Research Journal of Mathematical and Statistical Sciences ______________________________ISSN 2320-6047
Vol. **6(1),** 1-5, January **(2018)** *Res. J. Mathematical and Statistical Sci.*

On splitting of sizes for sampling procedure with inclusion probabilities proportional to size

Vijai Kumar Dwivedi
Department of Statistics, University of Botswana, Botswana
vijai.dwivedi@mopipi.ub.bw

Available online at: www.isca.in, www.isca.me
Received 25th October 2017, revised 3rd January 2018, accepted 10th January 2018

Abstract

The efficiency of Inclusion Probabilities Proportional to Size (πPS) sampling schemes varies from one another due to different sets of joint inclusion probabilities (i.e. π_{ij} 's). Dwivedi suggested an algorithm-I which control on $\pi_{i,i+1}$ (i=1,2,...,N-1) i.e. on diagonal value of π_{ij}. The other values of π_{ij}'s (j>i+1) remains more or less equal. Thus it is desirable to have control on other values of π_{ij}'s as well besides the diagonal ones. This paper provides a modified version of algorithm-I named as algorithm-II for a proper split of sizes X_i's which starts with simultaneous control on π_{ij}'s. The algorithm-II basically contains a maximum of N-n stages, and in the s-th stage, complete splitting of X_{ss} is achieved such that resulting π_{sj}'s satisfies condition $\phi_{ij}<1$ i.e. the non-negativity of variance estimates condition. It is presented that on an average the relative efficiency of proposed algorithm-II demonstrates the supremacy over probability proportional to size with replacement sampling scheme (PPSW).

Keywords: Selection probabilities, unequal probabilities, inclusion probabilities.

Introduction

Utilizing the nature of non-negativity condition of variance estimator ($\phi_{ij}>0$) approach, Dwivedi provided split of sizes with fewer number of attempts resulting a set of π_{ij} 's satisfying the non-negativity condition $\phi_{ij} < 1$ as suggested by Hanurav[1]. Dwivedi[2] suggested an algorithm named as algorithm-I which control on $\pi_{i,i+1}$ (i=1,2,...,N-1) i.e. diagonal value of π_{ij}. The other values of π_{ij}'s (j>i+1) remains more or less equal. It is known that the efficiency of Inclusion Probabilities Proportional to Size (π PS) sampling schemes varies from one another due to different sets of joint inclusion probabilities (i.e. π_{ij} 's). Thus it is desirable to have control on other values of π_{ij}'s as well besides the diagonal ones. The control on π_{ij} may be exercised with the help of corresponding $\pi_i \pi_j$ value. Because of condition $0 < X_i \leq X_{i+1}$ (i=1,2,3,...,N-1) the values of elements of matrix $\Phi_1 =((\pi_i \pi_j))$ increase horizontally from left to right and vertically downwards. Hence if the control on π_{ij}'s are exercised systematically, such that the resulting matrix $((\pi_{ij}))$ also follows more or less the same trend as Φ_1, then the values of

$\emptyset_{ij}{}'s \left[\frac{\pi_{ij}}{\pi_i \pi_j}\right]$ (i=1,2,,…,≤ (N-n); j= i=1,…,N)

are more likely to lie within the reasonable limits. While the remaining values of π_{ij}'s (i ≤ N-n+1, …, N-1; j= i+1,…,N) could be controlled with a little effort by shifting of elements from one to other columns. Thus still a little trial and error is required and this could not be completely eliminated. The proposed algorithm (named as algorithm-II) thus provides control on π_{ij}'s satisfying condition $\phi_{ij} < 1$. The approach of algorithm-II also depends on splitting of X_i's but it basically differs from the approaches of Srivastava and Singh's[3] method and Dwivedi's[2] algorithm-I in the sense that here the splitting of X_i's starts with simultaneous control on π_{ij}'s.

The algorithm-II basically contains a maximum of N-n stages, and in the s-th stage, complete splitting of X_{ss} is achieved such that π_{sj}'s satisfies condition $\phi_{ij} < 1$. Here X_{ss} is defined as follows along with some other terms which are needed in further workout.

U : N-n, the upper limit where the splitting process terminates. X_{si} : The size of i-th unit at s-th stage. Obviously, X_{1i} is the given X_i of i-th unit (s=1,2,3,...,, U); (i=s,s+1, ...,N). X'_{st} : The split value of X_{ss} in the i-th column in s-th stage and in s-th row (t=s,s+1, ...,U(s); t ≤. $\lessgtr$ s; s ≥ i).

Where to satisfy the condition of minimum n units in each split, U(s) = (N-n+2) – s are the maximum number of columns which can accommodate the split of X_{ss}. For the sake of computational ease it is assumed here that X'_{st} takes only integer values. If X'_{st} takes a non-integer values then the integer portion will be retained to satisfy this assumption.

Arrange X_{1i} in ascending order and put them in the column of sizes, such that i-th unit is placed in the i-th row. The steps of algorithm-II are then as follows:

Step-1: Calculate $\pi_i = n X_{1i} X^{-1}$ (i=1,2,...,N), and U(s) = (N-n+2)-s (s=, 1,2,...,U))

Where: $X = \sum_{i=1}^{N} X_i$.

Select a value of R < 1 being the desired level of ϕ_{12}.

Then $\phi_{12} \approx R = m X'_{11}\left[\pi_1 \pi_2 \ (N-1)\right]^{-1}$

i.e. $X'_{11} = \pi_1 \pi_2 \ (N-1) R m^{-1}$

$where,\ m = n(n-1) X^{-1}.$

Since X'_{11} should be at least unity, we redefine,

$$X'_{11} = X'_{11} \quad if\ X'_{11} \geq 1$$
$$=1 \quad if\ X'_{11} < 1.$$

Further, let $X''_{11} = X_{11} - X'_{11}$

Step-2: Define, UU (1) = U(1) – 1, being the total number of columns over which X''_{11} can be distributed over a total number of columns not exceeding X''_{11}, we redefine

UU(1) = UU(1) if UU(1) ≤ X''_{11}

= X''_{11} if UU(1) ≥ X''_{11}

Step-3: Let a(1) = UU(1) + 1

Let $X'_{12} + X'_{13} = ... = X'_{1,UU(1)} = X''_{11} / UU(1)$

$$and,\ X'_{1,a(1)} = X_{11} - X'_{11} - X'_{12} - ... - X'_{1,UU(1)}$$
$$= X_{11} - X'_{11} - \left[UU(1)-1\right] X''_{11} / UU(1)$$

Step-4: $Calculate,\ \pi_{12} = m X'_{11} (N-1)^{-1}$

$$and,\ \phi_{12} = \pi_{12}\left(\pi_1 \pi_2\right)^{-1}$$

Step-5: Let a(2) = UU(1) +2. Calculate for i=3,4,…,a(2)

$$\pi_{1i} = \pi_{1,i-1} + m X'_{1,i-1} (N+1-i)^{-1}$$

$$and,\ \phi_{1i} = \pi_{1i}\left(\pi_1 \pi_i\right)^{-1}$$

If $\pi_{1i} \geq 1$ then proceed to Step-7

Step-6: If n=2, then proceed to Step-8, otherwise define a (3) = UU(1) + 3; Calculate for i=3,4,…,a(3)

$$\pi_{1i} = \pi_{1,a(2)}$$

$$and,\ \phi_{1i} = \pi_{1i}\left(\pi_1 \pi_i\right)^{-1}$$

then proceed to Step-8

Step-7: $Calculate,\ X'_{11} = X'_{11} + 1$

$$X'_{1,i-1} = 1 / X'_{1,i-1},$$

and return to Step-4

Step-8: Up to these stage we have split X_{11} completely and X_{1i}, i>1, partially along with exercising control on π_{12}, π_{13}, ..., π_{1N}. The residual stock at the Ist stage are calculated as follows. For i= 2, 3, ...,a(2)

$$X_{2i} = X_{1i} - \sum_{j=1}^{i-1} X'_{1j}$$

and for n>2 and i= a(3), ...,N

$$X_{2i} = X_{1i} - X_{11}$$

These residual stock will provide other values of π_{ij}'s (i>1, j=i+1,…,N) after suitable shifting. These residual stocks will be the starting stocks for the second stage, where X_{22} will be split completely and X_{2i} (i=3,4,…,N) will be partially exhausted. This splitting is is done in such a way that the resulting values of π_{23}, π_{24},..., π_{2N} satisfy condition 3.7. It is to be noted here that this type of splitting $\pi_{i,i+1}$ will contribute towards the values of $\pi_{i+1,i+2}$.

Step-9: UU(s) is the total number of columns over which X_{ss} can be distributed. Obviously for the second stage, s=2. Since X_{ss} can be distributed over a total number of columns, not exceeding X_{ss}, following Step-2, we redefine,

UU(s) = UU(s) if UU(s) ≤ X_{ss}
UU(s) = X_{ss} if UU(s) > X_{ss}

Step-10: Let a(4)=UU(s) and a(5)= UU(s)-1

Let $X'_{s1} = X'_{s2} = ... = X'_{s,a(5)} = X_{ss} / UU(s)$

$$and,\ X'_{s,a(4)} = X_{ss} - X'_{s1} - X'_{s2} - ... - X'_{s,a(5)}$$
$$= X_{ss} - a(5) X_{ss} / UU(s)$$

Step-11: For the type of split considered here the diagonal elements of $((\pi_{ij}))$ matrix can be obtained using the recurrence relation given below:

$$X_{s,s+1} = \sum_{j=1}^{s-1} \pi_{js} - \sum_{j=1}^{s-2} \pi_{j,s-1} + m X'_{s1} (N-s)^{-1}$$

also,

$$\phi_{s,s+1} = \pi_{s,s+1} (\pi_s, \pi_{s+1})^{-1}$$

Step-12: Let a(6) = UU(s) +s. The off diagonal elements of above matrix $((\pi_{ij}))$ can be obtained as follows:

$$\pi_{si} = \pi_{s,i-1} + m X'_{s,i-s} (N-i+1)^{-1}$$

$$\phi_{si} = \pi_{si} (\pi_s, \pi_i)^{-1} \qquad ;i=s+2,\ldots, a(6)$$

If $\pi_{si} \geq 1$, then proceed to Step -14

Step-13: If n=2, then proceed to Step-15 otherwise define a (7)=UU(s) +s+1 and calculate for i=a(7),.., N

$$\pi_{si} = \pi_{s,a(6)}$$

$$\phi_{si} = \pi_{si} (\pi_s, \pi_i)^{-1}$$ and proceed to Step-15.

Step-14: Calculate,

$$X'_{s1} = X'_{s1} + 1$$

$$X'_{s,i-s} = X'_{s,i-s} - 1$$ and return to Step - 11

Step-15: In this way complete splitting of X_{22} and partial splitting of X_{2i} (i>2) have been achieved along with control on $\pi_{23}, \pi_{24},\ldots, \pi_{2N}$; such that the condition 3.7 is satisfied. Now the residual stocks from the second stage are used as starting stock for 3-rd stage. These are obtained using the following formulae;

$$X_{s+1,i} = X_{si} - \sum_{j=1}^{i-s} X'_{sj} \;; \qquad \text{(i=s+1,…, UU(s) +s+1)}$$

and for n >2,

$$X_{s+1,i} = X_{si} - X_{ss} \qquad \text{(i= = UU(s)+s+2, …, N)}$$

The Steps – 9 to 15 are then repeated. This procedure is continued until all the values of s are exhausted i.e. s=3,4,…, etc.

If at any stage some $X_{s+1,i}$ obtained using the above formulae, becomes negative then the process is terminated after completing the (s-1)-th stage.

Step-16: Thus this process exhausts the first ≤N-n stocks with simultaneous control on π_{ij}'s. Only n'≥n residual stock are left which are to be adjusted in such fashion that condition of non-negativity condition is satisfied. These residual stocks can be adjusted with little effort by shifting some of the elements from one to other columns. After achieving complete shifting (i.e. final splitting) in this manner, sampling scheme Srivastava and Singh[3] applied.

Example of splitting and corresponding set of π_{ij} 's: To illustrate the above suggested algorithm-II, we consider a numerical example consisting of three populations A, B and C given by Yates and Grundy[4] as follows:

Table-1: Three artificial population of size, N = 4.

Population	U_i	U_1	U_2	U_3	U_4
Size	X_i	1	2	3	4
Population A	Y_i	0.5	1.2	2.1	3.2
Population B	Y_i	0.8	1.4	1.8	2.0
Population C	Y_i	0.2	0.6	0.9	0.8

The scale of X_i's has been changed by a constant multiplier 10. After arranging in the ascending order the values are 10, 20, 30 and 40 respectively. Here N=4, n=2, U=2, U(1)=3 and U(2)=2. The process terminates after 2nd stage. The π_1, π_2, π_3 and π_4 values are 0.2, 0.4, 0.6 and 0.8 respectively. The value of R is chosen as 0.5. Following Step-1, $X'_{11} = 7$ and $X''_{11} = 3$ also Step-2 gives UU(1) = 2. Then following Step-3, $X'_{12} = 1$ and $X'_{13} = 2$. Then Steps-4 and 5 provide π_{12}=0.047, π_{13} =0.057 and π_{14}=0.05. Residual stock from the first stage, being the starting stocks for the second stage are then given by X_{22} = 13, X_{23}=22 and X_{24}= 30 (Step-9). Similarly split of X_{22} is obtained as $X'_{21} = 6$ and $X'_{22} = 7$ and the starting stock for 3rd stage are X_{33}=16 and X_{34}= 17 (Steps-9 to 15). Since U(3) = 0, the process terminates. Here X_{31}'s are residual stocks which are adjusted with a little effort in such a manner that condition suggested by Srivastava and Singh[3] is satisfied. The final split achieved is depicted below in Table-2.

Table-2: Final split of sizes with at least 2 nonzero elements in each column.

Population units	Sizes (X_i's)	Columns (Groups)					
		1	2	3	4	5	6
1	10	7	1	2			
2	20	7			5	8	
3	30	7	1		5		17
4	40	7	1	2	5	8	17
Total	100	28	3	4	15	16	34

Using the formulae given by Srivastava and Singh[3] the corresponding Φ matrix for split in Table-2 thus obtained and is given below:

$$\Phi = \begin{bmatrix} 0.58 & 0.47 & 0.60 \\ & 0.40 & 0.80 \\ & & 0.93 \end{bmatrix} \tag{1}$$

Following the initial split obtained using the method of Srivastava and Singh[3] the corresponding Φ thus obtained is given below

$$\Phi = \begin{bmatrix} 0.83 & 0.55 & 0.42 \\ & 0.28 & 0.83 \\ & & 0.97 \end{bmatrix} \tag{2}$$

It is clear from matrix (1) that algorithm-II provides a set of π_{ij}'s for which ϕ_{ij}'s lie within a closer limit than that obtained using the method of Srivastava and Singh[3] as given in matrix (2).

The above numerical example having small value of N and n is chosen to illustrate the various steps involved in algorithm-II. The method Srivastava and Singh[3] provided a split in which the range of ϕ is between 0.28 and 0.97 whereas the proposed algorithm-II provided a split for which these values lie between 0.43 and 0.93 which is more desirable from the efficiency point of view. Thus algorithm-II is clearly preferable.

Relative efficiency

For the population depicted in Table-1, the exact variance of Horvitz Thompson (HT) estimator of the population total Y resulting from algorithm-II is compared with probability proportional to size with replacement (PPS WR) sampling scheme and algorithm-I is presented in Table-3.

Conclusion

On an average the relative efficiency of both the algorithms shows the supremacy over PPSWR. Algorithm-I is more subjective than algorithm-II and also more sensitive to the population characteristics. The algorithm-II demonstrated better for population C (Table-3).

Acknowledgement

Author is thankful and indebted to Prof. A.K. Srivastava, Former Principal Scientist, Indian Agricultural Statistics Research Institute (IASRI), New Delhi for his valued supervision to carry out this research work.

Table-3: Variance and relative efficiency of algorithm-II over PPS WR and algorithm-I.

Population	PPS WR	Algorithm-I			Algorithm-II		
		R- Values			R- Values		
		0.4	0.5	0.6	0.4	0.5	0.6
A	0.500	0.283	0.312	0.333	0.292	0.322	0.338
B	0.500	0.283	0.312	0.333	0.292	0.322	0.338
C	0.125	0.067	0.053	0.047	0.057	0.048	0.042
Average	0.375	0.188	0.225	0.238	0.213	0.231	0.239
Relative Efficiency	100	198	166	157	175	162	156

References

1. Hanurav T.V. (1967). Optimum utilization of Auxiliary Information: π PS sampling of two Units from a Stratum. *J. R. Statist. Soc*., B29, 374-391.
2. Dwivedi V.K. (2016). Algorithm-I on Splitting of Sizes for Sampling Procedure with Inclusion Probabilities Proportional to Size. *Res. J. Mathematical and Statistical Sci*., 4(6), 1-5.
3. Srivastava A.K. and Singh D. (1981). A Sampling Procedure with Inclusion Probabilities Proportional to Size. *Biometrika*, 68(3), 732-734.
4. Yates F and Grundy P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Jour. Roy. Stat. Soc.,* B15, 253-261.