SELECTION PROCEDURE 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

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

and without replacement.

The probability of inclusion of ith unit \( \pi_i \) is as under,

\[ \pi_i = 2P_i \]  (1.1)

and the joint probability of inclusion of ith & jth units \( \pi_{ij} \) is as under,

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

\[ P_i \left( \frac{1}{1 - P_i} \right) \]

- Select second unit with probability proportional to

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

Derivation of \( \pi_i \) and \( \pi_{ij} \)

The probability of inclusion of ith unit \( \pi_i \) is derived as follow,

\[ \pi_i = P(\text{ith at 1st}) + \sum_{j \neq 1}^{N} P(\text{jth at 1st}) \cdot P(\text{ith at 2nd} / \text{jth at 1st}) \]

We know that,

\[ P_i = \frac{Z_i}{\sum_{i=1}^{N} Z_i} \]

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

On simplification,

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

Where,

\[ \sum_{i=1}^{N} \frac{P_i}{1 - P_i} = B \] (say)
Now for joint probability of inclusion of two units \( \pi_{ij} \), we proceed as follow,

\[
\pi_{ij} = P(\text{ith at } 1^{st}) P(\text{jth at 2\textsuperscript{nd} / ith at } 1^{st}) + P(\text{jth at } 1^{st}) P(\text{ith at 2\textsuperscript{nd} / jth at } 1^{st})
\]

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

On simplification we get,

\[
\pi_{ij} = \frac{P_i P_j}{BK} \left[ \frac{1}{1-2P_i} + \frac{1}{1-2P_j} \right] \left[ \frac{1}{1-P_i} + \frac{1}{1-P_j} \right] \tag{2.2.2}
\]

**Verification of Some Results for the New Selection Procedure:**
In this section some of the important results regarding \( \pi_i \) & \( \pi_{ij} \) have been verified which are essential for the validation of a procedure.

**Result - 1:** The results of \( \pi_i \) and \( \pi_{ij} \) reduce to Simple Random Sampling for \( P_i = 1/N = P_j \)

**Proof:**
Putting \( P_i = 1/N = P_j \) in the equation (2.2.1) & (2.2.2)

\[
K = 2(N-1)/(N-2)
\]

\[
B = \sum_{i=1}^{N} \frac{1}{1-P_i} = \sum_{i=1}^{N} \frac{1}{1-1/N}
\]

\[
B = \frac{N}{N-1}
\]

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

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

\[
\pi_i = \frac{2}{N}
\]

Required result of Simple Random Sampling for \( n=2 \)

Now, Putting in \( \pi_{ij} \) result,
\[
\pi_j = \frac{1/N^2}{(N/N - 1)(2(N-1)/(N-2)}\left[\frac{1}{1-2/N} + \frac{1}{1-2/N}\right]\left[\frac{1}{1-1/N} + \frac{1}{1-1/N}\right]
\]

\[
\pi_j = \frac{4N^2(N-2)}{2N^3(N-1)(N-2)}
\]

\[
\pi_j = \frac{2}{N(N-1)}
\]

Required result of Simple Random Sampling for n=2.

Result - 2:

\[
\sum_{i=1}^{N} \pi_i = 2 \quad \text{for } n=2
\]

Proof:

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

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

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

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

\[
= 1 + \frac{1}{K} (K)
\]

\[
\sum_{i=1}^{N} \pi_i = 2 \quad \text{Required result is achieved}
\]

Result - 3:

\[
\sum_{j \neq i}^{N} \pi_j = (n-1) \pi_i
\]

\[
\sum_{j \neq i}^{N} \pi_j = \pi_i \quad \text{for } n=2
\]

Proof:

\[
\sum_{j \neq i}^{N} \pi_j = \sum_{j \neq i}^{N} \frac{P_i P_j}{BK} \left[\frac{1}{1-2P_i} + \frac{1}{1-2P_j}\right]\left[\frac{1}{1-P_i} + \frac{1}{1-P_j}\right]
\]
\[
\frac{N}{i=1} \sum_{j=1}^{N} \pi_{ij} = n(n-1) \\
\frac{N}{i=1} \sum_{j=1}^{N} \pi_{ij} = 2 \quad \text{for } n=2
\]

**Proof:**

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

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

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

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

\[
\frac{N}{i=1} \sum_{j=1}^{N} \pi_{ij} = 2 \quad \text{Required result is achieved}
\]

**Result - 5:**

\[
\frac{N}{i=1} \sum_{j=1}^{N} \pi_{ij} \pi_{ij} = n^2 - \sum_{i=1}^{N} \pi_{i}^2
\]
\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j = 4 - \sum_{i=1}^{N} \pi_i^2 \text{ for } n=2 \]

**Proof:**

\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j = \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \pi_j \]

Firstly we prove that,

\[ \sum_{j \neq i}^{N} \pi_j = 2 - \pi_i \text{ for } n=2 \]

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

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

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,

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

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

Now,
\[
\sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j = \sum_{i=1}^{N} \pi_i (2 - \pi_i) = 4 - \sum_{i=1}^{N} \pi_i^2 \quad \text{(Using Result-2) Required result is achieved}
\]

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

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,
\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2
\]

Where,
\[
Y = \text{Ranks of the procedure}
\]
\[
X_1 = \rho_{yz}
\]
\[
X_2 = C.V(Z)
\]

The fitted model is,
\[
Y = 5.840 - 2.814 X_1 + 1.062 X_2
\]
\( \beta_1 = -2.814 \) shows that against one unit change in \( \rho_{yz} \), there is a decrement of 2.814 units in the rank of the procedure.

\( \beta_2 = 1.062 \) 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,

\[
Y = \text{Rank of the procedure} \\
X_1 = \text{Coefficient of Kurtosis of ‘Z’} \\
X_2 = \text{Coefficient of Skewness of ‘Z’}
\]

\( \beta_1 = 0.894 \) 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.

\( \beta_2 = -0.291 \) 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>
<thead>
<tr>
<th>Between</th>
<th>Coefficients of Rank Correlation obtained under This Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_{yz} ) &amp; C.V</td>
<td>-0.40501</td>
</tr>
<tr>
<td>Sk(Z) &amp; C.V</td>
<td>0.3824</td>
</tr>
<tr>
<td>Ku(Z) &amp; C.V</td>
<td>0.2652</td>
</tr>
</tbody>
</table>
As the coefficient of rank correlation between $\rho_{yz}$ & C.V is –0.40501, which indicates that there is indirect degree of association between $\rho_{yz}$ & 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 $\rho_{yz}$ & C.V. It shows that New procedure is more effected by coefficient of correlation as compare to kurtosis and skewness.

REFERENCES


