A NEW METHOD OF SAMPLING WITH UNEQUAL PROBABILITIES

M.N DESHPANDE

Institute of Science, Aurangabad

(Received : September, 1980)

SUMMARY

A new method of drawing a sample containing exactly n units with unequal probabilities is presented in this paper. A few properties of the new method are also presented.

INTRODUCTION

Let $U=(U_1, U_2 \ldots\ldots U_N)$ be a finite population of N units. A sample s is a subset of U containing n distinct units. A probability measure P on collection of all possible samples, S, is called as sampling design and corresponding first and second order inclusion probability are denoted by π_i and π_{ij}' Y_i and X_i are study and auxiliary variables defined on ith unit. Further each X_i is positive.

New Sampling method : Let $\underset{\sim}{A}=(a_1, a_2 \ldots, a_N)$ be a vector of real numbers associated with population $(U_1, U_2 \ldots, U_n)$. Let $a=\Sigma a_i$ and α and β are two real numbers such that

$$\alpha \leqslant \underset{s}{\text{Min}} \left(\sum_{i \in s} a_i\right) \leqslant \underset{s}{\text{Max}} \left(\sum_{i \in s} a_i\right) \leqslant \beta \qquad \ldots \text{I}$$

The proposed new method $M(\underset{\sim}{A}, \alpha, \beta)$ is as follows.

(*i*) Select a sets of n units from U by SRSWOR

(*ii*) Choose a random number R in between α and β

(*iii*) If $R \leqslant \sum_{i \in s} a_i$, then accept s as a sample otherwise discard it and repeat till a set if accepted as sample.

Properties for method $M\ (A, \alpha, \beta)$

(*i*) $p^{(0)} = \left[(\sum_{i \in s} a_i - \alpha) / \left\{ \binom{N}{n} \left(\frac{n}{N} a - \alpha \right) \right\} \right.$

(*ii*) $\pi_j = \left(\frac{n(n-1)}{N(N-1)} (a + \frac{N-n}{n-1} a_i) - \frac{n}{N} \alpha \right/ \left(\frac{n}{N} a - \alpha \right)$

(*iii*) $\pi_{ij} = \frac{n-1}{N-2} (\pi_i + \pi_j) - \frac{n(n-1)}{(N-1)(N-2)}$

(*iv*) for each sample s

$$\sum_{i \in s} \pi_i \geqslant \frac{n(n-1)}{N-1}$$

[Remark : The lower bound in (*iv*) is independent of A, α and β]

(*v*) If we choose $a_i = x_i$ and α, β, satatisfying I then it is easy to see that

$$e = \frac{\bar{y}}{\bar{x} - \alpha/n} \ (\bar{X} - \alpha/n)$$

is an unbiased estimator $\bar{Y}$ (Notations have usual meaning]. Its variance and estimator of variance can be easily obtained.

(*vi*) This is a multi trial sampling method (2) and here expected number of trials $E\ (T)$ is given by

$$(\beta - \alpha) \bigg/ \left[\frac{n}{N} a - \alpha \right] \quad \ldots \quad \ldots \quad \ldots \text{II}$$

Main Result : In sampling with unequal probabilities it is desirable to get $\pi_i \propto X_i$ for each i. As regards to this we have the following ;

Theorem : A necessary and sufficient condition that $M\ (\underline{A}, \alpha, \beta)$ leads to $\pi_i \propto X_i$ for each i is that

$$\underset{s}{\text{Min}} \left[\sum_{i \in s} x_i \right] \geqslant \frac{(n-1)\ N \bar{X}}{N-1} \quad \ldots \text{III}$$

Proof : Necessary part follows from property (*iv*) For sufficiency, choose real numbers *c* and *d* such that $(nc-Nd)>0$ and consider *M* $(\underset{\sim}{A}^*, \alpha^*, \beta^*,)$ with

$$a_i^* = \left[\left(\frac{n}{N}c-d\right) \frac{nx_i}{N} + \frac{n}{N}d - \frac{n(n-1)}{N(N-1)}c \right] \frac{(N-1)\,N}{(N-n)\,n}$$

$$\alpha^*=d,\ \beta^* \geqslant \underset{s}{\text{Max}} \left(\sum_{i\epsilon s} a_i^*\right)$$

If III holds good then it is easy to see that for $M(\underset{\sim}{A}^*, \alpha^*, \beta^*)$

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

(2) $\sum_{i\epsilon s} a_i^* \geqslant \alpha^*$

(3) $\pi_i = \frac{nx_i}{N\bar{x}}$

From the above proof it is noted that if III is satisfied then $(A^*, \alpha^*, \beta^*,)$ can be choosen in many ways via different choices of c and *d*. What can be the best choices of *c*, *d* and β^* ? Since it is multitrial sampling method one would like these constants so that *E* (*T*) is minimized. From property (*vi*) *E* (*T*) is decreasing function of β so one should choose

$$\beta^* = \beta^{**} = \underset{s}{\text{Max}} \left(\sum_{i\epsilon s} a_i^*\right)$$

A little algebra will prove that for $M(\underset{\sim}{A}^*, \alpha^*, \beta^{**})$

$$E(T) = \frac{N(N-1)}{N-n} \left[\frac{x_m}{N\bar{X}} - \frac{n-1}{N-1} \right] \qquad \ldots IV$$

where $X_m = \underset{sfs}{\text{Max}} \left(\sum_{i\epsilon s} X_i\right)$

It is indeed surprising to see that *E* (*T*) in (*IV*) is independent of choice of *c* and *d*.

Remarks : It can be easily noted that the sampling methods proposed by Sankar Narayanan [3] and by Deshpande [1] are particular cases of $M(\underset{\sim}{A}, \alpha, \beta)$

ACKNOWLEDGEMENTS

Author is thankful to Dr. S.G. Prabhu Ajgaonkar, Marathwada University, Aurangabad and the referees for many useful suggestions.

REFERENCES

[1] Deshpande, M.N. (1977). : A new sampling procedure with varying probabilities J.I.S.A.S. 30, 110-114.

[2] Deshpande, M.N. and Ajgaonkar, S.G.P. (1977) : On multi trial sampling methods Biometrika 64, 422-424.

[3] Sankarnarayannan K (1969) : An IPPS sampling scheme using Lahiris method of election J.I.S.A,S., 21. 58-66.