Model-Assisted Nonnegative Variance Estimator of the Ratio Estimator under the Midzuno-Sen Sampling Scheme

Praful A Patel
Sardar Patel University
Vallbh Vidyanagar 388 120
Gujarat, India
patelpraful_a@indiatimes.com

Jigna S Patel
Sardar Patel University
Vallbh Vidyanagar 388 120
Gujarat, India
jigna.stat@gmail.com

Abstract

This article suggests a model-assisted variance estimator for the ordinary ratio estimator under the Midzuno-Sen sampling scheme. To compare the suggested estimator empirically with available estimators a Monte Carlo comparison is carried out using real populations. The suggested estimator has performed very well and has taken non-negative values with probability 1.

Keywords: *M-S sampling scheme, Model-Assisted estimator, Non-negative variance estimation.*

1. Introduction

A model-based approach for estimation of finite population parameters has been used by many authors, see, e.g., Brewer (1963), Royall (1970), Ghosh and Meeden (1997), Valliant et al. (2000), Rao (2003), among others. Valliant et al. (2000) give a comprehensive account of the theory, including estimation of the (conditional) model variance of the estimator which varies with sample. By imposing a super population model on the actual finite population, several model-based variance estimators for the ratio estimator under SRSWOR have been proposed and studied (see, Royall and Eberhardt (1975), Royall and Cumberland (1978, 1981a, 1981b)). Rao (1972, 1977) addressed the problem of variance estimation for the ratio estimator under the Midzuno–Sen (M-S) sampling scheme whereas Chaudhuri (1976, 1981), Rao and Vijayan (1977), Rao (1979), and Chaudhuri and Arnab (1981) studied the problem of non-negative variance estimation for the same and proposed sufficient conditions for non-negativity. Recently, Patel and Patel (2009,10) have suggested a number of variance estimators for the ratio estimator under the M-S sampling. The estimation of the variance of the ratio estimator is still not resolved. In this article we focus on the same problem.

Consider a finite population $U=\{1,...,N\}$ of N identifiable units together with unknown values $\underline{y}=(y_1,...,y_N)$ and known values $(x_1,...,x_N)$ from which a sample s of size n is to be selected using a sampling design $p(s)$ with positive inclusion probabilities $\pi_i=P(i\in s)$ and $\pi_{ij}=P(i,j\in s)$ for every i and j. If $A\subseteq U$, we write Σ_A for $\Sigma_{i\in A}$ and $\Sigma\Sigma_A$ for $\Sigma\Sigma_{i\neq j\in A}$. We seek to estimate the variance of the ratio estimator $\hat{Y}_R=\dfrac{\sum_s y_i}{\sum_s p_i}$ of the population total $Y=\sum_U y_i$, where $p_i=\dfrac{x_i}{X}$ with $X=\Sigma_U x_i$, $(i=1,2,...,N)$. The design variance of $\hat{Y}_R$, suggested by Midzuno (1950), is given by

$$V_1\left(\hat{Y}_R\right)=\sum\nolimits_U \Lambda(s,ii)y_i^2+\sum\sum\nolimits_U \Lambda(s,ij)y_i y_j$$

where, for $i,j=1,...,N,\ (i\neq j)$,

$$\Lambda(s,ij)=\left(\frac{1}{M_1}\sum_{s\ni i}\frac{1}{P_s}-1\right)\ \text{if } i=j \ \text{ and } \ =\left(\frac{1}{M_1}\sum_{s\ni i,j}\frac{1}{P_s}-1\right)\ \text{if } i\neq j$$

with $P_s=x_s/X$, $x_s=\sum_s x_i$, $M_k=\begin{pmatrix}N-k\\ n-k\end{pmatrix}$, $k=1,2,...$

Rao (1972) proposed

$$v(\hat{Y}_R)=\sum\nolimits_s \Lambda(s,ii)\frac{y_i^2}{\pi_i}+\sum\sum\nolimits_s \Lambda(s,ij)\frac{y_i y_j}{\pi_{ij}}=v_1 \text{ (say)}$$

as an unbiased estimator of V_{YR}.

The following variance estimators of the ratio estimator under the M-S sampling scheme are available in the literature.

$$v_{11}=\sum\sum_{i<j\in s}a_{ij}\frac{(N-1)X}{(n-1)x_s}\left\{1-\frac{1}{M_1}\sum_{s\ni i,j}\frac{X}{x_s}\right\} \quad \text{(cf. Chaudhuri, 1981)}$$

$$v_{12}=v_{32}=\sum\sum_{i<j\in s}a_{ij}\left\{\frac{M_1X}{M_2x_s}-\frac{1}{\pi_{ij}M_1}\sum_{s\ni i,j}\frac{X}{x_s}\right\} \quad \text{(cf. Chaudhuri, 1981)}$$

$$v_{23}=\sum\sum_{i<j\in s}a_{ij}\left\{\frac{1}{\pi_{ij}}-\frac{X}{M_2x_s}\sum_{s\ni i,j}\frac{X}{x_s}\right\} \quad \text{(cf. Chaudhuri, 1981)}$$

$$v_{10}=v_{30}=\sum\sum_{i<j\in s}a_{ij}\left\{\frac{1}{M_2P(s)}-\frac{P(s/i).P(s/j)}{(P(s))^2}\right\} \quad \text{(cf. Rao and Vijayan, 1977)}$$

$$v_{20} = \sum\sum_{i<j\in s} a_{ij}\left\{\frac{1}{\pi_{ij}} - \left(\frac{X}{x_s}\right)^2\right\}$$ (cf. Vijayan et al., 1995)

$$v_{22} = \sum\sum_{i<j\in s} \frac{a_{ij}}{\pi_{ij}}\left\{1 - \frac{1}{M_1}\sum_{s\ni ij}\frac{X}{x_s}\right\}$$ (cf. Rao and Vijayan, 1977)

where

$$a_{ij} = \left\{\frac{y_i}{p_i} - \frac{y_j}{p_j}\right\}^2 p_i p_j .$$

The rest of the article is arranged as follows. Section 2 suggests a new model-based variance estimator of the ratio estimator. A Monte Carlo comparison of the suggested estimator with the available estimators is made in Section 3. The conclusions are given in Section 4.

2. The Model-Assisted Variance Estimator

For the ratio estimator under SRSWOR, some theoretical and empirical evidence (Wu, 1982; Wu & Deng, 1983) shows that the method of obtaining variance estimator just by replacing the population quantities in the variance expression by their sample analogues can be inefficient and that a simple modification of the standard variance estimator can bring about substantial gains in efficiency. Motivated by this, in this section, we construct an estimator of the variance of the ratio estimator, under the Midzuno-Sen sampling, by attaching the design weights to the model-based variance estimator.

Consider the model

$$Y_i = \beta x_i + e_i, \quad e_i \sim (0, \psi_i), \quad i = 1,\dots,N \tag{1}$$

with $\psi_i = \sigma^2 x_i$. It is well-known that the weighted least squares estimator of $\beta = \sum_U y_i \big/ \sum_U x_i$ is $\hat\beta = \dfrac{\sum_s Y_i}{\sum_s x_i}.$ Since $\hat Y_i = \hat\beta x_{i,}\ \ i \in s,$ the best linear unbiased predictor (BLUP) of $Y = \sum_s Y_i + \sum_{\bar s} Y_j$ is the ratio estimator

$$\hat Y_{BLUP} = \sum_s Y_i + \sum_{\bar s} \hat Y_j = \frac{\sum_s Y_i}{\sum_s x_i} X = \hat Y_R$$

where $\bar s = \{1,\dots,N\} - s.$ As the fitted values, dependent on sample y-values, are independent of the non-sample y -values, the error variance of $\hat Y = \sum_s Y_i + \sum_{\bar s} Y_j$, under Model (1), is written as

$$Var_M(\hat Y - Y) = V_I + V_{II}$$

where $V_I = Var_M(\sum_{\bar{s}} \hat{Y}_i)$ and $V_I = Var_M(\sum_{\bar{s}} Y_j)$

But, under Model (1), V_I becomes

$$V_I = \left(\frac{\sum_{\bar{s}} x_j}{\sum_s x_i} \right)^2 \sum_s \psi_i$$

and, therefore, the sandwich estimator of it is given by

$$v_I = \left(\frac{\sum_{\bar{s}} x_j}{\sum_s x_i} \right)^2 \sum_s r_i^2 \tag{2}$$

where $r_i = Y_i - \hat{\beta} x_i, i \in s$, the sample fit residuals. Again, under Model (1), V_{II} becomes

$$V_{II} = \sum_{\bar{s}} \psi_j$$

Since we don't have residuals for the non-sample units, as $Y_j, j \in \bar{s}$ are unobservable, and so a straight plug-in estimator id not possible. A reasonable strategy takes a plug-in estimator of $\sum_s \psi_i$ under the working model, and blows it up to non-sample size, using a multiplier

$$\frac{Var_M(\sum_{\bar{s}} Y_j)}{Var_M(\sum_s Y_i)} = \frac{\sum_{\bar{s}} x_j}{\sum_s x_i}$$

In this way we get an estimator of V_{II} as

$$v_{II} = \left(\frac{\sum_{\bar{s}} x_j}{\sum_s x_i} \right) \sum_s r_i^2 \tag{3}$$

Combining (2) and (3) the model-based estimator $\hat{Y}$ is then given by

$$v_M = v_I + v_{II} = \left(\frac{\sum_{\bar{s}} x_j}{\sum_s x_i} \right)^2 \sum_s r_i^2 + \left(\frac{\sum_{\bar{s}} x_j}{\sum_s x_i} \right) \sum_s r_i^2 \tag{4}$$

Injecting the design weights in v_M, given in (4) we suggest the following estimator for the variance of the ordinary ratio estimator as

$$v_{MA} = \left(\frac{\sum_{\bar{s}} x_j}{\sum_s \frac{x_i}{\pi_i}} \right)^2 \sum_s \frac{r_i^2}{\pi_i^2} + \left(\frac{\sum_{\bar{s}} x_j}{\sum_s \frac{x_i}{\pi_i}} \right) \sum_s \frac{r_i^2}{\pi_i} \tag{5}$$

Remark: We do not assume (1) to be the true superpopulation model that generates y-values. As our approach is design-based, $\beta = \sum_U y_i / \sum_U x_i$ is estimated by the sample weighted analogue $\hat{\beta} = (\sum_s y_i/\pi_i)/(\sum_s x_i/\pi_i)$. In this case the resultant model-based predictor of Y becomes the generalized ratio estimator $\hat{Y}_{GR} = \frac{\sum_s Y_i/\pi_i}{\sum_s x_i/\pi_i}$. However, we have taken (5) as the variance estimator of the ordinary ratio estimator under the M-S sampling.

3. Simulation Study

Recently, Patel and Patel (2009) show empirically that the estimators v_{12}, v_{13} and v_{22} have performed well among available estimators, and therefore, here we include these estimators for the comparison. Calculations of variance and variance estimates for the ratio estimator become more and more cumbersome with increasing population and sample size. For this reason we prefer to restrict ourselves to carry out simulation for small populations. From each population listed in Table 4 of Appendix A, a sample of size $n = 4$ is drawn using the M-S sampling design. This process is repeated $M = 10{,}000$ times. The suggested variance estimator v_{MA} given in (5) along with v_1, v_{12}, v_{13} and v_{22} given at (2) are computed for each sample and their performances are measured in terms of the relative efficiency (RE), the relative biased (RB) in percentage and the probabilities of taking negative values. The simulated results are presented in Table 1, Table 2 and Table 3, respectively.

For each variance estimate v, its relative percentage bias is calculated as

$$RB(v) = 100 * \frac{\bar{v} - V_1}{V_1} ,$$

the relative efficiency as

$$RE(v) = \frac{MSE(v_1)}{MSE(v)}$$

where $\bar{v} = \frac{1}{M}\sum_{j=1}^{M} v_{(j)}$, $MSE(v) = \frac{1}{M}\sum_{j=1}^{M}(v_{(j)} - V_1)^2$

Table 1: RE under the M-S sampling scheme

No.	v_1	v_{MA}	v_{22}	v_{13}	v_{12}	No.	v_1	v_{MA}	v_{22}	v_{13}	v_{12}
1	1	5.425	2.348	2.728	3.228	12	1	35.787	39.568	46.821	36.671
2	1	3.145	0.911	1.145	1.606	13	1	50.812	29.968	35.526	40.960
3	1	13.214	8.690	9.347	9.188	14	1	31.987	17.524	19.918	20.781
4	1	18.413	7.365	8.597	10.272	15	1	43.925	26.859	30.658	33.065
5	1	19.202	7.517	8.626	9.593	16	1	72.948	30.426	35.488	40.199
6	1	46.926	24.422	27.948	34.450	17	1	1872.179	1227.682	1236.281	1236.112
7	1	31.893	17.708	21.554	24.085	18	1	142.423	51.173	59.935	79.728
8	1	44.633	23.422	25.854	23.881	19	1	51.798	60.912	65.196	53.998
9	1	55.737	20.942	26.719	35.043	20	1	251.003	254.109	248.005	179.166
10	1	56.475	18.701	22.987	28.665	21	1	1898.046	498.027	627.216	898.607
11	1	70.971	34.808	41.255	47.515	22	1	3368.628	3002.737	3077.814	3205.117

From the simulation study the following are noticed:

(1) The estimators $v_{11} \equiv v_{13}, v_{12}, v_{22}$ are comparable among each other. However we can rank them from RE point of view as follow:

$$v_{22} \succ\succ v_{13} \succ\succ v_{12}.$$

However, v_{12} has taken less no. of times the negative value.

(2) The model-based estimator v_{MA} is the only non-negative estimator performed very well for all the populations considered here. However, it has the large absolute RB% compare to other and underestimate the true variance.

(3) The gain in efficiency of v_{MA} increases with increasing the correlation between the study variable and auxiliary variable.

Conclusions

Many researchers have studied the problem of variance estimation for the ordinary ratio estimator under the M-S sampling scheme and suggested various estimators. Also, they have studied non-negativity of their estimators and have derived necessary conditions. However, this problem of non-negative variance estimation is still not resolved. In this article we make an attempt to study the empirical properties of the estimators available in the literature and the estimators suggested in this article.

In many survey populations, the relationship between y_i and x_i is often a straight line through the origin with a general variance structure $v(y_i) = \sigma_i^2 = \sigma^2 x_i^\gamma, 1 \le \gamma \le 2$, which is usually the case in survey populations. Exploiting this relationship some variance estimators of H-T estimator have been suggested and to compare their performances empirical study has been carried out. The conclusions emerging from the simulation study can be summarized as follow.

(1) Implementation of the suggested estimators requires the complete auxiliary information, i.e., values of x variables for the entire finite populations.

(2) This limited simulation seems to suggest v_{MA} as the preferred variance estimator in conjunction with the routine use of the ratio estimator under the M-S sampling.

(3) The estimators $v_{11} \equiv v_{13}, v_{12}, v_{22}$ are comparable among each other. However we can rank them from RE point of view as follow:

$v_{22} \succ\succ v_{13} \succ\succ v_{12}$.

Further empirical work through Monte Carlo simulation for different populations is need to assess the properties of v_{MA}.

Appendix: A

Table 2: RB(%) under the M-S sampling scheme

No.	v_1	v_{MA}	v_{22}	v_{13}	v_{12}	No.	v_1	v_{MA}	v_{22}	v_{13}	v_{12}
1	6.359	-29.250	-5.897	-4.897	-1.224	12	-20.567	-6.459	4.983	4.484	2.675
2	-10.554	-38.669	14.544	11.103	1.407	13	0.538	-26.291	2.685	2.540	0.379
3	9.649	-39.582	-18.316	-17.274	-14.023	14	35.196	-13.196	5.782	6.141	7.027
4	-42.311	-34.360	8.208	6.173	-0.094	15	-77.145	-15.607	9.770	8.231	1.443
5	-11.366	-32.982	9.312	8.574	5.947	16	55.711	-33.421	2.200	2.659	3.711
6	33.770	-32.409	-12.728	-11.451	-7.068	17	222.545	-24.938	-1.453	-1.251	-0.559
7	34.873	-30.025	-6.676	-4.713	2.107	18	21.895	-35.317	0.992	1.512	3.163
8	59.674	-19.113	0.296	2.252	9.715	19	274.606	5.021	-6.929	-3.875	12.749
9	-88.049	-37.916	13.008	9.245	-1.283	20	175.971	-30.673	-9.938	-9.002	-5.198
10	-56.075	-33.561	10.643	8.551	2.536	21	-482.437	-53.316	12.880	9.878	2.590
11	-16.087	-33.829	-1.317	-1.366	-1.745	22	365.689	-25.168	-26.849	-25.433	-16.234

Table 3: Probability of taking negative values

No.	ν_1	ν_{MA}	ν_{22}	ν_{13}	ν_{12}	No.	ν_1	ν_{MA}	ν_{22}	ν_{13}	ν_{12}
1	0.1855	0.000	0.022	0.019	0.007	12	0.3186	0.000	0.029	0.025	0.017
2	0.1734	0.000	0.021	0.018	0.003	13	0.3011	0.000	0.037	0.032	0.012
3	0.2626	0.000	0.004	0.003	0.001	14	0.288	0.000	0.019	0.017	0.005
4	0.3458	0.000	0.003	0.002	0.000	15	0.3664	0.000	0.005	0.003	0.002
5	0.2975	0.000	0.014	0.010	0.001	16	0.3278	0.000	0.051	0.045	0.018
6	0.3029	0.000	0.034	0.030	0.002	17	0.4424	0.000	0.000	0.000	0.000
7	0.2531	0.000	0.035	0.031	0.001	18	0.3177	0.000	0.068	0.056	0.014
8	0.3043	0.000	0.024	0.021	0.012	19	0.2493	0.000	0.063	0.056	0.016
9	0.4069	0.000	0.030	0.024	0.007	20	0.371	0.000	0.000	0.000	0.000
10	0.3875	0.000	0.032	0.029	0.008	21	0.4359	0.000	0.048	0.038	0.005
11	0.3405	0.000	0.045	0.039	0.008	22	0.3258	0.000	0.113	0.097	0.051

Table 4: Study Population (Small)

Population	Source	x	y	N	CV(x)	CV(y)	ρ(x,y)
1	Murthy(1967).p.130 (107-128)	area in sq. miles	no. of cultivators	22	0.509	0.708	0.544
2	Murthy(1967).p.127 (1-21)	number of persons (1961)	workers at household industry	21	0.643	0.93	0.592
3	Murthy(1967).p.128-129 (66-88)	number of persons (1951)	no. of cultivators	23	0.39	0.576	0.772
4	Murthy(1967).p.128 (66-90)	number of persons (1961)	no. of cultivators	25	0.447	0.628	0.807
5	Murthy(1967).p.127 (23-41)	number of persons (1961)	no. of cultivators	19	0.524	0.647	0.829
6	Murthy(1967).p.128 (52-71)	number of persons (1961)	no. of cultivators	20	0.54	0.549	0.835
7	Murthy(1967).p.399 (18-34)	cultivated area (1961)	area under wheat (1964)	17	0.607	0.712	0.853
8	Murthy(1967).p.127 (1-17)	area in sq. miles	no. of cultivators	17	0.53	0.544	0.859
9	Murthy(1967).p.127 (1-21)	number of persons (1951)	no. of cultivators	21	0.616	0.633	0.864
10	Murthy(1967).p.127 (1-25)	area in sq. miles	no. of cultivators	25	0.607	0.657	0.868
11	Murthy(1967).p.128 (47-65)	number of persons (1951)	no. of cultivators	19	0.603	0.592	0.87
12	Murthy(1967).p.128 (42-62)	area in sq. miles	no. of cultivators	21	0.526	0.618	0.879
13	Murthy(1967).p.130 (106-128)	number of persons (1951)	no. of cultivators	23	0.642	0.702	0.881
14	Murthy(1967).p.127-128 (22-46)	number of persons (1951)	no. of cultivators	25	0.547	0.631	0.895
15	Murthy(1967).p.128 (32-53)	number of persons (1961)	no. of cultivators	22	0.502	0.602	0.896
16	Murthy(1967).p.127 (1-22)	number of persons (1961)	no. of cultivators	22	0.639	0.629	0.901
17	Murthy(1967).p.228 (22-54)	Fixed Capital	Output for Factories in a region	25	0.117	0.124	0.92
18	Murthy(1967).p.127 (18-41)	area in sq. miles	no. of cultivators	24	0.614	0.705	0.923
19	Murthy(1967).p.127 (26-50)	area in sq. miles	no. of cultivators	25	0.596	0.623	0.941
20	Murthy(1967).p.228 (57-80)	Fixed Capital	Output for Factories in a region	24	0.235	0.098	0.968
21	Murthy(1967).p.399 (1-17)	area under wheat (1963)	area under wheat (1964)	17	0.732	0.76	0.977
22	Murthy(1967).p.399 (18-34)	area under wheat (1963)	area under wheat (1964)	17	0.698	0.712	0.988

References

1. Brewer KRW (1963). Ratio Estimation in Finite Populations: Some Results Deducible from the Assumption of an underlying Stochastic Process. Aust. J. Statist., 5, 93-105.
2. Chaudhuri A (1976). A non-negativity criterion for a certain variance estimator. *Metrika*, 23, 201-205.
3. Chaudhuri A (1981). Non- negative unbiased variance estimators. Current topics in survey sampling, Academic Press, Inc., 317-328.
4. Chaudhuri A, Arnab R. (1981). On non-negative variance estimation. *Metrika*, 28, 1-12.
5. Ghosh M, Meeden G. (1997). Bayesian Methods for Finite Population Sampling. London: Chapman & Hall.
6. Midzuno H (1950). An outline of the theory of sampling systems. *Annals Institute of Statistics and Mathematics*, Tokyo, 1, 149-156.
7. Murthy MN (1967). Sampling Theory and Methods, Statistical Publishing House: Calcutta.
8. Patel PA, Patel JS (2009). On Non-Negative and Improved variance estimation for the ratio estimator under the Midzuno-Sen Sampling scheme. Statistics in Transition-new series, 10 (3), 371-385.
9. Patel PA, Patel JS (2010). Comparison of some variance estimators of the ratio estimator in presence of two auxiliary variables. Inter Stat journal, January 2010.
10. Rao JNK (2003). Small Area Estimation. New York: Wiley.
11. Rao JNK, Vijayan K (1977). On estimating the variance in sampling with probability proportional to aggregate size. *Journal of the American Statistical Association*, 72, 579-584.
12. Rao JNK (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling. Journal of the Indian Statistical Association, 17, 125-136.
13. Rao TJ (1972). On the variance of the ratio estimator for the Midzuno-Sen sampling scheme. Metrika, 18, p. 209-215.
14. Rao TJ (1977). Estimating the variance of the ratio estimator for the Midzuno- Sen sampling scheme. Metrika, 24, 203-215.
15. Royall RM (1970). On Finite Population Sampling Theory Under Certain Linear Regression Models, Biometrika, 57, 377-387.
16. Royall RM, Cumberland WG (1978). Variance estimation in finite population sampling. *Journal of the American Statistical Association*, 73, 351-358.
17. Royall RM and Cumberland WG (1981a). An empirical study of the ratio estimator and estimators of its variance. J. Amer. Statist. Assoc., 76, 66-88.
18. Royall RM and Cumberland WG (1981b). The finite population linear regression estimator and estimators of its variance- An empirical study. J. Amer. Statist. Assoc., 76, 924--930.

19. Royall RM, Eberhadt, KR (1975). Variance estimates for ratio estimator. *Sankhya, Series C*, 37, 43-52.
20. Valliant R, Dorfam AH, Royall RM (2000). *Finite population sampling and inference*, Wiley, New York
21. Vijayan K, Mukhopadhyay P, Bhattacharya S (1995). On non-negative unbiased estimation of quadratic forms in finite population sampling, Aust. Jour. Statist. 37(2), 169-178.
22. Wu CFJ (1982). Estimation of variance of the ratio estimator, Biometrika. 69, 183-189.
23. Wu CFJ, Deng LY (1983). Estimation of variance of the ratio estimator: An empirical study. In scientific Inference Data Analysis and Robustness, Ed. G.E.P. Box et al. New York Academic Press, 579-584.