optimum values of k_1 and k_2, determined by minimizing (11), are

$$k_{1(\mathrm{opt})} = \frac{V_{0200}}{(V_{0200}V_{2000} - V_{1100}^2 + V_{0200})}$$

$$k_{2(\mathrm{opt})} = \frac{F(y)V_{1100}}{F(x)(V_{2000}V_{0200} - V_{1100}^2 + V_{0200})}.$$

The minimum MSE of $\hat{F}_5(y)$ at the optimum values of k_1 and k_2 is

$$\mathrm{MSE}_{\min}\left(\hat{F}_5(y)\right) = \frac{F^2(y)(V_{2000}V_{0200} - V_{1100}^2)}{(V_{2000}V_{0200} - V_{1100}^2 + V_{0200})}. \quad (12)$$

Here, (12) may be written as

$$\mathrm{MSE}_{\min}\left(\hat{F}_5(y)\right) = \frac{F^2(y)V_{2000}(1 - R_{12}^2)}{1 + V_{2000}(1 - R_{12}^2)}. \quad (13)$$

6. Singh et al. [33] adapted generalized ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_6(y) = \hat{F}(y)\exp\left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) + \hat{F}(x)) + 2b}\right), \quad (14)$$

where a and b are known constants. The bias and MSE of $\hat{F}_6(y)$, to the first order of approximation, are

$$\mathrm{Bias}\left(\hat{F}_6(y)\right) \cong F(y)\left(\frac{3}{8}\theta^2 V_{0200} - \frac{1}{2}\theta V_{1100}\right),$$

$$\mathrm{MSE}\left(\hat{F}_6(y)\right) \cong \frac{F^2(y)}{4}\left(4V_{2000} + \theta^2 V_{0200} - 4\theta V_{1100}\right), \quad (15)$$

where $\theta = aF(x)/(aF(x)+b)$.

7. Grover and Kaur [11] adapted generalized class of ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_7(y) = \left\{k_3\hat{F}(y) + k_4(F(x) - \hat{F}(x))\right\}\exp\left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) + \hat{F}(x)) + 2b}\right), \quad (16)$$

where k_3 and k_4 are unknown constants. The bias and MSE of $\hat{F}_7(y)$, to the first order of

approximation, are

$$\text{Bias}\left(\hat{F}_7(y)\right) \cong F(y)(k_3 - 1) + \frac{3}{8}\theta^2 k_3 F(y) + \frac{1}{2}\theta k_4 F(x) V_{0200} - \frac{1}{2}\theta F(y) V_{1100},$$

$$\begin{aligned}\text{MSE}(\hat{F}_7(y)) \cong\ & k_4^2 F^2(x) V_{0200} + k_3^2 F^2(y) V_{2000} + 2\theta k_3 k_4 F(y) F(x) V_{0200}\\ & -2k_3 k_4 F(y) F(x) V_{1100} + F^2(y) - 2k_3 F^2(y) + \theta k_3^2 F^2(y)\\ & +k_3 F^2(y) V_{1100} - \theta k_4 F(y) F(x) V_{0200} - 2\theta k_3^2 F^2(y) V_{1100}\\ & -\frac{3}{4}\theta^2 k_3 F^2(y) V_{0200} + \theta^2 k_3^2 F^2(y) V_{0200}\end{aligned} \tag{17}$$

The optimum values of k_3 and k_4, determined by minimizing (17), are

$$k_{3(\text{opt})} = \frac{V_{0200}(\theta^2 V_{0200} - 8)}{8(-V_{2000} V_{0200} + V_{1100}^2 - V_{0200})},$$

$$k_{4(\text{opt})} = \frac{F(y)(\theta^3 V_{0200}^2 - \theta^2 V_{0200} V_{1100} + 4\theta V_{2000} V_{0200} - 4\theta V_{1100}^2 - 4\theta V_{0200} + 8 V_{1100})}{8F(x)(V_{2000} V_{0200} - V_{1100}^2 + V_{0200})}.$$

The simplified minimum MSE of $\hat{F}_7(y)$ at the optimum values of k_3 and k_4 is

$$\text{MSE}_{\min}\left(\hat{F}_7(y)\right) \cong \frac{F^2(y)}{64}\left(64 - 16\theta^2 V_{0200} - \frac{V_{0200}(-8 + \theta^2 V_{0200})^2}{V_{0200}(1 + V_{2000}) - V_{1100}^2}\right). \tag{18}$$

Here, (18) may be written as

$$\text{MSE}_{\min}\left(\hat{F}_7(y)\right) \cong \text{Var}_{\min}\left(\hat{F}_4(y)\right) - \frac{F^2(y)(\theta^2 V_{0200}^2 - 8V_{1100}^2 + 8V_{0200} V_{2000})^2}{64 V_{0200}^2\{1 + V_{2000}(1 - R_{12}^2)\}}, \tag{19}$$

which shows that $\hat{F}_7(y)$ is more precise than $\hat{F}_4(y)$.

4 Proposed estimators in simple random sampling

The precision of an estimator increases by using suitable auxiliary information at the estimation stage. In previous studies, the sample distribution function of the auxiliary variable was used to improve the efficiencies of the existing distribution function estimators. In a recent study, Haq et al. [16] suggested using ranks of the auxiliary variable as an additional auxiliary variable to increase the precision of an estimator of the population mean. On similar lines, we use additional auxiliary information on sample means of the auxiliary and ranked-auxiliary variables along with the sample distribution function estimators of $F(y)$ and $F(x)$ to estimate the finite CDF. For this purpose, we propose two families of estimators for estimating $F(y)$.

4.1 First proposed family of estimators

For the first family of estimators, the sample mean of the auxiliary variable is used as an additional auxiliary variable; whilst in the second family of estimators, the sample mean of the ranked auxiliary variable is used as an additional auxiliary variable.

On the lines of $\hat{F}_5(y)$ and $\hat{F}_6(y)$, first proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_8(y) = \left\{ k_5\hat{F}(y) + k_6\left(\frac{F(x) - \hat{F}(x)}{F(x)}\right) + k_7\left(\frac{\bar{X} - \hat{\bar{X}}}{\bar{X}}\right) \right\} \exp\left(\frac{a(F(x) - \hat{F}(x))}{a(F(x) + \hat{F}(x)) + 2b}\right), \quad (20)$$

where k_5, k_6 and k_7 are unknown constants, $a(\neq 0)$ and b are either two real numbers or function of known population parameters of $I(X \leq x)$, like R_{12}, β_2 (coefficient of kurtosis), C_2, etc.

The estimator $\hat{F}_8(y)$ can also be written as

$$\hat{F}_8(y) = \{k_5F(y)(1 + e_1) - k_6e_2 - k_7e_3\}\left(1 - \frac{1}{2}\theta e_2 + \frac{3}{8}\theta^2 e_2^2 + \cdots\right). \quad (21)$$

Simplifying (21) and keeping terms only up to the second power of e_i's, we can write

$$\left(\hat{F}_8(y) - F(y)\right) = -F(y) + k_5F(y) + k_5F(y)e_1 - \frac{1}{2}\theta k_5F(y)e_2 - k_6e_2 - k_7e_3$$

$$+\frac{3}{8}\theta^2 k_5F(y)e_2^2 + \frac{1}{2}\theta k_6e_2^2 - \frac{1}{2}\theta k_5F(y)e_1e_2 + \frac{1}{2}\theta k_7e_2e_3. \quad (22)$$

The bias and MSE of $\hat{F}_8(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_8(y)\right) \cong F(y)(k_5 - 1) + \frac{3}{8}\theta^2 k_5F(y)V_{0200} + \frac{1}{2}\theta k_6V_{0200} - \frac{1}{2}\theta k_5F(y)V_{1100} + \frac{1}{2}\theta k_7V_{0110},$$

$$\text{MSE}(\hat{F}_8(y)) \cong F^2(y)(k_5 - 1)^2 + k_5^2F^2(y)V_{2000} + k_6^2V_{0200} + k_7^2V_{0020} + \theta^2k_5^2F^2(y)V_{0200}$$

$$-\theta k_6F(y)V_{0200} + 2\theta k_5k_6F(y)V_{0200} - \frac{3}{4}\theta^2k_5F^2(y)V_{0200} + \theta k_5F^2(y)V_{1100}$$

$$-2\theta k_5^2F^2(y)V_{1100} - 2k_5k_6F(y)V_{1100} - 2k_5k_7F(y)V_{1010} - \theta k_7F(y)V_{0110}$$

$$+2\theta k_5k_7F(y)V_{0110} - 2k_6k_7V_{0110}. \quad (23)$$

The optimum values of k_5, k_6 and k_7, determined by minimizing (23), are

$$k_{5(\text{opt})} = \frac{8 - \theta^2V_{0200}}{8\{1 + V_{2000}(1 - R_{1.23}^2)\}},$$

$$k_{6(\text{opt})} = \frac{F(y)\left[\begin{array}{c}\theta^3V_{0200}^{3/2}(R_{23}^2 - 1) + V_{2000}^{1/2}(-8 + \theta^2V_{0200})(R_{12} - R_{23}R_{13})\\ +4\theta V_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + V_{2000}(1 - R_{1.23}^2)\}\end{array}\right]}{8V_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + V_{2000}(1 - R_{1.23}^2)\}},$$

$$k_{7(\text{opt})} = \frac{F(y)V_{2000}^{1/2}(8 - \theta^2V_{0200})(R_{12} - R_{23}R_{13})}{8V_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + V_{2000}(1 - R_{1.23}^2)\}}.$$

The simplified minimum MSE of $\hat{F}_8(y)$ at the optimum values of k_5, k_6 and k_7 is

$$\mathrm{MSE}_{\min}\left(\hat{F}_8(y)\right) \cong \frac{F^2(y)\{64V_{2000}(1-R^2_{1.23}) - \theta^4 V^2_{0200} - 16\theta^2 V_{0200}V_{2000}(1-R^2_{1.23})\}}{64\{1+V_{2000}(1-R^2_{1.23})\}}, \quad (24)$$

where $R^2_{1.23} = \left(\frac{V^2_{1100}V_{0020}+V^2_{1010}V_{0200}-2V_{1010}V_{1100}V_{0110}}{V_{2000}(V_{0200}V_{0020}-V^2_{0110})}\right)$.

Here, (24) may be written as

$$\mathrm{MSE}_{\min}(\hat{F}_8(y)) \cong \mathrm{Var}_{\min}(\hat{F}_4(y)) - H_1 - H_2, \quad (25)$$

where

$$H_1 = \frac{F^2(y)(\theta^2 V^2_{0200} - 8V^2_{1100} + 8V_{0200}V_{2000})^2}{64V^2_{0200}\{1+V_{2000}(1-R^2_{12})\}} \text{ and}$$

$$H_2 = \frac{F^2(y)(\theta^2 V_{0200} - 8)^2(V_{0200}V_{1010} - V_{0110}V_{1100})^2}{64V^2_{0200}V_{0020}(1-R^2_{23})\{1+V_{2000}(1-R^2_{12})\}\{1+V_{2000}(1-R^2_{1.23})\}}.$$

It can be seen that $\hat{F}_8(y)$ is more precise than $\hat{F}_4(y)$.

4.2 Second proposed family of estimators

On similar lines, second proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_9(y) = \left\{k_8\hat{F}(y) + k_9\left(\frac{F(x)-\hat{F}(x)}{F(x)}\right) + k_{10}\left(\frac{\bar{Z}-\hat{\bar{Z}}}{\bar{Z}}\right)\right\}\exp\left(\frac{a(F(x)-\hat{F}(x))}{a(F(x)-\hat{F}(x))+2b}\right), \quad (26)$$

where k_8, k_9 and k_{10} are unknown constants, $a(\neq 0)$ and b are either two real numbers or functions of known population parameters of $I(X \leq x)$, like R_{12}, β_2 (coefficient of kurtosis), C_2, etc. The estimator $\hat{F}_9(y)$ can also be written as

$$\hat{F}_9(y) = \{k_8F(y)(1+e_1) - k_9e_2 - k_{10}e_4\}\left(1 - \frac{1}{2}\theta e_2 + \frac{3}{8}\theta^2 e_2^2 + \cdots\right). \quad (27)$$

Simplifying (27) and keeping terms only up to the second power of e_i's, we can write

$$\left(\hat{F}_9(y) - F(y)\right) = -F(y) + k_8F(y) + k_8F(y)e_1 - \frac{1}{2}\theta k_8F(y)e_2 - k_9e_2 - k_{10}e_4$$

$$+\frac{3}{8}\theta^2 k_8F(y)e_2^2 + \frac{1}{2}\theta k_9e_2^2 - \frac{1}{2}\theta k_8F(y)e_1e_2 + \frac{1}{2}\theta k_{10}e_2e_4. \quad (28)$$

The bias and MSE of $\hat{F}_9(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_9(y)\right) \cong F(y)(k_8 - 1) + \frac{3}{8}\theta^2 k_8 F(y) V_{0200} + \frac{1}{2}\theta k_9 V_{0200} - \frac{1}{2}\theta k_8 F(y) V_{1100} + \frac{1}{2}\theta k_9 V_{0101},$$

$$\begin{aligned}\text{MSE}(\hat{F}_9(y)) \cong\; & F^2(y)(k_8 - 1)^2 + k_8^2 F^2(y) V_{2000} + k_9^2 V_{0200} + k_{10}^2 V_{0002} + \theta^2 k_8^2 F^2(y) V_{0200} \\ & -\theta k_9 F(y) V_{0200} + 2\theta k_8 k_9 F(y) V_{0200} - \frac{3}{4}\theta^2 k_8 F^2(y) V_{0200} + \theta k_8 F^2(y) V_{1100} \\ & -2\theta k_8^2 F^2(y) V_{1100} - 2k_8 k_9 F(y) V_{1100} - 2k_8 k_{10} F(y) V_{1001} - \theta k_{10} F(y) V_{0101} \\ & +2\theta k_8 k_{10} F(y) V_{0101} - 2k_9 k_{10} V_{0101}.\end{aligned} \tag{29}$$

The optimum values of k_8, k_9 and k_{10}, determined by minimizing (29), are

$$k_{8(\text{opt})} = \frac{8 - \theta^2 V_{0200}}{8\{1 + V_{2000}(1 - R_{1.24}^2)\}},$$

$$k_{9(\text{opt})} = \frac{F(y)\left[\begin{array}{l}\theta^3 V_{0200}^{3/2}(R_{24}^2 - 1) + V_{2000}^{1/2}(-8 + \theta^2 V_{0200})(R_{12} - R_{24}R_{14}) \\ +4\theta V_{0200}^{1/2}(R_{24}^2 - 1)\{-1 + V_{2000}(1 - R_{1.24}^2)\}\end{array}\right]}{8V_{0200}^{1/2}(R_{24}^2 - 1)\{-1 + V_{2000}(1 - R_{1.24}^2)\}},$$

$$k_{10(\text{opt})} = \frac{F(y) V_{2000}^{1/2}(8 - \theta^2 V_{0200})(R_{12} - R_{24}R_{14})}{8V_{0200}^{1/2}(R_{24}^2 - 1)\{-1 + V_{2000}(1 - R_{1.24}^2)\}}.$$

The simplified minimum MSE of $\hat{F}_9(y)$ at the optimum values of k_8, k_9 and k_{10} is

$$\text{MSE}_{\min}\left(\hat{F}_9(y)\right) \cong \frac{F^2(y)\{64V_{2000}(1 - R_{1.24}^2) - \theta^4 V_{0200}^2 - 16\theta V_{0200} V_{2000}(1 - R_{1.24}^2)\}}{64\{1 + V_{2000}(1 - R_{1.24}^2)\}}, \tag{30}$$

where $R_{1.24}^2 = \left(\frac{V_{1100}^2 V_{0002} + V_{1001}^2 V_{0200} - 2V_{1001} V_{1100} V_{0101}}{V_{2000}(V_{0200} V_{0002} - V_{0101}^2)}\right)$.

Here, (30) may be written as

$$\text{MSE}_{\min}(\hat{F}_9(y)) \cong \text{Var}_{\min}(\hat{F}_4(y)) - H_1 - H_3, \tag{31}$$

where

$$H_1 = \frac{F^2(y)(\theta^2 V_{0200}^2 - 8V_{1100}^2 + 8V_{0200})^2}{64V_{0200}^2\{1 + V_{2000}(1 - R_{12}^2)\}} \text{ and}$$

$$H_3 = \frac{F^2(y)(\theta^2 V_{0200} - 8)^2 (V_{0200} V_{1001} - V_{0101} V_{1100})^2}{64V_{0200}^2 V_{0002}(1 - R_{24}^2)\{1 + V_{2000}(1 - R_{12}^2)\}\{1 + V_{2000}(1 - R_{1.24}^2)\}}.$$

It is clear that $\hat{F}_9(y)$ is more precise than $\hat{F}_4(y)$.

In Table 1, we put some members of the Singh et al. [33], Grover and Kaur [11], and proposed families of estimators with selected choices of a and b.

Table 1. Some members of the adapted and proposed distribution function estimators.

a	**b*b***	$\hat{F}_6(y)$	$\hat{F}_7(y)$	$\hat{F}_8(y)$	$\hat{F}_9(y)$
1	CC_2	$\hat{F}_6^{(1)}(y)$	$\hat{F}_7^{(1)}(y)$	$\hat{F}_8^{(1)}(y)$	$\hat{F}_9^{(1)}(y)$
1	β_2	$\hat{F}_6^{(1)}(y)$	$\hat{F}_7^{(1)}(y)$	$\hat{F}_8^{(1)}(y)$	$\hat{F}_9^{(1)}(y)$
β_2	C_2	$\hat{F}_6^{(3)}(y)$	$\hat{F}_7^{(3)}(y)$	$\hat{F}_8^{(3)}(y)$	$\hat{F}_9^{(3)}(y)$
C_2	β_2	$\hat{F}_6^{(4)}(y)$	$\hat{F}_7^{(4)}(y)$	$\hat{F}_8^{(4)}(y)$	$\hat{F}_9^{(4)}(y)$
1	R_{12}	$\hat{F}_6^{(5)}(y)$	$\hat{F}_7^{(5)}(y)$	$\hat{F}_8^{(5)}(y)$	$\hat{F}_9^{(5)}(y)$
CC_2	R_{12}	$\hat{F}_6^{(6)}(y)$	$\hat{F}_7^{(6)}(y)$	$\hat{F}_8^{(6)}(y)$	$\hat{F}_9^{(6)}(y)$
R_{12}	C_2	$\hat{F}_6^{(7)}(y)$	$\hat{F}_7^{(7)}(y)$	$\hat{F}_8^{(7)}(y)$	$\hat{F}_9^{(7)}(y)$
β_2	R_{12}	$\hat{F}_6^{(8)}(y)$	$\hat{F}_7^{(8)}(y)$	$\hat{F}_8^{(8)}(y)$	$\hat{F}_9^{(8)}(y)$
R_{12}	β_2	$\hat{F}_6^{(9)}(y)$	$\hat{F}_7^{(9)}(y)$	$\hat{F}_8^{(9)}(y)$	$\hat{F}_9^{(9)}(y)$
1	$NF(x)$	$\hat{F}_6^{(10)}(y)$	$\hat{F}_7^{(10)}(y)$	$\hat{F}_8^{(1)}(y)$	$\hat{F}_9^{(10)}(y)$

https://doi.org/10.1371/journal.pone.0239098.t001

5 Efficiency comparisons in simple random sampling

In this section, the adapted and proposed estimators of $F(y)$ are compared in terms of the minimum MSEs. [(i)]

1. From (2) and (25),

$$\mathrm{MSE}_{\min}(\hat{F}_8(y)) < \mathrm{Var}(\hat{F}_1(y))\mathrm{if}$$

$$F^2(y)V_{2000}R_{12}^2 + H_1 + H_2 > 0.$$

2. From (4) and (25),

$$\mathrm{MSE}_{\min}(\hat{F}_8(y)) < \mathrm{MSE}(\hat{F}_2(y))\mathrm{if}$$

$$\frac{F^2(y)}{V_{2000}}(V_{0200} - V_{1100})^2 + H_1 + H_2 > 0.$$

3. From (6) and (25),

$$\mathrm{MSE}_{\min}(\hat{F}_8(y)) < \mathrm{MSE}(\hat{F}_3(y))\mathrm{if}$$

$$\frac{F^2(y)}{V_{2000}}(V_{0200} + V_{1100})^2 + H_1 + H_2 > 0.$$

4. From (9) and (25),

$$\mathrm{MSE}_{\min}(\hat{F}_8(y)) < \mathrm{MSE}_{\min}(\hat{F}_4(y))\mathrm{if}$$

$$H_1 + H_2 > 0.$$

5. From (13) and (25),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_5(y))\text{if}$$

$$\frac{F^2(y)\theta^2 V_{0200}\{\theta^2 V_{0200} + 16V_{2000}(1 - R_{12}^2)\}}{64\{1 + V_{2000}(1 - R_{12}^2)\}} + H_2 > 0.$$

6. From (15) and (25),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}(\hat{F}_6(y))\text{if}$$

$$\frac{F^2(y)}{V_{2000}}\left(\frac{\theta V_{0200}}{2} - V_{1100}\right)^2 + H_1 + H_2 > 0.$$

7. From (19) and (25),

$$\text{MSE}_{\min}(\hat{F}_8(y)) < \text{MSE}_{\min}(\hat{F}_7(y))\text{if}$$

$$H_2 > 0.$$

8. From (2) and (31),

$$\text{MSE}_{\min}(\hat{F}_9(y)) < \text{Var}(\hat{F}_1(y))\text{if}$$

$$\frac{1}{V_{0200}}F^2(y)V_{1100}^2 + H_1 + H_3 > 0.$$

9. From (4) and (31),

$$\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}(\hat{F}_2(y))\text{if}$$

$$\frac{F^2(y)}{V_{2000}}(V_{0200} - V_{1100})^2 + H_1 + H_3 > 0.$$

10. From (6) and (31),

$$\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}(\hat{F}_3(y))\text{if}$$

$$\frac{F^2(y)}{V_{2000}}(V_{0200} + V_{1100})^2 + H_1 + H_3 > 0.$$

11. From (9) and (31),

$$\text{MSE}_{\min}(\hat{F}_9(y)) < \text{MSE}_{\min}(\hat{F}_4(y))\text{if}$$

$$H_1 + H_3 > 0.$$

12. From (13) and (31),

$$\mathrm{MSE}_{\min}(\hat{F}_9(y)) < \mathrm{MSE}_{\min}(\hat{F}_5(y))\mathrm{if}$$

$$\frac{F^2(y)\theta^2 V_{0200}\{\theta^2 V_{0200} + 16V_{2000}(1-R_{12}^2)\}}{64\{1+V_{2000}(1-R_{12}^2)\}} + H_3 > 0.$$

13. From (15) and (31),

$$\mathrm{MSE}_{\min}(\hat{F}_9(y)) < \mathrm{MSE}_{\min}(\hat{F}_6(y))\mathrm{if}$$

$$\frac{F^2(y)}{V_{2000}}\left(\frac{\theta V_{0200}}{2} - V_{1100}\right)^2 + H_1 + H_3 > 0.$$

14. From (19) and (31),

$$\mathrm{MSE}_{\min}(\hat{F}_9(y)) < \mathrm{MSE}_{\min}(\hat{F}_7(y))\mathrm{if}$$

$$H_3 > 0.$$

The proposed families of estimators are always more precise than the adapted estimators as the above conditions (i)–(xiv) are always true.

6 Empirical study in simple random sampling

In this section, we conduct a numerical study to investigate the performances of the adapted and CDF estimators. For this purpose, five populations are considered. The summary statistics of these populations are reported in Table 2. The percentage relative efficiency (PRE) of an estimator $\hat{F}_i(y)$ with respect to $\hat{F}_1(y)$ is

$$\mathrm{PRE}\big(\hat{F}_i(y), \hat{F}_1(y)\big) = \frac{\mathrm{Var}(\hat{F}_1(y))}{\mathrm{MSE}_{\min}(\hat{F}_i(y))} \times 100,$$

where i = 2,3,. . .,9.

The PREs of distribution function estimators, computed from five populations, are given in Tables 7–11.

Population I (Source: Singh, [6])
Y: Duration of sleep (in minutes) and
X: Age of old persons.
Population II (Source: Gujarati, [34])
Y: The eggs produced in 1990 (millions) and
X: The price per dozen (cents) in 1990.
Population III (Source: Murthy, [1])
Y: The output of the factory and
X: The number of workers.
Population IV (Source: Sarndal, [35])
Y: Population in 1983 (in million) and
X: Population in 1980 (in million).
Population V (Source: Koyuncu and Kadilar, [36])
Y: Number of teachers and
X: number of students.

Table 2. Summary statistics for Populations I to V.

Population	I	II	III	IV	V
N	30	50	80	120	923
n	5	5	10	20	180
λ	0.16667	0.18	0.0875	0.04167	0.00447
$\bar{X}$	67.267	78.29	285.125	33.535	11440.5
C_3	0.13725	0.27229	0.94846	3.32354	1.86453
$\bar{R}$	15.5	25.5	40.5	60.5	462
C_4	0.56695	0.57166	0.57377	0.57493	0.57703
$m_{2(y)}$	387	831	5105	9.25	171
$m_{2(x)}$	66.5	75.35	148	8.6	4123
$R^2_{1.23}$	0.5751	0.06242	0.90458	0.93453	0.72087
$R^2_{1.24}$	0.5783	0.0403	0.907	0.93725	0.75269
$F(y)$	0.5	0.5	0.5	0.5	0.50163
C_1	1.01709	1.01015	1.00631	1.00419	0.99729
$F(x)$	0.5	0.5	0.5	0.5	0.50054
C_2	1.01709	1.01015	1.00631	1.00419	0.99946
R_{12}	−0.73333	−0.12000	0.95	0.96667	0.84616
R_{13}	0.71975	0.22925	−0.71920	- 0.26749	-0.44165
R_{23}	−0.83720	−0.78936	−0.72395	−0.26760	−0.44809
R_{14}	0.73622	0.18435	−0.85636	−0.86368	−0.82860
R_{24}	−0.86727	−0.86630	−0.86610	−0.86609	−0.86603
β_2	1	1	1	1	1

https://doi.org/10.1371/journal.pone.0239098.t002

From the numerical results, presented in Tables 3–7, it is observed that the PREs of all families of estimators change with the choice of a or b. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of PRE. It can be seen that, for data sets I, III, IV and V, the second proposed family of the estimators perform better than the first proposed family of estimators, and for data set-II, it is also observed that the first proposed family of the estimators perform better than the second proposed family of estimators.

Table 3. PREs of distribution function estimators using Population I.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	78.81	$\hat{F}_7^{(1)}(y)$	234.70	$\hat{F}_8^{(1)}(y)$	253.84	$\hat{F}_9^{(1)}(y)$	255.59
$\hat{F}_2(y)$	28.84	$\hat{F}_6^{(2)}(y)$	78.60	$\hat{F}_7^{(2)}(y)$	234.72	$\hat{F}_8^{(2)}(y)$	253.87	$\hat{F}_9^{(2)}(y)$	255.62
$\hat{F}_3(y)$	187.50	$\hat{F}_6^{(3)}(y)$	78.81	$\hat{F}_7^{(3)}(y)$	234.70	$\hat{F}_8^{(3)}(y)$	253.84	$\hat{F}_9^{(3)}(y)$	255.59
$\hat{F}_4(y)$	216.34	$\hat{F}_6^{(4)}(y)$	78.39	$\hat{F}_7^{(4)}(y)$	234.75	$\hat{F}_8^{(4)}(y)$	253.90	$\hat{F}_9^{(4)}(y)$	255.65
$\hat{F}_5(y)$	233.58	$\hat{F}_6^{(5)}(y)$	173.45	$\hat{F}_7^{(5)}(y)$	343.92	$\hat{F}_8^{(5)}(y)$	377.99	$\hat{F}_9^{(5)}(y)$	381.16
		$\hat{F}_6^{(6)}(y)$	161.15	$\hat{F}_7^{(6)}(y)$	372.72	$\hat{F}_8^{(6)}(y)$	411.96	$\hat{F}_9^{(6)}(y)$	415.63
		$\hat{F}_6^{(7)}(y)$	150.14	$\hat{F}_7^{(7)}(y)$	236.97	$\hat{F}_8^{(7)}(y)$	256.31	$\hat{F}_9^{(7)}(y)$	258.08
		$\hat{F}_6^{(8)}(y)$	173.45	$\hat{F}_7^{(8)}(y)$	343.92	$\hat{F}_8^{(8)}(y)$	377.99	$\hat{F}_9^{(8)}(y)$	381.16
		$\hat{F}_6^{(9)}(y)$	151.69	$\hat{F}_7^{(9)}(y)$	237.16	$\hat{F}_8^{(9)}(y)$	256.52	$\hat{F}_9^{(9)}(y)$	258.29
		$\hat{F}_6^{(10)}(y)$	97.66	$\hat{F}_7^{(10)}(y)$	233.59	$\hat{F}_8^{(10)}(y)$	252.65	$\hat{F}_9^{(10)}(y)$	254.39

https://doi.org/10.1371/journal.pone.0239098.t003

Table 4. PREs of distribution function estimators using Population II.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	93.70	$\hat{F}_7^{(1)}(y)$	120.43	$\hat{F}_8^{(1)}(y)$	125.66	$\hat{F}_9^{(1)}(y)$	123.19
$\hat{F}_2(y)$	44.642	$\hat{F}_6^{(2)}(y)$	93.65	$\hat{F}_7^{(2)}(y)$	120.44	$\hat{F}_8^{(2)}(y)$	125.67	$\hat{F}_9^{(2)}(y)$	123.19
$\hat{F}_3(y)$	56.810	$\hat{F}_6^{(3)}(y)$	93.70	$\hat{F}_7^{(3)}(y)$	120.43	$\hat{F}_8^{(3)}(y)$	125.66	$\hat{F}_9^{(3)}(y)$	123.19
$\hat{F}_4(y)$	101.46	$\hat{F}_6^{(4)}(y)$	93.59	$\hat{F}_7^{(4)}(y)$	120.45	$\hat{F}_8^{(4)}(y)$	125.67	$\hat{F}_9^{(4)}(y)$	123.20
$\hat{F}_5(y)$	119.82	$\hat{F}_6^{(5)}(y)$	62.86	$\hat{F}_7^{(5)}(y)$	131.42	$\hat{F}_8^{(5)}(y)$	137.18	$\hat{F}_9^{(5)}(y)$	134.46
		$\hat{F}_6^{(6)}(y)$	62.99	$\hat{F}_7^{(6)}(y)$	131.33	$\hat{F}_8^{(6)}(y)$	137.09	$\hat{F}_9^{(6)}(y)$	134.37
		$\hat{F}_6^{(7)}(y)$	100.66	$\hat{F}_7^{(7)}(y)$	119.85	$\hat{F}_8^{(7)}(y)$	125.04	$\hat{F}_9^{(7)}(y)$	122.58
		$\hat{F}_6^{(8)}(y)$	62.86	$\hat{F}_7^{(8)}(y)$	131.42	$\hat{F}_8^{(8)}(y)$	137.18	$\hat{F}_9^{(8)}(y)$	134.46
		$\hat{F}_6^{(9)}(y)$	100.66	$\hat{F}_7^{(9)}(y)$	119.85	$\hat{F}_8^{(9)}(y)$	125.04	$\hat{F}_9^{(9)}(y)$	122.58
		$\hat{F}_6^{(10)}(y)$	99.75	$\hat{F}_7^{(10)}(y)$	119.83	$\hat{F}_8^{(10)}(y)$	125.02	$\hat{F}_9^{(10)}(y)$	122.56

https://doi.org/10.1371/journal.pone.0239098.t004

Here we take five data sets for numerical illustration. We selected different sample sizes from these populations and, then, we used simple random sampling. The MSEs (minimum) of the proposed families of estimators are pointed out in Eqs 25 and 31. Finally, the adapted estimators and proposed families of estimators were compared with each other with respect to their PRE values. These results are set out in Tables 3–7. In Table 2, we see the summary statistics about the populations. We can also note from the numerical results presented in Tables 3–7 that the PREs of all families of estimators change with the choice of a or b. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of the PRE. It can be seen that, for data sets I, III, IV and V, the second proposed family of the estimators perform better than the first proposed family of estimators, and, for data set II, it is also observed that the first proposed family of the estimators performs better than the second proposed family of estimators.

7 Notation in stratified random sampling

Consider a finite population $\Omega = \{1,2,...,N\}$ of N distinct units, which is divided into L homogeneous strata, where the size of hth stratum is N_h, for $h = 1,2,...,L$, such that $\sum_{h=1}^{L} N_h = N$.

Table 5. PREs of distribution function estimators using Population III.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	140.40	$\hat{F}_7^{(1)}(y)$	1037.21	$\hat{F}_8^{(1)}(y)$	1059.61	$\hat{F}_9^{(1)}(y)$	1087.07
$\hat{F}_2(y)$	1000	$\hat{F}_6^{(2)}(y)$	140.62	$\hat{F}_7^{(2)}(y)$	1037.23	$\hat{F}_8^{(2)}(y)$	1059.64	$\hat{F}_9^{(2)}(y)$	1087.09
$\hat{F}_3(y)$	25.64	$\hat{F}_6^{(3)}(y)$	140.40	$\hat{F}_7^{(3)}(y)$	1037.21	$\hat{F}_8^{(3)}(y)$	1059.61	$\hat{F}_9^{(3)}(y)$	1087.07
$\hat{F}_4(y)$	1025.64	$\hat{F}_6^{(4)}(y)$	140.84	$\hat{F}_7^{(4)}(y)$	1037.26	$\hat{F}_8^{(4)}(y)$	1059.66	$\hat{F}_9^{(4)}(y)$	1087.12
$\hat{F}_5(y)$	1034.50	$\hat{F}_6^{(5)}(y)$	142.42	$\hat{F}_7^{(5)}(y)$	1037.44	$\hat{F}_8^{(5)}(y)$	1059.85	$\hat{F}_9^{(5)}(y)$	1087.31
		$\hat{F}_6^{(6)}(y)$	142.64	$\hat{F}_7^{(6)}(y)$	1037.46	$\hat{F}_8^{(6)}(y)$	1059.88	$\hat{F}_9^{(6)}(y)$	1087.34
		$\hat{F}_6^{(7)}(y)$	138.68	$\hat{F}_7^{(7)}(y)$	1037.02	$\hat{F}_8^{(7)}(y)$	1059.42	$\hat{F}_9^{(7)}(y)$	1086.87
		$\hat{F}_6^{(8)}(y)$	142.42	$\hat{F}_7^{(8)}(y)$	1037.44	$\hat{F}_8^{(8)}(y)$	1059.85	$\hat{F}_9^{(8)}(y)$	1087.31
		$\hat{F}_6^{(9)}(y)$	138.89	$\hat{F}_7^{(9)}(y)$	1037.04	$\hat{F}_8^{(9)}(y)$	1059.4	$\hat{F}_9^{(9)}(y)$	1086.89
		$\hat{F}_6^{(10)}(y)$	101.18	$\hat{F}_7^{(10)}(y)$	1034.50	$\hat{F}_8^{(10)}(y)$	1056.84	$\hat{F}_9^{(10)}(y)$	1084.22

https://doi.org/10.1371/journal.pone.0239098.t005

Table 6. PREs of distribution function estimators using Population IV.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	141.58	$\hat{F}_7^{(1)}(y)$	1531.59	$\hat{F}_8^{(1)}(y)$	1533.54	$\hat{F}_9^{(1)}(y)$	1599.81
$\hat{F}_2(y)$	1500.00	$\hat{F}_6^{(2)}(y)$	141.73	$\hat{F}_7^{(2)}(y)$	1531.60	$\hat{F}_8^{(2)}(y)$	1533.55	$\hat{F}_9^{(2)}(y)$	1599.82
$\hat{F}_3(y)$	25.42	$\hat{F}_6^{(3)}(y)$	141.58	$\hat{F}_7^{(3)}(y)$	1531.59	$\hat{F}_8^{(3)}(y)$	1533.54	$\hat{F}_9^{(3)}(y)$	1599.81
$\hat{F}_4(y)$	1525.42	$\hat{F}_6^{(4)}(y)$	141.88	$\hat{F}_7^{(4)}(y)$	1531.61	$\hat{F}_8^{(4)}(y)$	1533.56	$\hat{F}_9^{(4)}(y)$	1599.84
$\hat{F}_5(y)$	1529.62	$\hat{F}_6^{(5)}(y)$	142.95	$\hat{F}_7^{(5)}(y)$	1531.70	$\hat{F}_8^{(5)}(y)$	1533.65	$\hat{F}_9^{(5)}(y)$	1599.93
		$\hat{F}_6^{(6)}(y)$	143.11	$\hat{F}_7^{(6)}(y)$	1531.71	$\hat{F}_8^{(6)}(y)$	1533.66	$\hat{F}_9^{(6)}(y)$	1599.94
		$\hat{F}_6^{(7)}(y)$	140.39	$\hat{F}_7^{(7)}(y)$	1531.49	$\hat{F}_8^{(7)}(y)$	1533.44	$\hat{F}_9^{(7)}(y)$	1599.71
		$\hat{F}_6^{(8)}(y)$	142.95	$\hat{F}_7^{(8)}(y)$	1531.70	$\hat{F}_8^{(8)}(y)$	1533.65	$\hat{F}_9^{(8)}(y)$	1599.93
		$\hat{F}_6^{(9)}(y)$	140.53	$\hat{F}_7^{(9)}(y)$	1531.50	$\hat{F}_8^{(9)}(y)$	1533.45	$\hat{F}_9^{(9)}(y)$	1599.72
		$\hat{F}_6^{(10)}(y)$	100.80	$\hat{F}_7^{(10)}(y)$	1529.62	$\hat{F}_8^{(10)}(y)$	1531.57	$\hat{F}_9^{(10)}(y)$	1597.75

https://doi.org/10.1371/journal.pone.0239098.t006

Let Y and X be the study and auxiliary variables which take values y_h and x_h, respectively, where $i = 1,2,\ldots,N_h$ and $h = 1,2,\ldots,L$, for estimating finite population distribution function, assume a sample of size n_h is drawn from the hth stratum using simple random sampling without replacement, such that $\sum_{h=1}^{L} n_h = n$, where n is the sample size.

Let $F_{st}(y) = F(y) = \sum_{h=1}^{L} W_h F_h(y)$ and $F_{st}(x) = F(x) = \sum_{h=1}^{L} W_h F_h(x)$ $(\hat{F}_{st}(y) = \hat{F}(y) = \sum_{h=1}^{L} W_h \hat{F}_h(y)$ and $\hat{F}_{st}(x) = \hat{F}(x) = \sum_{h=1}^{L} W_h \hat{F}_h(x))$ be the population (sample) distribution functions of Y and X under stratified random sampling, respectively, where $W_h = N_h/N$, $F_h(y) = \sum_{i=1}^{N_h} I(Y_{ih} \leq y)/N_h$, $F_h(x) = \sum_{i=1}^{N_h} I(X_{ih} \leq y)/N_h$, $\hat{F}_h(y) = \sum_{i=1}^{n_h} I(Y_{ih} \leq y)/n_h$,

$\hat{F}_h(x) = \sum_{i=1}^{n_h} I(X_{ih} \leq y)/n_h$. Let $\bar{X}_{st} = \bar{X} = \sum_{h=1}^{L} W_h \bar{X}_h$ and $\bar{Z}_{st} = \bar{Z} = \sum_{h=1}^{L} W_h \bar{Z}_h$ $(\hat{\bar{X}}_{st} = \hat{\bar{X}} = \sum_{h=1}^{L} W_h \hat{\bar{X}}_h$ and $\hat{\bar{Z}}_h = \hat{\bar{Z}} = \sum_{i=1}^{n_h} Z_{ih}/n_h)$ be the population (sample) means of X and Z under stratified random sampling, respectively, where $\bar{X}_h = \sum_{i=1}^{N_h} X_{ih}/N_h$, $\bar{Z}_h = \sum_{i=1}^{N_h} Z_{ih}/N_h$, $\hat{\bar{X}}_h = \sum_{i=1}^{n_h} X_{ih}/n_h$, $\hat{\bar{Z}}_h = \sum_{i=1}^{n_h} Z_{ih}/n_h$.

Table 7. PREs of distribution function estimators using Population V.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	134.23	$\hat{F}_7^{(1)}(y)$	352.57	$\hat{F}_8^{(1)}(y)$	358.74	$\hat{F}_9^{(1)}(y)$	404.86
$\hat{F}_2(y)$	324.29	$\hat{F}_6^{(2)}(y)$	134.21	$\hat{F}_7^{(2)}(y)$	352.57	$\hat{F}_8^{(2)}(y)$	358.74	$\hat{F}_9^{(2)}(y)$	404.86
$\hat{F}_3(y)$	27.02	$\hat{F}_6^{(3)}(y)$	134.23	$\hat{F}_7^{(3)}(y)$	352.57	$\hat{F}_8^{(3)}(y)$	358.74	$\hat{F}_9^{(3)}(y)$	404.86
$\hat{F}_4(y)$	352.08	$\hat{F}_6^{(4)}(y)$	134.20	$\hat{F}_7^{(4)}(y)$	352.57	$\hat{F}_8^{(4)}(y)$	358.74	$\hat{F}_9^{(4)}(y)$	404.86
$\hat{F}_5(y)$	352.53	$\hat{F}_6^{(5)}(y)$	38.98	$\hat{F}_7^{(5)}(y)$	352.58	$\hat{F}_8^{(5)}(y)$	358.75	$\hat{F}_9^{(5)}(y)$	404.87
		$\hat{F}_6^{(6)}(y)$	138.96	$\hat{F}_7^{(6)}(y)$	352.58	$\hat{F}_8^{(6)}(y)$	358.75	$\hat{F}_9^{(6)}(y)$	404.87
		$\hat{F}_6^{(7)}(y)$	129.89	$\hat{F}_7^{(7)}(y)$	352.56	$\hat{F}_8^{(7)}(y)$	358.73	$\hat{F}_9^{(7)}(y)$	404.85
		$\hat{F}_6^{(8)}(y)$	138.98	$\hat{F}_7^{(8)}(y)$	352.58	$\hat{F}_8^{(8)}(y)$	358.75	$\hat{F}_9^{(8)}(y)$	404.87
		$\hat{F}_6^{(9)}(y)$	129.88	$\hat{F}_7^{(9)}(y)$	352.56	$\hat{F}_8^{(9)}(y)$	358.73	$\hat{F}_9^{(9)}(y)$	404.85
		$\hat{F}_6^{(10)}(y)$	100.09	$\hat{F}_7^{(10)}(y)$	352.53	$\hat{F}_8^{(10)}(y)$	358.69	$\hat{F}_9^{(10)}(y)$	404.80

https://doi.org/10.1371/journal.pone.0239098.t007

Table 8. Some members of the proposed and adapted estimators.

a	b	$\hat{F}_{S_{st}}(y)$	$\hat{F}_{G,K_{st}}(y)$	$\hat{F}_{Prop1_{st}}(y)$	$\hat{F}_9(y)$
1	C_2	$\hat{F}^{(1)}_{S_{st}}(y)$	$\hat{F}^{(1)}_{G,K_{st}}(y)$	$\hat{F}^{(1)}_{Prop1_{st}}(y)$	$\hat{F}^{(1)}_{Prop2_{st}}(y)$
1	$\beta_{2(st)}$	$\hat{F}^{(2)}_{S_{st}}(y)$	$\hat{F}^{(2)}_{G,K_{st}}(y)$	$\hat{F}^{(2)}_{Prop1_{st}}(y)$	$\hat{F}^{(2)}_{Prop2_{st}}(y)$
$\beta_{2(st)}$	C_2	$\hat{F}^{(3)}_{S_{st}}(y)$	$\hat{F}^{(3)}_{G,K_{st}}(y)$	$\hat{F}^{(3)}_{Prop1_{st}}(ty)$	$\hat{F}^{(3)}_{Prop2_{st}}(y)$
C_2	$\beta_{2(st)}$	$\hat{F}^{(4)}_{S_{st}}(y)$	$\hat{F}^{(4)}_{G,K_{st}}(y)$	$\hat{F}^{(4)}_{Prop1_{st}}(y)$	$\hat{F}^{(4)}_{Prop2_{st}}(y)$
1	R_{12}	$\hat{F}^{(5)}_{S_{st}}(y)$	$\hat{F}^{(5)}_{G,K_{st}}(y)$	$\hat{F}^{(5)}_{Prop1_{st}}(y)$	$\hat{F}^{(5)}_{Prop2_{st}}(y)$
C_2	R_{12}	$\hat{F}^{(6)}_{S_{st}}(y)$	$\hat{F}^{(6)}_{G,K_{st}}(y)$	$\hat{F}^{(6)}_{Prop1_{st}}(y)$	$\hat{F}^{(6)}_{Prop2_{st}}(y)$
R_{12}	C_2	$\hat{F}^{(7)}_{S_{st}}(y)$	$\hat{F}^{(7)}_{G,K_{st}}(y)$	$\hat{F}^{(7)}_{Prop1_{st}}(y)$	$\hat{F}^{(7)}_{Prop2_{st}}(y)$
$\beta_{2(st)}$	R_{12}	$\hat{F}^{(8)}_{S_{st}}(y)$	$\hat{F}^{(8)}_{G,K_{st}}(y)$	$\hat{F}^{(8)}_{Prop1_{st}}(y)$	$\hat{F}^{(8)}_{Prop2_{st}}(y)$
R_{12}	$\beta_{2(st)}$	$\hat{F}^{(9)}_{S_{st}}(y)$	$\hat{F}^{(9)}_{G,K_{st}}(y)$	$\hat{F}^{(9)}_{Prop1_{st}}(y)$	$\hat{F}^{(9)}_{Prop2_{st}}(y)$
1	$NF(x)$	$\hat{F}^{(10)}_{S_{st}}(y)$	$\hat{F}^{(10)}_{G,K_{st}}(y)$	$\hat{F}^{(10)}_{Prop1_{st}}(y)$	$\hat{F}^{(10)}_{Prop2_{st}}(y)$

https://doi.org/10.1371/journal.pone.0239098.t008

In order to obtain the biases and MSEs of the adapted and proposed estimators of $F(y)$ under stratified random sampling, the following relative error terms are considered. Let

$$\zeta_1 = \frac{\hat{F}_{st}(y) - F(y)}{F(y)},\ \zeta_2 = \frac{\hat{F}_{st}(x) - F(x)}{F(x)},\ \zeta_3 = \frac{\hat{\bar{X}}_{st} - \bar{X}}{\bar{X}} \text{ and } \zeta_4 = \frac{\hat{\bar{Z}}_{st} - \bar{Z}}{\bar{Z}},$$

such that $E(e_i) = 0$ for $i = 1,2,3,4$, where $E(\cdot)$ is the mathematical expectation of $(\cdot)$. Let

$$V_{rstu} = E[\zeta_1^r\ \zeta_2^s\ \zeta_3^t\ \zeta_4^u],$$

Table 9. Summary statistics for Population I.

h	N_h	n_h	W_h	λ_h	$F(y_h)$	$F(x_h)$	$\bar{X}_h$	$\bar{Z}_h$
1	127	31	0.1375	0.0244	0.3543	0.3779	20805	64
2	117	21	0.1267	0.0390	0.4188	0.4872	9212	59
3	103	29	0.1115	0.0248	0.4272	0.4660	14309	52
4	170	38	0.1841	0.0204	0.5765	0.6118	9479	86
5	205	22	0.2221	0.0406	0.6146	0.6537	5570	103
6	201	39	0.2177	0.0207	0.5025	0.3532	12998	101
S_{1h}	S_{2h}	S_{3h}	S_{4h}	R_{12h}	R_{13h}	R_{23h}	R_{14h}	R_{24h}
.4802	0.4868	30487	36.806	0.9164	−0.4602	−0.4832	−0.8155	−0.8399
.4955	0.5019	15181	33.919	0.8709	−0.4147	−0.4600	−0.8437	−0.8658
.4970	0.5013	27550	29.877	0.9244	−0.3928	−0.4198	−0.8489	−0.8640
.4956	0.4888	18219	49.219	0.8805	−0.5074	−0.5396	−0.8417	−0.8441
.4879	0.4769	8498	59.322	0.8772	−0.5579	−0.5909	−0.8305	−0.8241
.5012	0.4792	23094	58.168	0.7145	−0.4334	−0.3554	−0.8125	−0.8279

https://doi.org/10.1371/journal.pone.0239098.t009

Table 10. Summary statistics for Population II.

h	N_h	n_h	W_h	λ_h	$F(y_h)$	$F(x_h)$	$\bar{X}_h$	$\bar{Z}_h$
1	127	31	0.1375	0.0244	0.3543	0.3700	498.276	64
2	117	21	0.1267	0.0391	0.4188	0.4700	318.333	59
3	103	29	0.1115	0.0248	0.4272	0.4272	431.359	52
4	170	38	0.1841	0.0204	0.5765	0.5882	311.324	86
5	205	22	0.2221	0.0406	0.6146	0.6146	227.195	103
6	201	39	0.2177	0.0207	0.5025	0.4527	312.706	101
S_{1h}	S_{2h}	S_{3h}	S_{4h}	R_{12h}	R_{13h}	R_{23h}	R_{14h}	R_{24h}
.4802	0.4847	555.58	36.805	0.8983	−0.5398	−0.5580	−0.8114	−0.8363
.4955	0.5013	365.45	33.918	0.8666	−0.5205	−0.5589	−0.8382	−0.8645
.4970	0.4970	613.95	29.877	0.9603	−0.4803	−0.4823	−0.8519	−0.8568
.4956	0.4936	458.02	49.217	0.9277	−0.5818	−0.5934	−0.8490	−0.8525
.4879	0.4879	260.85	59.321	0.8764	−0.6395	−0.6457	−0.8285	−0.8429
.5012	0.4990	397.04	58.167	0.8450	−0.5063	−0.4873	−0.8343	−0.8622

https://doi.org/10.1371/journal.pone.0239098.t010

Where

$$E(\zeta_1^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{1h}^2 = \psi_{2000}, E(\zeta_2^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{2h}^2 = \psi_{0200},$$

$$E(\zeta_3^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{3h}^2 = \psi_{0020}, E(\zeta_4^2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 C_{4h}^2 = \psi_{0002},$$

$$E(\zeta_1\zeta_2) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 R_{12h} C_{1h} C_{2h} = \psi_{1100}, E(\zeta_1\zeta_3) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 R_{13h} C_{1h} C_{3h} = \psi_{1010},$$

$$E(\zeta_2\zeta_3) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 R_{23h} C_{2h} C_{3h} = \psi_{0110}, E(\zeta_1\zeta_4) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 R_{14h} C_{1h} C_{4h} = \psi_{1001},$$

$$E(\zeta_2\zeta_4) = \sum_{h=1}^{L} W_h^2 \lambda_h^2 R_{24h} C_{2h} C_{4h} = \psi_{0101}.$$

Table 11. Summary statistics for Population III.

h	N_h	n_h	W_h	λ_h	$F(y_h)$	$F(x_h)$	$\bar{X}_h$	$\bar{Z}_h$
1	106	9	0.1241	0.1017	0.5849	0.5472	24376	54
2	106	17	0.1241	0.0494	0.5189	0.5660	27422	54
3	94	38	0.1100	0.0157	0.3298	0.3404	72410	48
4	171	67	0.2002	0.0090	0.3684	0.3801	74365	87
5	204	7	0.2389	0.1379	0.4657	0.4657	26442	103
6	173	2	0.2026	0.4942	0.7052	0.7225	9844	87
S_{1h}	S_{2h}	S_{3h}	S_{4h}	R_{12h}	R_{13h}	R_{23h}	R_{14h}	R_{24h}
.4950	0.5001	49189	30.743	0.7722	−0.4470	−0.4523	−0.7665	−0.8622
.5020	0.4979	5746	30.743	0.8330	−0.4370	−0.4816	−0.8285	−0.8585
.4727	0.4764	160757	27.279	0.7854	−0.2957	−0.3087	−0.7509	−0.8208
.4838	0.4868	285603	49.507	0.7755	−0.1848	−0.1936	−0.7535	−0.8408
.5000	0.4965	45403	59.033	0.6750	−0.3929	−0.4129	−0.7218	−0.8578
.4573	0.4490	18794	50.084	0.7319	−0.5598	−0.6102	−0.7290	−0.7755

https://doi.org/10.1371/journal.pone.0239098.t011

Table 12. Summary statistics for Population IV.

h	N_h	n_h	W_h	λ_h	$F(y_h)$	$F(x_h)$	$\bar{X}_h$	$\bar{Z}_h$
1	106	9	0.1241	0.1017	0.5849	0.5189	24712	54
2	106	17	0.1241	0.0494	0.5189	0.5660	26840	54
3	94	38	0.1100	0.0157	0.3298	0.3404	72722	48
4	171	67	0.2002	0.0090	0.3684	0.3743	73191	87
5	204	7	0.2389	0.1379	0.4657	0.4363	26834	103
6	173	2	0.2026	0.4942	0.7052	0.7341	9903	87
S_{1h}	S_{2h}	S_{3h}	S_{4h}	R_{12h}	R_{13h}	R_{23h}	R_{14h}	R_{24h}
.4950	0.5020	49135	30.743	0.7598	−0.4439	−0.4360	−0.7474	−0.8655
.5020	0.4979	53979	30.743	0.8330	−0.5011	−0.4816	−0.8303	−0.8585
.4727	0.4764	161110	27.279	0.7376	−0.2957	−0.3093	−0.7460	−0.8208
.4838	0.4854	26249	49.507	0.7871	−0.1974	−0.2050	−0.7516	−0.8382
.5000	0.4971	45174	59.033	0.6690	−0.4011	−0.4266	−0.7164	−0.8590
.4573	0.4430	18977	50.084	0.7299	−0.5399	−0.6241	−0.7002	−0.7653

https://doi.org/10.1371/journal.pone.0239098.t012

8 Adapted estimators in stratified random sampling

In this section, some estimators of finite population mean are adapted for estimating the finite CDF under stratified random sampling. The biases and MSEs of these adapted estimators are derived under the first order of approximation.

1. The traditional unbiased estimator of $F(y)$ is

$$\hat{F}_{SRS_{st}}(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq y). \tag{32}$$

The variance of $\hat{F}_{SRS_{st}}(y)$ is

$$\mathrm{Var}(\hat{F}_{SRS_{st}}(y)) = F^2(y)\psi_{2000}. \tag{33}$$

Table 13. PREs of distribution function estimators using Population I.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	133.03	$\hat{F}_7^{(1)}(y)$	364.28	$\hat{F}_8^{(1)}(y)$	370.30	$\hat{F}_9^{(1)}(y)$	428.83
$\hat{F}_2(y)$	341.19	$\hat{F}_6^{(2)}(y)$	128.23	$\hat{F}_7^{(2)}(y)$	364.27	$\hat{F}_8^{(2)}(y)$	370.29	$\hat{F}_9^{(2)}(y)$	428.82
$\hat{F}_3(y)$	27.38	$\hat{F}_6^{(3)}(y)$	139.35	$\hat{F}_7^{(3)}(y)$	364.29	$\hat{F}_8^{(3)}(y)$	370.32	$\hat{F}_9^{(3)}(y)$	428.85
$\hat{F}_4(y)$	363.74	$\hat{F}_6^{(4)}(y)$	128.99	$\hat{F}_7^{(4)}(y)$	364.27	$\hat{F}_8^{(4)}(y)$	370.30	$\hat{F}_9^{(4)}(y)$	428.82
$\hat{F}_5(y)$	364.23	$\hat{F}_6^{(5)}(y)$	138.40	$\hat{F}_7^{(5)}(y)$	364.29	$\hat{F}_8^{(5)}(y)$	370.32	$\hat{F}_9^{(5)}(y)$	428.85
		$\hat{F}_6^{(6)}(y)$	139.35	$\hat{F}_7^{(6)}(y)$	364.29	$\hat{F}_8^{(6)}(y)$	370.32	$\hat{F}_9^{(6)}(y)$	428.85
		$\hat{F}_6^{(7)}(y)$	128.99	$\hat{F}_7^{(7)}(y)$	364.27	$\hat{F}_8^{(7)}(y)$	370.29	$\hat{F}_9^{(7)}(y)$	428.82
		$\hat{F}_6^{(8)}(y)$	145.36	$\hat{F}_7^{(8)}(y)$	364.31	$\hat{F}_8^{(8)}(y)$	370.34	$\hat{F}_9^{(8)}(y)$	428.87
		$\hat{F}_6^{(9)}(y)$	124.67	$\hat{F}_7^{(9)}(y)$	364.26	$\hat{F}_8^{(9)}(y)$	370.29	$\hat{F}_9^{(9)}(y)$	428.81
		$\hat{F}_6^{(10)}(y)$	100.09	$\hat{F}_7^{(10)}(y)$	364.23	$\hat{F}_8^{(10)}(y)$	370.26	$\hat{F}_9^{(10)}(y)$	428.78

https://doi.org/10.1371/journal.pone.0239098.t013

Table 14. PREs of distribution function estimators using Population II.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	136.01	$\hat{F}_7^{(1)}(y)$	454.79	$\hat{F}_8^{(1)}(y)$	464.37	$\hat{F}_9^{(1)}(y)$	502.91
$\hat{F}_2(y)$	427.32	$\hat{F}_6^{(2)}(y)$	132.94	$\hat{F}_7^{(2)}(y)$	454.78	$\hat{F}_8^{(2)}(y)$	464.36	$\hat{F}_9^{(2)}(y)$	502.90
$\hat{F}_3(y)$	26.53	$\hat{F}_6^{(3)}(y)$	139.80	$\hat{F}_7^{(3)}(y)$	454.80	$\hat{F}_8^{(3)}(y)$	464.38	$\hat{F}_9^{(3)}(y)$	502.92
$\hat{F}_4(y)$	454.24	$\hat{F}_6^{(4)}(y)$	133.38	$\hat{F}_7^{(4)}(y)$	454.78	$\hat{F}_8^{(4)}(y)$	464.36	$\hat{F}_9^{(4)}(y)$	502.90
$\hat{F}_5(y)$	454.73	$\hat{F}_6^{(5)}(y)$	140.06	$\hat{F}_7^{(5)}(y)$	454.80	$\hat{F}_8^{(5)}(y)$	464.38	$\hat{F}_9^{(5)}(y)$	502.90
		$\hat{F}_6^{(6)}(y)$	140.57	$\hat{F}_7^{(6)}(y)$	454.81	$\hat{F}_8^{(6)}(y)$	464.38	$\hat{F}_9^{(6)}(y)$	502.92
		$\hat{F}_6^{(7)}(y)$	132.71	$\hat{F}_7^{(7)}(y)$	454.78	$\hat{F}_8^{(7)}(y)$	464.36	$\hat{F}_9^{(7)}(y)$	502.92
		$\hat{F}_6^{(8)}(y)$	144.15	$\hat{F}_7^{(8)}(y)$	454.82	$\hat{F}_8^{(8)}(y)$	464.39	$\hat{F}_9^{(8)}(y)$	502.93
		$\hat{F}_6^{(9)}(y)$	129.86	$\hat{F}_7^{(9)}(y)$	454.77	$\hat{F}_8^{(9)}(y)$	464.35	$\hat{F}_9^{(9)}(y)$	502.89
		$\hat{F}_6^{(10)}(y)$	100.10	$\hat{F}_7^{(10)}(y)$	454.73	$\hat{F}_8^{(10)}(y)$	464.30	$\hat{F}_9^{(10)}(y)$	502.84

https://doi.org/10.1371/journal.pone.0239098.t014

2. Cochran [31] adapted ratio estimator of $F(y)$ is

$$\hat{F}_{R_{st}}(y) = \hat{F}_{st}(y)\left(\frac{F(x)}{\hat{F}_{st}(x)}\right). \tag{34}$$

The bias and MSE of $\hat{F}_{R_{st}}(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_{R_{st}}(y)) \cong F(y)(\psi_{0200} - \psi_{1100}),$$

$$\text{MSE}(\hat{F}_{R_{st}}(y)) \cong F^2(y)(\psi_{2000} + \psi_{0200} - 2\psi_{1100}). \tag{35}$$

Table 15. PREs of distribution function estimators using Population III.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	126.05	$\hat{F}_7^{(1)}(y)$	211.54	$\hat{F}_8^{(1)}(y)$	214.04	$\hat{F}_9^{(1)}(y)$	241.14
$\hat{F}_2(y)$	181.83	$\hat{F}_6^{(2)}(y)$	121.46	$\hat{F}_7^{(2)}(y)$	211.49	$\hat{F}_8^{(2)}(y)$	213.99	$\hat{F}_9^{(2)}(y)$	241.09
$\hat{F}_3(y)$	29.41	$\hat{F}_6^{(3)}(y)$	131.91	$\hat{F}_7^{(3)}(y)$	211.61	$\hat{F}_8^{(3)}(y)$	214.10	$\hat{F}_9^{(3)}(y)$	241.22
$\hat{F}_4(y)$	208.64	$\hat{F}_6^{(4)}(y)$	122.03	$\hat{F}_7^{(4)}(y)$	211.50	$\hat{F}_8^{(4)}(y)$	214.00	$\hat{F}_9^{(4)}(y)$	241.10
$\hat{F}_5(y)$	211.39	$\hat{F}_6^{(5)}(y)$	132.66	$\hat{F}_7^{(5)}(y)$	211.62	$\hat{F}_8^{(5)}(y)$	214.12	$\hat{F}_9^{(5)}(y)$	241.23
		$\hat{F}_6^{(6)}(y)$	133.40	$\hat{F}_7^{(6)}(y)$	211.63	$\hat{F}_8^{(6)}(y)$	214.13	$\hat{F}_9^{(6)}(y)$	241.24
		$\hat{F}_6^{(7)}(y)$	120.88	$\hat{F}_7^{(7)}(y)$	211.49	$\hat{F}_8^{(7)}(y)$	213.99	$\hat{F}_9^{(7)}(y)$	241.08
		$\hat{F}_6^{(8)}(y)$	139.14	$\hat{F}_7^{(8)}(y)$	211.71	$\hat{F}_8^{(8)}(y)$	214.21	$\hat{F}_9^{(8)}(y)$	241.34
		$\hat{F}_6^{(9)}(y)$	116.98	$\hat{F}_7^{(9)}(y)$	211.46	$\hat{F}_8^{(9)}(y)$	213.95	$\hat{F}_9^{(9)}(y)$	241.05
		$\hat{F}_6^{(10)}(y)$	100.08	$\hat{F}_7^{(10)}(y)$	211.39	$\hat{F}_8^{(10)}(y)$	213.88	$\hat{F}_9^{(10)}(y)$	240.97

https://doi.org/10.1371/journal.pone.0239098.t015

Table 16. PREs of distribution function estimators using Population IV.

Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value	Estimator	Value
$\hat{F}_1(y)$	100	$\hat{F}_6^{(1)}(y)$	125.53	$\hat{F}_7^{(1)}(y)$	208.38	$\hat{F}_8^{(1)}(y)$	210.61	$\hat{F}_9^{(1)}(y)$	231.39
$\hat{F}_2(y)$	179.42	$\hat{F}_6^{(2)}(y)$	120.78	$\hat{F}_7^{(2)}(y)$	208.34	$\hat{F}_8^{(2)}(y)$	210.56	$\hat{F}_9^{(2)}(y)$	231.34
$\hat{F}_3(y)$	29.65	$\hat{F}_6^{(3)}(y)$	131.68	$\hat{F}_7^{(3)}(y)$	208.45	$\hat{F}_8^{(3)}(y)$	210.68	$\hat{F}_9^{(3)}(y)$	231.47
$\hat{F}_4(y)$	205.48	$\hat{F}_6^{(4)}(y)$	121.39	$\hat{F}_7^{(4)}(y)$	208.34	$\hat{F}_8^{(4)}(y)$	210.56	$\hat{F}_9^{(4)}(y)$	231.35
$\hat{F}_5(y)$	208.23	$\hat{F}_6^{(5)}(y)$	132.28	$\hat{F}_7^{(5)}(y)$	208.46	$\hat{F}_8^{(5)}(y)$	210.69	$\hat{F}_9^{(5)}(y)$	231.48
		$\hat{F}_6^{(6)}(y)$	133.08	$\hat{F}_7^{(6)}(y)$	208.47	$\hat{F}_8^{(6)}(y)$	210.70	$\hat{F}_9^{(6)}(y)$	231.49
		$\hat{F}_6^{(7)}(y)$	120.31	$\hat{F}_7^{(7)}(y)$	208.33	$\hat{F}_8^{(7)}(y)$	210.55	$\hat{F}_9^{(7)}(y)$	231.34
		$\hat{F}_6^{(8)}(y)$	139.07	$\hat{F}_7^{(8)}(y)$	208.56	$\hat{F}_8^{(8)}(y)$	210.78	$\hat{F}_9^{(8)}(y)$	231.59
		$\hat{F}_6^{(9)}(y)$	116.30	$\hat{F}_7^{(9)}(y)$	208.30	$\hat{F}_8^{(9)}(y)$	210.52	$\hat{F}_9^{(9)}(y)$	231.30
		$\hat{F}_6^{(10)}(y)$	100.08	$\hat{F}_7^{(10)}(y)$	208.23	$\hat{F}_8^{(10)}(y)$	210.46	$\hat{F}_9^{(10)}(y)$	231.23

https://doi.org/10.1371/journal.pone.0239098.t016

3. Murthy [32] adapted product estimator of $F(y)$ is

$$\hat{F}_{P_{st}}(y) = \hat{F}_{st}(y)\left(\frac{\hat{F}_{st}(x)}{F(x)}\right). \tag{36}$$

The bias and MSE of $\hat{F}_{P_{st}}(y)$, to the first order of approximation, are

$$\text{Bias}(\hat{F}_{P_{st}}(y)) = F(y)\psi_{1100},$$

$$\text{MSE}(\hat{F}_{P_{st}}(y)) \cong F^2(y)(\psi_{2000} + \psi_{0200} + 2\psi_{1100}). \tag{37}$$

4. The adapted difference estimator of $F(y)$ is

$$\hat{F}_{Reg_{st}}(y) = \hat{F}_{st}(y) + m(F(x) - \hat{F}_{st}(x)), \tag{38}$$

where m is an unknown constant. Here, $\hat{F}_{Reg_{st}}(y)$ is an unbiased estimator of $F(y)$. The minimum variance of $\hat{F}_{Reg_{st}}(y)$ at the optimum value $m_{(\text{opt})} = (F(y)\psi_{1100})/(F(x)\psi_{0200})$ is

$$\text{Var}_{\min}\left(\hat{F}_{Reg_{st}}(y)\right) = \frac{F^2(y)(\psi_{2000}\psi_{0200} - \psi_{1100}^2)}{\psi_{0200}}. \tag{39}$$

Here, (39) may be written as

$$\text{Var}_{\min}(\hat{F}_{Reg_{st}}(y)) = F^2(y)\psi_{2000}(1 - R_{12}^2). \tag{40}$$

5. Rao [4] adapted difference-type estimator of $F(y)$ is

$$\hat{F}_{R,D_{st}}(y) = m_1\hat{F}_{st}(y) + m_2(F(x) - \hat{F}_{st}(x)), \tag{41}$$

where m_1 and m_2 are unknown constants. The bias and MSE of $\hat{F}_{R,D_{st}}(y)$, to the first order of

approximation, are

$$\text{Bias}(\hat{F}_5(y)) = F(y)(m_1 - 1),$$

$$\text{MSE}(\hat{F}_{R,D_{st}}(y)) \cong F^2(y) - 2m_1F^2(y) + m_1^2F^2(y) + m_1^2F^2(y)\psi_{2000}$$

$$-2m_1m_2F(y)F(x)\psi_{1100} + m_2^2F^2(x)\psi_{0200}. \quad (42)$$

The optimum values of m_1 and m_2, determined by minimizing (42), are

$$m_{1(\text{opt})} = \frac{\psi_{0200}}{(\psi_{0200}\psi_{2000} - \psi_{1100}^2 + \psi_{0200})},$$

$$m_{2(\text{opt})} = \frac{F(y)\psi_{1100}}{F(x)(\psi_{2000}\psi_{0200} - \psi_{1100}^2 + \psi_{0200})}.$$

The minimum MSE of $\hat{F}_{R,D_{st}}(y)$ at the optimum values of m_1 and m_2 is

$$\text{MSE}_{\min}\left(\hat{F}_{R,D_{st}}(y)\right) = \frac{F^2(y)(\psi_{2000}\psi_{0200} - \psi_{1100}^2)}{(\psi_{2000}\psi_{0200} - \psi_{1100}^2 + \psi_{0200})}. \quad (43)$$

Here, (43) may be written as

$$\text{MSE}_{\min}\left(\hat{F}_{R,D_{st}}(y)\right) = \frac{F^2(y)\psi_{2000}(1 - R_{12}^2)}{1 + \psi_{2000}(1 - R_{12}^2)}. \quad (44)$$

6. Singh et al. [33] adapted generalized ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_{S_{st}}(y) = \hat{F}_{st}(y)\exp\left(\frac{a_{st}(F(x) - \hat{F}_{st}(x))}{a_{st}(F(x) + \hat{F}_{st}(x)) + 2b_{st}}\right), \quad (45)$$

where a_{st} and b_{st} are known constants. The bias and MSE of $\hat{F}_{S_{st}}(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_{S_{st}}(y)\right) \cong F(y)\left(\frac{3}{8}\Theta^2\psi_{0200} - \frac{1}{2}\Theta\psi_{1100}\right),$$

$$\text{MSE}\left(\hat{F}_{S_{st}}(y)\right) \cong \frac{F^2(y)}{4}(4\psi_{2000} + \Theta^2\psi_{0200} - 4\Theta\psi_{1100}), \quad (46)$$

where $\Theta = a_{st}F(x)/(a_{st}F(x)+b_{st})$.

7. Grover and Kaur [11] adapted generalized class of ratio-type exponential estimator of $F(y)$ is

$$\hat{F}_{G,K_{st}}(y) = \left\{m_3\hat{F}_{st}(y) + m_4(F(x) - \hat{F}_{st}(x))\right\}\exp\left(\frac{a_{st}(F(x) - \hat{F}_{st}(x))}{a_{st}(F(x) + \hat{F}_{st}(x)) + 2b_{st}}\right), \quad (47)$$

where m_3 and m_4 are unknown constants. The bias and MSE of $\hat{F}_{G,K_{st}}(y)$, to the first order of

approximation, are

$$\text{Bias}\left(\hat{F}_{G,K_{st}}(y)\right) \cong F(y)(m_3 - 1) + \frac{3}{8}\Theta^2 m_3 F(y) + \frac{1}{2}\Theta m_4 F(x)\psi_{0200} - \frac{1}{2}\Theta F(y)\psi_{1100},$$

$$\begin{aligned}\text{MSE}(\hat{F}_{G,K_{st}}(y)) \cong\ & m_4^2 F^2(x)\psi_{0200} + m_3^2 F^2(y)\psi_{2000} + 2\Theta m_3 m_4 F(y)F(x)\psi_{0200} \\ & -2m_3 m_4 F(y)F(x)\psi_{1100} + F^2(y) - 2m_3 F^2(y) + \Theta m_3^2 F^2(y) \\ & +m_3 F^2(y)\psi_{1100} - \Theta m_4 F(y)F(x)\psi_{0200} - 2\Theta m_3^2 F^2(y)\psi_{1100} \\ & -\frac{3}{4}\Theta^2 m_3 F^2(y)\psi_{0200} + \Theta^2 m_3^2 F^2(y)\psi_{0200}. \end{aligned} \quad (48)$$

The optimum values of m_3 and m_4, determined by minimizing (48), are

$$m_{3(\text{opt})} = \frac{\psi_{0200}(\Theta^2\psi_{0200} - 8)}{8(-\psi_{2000}\psi_{0200} + \psi_{1100}^2 - \psi_{0200})},$$

$$m_{4(\text{opt})} = \frac{F(y)(\Theta^3\psi_{0200}^2 - \Theta^2\psi_{0200}\psi_{1100} + 4\Theta\psi_{2000}\psi_{0200} - 4\Theta\psi_{1100}^2 - 4\Theta\psi_{0200} + 8\psi_{1100})}{8F(x)(\psi_{2000}\psi_{0200} - \psi_{1100}^2 + \psi_{0200})}.$$

The simplified minimum MSE of $\hat{F}_{G,K_{st}}(y)$ at the optimum values of m_3 and m_4 is

$$\text{MSE}_{\min}\left(\hat{F}_{G,K_{st}}(y)\right) \cong \frac{F^2(y)}{64}\left(64 - 16\Theta^2\psi_{0200} - \frac{\psi_{0200}(-8 + \Theta^2\psi_{0200})^2}{\psi_{0200}(1 + \psi_{2000}) - V_{1100}^2}\right). \quad (49)$$

Here, (49) may be written as

$$\text{MSE}_{\min}\left(\hat{F}_{G,K_{st}}(y)\right) \cong \text{Var}_{\min}\left(\hat{F}_{Reg_{st}}(y)\right) - \frac{F^2(y)(\Theta^2\psi_{0200}^2 - 8\psi_{1100}^2 + 8\psi_{0200}\psi_{2000})^2}{64\psi_{0200}^2\{1 + \psi_{2000}(1 - R_{12}^2)\}}, \quad (50)$$

which shows that $\hat{F}_{G,K_{st}}(y)$ is more precise than $\hat{F}_{Reg_{st}}(y)$.

9 Proposed estimators in stratified random sampling

9.1 First proposed family of estimators

On the lines of $\hat{F}_{R,D_{st}}(y)$ and $\hat{F}_{S_{st}}(y)$, first proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_{Prop1_{st}}(y) = \left\{m_5\hat{F}_{st}(y) + m_6\left(\frac{F(x) - \hat{F}_{st}(x)}{F(x)}\right) + m_7\left(\frac{\bar{X} - \hat{\bar{X}}_{st}}{\bar{X}}\right)\right\}\exp\left(\frac{a_{st}(F(x) - \hat{F}_{st}(x))}{a_{st}(F(x) + \hat{F}_{st}(x)) + 2b_{st}}\right), \quad (51)$$

where m_5, m_6 and m_7 are unknown constants, $a_{st}(\neq 0)$ and b_{st} are either two real numbers or functions of known population parameters of $I(X \leq x)$, like R_{12}, $\beta_{2(st)} = \sum_{i=1}^{L} W_h\beta_{2h(x)}$ (coefficient of kurtosis), $C_2 = \sum_{i=1}^{L} W_h C_{2h}$, etc.

The estimator $\hat{F}_{Prop1_{st}}(y)$ can also be written as

$$\hat{F}_{Prop1_{st}}(y) = \{m_5 F(y)(1+\zeta_1) - m_6\zeta_2 - m_7\zeta_3\}\left(1 - \frac{1}{2}\Theta\zeta_2 + \frac{3}{8}\Theta^2\zeta_2^2 + \cdots\right). \quad (52)$$

Simplifying (52) and keeping terms only up to the second power of ζ_i's, we can write

$$\left(\hat{F}_{Prop1_{st}}(y) - F(y)\right) = -F(y) + m_5F(y) + m_5F(y)\zeta_1 - \frac{1}{2}\Theta m_5 F(y)\zeta_2 - m_6\zeta_2 - m_7\zeta_3$$

$$+\frac{3}{8}\Theta^2 m_5 F(y)\zeta_2^2 + \frac{1}{2}\Theta m_6\zeta_2^2 - \frac{1}{2}\Theta m_5 F(y)\zeta_1\zeta_2 + \frac{1}{2}\Theta m_7\zeta_2\zeta_3. \quad (53)$$

The bias and MSE of $\hat{F}_{Prop1_{st}}(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_{Prop1_{st}}(y)\right)$$

$$\cong F(y)(m_5 - 1) + \frac{3}{8}\Theta^2 m_5 F(y)\psi_{0200} + \frac{1}{2}\Theta m_6\psi_{0200} - \frac{1}{2}\Theta m_5 F(y)\psi_{1100} + \frac{1}{2}\Theta m_7\psi_{0110},$$

$$\text{MSE}(\hat{F}_{Prop1_{st}}(y)) \cong F^2(y)(m_5-1)^2 + m_5^2F^2(y)\psi_{2000} + m_6^2\psi_{0200} + m_7^2\psi_{0020} + \Theta^2k_5^2F^2(y)\psi_{0200}$$

$$-\Theta m_6 F(y)\psi_{0200} + 2\Theta m_5 m_6 F(y)\psi_{0200} - \frac{3}{4}\Theta^2 m_5 F^2(y)\psi_{0200} + \Theta m_5 F^2(y)\psi_{1100}$$

$$-2\Theta k_5^2 F^2(y)\psi_{1100} - 2m_5m_6F(y)\psi_{1100} - 2m_5m_7F(y)\psi_{1010} - \Theta m_7 F(y)\psi_{0110}$$

$$+2\Theta m_5 m_7 F(y)\psi_{0110} - 2m_6m_7\psi_{0110}. \quad (54)$$

The optimum values of m_5, m_6 and m_7, determined by minimizing (54), are

$$m_{5(\text{opt})} = \frac{8 - \Theta^2\psi_{0200}}{8\{1 + \psi_{2000}(1 - R_{1.23}^2)\}},$$

$$m_{6(\text{opt})} = \frac{F(y)\begin{bmatrix} \Theta^3\psi_{0200}^{3/2}(R_{23}^2 - 1) + \psi_{2000}^{1/2}(-8 + \Theta^2\psi_{0200})(R_{12} - R_{23}R_{13}) \\ +4\Theta\psi_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + \psi_{2000}(1 - R_{1.23}^2)\} \end{bmatrix}}{8\psi_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + \psi_{2000}(1 - R_{1.23}^2)\}},$$

$$m_{7(\text{opt})} = \frac{F(y)\psi_{2000}^{1/2}(8 - \Theta^2\psi_{0200})(R_{12} - R_{23}R_{13})}{8\psi_{0200}^{1/2}(R_{23}^2 - 1)\{-1 + \psi_{2000}(1 - R_{1.23}^2)\}}.$$

The simplified minimum MSE of $\hat{F}_{Prop1_{st}}(y)$ at the optimum values of m_5, m_6 and m_7 is

$$\text{MSE}_{\min}\left(\hat{F}_{Prop1_{st}}(y)\right) \cong \frac{F^2(y)\{64\psi_{2000}(1 - R_{1.23}^2) - \Theta^4\psi_{0200}^2 - 16\Theta^2\psi_{0200}\psi_{2000}(1 - R_{1.23}^2)\}}{64\{1 + \psi_{2000}(1 - R_{1.23}^2)\}}, \quad (55)$$

where $R_{1.23}^2 = \left(\frac{\psi_{1100}^2\psi_{0020} + \psi_{1010}^2\psi_{0200} - 2\psi_{1010}\psi_{1100}\psi_{0110}}{\psi_{2000}(\psi_{0200}\psi_{0020} - \psi_{0110}^2)}\right)$.

Here, (55) may be written as

$$\text{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) \cong \text{Var}_{\min}(\hat{F}_{Reg_{st}}(y)) - T_1 - T_2 \quad (56)$$

where

$$T_1 = \frac{F^2(y)(\Theta^2\psi_{0200}^2 - 8\psi_{1100}^2 + 8\psi_{0200}\psi_{2000})^2}{64\psi_{0200}^2\{1+\psi_{2000}(1-R_{12}^2)\}} \text{ and}$$

$$T_2 = \frac{F^2(y)(\Theta^2\psi_{0200} - 8)^2(\psi_{0200}\psi_{1010} - \psi_{0110}\psi_{1100})^2}{64\psi_{0200}^2\psi_{0020}(1-R_{23}^2)\{1+\psi_{2000}(1-R_{12}^2)\}\{1+\psi_{2000}(1-R_{1.23}^2)\}}.$$

It can be seen that $\hat{F}_{Prop1_{st}}(y)$ is more precise than $\hat{F}_{Reg_{st}}(y)$.

9.2 Second proposed family of estimators

On similar lines, second proposed family of estimators for estimating $F(y)$ is given by

$$\hat{F}_{Prop2_{st}}(y) = \left\{m_8\hat{F}_{st}(y) + m_9\left(\frac{F(x)-\hat{F}_{st}(x)}{F(x)}\right) + m_{10}\left(\frac{\bar{Z}-\hat{\bar{Z}}_{st}}{\bar{Z}}\right)\right\}\exp\left(\frac{a_{st}(F(x)-\hat{F}_{st}(x))}{a_{st}(F(x)-\hat{F}_{st}(x))+2b_{st}}\right), \quad (57)$$

where m_8, m_9 and m_{10} are unknown constants, a_{st} ($\neq 0$) and b_{st} are either two real numbers or functions of known population parameters of $I(X\leq x)$, like R_{12}, $\beta_{2(st)} = \sum_{i=1}^{L} W_h\beta_{2h(x)}$ (coefficient of kurtosis), $C_2 = \sum_{i=1}^{L} W_h C_{2h}$, etc.

The estimator $\hat{F}_{Prop2_{st}}(y)$ can also be written as

$$\hat{F}_{Prop2_{st}}(y) = \{m_8F(y)(1+\zeta_1) - m_9\zeta_2 - m_{10}\zeta_4\}\left(1 - \frac{1}{2}\Theta\zeta_2 + \frac{3}{8}\Theta^2\zeta_2^2 + \cdots\right). \quad (58)$$

Simplifying (58) and keeping terms only up to the second power of ζ_i's, we can write

$$\left(\hat{F}_{Prop2_{st}}(y) - F(y)\right) = -F(y) + m_8F(y) + m_8F(y)\zeta_1 - \frac{1}{2}\Theta m_8F(y)\zeta_2 - m_9\zeta_2 - m_{10}\zeta_4$$

$$+\frac{3}{8}\Theta^2 m_8F(y)\zeta_2^2 + \frac{1}{2}\Theta m_9\zeta_2^2 - \frac{1}{2}\Theta m_8F(y)\zeta_1\zeta_2 + \frac{1}{2}\Theta m_{10}\zeta_2\zeta_4. \quad (59)$$

The bias and MSE of $\hat{F}_{Prop2_{st}}(y)$, to the first order of approximation, are

$$\text{Bias}\left(\hat{F}_{Prop2_{st}}(y)\right)$$

$$\cong F(y)(m_8-1) + \frac{3}{8}\Theta^2 m_8F(y)\psi_{0200} + \frac{1}{2}\Theta m_9\psi_{0200} - \frac{1}{2}\Theta m_8F(y)\psi_{1100} + \frac{1}{2}\Theta m_9\psi_{0101},$$

$$\text{MSE}(\hat{F}_{Prop2_{st}}(y)) \cong F^2(y)(m_8-1)^2 + m_8^2F^2(y)\psi_{2000} + m_9^2\psi_{0200} + m_{10}^2\psi_{0002} + \Theta^2k_8^2F^2(y)\psi_{0200}$$

$$-\Theta m_9F(y)\psi_{0200} + 2\Theta m_8m_9F(y)\psi_{0200} - \frac{3}{4}\Theta^2 m_8F^2(y)\psi_{0200} + \Theta m_8F^2(y)\psi_{1100}$$

$$-2\Theta k_8^2F^2(y)\psi_{1100} - 2m_8m_9F(y)\psi_{1100} - 2m_8m_{10}F(y)\psi_{1001} - \Theta m_{10}F(y)\psi_{0101}$$

$$+2\Theta m_8m_{10}F(y)\psi_{0101} - 2m_9m_{10}\psi_{0101}. \quad (60)$$

The optimum values of m_8, m_9 and m_{10}, determined by minimizing (60), are

$$m_{8(\text{opt})} = \frac{8 - \Theta^2\psi_{0200}}{8\{1 + \psi_{2000}(1 - R^2_{1.24})\}},$$

$$m_{9(\text{opt})} = \frac{F(y)\left[\begin{array}{c}\Theta^3\psi^{3/2}_{0200}(R^2_{24} - 1) + \psi^{1/2}_{2000}(-8 + \Theta^2\psi_{0200})(R_{12} - R_{24}R_{14}) \\ +4\Theta\psi^{1/2}_{0200}(R^2_{24} - 1)\{-1 + \psi_{2000}(1 - R^2_{1.24})\}\end{array}\right]}{8\psi^{1/2}_{0200}(R^2_{24} - 1)\{-1 + \psi_{2000}(1 - R^2_{1.24})\}},$$

$$m_{10(\text{opt})} = \frac{F(y)\psi^{1/2}_{2000}(8 - \Theta^2\psi_{0200})(R_{12} - R_{24}R_{14})}{8\psi^{1/2}_{0200}(R^2_{24} - 1)\{-1 + \psi_{2000}(1 - R^2_{1.24})\}}.$$

The simplified minimum MSE of $\hat{F}_{Prop2_{st}}(y)$ at the optimum values of m_8, m_9 and m_{10} is

$$\text{MSE}_{\min}\left(\hat{F}_{Prop2_{st}}(y)\right) \cong \frac{F^2(y)\{64\psi_{2000}(1 - R^2_{1.24}) - \Theta^4\psi^2_{0200} - 16\Theta\psi_{0200}\psi_{2000}(1 - R^2_{1.24})\}}{64\{1 + \psi_{2000}(1 - R^2_{1.24})\}}, \quad (61)$$

where $R^2_{1.24} = \left(\frac{\psi^2_{1100}\psi_{0002} + \psi^2_{1001}\psi_{0200} - 2\psi_{1001}\psi_{1100}\psi_{0101}}{\psi_{2000}(\psi_{0200}\psi_{0002} - \psi^2_{0101})}\right)$.

Here, (61) may be written as

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) \cong \text{Var}_{\min}(\hat{F}_{Reg_{st}}(y)) - T_1 - T_3, \quad (62)$$

where

$$T_1 = \frac{F^2(y)(\Theta^2\psi^2_{0200} - 8\psi^2_{1100} + 8\psi_{0200})^2}{64\psi^2_{0200}\{1 + \psi_{2000}(1 - R^2_{12})\}} \text{ and}$$

$$T_3 = \frac{F^2(y)(\Theta^2\psi_{0200} - 8)^2(\psi_{0200}\psi_{1001} - \psi_{0101}\psi_{1100})^2}{64\psi^2_{0200}\psi_{0002}(1 - R^2_{24})\{1 + \psi_{2000}(1 - R^2_{12})\}\{1 + \psi_{2000}(1 - R^2_{1.24})\}}.$$

respectively, which shows that $\hat{F}_{Prop2_{st}}(y)$ is more precise than $\hat{F}_{Reg_{st}}(y)$.

In Table 8, we put some members of the Singh et al. [33], Grover and Kaur [11], and proposed families of estimators with selected choices of a and b.

10 Efficiency comparisons in stratified random sampling

In this section, the adapted and proposed estimators of $F(y)$ are compared in terms of the minimum MSEs. [(i)]

1. From (33) and (56),

$$\text{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \text{Var}(\hat{F}_{SRS_{st}}(y))\text{if}$$

$$F^2(y)\psi_{2000}R^2_{12} + T_1 + T_2 > 0.$$

2. From (35) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}(\hat{F}_{R_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}(\psi_{0200} - \psi_{1100})^2 + T_1 + T_2 > 0.$$

3. From (37) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}(\hat{F}_{P_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}(\psi_{0200} + \psi_{1100})^2 + T_1 + T_2 > 0.$$

4. From (40) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}_{\min}(\hat{F}_{Reg_{st}}(y))\text{if}$$

$$T_1 + T_2 > 0.$$

5. From (44) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}_{\min}(\hat{F}_{R,D_{st}}(y))\text{if}$$

$$\frac{F^2(y)\Theta^2\psi_{0200}\{\Theta^2\psi_{0200} + 16\psi_{2000}(1 - R_{12}^2)\}}{64\{1 + \psi_{2000}(1 - R_{12}^2)\}} + T_2 > 0.$$

6. From (46) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}(\hat{F}_{S_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}\left(\frac{\Theta\psi_{0200}}{2} - \psi_{1100}\right)^2 + T_1 + T_2 > 0.$$

7. From (50) and (56),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop1_{st}}(y)) < \mathrm{MSE}_{\min}(\hat{F}_{G,K_{st}}(y))\text{if}$$

$$T_2 > 0.$$

8. From (33) and (62),

$$\mathrm{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \mathrm{Var}(\hat{F}_{SRS_{st}}(y))\text{if}$$

$$\frac{1}{\psi_{0200}}F^2(y)\psi_{1100}^2 + T_1 + T_3 > 0.$$

9. From (35) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}(\hat{F}_{R_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}(\psi_{0200} - \psi_{1100})^2 + T_1 + T_3 > 0.$$

10. From (37) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}(\hat{F}_{P_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}(\psi_{0200} + \psi_{1100})^2 + T_1 + T_3 > 0.$$

11. From (40) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}_{\min}(\hat{F}_{Reg_{st}}(y))\text{if}$$

$$T_1 + T_3 > 0.$$

12. From (44) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}_{\min}(\hat{F}_{R,D_{st}}(y))\text{if}$$

$$\frac{F^2(y)\Theta^2\psi_{0200}\{\Theta^2\psi_{0200} + 16\psi_{2000}(1 - R_{12}^2)\}}{64\{1 + \psi_{2000}(1 - R_{12}^2)\}} + T_3 > 0.$$

13. From (46) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}_{\min}(\hat{F}_{S_{st}}(y))\text{if}$$

$$\frac{F^2(y)}{\psi_{2000}}\left(\frac{\Theta\psi_{0200}}{2} - \psi_{1100}\right)^2 + T_1 + T_3 > 0.$$

14. From (50) and (62),

$$\text{MSE}_{\min}(\hat{F}_{Prop2_{st}}(y)) < \text{MSE}_{\min}(\hat{F}_{G,K_{st}}(y))\text{if}$$

$$T_3 > 0.$$

The proposed families of estimators are always more precise than the adapted estimators as the above conditions (i)–(xiv) are always true.

11 Empirical study in stratified random sampling

In this section, we conduct a numerical study to investigate the performances of the adapted and proposed CDF estimators. For this purpose, five populations are considered. The summary statistics of these populations are reported in Tables 9–12. The PRE of an estimator $\hat{F}_i(y)$

with respect to $\hat{F}_1(y)$ is

$$\text{PRE}\left(\hat{F}_i(y), \hat{F}_1(y)\right) = \frac{\text{Var}(\hat{F}_1(y))}{\text{MSE}_{\min}(\hat{F}_i(y))} \times 100$$

where i = 2,3,. . .,9.

The PREs of distribution function estimators, computed from four populations, are given in Tables 9–12.

Population I (Source: Koyuncu and Kadilar, [36])

Y: The number of teachers and

X: The number of students in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

Population II (Source: Koyuncu and Kadilar, [36])

Y: The number of teachers and

X: The number of classes in both primary and secondary schools in Turkey in 2007 for 923 districts in six regions.

Population III (Source: Kadilar and Cingi, [37])

Y: Apple production amount in 1999 and

X: The number of apple trees in 1999.

Population IV (Source: Kadilar and Cingi, [37])

Y: Apple production amount in 1999 and

X: Apple production amount in 1998.

From the numerical results, presented in Tables 13–16, it is observed that the PREs of all families of estimators change with the choices of *a* and *b*. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of PRE. It can be seen that, for all data sets, the second proposed family of the estimators perform better than the first proposed family of estimators.

We computed sample size in stratum *h*. Here we took sample sizes of 180, 180, 140 and 140 from four populations. Then we used stratified random sampling, and the MSEs (minimum) of the proposed families of estimators were computed in Eqs 56 and 62, respectively. Lastly, the adapted estimators and proposed families of estimators were compared with each other with respect to their PRE values. The PRE results are shown in Tables 13–16. In Tables 9–12, we can observe the descriptive statistics regarding the populations, strata, and sample size. From the numerical results, presented in Tables 13–16, it is observed that the PREs of all families of estimators change with the choices of a and b. It is further noted that the proposed families of estimators are more precise than the adapted distribution function estimators of Cochran [31], Murthy [32], Rao [4], Singh et al. [33] and Grover and Kaur [11], in terms of the PRE. It can be seen that, for all data sets, the second proposed family of estimators perform better than the first proposed family of estimators. We can also see a rise in the value of PREs when the value of a = 1; and b = R_{12}; a = R_{12} and b = C_2; a = R_{12} and b = $\beta_{2(\text{st})}$, and see a slight fall in PREs when a = 1 and b = NF(x)).

12 Conclusion

In this paper, we have proposed two new families of estimators for estimating the finite population distribution function under simple and stratified random sampling schemes. The proposed estimators required supplementary information about the sample mean and the ranks of the auxiliary variable. The biases and MSEs of the proposed families of estimators were derived using the first order approximation. Based on theoretical and numerical comparative studies, it can be concludethat the proposed families of estimators are more precise than their existing counterparts. Thus, we recommend using the sample mean and the ranks of the auxiliary variable with the proposed families of estimators for estimating the finite population distribution function under simple or stratified random sampling. It would be interesting to extend the suggested estimators to two-phase and stratified two-phase sampling schemes. Furthermore, the proposed estimators could also be generalised by utilising information about multi-auxillary variables.

Author Contributions

Conceptualization: Sardar Hussain.

Formal analysis: Sardar Hussain.

Funding acquisition: Sardar Hussain.

Investigation: Sardar Hussain.

Methodology: Sardar Hussain, Sohaib Ahmad.

Validation: Sardar Hussain.

Writing – original draft: Sardar Hussain, Sohaib Ahmad.

Writing – review & editing: Mariyam Saleem, Sohail Akhtar.

References

1. Murthy MN. Sampling Theory and Methods. Statistical Pub. 1967 Society, Calcutta.
2. Sisodia BV, Dwivedi VK. Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics. 1981.
3. Srivastava SK, Jhajj HS. A class of estimators of the population mean using multiauxiliary information. Calcutta Statistical Association Bulletin. 1983 Mar; 32(1–2):47–56.
4. Rao TJ. On certail methods of improving ration and regression estimators. Communications in Statistics-Theory and Methods. 1991 Jan 1; 20(10):3325–40.
5. Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 1999 Sep; 41(5):627–36.
6. Singh S. Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy. Springer Science & Business Media; 2003.
7. Kadilar C, Cingi H. Estimator of a population mean using two auxiliary variables in simple random sampling. International Mathematical Journal. 2004; 5:357–67.
8. Kadilar C, Cingi H. Improvement in estimating the population mean in simple random sampling. Applied Mathematics Letters. 2006 Jan 1; 19(1):75–9.
9. Gupta S, Shabbir J. On improvement in estimating the population mean in simple random sampling. Journal of Applied Statistics. 2008 May 1; 35(5):559–66.
10. Grover LK, Kaur P. Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: theory and methods. South African Statistical Journal. 2011 Jan 1; 45(2):205–30.
11. Grover LK, Kaur P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. Communications in Statistics-Simulation and Computation. 2014 Jan 1; 43(7):1552–74.
12. Lu J. Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering. Mathematical Problems in Engineering. 2017 Jan 1;2017. https://doi.org/10.1155/2017/2696108 PMID: 29578548
13. Muneer S, Shabbir J, Khalil A. Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. Communications in Statistics-Theory and Methods. 2017 Mar 4; 46(5):2181–92.
14. Shabbir J, Gupta S. Estimation of finite population mean in simple and stratified random sampling using two auxiliary variables. Communications in Statistics-Theory and Methods. 2017 Oct 18; 46 (20):10135–48.

15. Gupta RK, Yadav SK. Improved estimation of population mean using information on size of the sample. American Journal of Mathematics and Statistics. 2018; 8(2):27–35.

16. Haq A, Khan M, Hussain Z. A new estimator of finite population mean based on the dual use of the auxiliary information. Communications in Statistics-Theory and Methods. 2017 May 3; 46(9):4425–36.

17. Chambers RL, Dunstan R. Estimating distribution functions from survey data. Biometrika. 1986 Dec 1; 73(3):597–604.

18. Rao JN, Kovar JG, Mantel HJ. On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika. 1990 Jun 1:365–75.

19. Rao JN. Estimating totals and distribution functions using auxiliary information at the estimation stage. Journal of Official Statistics. 1994 Jun 1; 10(2):153.

20. Kuk AY. A kernel method for estimating finite population distribution functions using auxiliary information. Biometrika. 1993 Jun 1; 80(2):385–92.

21. Ahmed MS. and Abu-Dayyeh W. Estimation of finite-population distribution function using multivariate auxiliary information. Statistics in Transition. 2001 5(3):501–507.

22. Rueda M, Martínez S, Martínez H, Arcos A. Estimation of the distribution function with calibration methods. Journal of statistical planning and inference. 2007 Feb 1; 137(2):435–48.

23. Singh HP, Singh S, Kozak M. A family of estimators of finite-population distribution function using auxiliary information. Acta applicandae mathematicae. 2008 Nov 1; 104(2):115–30.

24. Yaqub M, Shabbir J. Estimation of population distribution function in the presence of non-response. Hacettepe Journal of Mathematics and Statistics. 2018 Apr 1; 47(2):471–511.

25. Chen F, Chen S. Injury severities of truck drivers in single-and multi-vehicle accidents on rural highways. Accident Analysis & Prevention. 2011 Sep 1; 43(5):1677–88.

26. Zeng Q, Wen H, Huang H, Pei X, Wong SC. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. Accident Analysis & Prevention. 2017 Feb 1; 99:184–91.

27. Dong B, Ma X, Chen F, Chen S. Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model. Journal of Advanced Transportation. 2018 Jan 1;2018. https://doi.org/10.1155/2018/7905140 PMID: 32908973

28. Chen F, Chen S, Ma X. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. Journal of safety research. 2018 Jun 1; 65:153–9. https://doi.org/10.1016/j.jsr.2018.02.010 PMID: 29776524

29. Zeng Q, Guo Q, Wong SC, Wen H, Huang H, Pei X. Jointly modeling area-level crash rates by severity: a Bayesian multivariate random-parameters spatio-temporal Tobit regression. Transportmetrica A: Transport Science. 2019 Nov 29; 15(2):1867–84.

30. Zeng Q, Wen H, Wong SC, Huang H, Guo Q, Pei X. Spatial joint analysis for zonal daytime and nighttime crash frequencies using a Bayesian bivariate conditional autoregressive model. Journal of Transportation Safety & Security. 2020 Apr 20; 12(4):566–85. https://doi.org/10.4271/2016-01-1439 PMID: 27648455

31. Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. The journal of agricultural science. 1940 Apr; 30(2):262–75.

32. Murthy MN. Product method of estimation. Sankhyā: The Indian Journal of Statistics, Series A. 1964 Jul 1:69–74.

33. Singh R, Chauhan P, Sawan N, Smarandache F. Improvement in estimating the population mean using exponential estimator in simple random sampling. International Journal of Statistics & Economics. 2009 3(A09):13–18

34. Gujarati DN. Basic econometrics. Tata McGraw-Hill Education; 2009.

35. Särndal CE. Methods for estimating the precision of survey estimates when imputation has been used. Survey methodology. 1992; 18(2):241–52.

36. Koyuncu N, Kadilar C. Ratio and product estimators in stratified random sampling. Journal of statistical planning and inference. 2009 Aug 1; 139(8):2552–8.

37. Kadilar C, Cingi HB. Ratio estimators in stratified random sampling. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2003 Mar; 45(2):218–25.