# A COMPARISON AMONG SOME SAMPLING STRATEGIES UNDER A MODEL

BY

ARIJIT CHAUDHURI
*Indian Statistical Institute, Calcutta*
(Received : January, 1977)

## 1. INTRODUCTION

Various results are well-known in the literature [vide Durbin (7), Rao (14), Rao and Webster (16), Rao and Ramchandran (17), Rao (18), Rao and Rao (19), Singh (21) and Chakrabarti (2) among others] concerning the relative efficiencies, under various circumstances, of several sampling strategies for estimating a finite population mean $\bar{Y}=1/N \sum_i y_i$ for a variate $y$ for which the value $y_i$ on the $i$th unit of the population satisfies the following general model:

$$y_i=\alpha+\beta\, x_i+u_i \qquad (i=1, \ldots, N), \qquad \ldots(1.1)$$

where $u_i$'s are random variables with conditional expectations

$$\varepsilon\,(u_i/x_i)=0,$$

$$\varepsilon\,(u_i^2\,/x_i)=\delta\, x_i\, g \forall_i$$

$$(\delta>0,\ g=\text{a real number}),$$

$$\varepsilon\,(u_i\, u_j/x_i,\, x_j)=0 \ \forall\ i, j\ (i\neq j),$$

$x_i$'s are known real quantities which are either fixed (non-stochastic) constants or realizations on random variates usually with incompletely specified distributions. In this paper our modest purpose is to compare the relative efficiencies of a few customary sampling strategies under a rather simplistic special case of the above model [say, Model I, assumed, essentially following Chakrabarti (2), but stretching his ideas a little] for which $g=0$ and $x_i$s are identically and independently gamma-distributed with a mean parameter $m$ taken equal to [may not be too disconcerting if one assumes, as we do, $N$ to be large, in view of the Central Limit Theorem] the known value $\bar{X}=1/N \sum x_i$. The results (which are exact and valid for any sample-size) we derive in the sequel seem to be interesting and point to the appropriateness of a particular one among the sampling strategies we study under various circumstances. These, of course, do not follow as corollaries from results already available in the literature under more general models with $g \neq 0$ and $x_i$s having different and incompletely specified distributions. Obviously, our results are subject to severe limitations in respect of the nature of the $y$-variate-values to which they apply. Yet we present them because the model I we postulate is already receiving attention in the literature and we believe it is worthwhile to note a few of its implications on the relative efficiencies of several popular sampling strategies. We are, however, as yet unable to report more interesting results which should apply to situations with incompletely specified gamma distributions for $x_i$s and with the stipulation $g=0$ relaxed, but research along the line indicated is in progress.

In the end we present a few asymptotic results concerning the relative efficiencies of the same sampling strategies under an extended model (to be called Model II) following Singh's (21) works, on dispensing with the assumption about the gamma form of the distributions of $x_i$s.

In this note we shall denote a sample (typically, of size $n$ as treated throughout) by $s$, its selection-probability by $p(s)$ for a design $p$ (generically), and the expectation-operator (for the design) by $E$. Also, we shall use $\varepsilon$ for expectation over $u$ (with fixed $x$), $\varepsilon_x$ for expectation over the distribution of $x$, $\bar{\varepsilon}=\varepsilon_x\,\varepsilon$ for the two-step expectation (for $u$ with fixed $x$ and then over $x$) and $e=\bar{\varepsilon}\,E=\varepsilon_x\,\varepsilon\,E$ to denote the three-step (including the one over the sampling-design $p$) expectation-operator.

Here we study only six strategies consecutively numbered as 1, ..., 6 which respectively involve the estimators $t_1=\bar{X}\dfrac{\bar{y}}{\bar{x}}$ ($\bar{y}$, $\bar{x}$ being the same means), $t_2=\bar{X}\ 1/n \sum_s \dfrac{y_i}{x_i}$, both based on the SRSWOR scheme, $t_3=t_1$ based on Midzuno-Sen-Lahiri (11) scheme, the Horvitz-Thompson (10) estimator $t_4=1/N \sum_s y_i/\pi_i$ based on a $\pi ps$ design with inclusion-probability $\pi_i=n\dfrac{x_i}{NX}=np_i\ \forall_i$ (say), and $t_5$, $t_6$, the usual Rao-Hartley-Cochran (RHC. in brief) (15) and Hansen-Hurwitz estimators (HHE, in brief) (8) involving normed size-measures $p_i$s. The expectations, biases and mean-square-errors of the estimators will be denoted respectively as

$$E(t_i),\ \epsilon_i = e(t_i) = \bar{\varepsilon}\, E(t_i) = \varepsilon_x\, \varepsilon\, E(t_i) = \varepsilon_x E\, \varepsilon\, (t_i),$$

$$B_i = B(t_i) = e\,(t_i - \bar{Y}) = \bar{\varepsilon}\,[E(t_i) - \bar{Y}] = \varepsilon_x\,[\varepsilon\,(E(t_i) - \bar{Y})]$$

and $$M_i = \bar{\varepsilon}\,[E\,(t_i - \bar{Y})^2] = \varepsilon_x\,[\varepsilon\,\{E(t_i - \bar{Y})^2\}]$$

(in case $\alpha = 0$ in the model (1.1), we shall write $M_i'$ for $M_i$), $i = 1, \ldots, 6$.

## 2. New and Exact (small-sample) Results Concerning Relative Efficiencies of Sampling Strategies under Model I and Related Earlier Results under Alternative Models

Utilizing the properties of a gamma distribution with density $f(x) = 1/\Gamma(m)\ e^{-x}\, x^{m-1}$, $x > 0$, $m > 0$, with parameter $m$, we get, after some algebraic manipulations whenever needed :

$$\epsilon_1 = \alpha \frac{nm}{nm-1} + \beta m,$$

$$B_1 = \alpha\left(\frac{1}{nm-1}\right)$$

$$M_1 = \alpha^2 \frac{nm+2}{(nm-1)(nm-2)} + \frac{\delta\, nm^2}{(nm-1)(nm-2)} - \frac{\delta}{N}\frac{nm+1}{nm-1}$$

$$\epsilon_2 = \frac{\alpha m}{m-1} + \beta m,$$

$$B_2 = \alpha\,(1/m-1)$$

$$M_2 = \alpha^2\left[\frac{m^2}{n}\,\frac{1}{(m-1)(m-2)} + \frac{n-1}{(m-1)^2} + 1 - \frac{2m}{m-1}\right]$$
$$+ \frac{\delta}{n}\,\frac{m^2}{(m-1)(m-2)} - \frac{2m}{N}\,\frac{\delta}{(m-1)} + \frac{\delta}{N},$$

$$\epsilon_3 = \epsilon_4 = \epsilon_5 = \epsilon_6 = \alpha + \beta m,$$

$$M_3 = \frac{\alpha^2}{(nm-1)} + \delta\,\frac{m}{(nm-1)} - \frac{\delta}{N},$$

$$M_4 = \frac{m}{(m-1)}\,\frac{\delta}{n} - \frac{\delta}{N}$$

($M_4$ is complicated and hence not included)

$$M_5 = \frac{N-1}{N-n}\,\frac{1}{n}\left(\frac{\alpha^2}{m-1} + \delta\,\frac{m}{m-1} - \frac{\delta}{N}\right),$$

$$M_6 = \frac{N-1}{N-n}\,M_5.$$

The expression for $M_2$ is complicated. So, it is difficult to compare it with other $M_i$s. However, if $m$ is so large that we may neglect the error in writing $\frac{1}{m-2}$ for $\frac{1}{m-1}$ neglecting terms $0(1/m^2)$ then an approximate expression for it written $\overline{M}$ turns out to be

$$\overline{M_2}=\alpha^2\frac{m+2}{(m-1)(m-2)}+\frac{\delta}{n}\frac{m^2}{(m-1)(m-2)}-\frac{\delta}{n}\frac{m+1}{m-1}$$

(in case $\alpha=0$, we shall write $\overline{M}'_2$ for $\overline{M_2}$).

We shall use $M_2$ (and $M_2'$) to compare the efficiency of the strategy 2 with others. As this approximation is introduced for the sake of sheer simplicity alone we shall not make the corresponding approximations to other $M_i$'s while comparing them.

The relative efficiencies of the strategies under Model I turn out, suppressing the algebraic steps, as follows :

$\overline{M_2}'>M_1'$, if $m>2$ (a condition we assume throughout to be true) ;

$\overline{M_2}>M_1$ if $m\geqslant 4$; $|B_2|>|B_1|$ ; $M_1>M_3$; $\overline{M_2}>M_3$ ;
$M_1'<M_4'\ \forall\ n\geqslant 3$; $\overline{M_2}'>M_4'$ ;

$M_3'<M_4'$; $M_1'<M_5'$ if $n>2$; $M_1<M_6$ if $n\geqslant 5$; $M_3'<M_5'$ ;
$M_3<M_6$ ; $M_5'<M_4'$.

*Remark I.* Under the Model I the unbiased ratio-estimator based on Ikeda-Sen-Midzuno-Lahiri (11) scheme fares best among its competitors considered here.

The Model I with the assumption in particular that $m=\bar{X}$, renders peculiarities to the relative efficiencies of the strategies. For some other models studied in the literature their standings are not quite alike and some of the relevant results are noted below.

Taking $\alpha=0$ and $0\leqslant g\leqslant 2$ and without assuming a particular form for the distribution of $x_i$s, Rao (13) found, inter alia, that

$M_3'<M_6'$ if $g\leqslant 1$ ; $M_4'<M_6'\ \forall\ g$ ; $M_3' \gtreqless M_4'$ according as $g\gtreqless 1$ ;

$M_5' \gtreqless M_4'$ if $g \gtreqless 1$ ; $M_3' \gtrless M_5'$ according as $g\gtrless 1$, only if $n$ is large ;

$M_3'=M_4'=M_5'$ if $g=1$ ; $M_4'<M_5'$, $M_4'<M_3'$ if $1<g<2$ and if $N$ is large and Max $p_i$ is $0\ (1/N)\ \forall\ i$ ; $M_4'<M_5'$, $M_4'<M_3'$ if $g=2$.

For the same model Raj (12) showed earlier that $M_3'<M_6'$ if $n=2$ and $g=0$ or $g=1$. Hanurav (9) and Rao (20) showed respectively that $M_4' \lesseqgtr M_5'$ and $M_4' \lesseqgtr M_3'$ according as $g \gtreqless 1$. Chaudhuri (4) showed that $M_1'>M_4'$ if $g \geqslant 1$ and Chaudhuri and Arnab (5) have shown that $M_4' \lesseqgtr M_5' \lesseqgtr M_3'$ according as $g \gtreqless 1$.

Rao (18), on the other hand, assumed a model different from I only in relaxing the restriction that $m=\bar{X}$ and observed the following:

$M_3<M_1$ if $0 \leqslant g \leqslant 2$; $M_3<M_5$ if $0 \leqslant g \leqslant 1$; $M_1<M_5$ if $g=0$, $n>4$; $M_3'<M_1'$ if $0 \leqslant g \leqslant 2$; $M_5'>M_3'$ if $0 \leqslant g \leqslant 1$; $M_3' \leqslant M_5'$ if $g=1$; $M_5'<M_3'$ if $1 < g \leqslant 2$; $M_1'<M_5'$ if $g=0$, $n>2$; $M_5'<M_1'$ if $1 \leqslant g \leqslant 2$.

It is interesting to note the difference caused by taking $m=\bar{X}$ in the model.

### 3. A Few Asymptotic Results

In a recent article Singh (21) compared the strategies 1 and 3 by considering asymptotic expressions for the mean square errors of $t_1$ and $t_3$ and also on assuming in addition normality of the joint distribution of $x_i$ and $y_i$ for $i=1, \ldots\ldots, N$ and also under some other assumptions. Here we consider similar asymptotic expressions for these mean square errors assuming Chakrabarti's (2) model with the relaxation that $x_i$s are identically and independently distributed with a mean $m=\bar{X}$ (known), the common distribution being not necessarily of the gamma form the resulting generalized model being denoted as II. Naturally, the (asymptotic) expressions for the mean square errors do not agree with those due to Singh (21) because of the simplification we achieve by virtue of the assumption that $m$ equals $\bar{X}$ which is a fixed known quantity but we still have positive conclusions concerning their relative magnitudes as shown below.

Let us write

$$\bar{y}=\bar{Y}+\xi,$$

$$\bar{x}=\bar{X}-\eta,$$

$$\bar{u}=\frac{1}{n}\sum_{s} u_i,$$

$$\bar{U}=1/N \sum_{1}^{N} u_i,$$

$$R=\bar{Y}/\bar{X}=\bar{Y}/m=\beta+\frac{\alpha+\bar{U}}{m}.$$

Following Singh (21) we assume that we may write

$(\eta^r \zeta^q) \simeq 0$ for $r+q \geqslant 5$, with $r$, $q$ positive integers. ...(3.1)

Then assuming the model outlined above we have

$M_1 - M_3 = 2\delta\ (1/n - 1/N)\ \mu_2 + (\alpha^2 + \delta/N)\ (2\mu_4 - \mu_3)$ ...(3.2)

where we write

$$\mu_r = e \frac{(\bar{x}-m)^r}{m} \text{ for } r = 0, 1, 2, \ldots$$

Writing $\nu_r$ for $e\ (/Z/^r)$ where $z$ is any variate with a zero mean,

$$\beta_1 = \frac{\mu_3 2}{\mu_2 3}, \beta_2 = \frac{\mu_4}{\mu_2 2}, \text{ and remembering that}$$

$\beta_2 \geqslant \beta_1 + 1$, $\nu_r^{1/r} \leqslant \nu_{r+1}^{1/r+1}\ \forall\ r$ and that $m > 2$, we get immediately from (3.2) that $M_1 > M_3$, whatever identical distributions $x_i$s may have with a mean $m = \bar{X}$, provided their moments $\mu_r$ exist for $r = 4$.

If instead of (3.1) we assume more stringently that we may neglect the error in writing

$\bar{X} = m$ for $\bar{x}$ for each sample $s$ ...... ...(3.3)

as is often done in text books (*e.g.* Cochran's (6)), then under the model II outlined above we get the following results:

$$M_3' \ \delta\ (1/n - 1/N),$$

$$M_4' = \delta\ (\bar{X}/Nn\Sigma 1/x_i - 1/N),$$

$$M_5' = \frac{N-n}{N-1} M_6',$$

where

$$M_6' = \frac{\delta}{n} (\bar{X}/N\Sigma 1/x_i - 1/N); \text{ hence we conclude that}$$

$$M_6' > M_4' > M_3' \text{ and } M_5' > M_3'.$$

*Remark II.* Finally, we may mention that some results concerning these strategies are available in the literature under the assumption of a different type of alternative model introduced by Avadhani and Sukhatme (1) and refer the reader to this paper and also to the ones by Chaudhuri [(3), (4)] and Chaudhuri and Arnab (5) to take note of the results under that model.

## 5. Numerical Values of Relative Efficiencies of Strategies

Defining the efficiencies of the strategies as $E_i=100\ \bar{M}_2/M_i$ (or as $E_i=100\ \bar{M}'_2/M'_i$ in case $\alpha=0$), $i=1, \ldots, 6$, we present below the values of the relative efficiencies of the strategies for a few combinations of the parametric values.

| *Parametric combination* | | | *Efficiencies of the strategies* | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ | $E_6$ |
| *I* | $\alpha=0.5$, $\delta=2.0$, $\eta=6$, | $\beta=1.0$ $m=5$ $N=20$ | 256 | 100 | 276 | — | 246 | 168 |
| *II* | $\alpha=1.5$, $\delta=2.0$, $n=4$, | $\beta=0.5$ $m=8$ $N=40$ | 212 | 100 | 228 | — | 208 | 192 |
| *III* | $\alpha=0$, $\eta=2.5$, $n=5$, | $\beta=1.5$ $m=6$ $N=10$ | 189 | 100 | 204 | 157 | 177 | 100 |

Thus, in each $E_3>E_1>E_5>E_6\geqslant E_2$; our theoretical findings in favour of strategy 3 corroborated.

## Acknowledgment

The author is grateful to the referees for their valuable suggestions which led to a considerable improvement on earlier drafts of the paper.

## Summary

In estimating a finite population mean, we investigate the "small sample" relative efficiencies of several well-known sampling strategies under Chakrabarti's (2) model and find the one due to Ikeda-Sen-Midzuno-Lahiri (11) to be the most promising among them. A similar conclusion is reached with a slight generalization of the model coupled with large sample approximations.

## References

[1] Avadhani, M.S. and Sukhatme, B.V. (1970) : A comparison of two sampling procedures with an application to successive sampling. *Appl. Statist.* 19, 251-59.

[2] Chakrabarti, R.P. (1973) : A note on the small sample theory of the ratio estimator in certain specified population. *J. Ind. Soc. Agri. Statist.* 25, 49-57.

[3] Chaudhuri. Arijit (1976) : On the efficiencies of some sampling strategies. *Cal. Statist. Assoc. Bull.* **25**, 119-29.

[4]——(1977) : On some properties of the Horvitz-Thompson estimator based on Midzuno's $\pi$ps sampling scheme. *J. Ind. Soc. Agri. Statist.* **29**, 47-52.

[5] Chaudhuri, Arijit and Arnab, Raghunath (1971) : On the relative efficiencies of sampling strategies under two customary models. *Proc. 65th Session, Ind. Sc. Cong.*

[6] Durbin, J. (1959) : A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, **46**, 477-80.

[7] Cochran, W.G. (1963) : Sampling Techniques. *John Wiley and Sons*, N.Y.

[8] Hansen, M.H. and Hurvtiz, W.N. (1943) : On the theory of sampling from finite popula tions. *Ann. Math. Statist.* **14**, 333-62.

[9] Hanurav, T.V. (1967) : Optimum utilization of auxiliary information : $\pi$ps sampling of two units from a stratum. *J. Roy. Statist. Soc. B.* **29**, 379-91.

[10] Horvitz, D.G. and Thompson, D.J. (1952) : A generalization of sampling without replacement from finite universes. *J. Amer. Statist. Assoc.* **47**. 663-85.

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

[12] Raj, D. (1954) : On sampling with probability proportional to aggregate size. *J. Ind. Soc. Agri. Statist.* **16**, 317-19.

[13] Rao, J.N.K. (1966) : On the relative efficiency of some estimators in pps sampling for multiple characteristics. *Sankhya. Ser. A.* **28**, 61-70.

[14 ——(1968) : Ratio and regression estimators. New Developments in survey sampling, edited by N.L. Johnson and H. Smith (jr.). *Wiley Interscience.*

[15] Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962) : On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser. B.* 24, 482-49.

[16] Rao, J.N.K. and Webster, J.T. (1966) : On two methods of bias reduction in the estimation of ratios. *Biometrika.* 53, 571-77.

[17] Rao, J.N.K. and Ramachandran, V. (1974) : Comparison of the separate and combined ratio estimators. *Sankhya, Ser. C.* **36**, 151-56.

[18] Rao, P.S.R.S. (1968) : On three procedures of sampling from finite populations. *Biometrika.* **55**, 438-41.

[19] Rao, P.S.R.S. and Rao, J.N.K. (1971) : Small sample results for ratio estimators. *Biometrika.* 58, 625-30.

[20] Rao, T.J. (1967) : On the choice of a strategy for the ratio method of estimation. *J. Roy. Statist. Soc. Ser. B.* **29**, 392-97.

[21] Singh, Ravindra (1975) : A note on the efficiency of ratio estimate with Midzuno's scheme of sampling. *Sankhya. Ser. C.* **37**, 211-14.