ASYMPTOTIC EFFICIENCY OF THE SYMMETRIZED DES RAJ STRATEGY-II

By P. K. PATHAK

University of Illinois and *Indian Statistical Institute*

SUMMARY. The symmetrized Des Raj strategy (SDR strategy) (Murthy, 1957 ; Pathak, 1961) is the Rao-Blackwellization of Des Raj's (1956) strategy. Consequently the SDR strategy possesses smaller sampling variance than Des Raj's strategy. The principal result of the present paper states that the reduction in variance due to the symmetrization of Des Raj's strategy is, under certain regularity conditions, asymptotically negligible.

1. Introduction and statement of the problem

It is believed that the symmetrized Des Raj strategy (SDR strategy), the Rao-Blackwellization of Des Raj's strategy (Pathak, 1961), provides estimates of high precision in many situations commonly met in parctice. Unfortunately except for the case when the sample size $n=2$ (Vijayan, 1966) no theoretical justification exists in the literature to substantiate this belief. It would, therefore, be of some interest to study the efficiency of this strategy in some detail. As an attempt in this direction, we obtain here, under certain regularity conditions, an asymptotic expression for the sampling variance of the SDR strategy. We also show that under certain regularity conditions the SDR and Des Raj's strategies are asymptotically equally efficient.

For each natural number k, let $\pi_k = \{U_{k1}, ..., U_{kj}, ..., U_{kN_k}\}$ be a population of N_k units, where U_{kj} denotes the j-th population unit. Unless otherwise stated the subscripts j and j' run from 1 through N_k. Let Y_{kj} and P_{kj} respectively denote the Y-characteristic value and the probability of selection associated with U_{kj}. We follow here the notation of the author's earlier papers (Pathak, 1961 and 1967). Suppose that a sample of n_k units is selected from π_k according to the probabilities P_{kj}'s and without replacement. Then under the SDR strategy, the estimator of the population total $Y_k = \sum_j Y_{kj}$ is given by

$$t_{k(m)} = \sum_i y_{k(i)} \frac{P(T_k|i)}{P(T_k)} \quad \text{(Murthy, 1957; Pathak, 1961)} \quad \dots \quad (1)$$

where $T_k = (u_{k(1)}, ..., u_{k(i)}, ..., u_{k(n_k)})$ denotes the order-statistic obtained from the observed sample $s_k = (u_{k1}, ..., u_{ki}, ..., u_{kn_k})$, and $P(T_k)$ and $P(T_k|i)$ respectively denote the probability of getting the order-statistic T_k and the conditional probability of getting the order-statistic T_k given that the i-th order unit, $u_{k(i)}$, was drawn first. The subscripts i and i' run from 1 through n_k unless otherwise stated.

An unbiased estimator of $V(t_{k(m)})$ is given by Pathak and Shukla (1956)

$$v(t_{k(m)}) = \frac{1}{2} \sum_{i,i'} \left(\frac{y_{k(i)}}{p_{k(i)}} - \frac{y_{k(i')}}{p_{k(i')}}\right)^2 \frac{p_{k(i)}p_{k(i')}[P(T_k)P(T_k|i,i') - P(T_k|i)P(T_k|i')]}{P^2(T_k)} \quad \ldots(2)$$

where $P(T_k|i, i')$ denotes the conditional probability of getting the order-statistic T_k given that the i-th and i'-th order units were drawn in the first two draws.

Hence

$$V(t_{k(m)}) = \frac{1}{2} \sum_{j,j'} \left(\frac{Y_{kj}}{P_{kj}} - \frac{Y_{kj'}}{P_{kj'}}\right)^2 P_{kj}P_{kj'}\left[1 - \sum_{jj'} \frac{P(T_k|j)P(T_k|j')}{P(T_k)}\right] \quad \ldots (3)$$

where the summation $\sum_{jj'}$ is taken over all samples cotaining U_{kj} and $U_{kj'}$.

2. An asymptotic expression for $V(t_{k(m)})$

We now proceed to obtain a simple asymptotic expression for $V(t_{k(m)})$ under certain regularity conditions. The following lemmas will be found useful for this purpose.

Lemma 2.1 : *If*

(i) $N_k \to \infty$ *as* $k \to \infty$,

(ii) $\overline{\lim}\, n_k < \infty$ *and* $\quad \ldots (4)$

(iii) $\max_j P_{kj} = 0(N_k^{-1})$,

then

$$P(T_k) = n_k!\, p_{k(1)} \ldots p_{k(n_k)} \left\{1 + \frac{(n_k-1)}{2}\left[p_{k(1)} + \ldots + p_{k(n_k)}\right] + 0(N_k^{-2})\right\} \quad \ldots (5)$$

where $P(T_k)$ *has been defined in* (1).

Proof : Let $s_k = \left(u_{k1}, \ldots, u_{kn_k}\right)$ denote an observed sample that gives rise to T_k. Then

$$P(s_k) = \left\{p_{k1} \ldots p_{kn_k}\right\}\left\{\frac{1}{1-p_{k1}) \ldots \left(1 - p_{k1} - \ldots - p_{k(n_k-1)}\right)}\right\} \quad \ldots (6)$$

since $(1-p_{k1}-\ldots-p_{kt})^{-1} = \exp\{p_{kt} + 0(N_k^{-2})\}$ under the regularity conditions (i)—(iii), we have

$$P(s_k) = \{p_{k1} \ldots p_{kn_k}\} \exp\{(n_k-1)p_{k1} + (n_k-2)p_{k2} + \ldots + p_{k(n_k-1)} + 0(N_k^{-2})\}$$

$$= \{p_{k1} \ldots p_{kn_k}\}\{1 + (n_k-1)p_{k1} + (n_k-2)p_{k2} + \ldots + p_{k(n_k-1)} + 0(N_N^{-2})\}. \quad \ldots (7)$$

The lemma follows on noting that $P(T_k) = \Sigma' P(s_k)$, where the summation is taken over $n_k!$ permutations of $p_{k1}, \ldots, p_{kn_k}$.

Lemma 2.2 : *Under the regularity conditions of Lemma* 2.1, *we have*

$$P(T_k \mid i) = \frac{n_k!\, p_{k(1)} p_{k(n_k)}}{n_k p_{k(i)}} \left\{ 1 + \frac{n_k}{2} p_{k(i)} + \frac{(n_k-2)}{2} (p_{k(1)} + \ldots + p_{k(n_k)}) + 0(N_k^{-2}) \right\} \quad \ldots \quad (8)$$

where $P(T_k \mid i)$ *is defined in* (1).

Proof : Clear from (5).

Lemma 2.3 : *Let* $\pi_k(U_{kj})$ *denote the probability of inclusion of* U_{kj} *in the sample. Then under the regularity conditions of Lemma* 2.1, *we have*

$$\pi_k(U_{kj}) = P_{kj} \left\{ n_k + \binom{n_k}{2} (S_{k2} - P_{kj}) + 0(N_k^{-2}) \right\} \quad \ldots \quad (9)$$

where $S_{k2} = \Sigma P_{kj}^2$.

Proof : Lemma 2.3 evidently holds for $n = 1$. Now if the lemma holds for $(n_k - 1)$, then

$$\pi_k(U_{kj}) = P_{kj} + \sum_{j' \neq j} P_{kj'} \frac{P_{kj}}{(1-P_{kj'})} \left\{ (n_k - 1) + \binom{n_k - 1}{2} \left(\frac{S_{k2}}{(1-P_{kj'})^2} - \frac{P_{kj'}^2}{(1-P_{kj'})^2} - \frac{P_{kj}}{(1-P_{kj'})} \right) + 0(N_k^{-2}) \right\}. \quad \ldots \quad (10)$$

An easy calculation shows that

$$\pi_k(U_{kj}) = P_{kj} \left\{ n_k + \binom{n_k}{2} (S_{k2} - P_{kj}) + 0(N_k^{-2}) \right\}. \quad \ldots \quad (11)$$

This completes the proof.

Lemma 2.4 : *Let* $\pi_k(U_{kj}, \overline{U}_{kj'})$ *denote the probability of inclusion of* U_{kj} *and non-inclusion of* $U_{kj'}$ *in the sample. Then under the regularity conditions of Lemma* 2.1

$$\pi_k(U_{kj}, \overline{U}_{kj'}) = P_{kj} \left\{ n_k + \binom{n_k}{2} (S_{k2} - P_{kj} - 2P_{kj'}) + 0(N_k^{-2}) \right\} \quad \ldots \quad (12)$$

Proof : Similar to the proof of Lemma 1.3.

The following theorem now gives us a simple asymptotic expression for $V(t_{k(m)})$.

Theorem 1 : *Under the regularity conditions of Lemma 2.1, we have*

$$V(t_{k(m)}) = \frac{1}{2n_k} \sum_{j,j'} \left(\frac{Y_{kj}}{P_{kj}} - \frac{Y_{kj'}}{P_{kj'}}\right)^2 P_{kj}\, P_{kj'} \left\{1 - \frac{(n_k-1)}{2}(P_{kj}+P_{kj'}) + 0(N_k^{-2})\right\}$$

$$= [1+0(N_k^{-2})]\left[\frac{1}{n_k}\sum_j P_{kj}\left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2\right] - \left[\frac{(n_k-1)}{2n_k}\right]\left[\left(\sum_j P_{kj}^2\right)\right.$$

$$\left.\sum P_{kj}\left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2 + \sum P_{kj}^2\left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2\right]. \quad \ldots \quad (13)$$

Proof : From Lemma 2.1 and (8) it is evident that

$$\sum_{jj'} \frac{P(T_k|j)\, P(T_k|j')}{P(T_k)}$$

$$= \frac{1}{2n_kP_{kj}} \sum_{jj'} \{1+n_kP_{kj}+[1-p_{k(1)}]-\ldots-p_{k(n_k)}+0(N_k^{-2})\}\; P(T_k|j') \quad \ldots \quad (14)$$

where $\sum_{jj'}$ has been defined in (3) and $p_{k(1)}, \ldots, p_{k(n_k)}$ are the probabilities of selection of the n_k units of T_k.

$$= \frac{1}{2n_kP_{kj}} \{1+n_kP_{kj}+0(N_k^{-2})\}\; \pi_k(U_{kj}|j') + \sum_{l \neq j \neq j' = 1}^{N_k} \frac{P_l}{2n_kP_{kj}} \pi_k(U_{kj}, \bar{U}_{kl}|j') \quad \ldots \quad (15)$$

where $\pi_k(U_{kj}|j')$ denotes the conditional probability of inclusion of U_{kj} given that U_{kj}' was selected at the first draw and $\pi_k(U_{kj}, \bar{U}_{kl}|j')$, the conditional probability of inclusion of U_{kj} and non-inclusion of $U_{kl'}$ given than U_{kj} was selected at the first draw.

From Lemmas 2.3 and 2.4 it follows that

$$\pi_k(U_{kj}|j') = (n_k-1)\, P_{kj}\left\{1+P_{kj'}+\frac{(n_k-2)}{2}(S_{k2}-P_{kj})+0(N_k^{-2})\right\} \quad \ldots \quad (16)$$

and

$$\pi_k(U_{kj}, \bar{U}_{kl}|j') = (n_k-1)P_{kj}\left\{1+P_{kj}+\frac{(n_k-2)}{2}(S_{k2}-P_{kj}-P_{kl})+0(N_k^{-2})\right\}. \quad \ldots \quad (17)$$

On substituting these values of $\pi_k(U_{kj}|j')$ and $\pi_k(U_{kj}, \bar{U}_{kl}|j')$ in (15), we get the following after some simplification.

$$\sum_{jj'} \frac{P(T_k|j)P(T_k|j')}{P(T_k)} = \frac{(n_k-1)}{2n_k}\{2+(P_{kj}+P_{kj'})+0(N_k^{-2})\}. \quad \ldots \quad (18)$$

Theorem 1 now follows from (3) and (18). This proves Theorem 1.

Under the regularity conditions of Lemma 2.1, Theorem 1 provides us with a simple asymptotic expression for $V(t_{k(m)})$. It is important to note here that the assumption $\overline{\lim}\ n_k < \infty$ has played a crucial role in our development and we do not know if this asymptotic expression to $V(t_{k(m)})$ remains valid when the assumption $\overline{\lim}\ n_k < \infty$ is violated. It would therefore by very interesting and rewarding to obtain a similar asymptotic expression to $V(t_{k(m)})$ when the assumption $\overline{\lim}\ n_k < \infty$ is, for example, relaxed by an assumption of the kind $\overline{\lim}(n_k \mid N_k) = 0$. In this connexion the following argument shows that the right side of (13) still provides us with an upper bound to $V(t_{k(m)})$ when the assumption $\overline{\lim}\ n_k < \infty$ is replaced by the assumption $\overline{\lim}\ (n_k \mid N_k) = 0$. Under the relaxed assumption the variance of Des Raj's estimator, $\bar{t}_{k(d)}$, satisfies the following inequiality (Pathak, 1967)

$$V(\bar{t}_{k(d)}) \leqslant \frac{1}{n_k} \sum_j P_{kj} \left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2 + \left\{1 + 0\left(\frac{n_k}{N_k}\right)\right\}\left\{\sum_j P_{kj}^2 \left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2 + \left(\sum_j P_{kj}^2\right) \sum_j P_{kj}\left(\frac{Y_{kj}}{P_{kj}} - Y_k\right)^2\right\}. \quad \ldots (19)$$

Since $E(\bar{t}_{k(d)} \mid T_k) = t_{k(m)}$ (Pathak, 1961), $V(t_{k(m)}) \leqslant V(\bar{t}_{k(d)})$. This shows that the right side of (13) provides an upper to $V(t_{k(m)})$ when $\overline{\lim}\ (n_k \mid N_k) = 0$.

In the next section we compare the SDR strategy with Des Raj's strategy.

3. Comparison with Des Raj's strategy

Let $\Sigma\, P_{kj}(Y_{kj}/P_{kj} - Y_k)^2 = 0(N_k^2)$. Then it is seen from (13) and (12) that $\lim\ N_k^{-1}\{V(\bar{t}_{k(d)}) - V(t_{k(m)})\} = 0$ when the regularity conditions of Lemma 2.1 hold. Consequently in this situation the two strategies can be taken to be asymptotically equally efficient.

Since the SDR strategy is computationally cumbersome when $n_k > 2$, it is perhaps appropriate, as a consequence of the above equivalence of the efficiencies of the two strategies, to prefer Des Raj's strategy to the SDR strategy when $n_k > 2$. The SDR strategy presents no computational difficulty when $n_k = 2$. Hence in this case the SDR strategy should be preferred to Des Raj's strategy.

4. Acknowledgement

The author wishes to thank the referee for his many helpful suggestions.

REFERENCES

DES RAJ (1956) : Some estimators in sampling with varying probabilities without replacement. *J. Amer. Stat. Assoc.*, 51, 269-84.

MURTHY, M. N. (1957) : Ordered and unordered estimators in sampling without replacement. *Sankhyā* 18, 379-90.

PATHAK, P. K. (1961) : Use of order-statistic in sampling without replacement. *Sankhyā*, Series A, 23, 409-14.

——— (1967) : Asymptotic efficiency of Des Raj's strategy–I. *Sankhyā*, Series A, 29, 283-298.

PATHAK, P. K. and SHUKLA, N. D. (1966) : Non-negativity of a variance estimator. *Sankhyā*, Series A 28, 41-46.

VIJAYAN, K. (1966) : On Horvitz-Thompson and Des Raj estimators. *Sankhyā*, Series A 28, 87-92.

Paper received : February, 1967.