ON MIXING OF RATIO TYPE ESTIMATORS

M.K. PANDE,SPERSH BHATT

Abstract: In literature several ratio type estimators have been proposed in order to estimate the mean of the study variable. In the present paper, some variants of ratio estimators are proposed by mixing the aforesaid estimators, following the technique of Vos(1980).Thereafter, we utilize the Monte Carlo Simulation Sampling from a hypothetical population that are bi-variate normal. The comparisons between the proposed and conventional estimators are hereby presented via computer aided empirical studies.

Keywords: Coefficient of variation, Empirical probability, Mean Square Error(MSE), Simulation

Introduction: Srivastava(1967), Reddy(1974) and Sahai(1979) proposed modified ratio type estimators to make them more efficientby incorporating a non stochastic design parameter. Thereby they investigated the possibility of gaining efficiency of proposed estimators over conventional ratio estimator. The algebraic comparisons given by them were not only inconclusive but were algebraically complicated and rather intractable. In the present study, modified estimators have been proposed using the technique of Vos(1980). The proposed estimators are then compared with the usual ratio and mean per unit estimator. Interestingly the proposed estimators turn out to be more efficient than the conventional estimators. We assume in our discussion that the sampled bi-variate population, could be, very nearly regarded as normal.Consequently, via computer program using Monte-Carlo simulation leading to a 1000 pseudo random samples of size 15, 30 & 50 from hypothetical bivariate population, we have computed the resultant actual MSE's and hence the percentage relative efficiencies of the estimators(relative to the mean per unit estimator).

The reason for computer simulation study is two fold. Firstly, a closed form of the algebraic expression of the estimators proposed by Srivastava(1967), Reddy(1974) and Sahai(1979) is unavailable. Moreover if we take a first order , ($o(n^{-1})$), large sample approximation to the MSE' of these estimators,the comparisons are algebraically quite intricate. Furthur the issue depends on many population parameters which are unknown. It is thus difficult to conclude which of these estimators is efficient and when.In casethe sample is so large so as to justify this first order approximation, regression estimator is better motivated than ratio or product type estimators. As such when the sample is just fairly large and the approximation of the regression estimate being rather unpredictable we are motivated to use the estimators like Srivastava(1967), Reddy(1974) and Sahai(1979) . Then we have to go for atleast a second order,($o(n^{-2})$), large sample approximation for the MSE's of these estimators. These approximations to the MSE's turn out to be algebraically intricate and comparisons are just impossible. Hence computer simulation study is the only alternative in case of the estimators quoted above. The same follows for the estimators proposed by us in this paper. It may be pointed out in this context that the regression estimator dominates all other estimators. It cannot be decided easily upto which extent the relative advantage of the regression estimator can compensate the difficulty of calculating it for surveys which have complicated designs.

Using information on one or more auxiliary variable often ratio and product type estimators are used to estimate the population mean, toatl etc. In this context various estimation techniques have been studied by Goodman and Hartley(1958), Tin(1965), Rao(1967), Hutchinson(1971),Rao(1981), Royall and Cumerland(1981) among others. Tuschprow(1953) and Naylor, Balintfy and Chu(1969) have utilized relevant computer techniques which are useful in the present context.

For simplicity of illustration, we assume that the auxiliary information is available on only one auxiliary variable say X, which is positivley correlated with the study variable say, Y. $\overline{X}$ and $\overline{Y}$($\bar{x}$ and $\bar{y}=d_1$ say) be the population(sample) means of the two variables. A simple random sample of size n is taken without replacement and is assumed to have come from a hypothetical bi-variate normal population.

The conventional ratio estimator is given as $(\bar{y}/\bar{x})\overline{X} = d_2$. Let in addition $C(Z)=C(Y)/C(X)$ and $C=R\{C(Y)/C(X)\}$ and ignoring the finite population correction factor the MSE of estimator d_2 is given as

$M_1(d_2)=V_1[1+\{(1-2C/C^2(Z))\}]$

Where $V_1=Var(d_1)= Y^2[C^2(Y)/n]$

And as mentioned earlier $\bar{y}=d_1$

The relative efficiency(REF) of d_2 with respect to d_1 to the first degree of approximation is given as

$REF(d_2)=M_1(d_1)/M_2(d_2)$

The possibility of more efficient estimation of$\overline{Y}$ motivated the proposition of the following estimators

$d_3= d_1\overline{X}/\{\overline{X}+ a(\overline{X}-\bar{x})$ Reddy(1974)

$d_4= d_1 (\overline{X}/\bar{x})^a$ Srivastava(1967)

$d_5= d_1\{(1-a)\bar{x} + (1+a)\overline{X}\}/\{(1+a)\bar{x} + (1-a)\overline{X}\}$ Sahai(1979)

To the first degree of approximation

$M_1(d_3)=M_1(d_5)= V_1.P$ where $P=1 + a(a-2C)/C^2(Z)$ and 'a' is the non stochastic design parameter.

The estimators d_4 and d_5 are self dual but d_3 does not possess this characteristic so we define an estimator which is dual to d_3 as follows

$d_6 = (1-a)d_1 + ad_2$ (1.8)

Sahai and Ray(1980) proposed the following family of estimators

Sahai and Ray(1980) proposed the following family of estimators

$d_7 = 2d_1 - d_1 (\bar{x}/\bar{X})^a$ (1.9)

Here again 'a' is the non stochastic design parameter. The maximum value of MSE of the above estimators is obtained when a=C.

Let the relative error in guessing C be denoted by REG(C)=(g-C)/C. Here g is the guessed value of C which is assumed to be unknown. In our study we replace a by g in (1.5), (1.6), (1.7), (1.9) and (1.10). We know that product estimator is preferred when C < -0.5 and the ratio estimator when C > 0.5. The usual unbiased estimator is thrown rather in between the two in the case when C ϵ (-0.5,0.5).

The Estimators:

The following estimators along with the already mentioned ones are taken up in the simulation study.

$D_1 = (1-a)d_3 + ad_4$ (2.1)

$D_2 = ad_3 + (1-a)d_4$ (2.2)

$D_3 = (1-a)d_3 + ad_5$ (2.3)

$D_4 = ad_3 + (1-a)d_5$ (2.4)

$D_5 = (1-a)d_4 + ad_5$ (2.5)

$D_6 = ad_4 + (1-a)d_5$ (2.6)

$D_7 = (1-a)d_6 + ad_4$ (2.7)

$D_8 = ad_6 + (1-a)d_4$ (2.8)

$D_9 = (1-a)d_6 + ad_3$ (2.9)

$D_{10} = ad_6 + (1-a)d_3$ (2.10)

$D_{11} = (1-a)d_6 + ad_7$ (2.11)

$D_{12} = ad_6 + (1-a) d_7$

$D_{13} = (1+a)d_1 - ad_3$ (2.12)

(2.13)

Here D_{13} being a linear combination of d_1 and d_3.

$D_{14} = (1+a)d_1 - ad_5$ (2.14)

Again, D_{14} is a linear combination of d_1 and d_5

The minimum value for MSE of both D_{13} and D_{14} is obtained when a= ±√C. Thus we replace 'a' by its guessed value ±√g in (2.13) and (2.14). We thus obtain two more estimators corresponding to D_{13} and D_{14}, viz $D_{13.1}$ and $D_{13.2}$ ($D_{14.1}$ and $D_{14.2}$).

Considering the case when R < 0 , the product estimator is better than the ratio estimator. Here too similar computer aided efficiency comparisons can be carried out for construction of suitable product type estimators using Vos(1980).

The Simulation Study

In the following simulation study, the random variables are generated via the central limit theorem using uniformly distributed numbers between 0 and 1. To increase the reliability of simulation we have used Tschuprow(1953) and Naylor, Burdik and Chu(1969) approximation techniques to care of the tail probabilities.

We have assumed that the sampled bi- variate population could be very nearly regarded as normal. We consider four example levels of $\bar{Y}$ (1.0, 2.0, 3.0, 4.0): two each for population standard deviations σ(X) and σ(Y) namely σ(X) = 1.0, 2.0; σ(Y) = 2.0, 4.0. The value of $\bar{X}$ is fixed viz $\bar{X}$ = 1.0 . This is because we can divide each observation corresponding to the auxiliary variable by $\bar{X}$(assumed to be known) and carry out the study with the new set of auxiliary observations. Together with this, five values of REG(C) = -0.2, -0.1, 0.0, 0.1, 0.2 are taken up. Correspondingly we take five values of R(0.2, 0.4, 0.6, 0.8, 0.9). Thus with the example levels mentioned above of $\bar{Y}$,σ(X) , σ(Y), REG(C) and R we get 400 parametric value combinations, 100 pseudo-random samples each of size 15, 30 and 50 are generated using Box Muller(1958) approach. We compute the resultant actual MSE's and hence the relative efficiencies of estimators. Thus we have 400 X 3(sample sizes:15, 30 and 50) occasions of comparing the estimators.

Let E(.) be the empirical probability which is found out by counting the number of times a particular estimator has the maximum relative efficiency.

With reference to the tables given below we note the number of times a particular estimator has the maximum efficiency. We define E(.) as follows

E(.) = Empirical probability of being the winner

The comparison of the estimators is done for five levels of REG(C) viz. -0.2, -0.1, 0.0, 0.1 and 0.2. A finer description of the estimators with regard to empirical probability is provided in the tables below for different levels of over/ under guess of C

. Table 3.1

Estimator →	d_1	d_2	D_1
REG(C)↓			
-0.2	0.1240	0.1926	0.6833
-0.1	0.1168	0.1000	0.7831
0.0	0.1075	0.0714	0.8210
0.1	0.0796	0.0921	0.8282
0.2	0.0545	0.0955	0.8499

E(.)'s for d_1, d_2, and D_1, R > 0

From the above table we conclude that when corrrelation is positive the estimator D_1 is strongly recommended as compared to d_1 and d_2 in case C is overguessed. The highest E(.) attained by D_1 is 0.8499.

Table 3.2

E(.)'s for d_1, d_2, and D_1, R > 0

Estimator	d_1	d_2	D_1
REG(C)↓			
-0.2	0.1718	0.2161	0.6120
-0.1	0.1621	0.1150	0.7228
0.0	0.1457	0.0959	0.7583
0.1	0.1000	0.1250	0.7750
0.2	0.0705	0.1127	0.8167

From table 3.2 it follows that for estimator D_2 overguess is better than under guess. The maximum probability being 0.8167 when REG(C) = 0.2. The table 3.3 depicts one by one the performance of estimators from D_3 to D_{12} in addition to $D_{13.1}$, $D_{13.2}$, $D_{14.1}$ and $D_{14.2}$ vis a vis d_1 and d_2.

A brief idea of results obtained therein is presented in the table. The first row shows the nature of the guess of C (under guess or over guess) when the proposed estimator is better than d_1 and d_2. The second row depicts the maximum efficiency while the third row mentions the corresponding values of REG(C).

Hence more often than not the proposed estimator exhibit superiority over their rivals viz sample mean and ratio estimator.

We consider the six sub ranges viz

Range 1: 0.0 < C < 0.2 Range 2: 0.2 < C < 0.4

Range 3: 0.4 < C < 0.6 Range 4: 0.6 ≤ C < 0.8

Range 5: 0.8 ≤ C < 1.0

Range 6: Integrated range C > 1.0

Range 1: 0.0 < C < 0.2

The most probable winner is the estimator $D_{14.1}$. For REG(C) = 0.0, E(.) is maximum viz 0.5714. The overguess is more favourable than the underguess(of C) for this estimator.As expected the ratio estimator is out of competition because C > 0.5.

Range 2: 0.2 < C < 0.4

AgIn in this range $D_{13.1}$ is the most probable winner. Maximum value of E(.) for $D_{14.1}$ is 0.3958 for REG(C) = -0.1. In this case, underguess is more favorable to the estimator than the over guess of C. The second best estimator is $D_{14.1}$, which stands higher chance of winning when C is moderatley over guessed (REG(C) = 0.1). Once again, as expected ratio estimator is out of competition because C > 0.5

Range 3: 0.4 < C < 0.6

In this case as C is not far from 0.5 , the ratio estimators out of competition. The estimator $D_{14.1}$ has been able to take over the estimator $D_{13.1}$ and turns out to be the best, most probably.Underguess of C is more favorable to the winning estimator than overguess. Chance of winning happens to be maximum for the estimator when C is highly underguessed.REG(C) = -0.2)

Range 4: 0.6 < C ≤ 0.8

The estimator $D_{14.1}$ conolidates its winning spirit and the probability profile of its winning over the range of REG(C) is again more responsive to the underguess. As C > 0.5, the ratio estimator has been able to show up but only in one case when REG(C) = 0.2.

Range 5: 0.8 ≤ C < 1. 0

In this range the ratio estimator performs very nicely, so much so that it turns out to be the best when C is highly overguessed.[REG(C) = 0.2]. The estimator $D_{14.1}$ which had been the winner in other subranges, sustains itself only in situation when C is highly underguessed. When C is moderatley underguessed or exactly guessed, the most probable winner is the ratio estimator and when C is moderatley overguessed, D_4 is better than D_2 most probably. We illustrate briefly the results for estimators scoring appreciable values of E(.) along with d_1 and ratio estimator d_2 by presenting a part of the full table

, Table 3.3
Efficiency Comparison Of Estimator, R > 0
Proposed Estimator(In Competititon With Sample Mean And Ratio Estimator)

Estimator→	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	$D_{13.1}$	$D_{13.2}$	$D_{14.1}$	$D_{14.2}$
Nature of the guess of C when proposed estimator is better	OG	OG	OG	OG	OG	OG	OG	OG & UG	OG & UG	OG & UG	OG & UG	OG & UG	OG & UG	OG & UG	OG	OG & UG
Max E (.) REG(C)	0.8500 0.2	0.8166 0.2	0.9250 0.0	0.8250 0.0	0.7833 0.2	0.7833 0.0	0.7625 0.0	0.9250 0.0	0.9208 0.0	0.9208 0.0	0.9208 0.0	0.9250 0.0	0.8583 0.0	0.6432 0.2	0.9208 0.0	0.9250 0.0

Table 3.4
E(.)'s for estimators , R > 0, 0.8≤ C < 0.1

Estimator REG(C)↓	d_1	d_2	D_2	D_4	D_8	$D_{13.1}$	$D_{14.1}$
-0.2	0.0305	0.0	0.3937	0.0	0.0303	0.0909	0.4244
-0.1	0.0194	0.0395	0.3925	0.0	0.0198	0.1178	0.2448
0.0	0.0194	0.0395	0.2745	0.0392	0.0395	0.1373	0.2156
0.1	0.0	0.1113	0.3333	0.3885	0.0	0.0	0.0556
0.2	0.05	0.5	0.0	0.0	0.0	0.0833	0.3335

Range 6: C > 1.0

In this range , we find that the only probability for the conventional ratio estimator to come up is provided by the cases, when C is under guessed. However the most probable winner in such situations is the estimator D_6 which is also best suited when REG(C) = 0.0. The estimator D_{10} is the winner when the overguess is moderate and D_3 when the overguess is high.

A part of the table presented below depicts along with empirical probability of d_1 and d_2, the probability pattern of estimators which are top winners.

Table 3.5
E(.)'s for estimators , R > 0
C > 1.0

Estimator REG(C)↓	d_1	d_2	D_3	D_4	D_6	D_{10}
-0.2	0.0	0.1815	0.0	0.2578	0.3030	0.0606
-0.1	0.0	0.1251	0.0	0.0833	0.3333	0.0833
0.0	0.0	0.0415	0.0416	0.0416	0.2085	0.1455
0.1	0.0205	0.0	0.0205	0.0417	0.6251	0.1042
0.2	0.0278	0.0	0.4722	0.0278	0.0278	0.0557

Total Range: C > 0

The table constructed for C > 0 , indicates that the estimator $D_{14.1}$ is most probable winner in entire range except when C is overguessed , when $D_{13.1}$ has edge over $D_{14.1}$.

References:

1. Tschpurow, D.(1953). Distribution of sampling with high speed computer. Ph.D Thesis, University of North Carolina.
2. Box, G.E.P and Muller, M.E (1958). A note on the generation of Normal Deviates. Ann. Math. Statist, 610-11
3. Goodman, L.A and Hartley, H.O(1958). The precision of unbiased ratio type estimators. Jour. Amer. Stat. Assoc., 53, 491-508
4. Tin, M.(1965). Comparison of some ratio estimators. Jour. Amer. Stat. Assoc.,60, 294-307
5. Srivastava, S.K.(1967). A class of estimators using auxiliary information in sample surveys. Canad. Jour. Statist., 8, 253-254
6. Naylor, T. H., Blaintfy, J.L., Burdick, D.S. and Chu, K.(1969). Computer Simulation Techniques. John Wiley and Sons Inc., New York, London, Sydney.
7. Reddy, V.N.(1974). On a transformed ratio method of estimation., Sankhya(C), 36, 59-70
8. Royall, R.M. and Cumberland, W.G.(1981). An empirical study of the ratio estimator and the estimators of its variance., Jour. Amer. Stat. Assoc., 76, 66-68
9. Sahai, A.(1979). An efficient variant of the ratio and product estimators. Statistica Neerlandica, 33, 27-35

Vos, J.W.E.(1980). Mixing of direct, ratio and product method estimators. Statistica Neerlandica, 34(4), 209-218

***

Professor and Head, Department of Statistics, DSB Campus, Kumaun University,Nainital. 263002
e-mail: pandemanoj8@gmail.com
Assitant Professor ,Department of Statistics,DSB Campus, Kumaun University,Nainital. 263002
e-mail: kustatntl@gmail.com