ISSN: 2186-845X ISSN: 2186-8441 Print
Vol. 1. No. 1. April 2012

SELECTION PRCEDURE FOR SAMPLE SIZE TWO USING HORVITZ THOMPSON ESTIMATOR IN CASE OF PROBABILITY PROPORTIONAL TO SIZE SAMPLING

Abdul Qayyum, Muhammad Riaz Ahmad, Wasif Khurshid

Department of Statistics Govt. College University, Lahore, Additional Secretary (Planning), Agriculture Department, Punjab, PAKISTAN.
qayyum.ravian@gmail.com, qayyum_so@hotmail.com,
wasifkhurshid@gmail.com

ABSTRACT

A selection procedure has been developed taking the second unit selection criteria from Durbin (1967a) draw by draw procedure and using a new criteria of selection for the first unit of the sample. A comparative study, using Horvitz – Thompson estimator, has been carried out using eight different selection procedures to see the performance of this new proposed procedure. It shows comparatively better performance for the populations of various natures.

Keywords: Unequal probability sampling, Probability proportional to size sampling, Horvitz – Thompson estimator, Probability of inclusion, Joint probability of inclusion

INTRODUCTION

Durbin (1967a) proposed a draw – by – draw procedure for sample size two as follow,

- Select first unit with probability proportional to size.
- Select second unit with probability proportional to

$$P_j\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]$$

and without replacement.

The probability of inclusion of ith unit π_i is as under,

$$\pi_i = 2\,P_i \tag{1.1}$$

and the joint probability of inclusion of ith & jth units π_{ij} is as under,

$$\pi_{ij} = \frac{4P_iP_j}{K}\left[\frac{1-P_i-P_j}{(1-2P_i)(1-2P_j)}\right] \tag{1.2}$$

where,

$$K = \sum_{i=1}^{N} \frac{P_i}{1-2P_i} + 1$$

THE PROPOSED PROCEDURE

In this section a new procedure (draw – by – draw) has been developed by using a new criterion of selection for the first unit of the sample and taking the same selection criteria for the second unit from the Durbin (1967a) procedure. It is strictly based on sampling without replacement.

Statement of The New Procedure

The statement of the new procedure is stated as follow,

- Select first unit with probability proportional to

$$P_i \big/ (1-P_i)$$

- Select second unit with probability proportional to

$$P_j\left[\frac{1}{1-2P_i} + \frac{1}{1-2P_j}\right]$$

Derivation of π_i and π_{ij}

The probability of inclusion of ith unit π_i is derived as follow,

$$\pi_i = \text{P(ith at 1}^{\text{st}}) + \sum_{j\neq i=1}^{N} \text{P(jth at 1}^{\text{st}}).\ \text{P(ith at 2}^{\text{nd}} / \text{jth at 1}^{\text{st}})$$

We know that,

$$P_i = Z_i \Big/ \sum_{i=1}^{N} Z_i$$

$$\pi_i = \frac{P_i/(1-P_i)}{\sum_{i=1}^{N} P_i/(1-P_i)} + \sum_{j\neq i=1}^{N} \frac{P_j/(1-P_j)}{\sum_{j=1}^{N} P_j/(1-P_j)} \frac{P_i\left[\frac{1}{1-2P_i} + \frac{1}{1-2P_j}\right]}{\sum_{i\neq j=1}^{N} P_i\left[\frac{1}{1-2P_i} + \frac{1}{1-2P_j}\right]}$$

On simplification,

$$\pi_i = \frac{P_i}{B}\left[\frac{1}{1-P_i} + \frac{1}{K}\left[\frac{B}{1-2P_i} - \frac{2P_i}{(1-P_i)(1-2P_i)} + \sum_{j=1}^{N} \frac{P_j}{(1-P_j)(1-2P_j)}\right]\right] \qquad (2.2.1)$$

Where,

$$\sum_{i=1}^{N} \frac{P_i}{1-P_i} = B \qquad \text{(say)}$$

……………………………………………………………………………………………

Leena and Luna International, Oyama, Japan.
Copyright © 2012

ISSN: 2186-845X ISSN: 2186-8441 Print
Vol. 1. No. 1. April 2012

Now for joint probability of inclusion of two units π_{ij}, we proceed as follow,

π_{ij} =P (ith at 1st) P (jth at 2nd / ith at 1st) + P (jth at 1st) P (ith at 2nd / jth at 1st)

$$\pi_{ij} = \frac{P_i/(1-P_i)}{\sum_{i=1}^{N} P_i/(1-P_i)} \frac{P_j\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]}{\sum_{j\neq i=1}^{N} P_j\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]} + \frac{P_j/(1-P_j)}{\sum_{j=1}^{N} P_j/(1-P_j)} \frac{P_i\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]}{\sum_{i\neq j=1}^{N} P_i\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]}$$

On simplification we get,

$$\pi_{ij} = \frac{P_i P_j}{BK}\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]\left[\frac{1}{1-P_i}+\frac{1}{1-P_j}\right] \qquad (2.2.2)$$

Verification of Some Results for the New Selection Procedure:

In this section some of the important results regarding π_i & π_{ij} have been verified which are essential for the validation of a procedure.

Result - 1: The results of π_i and π_{ij} reduce to Simple Random Sampling for

$$P_i = 1/\text{N} = P_j$$

Proof:

Putting $P_i = 1/\text{N} = P_j$ in the equation (2.2.1) & (2.2.2)

K=2(N-1)/(N-2)

$$\text{B} = \sum_{i=1}^{N} \frac{P_i}{1-P_i} = \sum_{i=1}^{N} \frac{1/N}{1-1/N}$$

$$\text{B} = \frac{N}{N-1}$$

$$\pi_i = \frac{N-1}{N^2}\left[\frac{N}{N-1}+\frac{N-2}{2(N-1)}\left[\frac{N^2}{(N-1)(N-2)}-\frac{2N}{(N-1(N-2)}+\frac{N^2}{(N-1)(N-2)}\right]\right]$$

$$\pi_i = \frac{1}{N^2}\left[\text{N}+\frac{2N(N-2)}{2(N-1)(N-2)}[N-1]\right]$$

$$\pi_i = \frac{2}{N}$$

Required result of Simple Random Sampling for n=2

Now, Putting in π_{ij} result,

$$\pi_{ij} = \frac{1/N^2}{(N/N-1)(2(N-1)/(N-2)}\left[\frac{1}{1-2/N}+\frac{1}{1-2/N}\right]\left[\frac{1}{1-1/N}+\frac{1}{1-1/N}\right]$$

$$\pi_{ij} = \frac{4N^2(N-2)}{2N^3(N-1)(N-2)}$$

$$\pi_{ij} = \frac{2}{N(N-1)}$$

Required result of Simple Random Sampling for n=2.

Result - 2:

$$\sum_{i=1}^{N}\pi_i = 2 \qquad \text{for n=2}$$

Proof:

$$\sum_{i=1}^{N}\pi_i = \sum_{i=1}^{N}\frac{P_i}{B}\left[\frac{1}{1-P_i}+\frac{1}{K}\left[\frac{B}{1-2P_i}-\frac{2P_i}{(1-P_i)(1-2P_i)}+\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\right]\right]$$

$$=\frac{1}{B}\sum_{i=1}^{N}\frac{P_i}{1-P_i}+\frac{1}{K}\sum_{i=1}^{N}\frac{P_i}{1-2P_i}-\frac{2}{KB}\sum_{i=1}^{N}\frac{P_i^2}{(1-P_i)(1-2P_i)}+\frac{1}{KB}\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\sum_{i=1}^{N}P_i$$

$$=1+\frac{1}{K}\sum_{i=1}^{N}\frac{P_i}{1-2P_i}+\frac{1}{KB}\sum_{i=1}^{N}\frac{P_i}{(1-P_i)}$$

$$=1+\frac{1}{K}\left(\sum_{i=1}^{N}\frac{P_i}{1-2P_i}+1\right)$$

$$=1+\frac{1}{K}(K)$$

$$\sum_{i=1}^{N}\pi_i = 2 \qquad \text{Required result is achieved}$$

Result - 3:

$$\sum_{j\neq i=1}^{N}\pi_{ij} = (\text{n-1})\ \pi_i$$

$$\sum_{j\neq i=1}^{N}\pi_{ij} = \pi_i \qquad \text{for n=2}$$

Proof:

$$\sum_{j\neq i=1}^{N}\pi_{ij} = \sum_{j\neq i=1}^{N}\frac{P_iP_j}{BK}\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]\left[\frac{1}{1-P_i}+\frac{1}{1-P_j}\right]$$

Leena and Luna International, Oyama, Japan.
Copyright © 2012

ISSN: 2186-845X ISSN: 2186-8441 Print
Vol. 1. No. 1. April 2012

$$=\frac{P_i}{BK}\left[\begin{array}{l}\frac{1}{(1-2P_i)(1-P_i)}\sum_{j\neq i=1}^{N}P_j+\frac{1}{(1-2P_i)}\sum_{j\neq i=1}^{N}\frac{P_j}{1-P_j}+\frac{1}{(1-P_i)}\sum_{j\neq i=1}^{N}\frac{P_j}{1-2P_j}\\+\sum_{j\neq i=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\end{array}\right]$$

$$=\frac{P_i}{BK}\left[\frac{K}{1-P_i}-\frac{1}{1-P_i}+\frac{1}{1-2P_i}+\frac{B}{1-2P_i}-\frac{3P_i}{(1-P_i)(1-2P_i)}+\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\right]$$

$$=\frac{P_i}{B}\left[\frac{1}{1-P_i}+\frac{1}{K}\left[\frac{B}{1-2P_i}-\frac{2P_i}{(1-P_i)(1-2P_i)}+\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\right]\right]$$

$$\sum_{j\neq i=1}^{N}\pi_{ij}=\pi_i \qquad \text{for n=2} \qquad \text{Required result is achieved}$$

……………………………………………………………………………………

Result - 4:

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_{ij}=\text{n(n-1)}$$

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_{ij}=2 \qquad \text{for n=2}$$

Proof:

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_{ij}=\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\frac{P_iP_j}{BK}\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]\left[\frac{1}{1-P_i}+\frac{1}{1-P_j}\right]$$

$$=\frac{1}{BK}\sum_{i=1}^{N}P_i\sum_{j\neq i=1}^{N}P_j\left[\frac{1}{1-2P_i}+\frac{1}{1-2P_j}\right]\left[\frac{1}{1-P_i}+\frac{1}{1-P_j}\right]$$

We have already proved that,

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_{ij}=\frac{1}{BK}\sum_{i=1}^{N}P_i\left[\frac{K}{1-P_i}+\frac{B}{1-2P_i}-\frac{2P_i}{(1-P_i)(1-2P_i)}+\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\right]$$

$$=\frac{1}{BK}\left[K\sum_{i=1}^{N}\frac{P_i}{1-P_i}+B\sum_{i=1}^{N}\frac{P_i}{1-2P_i}-2\sum_{i=1}^{N}\frac{P_i^2}{(1-P_i)(1-2P_i)}+\sum_{i=1}^{N}\frac{P_i}{(1-P_i)(1-2P_i)}\sum_{i=1}^{N}P_i\right]$$

$$=\frac{1}{BK}\left[KB+B(K-1)+\sum_{i=1}^{N}\frac{P_i}{(1-P_i)(1-2P_i)}(1-2P_i)\right]$$

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_{ij}=2 \qquad \text{Required result is achieved}$$

Result - 5:

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_i\pi_j = n^2-\sum_{i=1}^{N}\pi_i^{\,2}$$

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_i\pi_j = 4-\sum_{i=1}^{N}\pi_i^2 \qquad \text{for n=2}$$

Proof:

$$\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\pi_i\pi_j = \sum_{i=1}^{N}\pi_i\sum_{j\neq i=1}^{N}\pi_j$$

Firstly we prove that,

$$\sum_{j\neq i=1}^{N}\pi_j = 2-\pi_i \qquad \text{for n=2}$$

$$\sum_{j\neq i=1}^{N}\pi_j = \sum_{j\neq i=1}^{N}\frac{P_j}{B}\left[\frac{1}{1-P_j}+\frac{1}{K}\left[\frac{B}{1-2P_j}-\frac{2P_j}{(1-P_j)(1-2P_j)}+\sum_{i=1}^{N}\frac{P_i}{(1-P_i)(1-2P_i)}\right]\right]$$

$$= \frac{1}{B}\left(\sum_{j=1}^{N}\frac{P_j}{1-P_j}-\frac{P_i}{1-P_i}\right)+\frac{1}{K}\left(\sum_{j=1}^{N}\frac{P_j}{1-2P_j}-\frac{P_i}{1-2P_i}\right)-\frac{2}{BK}\left(\sum_{j=1}^{N}\frac{P_j^2}{(1-P_j)(1-2P_j)}-\frac{P_i^2}{(1-P_i)(1-2P_i)}\right)+\frac{1}{BK}\left(\sum_{i=1}^{N}\frac{P_i}{(1-P_i)(1-2P_i)}(1-P_i)\right)$$

$$= 2-\frac{1}{B}\left(\frac{P_i}{1-P_i}\right)-\frac{1}{K}\left(\frac{P_i}{1-2P_i}\right)-\frac{2}{BK}\sum_{j=1}^{N}\frac{P_j^2}{(1-P_j)(1-2P_j)}+\frac{2}{BK}\frac{P_i^2}{(1-P_i)(1-2P_i)} +\left[\frac{1}{BK}(K-1)-\frac{1}{K}\right]$$

Taking,

$$\frac{1}{BK}(K-1)-\frac{1}{K}=\frac{1}{BK}(K-1-B)$$

$$=\frac{1}{BK}\left(\sum_{j=1}^{N}\frac{P_j}{1-2P_j}-\sum_{j=1}^{N}\frac{P_j}{1-P_j}\right)$$

$$\frac{1}{BK}(K-1)-\frac{1}{K} = \frac{1}{BK}\left(\sum_{j=1}^{N}\frac{P_j^2}{(1-P_j)(1-2P_j)}\right)$$

By substitution,

$$\sum_{j\neq i=1}^{N}\pi_j = 2-\frac{1}{B}\left(\frac{P_i}{1-P_i}\right)-\frac{1}{K}\left(\frac{P_i}{1-2P_i}\right)-\frac{2}{BK}\sum_{j=1}^{N}\frac{P_j^2}{(1-P_j)(1-2P_j)}+\frac{2}{BK}\frac{P_i^2}{(1-P_i)(1-2P_i)} +\frac{1}{BK}\sum_{j=1}^{N}\frac{P_j^2}{(1-P_j)(1-2P_j)}$$

$$= 2-\frac{P_i}{B}\left[\frac{1}{1-P_i}+\frac{1}{K}\left[\frac{B}{1-2P_i}-\frac{2P_i}{(1-P_i)(1-2P_i)}+\sum_{j=1}^{N}\frac{P_j}{(1-P_j)(1-2P_j)}\right]\right]$$

Leena and Luna International, Oyama, Japan.
Copyright © 2012

ISSN: 2186-845X ISSN: 2186-8441 Print
Vol. 1. No. 1. April 2012

$$\sum_{j \neq i=1}^{N} \pi_j = \quad 2 - \pi_i$$

Now,

$$\sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \pi_i \pi_j = \sum_{i=1}^{N} \pi_i (2 - \pi_i)$$

$$= 4 - \sum_{i=1}^{N} \pi_i^2 \qquad \text{(Using Result-2) Required result is achieved}$$

EMPIRICAL STUDY

In this section an empirical study has been carried out to check the performance of the new procedure. For this purpose variances of eight different selection procedures, using fifty natural populations, have been calculated applying Horvitz – Thompson estimator and then they are ranked in ascending order of variances. The results of average ranks are given below,

Table 3.1. Avg. Ranks of Variances of Horvitz Thompson estimator fifty natural populations

No	Procedures	Average Rank
1	Yates & Grundy (d – b – d)	2.92
2	Brewer / Durbin	3.23
3	This New procedure	4.21
4	Durbin (rejective)	4.30
5	Qayyum et al (2006)	4.45
6	Shahbaz et al (2002)	4.77
7	Midzuno – Sen	5.04
8	Simple Random Sampling	7.06

From the above table we can see that the performance of the New procedure is reasonably well as compare to other procedures under discussion.

Model Building

Now for the estimation of ranks of the newly developed procedure, we use the following model,

$$\mathrm{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Where,

$$\mathrm{Y} = \text{Ranks of the procedure}$$

$$X_1 = \rho_{yz}$$

$$X_2 = \mathrm{C.V(Z)}$$

The fitted model is,

$$\mathrm{Y} = 5.840 - 2.814\, X_1 + 1.062\, X_2$$

β_1 = -2.814 shows that against one unit change in ρ_{yz}, there is a decrement of 2.814 units in the rank of the procedure.

β_2 = 1.062 shows that against one unit change in C.V(Z), there is an increment of 1.062 unit in the rank of the procedure.

Similarly another model using skewness and kurtosis has been developed. The fitted model is,

$$Y = 3.716 + 0.894\, X_1 - 0.291\, X_2$$

Where, Y = Rank of the procedure

X_1 = Coefficient of Kurtosis of 'Z'

X_2 = Coefficient of Skewness of 'Z'

β_1 = 0.894 shows that against one unit change in coefficient of kurtosis, there is an increment of 0.894 unit in the ranking position of the New procedure. That is New performs better (variance decreases & consequently rank decreases) for leptokurtic situations. Hence we may recommend this procedure for leptokurtic measure of size.

β_2 = -0.291 shows that as coefficient of skewness of 'Z' changes one unit, there is a decrement in the rank of the procedure by 0.291 unit. That is for positively skewed measure of size, variance obtained from the procedure New decreases and consequently its rank decreases. Hence we may recommend the procedure New for populations having positively skewed measure of size 'Z'.

Rank Correlations

As we have seen different measures like coefficient of correlation between 'Y' and 'Z', kurtosis of 'Z' and skewness of 'Z' effect the variance of Horvitz – Thompson estimator using a particular selection procedure.

To compare the various selection procedures, we rank the coefficients of variation of different populations obtained under a particular selection procedure as well as also rank the above mentioned measures and find the coefficient of rank correlation to observe the degree of association between theme.

The summary of coefficients of rank correlation for natural populations is given as:

Table 3.2 Coefficients of Rank Correlation

Between	Coefficients of Rank Correlation obtained under This Procedure
ρ_{yz} & C.V	-0.40501
Sk(Z) & C.V	0.3824
Ku(Z) & C.V	0.2652

Copyright © 2012

As the coefficient of rank correlation between ρ_{yz} & C.V is –0.40501, which indicates that there is indirect degree of association between ρ_{yz} & C.V. That is this procedure gives good performance as coefficient of correlation increases.

The other two coefficients of rank correlation are positive. Their degrees are less than the coefficient of rank correlation between ρ_{yz} & C.V. It shows that New procedure is more effected by coefficient of correlation as compare to kurtosis and skewness.

REFERENCES

Beg M.A Hanif (1991) "Comparison of sampling with and without replacement using unequal probability sampling", Bull. Inter. Stat. Inst. 48(4), 118 – 119.

Brewer K. R. W. (1963a) "A model of systematic sampling with unequal probabilities" Aust. J. Stat. 5, 5 – 13.

Brewer K. R. W. and Hanif M (1983) "Sampling with unequal Probabilities" , lecture notes to Statistics , No. 15, Springer - Verlag.

Hanif M. and Brewer K. R. W. (1980) "On Unequal Probability Sampling without replacement; a review", Inter. Stat. Rev. 48(3), 317 – 335.

Hansen M. H. & Hurwitz W. N. (1943) "On the theory of sampling from a finite population", Ann. Math. Stat. 14, 333 – 362.

Horvitz D. G. and Thompson D. J. (1952) "A generalization of sampling without replacement from a finite universe" , J. Amer. Stat. Assoc. 47, 663 – 685.

Kish L. (1965) "Survey Sampling" , John Wiley and Sons Inc. , New York.

Raj D. (1956b) "A note in determination of optimum probabilities in sampling without replacement ", Sankhya, 17, 197 – 200.

Raj D. (1968) "Sampling Theory " , McGraw – Hill, New York.

Rao J. N. K. Hartely H. O. and Cochran W.G. (1962) "On a simple procedure of unequal probability sampling without replacement ", J. Roy. Stat. Soc. B, 24, 482 – 491.

Samiuddin M. and Asad H. (1981) "A simple procedure of unequal probability sampling " , Biometrika, 68 (3), 728 – 731.

Samiuddan M., Hanif M. and Asad H. (1978) "Some remarks on PPS sampling without replacement " , Presented at 4th Australian Statistical Conference, Canberra.

Sampford M. R. (1967) "On sampling without replacement with unequal probabilities of selection", Biometrika, 54, 499 – 513.

Som R. K. (1973) "A Manual of Sampling Techniques" , Heinemann Educational Book Ltd. U. K.