Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 2 (1986), pp. 153-160

A SAMPLING SCHEME WITH INCLUSION PROBABILITY PROPORTIONAL TO SIZE USING PPS SYSTEMATIC SAMPLING

RANJANA AGRAWAL
IASRI, New Delhi
and
PADAM SINGH
Planning Commission, New Delhi

(Received : April, 1983)

SUMMARY

A πPS sampling has been proposed by modifying the usual PPS systematic sampling. On comparing the efficiency of proposed sampling scheme with some of the existing sampling schemes it has been observed that the proposed sampling scheme is equally efficient.

Keywords : PPS systematic sampling, πPS scheme; inclusion probability.

Introduction

PPS systematic sampling suggested by Madow [6] besides being simple, provides inclusion probabilities for different units exactly proportional to size measure. But it has an obvious drawback that it is not possible to obtain unbiased estimator of the variance which is one of the basic requirements of a good sampling scheme. In this paper a modified PPS systematic sampling scheme has been suggested which maintains the simplicity, inclusion probabilities are proportional to size and is free from the above drawbaak. The efficiency of the proposed scheme has been compared empirically with some others in use.

2. Sampling Scheme

Let the population under study consists of N distinct and identifiable units and a sample of size n is desired to be drawn from it. Let Y_i and X_i denote the values of the character under study and auxiliary character respectively for the ith unit of the population. Conventionally small letters y_i and x_i will correspond to the ith unit in the sample. Further it is assumed that X_i's are known for all i's and set $P_i = X_i \Big/ \sum_{i=1}^{N} X_i$.

The proposed scheme consists of the following steps :

a. Select one unit from N units of the population with probability P_i'

b. Select $(n - 1)$ units from the remaining $(N - 1)$ units by usual PPS systematic sampling scheme with probability for ith unit proportional to P_i.

For this sampling scheme the inclusion probability for ith unit is given by

π_i = Probability that ith unit is selected at step (*a*)
+ Probability that ith unit is not selected at step (*a*)
$\times$ Probability that ith unit is selected at step (*b*)

$$= P_i' + (n - 1) \sum_{j \neq i}^{N} P_j' \frac{P_i}{1 - P_j} \tag{1}$$

For πPS i.e. $\pi_i = nP_i$, we should have

$$P_i' + (n - 1) \sum_{j \neq i}^{N} P_j' \frac{P_i}{1 - P_j} = nP_i$$

or
$$P_i' = \frac{P_i(1 - P_i)\{n - (n - 1)s\}}{1 - nP_i} \tag{2}$$

where

$$s = \sum_{j=1}^{N} \frac{P_j'}{1 - P_j}$$

Summing (2) over all units and using $\Sigma P_i' = 1$ it is found

$$n - (n - 1)s = \frac{1}{\sum_{i=1}^{N} \frac{P_i(1 - P_i)}{1 - nP_i}} \tag{3}$$

Substituting (3) in (2) the solution for revised probabilities is obtained as

$$P_i' = \frac{P_i\,(1 - P_i)/(1 - nP_i)}{\sum\limits_{i=1}^{N} \{P_i(1 - P_i)/(1 - nP_i)\}} \tag{4}$$

Since $nP_i < 1$ (an assumption for all πPS sampling schemes) it is obvious that P_i's will always be positive.

Inclusion probability for a pair of units (i, j) is given as

$$\begin{aligned}\pi_{ij} = &\ (\text{Probability that } i\text{th unit is selected at step } (a))\\ &\times (\text{Probability that } j\text{th unit is selected at step } (b))\\ &+ (\text{Probability that } j\text{th unit is selected at step } (a))\\ &\times (\text{Probability that } i\text{th unit is selected at step } (b))\\ &+ (\text{Probability that none of } i \text{ and } j \text{ is selected at step } (a))\\ &\times (\text{Probability that both } i \text{ and } j \text{ are selected at step } (b))\end{aligned}$$

$$= (n - 1)\left[P_i' \frac{P_j}{1 - P_i} + P_j' \frac{P_i}{1 - P_j}\right] + \sum_{\substack{k \neq i \\ \neq j}}^{N} P_k' \,\pi_{ij(k)} \tag{5}$$

Here $\pi_{ij(k)}$ is the inclusion probability of pair (i, j) in PPS systematic sample of size $(n - 1)$ from the remaining $(N - 1)$ units given that kth unit is selected at step (a) and is given by

$$\pi_{ij(k)} = \sum_{t=1}^{n-1-m} \pi_{ij(k)t} \tag{6}$$

$\pi_{ij(k)t}$ denotes inclusion probability for pair (i, j) at step (b) such that jth unit $(j > i)$ is selected at tth subsequent draw given that ith unit can be included not earlier than mth draw. Calculation of $\pi_{ij(k)t}$ is described below.

Let S_i's be the cumulative totals of sizes i.e. $S_i = \sum\limits_{j=1}^{i} X_j$ and K the integer given by a suitable multiple of $S_{N-1}/n - 1$.

Following Singh and Singh (1978), $\pi_{ij(k)t}$ is given as

$$\begin{aligned}\pi_{ij(k)t} &= 0 \text{ if } S_j - S_{i-1} \leqslant Kt\\ &= 0 \text{ if } S_j - S_{i-1} > Kt\end{aligned}$$

and $\quad S_{j-1} - S_i \geqslant Kt$

$$= \min\left\{\frac{Kt - (S_{j-1} - S_i)}{K}; \frac{(n - 1)\,P_i}{1 - P_k}\right\}$$

if $\quad S_j - S_{i-1} > Kt$

$\quad\quad S_{j-1} - S_i < Kt$

and $$S_{j-1} - S_{i-1} > Kt \tag{7}$$

$$= \min \left\{ \frac{S_j - S_{i-1} - Kt}{K} ; \frac{(n-1) P_i}{1 - P_k} \right\}$$

if $$S_j - S_{i-1} > Kt$$
$$S_{j-1} - S_i < Kt$$

and $$S_{j-1} - S_{i-1} \leqslant Kt$$

After obtaining these inclusion probabilities, usual Horwitz Thompson estimator for population total alongwith its variance and variance estimate can be used.

TABLE 1—DESCRIPTION OF POPULATIONS

Sl. No.	*Source*	*N*	*Y*	*X*	*CV(Y)*	*CV(X)*	ρ
1.	Horvitz & Thompson [5]	20	No. of households	Eye estimate of *Y*	0.44	0.40	0.87
2.	Sukhatme [10], p. 279-80	89	Wheat acreage	No. of villages	0.65	0.60	0.65
3.	Sukhatme [10], p. 183	34	Wheat acreage in 1937	Wheat acreage in 1936	0.36	0.44	0.78
4.	Sampford [7]	35	Oats acreage in 1957	Total acreage in 1947	0.71	0.71	0.83
5.	Cochran [2], p. 156	49	No. of persons in 1930	No. of persons in 1920	0.95	1.08	0.91
6.	Cochran [2], p. 325	10	No. of persons per block	No. of rooms per block	0.15	0.14	0.65
7.	Hanurav [3]	20	Population in 1967	Population in 1957	0.30	0.30	0.97
8.	,,	19	,,	,,	0.45	0.44	0.97
9.	,,	17	,,	,,	0.66	0.65	0.99
10.	,,	16	,,	,,	0.51	0.52	0.96

3. Empirical Comparisons

To see the performance of the suggesed scheme, it has been compared empirically with following existing sampling schemes

(i) Sampling with probability proportional to size with replacement.
(ii) Randomised probability proportional to size systematic sampling scheme.
(iii) Sampford's sampling scheme.
(iv) Singh and Singh Scheme [9].

Ten natural populations were considered for the purpose. The source of the populations, size of population, nature of Y and X, $C.V.$ of Y and X and correlation coefficient (ρ) between Y and X are given in Table 1.

Besides these populations, two sets of hypothetical populations were also considered.

The first set consisted of three populations each of size 13 given in Table 2.

TABLE 2—DATA ON THREE HYPOTHETICAL POPULATIONS

Sl. No.	X	Y values under population I1	I2	I3
1	2	16	28	16
2	2	17	27	17
3	3	27	39	27
4	3	28.5	37.5	28.5
5	4	40	48	40
6	4	42	46	42
7	5	55	55	55
8	6	69	63	63
9	6	72	60	60
10	7	87.5	66.5	66.5
11	7	91	63	63
12	8	108	68	68
13	8	112	64	64

The value of auxiliary variable is same for all the three populations.

These are similar to those considered by Cochran [2]. For the first population Y/X increases as X increases. For the second Y/X decreases as X increases whereas for the third population Y/X and X have no correlation. These populations are denoted by 11, 12 and 13 respectively.

The second set consisted of three populations each of size 15 generated by using the model

$$Y_i = \beta X_i + \epsilon_i$$

with $E(\epsilon_i/X_i) = 0$

and $E(\epsilon_i^2/X_i) = AX_i^g$

The three populations denoted by 14, 15 and 16 corresponding to values of g as 0, 1 and 2 respectively are given in the Table 3.

TABLE 3—GENERATED POPULATIONS TAKING g = 0, 1 AND 2

Sl. No.	*X*	*Y* values under population		
		14	*15*	*16*
1	13	25	22.34	13
2	13	26	26.00	26
3	13	27	29.66	39
4	14	27	24.26	14
5	14	28	28.00	28
6	14	29	31.74	42
7	15	29	26.13	15
8	15	30	30.00	30
9	15	31	33.87	45
10	16	31	28.00	16
11	16	32	32.00	32
12	16	33	36.00	48
13	17	33	29.88	17
14	17	34	34.00	34
15	17	35	38.12	51

Variances under the four above mentioned schemes (denoted by V_1, V_2, V_3 and V_4) and that under suggested scheme (V_5) for sample size 4 are given in Table 4 for various populations. In calculation of V_2 and V_3 approximate expressions given by Hartley and Rao [4] and Asok and Sukhatme [1] respectively have been used.

TABLE 4—VARIANCES UNDER DIFFERENT SAMPLING SCHEMES FOR SAMPLE SIZE 4

Population	V_1	V_2	V_3	V_4	V_5
1	1623.33	1272.56	1272.48	—	1369.47
2	816611840.00	793651230.00	793694680.00	—	679504470.00
3	432096.24	340921.96	341209.77	—	307210.85
4	55852.61	48979.33	48952.39	—	61525.92
5	2001224.30	1978711.10	1978921.80	—	1877901.70
6	3558.06	2452.14	2451.56	6681.02	1372.08
7	2113986.00	1778973.00	1778677.00	—	1953003.00
8	815632.62	651985.88	651975.75	—	1021072.40
9	188321.62	143819.94	143659.75	—	212330.16
10	28563.92	25865.52	25864.30	—	26811.54
11	3060.99	2354.60	2381.67	—	1293.59
12	3054.93	2349.95	2376.74	—	1144.76
13	987.97	759.98	687.93	—	371.36
14	89.63	71.70	30.33	31.56	25.59
15	617.66	494.13	452.19	472.51	366.60
16	8497.90	6798.30	6730.67	7055.70	2109.08

Perusal of the table indicated that out of 16 populations considered, the proposed scheme performed better than the existing schemes in 11 populations while in three populations the scheme was superior to the scheme with probability proportional to size with replacement but slightly less efficient than the Sampford's scheme and randomised probability proportional to size systematic sampling scheme.

REFERENCES

[1] Asok, C. and Sukhatme, B. V. (1976) : On Sampford's procedure of unequal probability sampling without replacement I, *Amer. Stat. Assoc.* 71 : 812-18.

[2] Cochran, W. G. (1978) : *Sampling Techniques*. 3rd Ed. John Wiley and Sons, Inc., New York.

[3] Hanurav, T. V. (1967) : Optimum utilization of auxiliary information : πPS sampling of two units from a stratum, *J. Roy. Stat. Soc.*, B-29 : 374-91.

[4] Hartley, H. O. and Rao, J. N. K. (1962) : Sampling with unequal probabilities and without replacement, *Ann. Math. Stat.*, **33** : 350-375.

[5] Horvitz, D. C. and Thompson, D. J. (1952) : A Generalization of sampling without replacement from a finite universe, *J. Amer. Stat. Assoc.*, **47** : 663-85.

[6] Madow, W. G. (1949) : On the theory of systematic sampling II, *Ann. Math. Stat.*, **20** : 333-54.

[7] Sampford, M. R. (1962) : Methods of cluster sampling with and without replacement of unequal sizes, *Biometrika*, **49** : 27-40.

[8] Sampford, M. R. (1967) : On sampling without replacement with unequal probabilities of selection, *Biometrika*, **54**.

[9] Singh, P. and Singh, R. (1978) : A sampling scheme with varying probabilities without replacement, *J. Ind. Soc. Agric. Stat.*, **XXX**(2) : 65-70.

[10] Sukhatme, P. V. and Sukhatme, P. V. (1954) : *Sampling Theory of Surveys with Applications*: Iowa State College Press, Ames, Iowa.