Pak. J. Statist.
2016 Vol. 32(6), 409-423

A HIGH ENTROPY PROCEDURE FOR UNEQUAL PROBABILITY SAMPLING

Aftab Ahmad and **Muhammad Hanif**
National College of Business Administration and Economics, Lahore, Pakistan
Email: aftab.stat@gmail.com, drmainhanif@gmail.com

ABSTRACT

In survey sampling we select a random sample according to some specified random fashion. We focused to apply innovative approach of maximum entropy sampling to develop a new easily executable procedure to determine the probability function in unequal probability sampling and it needs no iteration to compute inclusion probabilities of any order. The empirical comparison of this procedure shows that Horvitz & Thompson (1952) population total estimate has high entropy and lower variance than that of the Yates and Grundy (1953), Brewer (1963) and Prabhu and Ajgankar (1982) selection procedures.

KEY WORDS

Probability proportional to size sampling, Entropy, Horvitz & Thompson Population estimate, First & second (or joint) order probabilities of inclusion.

1. INTRODUCTION

In unequal probability sampling without replacement consider a finite population comprising of N elements or units. For each such element k, where $k = 1, 2, 3, \ldots, N$, two variables Y and Z are attached such that the values of the variable Y, called as benchmark or auxiliary variable, is known for all the values of k from 1 to N. The variable of main interest is denoted by Z and we want to estimate the population total $Z = \sum_{k=1}^{N} Z_k$. The benchmark or auxiliary variable Y is supposed to be related to the variable of interest Z. We select n distinct units as a random sample from the finite population and for those selected units k in the sample the values of the main variable Z_k are then known. The probability to select a sample s is represented by P_s. Since sampling is without replacement so there being altogether C_n^N distinct samples and $\sum_{s \in \Omega} P_s = 1$, where Ω denotes the collection of all possible samples. The first order inclusion probability or probability of inclusion of unit k in the sample is denoted by π_k where $\pi_k = \sum_{s \ni k} P_s$. For π_k the property $\sum_{k=1}^{N} \pi_k = n$, holds. With such π_k's a popular estimator of population total Z, suggested by the Horvitz & Thompson (1952) is

$$\hat{Z} = \sum_{k \in s} \frac{z_k}{\pi_k} \tag{1.1}$$

Expression (1.1) gives unbiased population total estimate Z and Horvitz & Thompson (1952) also suggested an variance expression of $\hat{Z}$ of the form

$$\mathrm{Var}(\hat{Z}) = \sum_{k=1}^{N} (1-\pi_k)(\pi_k)^{-1} z_k^2 - 2 \sum_{1 \le k \le l \le N} \Delta_k \, z_k z_l \tag{1.2}$$

where $\Delta_k = \left(\pi_k \pi_l - \pi_{kl}\right)\left(\pi_k \pi_l\right)^{-1}$. If $\pi_{kl} > 0$, Yates and Grundy (1953) provided an unbiased estimate $v(\hat{Z})$ of $\mathrm{Var}(\hat{Z})$ where

$$v(\hat{Z}) = \sum_{k < l \in s} \Delta_k \left(\frac{z_k}{\pi_k} - \frac{z_l}{\pi_l} \right)^2 \tag{1.3}$$

We define the unbiased estimate $\hat{Z}$ for a general set of first order inclusion probabilities π_k's, but in cases where there exists evidences that Y_k is closely correlated to Z_k then it seems better option to consider $\pi_k = \sum_{k \in s} \frac{n Y_k}{Y}$ where $Y = \sum_{k=1}^{N} Y_k$.

In unequal probability sampling literature we can find a variety of sampling schemes developed by several authors where the first order inclusion probability π_k's are used as pre-assigned values e.g. some references in this context are Brewer (1963), Durbin (1967), Sampford (1967) and Samiuddin and Asad (1981). But the major purpose of these authors was to develop such sampling schemes that can be executed with simplicity and ease. Hanif and Brewer (1980) elaborated fifty such schemes in their monograph "Sampling with unequal probabilities without replacement: a review". Hanif et al. (1992) added up the material and listed about seventy schemes but now more than hundred such sampling schemes have been reviewed by them.

But the issue was that when we fix $\pi_k = \sum_{s \ni k} P_s$, P_s cannot be determined properly and no attention was paid to solve this issue in a significant way. The first meaningful work in this direction seems to be a book of Hájek published in 1981. Hájek suggested the theory of Poisson sampling design. This design maximizes the entropy for first order inclusion probabilities but it suffers due to the variable sample size. Hájek suggested to use a fixed sample size instead of variable sample size and provided the idea of conditional Poisson sampling which is also known as rejective sampling. Hájek derived Conditional Poisson sampling by maximizing entropy criteria $-\sum_{s \in \Omega} P_s \ln(P_s)$ subject to two constraints $\pi_k = \sum_{s \ni k} P_s$, and $\sum_{s \in \Omega} P_s = 1$. Stern and Cover (1989) also worked on this model and applied it to study the Canadian Lotto lotteries (See also Joe (1990)).

2. DEVELOPMENT OF A SIMPLE NEW SELECTION PROCEDURE

Since sampling is without replacement we have altogether C_n^N samples of fixed size n and P_s is the probability to select a sample s such that $\sum_{s\in\Omega} P_s = 1$. The level of uncertainty or amount of randomness about the happening of a event or outcome s when one selects the sample according to some probability mechanism P_s is calculated by the term entropy, defined as H(P) = $-\sum_{s\in\Omega} P_s \ln(P_s)$. Where H(P), the term entropy represents on average the amount of information that a sampling design contains. It is interesting to mention that Shannon (1948) tried to calculate this average amount of information transferred from one point to another place and resulted with the same entropy expression, when he was working in Bell Telephone.

In unequal probability sampling when one selects the random sample and declares that the event or outcome suggests on average the amount of information equivalent to $-\sum_{s\in\Omega} P_s \ln(P_s)$ which is the same amount of information required to eliminate the uncertainty measured by the expression entropy**.** Samiuddin & Kattan (1991) and Berger (1996) also suggested Maximum Information Sampling abbreviated as (MIS) for this information which is more familiar in the statistical community. In the current context it means determining P_s such that the expression $-\sum_{s\in\Omega} P_s \ln(P_s)$ is maximized subject to single constraint, fixed first order inclusion probability $\pi_k = \sum_{s\ni k} P_s$, where $k = 1,2,3,...,N$ which is equivalent to maximize

$$-\sum_{s\in\Omega} P_s \ln(P_s) + \sum_{k=1}^{N} \mu_k \left(\sum_{s\ni k} P_s - \pi_k \right) \tag{2.1}$$

unconditionally. Differentiating expression (2.1) with respect to P_s and equating to zero leads to

$$-\ln(P_s) - 1 + \sum_{k\in s} \mu_k = 0 \Rightarrow \ln P_s = \sum_{k\in s} \left(\mu_k - \frac{1}{n} \right) = \sum_{k\in s} \lambda_k \text{, where } \lambda_k = \mu_k - \frac{1}{n}.$$

Finally this leads to

$$P_s = e^{\sum_{k\in s} \lambda_k} \tag{2.2}$$

with suitable choice of λ_k's satisfying $\pi_k = \sum_{s\ni k} P_s$,. The maximum entropy function with P_s given by the relation (2.2) is

$$H = -\sum_{s\in\Omega} P_s \ln P_s = -\sum_{s\in\Omega} P_s \left(\sum_{k\in s} \lambda_k \right) = -\sum_{k=1}^{N} \lambda_k \sum_{s\ni k} P_s = -\sum_{k=1}^{N} \lambda_k \pi_k \tag{2.3}$$

Generally to solve the relation (2.2) is tedious and time consuming. Chen et al. (1994) put the solution in a different form by suggesting an iterative procedure which links the Conditional or rejective Poisson sample with the exponential families framework. Chen et al. (1994) also derived a draw by draw sampling scheme to select a random sample. Due to such characteristics Maximum Information Sampling (MIS) is easily executable and hopefully in future it will be a widely used popular sampling scheme. We further enhance this work to minimize the computation labor in next sections. We start from a simple sample having two units and provide a complete solution.

3. THE SIMPLE CASE OF $n = 2$

Here easily we can write the P_s for a sample of two units $s = (k,l), k < l$ as

$$P_s = \pi_{kl} = \exp\left[\lambda_k + \lambda_l\right]$$

Now

$$\pi_k = \exp\lambda_k\left[A_1 - \exp\lambda_k\right] \text{ or } \pi_k = \mathrm{e}^{\lambda_k}\left[A_1 - \mathrm{e}^{\lambda_k}\right] \tag{3.1}$$

where $A_1 = \sum_{k=1}^{N}\exp\lambda_k$. In general it can be written as $A_r = \sum_{k=1}^{N}\exp\left[r\lambda_k\right]$, where $r = 1, 2, 3, \ldots$ in future.

Relation (3.1) leads to

$$(e^{\lambda_k})^2 - A_1 e^{\lambda_k} + \pi_k = 0 \tag{3.2}$$

Relation (3.2) is of quadratic nature, its solution is

$$e^{\lambda_k} = \frac{A_1}{2}\left[1 \pm \left(1 - \frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}\right] \text{for all} \quad k = 1, 2, 3, ..., N\,. \tag{3.3}$$

Also relation (3.2) leads to $A_1 = \left[\pi_k\, e^{-\lambda_k} + e^{\lambda_k}\right]$. We differentiate $\left[\pi_k e^{-\lambda_k} + e^{\lambda_k}\right]$ w.r.t. λ_k and equating it to zero to get its max / min, which leads to $\left[-\pi_k\, \mathrm{e}^{-\lambda_k} + \mathrm{e}^{\lambda_k}\right] = 0 \Rightarrow \mathrm{e}^{2\lambda_k} = \pi_k$. The second derivative is $\left[\pi_k\, \mathrm{e}^{-\lambda_k} + \mathrm{e}^{\lambda_k}\right] > 0$ which indicates that $e^{2\lambda_k} = \pi_k$ leads to $(4\pi_k)^{\frac{1}{2}}$ which is the minimum value of A_1. Thus we can write $A_1 \geq (4\pi_k)^{\frac{1}{2}}$ or $A_1^2 \geq 4\pi_k$ for all $k = 1, 2, \ldots, N$. Thus the roots obtained by relation (3.3) are real.

Also summing both sides of relation (3.3) we get

$$\sum_{k=1}^{N} e^{\lambda_k} = A_1 = \frac{A_1}{2}\left[N + \sum_{k=1}^{N}\left(\pm\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}\right)\right]$$

which simplifies to

$$N-2 = \sum_{k=1}^{N}\left(\pm\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}\right). \tag{3.4}$$

Expression (3.4) seems to indicate many possible forms but there are just two possible solutions, the first one is

$$N-2 = \sum_{k=1}^{N}\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}} \tag{3.5}$$

The relation (3.5) has N terms each of which is ≤ 1 and also it increases with the increase in A_1^2. Thus the terms on R.H.S. of relation (3.5) will lie between $\sum_{k=1}^{N}\left(1-\frac{4\pi_k}{\pi_N}\right)^{\frac{1}{2}}$ and N. We can find a value of A_1^2 satisfying relation (3.5) if and only if $\sum_{k=1}^{N}\left(1-\frac{4\pi_k}{\pi_N}\right)^{\frac{1}{2}} \leq (N-2)$. The other possibility is if $\sum_{k=1}^{N}\left(1-\frac{\pi_k}{\pi_N}\right)^{\frac{1}{2}} < (N-2)$ then one can observe that there may be one negative term only in the R.H.S. Let it be the r^{th} term then we have

$$(N-2)+\left(1-\frac{4\pi_N}{A_1^2}\right)^{\frac{1}{2}} = \sum_{k=1}^{N-1}\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}} \geq (N-2) \tag{3.6}$$

If we put $A_1^2 = 4\pi_N$ we get the maximum of $\sum_{k=1}^{N-1}\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}$ for this to happen we have $\sum_{k=1}^{N}\left(1-\frac{\pi_k}{\pi_N}\right)^{\frac{1}{2}} \geq (N-2)$. Thus if $\sum_{k=1}^{N}\left(1-\frac{\pi_k}{\pi_N}\right)^{\frac{1}{2}} \leq (N-2)$ solve relation (3.4) for A_1^2 and if $\sum_{k=1}^{N}\left(1-\frac{\pi_k}{\pi_N}\right)^{\frac{1}{2}} \geq (N-2)$ solve relation (3.5) for A_1^2.

We now illustrate these two cases with following two examples

Example-1:

The data of this example has been taken from Chen et al. (1994).

Here $\pi_1 = 0.1,\ \pi_2 = 0.4,\ \pi_3 = 0.7$ and $\pi_4 = 0.8$ and $\sum_{k=1}^{4}\left(1-\frac{\pi_k}{\pi_4}\right)^{\frac{1}{2}} = 1.9961 < 2$.

So it is possible to find A_1^2 such that $N-2 = 2 = \sum_{k=1}^{4}\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}$.

One can observe that $A_1^2 = 3.20005$ nearly satisfies the equation. For MIS the roots are given by $e^{\lambda_k} = \frac{A_1}{2}\left\{1-\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}\right\}$ which gives $e^{\lambda_1} = 0.057767$, $e^{\lambda_2} = 0.261969$, $e^{\lambda_3} = 0.578186$ and $e^{\lambda_4} = 0.890898$. The joint inclusion probability for MIS is given by $\pi_{kl} = e^{\lambda_k+\lambda_l}$, $k<l$. The values of these π_{kl} are given in the following table. This table also contains values of π_{kl} for Brewer sampling scheme where

$$\pi_{kl} = \left[\pi_k\pi_l\left(2-\pi_k-\pi_l\right)\right]\left[\left(1-\pi_k\right)^{-1}\left(1-\pi_l\right)^{-1}\right]\left[2+\sum_{k=1}^{4}\frac{\pi_k}{\left(1-\pi_l\right)}\right]^{-1}$$

Values of Joint Inclusion Probabilities for MIS and Brewer Sampling Scheme

	π_{12}	π_{13}	π_{14}	π_{23}	π_{24}	π_{34}
MIS	0.51330	0.033400	0.051465	0.151467	0.233388	0.515105
Brewer	0.012195	0.034146	0.053658	0.153658	0.234146	0.512195

The entropy value for MIS turns out to be 1.296867 and that for the Brewer's sampling scheme its entropy value is 1.296436. In this case entropy of MIS is very close to Brewer's sampling procedure.

Example-2:

$\pi_1 = 0.2,\ \pi_2 = 0.4,\ \pi_3 = 0.6$ and $\pi_4 = 0.8$.

Here $\sum_{k=1}^{4}\left(1-\frac{\pi_k}{\pi_4}\right)^{\frac{1}{2}} = 2.073130 > N-2 = 2$. Consequently

$$N-2=2=\sum_{k=1}^{3}\left(1-\frac{4\pi_k}{A_1^2}\right)^{\frac{1}{2}}-\left(1-\frac{4\pi_4}{A_1^2}\right)^{\frac{1}{2}} \tag{3.7}$$

For $A_1^2=3.2213$ RHS = 2.000009 which nearly satisfies the equation. These give $e^{\lambda_1}=0.119373, e^{\lambda_2}=0.260747, e^{\lambda_3}=0.444271$ and $e^{\lambda_4}=0.970372$.

Values of Joint Inclusion Probabilities for MIS and Brewer Sampling Scheme

	π_{12}	π_{13}	π_{14}	π_{23}	π_{24}	π_{34}
MIS	0.031126	0.053034	0.115836	0.115836	0.253022	0.431108
Brewer	0.027722	0.053465	0.118812	0.118812	0.253465	0.427772

The entropy for MIS is 1.473636 and that for the Brewer's sampling scheme is 1.473636. Again MIS is near to the Brewer's sampling procedure.

We can write the one by one draw procedure for a sample of size 2 for MIS as, draw first unit k with probability proportional to $e^{\lambda_k}(A_1-e^{\lambda_k})/2=\pi_k/2$. After the selection of first unit we choose the second unit l, $l \neq k$ at the 2nd draw with probability proportional to $e^{\lambda_l}(A_1-e^{\lambda_l})$.

4. THE GENERAL CASE

Consider a sample s containing three units (3, 5 and 9). The units can also be permuted so that $s \equiv (3,5,9) \equiv (3,9,5) \equiv (5,3,9) \equiv (5,9,3) \equiv (9,5,3) \equiv (9,3,5)$ following this pattern we construct a general sample of n units $s=(k_1,k_2,k_3....,k_{(n-1)},k_n)$ where $k_1,k_2,k_3....,k_{(n-1)},k_n$ are distinct identifiable units of sample. Now

$$P_S=\exp(\lambda_{k_1}+\lambda_{k_2}+...+\lambda_{k_n}), \text{or}$$

$$P_S=e^{\left[\lambda_{k_1}+\lambda_{k_2}+,....,+\lambda_{k_{(n-1)}}+\lambda_{k_n}\right]}\quad , k_1<k_2<k_3,.....,<k_n.$$

$$P_s=\frac{1}{n!}e^{\left[\lambda_{k_1}+\lambda_{k_2}+,.....,+\lambda_{k_{(n-1)}}+\lambda_{k_n}\right]}\ , k_1\neq k_2\neq,k_3\neq.....,\neq k_{(n-1)}\neq k_n. \tag{4.1}$$

Also

$$\pi_k=\sum_{s\ni k}P_s=\frac{e^{\lambda_k}}{(n-1)!}\sum_{k_1\neq k_2}^{k_1=1}\sum_{k_1\neq k_2\neq k_3}^{k_2=1}....\sum_{k_{(n-1)}\neq k_{(n-2)}\neq,...,\neq k_1}^{k_{(n-1)}=1}\sum_{k_n=1}e^{\left[\lambda_{k_1}+\lambda_{k_2}+...+\lambda_{k_{(n-1)}}\right]} \tag{4.2}$$

$$\pi_k=\frac{\text{n}\ e^{\lambda_k}\ s_{(n-1)}^{(k)}}{\text{n}!}$$

$$\pi_{kl} = \sum_{s \ni k,l,k \neq l} P_s = \frac{e^{\lambda_k+\lambda_l}}{(n-2)!} \sum_{k_1 \neq k_2}^{k_1=1} \sum_{k_2 \neq k_3}^{k_2=1} \ldots\ldots \sum_{k_{(n-3)} \neq k_{(n-2)} \neq,\ldots,\neq k_1}^{k_{(n-3)}=1} \sum_{k_{(n-2)}=1} e^{\left[\lambda_{k_1}+\lambda_{k_2}+\ldots+\lambda_{k_{(n-2)}}\right]} \tag{4.3}$$

and $\pi_{kl} = \dfrac{(n^2-n)e^{\left[\lambda_k+\lambda_l\right]} s_{(n-2)}^{(k,l)}}{\text{n!}}$, similarly inclusion probabilities of higher order i.e. $\pi_{klm,}\pi_{klmn,}\ldots.$ can be derived. Let

$$s_n = \sum_{k_1 \neq k_2} \sum_{k_1 \neq k_2 \neq k_3} \ldots\ldots \sum_{k_{(n-2)} \neq k_{(n-1)}} \sum_{k_{(n-1)} \neq k_n} e^{\left[\lambda_{k_1}+\lambda_{k_2}+\ldots.+\lambda_{k_{(n-2)}}+\lambda_{k_{(n-1)}}+\lambda_{k_n}\right]} \tag{4.4}$$

We expand the R.H.S. in relation (4.4) as

$$s_n = \sum_{k_1 \neq k_2} \sum_{k_2 \neq k_2 \neq k_3} \ldots \sum_{k_{(n-3)} \neq k_{(n-2)}} \sum_{\neq k_{(n-1)}} e^{\left[\lambda_{k_1}+\lambda_{k_2}+\ldots.+\lambda_{k_{(n-2)}}+\lambda_{k_{(n-1)}}\right]} \left[A_1 - \left\{e^{\lambda_{k_1}} + e^{\lambda_{k_2}},\ldots,+e^{\lambda_{k_{(n-1)}}}\right\}\right]$$

$$= A_1 s_{(n-1)} + \left[(-1)(n-1)\right] \sum_{k_1 \neq k_2} \sum_{k_1 \neq k_2 \neq k_3},\ldots., \sum_{k_{(n-2)} \neq k_{(n-1)}} \left[e^{\left\{\lambda_{k_1}+\lambda_{k_2},\ldots.,+\lambda_{k_{(n-2)}}+2\lambda_{k_{(n-1)}}\right\}}\right]$$

Finally we obtain

$$\begin{aligned} s_n = A_1 s_{(n-1)} + \left[(-1)(n-1)\right] A_2 s_{(n-2)} +,\ldots,+ \left[(-1)^r (n-1)\ldots\left\{n-\overline{r-1}\right\}\right] A_r s_{(n-r)} \\ +,\ldots,+ \left[(-1)^{n-1}(n-1)\ldots.2.1.\right] A_n\, s_0 \end{aligned} \tag{4.5}$$

We also have to investigate how a one by one procedure of selection of sample is to be done for $n > 2$.

5. APPLICATION OF THE GENERAL PROCEDURE AND SOME USEFUL RESULTS

Our sample for $n = 2$ will be

$$s = (k,l) \text{ and } P_s = \frac{e^{\lambda_k+\lambda_l}}{2} \text{ if } l \neq k\,. \tag{5.1}$$

$\sum_{s \in \Omega} P_s = \dfrac{1}{2} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ k \neq l}}^{N} e^{\lambda_k+\lambda_l} = \dfrac{1}{2} s_2 = 1$, inducting $s_2 = A_1^2 - A_2 \Rightarrow A_1^2 - A_2 = 2$.

π_k, the first order inclusion probability to select the k th unit in the sample of size 2 is

$$\pi_k = \exp\lambda_k \left(A_1 - \exp\lambda_k\right), \text{ where } A_1 = \sum_{k=1}^{N} \exp\lambda_k\,. \tag{5.2}$$

Summing over both sides of relation (5.2) we get

$$\sum_{k=1}^{N} \pi_k = A_1^2 - \sum_{k=1}^{N} \exp\left(2\lambda_k\right) = A_1^2 - A_2 = 2 \Rightarrow \sum_{k=1}^{N} \pi_k = 2 = n.$$

For a sample of size 2, the joint or second order inclusion probability is

$$P_s = \pi_{kl} = \frac{1}{2} e^{\left[\lambda_k + \lambda_l\right]} \tag{5.3}$$

Summing over both sides of relation (5.3) we have

$$\sum_{k=1}^{N} \sum_{\substack{l=1 \\ l \neq k}}^{N} \pi_{kl} = \frac{1}{2} \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l \neq k}}^{N} e^{\left[\lambda_k + \lambda_l\right]}$$

$$\Rightarrow \frac{1}{2} \sum_{k=1}^{N} e^{\lambda_k} \left[A_1 - e^{\lambda_k}\right] = \frac{1}{2}\left[A_1^2 - A_2\right] = \left[n-1\right] = 1$$

$$\Rightarrow \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l \neq k}}^{N} \pi_{kl} = \left[n-1\right] = 1.$$

Relation can be rewritten as (3.1) as

$$\pi_k = A_1 \exp\lambda_k - \exp\left[2\lambda_k\right].$$

If $\pi_k > \pi_l \qquad \Rightarrow A_1\left[\exp\lambda_k\right] - \exp\left[2\lambda_k\right] > A_1\left[\exp\lambda_l\right] - \exp\left[2\lambda_l\right]$

i) if $\exp\lambda_k > \exp\lambda_l \Rightarrow A_1 > \left[\exp\lambda_k + \exp\lambda_l\right]$ which is true.

ii) if $\exp\lambda_k < \exp\lambda_l \Rightarrow A_1 < \left[\exp\lambda_k + \exp\lambda_l\right]$ which is not true.

So $\pi_k > \pi_l$ implies that $\exp\lambda_k > \exp\lambda_l \Rightarrow \lambda_k > \lambda_l$

iii) if $\pi_k = \pi_l \Rightarrow A_1\left[\exp\lambda_k\right] - \left[\exp\left(2\lambda_k\right)\right] = A_1\left[\exp\lambda_l\right] - \left[\exp\left(2\lambda_l\right)\right]$

$$\Rightarrow A_1 = \left[\exp\lambda_k + \exp\lambda_l\right].$$

Thus $\pi_k = \pi_l \qquad \Rightarrow \exp\lambda_k = \exp\lambda_l \Rightarrow \lambda_k = \lambda_l$.

6. EMPIRICAL STUDIES AND CONCLUSIONS

In this section we conduct an empirical study with the aim to evaluate and compare the performance of Maximum Information Sampling procedure (MIS) with Yates & Grundy (1953) draw by draw procedure (YG procedure), Prabhu and Ajgonkar Procedure (1982) (PA procedure) and Brewer (1963) Procedure (B procedure) selected from sample survey literature. For this purpose we have gathered and worked out data of seventeen populations (among them fifteen are natural and two small artificial populations), found in sampling literature. The sources of these populations along with some major characteristics i.e. main variable or variable under study, benchmark or auxiliary variable, population size, variability and correlation found between these variables, of these populations are summarized in Table 1. The population sizes ranges from 4 to 20.

Table 1
Description of Populations Characteristics

Population	Source	N	y	x	CV (y)	CV (x)	ρ
1	Chen et al. (1994)	4	Small population used by Chen (1994)				
2	Sukhatme and Sukhatme (1970)	4	Small artificial population used Yates and Grundy (1953) and Raj (1956)		0.35	0.52	0.99
3	Cochran(1977) p-268	5	Small artificial population		0.68	0.50	0.99
4	Cochran (1977) p-203	10	Actual weight	Estimated weight	0.19	0.17	0.97
5	Cochran (1963) p-325	10	# of persons per block	# of rooms per block	0.15	0.14	0.65
6	Sukhatme and Sukhatme (1970)	10	Area under wheat in 1937	Area under wheat in 1936	0.93	0.94	0.99
7	Sukhatme and Sukhatme (1970)	10	Area under wheat in 1937	Area under wheat in 1936	0.65	0.59	0.98
8	Lahiri (1951)	10	Catch of fish in Kg	# of boats	-	-	-
9	Kish (1965)	14	-	-	1.46	1.11	0.98
10	Cochran (1963) p-156	15	# of people in 1930	# of people in 1920	0.67	0.69	0.94
11	Cochran (1977)	16	# of inhabitants (in 1000's) of cities in 1930	# of inhabitants (in 1000's) of cities in 1920	0.98	0.98	0.99
12	Sampford (1962) p-61	17	Oat acreage in 1957 (even units)	Total acreage in 1947	0.71	0.61	0.80
13	Sampford (1962) p-61	18	Oat acreage in 1957 (odd units)	Total acreage in 1947	0.75	0.73	0.91
14	Cochran (1977) p-272	19	Actual # of household in a block	Eye estimate of household in a block	-	-	-
15	Sukhatme (1970)	19	Wheat acreage	# of villages	0.63	0.50	0.59
16	Yates (1960)	20	-	-	0.56	0.49	0.75
17	Cochran (1977)	20	# of people in 1930	# of people in 1920	0.10	0.10	0.99

Table 2
Entropy Values for Different Schemes

Population	H_{MIS}	H_{YG}	H_{PA}	H_B
1	1.2968**2**439	1.2968**1**711	1.2964**3**906	1.2964**3**9055
2	1.473**6**0366	1.473**5**9942	1.473**3**1923	1.473**3**1931
3	2.0282**9**054	2.0282**8**263	2.028**1**3504	2.028**1**3528
4	3.777529**3**2	3.777529**3**2	3.777529**2**5	3.777529**2**4
5	3.788422**9**0	3.788422**8**8	3.788422**8**5	3.788422**8**5
6	2.90343**8**34	2.90343**5**14	2.903**3**5878	2.903**3**5878
7	3.456533**5**	3.456533**0**	3.45652**2**363	3.4565**2**363
8	3.34027**7**93	3.34027**6**69	3.3402**5**153	3.3402**5**154
9	3.40066**4**79	3.40066**2**75	3.4006**2**003	3.4006**2**003
10	4.23737**9**40	4.23737**7**903	4.23737**2**18	4.23737**2**18
11	4.01958**4**91	4.01958**1**86	4.0195**2**35	4.0195**2**35
12	4.436059**3**4	4.436059**1**9	4.43605**6**3	4.43605**6**33
13	4.527314**1**7	4.527314**0**5	4.52731**1**76	4.5273**0**458
14	4.9904906**8**2	4.9904906**7**7	4.9904905**6**	4.9904905**6**
15	4.876283**1**2	4.876283**0**9	4.876282**8**2	4.876282**8**1
16	5.0106740**5**	5.01067403**3**5	5.010673**8**0	5.010673**3**803
17	4.35805**5**54	4.35805**4**60	4.3580**3**695	4.3580**3**694

Table 2 displays the calculated set of entropy values of the four sampling procedures for each population. The abbreviations $H_{MIS,}$ $H_{YG,}$ H_{PA} and H_B respectively denotes the entropy values of MIS, YG (1953) procedure, PA(1982) procedure and B (1963) Procedure. In each of the entropy value of the four schemes one digit at some decimal value is kept bold which determines the place to differentiate that from here the entropy value of a specific scheme is smaller or greater than the other counterpart procedures. For an example in the following table for Population No. 1, the four schemes are arranged in the order of magnitude of the entropy values and we assign rank one to the highest value and allot rank two to the next higher value of entropy and so on.

Ranking of Schemes Based on Entropy Values				
Population #	MIS	YG Procedure	PA Procedure	B Procedure
1	1.2968**2**439 >	1.2968**1**711 >	1.2964**3**906 >	1.2964**3**906
Ranks	1	2	3	4

In case where the entropy values of two procedures for a same population are equal in magnitude, we allot them similar ranks e.g. in Population No. 5 PA (1982) and B (1963) procedures have same entropy values i.e. $H_B = H_{PA} = 3.78842285$ and we assign rank 3 to both of these sampling schemes.

In Table 3 we have allotted the ranks to all the seventeen populations for these four sampling procedures. The Maximum Information Sampling (MIS) design among the four schemes attains the highest entropy values for all the seventeen populations and so we allot rank one for each population and cumulative ranks for this scheme is seventeen. The

YG (1953) sampling procedure is very close competitor to MIS, the difference between entropy values of both these schemes is insignificant. For Population No. 4 its rank is one and for the remaining sixteen populations the entropy values are positioned at place second. Cumulative rank of the YG scheme for all the seventeen populations is 33.

Table 3
Ranking of Different Schemes with Respect to Entropy Values

Popu #	E_{MIS}	E_{YG}	E_{PA}	E_B
1	1	2	3	4
2	1	2	4	3
3	1	2	4	3
4	1	1	3	4
5	1	2	3	3
6	1	2	3	3
7	1	2	3	3
8	1	2	3	4
9	1	2	3	3
10	1	2	3	3
11	1	2	3	3
12	1	2	3	3
13	1	2	3	4
14	1	2	3	3
15	1	2	3	4
16	1	2	3	3
17	1	2	3	4
Total	**17**	**33**	**53**	**57**

When we discuss the PA (1982) selection procedure it reveals that entropy values for 15 populations are ranked 3 only two Populations No. 3 and 4 have attained same rank 4 and the cumulative rank of this procedure is 53. Evaluating the performance of the B (1963) procedure it turns out to be very close to that of the PA (1982) selection procedure. Out of 17 populations 11 have identical performance for both selection procedures. All these 11 populations attained rank three each, whereas populations No. 1,4,8,13,15 and 17 are positioned at rank four each. Thus the cumulative rank for B (1963) selection scheme turns out to be 57.

Thus the above discussion prompts that the Maximum Information Sampling (MIS) procedure having high entropy values for all the populations included in this study shows much randomness then other three schemes. The PA(1982) and B(1963) selection procedures are inferior to MIS to some extent in their performance. Performance of the YG (1953) procedure is also better than PA (1982) and B (1963) selection procedures. However the PA (1982) selection procedure and B (1963) selection procedure exhibits almost equal level of performance for these populations.

We have also evaluated and compared the performance of these procedures on the basis of the Horvitz and Thompson (1952) variance values (HT (1952)). Here we allot rank one to the smallest or minimum value of variance; rank two is attached to the second last variance value and ranks are assigned to remaining values in the same pattern. In Table 4 we have displayed the calculated values of the variances using these procedures for all the 17 populations and Table 5 displays the ranks assigned to the schemes under study according to their variance values. The summery data of this table reveals that the HT- variances values of thirteen populations out of seventeen with Maximum Information Sampling (MIS) procedure are smaller than their three counterpart procedures. We allot rank 1 to these populations. The two Populations No. 13 and No. 15 are positioned at rank 3 and the Populations No. 2 and No. 3 are ranked at number four. Thus the sum of ranks of this sampling procedure turns out to be 27.

Table 4
Horvitz and Thompson Variance Values For Different Schemes

Popu #	V_{MIS}	V_{YG}	V_{PA}	V_{B}
1	0.405868	0.407315	0.41463415	0.414635
2	0.288768	0.287748	0.282178	0.282178
3	0.252981	0.252188	0.24810811	0.248108
4	276.14606	276.1447	276.1410	276.1416
5	6373.243	6373.259	6373.319	6373.319
6	24139.243	24145.78	24172.07	24172.07
7	48660.68	48672.01	48713.50	48713.5
8	1866467.6	18666508	1867368.16	1867368
9	9166.601	9167.505	9170.75	9170.747
10	83776.298	83777.05	83779.68	83780
11	55360.58	55383.96	55380.06	55380.06
12	25552.574	25549.02	25536.94	25537
13	18323.626	18323.71	18324.015	18324.01
14	2887.196	2887.224	2887.317	2887.32
15	44054215	44053756	44052254	44052254
16	4591.594	4591.631	4591.75	4591.75
17	1701724	1701737	1701782.67	1701783

Table 5
Ranking of HT – Variance Values for Different Schemes

Popu #	V_{MIS}	V_{YG}	V_{PA}	V_B
1	1	2	3	4
2	4	3	1	1
3	4	3	2	1
4	1	1	1	1
5	1	1	1	1
6	1	2	3	3
7	1	2	3	3
8	1	2	4	3
9	1	2	3	3
10	1	2	3	4
11	1	3	2	2
12	3	2	1	1
13	1	2	4	3
14	1	1	1	1
15	3	4	1	1
16	1	1	1	1
17	1	2	3	3
Total	27	35	37	36

In the YG (1953) sampling procedure the HT (1952) - variance values for the populations at No. 4,5,14 and 16 are minimum so we allot rank one to each of them, nine populations due to their variance values are ranked at position number 2, among the remaining four populations, three are ranked at number 3 and only 1 Population No. 15 having larger variance value is ranked at number 4. Finally the total sum of ranks of the YG (1953) procedure turns out to be 35.

Similarly the cumulative rank of the Prabhu and Ajgonkar (1982) sampling procedure is 37 and that of the Brewer (1963) selection procedure is 36.

The cumulative rank of Maximum Information Sampling (MIS) procedure calculated using the HT - (1952) variance criteria is minimum from their counterpart procedures. This prompts that Maximum Information Sampling (MIS) sampling procedure on average produces smaller amount of the HT - variance and thus is superior in performance to its counterparts under study. The sum of ranks of YG (1953) sampling procedure, PA (1982) and B (1963) selection procedures vary from 35 to 37. The performance of these sampling schemes is mixed, none of them dominates or outperforms the other procedure and they nearly exhibit same level of performance. This empirical study is limited to small populations and sample sizes. Hopefully the results for of Maximum Information Sampling (MIS) procedure will improve with large populations and increased sample size.

ACKNOWLEDGEMENTS

The authors are thankful to Dr. Farrukh Shehzad and referees for significant suggestions that added value to the manuscript of this paper.

REFERENCES

1. Berger, Y.G. (1996). On sampling with unequal probabilities close to rejective sampling. *SSC Annual Meeting June 1996 Proceedings of the Survey Methods Section.*
2. Brewer, K.R.W. (1963). A model of systematic sampling with unequal probabilities. *Aust. J. Statist.*, 5, 5-13.
3. Chen, S-X, Dempster, A.P. and Liu, J.S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81, 457-469.
4. Deshpande, M.N., Prabhu–Ajgonkar, S.G. (1982). An IPPS sampling scheme. *Statistica*, 36(4), 209-212.
5. Durbin, J. (1967). Design of multistage surveys for the estimation of sampling errors. *Applied Statist*., 16, 152-164.
6. Hájek, J. (1981). *Sampling from a finite population*. New York: Marcel Dekker.
7. Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.
8. Hanif, M. and Brewer, K.R.W. (1980). Sampling with unequal probabilities without replacement: A Review. *International Statistical Review*, 48, 317-335.
9. Joe, H. (1990). A winning strategy for lotto games? *Canad. J. Statist.*, 18, 233-244.
10. Samiuddin, M. and Asad, H. (1981). A simple procedure of unequal Probability Sampling. *Biometrika*, 68(3), 728-731.
11. Samiuddin, M. and Kattan, A.K. (1991). A Procedure of Unequal Probability Sampling. *Pak. J. Statist.*, 7(3)B, 1-7.
12. Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 499-513.
13. Stern, H. and Cover, T.M. (1989). Maximum entropy and the lottery. *J. Amer. Statist. Assoc.*, 84, 980-85.
14. Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technology Journal*, 27, 623-656.
15. Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B,* 15, 235-261.