Development of Efficient Estimation Technique for Population Mean in Two Phase Sampling Using Fuzzy Tools

P. PARICHHA, K. BASU, A. BANDYOPADHYAY
AND P. MUKHOPADHYAY

Abstract

The present investigation deals with the problem of estimation of population mean in two-phase (double) sampling. Utilizing information on two auxiliary variables, one chain exponential ratio and regression type estimator has been proposed and its properties are studied under two different structures of two-phase sampling. To make the estimator practicable, unbiased version of the proposed strategy has also been developed. The dominance of the suggested estimator over some contemporary estimators of population mean has been established through numerical illustrations carried over the data set of some natural population and artificially generated population. Categorization of the dominance ranges of the proposed estimation strategies are deployed through defuzzification tools, which are followed by suitable recommendations.

Mathematics Subject Classification 2000: 62D05
Keywords: Double sampling, study variable, auxiliary variable, chain-type, regression, bias, fuzzy applications, variance, efficiency

1. INTRODUCTION

Information on variables correlated with the study variable is popularly known as auxiliary information. The use of supplementary information on auxiliary variable for estimating the finite population mean of the variable under study has played an eminent role in sampling theory and practices. Auxiliary information may be truthfully utilized at planning, design and estimation stages to develop improved estimation procedures in sample surveys. Use of auxiliary information at estimation stage was introduced during the 1940's with a comprehensive theory provided by Cochran. Sometimes, information on auxiliary variable may be readily available for all the units of population; for example, tonnage (or seat capacity) of each vehicle or ship is known in survey sampling of transportation and number of beds available in different hospitals may be known well in advance in health care surveys. If such

10.1515/jamsi-2017-0006

information lacks, it is sometimes, relatively cheap to take a large preliminary sample where auxiliary variable alone is measured, such practice is applicable in two-phase (or double) sampling. Two-phase sampling happens to be a powerful and cost effective (economical) technique for obtaining the reliable estimate in first-phase (preliminary) sample for the unknown parameters of the auxiliary variables. For example, Sukhatme (1962) mentioned that in a survey to estimate the production of lime crop based on orchards as sampling units, a comparatively larger sample is drawn to determine the acreage under the crop while the yield rate is determined from a sub sample of the orchards selected for determining acreage.

In order to construct an efficient estimator of the population mean of the auxiliary variable in first-phase (preliminary) sample, Chand (1975) introduced a technique of chaining another auxiliary variable with the first auxiliary variable by using the ratio estimator in the first phase sample. The estimator is known as chain-type ratio estimator. This work was further extended by Kiregyera (1980, 1984), Sahoo *et al.* (1994), Tracy *et al.* (1996), Singh and Espejo (2007), Gupta and Shabbir (2007), Shukla *et al.* (2012), Choudhury and Singh (2012) and among others where they proposed various chain-type ratio and regression estimators. It may be noted that the most of these estimation procedures of population mean in two-phase sampling are biased which become a serious drawback for their practical applications. Apart from the bias estimates, it may also be observed that dominance ranges of the recent developed ones over the conventional ones are not clearly mentioned. Simulation study carried over the data set of natural and artificially generated population have been utilized to obtain the trend of efficacy of these recent developed strategies. The dominance conditions of the newly developed estimators are very essential for their recommendations to the real life problems. Motivated with this argument, only Chatterjee *et al.* (2015) has given a new direction to find the dominance range of the estimation strategy through fuzzy tools in successive sampling and it may be noted that no attempt has been made yet to find the dominance ranges of the estimators in two phase sampling scheme for the estimation of population mean in sample surveys.

Encouraged and fascinated with the work discussed earlier, we have proposed chain exponential ratio and regression type estimator of population mean and studied its properties under two different structures of two-phase sampling. Considering the realistic situations, we have developed unbiased version of the proposed estimators. Performances of the proposed estimator have been examined through theoretical and numerical illustrations, which presents the effectiveness of the proposed strategies. To categorize the dominance ranges of the proposed estimation strategic, fuzzification and defuzzification rule are employed. Recommendations of the proposed estimation strategy have been put forward to the survey statisticians.

2. FORMULATION OF THE CLASS OF ESTIMATOR

2.1. Sample Structure and Some Existing Estimation Procedures

Let y_k, x_k and z_k be the values of the study variable y, first auxiliary variable x and second auxiliary variable z respectively associated with the k^{th} unit of the finite population $U = (U_1, U_2, U_3, \ldots, U_N)$. We wish to estimate the population mean $\bar{Y}$ of the study variable y in presence of auxiliary variables x and z, when the population mean $\bar{X}$ of x is unknown but information on z is readily available for all the units of population.

Thus, to estimate $\bar{Y}$, a first phase sample $S'(S' \subset U)$ of size n is drawn by simple random sampling without replacement scheme (SRSWOR) from the entire population U and observed for the auxiliary variables x and z to furnish the estimate of $\bar{X}$. Again a second-phase sample S of size m $(m \leq n)$ is drawn from the first phase sample SRSWOR scheme to observe the study variable y.

Hence onwards, we use the following notations:

$\bar{Z}$: Population mean of the auxiliary variable z.

$\bar{x}', \bar{z}'$: Sample means of the respective variables based on the first phase sample of size n.

$\bar{x}, \bar{y}, \bar{z}$: Sample means of the respective variables based on the second phase sample of size m. $b_{yx}(n)$, $b_{xz}(n)$: Sample regression coefficients between the variables shown in subscripts and based on the sample sizes indicated in the braces.

$s_{yx}(n), s_{xz}(n), s_{yz}(n)$: Sample covariance between the variables shown in subscripts and based on the second phase sample of size n.

s_x^2 : Sample mean square of the variable x based on the second phase sample of size n.

$\beta_{yx}, \beta_{xz}, \beta_{yz}$: Population regression coefficients between the variables shown in subscripts.

To estimate the population mean $\bar{Y}$, the classical ratio estimator is presented as

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}} \bar{X}. \tag{1}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of variables y and x respectively based on the second phase sample S.

If $\bar{X}$ is unknown, we estimate $\bar{Y}$ under two-phase sampling set up as

$$t_1 = \frac{\bar{y}}{\bar{x}} \bar{x}' \tag{2}$$

where $\bar{x}'$ is the sample mean of the auxiliary variable x based on the first-phase (preliminary) sample S'.

Srivastava (1970) generalized the ratio method of estimation and its structure in two-phase sampling is given by

$$t_2 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha}, \tag{3}$$

where α is a real scalar, which can be suitably determined by minimizing the mean square error (M. S. E.) of the estimator t_2.

The way in which the estimate of $\bar{Y}$ is improved using the auxiliary information on x can also be extended to improve the estimate of $\bar{X}$ in the first-phase sample, if another auxiliary variable z closely related to x but remotely related to y is used. Thus, assuming that the population mean of the auxiliary variable z is known, Chand (1975) proposed a chain-type ratio estimator as

$$t_c = \frac{\bar{y}}{\bar{x}} \bar{x}'_{rd} \tag{4}$$

where $\bar{x}'_{rd} = \frac{\bar{x}'}{\bar{z}'}\bar{Z}$, $\bar{z}'$ and $\bar{Z}$ are the sample mean based on the first phase sample of size n and population mean of the auxiliary variable z respectively.

Singh and Espejo (2007) considered a ratio - product type estimator in double sampling as

$$t_3 = \bar{y}\left[p\frac{\bar{x}'}{\bar{x}} + (1-p)\frac{\bar{x}}{\bar{x}'}\right] \tag{5}$$

where p is a real scalar which may be suitably determined to minimize the mean square error of the estimator t_4.

Singh and Vishwakarma (2007) constructed exponential ratio and product type estimator of population mean in two phase sampling as

$$t_4 = \bar{y}\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) \tag{6}$$

and

$$t_5 = \bar{y}\exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \tag{7}$$

respectively.

2.2. Proposed Class of Estimator

Motivated with the earlier work, we have defined a class of chain exponential ratio and regression type estimators as

$$t_p = \bar{y}\left\{(1-k)\frac{\bar{x}'}{\bar{x}} + k\exp\left(\frac{\bar{x}'_{ld} - \bar{x}}{\bar{x}'_{ld} + \bar{x}}\right)\right\} \tag{8}$$

where k is a real constant which can be suitably determined by minimizing the M. S. E. of the class of estimators t_p and $\bar{x}'_{id} = \bar{x}' + b_{xz}(n)\left(\bar{Z} - \bar{z}'\right)$.

3. FORMULATION OF THE CLASS OF ESTIMATOR … BIAS AND MEAN SQUARE ERRORS OF THE PROPOSED CLASS OF ESTIMATOR t_p

It can be easily noted that the proposed class of estimators t_p defined in equations (9) is chain exponential ratio and regression type estimator. Therefore, it is biased for population mean $\bar{Y}$. Hence, we proceed to obtain their biases and mean square errors under large sample approximations using the following transformations:

$\bar{y}=\bar{Y}(1+e_1)$, $\bar{x}=\bar{X}(1+e_2)$, $\bar{x}'=\bar{X}(1+e_3)$, $\bar{z}'=\bar{Z}(1+e_4)$,
$s'_{xz}=S_{xz}(1+e_5)$, $s^2_{z'}=S^2_z(1+e_6)$

where $E(e_i)=0$ for (i = 1, 2, . . ., 6), e_i for (i = 1, 2, . . ., 6) are relative error term.

Under above transformations the class of estimators t_p can be represented as

$$t_p=\bar{Y}(1+e_1)\left[(1-k)\left\{(1+e_3)(1+e_2)^{-1}\right\}+\frac{k}{2}\exp\left\{\left((e_3-e_2)-\frac{\bar{Z}}{\bar{X}}\beta_{xz}(e_4+e_4e_5-e_4e_6)\right)\left(1+\frac{e_2+e_3}{2}-\frac{\bar{Z}}{\bar{X}}\beta_{xz}(e_4+e_4e_5-e_4e_6)\right)^{-1}\right\}\right] \quad (9)$$

We have derived the expressions for bias and mean square error of the proposed class of estimators t_p and presented them below.

We have the following expectations of the sample statistics under two - phase sampling set up as

$$\left.\begin{array}{l}
E(e_1^2)= f_1C_y^2,\ E(e_2^2) = f_1C_x^2,\ E(e_3^2) = f_2C_x^2,\ E(e_4^2) = f_2C_z^2 \\
E(e_1e_2)= f_1\rho_{yx}C_yC_x,\ E(e_1e_3) = f_2\rho_{yx}C_yC_x, \\
E(e_2e_3)= f_2C_x^2,\ E(e_2e_4) = E(e_3e_4)= f_2\rho_{xz}C_xC_z, \\
E(e_4e_5) = f_2\ \dfrac{\mu_{102}}{\bar{Z}S_{xz}},\ E(e_4e_6) = f_2\ \dfrac{\mu_{003}}{\bar{Z}S_z^2}, \\
E(e_2e_5) = f_2\ \dfrac{\mu_{201}}{\bar{X}S_{xz}},\ E(e_2e_6) = f_2\ \dfrac{\mu_{102}}{\bar{X}S_z^2}, \\
E(e_1e_4)= f_2\rho_{yz}C_yC_z.
\end{array}\right\} \tag{10}$$

where

$$f_1 = \frac{1}{n} - \frac{1}{N},\ f_2 = \frac{1}{m} - \frac{1}{N},\ f_3 = \frac{1}{n} - \frac{1}{m},$$

$$\mu_{pqr} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^p(y_i - \bar{Y})^q(z_i - \bar{Z})^r;\ (p, q, r \geq 0)$$

Expanding binomially, exponentially, using results from equation (9) and retaining the terms up to first order of sample size, we have derived the expressions of bias B(.) and mean square error M(.) of the class of estimator t_p as

$$B(t_p)= E\left(t_p-\bar{Y}\right)= \bar{Y}\left[kf_2\left(\frac{S_{xz}^2}{\bar{X}^2S_z^2} - \frac{\mu_{102}}{\bar{X}S_z^2} + \frac{\mu_{003}}{\bar{X}S_z^4} - \frac{S_{yz}}{\bar{Y}\bar{X}}\right)+f_3\left(\frac{S_x^2}{\bar{X}^2}-\frac{S_{yx}}{\bar{Y}\bar{X}}\right)\right] \tag{11}$$

$$M\left(t_p\right)=E\left[t_p-\bar{Y}\right]^2 =\bar{Y}^2\left[f_1C_y^2+ (k^2/4)a+kb +c\right] \tag{12}$$

where

$$a = \left(f_3+ f_2\rho_{xz}^2\right)C_x^2 \quad \text{and} \quad b = -f_2\rho_{yz}\rho_{xz}C_yC_x + f_3\rho_{yx}C_yC_x - f_3C_x^2\ ,\ c = f_3C_x^2 - 2\, f_3\rho_{yx}C_yC_x$$

4. BIAS REDUCTION FOR THE PROPOSED CLASS OF ESTIMATOR

In some situations of practical importance, bias becomes a serious drawback. Therefore, unbiased versions of the proposed classes of estimators are more desirable. Motivated with this argument and influenced by the bias correction techniques of Tracy *et al.* (1996) and Bandyopadhyay and Singh (2014) we proceed to derive the unbiased version of our proposed class of estimator t_p.

From equation (12), we observe that the expression of bias of the estimator t_p contains the population parameters such as μ_{003}, μ_{102}, S_{yx}, S_{yz}, S_x^2, S_y^2, $\bar{Y}$, $\bar{X}$, S_{yz} and S_z^2. Since S_z^2 is known while μ_{003}, μ_{102}, S_{yx}, S_{yz}, S_x^2, S_y^2, $\bar{Y}$, $\bar{X}$ and S_{yz} are unknown, replacing μ_{003}, μ_{102}, S_{yx}, S_{yz}, S_x^2, S_y^2, $\bar{Y}$, $\bar{X}$ and S_{yz} by their respective sample estimator (based on the second phase sample of size m) m_{003}, m_{102}, s_{yz}, s_x^2, s_y^2, $\bar{y}$, $\bar{x}$ and s_{yz}, we get an estimator of $B\left(t_p\right)$ as

$$b(t_p) = \bar{y}\,k\left[f_2\left(\frac{s_{xz}^2}{\bar{x}^2 s_z^2} - \frac{m_{102}}{\bar{x}s_z^2} - \frac{m_{003}}{\bar{x}s_z^4} - \frac{s_{yz}}{\overline{yx}}\right) + f_3\left(\frac{s_x^2}{\bar{x}^2} - \frac{s_{yx}}{\overline{yx}}\right)\right] \tag{13}$$

where $m_{pqr} = \frac{1}{m}\sum_{i=1}^{m}\left(x_i - \bar{x}_m\right)^p\left(y_i - \bar{y}_m\right)^q\left(z_i - \bar{z}_m\right)^r$.

Now motivating with the bias reduction techniques of Tracy *et al.* (1996) and Bandyopadhyay and Singh (2014), we have derived the unbiased version of the proposed class of estimator t_p to the first order of approximations as

$$t'_p = t_p - b(t_p)$$

which becomes

$$t_p' = \bar{y}\left\{(1-k)\frac{\bar{x}'}{\bar{x}} + k\exp\left(\frac{\bar{x}'_{ld} - \bar{x}}{\bar{x}'_{ld} + \bar{x}}\right)\right\} - \bar{y}\,k\left\{f_2\left(\frac{s_{xz}^2}{\bar{x}^2 s_z^2} - \frac{m_{102}}{\bar{x}s_z^2} - \frac{m_{003}}{\bar{x}s_z^4} - \frac{s_{yz}}{\overline{yx}}\right) + f_3\left(\frac{s_x^2}{\bar{x}^2} - \frac{s_{yx}}{\overline{yx}}\right)\right\} \tag{14}$$

Thus, the variance of t_p' to the first order of approximation are obtained as

$$V\left(t_p'\right) = M\left(t_p\right) = \bar{Y}^2\left[f_1 C_y^2 + (k^2/4)a + kb + c\right] \tag{15}$$

Thus, from equations and (12) and (15) it is to be noted that the class of estimators t_p' is preferable over the class of estimator t_p as t_p' is unbiased (up to first order of sample size) class of estimators of $\bar{Y}$ while the class of estimator t_p is biased.

5. MINIMUM VARIANCE OF PROPOSED CLASS OF ESTIMATOR

It is obvious from the equation (15) that the variance of the proposed class of estimator t'_p depends on the value of the constant k. Therefore, we desire to minimize their variances and discussed them below.

The optimality condition under which proposed class of estimators t'_p have minimum variance is obtained as

$$k = -2b/a \tag{16}$$

where

Substituting the optimum value of the constant k in equation (15), we have the minimum variance of the class of estimator t'_p as

$$\text{Min. } V(t'_p) = \bar{Y}^2\left[f_1 C_y^2 - \frac{b^2}{a} + C\right] \tag{17}$$

REMARK 5.1: It is to mentioned that the optimum value of k depends on unknown population parameters such as C_x, C_y, C_z, ρ_{yx} and ρ_{xz}. Thus, to make the class of estimators practicable, these unknown population parameters may be estimated with their respective sample estimates or from past data or guessed from experience gathered over time. Such problems are also considered by Reddy (1978), Tracy *et al.* (1996) and Singh *et al.* (2007).

6. EFFICIENCY COMPARISONS OF THE PROPOSED CLASS OF ESTIMATOR t'_p

To examine the performances of the proposed class of estimators under two different cases of two - phase sampling set up as suggested in this paper, we have compared their efficiencies with some existing estimators of population mean such as $\bar{y}$ (sample mean estimator) and $t_i\,(i = 1, 2, \ldots, 5)$. The Variance/ M. S. E.s/ minimum M. S. E.s of the existing estimators t_i are obtained up to the first order of approximations under the Cases of the two phase - sampling set up and presented below.

$$M(t_1) = \bar{Y}^2\left[f_1C_y^2 + f_2C_x^2 - 2f_2\rho_{yx}C_yC_x\right]$$

$$\text{Min. } M(t_2) = S_y^2\left[f_1 - f_3\rho_{yx}^2\right]$$

$$\text{Min. } M(t_3) = \left[f_1\left(1 - \rho_{yx}^2\right) + f_2\rho_{yx}^2\right]S_y^2$$

$$M(t_4) = \bar{Y}^2\left[f_1C_y^2 + \frac{f_3}{4}C_x^2 - f_3\rho_{yx}C_yC_x\right]$$

$$M(t_5) = \bar{Y}^2\left[f_1C_y^2 + \frac{f_3}{4}C_x^2 + f_3\rho_{yx}C_yC_x\right]$$

We have demonstrated the superiority of the suggested estimator over the estimator t_i $(i = 1, 2, \ldots, 5)$ through numerical illustration and graphical interpretations.

6.1. Empirical Investigations through Natural Population

We have chosen four natural populations to illustrate the efficacious performance of our proposed classes of estimators. The source of the populations, the nature of the variables y, x, z and the values of the various parameters are as follows.

Population I - Source: Cochran (1977)

y: Number of 'placebo' children.

$N = 34$, $n = 15$, $m = 10$, $\bar{Y} = 4.92$, $C_y = 1.0123$, $C_x = 1.2318$, $C_z = 1.0720$, $\rho_{yx} = 0.7326$, $\rho_{yz} = 0.6430$ and $\rho_{xz} = 0.6837$.

Population II - Source: Shukla (1966)

y: Measurement of weight of children.

$N = 50$, $n = 15$, $m = 8$, $\bar{Y} = 2.584$, $C_y = 0.2943$, $C_x = 0.3410$, $C_z = 0.13038$, $\rho_{yx} = 0.48$, $\rho_{yz} = 0.37$ and $\rho_{xz} = 0.73$.

Population III - Source: Handique et al. (2011)

y: Forest timber volume in cubic meter (Cum) in 0.1 ha sample plot.

$N = 2500$, $n = 700$, $m = 80$, $\bar{Y} = 4.63$, $C_y = 0.95$, $C_x = 0.98$, $C_z = 0.64$, $\rho_{yx} = 0.79$, $\rho_{yz} = 0.72$ and $\rho_{xz} = 0.66$.

Population IV - Source: Sukhatme and Sukhatme (1970)

y: Area (acres) under wheat in 1937.

$N = 34$, $n = 10$, $m = 7$, $\bar{Y} = 201.41$, $C_y = 0.74$, $C_x = 0.76$, $C_z = 0.61$,
$\rho_{yx} = 0.93$, $\rho_{yz} = 0.9$ and $\rho_{xz} = 0.83$.

To have a tangible idea about the performance of the proposed class of estimator t'_p we have computed percent relative efficiencies (PREs) of them and the existing estimators t_i (i = 1, 2, . . ., 5) under similar realistic situations and the findings are displayed in Table 1 where PREs are designated as $PRE = \frac{V(\bar{y})}{V(T)} \times 100$ and $M(T)$ denote variance/minimum M. S. E. of an estimator T.

Table 1: PRE of various estimators

Estimator	Population -I	Population -II	Population -III	Population -IV
t_1	118.9748	90.7093	105.0521	213.8133
t_2	133.9482	114.6789	233.1262	148.5310
t_3	133.9482	114.6789	233.1262	148.5310
t_4	132.6546	113.9611	200.9033	135.3565
t_5	62.6649	66.8650	50.2741	68.4722
t_p'	177.9129	122.5387	243.5714	273.9723

6.2. Empirical Investigations through Artificially Generated Population

An important aspect of simulation is that one builds a simulation model to replicate the actual system. Simulation allows comparison of analytical techniques and helps in concluding whether a newly developed technique is better than the existing ones. Motivated by Singh and Deo (2003) and Singh *et al.*(2001) who have been adopted the artificial population generation techniques, we have generated five sets of independent random samples of size N (N = 100) namely x'_{1_k}, y'_{1_k}, x'_{2_k}, y'_{2_k} and z'_k (k =1, 2, 3, . . ., N) from a standard normal distribution with the help of R-software. By varying the correlation coefficients ρ_{yx} and ρ_{xz} , we have generated the following transformed variables of the population U with the values of $\sigma_y^2 = 50,\ \mu_y = 40,\ \sigma_x^2 = 25,\ \mu_x = 50,\ \sigma_z^2 = 9$ and $\mu_z = 30$ as

$$y_{1_k} = \mu_y + \sigma_y \left[\rho_{xy} x'_{1_k} + \left(\sqrt{1-\rho_{yx}^2} \right) y'_{1_k} \right], x_{1_k} = \mu_x + \sigma_x x'_{1_k} \text{ and } z_k = \mu_z + \sigma_z \left[\rho_{xz} x'_{1_k} + \left(\sqrt{1-\rho_{xz}^2} \right) z'_k \right].$$

Thus, we have derived following efficiency comparisons of our proposed strategy with the recent relevant ones with the above artificially generated population techniques as:

Table 2: PRE of various estimators

Estimators	PRE		
	$\rho_{yx}=0.7$, $\rho_{xz}=0.6$	$\rho_{yx}=0.9$, $\rho_{xz}=0.6$	$\rho_{yx}=0.7$, $\rho_{xz}=0.8$
t_1	120.9381	149.1971	109.6664
t_2	136.8180	172.3255	125.5277
t_3	133.3229	167.4536	102.8960
t_4	136.8180	172.3255	125.5277
t_5	133.0126	151.7422	124.6970
t_6	65.4381	62.0817	67.3860
t_p'	151.3304	188.2855	139.7335

7. ANALYSIS OF EMPIRICAL STUDY THROUGH FUZZY TOOLS

From the above empirical study, it is to be noted that the dominance of the proposed strategy over the existing ones have been established through data set of real life population and artificially generated population. However, the dominance conditions of the proposed strategy over the existing ones have not yet been obtained clearly. Therefore, we desire to derive the conditions where proposed estimators performs extremely well or dominate mildly the sample mean estimator $\bar{y}$. This investigation helps us in choosing the suitable population where our proposed work may be applied effectively which is very essential for the recommendations of our proposed work for their practical application. Motivated with this argument, we proceed to build up a decision making machinery through fuzzy tools which will enable us to measure the degree of efficiency of the estimator for different choices of correlations ρ_{yx}, ρ_{xz} and ρ_{yz}. Thus, we have computed the PRE of the proposed class estimators t'_p (under its respective optimum condition as discussed in section 5) with respect to the sample mean estimator $\bar{y}$ and presented them in Table 3.

REMARK: 7.1. Here we have considered the stability nature of the coefficient of variations of the study variable and auxiliary variables (Reddy 1978) and thus we have taken the coefficient of variations of y, x and z to be approximately equal.

Table 3: PRE of the proposed class of estimators

ρ_{yx} = 0.2 (fix.)		PRE	ρ_{yx} = 0.3 (fix.)		PRE	ρ_{yx} = 0.4 (fix.)		PRE	ρ_{yx} = 0.5 (fix.)		PRE
ρ_{xz}	ρ_{yz}		ρ_{xz}	ρ_{yz}		ρ_{xz}	ρ_{yz}		ρ_{xz}	ρ_{yz}	
0.4	0.3	106.62	0.4	0.3	109.87	0.4	0.3	114.61	0.4	0.3	121.20
	0.4	109.99		0.4	113.05		0.4	117.64		0.4	124.11
	0.5	113.71		0.5	116.56		0.5	120.97		0.5	127.31
	0.6	117.83		0.6	120.43		0.6	124.66		0.6	130.86
	0.7	122.41		0.7	124.73		0.7	128.75		0.7	134.80
	0.8	127.53		0.8	129.52		0.8	133.29		0.8	139.18
	0.9	133.28		0.9	134.87		0.9	138.36		0.9	144.06
0.5	0.3	106.22	0.5	0.3	109.78	0.5	0.3	114.80	0.5	0.3	121.65
	0.4	110.29		0.4	113.66		0.4	118.53		0.4	125.26
	0.5	114.89		0.5	118.03		0.5	122.73		0.5	129.34
	0.6	120.11		0.6	122.98		0.6	127.48		0.6	133.96
	0.7	126.05		0.7	128.60		0.7	132.87		0.7	139.21
	0.8	132.87		0.8	135.03		0.8	139.03		0.8	145.20
	0.9	140.77		0.9	142.44		0.9	146.10		0.9	152.07
0.6	0.3	105.38	0.6	0.3	109.29	0.6	0.3	114.63	0.6	0.3	121.79
	0.4	110.03		0.4	113.76		0.4	118.97		0.4	126.03
	0.5	115.37		0.5	118.89		0.5	123.95		0.5	130.92
	0.6	121.54		0.6	124.82		0.6	129.70		0.6	136.58
	0.7	128.75		0.7	131.70		0.7	136.38		0.7	143.15
	0.8	137.24		0.8	139.79		0.8	144.20		0.8	150.84
	0.9	147.35		0.9	149.36		0.9	153.44		0.9	159.91
0.7	0.3	104.24	0.7	0.3	108.52	0.7	0.3	114.19	0.7	0.3	121.67
	0.4	109.31		0.4	113.45		0.4	119.03		0.4	126.45
	0.5	115.23		0.5	119.20		0.5	124.68		0.5	132.07
	0.6	122.21		0.6	125.97		0.6	131.33		0.6	138.68
	0.7	130.50		0.7	133.99		0.7	139.20		0.7	146.53
	0.8	140.50		0.8	143.63		0.8	148.64		0.8	155.93
	0.9	152.74		0.9	155.36		0.9	160.08		0.9	167.31
0.8	0.3	102.91	0.8	0.3	107.54	0.8	0.3	113.55	0.8	0.3	121.35
	0.4	108.27		0.4	112.81		0.4	118.78		0.4	126.58
	0.5	114.62		0.5	119.06		0.5	124.98		0.5	132.81
	0.6	122.21		0.6	126.51		0.6	132.39		0.6	140.28
	0.7	131.40		0.7	135.50		0.7	141.33		0.7	149.30
	0.8	142.68		0.8	146.52		0.8	152.26		0.8	160.33
	0.9	156.83		0.9	160.26		0.9	165.85		0.9	174.03
0.9	0.3	101.49	0.9	0.3	106.45	0.9	0.3	112.78	0.9	0.3	120.89
	0.4	107.03		0.4	111.96		0.4	118.30		0.4	126.47
	0.5	113.67		0.5	118.56		0.5	124.93		0.5	133.21
	0.6	121.70		0.6	126.54		0.6	132.96		0.6	141.41
	0.7	131.56		0.7	136.32		0.7	142.81		0.7	151.47
	0.8	143.87		0.8	148.51		0.8	155.06		0.8	164.01
	0.9	159.61		0.9	164.03		0.9	170.63		0.9	179.95

Table 3 continued . . .

ρ_{yx} = 0.6 (fix.)		PRE	ρ_{yx} = 0.7 (fix.)		PRE	ρ_{yx} = 0.8 (fix.)		PRE	ρ_{xz} = 0.9 (fix.)		PRE
ρ_{xz}	ρ_{yz}		ρ_{xz}	ρ_{yz}		ρ_{xz}	ρ_{yz}		ρ_{yx}	ρ_{yz}	
0.4	0.3	130.26	0.4	0.3	142.75	0.4	0.3	160.38	0.4	0.3	186.35
	0.4	133.05		0.4	145.44		0.4	162.95		0.4	188.70
	0.5	136.16		0.5	148.46		0.5	165.88		0.5	191.47
	0.6	139.61		0.6	151.85		0.6	169.21		0.6	194.72
	0.7	143.45		0.7	155.63		0.7	172.97		0.7	198.47
	0.8	147.72		0.8	159.87		0.8	177.22		0.8	202.78
	0.9	152.49		0.9	164.61		0.9	182.02		0.9	207.72
0.5	0.3	130.94	0.5	0.3	143.66	0.5	0.3	161.48	0.5	0.3	187.57
	0.4	134.46		0.4	147.10		0.4	164.84		0.4	190.75
	0.5	138.46		0.5	151.05		0.5	168.75		0.5	194.59
	0.6	143.01		0.6	155.57		0.6	173.30		0.6	199.19
	0.7	148.19		0.7	160.76		0.7	178.57		0.7	204.62
	0.8	154.11		0.8	166.72		0.8	184.68		0.8	211.02
	0.9	160.91		0.9	173.58		0.9	191.76		0.9	218.54
0.6	0.3	131.36	0.6	0.3	144.35	0.6	0.3	162.44	0.6	0.3	188.72
	0.4	135.55		0.4	148.51		0.4	166.56		0.4	192.76
	0.5	140.41		0.5	153.37		0.5	171.48		0.5	197.75
	0.6	146.04		0.6	159.07		0.6	177.32		0.6	203.82
	0.7	152.62		0.7	165.75		0.7	184.25		0.7	211.17
	0.8	160.32		0.8	173.62		0.8	192.47		0.8	220.02
	0.9	169.40		0.9	182.93		0.9	202.27		0.9	230.69
0.7	0.3	131.56	0.7	0.3	144.85	0.7	0.3	163.23	0.7	0.3	189.79
	0.4	136.34		0.4	149.66		0.4	168.09		0.4	194.68
	0.5	141.98		0.5	155.40		0.5	174.00		0.5	200.83
	0.6	148.66		0.6	162.25		0.6	181.16		0.6	208.47
	0.7	156.61		0.7	170.46		0.7	189.83		0.7	217.89
	0.8	166.14		0.8	180.35		0.8	200.37		0.8	229.51
	0.9	177.70		0.9	192.39		0.9	213.28		0.9	243.93
0.8	0.3	131.56	0.8	0.3	145.17	0.8	0.3	163.88	0.8	0.3	190.76
	0.4	136.84		0.4	150.56		0.4	169.42		0.4	196.47
	0.5	143.18		0.5	157.10		0.5	176.27		0.5	203.77
	0.6	150.82		0.6	165.05		0.6	184.72		0.6	213.00
	0.7	160.09		0.7	174.76		0.7	195.17		0.7	224.61
	0.8	171.44		0.8	186.74		0.8	208.16		0.8	239.27
	0.9	185.57		0.9	201.69		0.9	224.51		0.9	257.97
0.9	0.3	131.41	0.9	0.3	145.34	0.9	0.3	164.39	0.9	0.3	191.61
	0.4	137.11		0.4	151.23		0.4	170.54		0.4	198.10
	0.5	144.06		0.5	158.50		0.5	178.27		0.5	206.52
	0.6	152.55		0.6	167.46		0.6	187.96		0.6	217.32
	0.7	163.03		0.7	178.61		0.7	200.16		0.7	231.17
	0.8	176.13		0.8	192.65		0.8	215.67		0.8	249.06
	0.9	192.81		0.9	210.61		0.9	235.71		0.9	272.52

Development of the Fuzzy Logic Controller (FLC) is accomplished by studying the empirical data furnished in Table 3 where the FLC produces the degree of efficiency for a given range of $\rho_{xy}, \rho_{xz} and \rho_{yz}$. They are taken as to be the three input fuzzy variables having 8 , 6 and 7 linguistics respectively (listed in Tables 4.a , 4.b & 4.c).

Entire range of ρ_{xy} [0.2≤ ρ_{xy} ≤0.9] is divided into 8 equal parts and that for ρ_{xz} [0.4≤ ρ_{xz} ≤0.9] into 6 equal parts and for ρ_{yz} [0.3≤ ρ_{yz} ≤0.9] into 7 equal parts. To each part a linguistic is assigned suitably for all three cases. Also the PRE is taken as the output fuzzy variable having the set of 16 linguistics (listed in Table 5) in the descending degree of efficiency. The range [101 ≤ PRE≤ 276], as it is furnished in Table 5**,** is divided into 16 equal parts as shown in the table.

Table 4.a

Ling	Range	A (mid.pt)	C (half width)
vwc	.15 - .25	0.2	.05
wc	.25 - .35	.3	.05
c1	.35 - .45	.4	.05
c2	.45 - .55	.5	.05
c3	.55 - .65	.6	.05
c4	.65 - .75	.7	.05
sc	.75 - .85	.8	.05
hsc	.85 - .95	.9	.05

Table 4.b

Ling	Range	A (mid.pt.)	C (half width)
Mc	.35 - .45	.4	.05
c1	.45 - .55	.5	.05
c2	.55 - .65	.6	.05
Sc	.65 - .75	.7	.05
Hc	.75 - .85	.8	.05
Vhc	.85 - .95	.9	.05

Table 4.c

Ling	Range	A(mid.pt)	C(half width)
vwc2	.25 - .35	.3	.05
wc2	.35 - .45	.4	.05
c21	.45 - .55	.5	.05
c22	.55 - .65	.6	.05
c23	.65 - .75	.7	.05
c24	.75 - .85	.8	.05
vsc2	.85 - .95	.9	.05

Table 5

Ling	Range	A(mid.pt)	C(half width)
T1	101 – 111	106	5
T2	112 – 122	117	5
T3	123 – 133	128	5
T4	134 – 144	139	5
T5	145 – 155	150	5
T6	156 – 166	161	5
T7	167 – 177	172	5
T8	178 – 188	183	5
T9	189 – 199	194	5
T10	200 – 210	205	5
T11	211 – 221	216	5
T12	222 – 232	227	5
T13	233 – 243	238	5
T14	244 – 254	249	5
T15	255 – 265	260	5
T16	266 - 276	271	5

The Mamdani Inference Model is adopted here as it is the most commonly used fuzzy methodology and was one among the first few control systems built using fuzzy set theory. Ebrahim Mamdani (1975) proposed and used it to control a steam engine and boiler combination by synthesizing a set of linguistic control rules obtained from experienced human operators. Mamdani's effort originated from Lotfi Zadeh's paper on fuzzy algorithms for complex systems and decision processes (1973).

The following standard operator set is used in this model:

Operators	Type	Default Function
AND	BINARY	MIN(a,b)
OR	BINARY	MAX(a,b)
IMPLICATION	BINARY	MIN(a,b)
ALSO	BINARY	MAX(a,b)
NOT	UNARY	1-a
STRONGLY	UNARY	a^2
MODERATELY	UNARY	a^(1/2)
SLIGHTLY	UNARY	4.a.(1-a)
DEFUZZIFICATION	DEFUZZIFICATION	Centre Of Area

A three-parameter (a, b, c) bell shaped continuous membership grade function is chosen for each linguistic of both input and output variables (This is a direct generalization of Cauchy Distribution) so that membership functions can be fine grained according to the necessity. The parameters a, b, c denote respectively the middle point of bell shaped curve (where the grade is max), the degree of peakedness (resembling the Kurtosis in Normal distribution) which taken to be 1 and half width of the membership function. c is kept constant=0.05 for ρ_{xy} , ρ_{xz} and ρ_{yz} . C takes values 5 for PRE and b is kept constant=1 throughout.

The function is defined by- $$f(x;a,b,c) = \frac{1}{1+\left|\frac{x-a}{c}\right|^{2b}}$$

The set of values for the parameters are computed from the set of data generated in Table 3.

Table 6.a

Sl. No.	Left end Of ρ_{xy}	Right end Of ρ_{xy}	Mid Point (a)	Linguistics	Interpretation	Half Width
1	.15	.25	0.2	Vwc	Very weekly correlated positive	0.05
2	.25	.35	.3	Wc	weekly positively correlated 3	0.05
3	.35	.45	.4	c1	moderately correlated 1	0.05
4	.45	.55	.5	c2	moderately correlated 2	0.05
5	.55	.65	.6	c3	moderately correlated 3	0.05
6	.65	.75	.7	c4	moderately correlated 4	0.05
7	.75	.85	.8	Sc	strongly correlated	0.05
8	.85	.95	.9	Hsc	Highly strongly correlated	0.05

Table 6.b

Sl. No.	Left end Of ρ_{xz}	Right end Of ρ_{xz}	Mid Point (a)	Linguistics	Interpretation	Half Width
1	.35	.45	.4	Mc	mildly correlated	0.05
2	.45	.55	.5	c1	moderately correlated 1	0.05
3	.55	.65	.6	c2	moderately correlated 2	0.05
4	.65	.75	.7	Sc	strongly correlated	0.05
5	.75	.85	.8	Hc	highly correlated	0.05
6	.85	.95	.9	Vhc	Very highly correlated	0.05

A 8x6x7 Fuzzy Association Matrix (FAM) is constructed which is the basis of FLC engine and the 'Centre of Area' method (which resembles the expected value computation in probability distribution) is adopted for defuzzification which is most widely used method and defined by-

$$z_{COA} = \int_A z\mu(z)dz \Big/ \int_A \mu(z)dz$$

All computations are done with the help of standard fuzzy software named XFuzzyVs3.0from IMSE-CNM which is available on internet (*vide:xfuzzy-team@imse.cnm.es*).

7.1. Categorization of Efficacy of Proposed Work

The above analysis of empirical study using fuzzy tools gives the advantage to find out the specific ranges of ρ_{xy}, ρ_{xz} and ρ_{yz} where our suggested estimator dominates

i. extremely
ii. mildly and
iii. equally

the sample mean estimator $\bar{y}$. To elucidate these particular regions of ρ_{xy}, ρ_{xz} & ρ_{yz}, the same graph of PRE in different views are presented below:

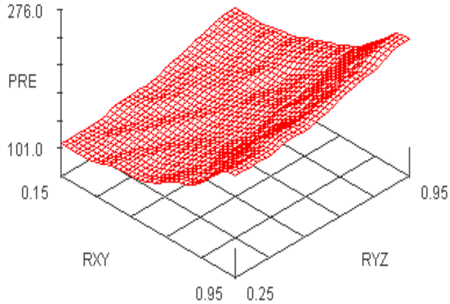

Fig. 1. Horizontal Front View $\rho_{xz} = 0.65$

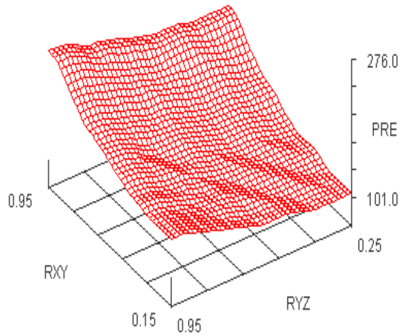

Fig. 2. Horizontal Rear View $\rho_{xz} = 0.65$

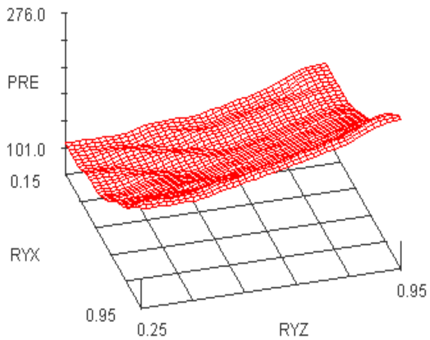

Fig. 3. Horizontal Front Rear View $\rho_{xz} = 0.8$

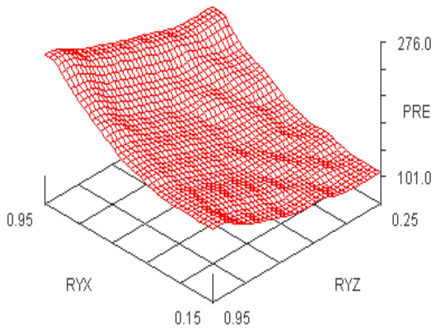

Fig. 4. Horizontal Front View $\rho_{xz} = 0.8$

Surface plot of PRE of the proposed class of estimators t_p' (under its respective optimum condition as discussed in section 5) against $\rho_{yx}, \rho_{yz}, \rho_{xz}$ from different angles.

(It may be noted from the Figs. 1 - 4 that $RYX = \rho_{yx}$, $RXZ = \rho_{xz}$ and $RYZ = \rho_{yz}$.)

8. CONCLUSION

The following interpretations can be read out from the present study:

The following conclusions may be read-out from the present study.

(a) Table 1 exhibits that for high positive values of the correlation coefficients, the proposed class of estimators t_p' yield impressive gains in efficiency over the existing estimators t_i(i=1, 2, . . ., 5). This pattern indicates that proposed class of estimators are more efficient than the existing ones which enhances their recommendations to survey statistician for their usage in real life problem.

(b) From table 2, it is observed that for fixed values of ρ_{xz}, the percent relative efficiencies of the class of estimators t_p' are increasing with the increasing values of ρ_{yx}. This behavior indicates that our proposed classes of estimators performs satisfactorily if highly correlated auxiliary variable present in population.

(c) From the above graphical representations of figures 1-4, it may be seen that PRE against ρ_{yx} and ρ_{yz} gives a clear idea about the efficiency of our proposed class of estimator t_p' over the sample mean estimator $\bar{y}$ under its respective optimal condition. It itself describes the specific ranges of ρ_{yx} and ρ_{yz} at where our estimator t_p' dominates (extremely, mildly or equally) $\bar{y}$. From the different views of the graph (taken from top and different sides), it is clear that the portion of the graph which is almost horizontal denotes that the proposed class of estimators t_p' is equally efficient with $\bar{y}$. Whereas the uprising portions denote the mildly efficient range and the peaks of the graph along with their neighbourhoods denote the extremely efficient range of the class of estimators t_p'.

The following conclusions may be drawn about the performance of t_p' with respect to the sample mean estimator $\bar{y}$.

i. Fig. 1 and Fig. 2 indicates that the class of estimators t_p' is mildly efficient than $\bar{y}$ when $0.15 \le \rho_{yx} \le 0.47,\ 0.39 \le \rho_{yz} \le 0.67$ and $\rho_{xz} = 0.65$.

ii. It is cleared from Fig. 1, Fig. 2 that the class of estimators t_p' is moderately efficient $\bar{y}$ when $0.47 \le \rho_{yx} \le 0.79,\ 0.67 \le \rho_{yz} \le 0.81$ and $\rho_{xz} = 0.65$.

iii. It can be observed from Fig.1,Fig.2 that the class of estimators t_p' is extremely efficient when $0.79 \le \rho_{yx} \le 0.95,\ 0.81 \le \rho_{yz} \le 0.95$ and $\rho_{xz} = 0.65$.

iv. Also from Fig. 3 and Fig. 4 the class of estimators t_p' is moderately efficient $\bar{y}$ when $0.31 \le \rho_{yx} \le 0.47,\ 0.25 \le \rho_{yz} \le 0.67$ and $\rho_{xz} = 0.8$.

v. It can be observed from Fig. 3, Fig. 4 that the class of estimators t_p' is extremely efficient when $0.47 \le \rho_{yx} \le 0.95,\ 0.67 \le \rho_{yz} \le 0.95$ and $\rho_{xz} = 0.8$.

Thus it is erected that the use of an auxiliary character is highly rewarding in terms of the proposed class of estimators. Moreover, the proposal of the class of estimators in the present study is justified as it unifies several desirable results including producing unbiased estimates (up to first order of sample size) and finding the dominance range of the proposed strategy. Looking at the nice behaviour of the proposed strategy, they are recommended to the survey statisticians for their applications in real life problems.

REFERENCES

[1] ANDERSON, T. W., 1958, An Introduction to Multivariate Statistical Analysis. *John Wiley & Sons, Inc. New York.*

[2] BANDYOPADHYAY, A. AND SINGH, G. N., 2014, Predictive estimation of population mean in two-phase sampling. *Communications in Statistics- Theory and Methods.* DOI: 10.1080/03610926.2014.919396, (Article in Press).

[3] CHAND, L., 1975, Some ratio type estimators based on two or more auxiliary variables.Unpublished Ph. D. thesis. Iowa State University, Ames, Iowa (USA).

[4] CHOUDHURY, S. AND SINGH, B. K., 2012, A class of chain ratio–product type estimators with two auxiliary variables under double sampling scheme. *Journal of the Korean Statistical Society*, 41, 247–256.

[5] CHATTERJEE, A., SINGH, G. N., BANDYOPADHYAY, A. AND MUKHOPADHYAY, P. 2015, A General Procedure for Estimating Population Variance in Successive Sampling using Fuzzy Tools. Hacettepe Journal of Mathematics and Statistics. ISSN: 1303-5010. (S. C. I. E. indexed) (Article in Press) DOI: 10.15672/HJMS.201510014169

[6] COCHRAN, W. G., 1940, The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce. *J. Agric. Soc*. 30, 262-275.

[7] COCHRAN, W. G., 1977, Sampling Techniques. *Wiley Eastern Limited*, *New Delhi, III, Edition.*

[8] GUPTA, S. AND SHABBIR, J., 2007, On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. *Journal of Statistical Planning and Inference,* 137, 1606 – 1611.

[9] HANDIQUE, B. K., DAS, G., AND KALITA, M. C., 2011, Comparative studies on forest sampling techniques with satellite remote sensing inputs. *Project Report, NESAC,* pp. 12-40.

[10] KIREGYERA, B., 1980, A chain ratio type estimators in finite population double sampling using two auxiliary variables. *Metrika*. 17, 217-223.

[11] KIREGYERA, B., 1984, Regression type estimators using two auxiliary variables and the model of double sampling from finite populations. *Metrika*. 31, 215-226.

[12] REDDY, V. N., 1978, A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya, Series C*. 40, 29–37.

[13] SHUKLA, D., PATHAK, SAND THAKUR, N. S., 2012, Estimation of population mean using two auxiliary sources in sample surveys. *Statistics in Transition*, 13(1), 21-36.

[14] SINGH, S., JOARDER, A. H. AND TRACY, D. S., 2001, Median estimation using double sampling. Australian & New Zealand Journal of Statistics, 43(1), 33-46.

[15] SINGH, S. AND DEO, B., 2003, Imputation by power transformation. *Statistical Papers*, 4, 555-579.

[16] SINGH, H. P. AND ESPEJO, M. R., 2007, Double sampling ratio-product estimator of a finite population mean in sampling surveys. *Journal of Applied Statistics*, 34(1), 71–85.

[17] SINGH, H. P. AND VISHWAKARMA, G. K., 2007, Modified Exponential Ratio and Product Estimators for Finite Population Mean in Double Sampling. *Austrian journal of statistics*, 36 (3), 217–225

[18] SINGH, H. P., CHANDRA, P., JOARDER, A. H. AND SINGH, S., 2007, Family of estimators of mean, ratio and product of a finite population using random nonresponse. *Test*, 16, 565–597.

[19] SRIVASTAVA, S. K., 1970, A two-phase sampling estimator in sample surveys. *Australian Journal of Statistics.* 12(1), *23-27.*

[20] SUKHATME, B., 1962, Some ratio-type estimators in two - phase sampling. *Journals of the American Statistics Associations.* 57, 628-632.

[21] SUKHATME, P. V. AND SUKHATME, B. V., 1970, Sampling theory of surveys with applications. *Asia Publishing House*, India.

[22] TRACY, D. S., SINGH, H. P., AND SINGH, R., 1996, An alternative to the ratio-cum-product estimator in sample surveys. *Journal of Statistical Planning and Inference*. 53, 375- 387.

P. Parichha
Department of Mathematics,
Asansol Engineering College, Asansol-713305, India.
Email: parthaparichha1989@gmail.com

K. Basu
Department of Mathematics,
National Institute of Technology, Durgapur –713209, India
Email: kajla.basu@gmail.com

A. Bandyopadhyay
Department of Mathematics,
Asansol Engineering College, Asansol-713305, India.
Email: arnabbandyopadhyay4@gmail.com

P. Mukhopadhyay
Department of Mathematics,
Asansol Engineering College, Asansol-713305, India.
Email: parthamukhopadhyay1967@gmail.com