Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 1, 74 – 88
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n1p74

SOME EFFICIENT SAMPLING STRATEGIES BASED ON RATIO TYPE ESTIMATOR

Shashi Bhushan*

Department of Statistics,PUC – Campus, Mizoram University, India

Received 20 June 2010; Accepted 11 February 2011
Available online 26 April 2012

Abstract: *In this paper we have proposed to use two sampling strategies based on the modified ratio estimator using the coefficient of kurtosis of the auxiliary variable by Singh et al [9] for estimating the population mean (total) of the study variable in a finite population. The main objective of this paper is to provide some ideas for unbiased estimation of population mean of the study variable by using prior information on auxiliary variable. A comparative study is made with usual sampling strategies utilizing the availability of the range prior information regarding the optimizing value of the characterizing scalar. Finally some concluding remarks are given and an empirical study is included as an illustration.*

Keywords: *Ratio estimator, coefficient of kurtosis, unbiasedness, mean square error, prior information.*

1. Introduction

In survey sampling literature, it is quite common to use the prior information about some auxiliary variable closely related to the study variable in order to estimate some unknown parameter of the study variable. Many transformed estimation procedures are also available in the literature for increasing the efficiency but, unfortunately, this is always done at the cost of unbiasedness. There are numerous instances where the bias of these estimation procedures is very large with respect to their mean square error and therefore may not be advisable. The aim of this paper is to introduce some efficient as well as unbiased sampling strategies for finite population. Further, it may be of interest to note that there are very few works in sampling literature wherein the focus is on devising better sampling strategies both in terms of efficiency

* E-mail: shashi.py@gmail.com

and unbiasedness. This may be done by some innovations in both the aspects of sampling strategy viz namely sampling scheme and estimation procedure. In this paper, we have used the prior information about coefficient of kurtosis at both the stages – estimation and sampling scheme so that the accuracy of the sampling strategy is improved.

The use of prior value of coefficient of kurtosis in estimating the population variance of characteristic under study y was first made by Singh et al [10]. Later, used by Searls and Intrapanich [6] and Singh et al [9] in the estimation of population mean of characteristic under study. The knowledge of coefficient of kurtosis of the character under study is seldom available. However, the coefficient of kurtosis of an auxiliary character may be easily known or obtained. It was mentioned by the above authors that the use of such prior information about the coefficient of kurtosis leads to more efficient estimation population mean and variance of the study variable.

The ratio estimator for estimating the population mean of the study variable y is given by:

$$\bar{y}_R = \bar{y}\bar{X}/\bar{x} = \hat{R}\bar{X} \tag{1}$$

where $\hat{R} = \bar{y}/\bar{x}$, $\bar{y}$ is the sample mean of the study variable y, $\bar{x}$ is the sample mean of the some auxiliary variable x and $\bar{X}$ is the population mean of the x which is assumed to be known.

If the population coefficient of kurtosis of x denoted by β_{2x} is known, then Singh et al [9] proposed a modified ratio estimator for estimating the population mean $\bar{Y}$ of the study variable given by:

$$\bar{y}_{SR} = \bar{y}\frac{\left(\bar{X} + \beta_{2x}\right)}{\left(\bar{x} + \beta_{2x}\right)} \tag{2}$$

The bias and mean square error of $\bar{y}_{SR}$ under simple random sampling are given by:

$$Bias\left(\bar{y}_{SR}\right) = \gamma_n \bar{Y}\left(\delta^2 C_x^2 - \rho\delta C_x C_y\right) \tag{3}$$

and

$$MSE\left(\bar{y}_{SR}\right) = \gamma_n \bar{Y}^2\left\{C_y^2 + \delta C_x^2\left(\delta - 2K\right)\right\} \tag{4}$$

where $\gamma_n = \left(1 - f\right)/n$, C_y is the population coefficient of variation of y, $\delta = \bar{X}/\left(\bar{X} + \beta_{2x}\right)$, $K = \rho C_y/C_x$ and ρ is the population correlation coefficient between x and y.

Similarly, a product type estimator $\bar{y}_{SP}$ may also be considered given by:

$$\bar{y}_{SP} = \bar{y}\frac{\left(\bar{x} + \beta_{2x}\right)}{\left(\bar{X} + \beta_{2x}\right)}$$

having bias and mean square error given by:

$$Bias\left(\overline{y}_{SP}\right)=\gamma_n\overline{Y}\rho\delta C_x C_y \text{ and}$$
$$MSE\left(\overline{y}_{SP}\right)=\gamma_n\overline{Y}^2\left\{C_y^2+\delta C_x^2\left(\delta+2K\right)\right\} \text{ respectively.} \tag{5}$$

Now, motivated by Walsh [12] and Reddy [5], in this paper we propose some improved sampling strategies such that the estimator of population mean $\overline{Y}$ is:

$$\overline{y}_{SA}=\overline{y}\frac{\left(\overline{X}+\beta_{2x}\right)}{\left\{A\left(\overline{x}+\beta_{2x}\right)+(1-A)\left(\overline{X}+\beta_{2x}\right)\right\}} \tag{6}$$

Note that for $A=1$ the proposed estimator reduces to $\overline{y}_{SR}$. We now consider this estimator under the following sampling schemes:

1. Simple random sampling without replacement (SRSWOR) along with the jack-knife technique and denote the resulting estimator as $\widehat{\overline{Y}_{SS}}^{*}$.
2. Midzuno [3] - Lahiri [2] - Sen [7] type sampling scheme and denote the resulting estimator by $\overline{y}_{SM}$.

Both the sampling strategies aim at getting some classes of better sampling strategies than the existing ones in the sense of unbiasedness and lesser mean square error.

2. Bias and MSE of Proposed Estimator under SRSWOR

Consider estimator $\overline{y}_{SA}$ under SRSWOR and denote it by $\overline{y}_{SS}$. Let $\overline{y}_s=\overline{Y}+e_0$ and $\overline{x}_s=\overline{X}+e_1$ such that:

$$E(e_0)=E(e_1)=0 \tag{7}$$

Putting these values (6) and simplifying, we have:

$$\overline{y}_{SS}-\overline{Y}=e_0-\frac{A\overline{Y}e_1}{\left(\overline{X}+\beta_{2x}\right)}+\frac{A^2\overline{Y}e_1^2}{\left(\overline{X}+\beta_{2x}\right)^2}-\frac{A\overline{Y}e_0e_1}{\left(\overline{X}+\beta_{2x}\right)}+\ldots \tag{8}$$

Taking expectation on both sides and using (7), we have:

$$Bias\left(\overline{y}_{SS}\right)=E\left(\overline{y}_{SS}\right)-\overline{Y}\cong\overline{Y}\left\{\frac{A^2\overline{Y}E\left(e_1^2\right)}{\left(\overline{X}+\beta_{2x}\right)^2}-\frac{AE\left(e_0e_1\right)}{\left(\overline{X}+\beta_{2x}\right)}\right\}$$

Since $E(e_0{}^2)=\gamma_n\overline{Y}^2C_y{}^2$, $E(e_1{}^2)=\gamma_n\overline{X}^2C_x{}^2$, $E(e_0e_1)=\gamma_n\overline{X}\overline{Y}\rho C_xC_y$ (9)

Therefore,

$$Bias\left(\bar{y}_{SS}\right) = \gamma_n \bar{Y}\left\{A^2\delta^2 C_x^{\,2} - A\delta\rho C_x C_y\right\} \tag{10}$$

Now, for mean square error, consider (8) up to the first order of approximation:

$$MSE(\bar{y}_{SS}) = E(\bar{y}_{SS} - \bar{Y})^2 \cong E\left\{e_0 - \frac{A\bar{Y}e_1}{\left(\bar{X}+\beta_{2x}\right)}\right\}^2$$

$$= E(e_0^{\,2}) + \frac{A^2\bar{Y}^2E(e_1^{\,2})}{\left(\bar{X}+\beta_{2x}\right)^2} - \frac{2A\bar{Y}E(e_0e_1)}{\left(\bar{X}+\beta_{2x}\right)}$$

$$= \gamma_n\bar{Y}^2\left\{C_y^{\,2} + \delta C_x^{\,2}\left(A^2\delta - 2AK\right)\right\} \tag{11}$$

Note that on putting $A=1$ and -1 in (11), we get expressions of mean square error of $\bar{y}_{SR}$ and $\bar{y}_{SP}$ respectively. The optimizing value of the characterizing scalar A is given by:

$$A = K/\delta = A_{opt} \text{ (say)} \tag{12}$$

The minimum mean square error under optimizing value of $A = A_{opt}$ is:

$$MSE\left(\bar{y}_{SS}\right) = \gamma_n\bar{Y}^2(1-\rho^2)C_y^{\,2} \tag{13}$$

which is same as the mean square error of the linear regression estimator. Also note that $Bias\left(\bar{y}_{SS}\right) = 0$ under the optimizing value of A.

3. The Proposed Jack-Knife Sampling Strategy

Let us now apply Quenouille's [4] method of Jack-Knife such that the sample of size $n = 2m$ from a population of size N is split up at random into two sub samples of size m each. For further details one may refer to Gray and Schucany [1]. Let us define:

$$\bar{y}_{ss}^{(i)} = \bar{y}_i\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{A\left(\bar{x}_i+\beta_{2x}\right)+(1-A)\left(\bar{X}+\beta_{2x}\right)\right\}};\, i = 1,2$$

$$\bar{y}_{ss}^{(3)} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{A\left(\bar{x}+\beta_{2x}\right)+(1-A)\left(\bar{X}+\beta_{2x}\right)\right\}} \tag{14}$$

where is the characterizing scalar to be chosen suitably such that $s = s_1 + s_2$, s_1 and s_2 be the two sub samples of size m each and $+$ denotes the disjoint union. $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}_s$ denote the sample means based on two sub samples of size m and the entire sample of size $n = 2m$ for characteristic y. $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_s$ denote the sample means based on two sub samples of size m and the entire sample of size $n = 2m$ for characteristic x. It can be easily seen that:

$$Bias\left(\bar{y}_{ss}^{(i)}\right) = \gamma_m \bar{Y}\left\{A^2\delta^2 C_x{}^2 - A\delta\rho C_x C_y\right\};\ i = 1,2$$

$$Bias\left(\bar{y}_{ss}^{(3)}\right) = \gamma_n \bar{Y}\left\{A^2\delta^2 C_x{}^2 - A\delta\rho C_x C_y\right\} = B_1(say) \tag{15}$$

Let us define $\widehat{\bar{Y}}_{SS}{}' = \left(\bar{y}_{SS}^{(1)} + \bar{y}_{SS}^{(2)}\right)\Big/2$ as an alternative estimator of the population mean $\bar{Y}$.

The bias of $\widehat{\bar{Y}}_{SS}{}'$ is:

$$Bias(\widehat{\bar{Y}}_{SS}{}') = \gamma_m \bar{Y}\left\{A^2\delta^2 C_x^2 - A\delta\rho C_x C_y\right\} = B_2(say) \tag{16}$$

We propose the jackknife estimator $\widehat{\bar{Y}}_{SS}{}^*$ for estimating population mean $\bar{Y}$ given by:

$$\widehat{\bar{Y}}_{SS}{}^* = \frac{\left(\bar{y}_{SS}^{(3)} - R\widehat{\bar{Y}}_{SS}{}'\right)}{\left(1-R\right)} = \left[\bar{y}_{SS}^{(3)} - \left\{\frac{N-2n}{2(N-n)}\right\}\widehat{\bar{Y}}_{SS}{}'\right]\Big/\left[1 - \left\{\frac{N-2n}{2(N-n)}\right\}\right] \tag{17}$$

where $R = B_1/B_2$. Taking expectation of (17) and using (15) and (16) we obtain $E(\widehat{\bar{Y}}_{SS}{}^*) = \bar{Y}$ showing that $\widehat{\bar{Y}}_{SS}{}^*$ is an unbiased estimator of population mean $\bar{Y}$ to the first order of approximation. Consider:

$$MSE(\widehat{\bar{Y}}_{SS}{}^*) = E\left\{\frac{\left(\bar{y}_{SS}^{(3)} - R\widehat{\bar{Y}}_{SS}{}'\right)}{\left(1-R\right)} - \bar{Y}\right\}^2$$

$$= \frac{E\left\{\bar{y}_{SS}^{(3)} - R\widehat{\bar{Y}}_{SS}{}' - \bar{Y} + R\bar{Y}\right\}^2}{(1-R)^2}$$

$$= \frac{\left\{E(\overset{(3)}{\bar{y}_{SS}} - \bar{Y})^2 + R^2 E(\widehat{\bar{Y}}_{SS}' - \bar{Y})^2 - 2RE(\overset{(3)}{\bar{y}_{SS}} - \bar{Y})(\widehat{\bar{Y}}_{SS}' - \bar{Y})\right\}}{(1-R)^2} \tag{18}$$

Also,

$$E(\overset{(3)}{\bar{y}_{SS}} - \bar{Y})^2 = MSE(\overset{(3)}{\bar{y}_{SS}}) = \gamma_n \bar{Y}^2 \left\{C_y{}^2 + \delta C_x{}^2 \left(A^2\delta - 2AK\right)\right\} \tag{19}$$

Further,

$$E(\widehat{\bar{Y}}_{SS}' - \bar{Y})^2 = E\left\{\frac{\left(\overset{(1)}{\bar{y}_{SS}} + \overset{(2)}{\bar{y}_{SS}}\right)}{2} - \bar{Y}\right\}^2$$

$$= \frac{\left\{E(\overset{(1)}{\bar{y}_{SS}} - \bar{Y})^2 + E(\overset{(2)}{\bar{y}_{SS}} - \bar{Y})^2 + 2E(\overset{(1)}{\bar{y}_{SS}} - \bar{Y})(\overset{(2)}{\bar{y}_{SS}} - \bar{Y})\right\}}{4} \tag{20}$$

Since:

$$E(\overset{(i)}{\bar{y}_{SS}} - \bar{Y})^2 = MSE(\overset{(i)}{\bar{y}_{SS}}) = \gamma_m \bar{Y}^2 \left\{C_y{}^2 + \delta C_x{}^2 \left(A^2\delta - 2AK\right)\right\}; i = 1,2 \tag{21}$$

Let $\bar{y}_i = \bar{Y} + e_0^{(i)}$ and $\bar{x}_i = \bar{X} + e_1^{(i)}$ such that $E\left(e_0^{(i)}\right) = E\left(e_1^{(i)}\right) = 0$, $i = 1,2$
Consider:

$$E(\overset{(1)}{\bar{y}_{SS}} - \bar{Y})(\overset{(2)}{\bar{y}_{SS}} - \bar{Y}) = E\left\{e_0^{(1)} - \frac{A\bar{Y}e_1^{(1)}}{\left(\bar{X} + \beta_{2x}\right)}\right\}\left(e_0^{(2)} - \frac{A\bar{Y}e_1^{(2)}}{\left(\bar{X} + \beta_{2x}\right)}\right)$$

$$= E(e_0^{(1)}e_0^{(2)}) - \frac{A\bar{Y}\left\{E(e_0^{(2)}e_1^{(1)}) + E(e_0^{(1)}e_1^{(2)})\right\}}{\left(\bar{X} + \beta_{2x}\right)} + \frac{A^2\bar{Y}^2 E(e_1^{(1)}e_1^{(2)})}{\left(\bar{X} + \beta_{2x}\right)^2}$$

Substituting the results in Sukhatme and Sukhatme [11]:

$E(e_0{}^{(1)}e_0{}^{(2)}) = -\bar{Y}^2 C_y{}^2 \big/ N$
$E(e_1{}^{(1)}e_1{}^{(2)}) = -\bar{X}^2 C_x{}^2 \big/ N$
$E(e_0{}^{(1)}e_1{}^{(2)}) = E(e_0{}^{(2)}e_1{}^{(1)}) = -\bar{X}\bar{Y}\rho C_x C_y \big/ N$, we have

$$E(\overset{(1)}{\bar{y}_{SS}} - \bar{Y})(\overset{(2)}{\bar{y}_{SS}} - \bar{Y}) = -\bar{Y}^2 \left\{C_y{}^2 + \delta C_x{}^2 \left(A^2\delta - 2AK\right)\right\} \big/ N \tag{22}$$

Putting the values from (21) and (22) in (20), we have:

$$E(\widehat{\overline{Y}}_{SS}^{'} - \overline{Y})^2 = \frac{\overline{Y}^2}{4}\left(2\gamma_m - \frac{2}{N}\right)\left\{C_y^{\,2} + \delta C_x^{\,2}\left(A^2\delta - 2AK\right)\right\}$$

$$= \gamma_n \overline{Y}^2\left\{C_y^{\,2} + \delta C_x^{\,2}\left(A^2\delta - 2AK\right)\right\} \tag{23}$$

Now consider:

$$E(\overset{(3)}{\overline{y}}_{SS} - \overline{Y})(\widehat{\overline{Y}}_{SS}^{'} - \overline{Y}) = E\left(\overset{(3)}{\overline{y}}_{SS} - \overline{Y}\right)\left\{\left(\overset{(1)}{\overline{y}}_{SS} + \overset{(2)}{\overline{y}}_{SS}\right)\Big/2 - \overline{Y}\right\}$$

$$= \{E(\overset{(3)}{\overline{y}}_{SS} - \overline{Y})(\overset{(1)}{\overline{y}}_{SS} - \overline{Y}) + E(\overset{(3)}{\overline{y}}_{SS} - \overline{Y})(\overset{(2)}{\overline{y}}_{SS} - \overline{Y})\}\Big/2$$

Since:

$$E(\overset{(3)}{\overline{y}}_{SS} - \overline{Y})(\overset{(i)}{\overline{y}}_{SS} - \overline{Y}) = E\left\{e_0 - \frac{A\overline{Y}e_1}{\left(\overline{X} + \beta_{2x}\right)}\right\}\left\{e_0^{(i)} - \frac{A\overline{Y}e_1^{(i)}}{\left(\overline{X} + \beta_{2x}\right)}\right\} \text{where } i = 1,2$$

$$= E(e_0 e_0^{(i)}) - \frac{A\overline{Y}\left\{E(e_0^{(i)}e_1) + E(e_0 e_1^{(i)})\right\}}{\left(\overline{X} + \beta_{2x}\right)} + \frac{A^2\overline{Y}^2 E(e_1 e_1^{(i)})}{\left(\overline{X} + \beta_{2x}\right)^2}$$

using the following results given in Sukhatme and Sukhatme [11]:

$$E(e_0 e_0^{(i)}) = \gamma_n \overline{Y}^2 C_y^{\,2}$$
$$E(e_1 e_1^{(i)}) = \gamma_n \overline{X}^2 C_x^{\,2}$$
$$E(e_0 e_1^{(i)}) = E(e_0^{(i)} e_1) = \gamma_n \overline{X}\overline{Y}\rho C_x C_y \text{ for } i = 1,2 \text{ we have}$$

$$E(\overset{(3)}{\overline{y}}_{SS} - \overline{Y})(\overset{(i)}{\overline{y}}_{SS} - \overline{Y}) = \gamma_n \overline{Y}^2\left\{C_y^{\,2} + \delta C_x^{\,2}\left(A^2\delta - 2AK\right)\right\} \tag{24}$$

Putting these values from (19), (23), (24) in (18), we have

$$MSE(\widehat{\overline{Y}}_{SS}^{*}) = \frac{\gamma_n \overline{Y}^2(1 + R^2 - 2R)\left\{C_y^2 + \delta C_x^2\left(A^2\delta - 2AK\right)\right\}}{(1-R)^2}$$

$$= \gamma_n \overline{Y}^2\left\{C_y^{\,2} + \delta C_x^{\,2}\left(A^2\delta - 2AK\right)\right\} \tag{25}$$

which is equal to the mean square error of $\overline{y}_{SS}$.
The optimizing value of the characterizing scalar A is given by (12) and the minimum mean square error $MSE(\widehat{\overline{Y}}_{SS}^{*})_{\min}$ under optimizing value of $A = A_{opt}$ is same as given by (13).

4. Proposed Midzuno-Lahiri-Sen Type Sampling Strategy

Let us consider $\bar{y}_{SA}$ under Midzuno [3] - Lahiri [2] - Sen [7] type sampling scheme and denote it by $\bar{y}_{SM}$. The proposed Midzuno - Lahiri - Sen type sampling scheme for selecting a sample s of size n deals with selecting first unit with probability proportional to $\bar{X}+\beta_{2x}+A(x_i-\bar{X})$ where x_i is the size of the first selected unit such that:

$$P(i) = \text{P(selecting first unit } i \text{ with size } x_i) = \frac{\left\{\bar{X}+\beta_{2x}+A(x_i-\bar{X})\right\}}{N\left(\bar{X}+\beta_{2x}\right)} \tag{26}$$

and selecting the remaining $n-1$ units in the sample from $N-1$ units in the population by simple random sampling without replacement. Thus probability of selecting a sample s of size n is given by:

$$P(s) = \frac{\left\{\bar{X}+\beta_{2x}+A(\bar{x}-\bar{X})\right\}}{(\bar{X}+\beta_{2x}).{}^{N}C_n} \tag{27}$$

Consider:

$$\begin{aligned}
E(\bar{y}_{SM}) &= E\left[\bar{y}(\bar{X}+\beta_{2x})\Big/\left\{\bar{X}+\beta_{2x}+A(\bar{x}-\bar{X})\right\}\right] \\
&= \sum_{s=1}^{{}^{N}C_n}\bar{y}(\bar{X}+\beta_{2x})P(s)\Big/\left\{\bar{X}+\beta_{2x}+A(\bar{x}-\bar{X})\right\} \\
&= \sum_{s=1}^{{}^{N}C_n}\bar{y}\Big/{}^{N}C_n \\
&= E(\bar{y}) \qquad\qquad \text{(under SRSWOR)} \\
&= \bar{Y}
\end{aligned} \tag{28}$$

showing that $\bar{y}_{SM}$ is an unbiased estimator of population mean $\bar{Y}$ for all values of A under the proposed Midzuno-Lahiri-Sen type sampling scheme. Now, since:

$$MSE(\bar{y}_{SM}) = V(\bar{y}_{SM}) = E(\bar{y}_{SM}^2) - \bar{Y}^2$$

where

$$E(\bar{y}_{SM}^2) = \sum_{s=1}^{{}^{N}C_n}\bar{y}_{SM}^2.P(s)$$

$$= \sum_{s=1}^{^NC_n} \left[\frac{\bar{y}(\bar{X}+\beta_{2x})}{\left\{\bar{X}+\beta_{2x}+A(\bar{x}-\bar{X})\right\}} \right]^2 P(s)$$

$$= \sum_{s=1}^{^NC_n} \bar{y}^2 \frac{(\bar{X}+\beta_{2x})}{\left[{}^NC_n \left\{\bar{X}+\beta_{2x}+A(\bar{x}-\bar{X})\right\}\right]}$$

$$= E\left[\left(\bar{Y}+e_0\right)^2 \left\{1+\frac{Ae_1}{\left(\bar{X}+\beta_{2x}\right)}\right\}^{-1} \right]$$

$$= E\left[\left(\bar{Y}^2+e_0^2+2\bar{Y}e_0\right) \left\{1-\frac{Ae_1}{\left(\bar{X}+\beta_{2x}\right)}+\frac{A^2e_1^2}{\left(\bar{X}+\beta_{2x}\right)^2}-\ldots\right\} \right]$$

$$= \bar{Y}^2 + E(e_0^2) + \frac{A^2\bar{Y}^2E(e_1^2)}{\left(\bar{X}+\beta_{2x}\right)^2} - 2\frac{A\bar{Y}E(e_0e_1)}{\left(\bar{X}+\beta_{2x}\right)} \text{ (upto 1}^{\text{st}}\text{ order of approximation)}$$

Therefore, by using (2.3), we have:

$$MSE(\bar{y}_{SM}) = \gamma_n \bar{Y}^2 \left\{C_y^2 + \delta C_x^2 \left(A^2\delta - 2AK\right)\right\} = MSE(\bar{y}_S) \text{ (say)} \tag{29}$$

which is same as that of (11) and (25). Therefore, the optimizing value of the characterizing scalar A is same as given by (12). Also, the minimum mean square error under optimizing value of $A = A_{opt}$ is same as given by (13).

5. Comparative Study

From (29), we have:

$$MSE(\bar{y}_S) = \gamma_n \bar{Y}^2 \left\{C_y^2 + \delta C_x^2 \left(A^2\delta - 2AK\right)\right\} \tag{30}$$

and further we know that:

$$MSE(\bar{y}) = \gamma_n \bar{Y}^2 C_y^2 \tag{31}$$

so that $MSE(\bar{y}_S) < MSE(\bar{y})$

if, $\quad -2AK < -A^2\delta \quad$ (32)

or

(i) if $A\delta > 0$ and $R > 0$, the efficiency condition (32) reduces to:

$$K > \frac{A\delta}{2} \tag{33}$$

(ii) if $A\delta < 0$ and $R > 0$, the efficiency condition (32) reduces to:

$$K < \frac{A\delta}{2} \tag{34}$$

Further, we know that mean square error of ratio estimator $\overline{y}_R$ to be:

$$MSE(\overline{y}_R) = \gamma_n \overline{Y}^2 \{C_y^2 + C_x^2 - 2\rho C_x C_y\} \tag{35}$$

Thus, $MSE(\overline{y}_S) < MSE(\overline{y}_R)$
if and only if:

$$(A^2\delta^2 - 1)C_x^2 - 2(A\delta - 1)\rho C_y C_x < 0 \text{ or } (A\delta - 1)(A\delta + 1 - 2K) < 0 \tag{36}$$

(i) When A is chosen such that $A\delta > 1$, the efficiency condition (36) reduces to:

$$K > \frac{(A\delta + 1)}{2} \tag{37}$$

(ii) When A is chosen such that $A\delta < 1$, the efficiency condition (36) reduces to:

$$K < \frac{(A\delta + 1)}{2} \tag{38}$$

Also, we know that the mean square error of the product estimator $\overline{y}_P$ is:

$$MSE(\overline{y}_P) = \gamma_n \overline{Y}^2 \{C_y^2 + C_x^2 + 2\rho C_x C_y\} \tag{39}$$

Therefore, $MSE(\overline{y}_S) < MSE(\overline{y}_P)$
if and only if:

$$(A^2\delta^2 - 1)C_x^2 - 2(A\delta + 1)\rho C_y C_x < 0 \text{ or } (A\delta + 1)(A\delta - 1 - 2C) < 0 \tag{40}$$

(i) When A is chosen such that $A\delta > -1$, the efficiency condition (40) is reduced to:

$$K > \frac{(A\delta - 1)}{2} \tag{41}$$

(ii) When A is chosen such that $A\delta < -1$, the efficiency condition (40) is reduced to:

$$K < \frac{(A\delta - 1)}{2} \tag{42}$$

From (13), we know that:

$$MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 (1-\rho^2) C_y^{\,2} \tag{43}$$

and the mean square error of linear regression estimator $\bar{y}_{lr}$ is:

$$MSE(\bar{y}_{lr}) = \gamma_n \bar{Y}^2 (1-\rho^2) C_y^{\,2} \tag{44}$$

Hence $MSE(\bar{y}_S)_{\min} = MSE(\bar{y}_{lr})$ to the first degree of approximation. Also $\bar{y}_S$ is unbiased for all values of A under the proposed jack-knife technique or proposed Midzuno-Lahiri-Sen type sampling scheme while the linear regression estimator $\bar{y}_{lr}$ is biased estimator under simple random sampling without replacement to the first degree of approximation.

6. Comparative Study

If the minimizing value $A = K/\delta = A_{opt}$ is known then, we have:

$$MSE(\bar{y}) - MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 \rho^2 C_y^{\,2} \geq 0 \tag{45}$$

$$MSE(\bar{y}_R) - MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 \left(C_x - \rho C_y\right)^2 \geq 0 \tag{46}$$

$$MSE(\bar{y}_P) - MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 \left(C_x + \rho C_y\right)^2 \geq 0 \tag{47}$$

$$MSE(\bar{y}_{SR}) - MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 \left(\delta C_x - \rho C_y\right)^2 \geq 0 \tag{48}$$

$$MSE(\bar{y}_{SP}) - MSE(\bar{y}_S)_{\min} = \gamma_n \bar{Y}^2 \left(\delta C_x + \rho C_y\right)^2 \geq 0 \tag{49}$$

Hence, under the optimizing value of the characterizing scalar $A = K/\delta = A_{opt}$, the proposed sampling strategies are always better than $\bar{y}$, $\bar{y}_{SR}$, $\bar{y}_{SP}$, $\bar{y}_R$, $\bar{y}_P$ and $\bar{y}_{lr}$ in sense of unbiasedness and gain in efficiency.
But since the minimizing value $A = K/\delta = A_{opt}$ depends on the exact value of K which may not always be known. However, range information about the stable value of K may be easily known

in practice, hence, using this information for K, we find efficient estimators in the sense of having lesser mean square error as follows.
As we know that mean per unit estimator $\bar{y}$ is preferred to ratio and product estimators when $0 < K < 1/2$ and $-1/2 < K < 0$ respectively. In such situations the valuable auxiliary information remains unutilized. From efficiency condition (33) and (34) we can get class of estimators, which are better than the mean per unit estimator even in situations when $0 < K < 1/2$ or $-1/2 < K < 0$.
Let the range information about K be known as $K > K_0$ where $0 < K_0 < 1/2$, then we choose $A\delta$ satisfying the efficiency condition (33) such that $A\delta/2 = K_0$ or $A\delta = 2K_0$ to get a class of estimators:

$$\bar{y}_{2K_0} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+2K_0\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{50}$$

which are better than the mean per unit estimator $\bar{y}$ in the sense of having smaller mean square error. Further if it is known that $K < K_1$ where $-1/2 < K_1 < 0$ then we choose $A\delta$ satisfying the efficiency condition (34) such that $A\delta = 2K_1 (< 0)$ so that the class of estimators:

$$\bar{y}_{2K_1} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+2K_1\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{51}$$

are better than $\bar{y}$ in the sense of having lesser mean square error. More specifically, for prior range information $K > 1/3$ choosing $A\delta = 2/3$ satisfying the efficiency (dominance) condition (34), we get more efficient unbiased estimator:

$$\bar{y}_{2/3} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+2\left(\bar{x}-\bar{X}\right)/3\delta\right\}}$$

than $\bar{y}$ in the sense of having smaller mean square error.
Let the range information about K be known as $K > K_0^* (> 1)$, then from the efficiency condition (37) we choose $\left(A\delta+1\right)/2 = K_0^*$ or $A\delta = 2K_0^* - 1$ to get the class of estimators:

$$\bar{y}_{2K_0^*-1} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+\left(2K_0^*-1\right)\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{52}$$

which are better than the ratio estimator in the sense of having lesser mean square error. Further if it is known that $K < K_1^* (< 1)$ then we choose $A\delta$ such that $A\delta = 2K_1^* - 1$ satisfying the efficiency condition (38) so that the class of estimators

$$\bar{y}_{2K_1^*-1} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+\left(2K_1^*-1\right)\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{53}$$

are better than the ratio estimator in the sense of having lesser mean square error. For example, if it is known that $K > 3/2$ we may choose $A\delta = 2$ satisfying the efficiency (dominance) condition (54) to obtain a more efficient unbiased estimator:

$$\bar{y}_2 = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+2\left(\bar{x}-\bar{X}\right)/\delta\right\}}$$

than ratio estimator $\bar{y}_R$ in the sense of having lesser mean square error.

If the range information about K be known as $K > K_0^{'}(>-1)$ then we choose $A\delta$ such that $\left(A\delta-1\right)/2 = K_0^{'}$ or $A\delta = 2K_0^{'}+1$ satisfying the efficiency condition (41) to get the class of estimators:

$$\bar{y}_{2K_0^{'}+1} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+\left(2K_0^{'}+1\right)\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{54}$$

which are better than the product estimator in the sense of having lesser mean square error. Further, if it is known that $K < K_1^{'}(<-1)$ then we choose $A\delta = 2K_1^{'}+1$ satisfying the efficiency condition (42) so that the class of estimators:

$$\bar{y}_{2K_1^{'}+1} = \bar{y}\frac{\left(\bar{X}+\beta_{2x}\right)}{\left\{\left(\bar{X}+\beta_{2x}\right)+\left(2K_1^{'}+1\right)\left(\bar{x}-\bar{X}\right)/\delta\right\}} \tag{55}$$

is more efficient than the product estimator $\bar{y}_P$ in the sense of having smaller mean square error.

In the light of these $\bar{y}_{SM}$ and $\widehat{\bar{Y}}_{SS}^{*}$ can be preferred to $\bar{y}_{SR}$, $\bar{y}_{SP}$, $\bar{y}_P$, $\bar{y}_R$ and $\bar{y}_{lr}$ in sense of unbiasedness and efficiency. Therefore, the proposed sampling strategies can be a better alternative in various practical situations.

7. Empirical Study

Let us consider the following example considered by Singh and Chaudhary (1986) wherein the following values were obtained $\bar{Y}$ = 22.62, $\bar{X}$ = 1467.55, C_x = 1042.46, C_y = 1.7459, β_{2x} =

5.5788, ρ = 0.9022. The bias and mean square errors of the sample mean $\bar{y}$, ratio estimator $\bar{y}_R$, product estimator $\bar{y}_P$, $\bar{y}_{SR}$, $\bar{y}_{SP}$ and $\bar{y}_{lr}$ are given by:

Table 1. Comparision with other sampling strategies.

Est.'s	$\bar{y}$	$\bar{y}_R$	$\bar{y}_P$	$\bar{y}_{lr}$	$\bar{y}_{SR}$	$\bar{y}_{SP}$	$\bar{y}_{SM}$ / $\widehat{\bar{Y}_{SS}}^*$
Bias	0	-0.1667	4.4352	3867.58	-0.3439	0.4810	0
MSE	6564590	0.5804	9.7843	0.5673	2.2918	3.9415	0.5673

Note that the above results are scaled by the factor γ_n. It may be easily verified from the table that only sample mean and the proposed sampling strategies are unbiased. Further, it can be easily observed that the linear regression estimator $\bar{y}_{lr}$ and the proposed sampling strategies attain the minimum mean square error but $\bar{y}_{lr}$ is biased. It is evident from the above empirical study that the proposed sampling strategies are better than the remaining sampling strategies both in terms of unbiasedness and mean square error.

8. Conclusion

It is evident from sections 5, 6 and 7 that the proposed sampling strategies are better than most of the commonly used estimators. The proposed sampling strategies are unbiased for all values of the characterizing scalar *A* and the gain in efficiency is substantial under the optimizing value of the characterizing scalar. Therefore, the proposed sampling strategies provide better alternative for estimating the population mean.

References

[1]. Gray H. L. and Schucany W. R. (1972), *The Generalized Jack-knife Statistic*. New York: Marcel Dekker.

[2]. Lahiri D. B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.*, 3, 133-140.

[3]. Midzuno H. (1952). On the sampling system with probability proportional to the sum of the sizes. *Ann. Inst. Stat. Math.*, 3, 99-107.

[4]. Quenouille M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

[5]. Reddy, V. N. (1974). On a transformed ratio method of estimation. *Sankhya*, C, 36, 59-70.

[6]. Searls, D.T. and Intrapanich, R. (1990). A note on an estimator for variance that utilizes the kurtosis. *Amer. Stat.*, 44(4), 295-296.

[7]. Sen A. R. (1952). Present status of probability sampling and its use in the estimation of a characteristic. *Econometrika*, 20 - 103.

[8]. Singh D. and Chaudhary F. S. (1986). *Theory and analysis of sampling survey design*. New Delhi: Wiley Eastern Limited.

[9]. Singh H. P., Tailor R. and Kakran M. S. (2004). An estimator of population mean using power transformation. *J. I. S. A. S.*, 58(2), 223-230.

[10]. Singh, J., Pandey, B.N. and Hirano, K. (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Ann. Inst. Statist. Math.*, 25, 51-55.
[11]. Sukhatme P. V. and Sukhatme B. V. (1997). *Sampling Theory of Surveys with Applications*. Delhi: Piyush Publications.
[12]. Walsh J. E. (1970). Generalization of ratio estimator for population total. *Sankhya*, A, 32, 99-106.