COMMUN. STATIST.-THEORY METH., 20(12), 3933-3941 (1991)

ON THE PERFORMANCE OF THE NEAREST PROPORTIONAL TO SIZE SAMPLING DESIGN

Arun Kumar Adhikary[1]
Department of Mathematics
University of Nairobi
P.O. Box 30197
Nairobi, Kenya.

Key Words and Phrases : Finite population; Generalized difference estimator; Nearest proportional to size sampling design; Ratio-cum-generalized difference estimator.

ABSTRACT

Assuming a super-population model the expected variance of the generalized difference estimator (Basu,1971) based on the nearest proportional to size sampling design introduced by Gabler(1987) is shown to be less than that of the same estimator based on an arbitrary sampling design from which the former design is realized. The former strategy is also shown to fare better than an

[1] On leave from Indian Statistical Institute,Calcutta.

unbiased ratio-cum-generalized difference estimator based on the nearest proportional to size sampling design in the sense of having less expected design variance under the same model.

1.INTRODUCTION

Consider a finite population U of size N and let $y_i(i=1,\ldots N)$be the values of a variate y under enquiry. our problem is to estimate the the population total $Y = \sum_{i=1}^{N} y_i$ on the basis of a sample of a fixed size n drawn from the population with a probability $p_o(s)$.

Gabler (1987) has introduced the nearest proportional to size sampling design $p^*(s)$defined as

$$p^*(s)= \left[\sum_{i\in s} \lambda_i \right] p_o(s) \tag{1.1}$$

where λ_i's $(i=1,\ldots,N)$ are all positive and are given by

$$\underset{\sim}{\Pi_o} \underset{\sim}{\lambda} = \underset{\sim}{\Pi}^* \tag{1.2}$$

where

$$\underset{\sim}{\Pi_o} = \begin{bmatrix} \pi_1^o & \pi_{12}^o & \cdots\pi_{1N}^o \\ \pi_{21}^o & \pi_2^o & \cdots & \pi_{2N}^o \\ \cdots & \cdots & \cdots & \cdots \\ \pi_{N1}^o & \pi_{N2}^o & \cdots & \pi_N^o \end{bmatrix}$$

$\underset{\sim}{\lambda} =(\lambda_1,\ldots,\lambda_N)^T$ and $\underset{\sim}{\Pi}^* =(\pi_1^*,\ldots,\pi_N^*)^T$, $\pi_i^o(\pi_i^*)$'s being the first order inclusion probabilities for the units for the sampling design $p_o(s)$ $\left[p^*(s)\right]$and π_{ij}^{o}'s being the second order inclusion probabilities for the pairs of units for the design $p_o(s)$.

Gabler(1987) has also discussed how to realise $p^*(s)$ starting from an arbitrary fixed sample size(n) design $p_0(s)$ and has called such a design a π^*ps design which satisfies $\sum_{i=1}^{N} \pi_i^* = n$.

Let $t_1 = \sum_{i\in s} \frac{y_i - \theta_i}{\pi_i^*} + \sum_{i=1}^{N} \theta_i$ be a generalized difference estimator (Basu,1971) based on such a π^*ps design for some real numbers $\theta_{i,}$ $i=1,\ldots,N$. known or otherwise. Our purpose here is to investigate whether t_1 fares better than the same estimator based on the original design $p_0(s)$ viz.

$$t_2 = \sum_{i\in s} \frac{y_i - \theta_i}{\pi_i^0} + \sum_{i=1}^{N} \theta_i \ .$$

As the classical ratio estimator is known to be unbiased under the Midzuno-Sen sampling scheme (Midzuno, 1952; Sen, 1953) and as π^*ps design is a proceeding of the Midzuno-Sen sampling scheme, we may consider the following ratio-cum-generalized difference estimator

$$t_3 = \frac{\sum_{i\in s} \frac{y_i - \theta_i}{\pi_i^0}}{\sum_{i\in s} \lambda_i} + \sum_{i=1}^{N} \theta_i \ .$$

which is also unbiased under $p^*(s)$.

The motivation for introducing the above ratio-cum-generalized difference estimator is eventually to compare its relative efficiency with that of t_1 and t_2.

2. A MODEL AND THE RESULTS

To compare the relative efficiencies of the above strategies we postulate a super-population model M specified by

$$E_m\,(y_i) = \theta_i\ ,\qquad V_m\,(y_i) = E_m(y_i - \theta_i)^2 = \sigma_i^2$$

and $\quad C_m(y_i\ ,\ y_j) = E_m(y_i-\theta_i)(y_j-\theta_j) = 0 \quad \forall\ i\neq j,$

where σ_i's are any positive real numbers $\forall\ i$.

Writing E_{p^*}(V_{p^*})as an operator for the expectation (variance) with respect to the sampling design p^*, we have the following theorem

Theorem 1. Under the model M, we have

$$E_m\,V_{p^*}\,(t_3) \geq E_m\,V_{p^*}\,(t_1) \tag{2.1}$$

Proof. Following Godambe and Thompson (1977), we can write

$$E_m\,V_{p^*}\,(t_1) = E_{p^*}V_m(t_1) + E_{p^*}\Delta_m^2(t_1) - V_m(Y)$$

$$\text{where}\quad \Delta_m(t_1) = E_m(t_1) - E_m(Y).$$

Now under the model M, we have $\Delta_m(t_1) = 0$ and hence

$$E_m V_{p^*}(t_1) = \sum_{i=1}^{N} \sigma_i^2 \left[\frac{1}{\pi_i^*} - 1\right] \tag{2.2}$$

Similarly we may check that

$$E_m\,V_{p^*}\,(t_3) = \sum_{i=1}^{N} \sigma_i^2 \left[\frac{1}{\pi_i^{o2}} \sum_{s\ni i} \frac{p_o(s)}{\sum_{i\in s} \lambda_i} - 1\right] \tag{2.3}$$

By Cauchy-Schwarz inequality it follows that

$$\left[\sum_{s\ni i}\left(\sum_{i\in s}\lambda_i\right)p_o(s)\right]\left[\sum_{s\ni i}\frac{p_o(s)}{\sum_{i\in s}\lambda_i}\right]\geq\left[\sum_{s\ni i}p_o(s)\right]^2$$

$$\Rightarrow\ \frac{1}{\pi_i^{o2}}\sum_{s\ni i}\frac{p_o(s)}{\sum_{i\in s}\lambda_i}\geq\frac{1}{\pi_i^{*}} \qquad (2.4)$$

Now, $E_m\,V_p{*}\,(t_3)-E_m\,V_p{*}\,(t_1)$

$$=\sum_{i=1}^{N}\sigma_i^2\left[\frac{1}{\pi_i^{o2}}\sum_{s\ni i}\frac{p_o(s)}{\sum_{i\in s}\lambda_i}-\frac{1}{\pi_i^{*}}\right]\geq 0$$

by using (2.4).

Remark 1. The equality holds when $\sum_{i\in s}\lambda_i$ is constant for all s with $p_o(s) > 0$ i.e when $\lambda_i = \frac{1}{n}\ \forall\ i$ which satisfies $\sum_{i=1}^{N}\lambda_i\pi_i^o = 1$ and in that case t_1 concides with t_3.

Let us now consider a simplified form of the above model (to be called model M_1) when $\sigma_i^2 = \sigma^2\lambda_i\pi_i^o\pi_i^*$. where σ is a positive real number.

Writing $E_{p_o}\left(V_{p_o}\right)$ as an operator for expectation (variance) with respect to the sampling design p_o, we have the follwing theorem.

Theorem 2. Under the model M_1,we have

$$E_m\,V_{p_o}\left(t_2\right)\geq E_m\,V_p{*}\left(t_1\right) \qquad (2.5)$$

Proof. We have

$$E_m\,V_{p_o}\left(t_2\right)=\sum_{i=1}^{N}\sigma_i^2\left(\frac{1}{\pi_i^o}-1\right)$$

so that

$$E_m V_{p_O}\left(t_2\right) - E_m V_{p^*}\left(t_1\right)$$

$$= \sum_{i=1}^{N} \sigma_i^2 \left[\frac{1}{\pi_i^o} - \frac{1}{\pi_i^*}\right]$$

$$= \sigma^2 \sum_{i=1}^{N} \lambda_i \left[\pi_i^* - \pi_i^o\right] \geq 0$$

because $$\sum_{i=1}^{N} \lambda_i \left[\pi_i^* - \pi_i^o\right]$$

$$= \sum_{i=1}^{N} \lambda_i \pi_i^* - 1 \quad \text{as} \quad \sum_{i=1}^{N} \lambda_i \pi_i^o = 1$$

$$= \sum_{i=1}^{N} \lambda_i \sum_{s \ni i} p^*(s) - 2+1$$

$$= \sum_{s} \left[\sum_{i \in s} \lambda_i\right] p^*(s) - 2\sum_{s} p^*(s) + 1$$

$$= \sum_{s} \left[\sum_{i \in s} \lambda_i\right]^2 p_o(s) - 2\sum_{s} \left[\sum_{i \in s} \lambda_i\right] p_o(s) + 1$$

$$= \sum_{s} \left[\sum_{i \in s} \lambda_i - 1\right]^2 p_o(s) \geq 0.$$

Remark 2. We may note that the quantity $\sum_{i=1}^{N} \lambda_i \left[\pi_i^* - \pi_i^o\right]$ is the directed distance from the design p_o to the design p^* as introduced by Gabler(1987).

Remark 3. Here the equality holds when $\sum_{i \in s} \lambda_i = 1$ for all s with $p_o(s) > 0$ i.e when $\lambda_i = \frac{1}{n} \forall i$ in which case π_i^* coincides with π_i^o resulting no difference between t_1 and t_2.

Under another simplified version of the above model M (to be called model M_2) when $\sigma_i^2 = \sigma_o^2 \pi_i^o \pi_i^*$ where σ_o is a positive real number, we have the following theorem.

Theorem 3. Under the model M_2, we have

$$E_m V_{p^*}\left[t_3\right] \ge E_m V_{p_o}\left[t_2\right] \qquad (2.6)$$

Proof. We have

$$E_m V_{p^*}\left[t_3\right] - E_m V_{p_o}\left[t_2\right]$$

$$= \sum_{i=1}^{N} \sigma_i^2 \left[\frac{1}{{\pi_i^o}^2} \sum_{s \ni i} \frac{p_o(s)}{\sum_{i \in s} \lambda_i} - \frac{1}{\pi_i^o}\right]$$

$$\ge \sum_{i=1}^{N} \sigma_i^2 \left[\frac{1}{\pi_i^*} - \frac{1}{\pi_i^o}\right] \quad \text{by using (2.4)}$$

$$= \sigma_o^2 \sum_{i=1}^{N} \left[\pi_i^o - \pi_i^*\right]$$

$$= 0.$$

Remark 4. Here also the equality holds when $\sum_{i \in s} \lambda_i$ is constant for all s with $p_o(s) > 0$ i.e. when $\lambda_i = \frac{1}{n} \; \forall \; i$ in which case there is nothing to choose between t_2 and t_3.

Remark 5. Under the model M_2, t_1 and t_2 have the same expected design-variance i.e. $E_m V_{p^*}\left[t_1\right] = E_m V_{p_o}\left[t_2\right]$.

Remark 6. We may note that unlike Theorem 1, in Theorem 2(Theorem 3), the model variance $V_m(y_i)$ is assumed to be proportional to $\lambda_i \pi_i^o \pi_i^* \left[\pi_i^o \pi_i^*\right]$ which is nearly proportional to $p_i^2(p_i)$, p_i's being the normed size measures of the units. Similar assumptions regarding σ_i^2 are also available in the literature. For example, Cassel, Särndal and Wretman(1976) investigated optimal strategies for estimating Y within a class of linear estimators under a super-population model in the sense of attaining a lower bound on the model-expected design-variance of an estimator.They found that the lower bound is attained by a generalized difference estimator based on a sampling design with inclusion probabilities proportional to known size-measures(W_i's,say) only when the model-expectations and standard deviations are proportional to W's.

ACKNOWLEDGEMENT

The author is grateful to the referee for helpful suggestions that led to an appreciable improvement over an earlier draft.

BIBLIOGRAPPHY

Basu, D.(1971). An essay on the logical foundations of survey sampling. PartI in : V.P. Godambe and D.A. Sprott edition, *Foundations of Statistical Inference*, Holt, Rinchart and Winston, Toronto, 203-242.

Cassel, C. M., Särndal, C. E. and Wretman, J. H. (1976). Some results in generalized difference estimation and generalized regression estimation for finite populations. *Biometrika,* 63, 615-620.

Gabler,S. (1987). The nearest proportional to size sampling design . *Comm. statist. - Theor. Meth.*, 16, No.4, 1117-1131.

Godambe, V. P. and Thompson, M. E. (1977). Robust near optimal estimation in survey practice. *Bull. Int. Statist. Inst.*,47, 129-146.

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

Sen, A. R.(1953). On the eatimation of variance in sampling with varying probabilities. *J. Ind. Soc. Agri. Statist.*, 5, 119-127.

Received June 1990; Revised April 1991.

Recommended by P. S. R. S. Rao, University of Rochester, Rochester, NY.