Simulation study to check the performance of various unequal probability sampling estimators

Nadeem Shafique Butt
nadeemshafique@hotmail.com
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore

Muhammad Qasier Shahbaz
qshahbaz@gmail.com
Department of Mathematics
COMSATS Institute of Information Technology, Lahore

ABSTRACT

A large number of unequal probability sampling estimators are available in literature. Theoretical comparison of available unequal probability sampling estimators is a hard task, so survey statisticians have conducted empirical studies to discuss the performances of various estimators. These empirical comparisons were based on limited number of populations available in the literature on survey sampling. The aim of this paper is to perform a simulation study on various popular unequal probability sampling estimators. This simulation study has been conducted by generating populations with given correlation structure from the Bi-Variate Normal Distribution. The study attempt to obtain a minimum variance estimator in unequal probability sampling for population with specific correlation structure.

KEY WORDS : Unequal Probability Sampling, Bi-variate Normal Data, Simulation

1. INTRODUCTION

Unequal probability sampling has been a popular method of sample selection for estimation of population characteristics. The paper deals with comprehensive comparison of some procedures of unequal probability sampling by Simulation. The available procedures are not mathematical comparable because of complex nature of equations and conditions. Lot of empirical comparisons has been done from time to time to obtain a minimum variance estimator or selection procedure for unequal probability sampling. These empirical comparisons have been limited to a specific set of populations and no correlation structures have been specified in these empirical comparison. In this article we have carried out a simulation study to search for an optimum selection procedure that can be used with the Horvitz – Thompson (1952) estimator. This simulation study has been carried out by generating random data from bi-variate normal population with specific correlation structure. The methods that are compared are given in the following section.

2. METHODS USED FOR COMPARISION

In this section the methods used for comparison are given. These methods are selected as they have been widely used in many empirical studies by number of survey statisticians. The methods are listed below:

2.1: Horvitz-Thompson Estimator (1952)

The Horvitz – Thompson (1952) estimator is given as:

Estimate : $$y^{/}_{HT} = \sum_{i \in S} \frac{Y_i}{\pi_i} \tag{2.1}$$

Variance: $$Var\left(y^{/}_{HT}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \tag{2.2}$$

The Horvitz – Thompson (1952) estimator has been used under following selection procedures:

2.2.1: Yates Grundy (1953) draw by draw Procedure:

$$\pi_i = p_i\left[1 + \sum_{j=1}^{N}\frac{p_j}{1-p_j} - \frac{p_i}{1-p_i}\right] \tag{2.3}$$

$$\pi_{ij} = p_i\, p_j\left[\frac{1}{1-p_i} + \frac{1}{1-p_j}\right] \tag{2.4}$$

2.2.2: Brewer (1963) Procedure:

$$\pi_i = 2\, p_i$$

$$\pi_{ij} = \frac{2\, p_i\, p_j}{k}\left[\frac{1}{1-2\, p_i} + \frac{1}{1-2\, p_j}\right] \text{ with } k = 1 + \sum_{j=1}^{N}\frac{p_j}{1-2\, p_j} \tag{2.5}$$

2.2.3: Yates – Grundy Rejective Procedure (1953)

$$\pi_i = \frac{2\, p_i \left(1 - p_i\right)}{1 - \sum_{j=1}^{N} p_j^2} \qquad (2.6)$$

$$\pi_{ij} = \frac{2\, p_i\, p_j}{1 - \sum_{j=1}^{N} p_j^2} \qquad (2.7)$$

2.2.4: Shahbaz and Hanif Procedure (2003)

$$\pi_{\mathrm{i}} = \frac{p_i}{d}\left[\frac{1}{1 - p_i} + \sum_{j=1}^{N} \frac{p_j}{\left(1 - p_j\right)\left(1 - 2\, p_j\right)}\right];\; = \; d = \sum_{i=1}^{N} \frac{p_i}{1 - 2 p_i} \qquad (2.8)$$

$$\pi_{\mathrm{ij}} = \frac{p_i\; p_j}{d}\left[\frac{1}{\left(1 - p_i\right)\left(1 - 2\, p_i\right)} + \frac{1}{\left(1 - p_j\right)\left(1 - 2\, p_j\right)}\right] \qquad (2.9)$$

2.2 Murthy (1957) Estimator:

$$\text{Estimate : } t_{symm} = \frac{1}{2 - p_i - p_j}\left[\frac{y_i}{p_i}\left(1 - p_j\right) + \frac{y_j}{p_j}\left(1 - p_i\right)\right] \qquad (2.10)$$

$$Var\left(t_{symm}\right) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}} \frac{P_i\; P_j\; \left(1 - P_i - P_j\right)}{2 - P_i - P_j} \cdot \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \qquad (2.11)$$

The simulation study is given in the following section.

3. SIMULATION STUDY

In this section the results of simulation study have been given. The study has been carried out by drawing random data from bi-variate normal distribution. The random data has been drawn by using various values of correlation coefficient. After drawing the random data; the variance of Horvitz-Thompson Estimator (1952) has been computed under various selection procedures alongside variance of Murthy estimator. The procedure has been replicated various number of times. After this replication the average variance of each method has been computed for various values of correlation co-efficient. The results of the study have been given in the table 3.1. Following abbreviations have been used in table 3.1

Vyg	Variance of Yates Grundy (1953) draw by draw procedure
Vr	Variance of Yates Grundy (1953) rejective procedure
Vb	Variance of Brower (1963) procedure
Vs	Variance of Shahbaz and Hanif (2003)
Vm	Variance of Murthy (1957) estimator
Raho	Population Correlation Coefficient

Table 3.1: Average Variances of Various Estimators for different number of replicates and different Correlation Structure

		Rho									
Replicate		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10	Vyg	◔ 367814	◔ 314255	◔ 261634	◔ 208861	◕ 203721	◕ 222588	◕ 155648	◕ 146938	◕ 128165	◕ 127014
	Vb	◕ 373242	◕ 316477	◕ 263982	◕ 209528	◕ 203734	◔ 221543	◔ 154544	◔ 145238	◔ 124730	◔ 121554
	Vr	○ 363146	○ 312431	○ 259699	○ 208400	● 203826	● 223634	● 156703	● 148521	● 131255	● 131881
	Vm	◑ 371720	◑ 315654	◑ 262921	◔ 209056	○ 203417	○ 220460	○ 154251	○ 144740	◔ 124219	◔ 121104
	Vs	● 378142	● 318557	● 266149	● 210238	● 203865	○ 220761	○ 153684	○ 143848	○ 121835	○ 116948
25	Vyg	◔ 322869	◔ 257688	◔ 230817	◔ 217205	◕ 221015	◕ 188407	◕ 157279	◕ 158859	◕ 112280	◕ 121020
	Vb	◕ 326943	◕ 260214	◕ 232414	◕ 217912	◔ 219976	◔ 187447	◔ 155622	◔ 156558	◔ 108731	◔ 116633
	Vr	○ 319415	○ 255588	○ 229532	○ 216683	● 222033	● 189344	● 158841	● 160969	● 115476	● 124943
	Vm	◑ 325910	◑ 259669	◑ 231733	◔ 217440	○ 219487	○ 187132	○ 155044	○ 155907	◔ 108379	◔ 115901
	Vs	● 330631	● 262540	● 233922	● 218645	○ 219196	○ 186703	○ 154299	○ 154644	○ 105751	○ 112912
50	Vyg	◔ 341313	◔ 276598	◔ 254699	◔ 257192	◕ 187098	◕ 181670	◕ 158427	◕ 155054	◕ 123453	◕ 105956
	Vb	◕ 346060	◕ 278795	◕ 255863	◕ 258298	◑ 187003	◔ 180741	◔ 156402	◔ 152696	◔ 120007	◔ 101695
	Vr	○ 337277	○ 274806	○ 253799	○ 256336	● 187280	● 182592	● 160295	● 157224	● 126555	● 109763
	Vm	◑ 344594	◑ 278257	◑ 255244	◑ 257662	○ 186648	○ 180232	◔ 155995	◔ 152052	◔ 119548	◔ 101196
	Vs	● 350346	● 280822	● 256992	● 259379	◑ 187024	○ 180041	○ 154747	○ 150742	○ 117105	○ 98089
100	Vyg	◔ 289818	◔ 271476	◔ 257316	◔ 241046	◕ 198177	◕ 172609	◕ 164012	◕ 153776	◕ 130060	◕ 110205
	Vb	◕ 293382	◕ 274201	◕ 259036	◕ 241583	◕ 198224	◔ 171567	◔ 162354	◔ 150788	◔ 126860	◔ 105909
	Vr	○ 286812	○ 269199	○ 255921	○ 240684	● 198240	● 173621	● 165557	● 156489	● 132939	● 114038
	Vm	◑ 292543	◑ 273500	◑ 258304	○ 240934	○ 197638	○ 171117	○ 161914	◔ 150374	◔ 126232	◔ 105375
	Vs	● 296622	● 276695	● 260648	● 242171	● 198373	○ 170759	○ 161005	○ 148297	○ 124180	○ 102277
250	Vyg	◔ 300222	◔ 282311	◔ 258937	◔ 228193	◕ 206228	◕ 184452	◕ 166560	◕ 148726	◕ 128352	◕ 108019
	Vb	◕ 303650	◕ 284474	◕ 260428	◕ 229203	◑ 206066	◔ 183428	◔ 164680	◔ 146109	◔ 124901	◔ 103788
	Vr	○ 297332	○ 280524	○ 257743	○ 227416	● 206472	● 185448	● 168303	● 151109	● 131458	● 111800
	Vm	◑ 302834	◑ 283813	◑ 259765	◑ 228570	○ 205578	○ 182983	◔ 164203	◔ 145622	◔ 124408	◔ 103262
	Vs	● 306763	● 286476	● 261846	● 230198	◑ 206035	○ 182645	○ 163146	○ 143932	○ 121995	○ 100203
500	Vyg	◔ 292662	◔ 274989	◔ 255172	◔ 225084	◕ 215332	◕ 194098	◕ 165507	◕ 150412	◕ 125342	◕ 104594
	Vb	◕ 295813	◕ 277466	◕ 256786	◕ 225681	◑ 215180	◔ 193002	◔ 163766	◔ 147704	◔ 121854	◔ 100526
	Vr	○ 290013	○ 272928	○ 253867	○ 224666	● 215570	● 195159	● 167131	● 152871	● 128482	● 108229
	Vm	◑ 295015	◑ 276743	◑ 256091	◔ 225111	○ 214647	○ 192501	○ 163251	◔ 147249	◔ 121376	◔ 100004
	Vs	● 298685	● 279742	● 258313	● 226312	◑ 215158	○ 192156	○ 162348	○ 145446	○ 118919	○ 97079
1000	Vyg	◔ 305239	◔ 276436	◔ 253448	◔ 230203	◕ 210368	◕ 187332	◕ 164127	◕ 145648	◕ 125993	◕ 102516
	Vb	◕ 308630	◕ 278862	◕ 254996	◕ 230899	◕ 210241	◔ 186398	◔ 162338	◔ 143169	◔ 122559	◔ 98435
	Vr	○ 302379	○ 274421	○ 252205	○ 229700	● 210587	● 188253	● 165790	● 147909	● 129083	● 106162
	Vm	◑ 307752	◑ 278101	◑ 254375	◔ 230312	○ 209661	○ 185929	◔ 161869	◔ 142693	◔ 122072	◔ 97955
	Vs	● 311712	● 281097	● 256463	● 231618	◕ 210241	○ 185691	○ 160884	○ 141109	○ 119671	○ 94978
2500	Vyg	◔ 297927	◔ 277918	◔ 255518	◔ 235577	◕ 210328	◕ 188142	◕ 166078	◕ 145403	◕ 125413	◕ 104538
	Vb	◕ 301263	◕ 280361	◕ 257121	◕ 236327	◕ 210231	◔ 187211	◔ 164328	◔ 142804	◔ 122030	◔ 100419
	Vr	○ 295116	○ 275888	○ 254226	○ 235032	● 210520	● 189059	● 167708	● 147772	● 128460	● 108217
	Vm	◑ 300398	◑ 279597	◑ 256473	◔ 235726	○ 209683	○ 186718	○ 163853	◔ 142317	◔ 121536	◔ 99924
	Vs	● 304298	● 282613	● 258636	● 237094	◕ 210256	○ 186507	○ 162907	○ 140640	○ 119183	○ 96930
5000	Vyg	◔ 304084	◔ 276182	◔ 257168	◔ 229185	◕ 208605	◕ 188005	◕ 167797	◕ 144873	◕ 125584	◕ 105463
	Vb	◕ 307459	◕ 278567	◕ 258784	◕ 229870	◑ 208435	◔ 187049	◔ 166012	◔ 142305	◔ 122196	◔ 101291
	Vr	○ 301237	○ 274203	○ 255865	○ 228694	● 208859	● 188944	● 169458	● 147211	● 128633	● 109190
	Vm	◑ 306574	◑ 277840	◑ 258116	◔ 229279	○ 207890	○ 186544	○ 165516	◔ 141831	◔ 121713	◔ 100794
	Vs	● 310529	● 280766	● 260309	● 230579	◑ 208397	○ 186323	○ 164561	○ 140169	○ 119345	○ 97756
10000	Vyg	◔ 299232	◔ 275627	◔ 254391	◔ 234400	◕ 208839	◕ 187887	◕ 167005	◕ 145727	◕ 125005	◕ 104743
	Vb	◕ 302605	◕ 278038	◕ 255938	◕ 235090	◑ 208680	◔ 186939	◔ 165229	◔ 143134	◔ 121637	◔ 100623
	Vr	○ 296388	○ 273625	○ 253146	○ 233905	● 209085	● 188821	● 168658	● 148090	● 128037	● 108425
	Vm	◑ 301757	◑ 277287	◑ 255269	◔ 234473	○ 208128	○ 186428	○ 164747	◔ 142662	◔ 121155	◔ 100118
	Vs	● 305671	● 280261	● 257405	● 235804	◑ 208651	○ 186220	○ 163786	○ 140977	○ 118803	○ 97132
25000	Vyg	◔ 299785	◔ 277551	◔ 253927	◔ 231460	◕ 209898	◕ 187516	◕ 165986	◕ 145536	◕ 124950	◕ 106295
	Vb	◕ 303119	◕ 280010	◕ 255512	◕ 232170	◑ 209748	◔ 186512	◔ 164196	◔ 142944	◔ 121572	◔ 102125
	Vr	○ 296975	○ 275509	○ 252649	○ 230949	● 210136	● 188499	● 167651	● 147897	● 127991	● 110021
	Vm	◑ 302269	◑ 279261	◑ 254841	◔ 231566	○ 209204	○ 186012	○ 163717	◔ 142470	◔ 121092	◔ 101607
	Vs	● 306150	● 282274	● 257011	● 232902	◑ 209728	○ 185744	○ 162740	○ 140788	○ 118730	○ 98593
50000	Vyg	◔ 300857	◔ 277028	◔ 254536	◔ 232143	◕ 209521	◕ 187660	◕ 166735	◕ 145375	◕ 125122	◕ 105320
	Vb	◕ 304207	◕ 279462	◕ 256116	◕ 232853	◑ 209379	◔ 186676	◔ 164940	◔ 142779	◔ 121739	◔ 101168
	Vr	○ 298033	○ 275008	○ 253262	○ 231631	● 209752	● 188625	● 168405	● 147739	● 128168	● 109029
	Vm	◑ 303354	◑ 278714	◑ 255451	◔ 232253	○ 208833	○ 186173	○ 164456	◔ 142306	◔ 121258	◔ 100662
	Vs	● 307254	● 281704	● 257610	● 233586	◑ 209365	○ 185925	○ 163480	○ 140620	○ 118893	○ 97651

* More shaded circle represent the larger variance

4. CONCLUSIONS

The results of the simulation study have been given in section 3 of the article. The table 3.1 contains the average variance of various selection procedures under different correlation structure. The table has been constructed by using various numbers of replicates from the bivariate normal distribution. From the table we can see that Yates–Grundy (1953) rejective procedure outperform other selection procedures involved in the study for populations having correlation between 0.1 to 0.5. The procedure given by Shahbaz–Hanif (2003) outperform other procedures involved in the study for populations having correlation of 0.6 to 1.0, and this procedure is closely followed by the Murthy (1957) estimator. The other procedures do not perform batter.

From the table 3.1; we can, therefore, conclude that the Yates–Grundy (1953) rejective procedure is a batter choice for estimation of population total by using the Horvitz–Thompson (1952) estimator for populations having a correlation coefficient of below 0.5. The Shahbaz – Hanif (2003) procedure is better for populations having high correlation; that is correlation of greater than or equal to 0.6.

REFERENCE

1. Brewer, K. R. W. (1963) "A model of systematic sampling with unequal probabilities", *Aust. J. Stat.* 5, 5 – 13.
2. 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.
3. Murthy, M. N. (1957) "Ordered and unordered estimators in sampling without replacement", *Sankhya*, 18, 379 – 390.
4. Shahbaz, M. Q. and Hanif, M. (2003) "A simple procedure for unequl probability sampling without replacement and a sample of size 2", *Pak. J. Stat.* 19(1), 151- 160
5. Yates, F. and Grundy, P. M. (1953) "Selection without replacement from within strata with probability proportional to size", *J. Roy. Stat. Soc.*, B, 15, 153 – 161.