African Multidisciplinary
Journal of Sciences and Artificial Intelligence

ISSN : 0000-0000

Index: Harvard, Boston, Sydney University, Dimensions, Lens, ResearchGet Scilit, Semantic, Google Scholar, Base etc

https://doi.org/10.58578/AMJSAI.v1i2.3799

# An Efficient Non-Linear Estimator for Estimating the Finite Population Distribution Function under Simple Random Sampling Design

**Adam Rabiu[*1], Ahmed Audu[2], and Ibrahim Abubakar[3]**

[*1]National Open University of Nigeria, Nigeria

[2]Usmanu Danfodiyo University, Sokoto, Nigeria

[3]Kaduna State University, Kaduna, Nigeria

adrabiu@noun.edu.ng

**Article Info:**

| Submitted: | Revised: | Accepted: | Published: |
|---|---|---|---|
| Aug 15, 2024 | Aug 27, 2024 | Aug 30, 2024 | Sep 3, 2024 |

**Abstract**

The main purpose of this paper is to propose an efficient non-linear estimator for estimating the population distribution function under Simple Random Sampling (SRS). The properties (Bias and Mean Square Error (MSE)) of the suggested estimator are obtained up to the first-order approximation using Taylor's series expansion approach. The performance of the proposed estimator over some existing estimators is theoretically compared and efficiency conditions under which the proposed estimator outperforms existing estimators were obtained. The theoretical findings are supported numerically by empirical studies using five different population data sets and the result shows that the proposed estimator performed better than the existing estimators considered in the literature.

**Keywords:** Auxiliary variable, Exponential estimator, Cumulative Distribution Function

**Symbols and Notations**

$I(.)$: Indicator variable

$Y$ : The study variable

$X$ : The auxiliary variable

$I(Y \leq t_y)$: Indicator variable based on $Y$

$I(X \leq t_x)$: Indicator variable based on $X$

$\hat{F}_Y(t_y) = \dfrac{\sum_{i=1}^{n} I(Y_i = t_y)}{n}$: The sample distribution function of $Y$

$\hat{F}_X(t_x) = \dfrac{\sum_{i=1}^{n} I(X_i = t_x)}{n}$ : The sample distribution function of $X$

$S^2_{F_Y(t_y)} = \dfrac{\sum_{i=1}^{N} \left[ I(Y_i \leq t_y) - F_Y(t_y) \right]^2}{(N-1)}$ : The population variance of $I(Y \leq t_y)$

$S^2_{F_X(t_x)} = \dfrac{\sum_{i=1}^{N} \left[ I(X_i \leq t_x) - F_X(t_x) \right]^2}{(N-1)}$ : The population variance of $I(X \leq t_x)$

$C_{F_Y(t_y)} = \dfrac{S_{F_Y(t_y)}}{F_Y(t_y)}$ : The population coefficient of variation of $I(Y \leq t_y)$

$C_{F_X(t_x)} = \dfrac{S_{F_X(t_x)}}{F_X(t_x)}$: The population coefficient of variation of $I(X \leq t_x)$

$S_{F_Y(t_y)F_X(t_x)} = \dfrac{\sum_{i=1}^{N} \left[ I(Y_i \leq t_y) - F_Y(t_y) \right] \left[ I(X_i \leq t_x) - F_X(t_x) \right]}{(N-1)}$: The population covariance between $I(Y \leq t_y)$ and $I(X \leq t_x)$

$R_{F_Y(t_y)F_X(t_x)} = \dfrac{S_{F_Y(t_y)F_X(t_x)}}{S_{F_Y(t_y)} S_{F_X(t_x)}}$: The population correlation coefficient between $I(Y \leq t_y)$ and $I(X \leq t_x)$.

## Introduction

Many scholars have examined the use of an auxiliary variable in the literature of survey sampling to improve the effectiveness of their developed estimators for estimating common parameters such as mean, total, variance, and among others, in such cases, traditional ratio, product and regression estimators produce reliable outcomes for parameters that are not known.

The ratio approach and the product method have been frequently utilized for estimating unknown population parameters where there is a high positive and negative correlation between the study parameter and supplementary variable. Several authors have already presented many ratio type and product type estimators applying various types of linear transformation of the original auxiliary variables. The drawback of these classic of estimators is that they require a very precise linear transformation of the auxiliary variable, which limits the scope of applications of this classic in practice. To address this limitation, we offer a non-linear transformation class estimator of the population distribution function in this paper.

The problem of estimating the finite population distribution function occurs when it is desired to know the proportion of study variable values that are less or equal to a given value. Estimating CDF is required in a variety of scenarios for economists, for instance, it is interesting to know the proportion of the population that 25% or more Nigerians do not have skills. Similarly, a soil scientist may be interested in determining the percentage of clay in the soil. Furthermore, policymakers may be interested in understanding the proportion of people living in a developing country below the poverty line.

An extensive literature is available on the estimation of population mean, total, etc. However, no effort has been devoted to the development of efficient methods for population cumulative distribution function; here our focus is on the estimation of the finite population CDF with supplementary variables. Chambers and Dunstan (1986) consider the estimation of the population CDF and quantiles with a model-based approach, on a similar line to Rao et al (1990) proposed ratio and regression estimators for estimating the CDF under a general sampling scheme, Singh et al (2008) consider the problem of estimating the CDF and quantiles with the use of auxiliary variable at the estimation stage of the survey. To the base of our knowledge, recent research may be seen in Martinez et al (2010), Mayer et al (2010), Berger and Munoz (2015), Hussain et al (2020), Yakubu and Shabbir (2020), GaribNath Singh and Mahammood Usman (2021), Sardar et al (2022), Rueda and Illescas (2022), Ahmad et al (2022), Mohsin Abbas and Abdul Haqto (2022), Sohaib Ahmad et al (2023) name a few.

## Materials and Methods

Consider a finite identifiable population $U=\{U_1,U_2,...,U_N\}$ having N distinct units. Let $y_i$ and $x_i$ be the $i^{th}$ values of the study variable $Y$ and auxiliary variable $X$ which takes the values such that $(y_i,x_i)\in U_i \text{ in } U$. We assumed that the information for all the units of $X$ is available in the population. The problem is to estimate the CDF of $F_Y\left(t_y\right)$ and $F_X\left(t_x\right)$ $Y$ and $X$ respectively which defined as

$$F_Y\left(t_y\right)=\frac{1}{N}\sum_{i\in u} I\left(Y\le t_y\right) \text{ and } F_X\left(t_x\right)=\frac{1}{N}\sum_{i\in u} I\left(X=t_x\right)$$

The usual unbiased estimator $F_{Y_{us}}\left(t_y\right)$ as well as its variance are given in (1) and (2) respectively

$$\hat{F}_{Y_{usual}}\left(t_y\right)=\frac{\sum_{i=1}^{n} I\left(Y_i=t_y\right)}{n} \tag{1}$$

$$Var\left(\hat{F}_{Y_{usual}}\left(t_y\right)\right)=F_Y^2\left(t_y\right)\xi_{20} \tag{2}$$

Following Cochran (1940), ratio type estimator of $F_Y\left(t_y\right)$ denoted by $\hat{F}_{Y_c}\left(t_y\right)$ as well as its bias and MSE are given in (3), (4), and (5), respectively

$$\hat{F}_{Y_c}\left(t_y\right)=\hat{F}_Y\left(t_y\right)\left[\frac{F_X\left(t_x\right)}{\hat{F}_X\left(t_x\right)}\right] \tag{3}$$

$$Bias\left[\hat{F}_{Y_c}\left(t_y\right)\right]=\hat{F}_Y\left(t_y\right)\left(\xi_{02}-\xi_{11}\right) \tag{4}$$

$$MSE\left[\hat{F}_{Y_c}\left(t_y\right)\right]=F_Y^2\left(t_y\right)\left(\xi_{20}+\xi_{02}-2\xi_{11}\right) \tag{5}$$

Following Murthy (1964), the product estimator of $F_Y\left(t_y\right)$ denoted by $\hat{F}_{Y_m}\left(t_y\right)$ as well as its bias and MSE are given in (6), (7), and (8), respectively

$$\hat{F}_{Y_m}(t_y) = \hat{F}_Y(t_y)\left[\frac{\hat{F}_X(t_x)}{F_X(t_x)}\right] \tag{6}$$

The bias and MSE of Equation (6) are

$$Bias\left[\hat{F}_{Y_m}(t_y)\right] = F_Y(t_y)\xi_{11} \tag{7}$$

$$MSE\left[\hat{F}_{Y_m}(t_y)\right] = F_Y^2(t_y)(\xi_{20} + \xi_{02} + 2\xi_{11}) \tag{8}$$

Following Bahl and Tuteja (1991), ratio type and product type exponential estimator of $F_Y(t_y)$ denoted by $\hat{F}_{Y_{BTR}}(t_y)$ as well as its bias and MSE are given in (9), (10), (11), (12), (13), and (14), respectively

$$\hat{F}_{Y_{BTR}}(t_y) = \hat{F}_Y(t_y)\exp\left[\frac{F_X(t_x) - \hat{F}_X(t_x)}{F_X(t_x) + \hat{F}_X(t_x)}\right] \tag{9}$$

$$\hat{F}_{Y_{BTP}}(t_y) = \hat{F}_Y(t_y)\exp\left[\frac{\hat{F}_X(t_x) - F_X(t_x)}{\hat{F}_X(t_x) + F_X(t_x)}\right] \tag{10}$$

The bias and MSE of equation (9) and (10) are given below

$$Bias\left[\hat{F}_{Y_{BTR}}(t_y)\right] = \frac{F_Y(t_y)}{4}\left(\frac{3}{2}\xi_{02} - 2\xi_{11}\right) \tag{11}$$

$$MSE\left[\hat{F}_{Y_{BTR}}(t_y)\right] = \frac{F_Y^2(t_y)}{4}(4\xi_{20} + \xi_{02} - 4\xi_{11}) \tag{12}$$

$$Bias\left[\hat{F}_{Y_{BTP}}(t_y)\right] = \frac{F_Y(t_y)}{4}\left(2\xi_{11} - \frac{1}{2}\xi_{02}\right) \tag{13}$$

$$MSE\left[\hat{F}_{Y_{BTP}}(t_y)\right] = \frac{F_Y^2(t_y)}{4}(4\xi_{20} + \xi_{02} + 4\xi_{11}) \tag{14}$$

Following Sisodia and Dwivesdi (1981) proposed ratio type estimator of $F_Y(t_y)$ denoted by $\hat{F}_{Y_{sd}}(t_y)$ as well as its bias and MSE are given in (15), (16), and (17), respectively

$$\hat{F}_{Y_{sd}}(t_y) = \hat{F}_Y(t_y)\left[\left(\frac{F_X(t_x) + C_{F_X(t_x)}}{\hat{F}_X(t_x) + C_{F_X(t_x)}}\right)\right] \tag{15}$$

The bias and MSE of equation (15) are given as

$$Bias\left[\hat{F}_{Y_{sd}}(t_y)\right] = F_Y(t_x)\left(\lambda^2\xi_{02} - 2\lambda\xi_{11}\right) \tag{16}$$

$$MSE\left[\hat{F}_{Y_{sd}}\left(t_y\right)\right]=F_Y^2\left(t_y\right)\left(\xi_{20}+\lambda^2\xi_{02}-2\lambda\xi_{11}\right) \tag{17}$$

where, $\lambda=\dfrac{F_X\left(t_x\right)}{F_X\left(t_x\right)+C_{F_X\left(t_x\right)}}$

Following Upadhyaya and Singh (1999) proposed ratio type estimator of $F_Y\left(t_y\right)$ denoted by $\hat{F}_{Y_{ups}}\left(t_y\right)$ as well as its bias and MSE are given in (18), (19), and (20), respectively

$$\hat{F}_{Y_{ups}}\left(t_y\right)=\hat{F}_Y\left(t_y\right)\left[\frac{F_X\left(t_x\right)C_{F_X\left(t_x\right)}+\beta_2}{\hat{F}_X\left(t_x\right)C_{F_X\left(t_x\right)}+\beta_2}\right] \tag{18}$$

The bias and MSE of equation (18) are given below

$$Bias\left[\hat{F}_{Y_{ups}}\left(t_y\right)\right]=F_Y\left(t_x\right)\left(\delta^2\xi_{02}-2\delta\xi_{11}\right) \tag{19}$$

$$MSE\left[\hat{F}_{Y_{ups}}\left(t_y\right)\right]=F_Y^2\left(t_y\right)\left(\xi_{20}+\delta^2\xi_{02}-2\delta\xi_{11}\right) \tag{20}$$

Where $\delta=\dfrac{F_X\left(t_x\right)C_{F_X\left(t_x\right)}}{F_X\left(t_x\right)C_{F_X\left(t_x\right)}+\beta_2}$

Following Singh and Tailor (2003) ratio estimator of $F_{Y_{us}}\left(t_y\right)$ denoted by $\hat{F}_{st}\left(t_y\right)$ as well as its bias and MSE are given in (21), (22), and (23), respectively

$$\hat{F}_{Y_{st}}\left(t_y\right)=\hat{F}_Y\left(t_y\right)\left[\frac{F_X\left(t_x\right)+R_{F_Y\left(t_y\right)F_X\left(t_x\right)}}{\hat{F}_X\left(t_x\right)+R_{F_Y\left(t_y\right)F_X\left(t_x\right)}}\right] \tag{21}$$

The bias and MSE of equation (21) is obtained as

$$Bias\left[\hat{F}_{Y_{st}}\left(t_y\right)\right]=F_Y\left(t_x\right)\left(\gamma^2\xi_{02}-2\gamma\xi_{11}\right) \tag{22}$$

$$MSE\left[\hat{F}_{Y_{st}}\left(t_y\right)\right]=F_Y^2\left(t_y\right)\left(\xi_{20}+\gamma^2\xi_{02}-2\gamma\xi_{11}\right) \tag{23}$$

where; $\gamma = \dfrac{F_X(t_x)}{F_X(t_x) + R_{F_Y(t_y)F_X(t_x)}}$

Following Singh *et al.* (2004) estimator of $F_{Y_{us}}(t_y)$ denoted by $\hat{F}_{Y_{se}}(t_y)$ as well as its bias and MSE are given in (24), (25) and (26), respectively

$$\hat{F}_{Y_{se}}(t_y) = \hat{F}_Y(t_y)\left[\frac{F_X(t_x) + \beta_2}{\hat{F}_X(t_x) + \beta_2}\right] \tag{24}$$

The bias and MSE of equation (24) are

$$Bias\left[\hat{F}_{Y_{se}}(t_y)\right] = F_Y(t_x)\left(\phi^2\xi_{02} - 2\phi\xi_{11}\right) \tag{25}$$

$$MSE\left[\hat{F}_{Y_{se}}(t_y)\right] = F_Y^2(t_y)\left(\xi_{20} + \phi^2\xi_{02} - 2\phi\xi_{11}\right) \tag{26}$$

where; $\phi = \dfrac{F_X(t_x)}{F_X(t_x) + \beta_2}$

Following Jerajuddin and Kishun (2016) estimator of population mean $F_{Y_{us}}(t_y)$ denoted by $\hat{F}_{Y_{jk}}(t_y)$ as well as its bias and MSE are given in (27), (28) and (29), respectively

$$\hat{F}_{Y_{jk}}(t_y) = \hat{F}_Y(t_y)\left[\frac{F_X(t_x) + n}{\hat{F}_X(t_x) + n}\right] \tag{27}$$

The bias and MSE of equation (27) are obtained as follows

$$Bias\left[\hat{F}_{Y_{jk}}(t_y)\right] = F_Y(t_x)\left(\tau^2\xi_{02} - 2\tau\xi_{11}\right) \tag{28}$$

$$MSE\left[\hat{F}_{Y_{jk}}(t_y)\right] = F_Y^2(t_y)\left(\xi_{20} + \tau^2\xi_{02} - 2\tau\xi_{11}\right) \tag{29}$$

Where $\tau = \dfrac{F_X(t_x) + n}{F_X(t_x) + n}$

Following Gupta and Yadav (2018) suggested the improved estimation based on Jerajuddin and Kishun (2016) estimator in population mean $F_{Y_{us}}(t_y)$ denoted by $\hat{F}_{Y_{jk}}(t_y)$ as well as its bias and MSE is given in (30), (31) and (32), respectively

$$\hat{F}_{Y_{gy}}\left(t_y\right)=\hat{F}_Y\left(t_y\right)\left[\alpha+(1-\alpha)\left(\frac{F_X\left(t_x\right)+n}{\hat{F}_X\left(t_x\right)+n}\right)\right] \tag{30}$$

The bias and MSE obtained as

$$Bias\left[\hat{F}_{Y_{gy}}\left(t_y\right)\right]=F_Y\left(t_x\right)\left(\alpha^2\xi_{02}-\eta\xi_{11}+\alpha\eta\xi_{11}-\alpha\eta^2\xi_{02}\right) \tag{31}$$

$$MSE\left[\hat{F}_{Y_{gy}}\left(t_y\right)\right]=F_Y^2\left(t_y\right)\left\{\xi_{20}+\eta^2\xi_{02}-2\eta\xi_{11}-\left[\frac{\left(\eta^2\xi_{02}-\eta\zeta_{11}\right)^2}{\eta^2\xi_{02}}\right]\right\} \tag{32}$$

Where; $\eta=\dfrac{F_X\left(t_x\right)+n}{F_X\left(t_x\right)+n}$

**Propose estimator**

motivated by Ahmad et al (2021), we proposed an efficient non-linear estimator of finite population cumulative distribution function as

$$\hat{F}_{Y_{AD}}\left(t_y\right)=\hat{F}_Y\left(t_y\right)\left\{\frac{1}{2}\left[\gamma^{\left(\frac{F_X(t_x)-\hat{F}_X(t_x)}{F_X(t_x)+\hat{F}_X(t_x)}\right)}+\exp\left(\frac{\hat{F}_X\left(t_x\right)-F_X\left(t_x\right)}{\hat{F}_X\left(t_x\right)+F_X\left(t_x\right)}\right)\right]\right\} \tag{33}$$

Where $\gamma$ is constant

To obtain the bias and MSE, we define

$E\left(e_0\right)=\dfrac{\hat{F}_Y\left(t_y\right)-F_Y\left(t_y\right)}{F_Y\left(t_y\right)}$ And $E\left(e_1\right)=\dfrac{\hat{F}_X\left(t_x\right)-F_X\left(t_x\right)}{F_X\left(t_x\right)}$ Such that $E\left(e_0\right)=E\left(e_1\right)=0$

$$\left.\begin{aligned}E\left(e_0^2\right)&=\left(\frac{1}{n}-\frac{1}{N}\right)C_{F_Y\left(t_y\right)}^2=\xi_{20}\\E\left(e_1^2\right)&=\left(\frac{1}{n}-\frac{1}{N}\right)C_{F_X\left(t_x\right)}^2=\xi_{02}\\E\left(e_0e_1\right)&=\left(\frac{1}{n}-\frac{1}{N}\right)R_{F_Y\left(t_y\right)F_X\left(t_x\right)}C_{F_Y\left(t_y\right)}C_{F_X\left(t_x\right)}=\xi_{11}\end{aligned}\right\} \tag{34}$$

Taylor's Expansion was used for the derivation of properties (Bias, Mean Square Error, and Minimum Mean Square Error) of the proposed estimator.

Now expressing (33) in error terms as:

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[\gamma^{\left(\frac{F_X\left(t_x\right)-F_X\left(t_x\right)\left(1+e_1\right)}{F_X\left(t_x\right)+F_X\left(t_x\right)\left(1+e_1\right)}\right)}+\exp\left(\frac{F_X\left(t_x\right)\left(1+e_1\right)-F_X\left(t_x\right)}{F_X\left(t_x\right)\left(1+e_1\right)+F_X\left(t_x\right)}\right)\right]\right\}$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[\gamma^{\left(\frac{-e_1}{2+e_1}\right)}+\exp\left(\frac{e_1}{2+e_1}\right)\right]\right\}$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[\gamma^{\left(\frac{-e_1}{2+e_1}\right)}+\exp\left(\frac{e_1}{2+e_1}\right)\right]\right\}$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[\gamma^{\left(\frac{-e_1}{2}\right)\left(1-\frac{e_1}{2}+\frac{e_1^2}{4}\right)}+\exp\left(\frac{e_1}{2}\right)\left(1-\frac{e_1}{2}+\frac{e_1^2}{4}\right)\right]\right\}$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[\exp\left(\ln\gamma^{\left(\frac{e_1}{2}+\frac{e_1^2}{4}\right)}\right)+\exp\left(\frac{e_1}{2}+\frac{e_1^2}{4}\right)\right]\right\}$$

Neglecting the term of order more than two because we restricted to first order approximation, we have,

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left\{\frac{1}{2}\left[2-\frac{(\ln\gamma)e_1}{2}+\frac{(\ln\gamma)e_1^2}{4}+\frac{(\ln\gamma)^2e_1^2}{8}+\frac{e_1}{2}-\frac{e_1^2}{8}\right]\right\}$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left(1+e_0\right)\left[1-\frac{(\ln\gamma)e_1}{4}+\frac{(\ln\gamma)e_1^2}{8}+\frac{(\ln\gamma)^2e_1^2}{16}+\frac{e_1}{4}-\frac{e_1^2}{16}\right]$$

$$\hat{F}_{Y_{AD}}\left(t_y\right)=F_Y\left(t_y\right)\left[\begin{array}{l}1-\frac{(\ln\gamma)e_1}{4}+\frac{(\ln\gamma)e_1^2}{8}+\frac{(\ln\gamma)^2e_1^2}{16}+\frac{e_1}{4}-\frac{e_1^2}{16}+e_0-\frac{e_0e_1(\ln\gamma)}{4}+\\ \frac{e_0e_1}{4}\end{array}\right] \quad (35)$$

Subtracting $F_Y\left(t_y\right)$ from both sides of (35) we have,

$$\hat{F}_{Y_{AD}}\left(t_y\right)-F_Y\left(t_y\right)=F_Y\left(t_y\right)\left[\begin{array}{l}e_0+\frac{e_1}{4}+\frac{e_0e_1}{4}-\frac{(\ln\gamma)e_1}{4}+\frac{(\ln\gamma)e_1^2}{8}+\frac{(\ln\gamma)^2e_1^2}{16}-\frac{e_1^2}{16}-\\ \frac{(\ln\gamma)e_0e_1}{4}\end{array}\right] \quad (36)$$

Taking Expectations of both sides of (36)

$$E\left[\hat{F}_{Y_{AD}}\left(t_y\right)-F_Y\left(t_y\right)\right]=F_Y\left(t_y\right)E\left[e_0+\frac{e_1}{4}+\frac{e_0e_1}{4}-\frac{e_1^2}{16}-\frac{(\ln\gamma)e_1}{4}+\frac{(\ln\gamma)e_1^2}{8}+\frac{(\ln\gamma)^2e_1^2}{16}-\frac{(\ln\gamma)e_0e_1}{4}\right]$$

Applying the results of (34) obtaining the $Bias\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]$ as

$$Bias\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]=\frac{F_Y\left(t_x\right)}{4}\left\{(1-\ln\gamma)\xi_{11}-\frac{\xi_{02}}{4}\left[1-2\ln\gamma-(\ln\gamma)^2\right]\right\} \quad (37)$$

To obtain the MSE, we square both sides of (36) as well as expectations to both sides we have,

$$E\left[\hat{F}_{Y_{AD}}\left(t_y\right)-F_Y\left(t_y\right)\right]^2=F_Y^2\left(t_y\right)E\left[e_0+\frac{e_1}{4}-\frac{(\ln\gamma)e_1}{4}\right]^2$$

By simplifying the above expression, we obtain the MSE of the proposed estimator $\hat{F}_{Y_{AD}}\left(t_y\right)$ as

$$MSE\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]=\frac{F_Y^2\left(t_y\right)}{4}\left\{4\xi_{20}+\frac{\xi_{02}}{4}\left[1+(\ln\gamma)^2-2\ln\gamma\right]+2\xi_{11}(1-\ln\gamma)\right\} \quad (38)$$

To obtain the minimum MSE of the proposed estimator, we differentiate (38) with respect to $\gamma$ and equate the derivative to zero as;

$$\frac{\partial MSE\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]}{\partial\gamma}=\frac{F_Y^2\left(t_y\right)}{4}\left[\ln\gamma\xi_{02}-\xi_{02}-4\xi_{11}\right]=0$$

$$\gamma_{opt}=e^{\psi} \text{ Where } \psi=1+\frac{4\xi_{11}}{\xi_{02}}$$

When we substitute $\gamma_{Opt}$ in (38) we have a minimum MSE of $\hat{F}_{Y_{AD}}\left(t_y\right)$ as

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]=\frac{F_Y^2\left(t_y\right)}{4}\left\{4\xi_{20}+\frac{\xi_{02}}{4}\left[1+\psi^2-2\psi\right]+2\xi_{11}(1-\psi)\right\}$$

**Efficiency condition**

In this section efficiency of the proposed estimator is compared with the efficiencies of some existing estimators in the literature. The proposed estimator $\hat{F}_{Y_{AD}}\left(t_y\right)$ is more efficient than

1. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < \hat{F}_{Y_{usual}}\left(t_y\right)$$

$$\frac{\xi_{02}}{16}\left(1+\psi^2-2\psi\right)-\frac{1}{2}\xi_{11}\left(1-\psi\right)<0$$

2. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < MSE\left[\hat{F}_{Y_c}\left(t_y\right)\right]$$

$$\frac{\left(\psi-2\right)}{\xi_{02}}\left(\psi\xi_{02}-8\xi_{11}\right)<15$$

3. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < MSE\left[\hat{F}_{Y_m}\left(t_y\right)\right]$$

$$\psi\left(\psi-\frac{8\xi_{11}}{\xi_{02}}-2\right)<15$$

4. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < MSE\left[\hat{F}_{Y_{BTR}}\left(t_y\right)\right]$$

$$\psi\left(\psi-2\right)-\frac{8\xi_{11}}{\xi_{02}}\left(\psi-3\right)<3$$

5. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < MSE\left[\hat{F}_{Y_{BTP}}\left(t_y\right)\right]$$

$$\psi\left(\psi-2\right)-\frac{8\xi_{11}}{\xi_{02}}\left(1+\psi\right)<3$$

6. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right] < MSE\left[\hat{F}_{Y_{SI}}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\lambda\left(\lambda\xi_{02}-\xi_{11}\right)$$

7. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]<MSE\left[\hat{F}_{UPS}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\delta\left(\delta\xi_{02}-2\xi_{11}\right)$$

8. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]<MSE\left[\hat{F}_{TS}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\gamma\left(\gamma\xi_{02}-2\xi_{11}\right)$$

9. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]<MSE\left[\hat{F}_{SE}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\phi\left(\phi\xi_{02}-2\xi_{11}\right)$$

10. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]<MSE\left[\hat{F}_{JK}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\tau\left(\tau\xi_{02}-2\xi_{11}\right)$$

11. If;

$$MSE_{\min}\left[\hat{F}_{Y_{AD}}\left(t_y\right)\right]<MSE\left[\hat{F}_{GY}\left(t_y\right)\right]$$

$$\xi_{02}\left(1+\psi^{2}-2\psi\right)+8\xi_{11}\left(1-\psi\right)<16\eta\left(\eta\xi_{02}-2\xi_{11}\right)$$

Where $\lambda=\dfrac{F_X\left(t_x\right)}{F_X\left(t_x\right)+C_{F_X\left(t_x\right)}}$, $\delta=\dfrac{F_X\left(t_x\right)C_{F_X\left(t_x\right)}}{F_X\left(t_x\right)C_{F_X\left(t_x\right)}+\beta_2}$, $\gamma=\dfrac{F_X\left(t_x\right)}{F_X\left(t_x\right)+R_{F_Y\left(t_y\right)F_X\left(t_x\right)}}$,

$$\phi=\frac{F_X\left(t_x\right)}{F_X\left(t_x\right)+\beta_2}$$

$$\tau=\frac{F_X\left(t_x\right)+n}{F_X\left(t_x\right)+n},\ \eta=\frac{F_X\left(t_x\right)+n}{F_X\left(t_x\right)+n}$$

**Empirical study**

In this section, an empirical evaluation is presented to evaluate the performance of the suggested estimator and some of the existing estimators by using five different populations.

**Population I**

[Source:Singh 2003]

Y: Duration of sleep (in minutes) and

X: Age of old persons.

**Population II**

[Source: Gujarati2009]

Y: The eggs produced in 1990 (millions) and

X: The price per dozen (cents) in 1990

**Population III**

[Source: Murthy, 1967]

Y: The output of the factory and

X: The number of workers

**Population IV**

[Source: Sarndal, 1992a]

Y: Population in 1983 (in million) and

X: Population in 1980 (in million)

**Population V**

[Source:Koyuncuand Kadilar, 2009]

Y: Number of teachers and

X: Number of students

**Table 1: Summary statistics for Population I-V**

| Parameters | Population I | Population II | Population III | Population IV | Population V |
|---|---|---|---|---|---|
| $N$ | 30 | 50 | 80 | 120 | 923 |
| $n$ | 5 | 5 | 10 | 20 | 180 |
| $\lambda$ | 0.16667 | 0.18 | 0.0875 | 0.04167 | 0.00447 |
| $Q_{2(y)}$ | 387 | 831 | 5105 | 9.25 | 171 |
| $Q_{2(x)}$ | 66.5 | 75.35 | 148 | 8.6 | 4123 |
| $F_Y(t_y)$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.50163 |
| $C_{F_Y(t_y)}$ | 1.01709 | 1.01015 | 1.00631 | 1.00419 | 0.99729 |
| $F_X(t_x)$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.50054 |
| $C_{F_X(t_x)}$ | 1.01709 | 1.01015 | 1.00631 | 1.00419 | 0.99946 |
| $R_{F_Y(t_y)F_X(t_x)}$ | -0.73333 | -0.12000 | 0.95 | 0.96667 | 0.84616 |
| $\beta_1$ | 1 | 1 | 1 | 1 | 1 |

**Table 2: MSE for Population I-V**

| Estimator | Population I | Population II | Population III | Population IV | Population V |
|---|---|---|---|---|---|
| $\hat{F}_{Y_{usual}}(t_y)$ | 0.0431039 | 0.0459181 | 0.0221519 | 0.0105050 | 0.0011187 |
| $\hat{F}_{Y_c}(t_y)$ | 0.1494264 | 0.1028566 | 0.0022152 | 0.0007003 | 0.0003450 |
| $\hat{F}_{Y_m}(t_y)$ | 0.0229890 | 0.0808159 | 0.0863925 | 0.0413197 | 0.0041396 |
| $\hat{F}_{Y_{BTR}}(t_y)$ | 0.0222701 | 0.0518875 | 0.0487343 | 0.0232861 | 0.0023483 |
| $\hat{F}_{Y_{BTP}}(t_y)$ | 0.0854892 | 0.0629079 | 0.0066456 | 0.0029764 | 0.0004509 |
| $\hat{F}_{Y_{SI}}(t_y)$ | 0.3459963 | 0.2536123 | 0.0268767 | 0.0119967 | 0.0018170 |
| $\hat{F}_{Y_{ups}}(t_y)$ | 0.3501308 | 0.2556228 | 0.0271748 | 0.0120886 | 0.0018156 |
| $\hat{F}_{Y_{st}}(t_y)$ | 0.0630276 | 0.0911751 | 0.0243117 | 0.0111936 | 0.0014454 |
| $\hat{F}_{Y_{se}}(t_y)$ | 0.3419568 | 0.2516314 | 0.0265823 | 0.0119055 | 0.0018184 |
| $\hat{F}_{Y_{jk}}(t_y)$ | 1.9741553 | 1.7650936 | 2.2395605 | 4.2166982 | 36.4674200 |
| $\hat{F}_{Y_{gy}}(t_y)$ | 1.916062 | 1.7070734 | 2.2219939 | 4.2092540 | 36.4666900 |
| $\hat{F}_{Y_{AD}}(t_y)$ | **0.0199238** | **0.0452569** | **0.0021598** | **0.0006886** | **0.0003170** |

**Table 3: Percentage Relative Efficiency (PRC) for Population I-V**

| Estimators | Population I | Population II | Population III | Population IV | Population V |
|---|---|---|---|---|---|
| $\hat{F}_{Y_{usual}}(t_y)$ | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $\hat{F}_{Y_c}(t_y)$ | 28.85 | 44.64 | 1000.00 | 1499.96 | 3044.93 |
| $\hat{F}_{Y_m}(t_y)$ | 187.50 | 56.82 | 25.64 | 25.42 | 253.77 |
| $\hat{F}_{Y_{BTR}}(t_y)$ | 193.55 | 88.50 | 45.45 | 45.11 | 447.37 |
| $\hat{F}_{Y_{BTP}}(t_y)$ | 50.42 | 72.99 | 333.33 | 352.94 | 2329.78 |
| $\hat{F}_{Y_{SI}}(t_y)$ | 12.46 | 18.11 | 82.42 | 87.57 | 578.60 |
| $\hat{F}_{Y_{ups}}(t_y)$ | 12.31 | 17.96 | 81.52 | 86.91 | 578.60 |
| $\hat{F}_{Y_{st}}(t_y)$ | 68.39 | 50.36 | 91.12 | 93.85 | 726.79 |
| $\hat{F}_{Y_{se}}(t_y)$ | 12.61 | 18.25 | 83.33 | 88.24 | 577.71 |
| $\hat{F}_{Y_{jk}}(t_y)$ | 2.18 | 2.60 | 0.99 | 0.25 | 0.03 |
| $\hat{F}_{Y_{gy}}(t_y)$ | 2.25 | 2.69 | 1.00 | 0.25 | 0.03 |
| $\hat{F}_{Y_{AD}}(t_y)$ | **216.34** | **101.46** | **1025.65** | **1525.56** | **3306.58** |

From the numerical results presented in Tables 2 and 3 show that the proposed estimator is more efficient than the reviewed existing estimators of CDF.

## Conclusion

This study proposed an efficient non-linear estimator for the estimation of a finite population CDF using simple random sampling and from theoretical and empirical comparisons, the proposed estimator was found to perform better, based on large PRE and smaller MSE. Therefore, our study suggested the use of the proposed estimator for estimating the CDF in practice than the existing estimators in the literature.

## Reference

Abbas, M., &Haq, A. (2022). Estimation of finite population distribution function with auxiliary information in a complex survey sampling. *SORT-Statistics and Operations Research Transactions*, 67-94.

Ahmad, S., Hussain, S., & Ahmad, S. (2021). Finite population distribution function estimation using auxiliary information under simple random sampling.*Statistics, Computing and Interdisciplinary Research*, *3*(1), 29-38.

Ahmad, S., Aamir, M., Hussain, S., Shabbir, J., Zahid, E., Subkrajang, K., &Jirawattanapanit, A. (2022). A new generalized class of exponential factor-type estimators for population distribution function using two auxiliary variables. *Mathematical Problems in Engineering*, *2022*.

Ahmad, S., Ullah, K., Zahid, E., Shabbir, J., Aamir, M., Alshanbari, H. M., & El-Bagoury, A. A. A. H. (2023). A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. *Scientific Reports*, *13*(1), 5415.

Berger, Y. G., & Munoz, J. F. (2015).On estimating quantiles using auxiliary information.*Journal of Official Statistics*, *31*(1), 101-119.

Chambers, R. L., & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, *73*(3), 597-604.

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The journal of agricultural science*, *30*(2), 262-275.

Gupta, R. K. and Yadav, S. K. (2018). Improved Estimation of Population Mean Using Information on Size of the Sample. *American Journal of Mathematics and Statistics*.8(2): 27-35.

Jerajuddin, M. and Kishun, J. (2016). Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected From Population, *IJSRSET*. (2)2:10-16.

Hussain, S., Akhtar, S., & El-Morshedy, M. (2022). Modified estimators of finite population distribution function based on dual use of auxiliary information under stratified random sampling. *Science Progress*, *105*(3), 00368504221128486.

Hussain, S., Ahmad, S., Akhtar, S., Javed, A., & Yasmeen, U. (2020). Estimation of finite population distribution function with dual use of auxiliary information under non-response. *Plos one*, *15*(12), e0243584.

Martínez, S., Rueda, M., Arcos, A., &Martínez, H. (2010). Optimum calibration points estimating distribution functions. *Journal of computational and applied mathematics*, *233*(9), 2265-2277.

Prasad, N. N., & Rao, J. N. (1990). The estimation of the mean squared error of small-area estimators.*Journal of the American statistical association*, *85*(409), 163-171.

Singh, G. N., & Usman, M. (2021). Enhanced estimation of population distribution function in the presence of non-response.*Ain Shams Engineering Journal*, *12*(3), 3109-3119.

Singh, H.P., Tailor, R. and Kakran, M. S. (2004). An Improved Estimation of Population Mean using Power Transformation. *Journal of the Indian Society of Agricultural Statistics*, 8(2): 223–230.

Singh, H.P. and Tailor, R. (2003). Use of known Correlation Coefficient in Estimating the Finite Population Means; *Statistics in Transition* 6 (4), 555-560.

Sisodia, B.V.S. and Dwivedi, V.K. (1981): A Modified Ratio Estimator using Coefficient of Variation of Auxiliary Variable; *Jour. of Indian. Soc. of Agri. Stat.*, 33(1).Pp. 13–18

Upadhyaya, L.N. and Singh, H.P. (1999): Use of Transformed Auxiliary Variable in Estimating the Finite Population Means, *Biometrical Journal*.41 (5), 627–636.

Yaqub, M., & Shabbir, J. (2020). Estimation of population distribution function involving measurement error in the presence of non-response.*Communications in Statistics-Theory and Methods*, *49* (10), 2540-2559.

Zakari, Y., Muili, J. O., Tela, M. N., Danchadi, N. S., &Audu, A. (2020).Use of unknown weight to enhance ratio-type estimator in simple random sampling.*Lapai Journal of Applied and Natural Sciences*, *5*(1),74-81.