Sankhyā: The Indian Journal of Statistics
2001, Volume 63, Series B, Pt. 3, pp. 375-382

ON THE ASSESSMENT OF AN ADAPTED ESTIMATION STRATEGY

By V.R. PADMAWAR
Indian Statistical Institute

SUMMARY. It is well known that the continuous survey sampling set-up facilitates the assessment of estimation strategies, especially of the mathematically cumbersome ones. This makes it easier to interpret and grasp some of the complexities of modern survey sampling theory of finite populations. The continuous set-up also gives rise to new estimation strategies. It would be worthwhile to investigate whether some of these new strategies can be adapted to the finite set-up and if so, how they compare with the existing known strategies in the finite set-up.

1. **Introduction**

Consider a finite population $U = \{1,\ 2,\ \cdots,\ N\}$. Let y be the study variate taking values y_i on units i, $1 \le i \le N$. Let x be an auxiliary variate, closely related to y, taking values x_i on units i, $1 \le i \le N$. We assume that $y_1,\ y_2,\ \cdots,\ y_{\text{N}}$ are a realization of the variables $Y_1,\ Y_2,\ \cdots,\ Y_{\text{N}}$; the joint distribution of which is not fully known but specified by the first two moments as follows. If $E_\xi,\ V_\xi$ denote expectation and variance, respectively, w.r.t. the model ξ that defines a class of distributions for $Y_1,\ Y_2,\ \cdots,\ Y_{\text{N}}$; then

$$\left.\begin{array}{rcll} E_\xi(Y_i) & = & \beta x_i & i = 1,\ 2,\ \cdots,\ N \\ V_\xi(Y_i) & = & \sigma^2 x_i^g & i = 1,\ 2,\ \cdots,\ N \\ E_\xi(Y_iY_j) & = & \beta^2 x_i x_j & i \ne j = 1,\ 2,\ \cdots,\ N \end{array}\right\} \qquad (1.1)$$

where $\sigma^2 > 0$ and β are the unknown model parameters and $g \in [0,\ 2]$ is known. Model (1.1) is the well known regression model.

Survey statistician has to estimate the population mean $\overline{Y} = \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} y_i$ using a strategy $(p,\ t)$, say, that consists of a noninformative design p and an estimator $t(s,\ \mathbf{y})$, where $\mathbf{y} = (y_1,\ y_2,\ \cdots,\ y_{\text{N}})$. We use the measure of uncertainty

$$M(\ p,\ t) = E_\xi E_p(t - \overline{Y}\ )^2, \qquad (1.2)$$

as a measure of performance of the strategy $(p,\ t)$, where E_p denotes expectation w.r.t. the design p.

Paper received April 1999; revised April 2000.
AMS (1991) *subject classification.* Primary 62D05; secondary 15A48.
Key words and phrases. Continuous set-up, finite set-up, regression model, unbiased strategy.

A strategy $(p,\ t)$ is said to be $p-$unbiased or design unbiased for estimating the population mean $\overline{Y}$ if

$$E_p(t(s,\ \mathbf{y})) = E_p(t) = \sum_{s\in S} p(s)t(s,\ \mathbf{y}) = \overline{Y} \quad \forall\, \mathbf{y},$$

where S is the collection of all samples.

A strategy $(p,\ t)$ is said to be $\xi-$unbiased or model unbiased for the population mean $\overline{Y}$ if

$$E_\xi(t(s,\ \mathbf{y})) = E_\xi(\overline{Y}) \quad \forall\, s \text{ with } p(s) > 0.$$

Now for a $p-$unbiased strategy $(p,\ t)$

$$M(p,\ t) = E_p V_\xi(t) + E_p\left(E_\xi(t - \overline{Y})\right)^2 - V_\xi(\overline{Y}),$$

and for a $p-$unbiased as well as $\xi-$unbiased strategy $(p,\ t)$

$$M(p,\ t) = E_p V_\xi(t) - V_\xi(\overline{Y}).$$

2. **Strategy** $(p_g,\ t_g)$

In the continuous framework, [vide Padmawar(1996, 1998)], the strategy $(p_g,\ t_g)$ was introduced and studied. The strategy $(p_g,\ t_g)$ consists of estimator t_g given by

$$t_g = \frac{\mu}{\sum_{i=1}^{n} x_i^{2-g}} \sum_{i=1}^{n} x_i^{1-g} y(x_i)\ ,\ g \in [0,\ 2]\ ,$$

and the sampling design p_g given by

$$p_g(\mathbf{x}) = k \prod_{i=1}^{n} x_i^{g-1} \sum x_i^{2-g}\ ,$$

where $k = \frac{1}{n\alpha}\left[\frac{\Gamma(\alpha)}{\Gamma(\alpha+g-1)}\right]^{n-1}$.

In this note our objective is to explore whether we can adapt the strategy $(p_g,\ t_g)$ to the finite set-up.

We would use the same notation $(p_g,\ t_g)$ for the strategy adapted to the finite set-up. We first state a theorem of alternatives due to Farkas (vide Mangasarian (1969), pp. 16-17).

Theorem 2.1. *Let B be any $m \times n$ matrix then*

either $B\mathbf{z} = \mathbf{b}$, $\mathbf{z} \geqq \mathbf{0}$ *has a solution* $\mathbf{z} \in R^n$,
or $B'\mathbf{w} \leqq \mathbf{0}$, $\mathbf{b}'\mathbf{w} > \mathbf{0}$ *has a solution* $\mathbf{w} \in R^m$,

but never both.

We now prove a lemma that would be used later in this section.

LEMMA 2.1. *Let* $x_1,\ x_2,\ \cdots,\ x_{\rm N} > 0$ *be such that*

$$x_m^{g-1} = \max_{1\le i\le {\rm N}} x_i^{g-1} \le \frac{1}{n}\sum_{i=1}^{\rm N} x_i^{g-1} \ .$$

If $w_1,\ w_2,\ \cdots,\ w_{\rm N},\ w_{\rm N+1}$ *are such that for every* n *distinct labels* $1 \le i_1,\ i_2,\ \cdots,\ i_n \le N$

$$\sum_{j=1}^{n} w_{i_j} + w_{\rm N+1}\sum_{j=1}^{n} x_{i_j}^{2-g} \le 0, \tag{2.1 }$$

then

$$\sum_{i=1}^{\rm N} w_i x_i^{g-1} + w_{\rm N+1}\sum_{i=1}^{\rm N} x_i \le 0. \tag{2.2 }$$

PROOF. Here x_i , $1 \le i \le N$, are such that

$$x_m^{g-1} = \max_{1\le i\le {\rm N}} x_i^{g-1} \le \frac{1}{n}\sum_{i=1}^{\rm N} x_i^{g-1}.$$

This condition guarantees the existence of a fixed size (n) $\pi p x^{g-1}$ design, say $q(s)$, where $\pi p x^a$ is a sampling design for which the inclusion probabilities satisfy $\pi_i \ \propto\ x_i^a$, $1 \le i \le N$.

From equation (2.1), we get, for every sample s of size n,

$$\sum_{i\in s} w_i + w_{\rm N+1}\sum_{i\in s} x_i^{2-g} \le 0.$$

Now taking expectation w.r.t. $\pi p x^{g-1}$ design q , we have,

$$\sum_{s\in S}\left\{\sum_{i\in s} w_i\right\} q(s) + w_{\rm N+1}\sum_{s\in S}\left\{\sum_{i\in s} x_i^{2-g}\right\} q(s) \le 0.$$

Changing the order of summation, we get,

$$\sum_{i=1}^{\rm N} w_i\left\{\sum_{s\ni i} q(s)\right\} + w_{\rm N+1}\sum_{i=1}^{\rm N} x_i^{2-g}\left\{\sum_{s\ni i} q(s)\right\} \le 0.$$

But $q(s)$ being a fixed size (n) $\pi p x^{g-1}$ design, $\sum\limits_{s\ni i} q(s)$ is proportional to x_i^{g-1} , $1 \le i \le N$. Therefore,

$$\sum_{i=1}^{\rm N} w_i x_i^{g-1} + w_{\rm N+1}\sum_{i=1}^{\rm N} x_i \le 0.$$

This proves the lemma.

Let us first adapt the estimator t_g to the finite set-up as

$$t_g(s,\ \mathbf{y}) = \frac{\overline{X}}{\sum\limits_{i\in s} x_i^{2-g}} \sum_{i\in s} x_i^{1-g} y_i\ , \tag{2.3 }$$

where $N\overline{X} = X = \sum\limits_{i=1}^{N} x_i$.

It is now natural to ask whether there exists a design that makes the estimator t_g given by (2.3) $p-$unbiased.

For $s \in S$ and $1 \le i \le N$, define indicator function $I_s(\cdot)$ as

$$\begin{aligned} I_s(i) &= 1 \text{ if } i \in s \\ &= 0 \text{ if } i \notin s. \end{aligned}$$

For convenience, let $d(s) = \sum\limits_{i\in s} x_i^{2-g}$ and let us denote the samples by $s = 1,\ 2,\ \cdots,\ M,$ where $M = \binom{N}{n}$.

Now the problem of finding a design that makes the estimator t_g given by (2.3) $p-$unbiased is equivalent to solving the following system of equations.

$$\begin{aligned} B\mathbf{z} &= \mathbf{b}, \\ \mathbf{z} &\ge \mathbf{0} \end{aligned} \tag{2.4 }$$

where

$$B = \begin{bmatrix} \frac{I_1(1)}{d(1)} & \frac{I_2(1)}{d(2)} & \cdots & \frac{I_M(1)}{d(M)} \\ \frac{I_1(2)}{d(1)} & \frac{I_2(2)}{d(2)} & \cdots & \frac{I_M(2)}{d(M)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{I_1(N)}{d(1)} & \frac{I_2(N)}{d(2)} & \cdots & \frac{I_M(N)}{d(M)} \\ 1 & 1 & \cdots & 1 \end{bmatrix},\quad \mathbf{z} = \begin{bmatrix} p(1) \\ p(2) \\ \vdots \\ p(M) \end{bmatrix} \text{ and } \mathbf{b} = \frac{1}{X}\begin{bmatrix} x_1^{g-1} \\ x_2^{g-1} \\ \vdots \\ x_N^{g-1} \\ X \end{bmatrix} \tag{2.5 }$$

THEOREM 2.2. *Let* $x_1,\ x_2,\ \cdots,\ x_N > 0$ *be such that*

$$x_m^{g-1} = \max_{1\le i\le N} x_i^{g-1} \le \frac{1}{n}\sum_{i=1}^{N} x_i^{g-1}\ ,$$

then the system of equations (2.4) is consistent.

PROOF. First of all observe that $\mathbf{b}$ given by (2.5) satisfies $\mathbf{b} > \mathbf{0}$. Therefore, $\mathbf{z} = \mathbf{0}$ is not a solution to (2.4). Now in view of Theorem 2.1 it is enough to show that $B'\mathbf{w} \leqq \mathbf{0}$ and $\mathbf{b}'\mathbf{w} > 0$ is not consistent where B and $\mathbf{b}$ are given by (2.5).

If possible, let there exist $\mathbf{w}' = (w_1,\ w_2,\ \cdots,\ w_{\mathrm{N}},\ w_{\mathrm{N+1}})$ such that $B'\mathbf{w} \leqq \mathbf{0}$ and $\mathbf{b}'\mathbf{w} > 0$.

$$\mathbf{b}'\mathbf{w} > 0 \Leftrightarrow \sum_{i=1}^{\mathrm{N}} w_i x_i^{g-1} + w_{\mathrm{N+1}} \sum_{i=1}^{\mathrm{N}} x_i > 0. \tag{2.6}$$

Further we have, for any n distinct labels $i_1,\ i_2,\ \cdots,\ i_n,$ because of $B'\mathbf{w} \leqq \mathbf{0}$,

$$\sum_{j=1}^{n} w_{i_j} + w_{\mathrm{N+1}} \sum_{j=1}^{n} x_{i_j}^{2-g} \leq 0. \tag{2.7}$$

In view of Lemma 2.1, from (2.7), we have,

$$\sum_{i=1}^{\mathrm{N}} w_i x_i^{g-1} + w_{\mathrm{N+1}} \sum_{i=1}^{\mathrm{N}} x_i \leq 0.$$

This is a contradiction to (2.6). This proves that the system (2.4) is consistent.

This enables us to obtain a design p that makes the estimator (2.3) $p-$unbiased.

The proof of existence of a design that makes the estimator (2.3) $p-$unbiased, in fact, establishes the existence of a fixed size (n) design that satisfies (2.4). Let p_1 be any fixed size (n) design satisfying (2.4). We now prove the following theorem.

THEOREM 2.3. *The strategy* $(p_1,\ t_g)$ *is unique up to design in the sense that* $M(p_1,\ t_g)$ *, under model (1.1), is same for all fixed size (n) designs* p_1 *satisfying (2.4).*

PROOF. It is enough to check that $E_{p_1}V_\xi(t_g)$ is same for all fixed size (n) designs p_1 satisfying (2.4).

$$\begin{aligned} E_{p_1}V_\xi(t_g) &= \sigma^2 \sum_{i=1}^{\mathrm{N}} x_i^g \sum_{s\ni i} \overline{X}^2 x_i^{2-2g} \frac{p_1(s)}{(d(s))^2} \\ &= \sigma^2 \overline{X}^2 \sum_{s\in S} \frac{p_1(s)}{d(s)}. \end{aligned}$$

But note that $(p_1,\ t_g)$ is $p-$unbiased, hence

$$\begin{aligned} & \sum_{s\ni i} \overline{X} x_i^{1-g} \frac{p_1(s)}{d(s)} = \frac{1}{N} \quad \forall i = 1,\ 2,\ \cdots,\ N \\ \Rightarrow \quad & \sum_{s\ni i} \frac{p_1(s)}{d(s)} = \frac{x_i^{g-1}}{X} \\ \Rightarrow \quad & \sum_{i=1}^{\mathrm{N}} \sum_{s\ni i} \frac{p_1(s)}{d(s)} = \sum_{i=1}^{\mathrm{N}} \frac{x_i^{g-1}}{X} \\ \text{or} \quad & \sum_{s\in S} \frac{p_1(s)}{d(s)} = \frac{1}{nX} \sum_{i=1}^{\mathrm{N}} x_i^{g-1}, \end{aligned}$$

which is independent of choice of p_1.

Thus $M(p_1, t_g)$ is same for all (p_1, t_g), p_1 being fixed size (n) design satisfying (2.4). This completes the proof.

Having established, theoretically, the existence of a design that makes the estimator t_g $p-$unbiased, we now try to see whether such a design can actually be constructed as in the continuous set-up.

Let us again assume that $x_1, x_2, \cdots, x_{\text{N}} > 0$ be such that

$$x_m^{g-1} = \max_{1 \le i \le \text{N}} x_i^{g-1} \le \frac{1}{n} \sum_{i=1}^{\text{N}} x_i^{g-1},$$

This condition guarantees the existence of a fixed size (n), $\pi p x^{g-1}$ design, say $q(s)$. Now define another fixed size (n) design $p_g(s)$ as,

$$p_g(s) = k q(s) d(s),$$

where

$$\begin{aligned} k^{-1} &= \sum_{s \in S} q(s) d(s) = \sum_{s \in S} \left\{ q(s) \sum_{i \in s} x_i^{2-g} \right\} \\ &= \sum_{i=1}^{\text{N}} \left\{ x_i^{2-g} \sum_{s \ni i} q(s) \right\} \\ &= \sum_{i=1}^{\text{N}} \left\{ x_i^{2-g} \frac{n x_i^{g-1}}{\sum_{i=1}^{\text{N}} x_i^{g-1}} \right\} = \frac{n \sum_{i=1}^{\text{N}} x_i}{\sum_{i=1}^{\text{N}} x_i^{g-1}} . \end{aligned}$$

We now prove

THEOREM 2.4. *The strategy* (p_g, t_g) *is* $p-$*unbiased for estimating the population mean* $\overline{Y}$.

PROOF. Observe that for $1 \le i \le N$,

$$\begin{aligned} \sum_{s \ni i} \frac{\overline{X} x_i^{1-g}}{d(s)} p_g(s) &= k \overline{X} x_i^{1-g} \sum_{s \ni i} q(s). = k \overline{X} x_i^{1-g} \frac{n x_i^{g-1}}{\sum_{i=1}^{\text{N}} x_i^{g-1}} \\ &= \left[\frac{n \sum_{i=1}^{\text{N}} x_i}{\sum_{i=1}^{\text{N}} x_i^{g-1}} \right]^{-1} \frac{n \overline{X}}{\sum_{i=1}^{\text{N}} x_i^{g-1}} = \frac{1}{N} . \end{aligned}$$

This proves that the strategy (p_g, t_g) is $p-$unbiased for estimating the population mean $\overline{Y}$.

Having successfully adapted the strategy (p_g, t_g) to the finite set-up it remains to see how it compares with some of the well known existing design unbiased strategies for estimating the population mean $\overline{Y}$.

We consider the strategies $(p_{\text{M}}, t_{\text{R}})$, $(\pi p x, t_{\text{HT}})$, and the Rao-Hartley-Cochran strategy $(p_{\text{RHC}}, t_{\text{RHC}})$ where

πpx : sampling design for which the inclusion probabilities satisfy $\pi_i \propto x_i$.

$p_{\rm M}$: Midzuno-Sen sampling design for which $p_{\rm M}(s) = \frac{\sum_{i\in s} x_i}{\binom{N-1}{n-1} X}$.

$t_{\rm R}$: ratio estimator $\frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i}\, \overline{X}$.

$t_{\rm HT}$: Horvitz-Thompson estimator given by $\frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i}$, based on the design p for which $\pi_i = \sum_{s\ni i} p(s) > 0,\ 1 \le i \le N$.

Before taking up the assessment of the strategy $(p_g,\ t_g)$ it is in order, in view of Theorem 2.3, to note the following.

Remark 2.1. As a consequence of the approach adopted in the Theorem 2.3 we have, for $g = 1$, the strategy $(p_g,\ t_g)$ coincides with the strategy $(p_{\rm M},\ t_{\rm R})$ and for $g = 2$, it coincides with the strategy $(\pi px,\ t_{HT})$.

We now take up the comparison of strategies. As a matter of fact, all these strategies under consideration are $p-$unbiased as well as $\xi-$unbiased.

It is easy to see that for the strategy $(p_g,\ t_g)$

$$M(p_g,\ t_g) = \frac{\sigma^2\overline{X}}{nN}\sum_{i=1}^{N} x_i^{g-1} - \frac{\sigma^2}{N^2}\sum_{i=1}^{N} x_i^{g}. \tag{2.8}$$

We know from Rao (1967) that,

$$M(p_{\rm M},\ t_{\rm R}) = \frac{\sigma^2\overline{X}}{N\binom{N-1}{n-1}}\sum_{s\in S}\frac{\sum_{i\in s} x_i^g}{\sum_{i\in s} x_i} - \frac{\sigma^2}{N^2}\sum_{i=1}^{N} x_i^{g} \tag{2.9}$$

$$M(\pi px,\ t_{\rm HT}) = \frac{\sigma^2\overline{X}}{nN}\sum_{i=1}^{N} x_i^{g-1} - \frac{\sigma^2}{N^2}\sum_{i=1}^{N} x_i^{g}, \tag{2.10}$$

and that under model (1.1) for $n \ge 2$,

$$M(p_{\rm M},\ t_{\rm R}) \lesseqgtr M(\pi px,\ t_{\rm HT}) \quad \text{according as} \quad g \lesseqgtr 1. \tag{2.10a}$$

We also know from Hanurav (1965) and J. N. K. Rao (1966) that for $\frac{N}{n}$ an integer,

$$M(p_{\rm RHC},\ t_{\rm RHC}) = \frac{(N-n)\sigma^2}{nN^2(N-1)}\left[\sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i^{g-1} - \sum_{i=1}^{N} x_i^{g}\right], \tag{2.11}$$

and that under model (1.1), for $n \ge 2$ and $\frac{N}{n}$ an integer,

$$M(p_{\rm RHC},\ t_{\rm RHC}) \lesseqgtr M(\pi px,\ t_{\rm HT}) \quad \text{according as} \quad g \lesseqgtr 1. \tag{2.11a}$$

In view of (2.8) through (2.11a) we have the following result.

Theorem 2.5. *Under model (1.1), the strategy* $(p_g,\ t_g)$ *is as good as the strategy* $(\pi px,\ t_{\rm HT})$ *w.r.t. the measure of uncertainty* $M(p,\ t)$ *and that for* $n \geq 2$,

$$M(p_{\rm M},\ t_{\rm R}) \ \substack{\leq \\ = \\ >}\ M(p_g,\ t_g) \qquad according \quad as \quad g \ \substack{\leq \\ = \\ >}\ 1.$$

and further, if $\frac{N}{n}$ *is an integer, then*

$$M(p_{\rm RHC},\ t_{\rm RHC}) \ \substack{\leq \\ = \\ >}\ M(p_g,\ t_g) \qquad according \quad as \quad g \ \substack{\leq \\ = \\ >}\ 1.$$

Remark 2.2. In the finite set-up the adapted strategy $(p_g,\ t_g)$ satisfies many of the properties satisfied by its counterpart in the continuous set-up [vide Padmawar(1998)]. The statement comparing the strategies $(p_{\rm RHC},\ t_{\rm RHC})$ and $(p_g,\ t_g)$ in the finite set-up does not, however, agree with its counterpart in the continuous set-up. The reason for this is the fact that the strategy $(p_{\rm RHC},\ t_{\rm RHC})$ defined in the continuous set-up in Padmawar(1996) is only a Rao-Hartley-Cochran 'type' strategy.

Remark 2.3. The strategy $(\pi px,\ t_{\rm HT})$ has been extensively studied in the literature. Its performance has also been compared with various well known strategies w.r.t. the measure of uncertainty $M(p,\ t)$ under the regression model [vide Padmawar (1981, 1982), Chaudhuri and Vos (1988)]. In view of Theorem 2.5, similar results would also hold for the strategy $(p_g,\ t_g)$.

References

Chaudhuri, A. and Vos, J.W.E. (1988). *Unified Theory and Strategies of Survey Sampling*, North-Holland, Amsterdam.

Hanurav, T.V. (1965). *Optimum sampling strategies and some related problems*, Ph.D. Thesis, Indian Statistical Institute.

Mangasarian, O.L. (1969). *Nonlinear Programming*, McGraw Hill, New York.

Padmawar, V.R. (1981). A note on the comparison of certain sampling strategies *Journal of the Royal Statistical Society, Series B*, **43**, 321-326.

−−−− (1982). *Optimal strategies under superpopulation models*, Ph.D. Thesis, Indian Statistical Institute.

−−−− (1996). Rao-Hartley-Cochran strategy in survey sampling of continuous populations, *Sankhyā Series B*, **57**, 90-104.

−−−− (1998). A note on comparison of estimation strategies in survey sampling of continuous populations, *Sankhyā Series B*, **60**, 301-314.

Rao, J.N.K. (1966). On the relative efficiency of some estimators in pps sampling for multiple characteristics, *Sankhyā Series A*, **43**, 61-70.

Rao, T.J. (1967). On the choice of a strategy for the ratio method of estimation, *Journal of the Royal Statistical Society, Series B*, **29**, 392-397.

V.R. Padmawar
Stat−Math Division
Indian Statistical Institute
8th Mile, Mysore Road
Bangalore, 560 059, India
E-mail: vrp@isibang.ac.in