ON THE RELATIVE EFFICIENCIES OF SOME SAMPLING STRATEGIES UNDER A SUPER-POPULATION MODEL

BY

RAGHUNATH ARNAB

Indian Statistical Institute, Calcutta

(Received : August, 1977)

INTRODUCTION

In a recent article Ravindra Singh [8] has considered the relative efficiencies of the strategies involving the ratio estimators based on Midzuno-Sen [4, 7] sampling scheme and SRSWOR scheme and obtained some interesting results on assuming normality for the joint distribution of the variate under study y and on auxiliary variate x. In this note we have compared the relative efficiencies of these strategies considering the usual super-population model concerning the relationship between y and x and in addition, assuming the x-variate values for the N units of finite population to be independently identically distributed gamma varitaes with an unknown parameter. The results under the present set-up tend to indicate that SRSWOR scheme of sampling should be preferred in most cases in practice in using the ratio estimator. At the end we also compare these strategies with the strategy of Horvitz-Thompson [3] method of estimation based on any fixed sample size π ps sampling scheme, sample-size being throughout taken to be an integer n.

2. NOTATIONS AND RESULTS

By Y_i, X_i we shall denote the value assumed by y and x respectively on the ith ($i=1, 2 \ldots N$) unit of a finite universe of N individuals where Xi's are the known positive numbers, our problem being to estimate

$$Y=\sum_{I}^{N} Y_i$$

on employing a suitable sampling strategy. Following Cochran (1963) and others we shall assume that we may write

$$Y_i=\beta x_i+e_i \qquad \forall i$$

such that

$$E(e_i/x_i)=0 \qquad \forall i$$
$$E(e_i^2/x_i)=\sigma^2 x_i^g \qquad \forall i$$
$$E(e_i e_j/x_i, x_j)=0 \qquad \text{for } i\neq j \tag{2.1}$$

where

$$0<\sigma^2<\infty, \quad g\geqslant 0, \quad \infty<\rho<\infty$$

Let $E(.|.)$ denote the conditional expectation with respect to the distribution of ei's over a hypothetical super-population of which the current universe is a random sample. Further from Durbin [2] Quenouille [9] and others we shall suppose that X_i's are identically independently distributed as gamma variates each with a common unknown parameter h such that g may be supposed negligible compared to Nh.

Writing

$$\frac{Y}{X}, \text{ where } X=\sum_{I}^{N} X_i$$

the usual approximate expression for the variance of ratio-estimator based on SRSWOR scheme is

$$V_1=\frac{N(N-n)}{n(N-1)}\sum_{I}^{N}\left(Y_i-RX_i\right)^2$$

$$=\frac{N(N-n)}{n(N-1)}\left[\sum_{I}^{N} e_i^2+\frac{(\sum_{I}^{N} e_i)^2}{X^2}\sum_{I}^{N} X_i^2-2\sum_{I}^{N} e_i\frac{X_i}{X}\sum_{I}^{N} e_i\right]$$

We shall write $(. \mid x)$ for the conditional expectation operator where X_i's are held fixed, $E_x(.)$ for the expectation operator with respect to the distribution of X_i's and E for the overall expectation, so that

$$E=E_x[E(\cdot \mid X)]$$

Then, we get

$$E(v_1 \mid x)=\frac{N(N-n)}{n(N-1)}\sigma^2\left[\sum x_i^g+\frac{\Sigma x_i^{g+2}}{x^2}+\sum_{i\neq j}\frac{x_i^g x_j^2}{x^2}\right.$$
$$\left.-2\sum\frac{\Sigma x_i^{g+1}}{x}\right]$$

$$E(v_1)=E_x[E(v_1 \mid x)]$$

$$=\frac{N(-n)}{n(N-1)}\sigma^2 E_x\left[Nx_i^g+\frac{Nx_i^{g+2}+N(N-1)x_i^g x_j^2}{x^2}-\frac{2Nx_i^{g+1}}{x}\right].$$

We may recall from Rao and Webster (1966) that for any non-negative integers a, b, c, and the mutually identically independently distributed gamma variates z_1, z_2, ... z_n with a parameter h one has

$$E\left[\frac{z_i^a z_j^b}{\left(\sum_1^n z_i\right)^c}\right]=\frac{\Gamma(a+h)\,\Gamma(b+h)}{(\Gamma h)^2}\frac{1}{\prod_{t=1}^{c}(m+a+b-t)}$$

(where $m=nh$)

which one may readily check to be valid also for a, b non-integers. Using this result we readily obtain

$$E_1=E(v_1)=\frac{N^2(N-n)}{n(N-1)}\sigma^2$$

$$\left[\frac{\Gamma(g+h)}{\Gamma h}+\frac{\Gamma(g+h+2)}{\Gamma h\,(Nh+g+2-1)\,(Nh+g+2-2)}+(N-1)\frac{\Gamma(g+h)\,\Gamma(2+h)}{(\Gamma h)\,(Nh+g+2-1)(Nh+g+2-2)}-2\frac{\Gamma(g+h+1)}{\Gamma n.(Nh+g)}\right]$$

which, after some simplifications, may be varified to reduce to

$$E_1=\sigma^2 N^2\frac{(N-n)}{n}\frac{\Gamma(g+h)}{\Gamma h}\frac{h(Nh+2)}{(Nh+g)(Nh+g+1)}$$

The variance expression for the ratio estimetor based on Sen-Midzuno scheme of sampling involving normed-size measure

$p_i=\dfrac{x_i}{x}$ is given by Rao [5]

$$V_2=\sum T_i Y_i^2+\sum_{i\neq j}\sum T_{ij}Y_iY_j-Y^2$$

where

$$T_i = \frac{x}{\binom{N-1}{n-1}} \sum_{1 \leqslant i_2 < i_3 < \ldots\ldots < in \leqslant N}^{(i)} \frac{1}{x_i + x_{i2} + \ldots 1 x_{in}}$$

$$T_{ij} = \frac{x}{\binom{N-1}{n-1}} \sum_{1 \leqslant i_3 < \ldots\ldots < in \leqslant N}^{(i,j)} \frac{1}{x_i + x_j + x_{i3} + \ldots + x_{in}}$$

$$\sum_{i_2 < i_3 < \ldots\ldots < in,}^{(i)} \qquad \sum_{i_3 < \ldots\ldots in}^{(i,j)}$$

denoting summations respectively over $(n-1)$ and $(n-2)$ distinct units except i, i and j respectively out of 1, 2, ... N.

Then we get

$$E(V_2 \mid x) = \sigma^2 \sum \left(T_i - 1 \right) x_i^g$$

Noting that

$$E_x\left(T_i x_i^g \right) = E_x\left(x_i^g \right) + \left(N-n \right) h E_x \left[\frac{x_i^g}{x_i + x_{i2} + \ldots\ldots + x_{in}} \right]$$

$$= E_x\left(x_i^g \right) + \left(N-n \right) h \frac{\Gamma g+h}{\Gamma h} \frac{1}{(nh+g-1)}$$

we have

$$E_2 = E(v_2) = E_x[E(v_2 \mid x)]$$

$$= \sigma^2 N \left(N-n \right) h \frac{\Gamma g+h}{\Gamma h} \frac{1}{(nh+g-1)}$$

Again, we may note that for any fixed sample size (n) πps sampling scheme, the variance V_3 of HTE we have

$$E\left(V_3 \mid x \right) = \sigma^2 \sum x_i^a \left(\frac{x}{nx_i} - 1 \right)$$

$$= \sigma^2 \left(\sum x_i^g \left(\frac{1}{n} - 1 \right) + \sum_{(i \neq j)} \sum \frac{x_i^{g-1} x_j}{n} \right)$$

so that we have

$$E_3 = E(V_3) = E_x[\,E(V_3 \mid x)\,]$$

$$= \sigma^2 \left\{ N\left(\frac{1}{n}-1\right)\frac{\overline{|g+h}}{|\overline{h}} + \frac{(N-1)N}{n}\,\frac{\overline{|g+h-1}}{|\overline{h}}\;h \right\}$$

which readliy simplifies to

$$E_3 = \sigma^2\,\frac{N}{h}\,\frac{\overline{|g+h}}{|\overline{h}}\;\frac{(N-n)h-(n-1)(g-1)}{g+h-1}$$

2.1. Comparison between ratio-estimators based on SRSWOR and on Sen-Midzuno (1953, 1952) sampling scheme under the super population model (2.1).

$$E(V_1)-E(V_2) = \sigma^2 N^2\,\frac{(N-n)}{n}\,\frac{\overline{|g+h}}{|\overline{h}}\cdot\frac{h(Nh+2)}{(Nh+g)(Nh+g+1)}$$

$$-\sigma^2 N\,\frac{(N-n)}{|\overline{h}}\;h\;\frac{\overline{|g+h}}{(nh+g-1)}$$

$$= \sigma^2 Nh\left(N-n\right)\frac{\overline{|g+h}}{|\overline{h}}\left(\frac{N(Nh+2)}{n(Nh+g)(Nh+g+1)} - \frac{1}{nh+g-1}\right)$$

$$= \sigma^2\,\frac{N}{n}\left(N-n\right)\frac{\overline{|g+h}}{|\overline{h}}\left\{\left(1+\frac{2}{Nh}\right)\left(1+\frac{g}{Nh}\right)^{-1}\left(1+\frac{g+1}{Nh}\right)^{-1} - \left(1+\frac{g-1}{nh}\right)^{-1}\right\}$$

$$= \sigma^2\,\frac{N}{n}\left(N-n\right)\frac{\overline{|g+h}}{|\overline{h}}\,\Delta\left(g\right)\left(\text{say}\right)$$

Now neglecting the terms involving

$$\left(\frac{1}{Nh}\right)^j \text{ for } j \geqslant 2$$

we get

$$\triangle\left(g \right)\simeq\left(1-\frac{2g+1}{Nh}+\frac{2}{Nh}\right)-\left(1+\frac{g-1}{nh}+\frac{(g-1)^2}{(nh)^2}\right)$$

$$=\left(\frac{1}{Nh}-\frac{1}{nh}\right)+g\left(\frac{1}{nh}-\frac{2}{Nh}\right)-\frac{(g-1)^2}{(nh)^2}$$

$$=\delta(g)\text{ (say)}$$

It is known that the situations where $g<1$ are more frequent in practice than when $g\geqslant 1$. And, we note that for $0\leqslant g\leqslant 1$, we have $\delta(g)<0$ and for $g>1$, we have $\delta(g)<0$, provided $\left(\frac{g-1}{nh}\right)^2$ is neglected and the sampling fraction $f=\frac{n}{N}$ exceeds $\frac{1}{2}$. For $g>\frac{1-f}{1-2f}$ with $f<\frac{1}{2}$ we have $\delta(g)>0$. Thus we get an idea about when SRSWOR is to be preferred to Midzuno-Sen scheme in employing ratio-method of estimation and vice versa when the assumed model holds.

2.2. Comparison between ratio estimator based on SRSWOR and Horvitz-Thomson estimator based on a πps sampling scheme under the super-population model (2.1).

Now,

$$E_1-E_3=\sigma^2\,\frac{N}{n}\,\frac{\lceil g+h}{\lceil h}\,f(g)\text{ (say)}$$

where

$$f\left(g \right)=\left(N-n\right)\frac{Nh(Nh+2)}{(Nh+g)(Nh+g+1)}-\frac{(N-n)h-(n-1)(g-1)}{h+g-1}$$

$$=(N-n)\left[\left(1+\frac{2}{Nh}\right)\left(1+\frac{g}{Nh}\right)^{-1}\left(1+\frac{g+1}{Nh}\right)^{-1}-\left(1-\frac{(n-1)(g-1)}{(N-n)h}\right)\left(1+\frac{g-1}{h}\right)^{-1}\right]$$

$$=(N-n)\left[\left(1+\frac{2}{Nh}\right)\left(1-\frac{2g+1}{Nh}+\frac{g(g+1)}{(Nh)^2}+\ldots\right)-\left(1-\frac{g-1}{h}-\frac{(n-1)(g-1)}{(N-n)h}+\frac{(n-1)^2(g-1)^2}{(N-n)h^2}+\ldots\right)\right]$$

$$\simeq(N-n)\left[\left(1+\frac{2}{Nh}-\frac{2g+1}{Nh}\right)-\left(1-\frac{g-1}{h}-\frac{(n-1)(g-1)}{(N-n)h}\right)\right]$$

$$=\frac{N-n}{h}\left[g\left(1-\frac{2}{N}+\frac{n-1}{N-n}\right)-\left(1+\frac{n-1}{N-n}-\frac{1}{N}\right)\right].$$

so

$$E_1 \leqslant E_3 \forall\ O \leqslant g \leqslant 1+\frac{1}{N-2+\frac{N(n-1)}{N-n}}$$

and

$$E_1 \geqslant E_3 \forall \quad g \geqslant 1+\frac{1}{N-2+\frac{N(n-1)}{N-n}}$$

In most cases g being less than 1, it is thus seen that with the present model the HTE may often turn out worse than the ratio estimator based on SRSWOR scheme if we have reasons to believe that the bias of the latter may be supposed to be negligible.

We do not compare here between E_2, and E_3 because their relative magnitudes are known, vide Rao [5] even without the assumption of the nature of the distribution of X_i's.

SUMMARY

Some results are derived concerning the relative efficiencies of sampling strategies involving the Horvitz-Thompson estimator based on some πps sampling scheme and the ratio estimator based on SRSWOR and Sen-Mitzuno sampling scheme for estimating the finite population total of a real variate. Our study is based on a customary super-population model along with an additional assumption of identical, and independent uniparametric gamma distribution of the auxiliary variate-values.

ACKNOWLEDGEMENT

The author wishes to thank Dr. Arijit Chaudhuri for his help and guidance.

REFERENCES

[1] Cochran, W. G. (1963) ; Sampling Technique, Wiley, New York.

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

[3] Horvitz, D.G. Thompson D. J. (1952) : A generalization of sampiing without replacement from a finite universe. J. Amer, Statist Assoc 4J, 663-685.

[4] Midzuno, H (1952) : On the sampling system with probability pre-proportional to sumof sizes. Ann. Inst. Statist Math. 3, 99-107.

[5] Rao, T. J. (1967) : On the choice of a strategy for the ratio method of estimation. J. Roy. Statist, Soc. Ser. B, 29. 392-397.

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

[7] Sen A.R. (1953) : On the estimation of variance in sampling with varying probabilities. J. Iad. Soc. Agr.-Statist, 5, 119-127.

[8] Singh, Rabindra (1975) : A note on the efficiency of ratio estimate with Midzuno's scheme of sampling. Sankhya, Ser. C 37. 3. 211-214.

[9] Quenouille, M. H. (1956) : Notes on bias in estimatian. Biometrika 43, 353-360.