Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 6(5): 5397–5409.
DOI:10.3934/math.2021318
Received: 13 November 2020
Accepted: 03 March 2021
Published: 15 March 2021

Research article

Estimation of finite population mean under PPS in presence of maximum and minimum values

Sanaa Al-Marzouki[1], Christophe Chesneau[2], Sohail Akhtar[3], Jamal Abdul Nasir[3], Sohaib Ahmad[4], Sardar Hussain[4], Farrukh Jamal[5,*], Mohammed Elgarhy[6] and M. El-Morshedy[7,8]

[1] Statistics Department, Faculty of Science, King Abdul Aziz University, Jeddah 21551, Saudi Arabia
[2] Department of Mathematics, Université de Caen, LMNO, Campus II, Science 3, 14032 Caen, France
[3] Department of Statistics, Government College University Lahore Lahore Pakistan
[4] Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
[5] Department of Statistics, Govt. S.A Postgraduate College Dera Nawab Sahib, Bahawalpur, Punjab 63100, Pakistan
[6] The Higher Institute of Commercial Sciences, Al Mahalla Al Kubra, 31951, Algarbia, Egypt
[7] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[8] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

* **Correspondence:** Email: drfarrukh1982@gmail.com; Tel: +923216802194.

Abstract: This article deals with the estimation of the finite population mean under probability proportional to size (PPS) sampling using information on the auxiliary variable along with the rank of the auxiliary variable. We propose a ratio, product and regression type estimators by incorporating the maximum and minimum values of the study variable and the auxiliary variable. The mathematical expressions of the proposed estimators are derived up to first order of approximation. Efficiency comparisons are made on the basis of real data sets.

Keywords: PPS; mean squared error; auxiliary variable; maximum and minimum values
Mathematics Subject Classification: 62D05

1. Introduction

In the survey sampling literature, researchers have attempted to obtain estimates for population quantities such as mean, total, median, etc., that possess maximum statistical properties. For this

purpose, a representative part of the population is needed. When the population of interest is homogeneous, then one can use a simple random sampling scheme for selecting units. On the other hand, when sampling units vary considerably in size, then units may be selected with probability proportional to size (PPS). Probability proportional to size sampling, usually called as PPS sampling, is an unequal probability sampling scheme, in which the probability of selection for each sampling unit in the population is proportional to an auxiliary variable. Let Y be a variable under consideration and X be a supplementary information. For instance, let us consider that we want to estimate the population in the villages in a particular district. Then we would select a variable on which we have information as the auxiliary variable, e.g.,
(a) Size of each village in the district (correlated with one study variable = 0.75, say).
(b) Number of households in each town in the district (correlation with a study variable = 0.95, say).
Based on the above information, we would select the ancillary variable that has the maximum correlation with the study variable. Thus the variable at (b) may be more useful as auxiliary variable when selecting a sample using probability proportional to size with replacement sampling.

Similarly, surveys in relation to income of households may differ in sizes; for a medical survey related to the number of patients, health units may vary in sizes. Similarly, in an agriculture context, fields may vary in sizes. Villages with larger geographical areas are likely to have large populations and covered large areas under food crops (see [21]). The number of persons in the previous period may be taken as a measure of size related to surveys of socio-economic characters, which are likely to be related to population (see [14]).

The use of auxiliary information can be used either at selection stage or at estimation stage, or at both stages. [15] proposed alternative estimators in PPS sampling for multiple characteristics. [22] proposed the regression type estimator with PPS sampling. The readers are also referred to the papers by [2, 16, 18, 20], and the references cited therein. The use of auxiliary information may increase the precision of estimators of the unknown population parameters such as population mean, variance, correlation coefficient, etc. Some common estimators that utilize the information about the auxiliary variable are highly correlated with the study variable. When the correlation between the study variable and the auxiliary variable is high, in such situation, the rank of the auxiliary variable is also correlated with the study variable.

Recently, many authors have proposed ratio type estimators by means of transforming the auxiliary variables. The readers can explore these research findings by looking [5, 6, 10, 11, 19, 23]. For estimating the finite population mean using maximum and minimum values, see [1, 3, 4, 7, 8, 9, 12, 13], among other.

In this article, we propose ratio, product and regression type estimators for estimating the finite population mean under PPS sampling scheme, using maximum and minimum values. Consider a finite population $U = \{1, 2, ..., N\}$. Let y_i and (x_i, z_i) be the values of the study variable (y) and the auxiliary variables (x, z), respectively. Let r_{xi} be the rank of the auxiliary variable corresponding to rank of x, i.e., (R_x).

Let a sample of size n is selected with probability proportional to size z_i with replacement (PPSWR), i.e., $P_i = \frac{z_i}{\sum_{i=1}^{N} z_i}$.

Suppose that u_i and v_i are the study and auxiliary for the PPS sampling. Let v_i^* denote the rank of v, that is

$$u_i = \frac{y_i}{NP_i}, \quad v_i = \frac{x_i}{NP_i}, \quad v_i^* = \frac{r_{(x)i}}{NP_i},$$

$$\bar{u} = \lambda \sum_{i=1}^{n} u_i, \quad \bar{v} = \lambda \sum_{i=1}^{n} v_i, \quad \bar{v}^* = \lambda \sum_{i=1}^{n} v_i^*.$$

Let

$$e_0 = \frac{\bar{u} - \bar{Y}}{\bar{Y}}, \quad e_1 = \frac{\bar{v} - \bar{X}}{\bar{X}}, \quad e_2 = \frac{\bar{v}^* - \bar{R}_x}{\bar{R}_x}.$$

such that

$$E(e_o) = E(e_1) = E(e_2) = 0.$$

Also

$$E(e_0^2) = \lambda C_u^2, \; E(e_1^2) = \lambda C_v^2, \; E(e_2^2) = \lambda C_{v^*}^2, \; E(e_0 e_1) = \lambda \rho_{uv} C_u C_v,$$
$$E(e_0 e_2) = \lambda \rho_{uv^*} C_u C_v^*, \; E(e_1 e_2) = \lambda \rho_{vv^*} C_v C_{v^*}, \; C_u^2 = \frac{S_u^2}{\bar{Y}^2}, \; C_v^2 = \frac{S_v^2}{\bar{X}^2}, \; C_{v^*}^2 = \frac{S_{v^*}^2}{\bar{R}_x^2},$$

where

$$\lambda = \frac{1}{n}, S_u^2 = \sum_{i=1}^{N} P_i (u_i - \bar{Y})^2, \quad S_v^2 = \sum_{i=1}^{N} P_i (v_i - \bar{X})^2, \quad S_{v^*}^2 = \sum_{i=1}^{N} P_i (v_i^* - \bar{R}_x)^2$$

$$S_{uv} = \sum_{i=1}^{N} P_i (u_i - \bar{Y})(v_i - \bar{X}), \quad S_{uv^*} = \sum_{i=1}^{N} P_i (u_i - \bar{Y})(v_i^* - \bar{R}_x),$$
$$S_{vv^*} = \sum_{i=1}^{N} P_i (v_i - \bar{X})(v_i^* - \bar{R}_x),$$

$$\rho_{uv} = \frac{\sum_{i=1}^{N} P_i (u_i - \bar{Y})(v_i - \bar{X})}{S_u S_v}, \quad \rho_{uv^*} = \frac{\sum_{i=1}^{N} P_i (u_i - \bar{Y})(v_i^* - \bar{R}_x)}{S_u S_{v^*}},$$
$$\rho_{vv^*} = \frac{\sum_{i=1}^{N} P_i (v_i - \bar{X})(v_i^* - \bar{R}_x)}{S_v S_{v^*}}.$$

Many real data sets carry unexpected large (y_{max}) or small (y_{min}) values. In estimation of finite population mean, the results will be sensitive when such types of values occur. Under these circumstances, when there exist y_{max} and y_{min}, then the results will be either overestimated or underestimated. To handle such type of situation, [17] suggested the following unbiased estimator for the estimation of finite population mean using maximum and minimum values:

$$\bar{y}_s = \begin{cases} \bar{y} + c, & \text{if samples contain } y_{min} \text{ but not } y_{max} \\ \bar{y} - c, & \text{if samples contain } y_{max} \text{ but not } y_{min} \\ \bar{y}, & \text{for all other samples.} \end{cases}$$

The variance of $\bar{y}_s$ is given by

$$V(\bar{y}_s) = \lambda S_y^2 - \frac{2\lambda nc}{N-1}(y_{max} - y_{min} - nc), \tag{1.1}$$

where S_y^2 is the population variance, and c is a constant. The optimum value of c is

$$c_{opt} = \frac{(y_{max} - y_{min})^2}{2n}.$$

The minimum variance of $\bar{y}_s$ is specified by

$$V(\bar{y}_s)_{min} = V(\bar{y}) - \frac{\lambda(y_{max} - y_{min})^2}{2(N-1)},$$

which is always smaller than the variance of $\bar{y}$. The usual ratio estimator under probability proportional to size (PPS) is

$$\bar{y}_{R(pps)} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)\left(\frac{\bar{R}_x}{\bar{v}^*}\right). \tag{1.2}$$

The bias and MSE of $\bar{y}_{R(pps)}$ up to the first order of approximation are given by

$$Bias(\bar{y}_{R(pps)}) \cong \lambda\left(S_v^2 + S_{v^*}^2 + \rho_{vv^*}S_vS_{v^*} - \rho_{uv^*}S_uS_{v^*} - \rho_{uv}S_uS_v\right) \tag{1.3}$$

and

$$MSE(\bar{y}_{R(pps)}) \cong \lambda\left[S_u^2 + R_1^2S_v^2 + R_2^2S_{v^*}^2 + 2R_1R_2S_{vv^*} - 2R_1S_{uv} - 2R_2S_{uv^*}\right], \tag{1.4}$$

respectively, where $R_1 = \frac{\bar{Y}}{\bar{X}}$ and $R_2 = \frac{\bar{Y}}{\bar{R}_x}$. The usual product estimator under PPS is given by

$$\bar{y}_{P(pps)} = \bar{u}\left(\frac{\bar{v}}{\bar{X}}\right)\left(\frac{\bar{v}^*}{\bar{R}_x}\right). \tag{1.5}$$

The bias and MSE of $\bar{y}_{P(pps)}$ up to first order of approximation are given by

$$Bias(\bar{y}_{P(pps)}) \cong \lambda\left(S_v^2 + S_{v^*}^2 + \rho_{vv^*}S_vS_{v^*} + \rho_{uv^*}S_uS_{v^*} + \rho_{uv}S_uS_v\right) \tag{1.6}$$

and

$$MSE(\bar{y}_{P(pps)}) \cong \lambda\left[S_u^2 + R_1^2S_v^2 + R_2^2S_{v^*}^2 + 2R_1R_2S_{vv^*} + 2R_2S_{uv^*} + 2R_1S_{uv}\right], \tag{1.7}$$

respectively. The usual regression estimator for estimating the unknown population mean under PPS sampling scheme is

$$\bar{y}_{lr(pps)} = \bar{u} + b_1(\bar{X} - \bar{v}) + b_2(\bar{R}_x - \bar{v}^*), \tag{1.8}$$

where $b_1 = \frac{s_{uv}}{s_v^2}$, and $b_2 = \frac{s_{uv^*}}{s_{v^*}^2}$, are the sample regression coefficients.

The MSE of $\bar{y}_{lr(pps)}$ up to first order of approximation is obtained as

$$MSE(\bar{y}_{lr(pps)}) = \lambda S_u^2\left(1 - \rho_{uv}^2 - \rho_{uv^*}^2 + 2\rho_{uv}\rho_{uv^*}\rho_{vv^*}\right). \tag{1.9}$$

2. Suggested estimators

Following the lines of [17], we propose a ratio, product and regression type estimators under PPS sampling utilizing the auxiliary variable along with rank of the auxiliary variable having stronger correlation with the study variable. We also incorporate the minimum and maximum values of the study and the auxiliary variables.

2.1. *First situation*

When the correlation between the study variable and the auxiliary variable is positive, then, for the selection of larger value of the auxiliary variable, a larger value of the study variable is to be selected. And, for the selection of the smaller value of the auxiliary variable, a smaller value of the study variable is to be selected. To utilize such type of information, we suggest the ratio type estimator using the auxiliary variable and rank of the auxiliary variable as

$$\hat{\bar{Y}}_{R(pps)} = \bar{u}\left(\frac{\bar{X}}{\bar{v}_{c_{21}}}\right)\left(\frac{\bar{R}_x}{\bar{v}^*_{c_{31}}}\right) \tag{2.1}$$

or

$$\hat{\bar{Y}}_{R(pps)} = \begin{cases} (\bar{u}+c_1)\left(\frac{\bar{X}}{\bar{v}+c_2}\right)\left(\frac{\bar{R}_x}{\bar{v}^*+c_3}\right), & \text{if samples contain } u_{min} \text{ and } v_{min}, v^*_{min} \\ (\bar{u}-c_1)\left(\frac{\bar{X}}{\bar{v}-c_2}\right)\left(\frac{\bar{R}_x}{\bar{v}^*-c_3}\right), & \text{if samples contain } u_{max} \text{ and } v_{max}, v^*_{max} \\ \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)\left(\frac{\bar{R}_x}{\bar{v}^*}\right), & \text{for all other samples} \end{cases} \tag{2.2}$$

The regression type estimator is

$$\hat{\bar{Y}}_{lr1(pps)} = \bar{u}_{c_{11}} + b_1(\bar{X} - \bar{v}_{c_{21}}) + b_2(\bar{R}_x - \bar{v}^*_{c_{31}}), \tag{2.3}$$

where $b_1 = \frac{s_{uv}}{s^2_v}$, $b_2 = \frac{s_{uv^*}}{s^2_{v^*}}$, $\bar{u}_{c_{11}} = \bar{u} + c_1$, $\bar{v}_{c_{21}} = \bar{v} + c_2$, $\bar{v^*}_{c_{31}} = \bar{v}^* + c_3$. If the sample contains u_{min} and (v_{min}, v^*_{min}), then $\bar{u}_{c_{21}} = \bar{u} - c_1$, $\bar{v}_{c_{22}} = \bar{v} - c_2$, $\bar{v}^*_{c_{32}} = \bar{v}^* - c_3$. If the samples contain u_{max} and (v_{max}, v^*_{max}), then $\bar{u}_{c_{11}} = \bar{u}$, $\bar{v}_{c_{21}} = \bar{v}$, $\bar{v}^*_{c_{31}} = \bar{v}^*$, for all other samples (here, we mean that, if we can take any value of sample, the ratio estimator gives us good result in term of MSEs as compared to the usual ratio and product estimators using two auxiliary variables.)

2.2. *Second situation*

While in this situation, when the correlation between the study variable and the auxiliary variable is negative, then, for the selection of larger value of the auxiliary variable, the smaller value of the study variable is to be selected. And for the selection of the smaller value of the auxiliary variable, the larger value of the study variable is to be selected. In such situation the proposed product type estimator using the auxiliary variable and rank of the auxiliary variable (x) is given by

$$\hat{\bar{Y}}_{P(pps)} = \bar{u}_{c_{12}}\left(\frac{\bar{v}_{c_{22}}}{\bar{X}}\right)\left(\frac{\bar{v}^*_{c_{32}}}{\bar{R}_x}\right) \tag{2.4}$$

or

$$\hat{\bar{Y}}_{P(pps)} = \begin{cases} (\bar{u}+c_1)\left(\frac{\bar{v}+c2}{\bar{X}}\right)\left(\frac{\bar{v}^*+c_3}{\bar{R}_x}\right), & \text{if samples contain } u_{min} \text{ and } v_{max}\ v^*_{max} \\ (\bar{u}-c_1)\left(\frac{\bar{v}+c2}{\bar{X}}\right)\left(\frac{\bar{v}^*+c_3}{\bar{R}_x}\right), & \text{if samples contain } u_{max} \text{ and } v_{min}, v^*_{min} \\ \bar{u}\left(\frac{\bar{v}}{\bar{X}}\right)\left(\frac{\bar{v}^*}{\bar{R}_x}\right), & \text{for all other samples} \end{cases} \tag{2.5}$$

The regression type estimator is

$$\hat{\bar{Y}}_{lr2(pps)} = \bar{u}_{c_{11}} + b_1(\bar{X} - \bar{v}_{c_{22}}) + b_2(\bar{R}_x - \bar{v}^*_{c_{32}}), \tag{2.6}$$

where $(\bar{u}_{c_{12}} = \bar{u} + c_1,\ \ \bar{v}_{c_{22}} = \bar{v} - c_2,\ \ \bar{v}^*_{c_{32}} = \bar{v}^* + c_3)$.
If the samples contain u_{min} and $(v_{max},\ v^*_{max})$, then $(\bar{u}_{c_{11}} = \bar{u} - c_1,\ \ \bar{v}_{c_{21}} = \bar{v} + c_2\ \ \bar{v}^*_{c_{32}} = \bar{v}^* + c_3)$.
If the samples contain u_{max} and $(v_{min},\ v^*_{min})$, then $(\bar{u}_{c_{12}} = \bar{u},\ \ \bar{v}_{c_{22}} = \bar{v},\ \ \bar{v}^*_{c_{32}} = \bar{v}^*)$ for all types of samples. Also $c_1,\ c_2,\ c_3$ are unknown constants. To obtain the biases and mean squared errors, we use the following relative errors terms and their expectations:

$$e_0 = \frac{\bar{u}_{c1} - \bar{Y}}{\bar{Y}}, \quad e_1 = \frac{\bar{u}_{c2} - \bar{X}}{\bar{X}}, \quad e_2 = \frac{\bar{v}^*_{c3} - \bar{R}_x}{\bar{R}_x},$$

such that $E(e_0) = E(e_1) = E(e_2) = 0$,

$$E(e_0^2) = \frac{\lambda}{\bar{Y}^2}\left[S_u^2 - \frac{2nc_1}{N-1}\left(u_{max} - u_{min} - nc_1\right)\right],$$

$$E(e_1^2) = \frac{\lambda}{\bar{X}^2}\left[S_v^2 - \frac{2nc_2}{N-1}\left(v_{max} - v_{min} - nc_2\right)\right],$$

$$E(e_2^2) = \frac{\lambda}{\bar{R}_x^2}\left[S_{v^*}^2 - \frac{2nc_3}{N-1}\left(v^*_{max} - v^*_{min} - nc_3\right)\right],$$

$$E(e_0e_1) = \frac{\lambda}{\bar{Y}\bar{X}}\left[S_{uv} - \frac{n}{N-1}\left\{c_2\left(u_{max} - u_{min}\right) + c_1\left(v_{max} - v_{min}\right) - 2nc_1c_2\right\}\right],$$

$$E(e_0e_2) = \frac{\lambda}{\bar{Y}\bar{R}_x}\left[S_{uv^*} - \frac{n}{N-1}\left\{c_3\left(u_{max} - u_{min}\right) + c_1\left(v^*_{max} - v^*_{min}\right) - 2nc_1c_3\right\}\right],$$

$$E(e_1e_2) = \frac{\lambda}{\bar{X}\bar{R}_x}\left[S_{uv^*} - \frac{n}{N-1}\left\{c_3\left(v_{max} - v_{min}\right) + c_2\left(v^*_{max} - v^*_{min}\right) - 2nc_2c_3\right\}\right].$$

Expressing (2.1) in terms of e_0, e_1 and e_2, we have

$$\hat{\bar{Y}}_{Rpps} = \bar{Y}(1 + e_0)(1 + e_1)^{-1}(1 + e_2)^{-1}. \tag{2.7}$$

Expressing (2.7) up to first order of approximation, we have

$$\hat{\bar{Y}}_{R(pps)} - \bar{Y} \cong \bar{Y}\left(e_0 - e_1 - e_2 + e_2^2 + e_1^2 + e_1e_2 - e_0e_2 - e_0e_1\right). \tag{2.8}$$

Taking expectation on both sides of (2.8), we have

$$\begin{aligned} Bias(\hat{\bar{Y}}_{R(pps)}) \cong \lambda\Bigg[&\frac{R_1}{\bar{X}\bar{R}_x}\left\{S_v^2 - \frac{2nc_2}{N-1}\left(v_{max} - v_{min} - nc_2\right)\right\} + R_2\Bigg\{S_{v^*}^2 \\ &- \frac{2nc_3}{N-1}\left(v^*_{max} - v^*_{min} - nc_3\right)\Bigg\} - R_1R_2\Bigg\{S_{uv^*} - \frac{n}{N-1}\{c_3(v_{max} \\ &- v_{min}) + c_2(v^*_{max} - v^*_{min}) - 2nc_2c_3\}\Bigg\} - R_1\Bigg\{S_{uv} - \frac{n}{N-1}\{c_2(u_{max} \\ &- u_{min}) + c_1(v_{max} - v_{min}) - 2nc_1c_2\}\Bigg\} - R_2\Bigg\{S_{uv^*} - \frac{n}{N-1}\{c_3(u_{max} \end{aligned}$$

$$- u_{min}) + c_2(v^*_{max} - v^*_{min}) - 2nc_2c_3\}\Big\}\Big].$$

Squaring (2.8), and then taking expectations, we have

$$\begin{aligned} MSE\left(\hat{\bar{Y}}_{R(pps)}\right) &\cong \Big[\lambda\left(S_u^2 + R_1^2S_v^2 + R_2^2S_{v^*}^2 + 2R_1R_2S_{uv} - 2R_2S_{uv^*}\right) \\ &\quad -\frac{2n\lambda\,(c_1R_2c_3 - R_1c_2)}{N-1}\Big\{(u_{max} - u_{min}) - R_1\,(v_{max} - v_{min}) \\ &\quad -R_2\,(v^*_{max} - v^*_{min}) - n\,(c_1 - R_2c_3 - R_1c_2)\Big\}\Big]. \end{aligned} \tag{2.9}$$

Differentiating (2.9) with respect to c_1, c_2, and c_3, we have

$$c_{1opt.} = \frac{u_{max} - u_{min}}{2n},$$

$$c_{2opt.} = \frac{v_{max} - v_{min}}{2n},$$

$$c_{3opt.} = \frac{v^*_{max} - v^*_{min}}{2n}.$$

Substituting the optimum value of c_1, c_2 and c_3 in (2.9), we get the minimum MSE of $(\hat{\bar{Y}}_{R(pps)})$ given by

$$\begin{aligned} MSE\left(\hat{\bar{Y}}_{R(pps)}\right)_{min} &\cong MSE\left(\bar{y}_{R(pps)}\right) - \frac{\lambda}{2(N-1)}\big[(u_{max} - u_{min}) - R_1\,(v_{max} - v_{min}) \\ &\quad -R_2\,(v^*_{max} - v^*_{min})\big]^2, \end{aligned} \tag{2.10}$$

where

$$MSE\left(\bar{y}_{R(pps)}\right) = \lambda\left[S_u^2 + R_1^2S_v^2 + R_2^2S_{v^*}^2 + 2R_1R_2S_{vv^*} - 2R_2S_{uv^*} - 2R_1S_{uv}\right].$$

Similarly, the bias and minimum MSE of product estimator in PPS sampling scheme is given by

$$\begin{aligned} Bias(\hat{\bar{Y}}_{P(pps)}) \cong \lambda\Big[&\frac{R_1}{\bar{X}\bar{R}_x}\Big\{S_v^2 - \frac{2nc_2}{N-1}\,(v_{max} - v_{min} - nc_2)\Big\} + R_2\Big\{S_{v^*}^2 \\ &- \frac{2nc_3}{N-1}\,(v^*_{max} - v^*_{min} - nc_3)\Big\} + R_1R_2\Big\{S_{uv^*} - \frac{n}{N-1}\{c_3(v_{max} \\ &- v_{min}) + c_2(v^*_{max} - v^*_{min}) - 2nc_2c_3\}\Big\} + R_1\Big\{S_{uv} - \frac{n}{N-1}\{c_2(u_{max} \\ &- u_{min}) + c_1(v_{max} - v_{min}) - 2nc_1c_2\}\Big\} + R_2\Big\{S_{uv^*} - \frac{n}{N-1}\{c_3(u_{max} \\ &- u_{min}) + c_2(v^*_{max} - v^*_{min}) - 2nc_{2c_3}\}\Big\}\Big] \end{aligned}$$

and

$$\begin{aligned} MSE\left(\hat{\bar{Y}}_{P(pps)}\right)_{min} &\cong MSE\left(\bar{y}_{P(pps)}\right) - \frac{\lambda}{2(N-1)}\big[(u_{max} - u_{min}) + R_1\,(v_{max} - v_{min}) \\ &\quad +R_2\,(v^*_{max} - v^*_{min})\big]^2, \end{aligned} \tag{2.11}$$

where

$$MSE\left(\bar{y}_{P(pps)}\right) = \lambda\left[S_u^2 + R_1^2 S_v^2 + R_2^2 S_{v^*}^2 + 2R_1R_2S_{vv^*} + 2R_2S_{uv^*} + 2R_1S_{uv}\right].$$

The minimum MSE of regression estimator in case of positive correlation is given by

$$\begin{aligned} MSE\left(\hat{\hat{Y}}_{lr1P(pps)}\right)_{min} &\cong MSE(\bar{y}_{lr(pps)}) - \frac{\lambda}{2(N-1)}\Big[(u_{max} - u_{min}) - \beta_1(v_{max} - v_{min}) \\ &\quad -\beta_2(v^*_{max} - v^*_{min})\Big]^2, \end{aligned} \tag{2.12}$$

where

$$MSE(\bar{y}_{lr(pps)}) = \lambda S_u^2\left(1 - \rho_{uv}^2 - \rho_{uv^*}^2 + 2\rho_{uv}\rho_{uv^*}\rho_{vv^*}\right)$$

and β_1 and β_2 are the population regression coefficients.

Similarly, the minimum MSE of $\hat{\hat{Y}}_{lr1P(pps)}$ in case of negative correlation is

$$\begin{aligned} MSE\left(\hat{\hat{Y}}_{lr1(pps)}\right)_{min} &\cong MSE(\bar{y}_{lr(pps)}) - \frac{\lambda}{2(N-1)}\Big[(u_{max} - u_{min}) + \beta_1(v_{max} - v_{min}) \\ &\quad +\beta_2(v^*_{max} - v^*_{min})\Big]^2. \end{aligned} \tag{2.13}$$

A general form for the MSE for the situation of both positive and negative correlation between the study and the auxiliary variable is given by

$$\begin{aligned} MSE\left(\hat{\hat{Y}}_{lr(g)pps}\right)_{min} &\cong MSE(\bar{y}_{lr(pps)}) - \frac{\lambda}{2(N-1)}\Big[(u_{max} - u_{min}) - |\beta_1|(v_{max} - v_{min}) \\ &\quad -|\beta_2|(v^*_{max} - v^*_{min})\Big]^2. \end{aligned} \tag{2.14}$$

3. Comparison of estimators

In this section, we compare the proposed estimators with usual ratio, product and regression estimators under PPS sampling scheme.

3.1. Condition (i)

By (1.4) and (2.10), we get

$$\left[MSE(\bar{y}_{R(pps)}) - MSE\left(\hat{\hat{Y}}_{R(pps)}\right)_{min}\right] \geq 0,$$

if

$$\left[(u_{max} - u_{min}) - R_1(v_{max} - v_{min}) - R_2(v^*_{max} - v^*_{min})\right]^2 \geq 0.$$

3.2. Condition (ii)

By (1.7) and (2.11), we obtain

$$\left[MSE(\bar{y}_{P(pps)}) - MSE\left(\hat{\hat{Y}}_{P(pps)}\right)_{min}\right] \geq 0,$$

if

$$\left[(u_{max} - u_{min}) + R_1(v_{max} - v_{min}) + R_2(v^*_{max} - v^*_{min})\right]^2 \geq 0.$$

3.3. Condition (iii)

By (1.9) and (2.14), it comes

$$\left[MSE(\bar{y}_{lr(pps)}) - MSE\left(\hat{\hat{Y}}_{lr(g)pps}\right)_{min}\right] \geq 0,$$

if

$$\left[\,(u_{max} - u_{min}) - |\beta_1|\,(v_{max} - v_{min}) - |\beta_2|\,(v^*_{max} - v^*_{min})\,\right]^2 \geq 0.$$

We observe that the proposed estimators perform better than the existing estimators if above conditions (i)–(iii) are satisfied.

4. Empirical study

We consider four real data sets for numerical comparisons which are described below.
Population 1: [Source: [14]]
y=Cultivation wheat in the region during 1964.
x=Cultivation wheat in the region during 1963.
z=Cultivated area in 1961.
R_x=Rank of the auxiliary variable (x).
$N = 34,\ n = 6$
$\bar{Y} = 199.4412,\ \ \bar{X} = 208.8824,\ \ \bar{Z} = 747.5882,\ \ \bar{R}_x = 17.47059,$
$C_u = 0.3630288,\ \ C_v = 0.3401158,\ \ C^*_v = 0.3598615,$
$C^2_u = 0.1317899,\ \ C^2_v = 0.1156788,\ \ C^2_{v^*} = 0.1295003,$
$\rho_{uv} = 0.7014932,\ \ \rho_{uv^*} = 0.4524924,\ \ \rho_{vv^*} = 0.5274431,$
$S_{uv} = 3387.898,\ \ S_{uv^*} = 233.6354,\ \ S_{vv^*} = 202.4327,$
$S^2_u = 3825.127,\ \ S^2_v = 3689.709,\ \ S^2_{v^*} = 27.96725,$
$\beta_1 = 0.918202,\ \ \ \beta_2 = 7.238203,$
$u_{max} = 315.2996,\ \ \ u_{min} = 46.72426,$
$v_{max} = 352.3937,\ \ \ v_{min} = 38.93689,$
$v^*_{max} = 32.81399,\ \ \ v^*_{min} = 7.457239.$

Population 2: [Source: Abreu-Lima and Joao(2009) Hospital, Faculty of Medicine, Porto University, Portugal].
VCG(vectocardiograms) data set:
http://archive.ics.uci.edu/ml/datasets/vcg.
y=The electrical axis corresponds to the half-area vector in the horizontal QRS Loop.
x=The electrical axis is determined by the amplitudes of the wave peaks (x, y) in the horizontal QRS Loop.
z=The electrical axis is determined by the time integrals ratio along x and y in the horizontal QRS Loop.
R_x=Rank of the auxiliary variable (x).
$N = 120,\ n = 15$
$\bar{Y} = 149.2458,\ \ \bar{X} = 151.0833,\ \ \bar{Z} = 154.3958,\ \ \bar{R}_x = 60.44167,$

$C_u = 0.7646985,\ \ C_v = 0.3704849,\ \ C_v^* = 0.5443165,$
$C_u^2 = 0.5847638,\ \ C_v^2 = 0.1372591,\ \ C_{v^*}^2 = 0.2962804,$
$\rho_{uv} = 0.2013749,\ \ \rho_{uv^*} = 0.06761159,\ \ \rho_{vv^*} = 0.1100916,$
$S_{uv} = 1327.191,\ \ S_{uv^*} = 325.0099,\ \ S_{vv^*} = 240.8604,$
$S_u^2 = 1918.083,\ \ S_v^2 = 1593.75,\ \ S_{v^*}^2 = 284.4695,$
$\beta_1 = 0.8327473,\ \ \ \beta_2 = 1.142512,$
$u_{max} = 1389.562,\ \ u_{min} = 0.441131,$
$v_{max} = 0.441131,\ \ v_{min} = 1.317745,$
$v^*_{max} = 463.1875,\ \ v^*_{min} = 0.4392485.$

Population 3: [Source: Herbert(2009), DEQ, Faculty of Engineering, Porto University, Portugal].
Wines dataset
http : //archive.ics.uci.edu/ml/datasets/wine.
y=Aspartame.
x=Leucine.
z=Isoleucine.
R_x=Rank of the Leucine.
$N = 67,\ n = 8$
$\bar{Y} = 23.63433,\ \ \bar{X} = 20.59851\ \ \bar{Z} = 9.792537,\ \ \bar{R}_x = 34,$
$C_u = 0.53356861,\ \ C_v = 0.6869816,\ \ C_v^* = 0.7242662,$
$C_u^2 = 0.2869596,\ \ C_v^2 = 0.4719437,\ \ C_{v^*}^2 = 0.5245616,$
$\rho_{uv} = 0.3171296,\ \ \rho_{uv^*} = 0.4236072,\ \ \rho_{vv^*} = 0.4236072,$
$S_{uv} = 71.13702,\ \ S_{uv^*} = 128.748,\ \ S_{vv^*} = 180.1726,$
$S_u^2 = 93.83272,\ \ S_v^2 = 107.6524,\ \ S_{v^*}^2 = 336.2949,$
$\beta_1 = 0.6608032,\ \ \ \beta_2 = 0.3828427,$
$u_{max} = 122.2254,\ \ u_{min} = 5.092119,$
$v_{max} = 131.6552,\ \ v_{min} = 0,$
$v^*_{max} = 213.9851,\ \ v^*_{min} = 1.052961.$
Population 4: [Source: [18]].
y=Number of tube wells.
x=Irrigated area (in hectares) for 69 villages of Doraha development block of Punjab, India.
z=Number of tractors.
R_x=Rank of the auxiliary variable (x).
$N = 69,\ n = 8,$
$\bar{Y} = 135.2609,\ \ \bar{X} = 345.7536,\ \ \bar{Z} = 21.23188,\ \ \bar{R}_x = 34.95652,$
$C_u = 0.6701015,\ \ C_v = 0.481404,\ \ C_v^* = 0.6560908,$
$C_u^2 = 0.449036,\ \ C_v^2 = 0.2317498,\ \ C_{v^*}^2 = 0.3076422,$
$\rho_{uv} = 0.1814586,\ \ \rho_{uv^*} = 0.1015504,\ \ \rho_{vv^*} = 0.4086618,$
$S_{uv} = 3442.406,\ \ S_{uv^*} = 1740.037,\ \ S_{vv^*} = 237.9886,$
$S_u^2 = 2529.325,\ \ S_v^2 = 17050.9,\ \ S_{v^*}^2 = 329.5376,$
$\beta_1 = 0.20189,\ \ \ \beta_2 = 0.7221896,$
$u_{max} = 859.8913,\ \ u_{min} = 57.45098,$
$v_{max} = 1167.754,\ \ v_{min} = 121.0217,$
$v^*_{max} = 115.5958,\ \ v^*_{min} = 4.246377.$

The results are given in Table 1.

Table 1. MSE values of all the considered estimators.

Estimator	Population 1	Population 2	Population 3	Population 4
$\bar{y}_{R(pps)}$	805.8873	143.7090	42.90262	1080.554
$\hat{\bar{Y}}_{R(pps)}$	550.8713	107.0965	11.79146	1079.505
$\bar{y}_{P(pps)}$	4503.025	707.3308	128.4614	1987.579
$\hat{\bar{Y}}_{P(pps)}$	4248.009	670.7184	97.3502	1986.530
$\bar{y}_{lr(pps)}$	406.7386	122.4856	10.3137	245.8052
$\hat{\bar{Y}}_{lr(g)(pps)}$	304.4522	88.85371	7.832164	55.42954

In Table 1, we observed that ($MSEs$) of the proposed estimators are smaller than the corresponding existing estimators, for all four populations. The performance of the regression estimator is the best among all other estimators.

5. Conclusions

In this paper, we have proposed ratio, product, and regression type estimators in presence of maximum and minimum values using the auxiliary variable and rank of the auxiliary variable X under PPS sampling scheme. The bias and mean squared error of the proposed estimators were derived under the first degree of approximation. Based on the theoretical and numerical investigations, it is observed that the proposed estimators are more efficient than the corresponding existing estimators for all populations which are used here. The performance of the suggested regression estimator is the best than existing estimators in terms of MSEs. Categorically, we recommend the use of our proposed estimators over the existing estimators considered in this paper for the new survey for estimating the finite population mean under probability proportional to size.

Acknowledgments

We appreciate the thoughtful comments and suggestions of the referees for enhancing the presentation of the article.

Conflict of interest

The authors declare no conflict of interest.

References

1. A. Y. Al-Hossain, M. Khan, Efficiency of ratio, product, and regression estimators under maximum and minimum values, using two auxiliary variables, *J. Appl. Math.*, **2014** (2014), 1–6.
2. S. Agarwal, P. Kumar, Combination of ratio and PPS estimators, *Journal of Indian Society of Agricultural Statistics*, **32** (1980), 81–86.

3. S. Ahmad, J. Shabbir, Use of extreme values to estimate finite population mean under PPS sampling scheme, *Journal of Reliability and Statistical Studies*, **11** (2018), 99–112.

4. X. Ding, H. Li, J. Lu, S. Wang, Optimal strategy estimation of random evolutionary boolean games, *IEEE T. Cybernetics*, 2021, doi: 10.1109/TCYB.2021.3050192.

5. L. K. Grover, P. Kaur, An improved estimator of the finite population mean in simple random sampling, *Model Assisted Statistics and Applications*, **6** (2011), 47–55.

6. S. Gupta, J. Shabbir, On improvement in estimating the population mean in simple random sampling, *J. Appl. Stat.*, **35** (2008), 559–566.

7. S. Hussain, S. Ahmad, S. Akhtar, A. Javed, U. Yasmeen, Estimation of finite population distribution function with dual use of auxiliary information under non-response, *Plos one*, **15** (2020), 1–24.

8. S. Hussain, S. Ahmad, M. Saleem, S. Akhtar, Finite population distribution function estimation with dual use of auxiliary information under simple and strati

   ed random sampling, *Plos one*, **15** (2020), 1–30.

9. S. Hussain, M. Zichuan, S. Hussain, A. Iftikhar, M. Asif, S. Akhtar, et al. On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling, *Journal of Probability and Statistics*, **2020** (2020), 1–13.

10. C. Kadilar, H. Cingi, Ratio estimators in simple random sampling, *Appl. Math. Comput.*, **151** (2004), 893–902.

11. C. Kadilar, H. Cingi, Improvement in estimating the population mean in simple random sampling, *Appl. Math. Lett.*, **19** (2006), 75–79.

12. M. Khan, J. Shabbir, Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values, *The Scientic World Journal*, **2013** (2013), 1–7.

13. Y. Li, H. Li, G. Zhao, Optimal state estimation for finite-field networks with stochastic disturbances, *Neurocomputing*, **414** (2020), 238–244.

14. M. N. Murthy, *Sampling theory and methods*, Statistical Publishing Society, 1967.

15. J. Rao, Alternative estimators in pps sampling for multiple characteristics, *Sankhya: The Indian Journal of Statistics, Series A*, **28** (1966), 47–60.

16. V. N. Reddy, T. J. Rao, Modified PPS method of estimation, *Sankhya*, **39** (1977), 185–197.

17. C. E. Särndal, Sample survey theory vs. general statistical theory: Estimation of the population mean, *International Statistical Review/Revue Internationale de Statistique*, **40** (1972), 1–12.

18. H. P. Singh, R. S. Solanki, An efficient class of estimators for the population mean using auxiliary information, *Commun. Stat. Theor. M.*, **42** (2013), 145–163.

19. R. Singh, N. S. Mangat, *Elements of survey sampling*, Springer Science & Business Media, 1996.

20. T. Srivenkataramana, D. Tracy, Transforming the study variate after PPS sampling, *Metron*, **37** (1979), 175–181.

21. P. V. Sukhatme, *Sampling theory of surveys with applications*, The Indian Society Of Agricultural Statistics, New Delhi, 1957.

22. T. Tripathi, A regression-type estimator in sampling with PPS and with replacement, *Aust. N. Z. J. Stat.*, **11** (1969), 140–148.

23. L. N. Upadhyaya, H. P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical J.*, **41** (1999), 627–636.

AIMS Press