JACOBS PUBLISHERS

OPEN ACCESS

Jacobs Journal of Biostatistics

Review article

Entropy of Selection Procedures for Unequal Probability Sampling

Anam Riaz[1], Abdul Basit[1*], Zafar Iqbal[1] and Munir Ahmad[1]

[1]*National College of Business Administration and Economics, Lahore, Pakistan*

*Corresponding author: *Abdul Basit, National College of Business Administration and Economics, Lahore, Pakistan,*

Email: basit_ravian917@hotmail.com

Received: *08-03-2016*

Accepted: *08-23-2016*

Published: *08-30-2016*

Abstract

Entropy measure has been used to compare the different selection procedures for unequal probability sampling. Basit and Shahbaz [1,2] derived the general class of selection procedures for sample size two and sample size 'n'. To compare the selection procedures, Shannon entropy has been used for these selection procedures. This study claims that selection procedure with a higher entropy will produce the smaller variance of Horvitz - Thompson [3] estimator and as well as variance of Murthy [4] estimator.

Keywords: Unequal Probability Sampling; Horvitz – Thompson Estimator; Murthy Estimator; Shannon Entropy

Introduction

Shannon [5] introduced the concept of information theory and gave the idea of information function & entropy measure. Information function is based on the logarithm of probability of an event. Entropy measure is the average of information function. In the literature different generalized entropy measures are available for the engineering sciences and reliability theory. Takahashi [6] addressed two main points regarding weighted probability of the events. The 1st main point is; probability of an event is non-linearly transformed into weighted probability which has concave and convex points. The 2nd main point he addressed; unknown probability distribution of the event.

Hansen and Hurwitz [7] firstly introduced the idea of unequal probability sampling early in forties. They gave the idea of unequal probability sampling with replacement. The theoretical framework of unequal probability sampling without replacement was introduced in early fifties. The estimator of population total proposed by Horvitz and Thompson [3] is:

$$y'_{HT} = \sum_{i\in s} \frac{Y_i}{\pi_i} , \qquad (1.1)$$

Where π_i is probability of inclusion of i-th unit in the sample.

Variance of this estimator derived the Sen [8] and Yates and Grundy [9] is given as:

$$V(y'_{HT}) = \sum_{j>i}^{N}\sum \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \qquad (1.2)$$

Where π_{ij} is joint probability of inclusion of i-th and j-th unit in the sample.

For the comprehensive review of (1.1) can be found in Brewer and Hanif [10].

The estimator of population total proposed by Murthy [4] is given as:

$$t_{symm} = \frac{1}{P(S)} \sum_{i=1}^{n} P(S/i)\, y_i \qquad (1.3)$$

The Murthy [4] estimator for a sample size two, under Yates – Grundy [9] procedure is:

$$t_{symm} = \frac{1}{2-p_i-p_j}\left[\frac{y_i}{p_i}\left(1-p_j\right) + \frac{y_j}{p_j}\left(1-p_i\right)\right] \qquad (1.4)$$

With

$$Var\left(t_{symm}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq i}} \frac{P_i P_j \left(1-P_i-P_j\right)}{2-P_i-P_j} \cdot \left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2 \qquad (1.5)$$

Selection Procedures

Basit and Shahbaz [1], derived a general class of selection procedure for unequal probability sampling for sample size 2. The probability of inclusion for the i-th unit in the sample for this selection procedure is given as:

$$\pi_i = \frac{1}{d}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)\left(1-2p_i\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + p_i \sum_{j=1}^{N} \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)\left(1-p_j\right)} \right] \quad (2.1)$$

Where

$$d = \sum_{i=1}^{N} \frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)}$$

The joint probability of inclusion for i-th and j-th units in the sample for this selection procedure is given as:

$$\pi_{ij} = \frac{1}{d}\left[\frac{p_i^{\alpha} p_j \left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha} p_i \left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)} \right] \quad (2.2)$$

Basit and Shahbaz [2], also derived a general selection procedure for unequal probability sampling for sample size ‘n’. The probability of inclusion for the i-th unit and joint probability i-th and j-th units for this selection procedure is given as:

$$\pi_i = \frac{1}{d}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)} \left\{ \frac{N-n}{N-1} \right\} + \frac{(n-1)}{N-1} d \right] \quad (2.3)$$

$$\pi_{ij} = \frac{(n-1)(N-n)}{d(N-1)(N-2)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{1-2p_i^{\beta}} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{1-2p_j^{\beta}} \right] + \frac{(n-1)(n-2)}{(N-1)(N-2)} \quad (2.4)$$

They compared the variances of Horvitz & Thompson [3] estimator using first selection procedure and compared the variances of Murthy [4] estimator using 2nd selection procedure. In both procedures, we have different selection procedures for the different values of α and β. In these procedures each pair of α and β provides a new selection procedure. Al-Jjararha [11] also derived a class of selection procedure for sampling two units with probability proportional to size. He also compared the variance of Horvitz – Thompson Estimator.

Entropy of Selection Procedure

Entropy is a measure to check the spread or randomness of the sampling design. In simple words entropy is the average of the amount of information. Entropy has different definitions e.g. entropy as a measure of diversity, degree of randomness and measure of the amount of disorder in a system.

$$H = -\sum_{s \in S} P(S) \ln P(S) \quad (3.1)$$

Where P(S) is the joint probability of inclusion. i.e. π_{ij}

Shannon entropy for Basit and Shahbaz [1] selection procedure is:

$$H_{2006} = -\sum_{s \in S} \frac{1}{d}\left(\frac{p_i^{\alpha} p_j \left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha} p_i \left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)} \right) \ln\left(\frac{1}{d}\left(\frac{p_i^{\alpha} p_j \left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha} p_i \left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)} \right) \right)$$

$$H_{2006} = \ln(d)\sum_{s \in S} \pi_{ij} - \sum_{s \in S} \pi_{ij} \ln\left(\frac{p_i^{\alpha} p_j \left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha} p_i \left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)} \right)$$

$$H_{2006} = \ln(d) - \sum_{s \in S} \ln\left(\frac{p_i^{\alpha} p_j \left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha} p_i \left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)} \right)^{\pi_{ij}} \quad (3.2)$$

Shannon entropy for Basit and Shahbaz [2] selection procedure using n=2 is:

$$H_{2007} = -\sum_{s \in S} \frac{1}{d}\left(\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)} \right) \ln\left(\frac{1}{d}\left(\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)} \right) \right)$$

$$H_{2007} = \ln(d) - \sum_{s \in S} \ln\left(\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)} + \frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)} \right)^{\pi_{ij}} \quad (3.3)$$

Empirical Study

In this section variance of Horvitz –Thompson estimator and Murthy estimator for both procedures has been calculated. Entropy of the each selection procedure has been calculated for different values of α and β. We used the values of α and β which were mentioned in Basit and Shahbaz [1]. For the empirical study, an artificial population has been used. For the comparison of the variance of Horvitz-Thompson estimator, we assigned the rank of lowest variance is 1 and so on, similarly assigned rank 1 for the minimum variance of Murthy estimator.

	β									
α	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
-5	96	97	98	99	100	91	92	93	94	95
-4	86	87	88	89	90	81	82	83	84	85
-3	76	77	78	79	80	71	72	73	74	75
-2	66	67	68	69	70	61	62	63	64	65
-1	**56**	**57**	**58**	**59**	**60**	**50**	**52**	**53**	**54**	**55**
1	**5**	**4**	**3**	**2**	**1**	**19**	**10**	**8**	**7**	**6**
2	**14**	**13**	**12**	**11**	**9**	**28**	**18**	**17**	**16**	**15**
3	24	23	22	21	20	36	29	27	26	25
4	34	33	32	31	30	43	39	38	37	35
5	45	44	42	41	40	51	49	48	47	46

Table 1. Rank of Variance of Horvitz – Thompson Estimator (Basit and Shahbaz [1]).

	β									
α	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
-5	1.4748	1.4747	1.4745	1.4732	1.4677	1.4952	1.4782	1.4755	1.4750	1.4748
-4	1.6064	1.6062	1.6057	1.6033	1.5932	1.6447	1.6129	1.6078	1.6067	1.6065
-3	1.8255	1.8253	1.8245	1.8207	1.8050	1.8848	1.8359	1.8278	1.8261	1.8257
-2	2.0978	2.0976	2.0966	2.0928	2.0765	2.1545	2.1087	2.1004	2.0985	2.0980
-1	**2.2734**	**2.2733**	**2.2731**	**2.2721**	**2.2669**	**2.2768**	**2.2755**	**2.2738**	**2.2735**	**2.2734**
1	**2.1092**	**2.1098**	**2.1114**	**2.1158**	**2.1289**	**2.0219**	**2.0924**	**2.1037**	**2.1071**	**2.1083**
2	**1.9683**	**1.9691**	**1.9712**	**1.9767**	**1.9916**	**1.8652**	**1.9469**	**1.9609**	**1.9654**	**1.9670**
3	1.8326	1.8335	1.8360	1.8422	1.8585	1.7254	1.8087	1.8240	1.8292	1.8311
4	1.7035	1.7044	1.7067	1.7128	1.7286	1.6072	1.6807	1.6952	1.7001	1.7019
5	1.5918	1.5925	1.5945	1.5996	1.6129	1.5155	1.5730	1.5849	1.5890	1.5905

Table 2. Shannon Entropy (Basit and Shahbaz [1]).

	β									
α	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
-5	96	97	98	99	100	91	92	93	94	95
-4	85	86	87	89	90	80	81	82	83	84
-3	75	76	77	78	79	69	71	72	73	74
-2	54	55	56	57	58	44	49	51	52	53
-1	**24**	**25**	**28**	**30**	**34**	**20**	**22**	**23**	**26**	**27**
1	**5**	**4**	**3**	**2**	**1**	**18**	**9**	**8**	**7**	**6**
2	**14**	**13**	**12**	**11**	**10**	**33**	**19**	**17**	**16**	**15**
3	35	32	31	29	21	42	39	38	37	36
4	46	45	43	41	40	61	50	48	47	65
5	64	63	62	60	59	88	70	68	67	65

Table 3. Rank of Variance of Horvitz – Thompson Estimator (Basit and Shahbaz [2]).

	β									
α	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
-5	1.4970	1.4970	1.4968	1.4957	1.4911	1.5144	1.5000	1.4977	1.4972	1.4971
-4	1.6084	1.6083	1.6078	1.6058	1.5972	1.6416	1.6140	1.6096	1.6087	1.6085
-3	1.7973	1.7971	1.7963	1.7930	1.7792	1.8512	1.8067	1.7994	1.7979	1.7975
-2	2.0445	2.0442	2.0432	2.0393	2.0235	2.1059	2.0560	2.0474	2.0453	2.0448
-1	**2.2421**	**2.2419**	**2.2412**	**2.2387**	**2.2291**	**2.2755**	**2.2496**	**2.2443**	**2.2428**	**2.2424**
1	**2.2648**	**2.2651**	**2.2660**	**2.2685**	**2.2748**	**2.2007**	**2.2543**	**2.2613**	**2.2634**	**2.2642**
2	**2.1690**	**2.1698**	**2.1719**	**2.1772**	**2.1907**	**2.0598**	**2.1475**	**2.1615**	**2.1661**	**2.1677**
3	2.0305	2.0316	2.0345	2.0418	2.0605	1.9007	2.0021	2.0203	2.0264	2.0286
4	1.8761	1.8772	1.8802	1.8877	1.9075	1.7537	1.8473	1.8656	1.8718	1.8741
5	1.7346	1.7356	1.7381	1.7447	1.7619	1.6356	1.7104	1.7257	1.7310	1.7329

Table 4. Shannon Entropy (Basit and Shahbaz [2]).

	β									
α	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
-5	96	97	98	99	100	91	92	93	94	95
-4	86	87	88	89	90	81	82	83	84	85
-3	75	76	77	78	79	69	71	72	73	74
-2	46	47	48	50	55	40	41	43	44	45
-1	**25**	**26**	**27**	**28**	**29**	**11**	**20**	**22**	**23**	**24**
1	**5**	**4**	**3**	**2**	**1**	**12**	**9**	**8**	**7**	**6**
2	**16**	**15**	**14**	**13**	**10**	**31**	**21**	**19**	**18**	**17**
3	35	34	33	32	30	49	39	38	37	36
4	54	53	52	51	42	61	59	58	57	56
5	65	64	63	62	60	80	70	68	67	66

Table 5. Rank of Variance of Murthy Estimator (Basit and Shahbaz [2]).

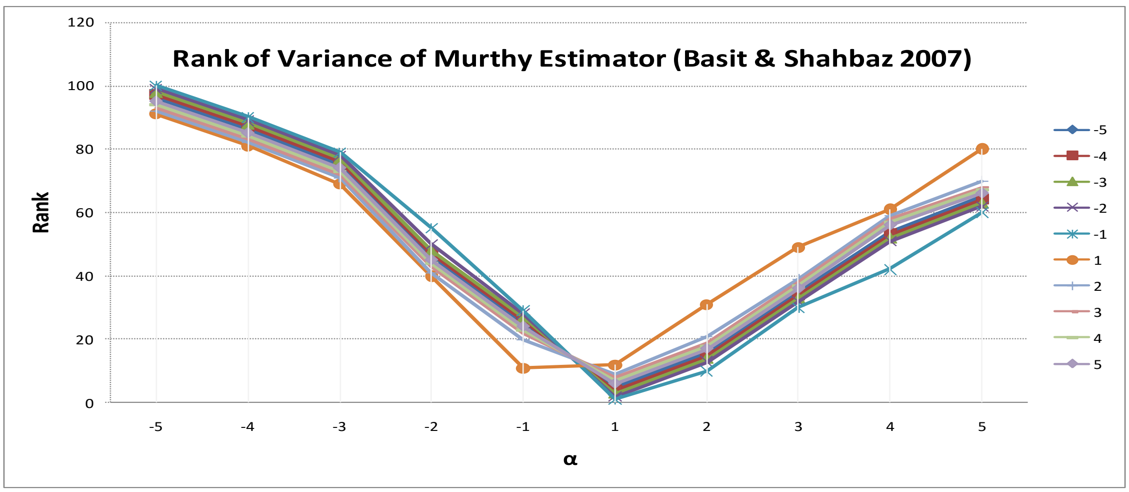

Figure 1. Basit and Shahbaz [2]

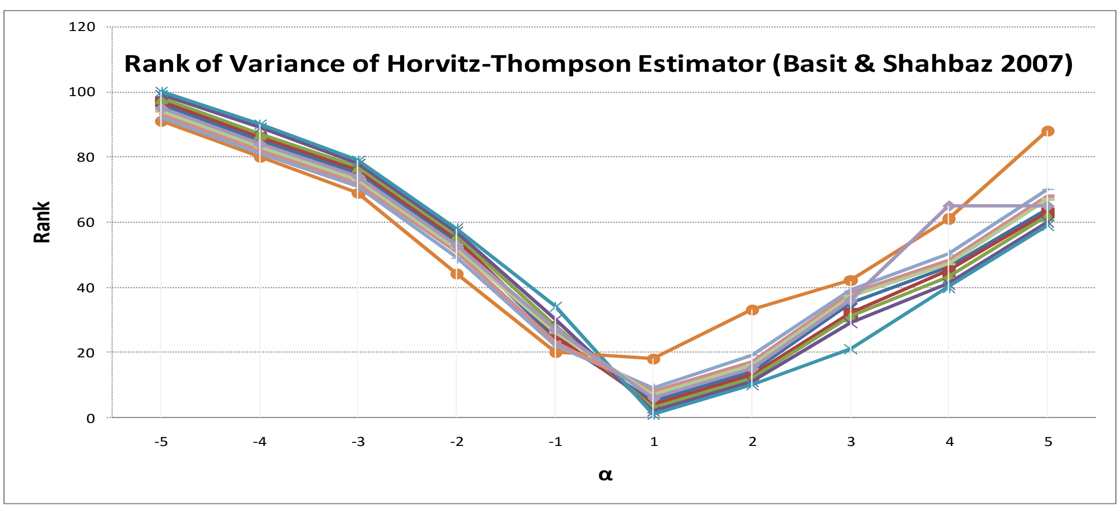

Figure 2. Basit and Shahbaz [2]

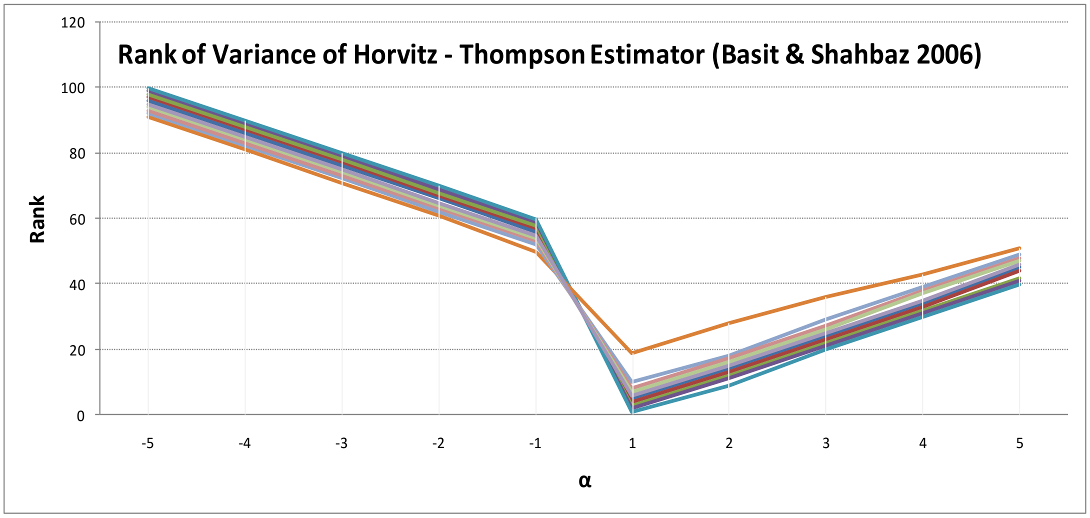

Figure 3. Basit and Shahbaz [1]

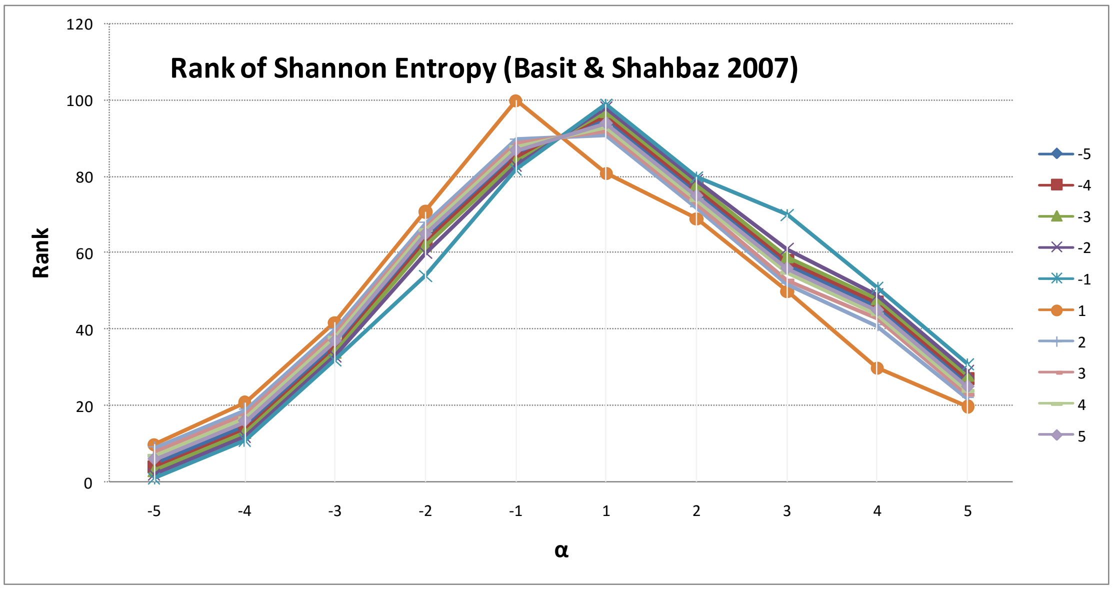

Figure 4. Basit and Shahbaz [2]

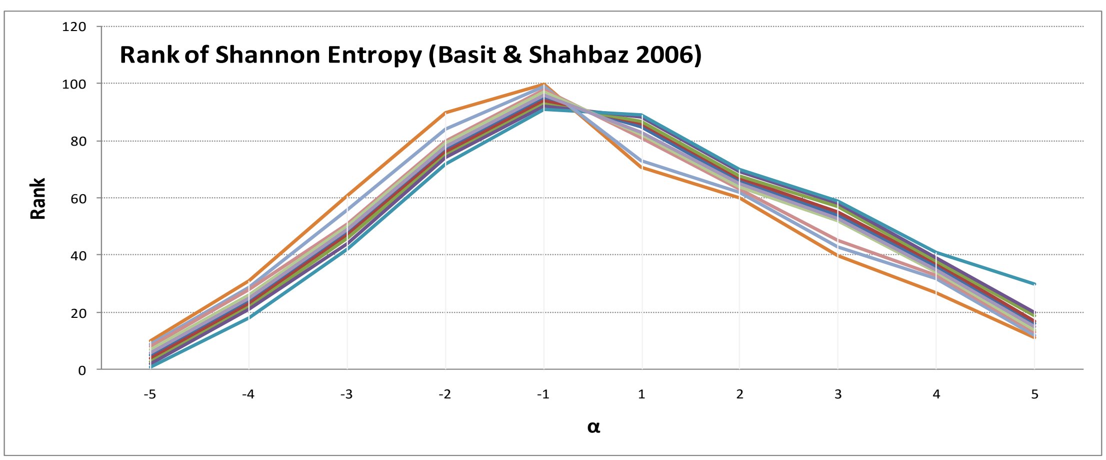

Figure 5. Basit and Shahbaz [1]

Figure 1-3 express the rank of variance of Horvitz – Thompson estimator and Murthy estimators for both selection procedures. Figure 4–5 shows the trend of entropy for both selection procedures. Rank of Shannon entropy for both procedures are high and variance of both estimators are smaller for α = -1, 1 and 2.

Conclusion

From the empirical study it is concluded that both selection procedures has the higher entropy for the values of α = -1, 1, 2 and any value of β. Variance of Horvitz – Thompson estimator and Murthy estimator for both procedures has minimum rank for the values of α = 1, 2 and any value of β. We found that entropy of selection procedures and the variance of estimators has the inverse relationship.

Acknowledgments

The authors' special thanks to editor and reviewer for their encouraging comments which led to improve the quality of the article.

References:

1. Basit A, Shahbaz QM. A class of selection procedure for unequal probability sampling without replacement and a Sample of Size 2. Journal of Statistics. 2006, 13(1): 26–32.

2. Basit A, Shahbaz QM. A general class of selection procedures and modified Murthy estimator. Journal of Statistics. 2007, 14: 13-19.

3. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Amer Stat Assoc. 1952, 47(260): 663–685.

4. Murthy MN. Ordered and unordered estimators in sampling without replacement. Sankhya. 1957, 18(3-4): 379–390.

5. Shannon CE. A mathematical theory of communication. Bell System Tech Journal. 1948, 27:379-423.

6. Takahashi T. Tsallis' non-extensive free energy as a subjective value of an uncertain reward. Physica A: Statistical Mechanics and Its Applications. 2009, 388(5): 715-719.

7. Hansen MH, Hurwitz WN. On the theory of sampling from a finite population. Ann Math Stat. 1943, 14(4): 333–362.

8. Sen AR. On the estimate of the variance in sampling with varying probabilities. J Ind Soc Agri Stat. 1953, 5:119–127.

9. Yates F, Grundy PM. Selection without replacement from within strata with probability proportional to size. J Roy Stat Soc. Series B. 1953, 15(2): 253–261.

10. Brewer KRW, Hanif M. Sampling with unequal probabilities. Lecture notes in Statistics series. Springer Verlag, New York. 1983, 15(1): 164.

11. Al-Jararha J. A Class of Sampling Two Units with Probability Proportional to Size. Communications in Statistics-Simulation and Computation. 2013, 42(8): 1906-1916.