A NEW SAMPLING PROCEDURE WITH VARYING PROBABILITIES

BY

M.N. DESHPANDE

Institute of Science, Nagpur

(Received : May, 1976)

1. INTRODUCTION

Let $U=(U_1, U_2...U_N)$ be a finite population on N identifiable units. A sample is a subset of U and collection of samples is called sample space and is denoted by S. The number of units in sample is called the sample size. In this note we shall confine to samples of fixed size, n. $P=(p(s), s\epsilon S)$ is a set such that $p(s) \geqslant 0$

and
$$\sum_{s \epsilon S} p(s)=1p$$

$P(s)$ is probability associated with s.

$$\text{Let} \quad \pi_i=\sum_{s \supset U_i} p(s) \text{ and } \pi_{ij}=\sum_{s \supset U_i U_j} p(s)$$

denote inclusion probabilities of (U_i) and $(U_i U_j)$ respectively. Y and X are two real valued functions defined on U such that $Y(U_i)=Y$, and $X(U_i)=X_i$. X_i's are called sizes and are known, while Y_i's are unknown. In this note we present a procedure for selecting a sample of size n with varying probabilities and also propose an

unbiased estimator of $\bar{Y}_N=\frac{1}{N}\sum_{i=1}^{N} Y_i,$

belonging to T_3 class proposed by Honvitz and Thompson (1952).

2. NEW SAMPLING PROCEDURE

Let $p=(p_1, p_2,...p_N)$ be a non-negative vector with

$$\sum_{i=1}^{N} p_i=1$$

(i) Select a set of n units from U by simple random sampling without replacement. Let the set be s.

(ii) Perform a binomial trial with probability of success.

$$=\sum_{i\notin s} p_i = 1-\sum_{i\in s} p_i$$

[Here $i \in s \Leftrightarrow U_i \in s$]

(iii) If the trial results in success then accept the set s as sample.

In case of failure, reject the set and perform (i), (ii) and (iii) till a set is accepted as sample.

It is easy to note the following properties.

(a) For sample s, $p(s)$ is proportional to $\sum_{i\notin s} p_i$ and the constant of proportion is $\binom{N-1}{n}^{-1}$.

$$\text{Thus } p(s)=\frac{1}{\binom{N-1}{n}} \sum_{i\notin s} p_i = \frac{1}{\binom{N-1}{n}} \left[1-\sum_{i\in s} p_i\right]$$

$$(b)\ \pi_i=\frac{n}{N-1}(1-p_i) \qquad 1\leq i\leq N \qquad \ldots\text{(I)}$$

$$\pi_{ij}=\frac{n(n-1)}{(N-1)(N-2)}(1-p_i-p_j) \qquad 1\leq i\neq j\leq N \qquad \ldots\text{(II)}$$

(c) From (I) and (II)

$$\pi_i\,\pi_j-\pi_{ij}=\frac{n(N-n-1)}{(N-1)^2(N-2)}(1-p_i-p_j)+\frac{n^2}{(N-1)}p_ip_j$$

Thus for this sampling procedure Yates and Grundy's (1953) estimator of variance of Horvitz and Thompson's estimator is non-negative.

3. Unbiased Estimator of $\bar{Y}_N$ in T_3 class.

Let the size vector $(X_1, X_2, \ldots, X_N)$ be non-negative. If we choose p_i to be proportional to X_i for every i, then the proposed estimator of $\bar{Y}_N$ is

$$\hat{\bar{Y}}_N=\frac{\bar{Y}_n}{\bar{X}_{N-n}}\bar{X}_N$$

where $\bar{Y}_n=\frac{1}{n}\sum_{i\in s} Y_i$, $\bar{X}_N=\frac{1}{N}\sum_{i=1}^{N} X_i$, $\bar{X}_{N-n}=\frac{1}{N-n}\sum_{i\notin s} X_i$

[It may be noted that $\hat{\bar{Y}}_N$ belongs to T_3 class of estimator proposed by Horvitz and Thompson (1952)].

Following results are immediate :

(a) $\hat{\bar{Y}}_N$ is unbiased estimator of $\bar{Y}_N$.

(b) $$\text{Var}(\hat{\bar{Y}}_N)=\frac{\bar{X}_N}{\binom{N}{n}}\sum_{s\in S}\frac{\bar{Y}_n^2}{\bar{X}_{N-n}}-\bar{Y}_N^2$$

(c) Unbiased estimator of Var $(\hat{\bar{Y}}_N)$

$$=\hat{\bar{Y}}_N^2-\frac{\bar{X}_N}{nN\,\bar{X}_{N-n}}\left[\sum_{i\in s}Y_i^2+\frac{N-1}{n-1}\sum_{i\neq j\in s}\sum Y_i\,Y_j\right]$$

4. USE FOR IPPS SAMPLING

A procedure is inclusion probability proportional to size (in short IPPS) sampling procedure if

$$\pi_i=\frac{nX_i}{N\bar{X}_N}\qquad\text{for } i=1, 2,\ldots,N$$

From I we note that for the proposed procedure

$$\underset{i}{Max}\,\pi_i\leq\frac{n}{N-1}$$

Thus the procedure can be used for IPPS sampling if the size vector is non-negative and

$$\underset{i}{Max}\,X_i\leq\frac{N}{N-1}\,\bar{X}_N \qquad \ldots[C_1]$$

However, this is quite a strong condition.

In the next section we present a modified procedure.

5. MODIFIED PROCEDURE

Let $P^*=(p_1^*\ ,p_2^*\ ,\ldots,\ p_i^*\ )$ be such that

(i) $$\sum_{i=1}^{N}p_i^*=1$$

and (ii) sum of any $(N=n)$ p_i^* 's is non-negative.

(A) Select a set of n units from U by simple random sampling without replacement. Let the set be s.

(*B*) Choose a random number R between 0 and M^* where

$$M^* = \underset{s \in S}{\text{Max}} \left[\sum_{i \notin s} p_i^* \right]$$

(*C*) If $R \leq \sum_{i \notin s} p_i$ then accept the set s as sample. Otherwise reject the set and perform (*A*), (*B*) and (*C*) till a set is accepted as sample.

For this modified procedure we get

$$(i) \quad p(s) = \binom{N-1}{n}^{-1} \sum_{i \notin s} p_i^*$$

$$(ii) \quad \pi_i = \frac{n}{N-1}\left(1 - p_i^*\right) \qquad \text{...(III)}$$

$$(iii) \quad \pi_{ij} = \frac{n\,(n-1)}{(N-1)(N-2)}\left(1 - p_i^* - p_j^*\right)$$

From III for each sample s

$$\sum_{i \in s} \pi_i = \frac{n}{N-1}\left(n - \sum_{i \in s} p_i^*\right)$$

$$= \frac{n(n-1)}{N-1} + \frac{n}{N-1} \sum_{i \notin s} p_i^*$$

But for each s, $\sum_{i \notin s} p_i^* \geqslant 0$

Hence $$\sum_{i \in s} \pi_i \geqslant \frac{n(n-1)}{N-1} \text{ for every } s \in S \qquad \text{...(IV)}$$

This leads to,

Theorem. The modified procedure can be applied for IPPS sampling if and only if the sizes satisfy

$$\bar{X}_n = \frac{1}{n} \sum_{i \in s} X_i \geqslant \frac{N(n-1)}{n(N-1)} \bar{X}_N \text{ for every } s \in S. \qquad \text{...}[C_2]$$

In condition $[C_2]$ putting

$$n\bar{X}_n = N\bar{X}_N - (N-n)\,\bar{X}_{N-n}$$

the same can be stated as,

$$\bar{X}_{N-n} \leq \frac{N}{N-1} \bar{X}_N \qquad \text{for every } s \in S \qquad \text{...}[C_2]$$

On comparison we note that the condition $[C_2]$ is weaker than the condition $[C_1]$.

Remark. It may be noted that Sankarnarayanan (1969) also obtained (IV) as condition for applicability of his procedure for IPPS sampling. However it is important to note that when the condition is satisfied, the expressions of $p(s)$ for Sankarnatayanan's scheme and modified procedure are same.

SUMMARY

In this note a new sampling procedure with varying probabilities is presented. Also an unbiased estimator of population mean is proposed. After studying the properties of the procedure a modification is suggested.

ACKNOWLEDGEMENT

Author is grateful to the referee for useful comments.

REFERENCES

[1] Horvitz, D.G. and Thomson, D.J. (1952) : A generalization of sampling without replacement from finite universe. *J.A.S.A.*, 47, pp. 663-685.

[2] Sankarnarayanan, K. (1969) : An IPPS sampling scheme using Lahiri's method of selection. *J.I.S.A.S.*, 21(2), pp. 58-66.

[3] Yates, F. and Grundy, P.M. (1953) : Selection without replacement from within strata with probability proportional to size. *J.R.S.S.*, (13) 15, pp. 253-261.