ON THE RELATIVE EFFICIENCIES OF SEVERAL STRATEGIES FOR ESTIMATING A FINITE POPULATION MEAN

By

ARIJIT CHAUDHURI AND ARUN KUMAR ADHIKARY

Indian Statistical Institute, Calcutta

(Received : August, 1977)

SUMMARY

Following Chakrabarti [1] and Chaudhuri [2] we compare the relative efficiencies of six well-known sampling strategies under models more general than those considered by them. Exact (small sample) as well as asymptotic results are presented along with numerical ones.

I. NTRODUCTION

Pursuing the works of Chakrabarti [1] and Chaudhuri [2], consider the following generalization of their models for estimating the mean

$$\bar{Y}=\frac{1}{N}\sum_{1}^{N} Y_i$$

of a finite population on the i-th unit of which a real variate y assumes the value y_i $(i=1,\ldots\ldots, N)$:

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

where $\alpha>0$, $\beta\geqslant 0$ and u_i's are random variables with conditional expectations

$$\varepsilon(u_i \mid x_i)=0,\; \epsilon\,(u_i^2 \mid x_i)=\delta x_i^g \;\; \forall i,$$

$$\text{(with } 0<\delta<\infty,\; 0<g<2),$$

$$\varepsilon\,(u_i\, u_i \mid x_i, x_j)=0 \;\; \forall \;\; i\neq j,$$

where x_i's are known positive quantities which are assumed to be realizations on random variates (also denoted as x_i's), with incompletely specified distributions. In section 2 the auxiliary variables x_i's are assumed to be identically and independently gamma-distributed with the common density

$$f(v)=\frac{1}{\Gamma(m)}\, e^{-v}\, v^{m-1},\; v>0,$$

with a single parameter (mean) m (>2) taken, following Chakrabarti [1] as equal to the known value

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{N} x_i$$

for the finite population (the sum being over the known x_i's). This model will be denoted as I in section 3 the forms of the distributions of x_i's are left unspecified except that they are assumed to be distributed independently and identically with a common mean $m=\bar{X}$ as above and their common moments μ_r' of orders r exist for r upto g, $(0 \leq g<2)$. The model in this case will be denoted as II. Our objective in this article is to compare the biases and mean square errors of six well-known sampling strategies under the models stipulated above, which are slight generalizations of those due to Chakrabarti [1] and Chaudhuri [2] in the sense that unlike them we allow g to be non-zero. Following Chaudhuri [2], in section 3 a few asymptotic results are presented assuming both the sample and the population sizes to be large. Our results differ from similar others available in the literature because of the peculiarities following from the assumption $m=\bar{X}$ which introduces a novelty suggested initially by Chakrabarti [1].

Our notations are as in Chaudhuri [2]: by s is meant a sample (typically, the sample-size being taken throughout as a fixed positive integer n), its selection-probability is $p(s)$ for a design p (generically), and the expectation-operator for the design is E. By ε we denote conditional expectation over the error term u (with fixed x) in the model (1.1), ε_x denotes expectation over the distribution of x (standing for x_i, $i=1, \ldots, N$), also $\bar{\epsilon}=\varepsilon_x\varepsilon$ is the two step expectation (for u with x fixed and then over x) and $e=\bar{\varepsilon}\, E=\varepsilon_x\, \varepsilon E$ is the three step (including the one over the sampling design p) expectation-operator.

The six strategies we consider will be numbered consecutively as 1, ..., 6 which respectively involve the estimators $t_1=\bar{X}\frac{\bar{y}}{\bar{x}}$ ($\bar{y}$, $\bar{x}$ denoting the sample means for y and x); $t_2=\bar{X}\frac{1}{n}\sum_{i\epsilon s}\frac{y_i}{x_i}$, both based on the SRSWOR scheme, $t_3 = t_1$ based on Midzuno-Sen-Lahiri [7] scheme, the Horvitz- Thompson [6] estimator $t_4 = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}$, based on a π_{ps} design with inclusion-probability $\pi_i = np_i$ i(say), $\left(p_i = \frac{x_i}{N\bar{X}} \right)$, and t_5, t_6 the usual Rao-Hartley-Cochran (RHC, in brief) [8], (for this we assume $N/n = k$ to be an integer and each group formed in applying this scheme is of size K), and Hansen-Hurwitze etimators (HHE). in brief, [5], both involving normed size-measures $p_i's$. The expectations, biases and mean square errors of the estimators will be denoted respectively as

$$\varepsilon_i = e(t_i) = \varepsilon E(t_i) = \varepsilon_x \varepsilon E(t_i)$$

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

$$M_i = \varepsilon [E(t_i - \bar{Y})^2] = \varepsilon_x [\varepsilon \{E(t_i - \bar{Y})^2\}],\ i = 1, \ldots, 6$$

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

2. Exact (Small Sample) Formulae Concerning Relative Efficiencies of Sampling Strategies Under Model I with Gamma-Distributed Auxiliary Variates

Omitting easily verifiable algebraic steps, to save space, and using Rao and Webster's [9] lemma concerning properties of gamma distrtbutions we get

$$B_1 = \frac{\alpha}{nm-1},\ B_2 = \frac{\alpha}{m-1},\ \text{so that } B_2 > B_1,$$

$$M_1 = \alpha^2 \frac{nm-2}{(nm-1)(nm-2)} + \delta\, nm^2 \frac{\Gamma(m+q)}{\Gamma(m)} \frac{1}{(nm+g-1)(nm+g-2)}$$

$$+ \frac{\delta}{N} \left[\frac{\Gamma(m+g)}{\Gamma(m)} \left\{ 1 - 2 \frac{nm}{nm+g-1} \right\} \right]$$

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

$$+ \frac{\delta m^2}{n} \cdot \frac{\Gamma(m+g-2)}{\Gamma(m)} + \frac{\delta}{N} \left[\frac{\Gamma(m+q-1)}{\Gamma(m)} (g-m-1) \right]$$

$$B_i = 0\ \forall\ i = 3, \ldots, 6,$$

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

In calculating M_4 we neglect the term $\alpha^2 \sum\sum_{i\neq j} \left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)$ and thus get a conservative expression on assuming $\pi_{ij}<\pi_i\pi_j \ \forall \ i, j$,

$$M_4=\alpha^2\left\{\frac{m}{n(m-1)}-\frac{1}{N}\right\}+M_4', \text{ where}$$

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

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

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

Noting the complicated form of the coefficient of α^2 in M_2 we assume m to be large and neglecting terms $0(1/m^2)$ for the sake of sheer simplicity in it (only) we approximate this coefficient by $\frac{(m+2)}{(m-1)(m-2)}$ and the quantity M_2 is approximated by

$$\overline{M}_2=\alpha^2\frac{(m+2)}{(m-1)(m-2)}+\frac{\delta}{n}m^2\frac{\Gamma(m+g-2)}{\Gamma(m)}$$
$$+\frac{\delta}{N}\frac{1}{\Gamma(m)}\left\{\Gamma(m+g)-2m\,\Gamma(m+g-1)\right\}.$$

3. Asymptotic Formulae Concerning Relative Efficiencies of Strategies Under Model II

As in Cochran [4], in what follows, we shall assume N and n large to such an extent that we may neglect the error in writing $\bar{x}=\bar{X}$ for every sample s with $p(s)>0$ for the strategies 1 and 3. For other strategies also N and n will be taken to be large whenever necessary (as discussed below) in comparing the efficiencies. For the Model II, we have the following formulae on omitting details of calculations, *viz.*

$$M_1=\frac{N-n}{N-n}\frac{1}{N-1}\left[N\beta^2\mu_2+\delta(N-1)\mu_g'\right],$$

where $\mu_2=\mu_2'-\mu_1'^2$,

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

TABLE 4.1

Showing relative performances of the strategies with variation in g (section I of the table represents results under model I and section II those for Model II)

I

$g<\frac{1}{2}$	$g\leqslant 1$	$g\leqslant g_0$	$1\leqslant g\leqslant 1.5$	$g>g_0$	$g\geqslant 1$	$0\leqslant g\leqslant 2$
$M_3<M_1$	$M'_1 \leqslant M'_4$				$M'_4 \leqslant M'_1$	$M'_4 \leqslant M'_2$ (if $N \lhd 2n$)
$M_1<M_4$	$M_3\leqslant M_4$				$M'_4 \leqslant M'_2$	$M'_4 \leqslant M'_6$
	$M_5\leqslant M_4$	$M'_1 \leqslant M'_6$ (for large N)	$M_1\leqslant \overline{M}_2$	$M'_6 \leqslant M'_1$ (for large N)	$M'_4 \leqslant M'_5 \leqslant M'_3$	
	$M'_3 \leqslant M'_5\ M'_4$	$M_1\leqslant M_6$ (if $n\geqslant 5$)			$M'_6 \leqslant M'_3$ (for large N)	
	$M_3\leqslant M_5$ (if $N>nm$ $M_3\leqslant M_6$				$M'_6 - \overline{M}'_2$	
	$M'_3 < M'_1$				$\overline{M}_2\leqslant M_3$	
	$\overline{M}'_2 \leqslant M'_6$					
	$M_3\leqslant \overline{M}_2$					

Table 4.1 (contd.) *II*

$g \leqslant 1$	$g \geqslant 1$	$0 \leqslant g \leqslant 2$
$M'_6 \leqslant M'_1$ (neglecting terms of 0 (1/N))	$M_3 \leqslant M_1$	$M'_6 \leqslant M'_2$ (neglecting terms 0 $1/N$)
$M'_3 \leqslant M'_2$ (if $N > 2n$)	$M'_4 \leqslant M'_1$	$M'_4 \leqslant M'_6$
$M'_3 \leqslant M'_5 \leqslant M'_4$	$M'_5 \leqslant M'_1$	
$M'_3 \leqslant M'_4$	$M'_5 \leqslant M'_2$ (neglecting terms of 0 $(1/N^2)$)	
$M_3 \leqslant M_4$	$M'_4 \leqslant M'_2$	
	$M'_4 \leqslant M'_3 \leqslant M'_3$	
$\left(\text{if } \frac{n-1}{n-1} \simeq 0 \right)$		
$M_3 \leqslant M_6$		

g_0 is a root in [0, 2] of $g^2-(n^2m-2nm+3)\, g(n^2m-3nm+2)=0$ if g less than 0 is permitted, then $M'_1 \leqslant M'_2$ if $\beta=0$.

$$M_3 = \delta \left(\frac{1}{n} - \frac{1}{N} \right) \mu_g'$$

On neglecting the term $\alpha^2 \sum\sum_{i \neq j} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$ as in section 2,

$$M_4 = \alpha^2 \left[\frac{m}{n} \epsilon_x (1/X) - 1/N \right] + M_4',$$

where

$$M_4' = \delta \left[\frac{m}{n} \mu_{g-1}' - \frac{1}{N} \mu_g' \right],$$

$$M_5 = \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \left[\alpha^2 \left(\frac{m}{m-1} - 1 \right) + \delta m \mu_{g-1}' - \frac{\delta}{N} \mu_g' \right]$$

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

4· Relative Efficiencies

With simple algebra (suppressing details to save space), noting inter alia that cov $(x, x^{g-1}) \gtreqless 0$ as $g \gtreqless 1$, cov $(x, x^{g-2}) \leqslant 0$ since $g \leqslant 2$, $\mu_1'^2 > \mu_1'$, since $\mu_1' = m > 2$, writing $\frac{1}{N-1} \approx \frac{1}{N}$, when necessary) etc. we can derive the results presented in Table 4.1. Using formulae in section 2 we also present results of numerical investigation in Table 4.2.

5. Concluding Remarks

I. Relative performances of these strategies under other alternative models are well-known and a relevant recent reference is Chaudhuri [2].

II. One may, in practice, choose among these six strategies with reference to the above according to situations believed to be in hand.

III. From the Table 4.2 presenting numerical values one may gain some insight into the relative performances of the strategies under certain particular combinations of the values of the parameters for the Model I, *e.g.*, for $g<1$, strategy 3 is found to fare best amongst all, for $g=1$, strategy 5 is as good and strategy 4 is also on a par if $\alpha=0$—consistently with theory developed here and elsewhere.

TABLE 4.2

Showing relative efficiencies for certain parametric combinations using formulae in section 2

$\left(\text{writing } E_i = 100\,\frac{\bar{M}_2}{M_i} \text{ or } 100\,\frac{M_2'}{M_i'} \text{ in case } \alpha=0\right)$

		E_1	E_2	E_3	E_4	E_5	E_6
I. $\alpha=0.5$,	$g=0.0$	211	100	236	17ↄ	200	168
$\delta=2.0$,	$g=0.5$	162	100	176	152	163	137
$m=5$	$g=1.0$	129	100	138	134	138	116
$n=4$,	$g=1.5$	107	100	112	121	120	101
$N=20$	$g=2.0$	92	100	94	111	107	90
II. $\alpha=1.5$	$g=0.0$	213	100	229	111	209	193
$\delta=2.0$	$g=0.5$	155	100	864	114	156	144
$m=8$	$g=1.0$	124	100	128	112	128	118
$n=4$	$g=1.5$	106	100	108	108	113	104
$N=40$	$g=2.0$	94	100	95	103	103	95
III. $\alpha=0$	$g=0.0$	192	100	206	157	180	100
$\delta=2.5$	$g=0.5$	156	100	167	146	157	87
$m=6$	$g=1.0$	131	100	140	140	140	78
$n=5$	$g=1.5$	113	100	121	138	128	71
$N=10$	$g=2.0$	100	100	107	140	119	66
IV. $\alpha=0.5$	$g=0.0$	204	100	227	172	192	177
$\delta=2.0$	$g=0.5$	157	100	170	148	157	145
$m=5$	$g=1.0$	126	100	133	130	133	123
$n=4$	$g=1.5$	105	100	108	116	116	107
$N=40$	$g=2.0$	90	100	91	105	104	96
V. $\alpha=0$	$g=0.0$	167	100	179	151	154	138
$\delta=2.5$	$g=0.5$	139	100	146	135	136	222
$m=6$	$g=1.0$	118	100	123	123	123	110
$n=5$	$g=1.5$	102	100	105	113	112	100
$N=40$	$g=2.0$	90	100	91	105	103	92

Note. From the numerical data we see that $M_1 > M_3\ \forall g$ and $M_2 > M_0'$ unless g=2. But we are unable to establish general results about M_1 vs M_2 vs M_3.

ACKNOWLEDGEMENT

The authors are grateful to the referees for their helpful suggestions.

REFERENCES

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

[2] Chaudhuri, Arijit (1979) : A comparison among sampling strategies under a model, *J. Ind. Soc. Agri. Statist.* **31**, 16-23.

[3] Chaudhuri, Arijit and Arnab, Raghunath (1979) : On the relative efficiencies of sampling strategies under a super population model. *Sankhya, Ser C*, **41, 40-43**

[4] Cochran, W.G. (1963) : Sampling Techniques.John Wiley & Sons, N.Y.

[5] Hansen, M.H. and Hurwitz, W.N. (1943) : On the theory of sampling from finite populations. *Ann. Math. Statist.* **14**, 333-362.

[6] Horvitz, D.G. and Thompson, D.J. (1952) : A generalization of sampling without replacement from finite universes, *J. Amer. Statist. Assoc.* **47**, 663-685.

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

[8] 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**, 482-491.

[9] Rao, J.N.K. and Webster, J.T. (1966) : On two methods of bias reduction in the estimation of ratios. *Biometrika*, **53**, 571-577.