A NEW FIXED SIZE SAMPLING PROCEDURE WITH UNEQUAL PROBABILITIES OF SELECTION

By
M.N. DESHPANDE
Institute of Science, Aurangabad
(Received : March, 1980)

SUMMARY

In this paper a new sampling procedure for drawing a sample of fixed size with unequal probabilities is presented and the important properties are discussed.

$U=(U_1, U_2, \cdots, U_N)$ is a finite population of N units. A subset s of U containing n units is called sample and collection of $\binom{N}{n}$ samples is called sample space and is denoted by S. P is a probability measure on S. π_i and π_{ij} are usual inclusion probabilities of unit U_i and pair of units (U_i, U_j). For each unit, U_i, we have a study variable Y_i and a size variable X_i. Y_i's are unknown and X_i's are known S_N is collection of non negative vectors $\underline{P}=(p_1, p_2, \cdot, p_N)$ such that

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

New Sampling Procedure

Let P be a vector belonging S_N.

(*a*) Select one integer from $(1, 2, \cdots, N)$ according to $\underline{P}$. Let the selected integer be i (*b*) select a sample of size n from

$$(U_1, U_2, \ldots, U_{i-1}, U_{i+1}, \ldots, U_N)$$

by method of simple randon sampling without replacement, (in short SRSWOR).

Hereafter this procedure is referred to as *New Procedure.*

Properties

Following properties of the New procedure can be easily proved.

(*i*) $p(s)$ is proportional to

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

and constant of proportionality is $\binom{N-1}{n}^{-1}$. Thus

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

for every $s \varepsilon S$.

(*ii*) $\pi_i = n(1-p_i)/(N-1)$

$$\pi_{ij} = n(n-1)(1-p_i-p_j)/(N-1)(N-2)$$

(*iii*) It is easy to see that

$$\pi_{ij} = \frac{n-1}{N-2} \left\{ \pi_i + \pi_j - \frac{n}{n-1} \right\}$$

(*iv*) $\pi_i \ \pi_j \ -\pi_{ij}$ is non-negative for each pair (i, j). Hence Yates and Grundy's estimator of variance of Horvitz-Thompson's estimator of population total is always non-negative.

(*v*) Generally it is desired to have π_i's proportional to X_i's. In this connection we have the following.

Theorem : Necessary and sufficient condition to get inclusion probabilities proportional to X_i is that

$$\text{Max } X_i \leqslant X/(N-1) \text{ where } X = \sum_{i=1}^{N} X_i.$$

Proof : Since p_i is nonnegative we get that Max $\pi_i \leqslant n/(N-1)$. If $\pi_i \propto X_i$ then $\pi_i = nX_i/X$. Combining the two we get the result.

Comparison with Midzuno's Procedure

Below we compare the New procedure with Midzuno's procedure (1952).

(*a*) In both the procedure, selection with unequal probabilities is done only once.

(*b*) The number of units selected by SRSWOR in Midzuno's procedure are $n-1$ while these are n in New procedure.

(*c*) The unit selected with unequal probabilities is present in sample, in case of Midzuno's procedure while it is absent in sample in case of New procedure.

(*d*) Property (*iii*) in section 4 is valid for both the procedure.

(*e*) In case of Midzuno's procedure we have

$$\pi_i \geqslant \frac{n-1}{N-1} \text{ for every } i$$

while in case of New procedure

$$\pi_i \leqslant \frac{n}{N-1} \text{ for every } i$$

(*f*) If $p_i \propto X_i$ in case of Midzuno's procedure then the usual ratio estimator is unbiased for population mean. Below we propose one ratio type estimator which is unbiased in case of New procedure if $p_i \propto X_i$. New ratio type estimator

$$\hat{\bar{Y}} = \bar{Y}_s \, \bar{X} / \bar{X}_{s^c}$$

where

$$\bar{Y}_s = \frac{1}{n} \sum_{i \in s} Y_i$$

$$\bar{X}_{s^c} = \frac{1}{N-n} \sum_{i \notin s} \bar{X}_i$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

We note that $N\bar{X} = n\bar{X}_s + (N-n)\,\bar{X}_{s^c}$.

It is easy to show that $\hat{\bar{Y}}$ is unbiased estimator of $\bar{Y}$ if sampling procedure is the New procedure with $p_i \propto X_i$. Variance and estimator of variance in this case can be easily derived as

$$V(\hat{\bar{Y}}) = \frac{1}{N\binom{N}{n}} \sum_{s \in S} \frac{\bar{Y}_s^2}{\bar{X}_{s^c}} \bar{X} - \bar{Y}^2$$

$$\text{Var}\,(\hat{\bar{Y}}) = (\bar{Y}^2) - \frac{\bar{X}}{nN\bar{X}_{s^c}} \left\{ \sum_{i \in s} Y_i^2 + \frac{N-1}{n-1} \sum\sum_{i \neq j \in s} Y_i Y_j \right\}$$

Extension

In this section we propose one extension of the new procedure.

a') Select a set $U_{(k)}$ of k units from U by a procedure-say P^* $(k<N-n)$

b') Select a sample of size n from $U-U_{(k)}$ by SRSWOR.

Let π_i^* be the inclusion probability of U_i in $U_{(k)}$ then for extended procedure we have

$$\pi_i=\frac{n}{N-k}(1-\pi_i^*)$$

Here we consider two special cases of P^*.

Case I. P^* as New procedure ; then

$$\pi_i^*=\frac{k}{N-1}(1-p_i)$$

and

$$\pi_i=\frac{n(N-k-1)}{(N-k)(N-1)}+\frac{nk}{(N-k)(N-1)}p_i$$

Thus in case I we have for every i

$$\pi_i \geqslant n(N-k-1)/(N-k)(N-1)$$

This lower bound is greater than the one in case of Midzuno's procedure.

Case II. P^* as Midzuno's procedure, then

$$\pi_i^*=\frac{k-1}{N-1}+\frac{N-k}{N-1}p_i$$

and

$$\pi_i=\frac{n}{N-1}(1-p_i)$$

It is a surprising result that in case II and New procedure the expressions of π_i are alike.

Next we consider a case III which is combination of Midzuno's and New procedure.

Case III. Select k $(k, <n)$ units by the New procedure and select $(n-k)$ units from remaining $(N-k)$ units by *SRSWOR*. One can easily prove that

$$\pi_i=\frac{n}{N-1}-\frac{n-k}{(N-1)(N-k)}-\frac{(N-n)k}{(N-k)(N-1)}p_i$$

Thus $$\pi_i \leqslant \frac{n}{N-1} - \frac{n-k}{(N-k)(N-1)} \text{ for every } i$$

Here this upper bound is less than that of the New procedure.

ACKNOWLEDGEMENT

Author is grateful to the referee for suggesting valuable changes in the original draft.

REFERENCE

Midzuno, H. (1952) : On the sampling system with probability proportional to sum of sizes, *Ann. Inst. Stat. Math.* **3**, 99-107.