

# Murthy's Estimator in Unequal Probability Inverse Adaptive Cluster Sampling

Prayad Sangngam<sup>1</sup>

<sup>1</sup> Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand

Correspondence: Prayad Sangngam, Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand. E-mail: prayad@su.ac.th

Received: October 3, 2012      Accepted: October 20, 2012      Online Published: October 25, 2012

doi:10.5539/mas.v6n11p20

URL: <http://dx.doi.org/10.5539/mas.v6n11p20>

## Abstract

This paper derives a Murthy's unbiased estimator of population total under unequal probability inverse sampling. A general unequal probability inverse sampling is combined with adaptive cluster sampling. An unbiased estimator of population total and its variance estimator are given using Murthy's approach. The general unequal probability inverse adaptive cluster sampling and general equal probability inverse adaptive cluster sampling are compared using simulation study based on real life data. The results indicate that the general unequal probability inverse adaptive cluster sampling has a small coefficient of variation for estimates compared to equal probability inverse adaptive cluster sampling. When the coefficients of correlation between study variable and probability of selection units increase, the coefficient of variation decreases.

**Keywords:** unequal probability sampling, adaptive cluster sampling, inverse sampling, unbiased estimator

## 1. Introduction

An efficient sampling design for estimating parameters of interest in a rare population is one of the most challenging areas for statisticians. Haldane (1945) used an inverse sampling for selecting units of interest in a population which contains a small fraction of the units of interest. Thompson (1990) proposed adaptive cluster sampling that is an efficient design when populations are rare and clustered because survey effort can be targeted to units of interest. One of problems encountered when adaptive cluster sampling from a rare population is that an initial sample may not contain a unit of interest. In order to get sampled units of interest, Christman and Lan (2001) applied inverse sampling to selecting an initial sample in adaptive cluster sampling. For any inverse sampling design, there is the possibility that we cannot select a fixed number ( $m$ ) of units of interest under a given resource. It might happen because  $m$  is too large or the number of units of interest in the population is too small. Salehi and Seber (2004) proposed a general inverse sampling with fixing the final sample size incorporated into adaptive cluster sampling.

The sampling designs mentioned above are taken under sampling with equal probability. If the probability of selecting a unit is highly correlated to the study variable, unequal probability sampling design can give higher efficiency than equal probability sampling design. Greco and Naddeo (2007) proposed an unequal probability inverse sampling with replacement. An unbiased estimator of the population total was derived by using conditional expectation of sample size approach. This paper derives an unbiased estimator of population total using Murthy's approach. A general unequal probability inverse sampling is combined with adaptive cluster sampling. An unbiased estimator of the population total and an unbiased variance estimator are given. A comparison of the sampling design with the general equal probability inverse adaptive cluster sampling design is considered using simulation study.

## 2. Murthy's Estimator

Murthy (1957) derived an unbiased estimator of population total as

$$\hat{\tau} = \sum_{i=1}^v y_i \frac{P(s|i)}{P(s)}, \quad (1)$$

where  $v$  is the number of distinct units in the sample,  $P(s|i)$  is the conditional probability of getting the sample  $s$ , given the  $i$ -th unit was drawn first. The variance of  $\hat{\tau}$  is given by

$$V(\hat{\tau}) = \sum_{i=1}^N \sum_{j<i}^N \left[ 1 - \sum_{s \ni i, j} \frac{P(s|i)P(s|j)}{P(s)} \right] \left( \frac{y_i}{z_i} - \frac{y_j}{z_j} \right)^2 z_i z_j. \quad (2)$$

An unbiased estimator of the variance of  $\hat{\tau}$  is

$$\hat{V}(\hat{\tau}) = \sum_{i=1}^v \sum_{j<i}^v \left[ \frac{P(s|i, j)}{P(s)} - \frac{P(s|i)P(s|j)}{[P(s)]^2} \right] \left( \frac{y_i}{z_i} - \frac{y_j}{z_j} \right)^2 z_i z_j, \quad (3)$$

where  $P(s|i, j)$  denotes the probability of getting sample  $s$  given that the  $i$ -th and  $j$ -th units were drawn in the first two draws.

Salehi and Seber (2004) proved that Murthy's estimator can be applied for sequential sampling. Using this approach, we derive unbiased estimators of population total and its variance estimator under considered sampling designs.

### 3. Unequal Probability Inverse Sampling

Assume that a finite population consists of  $N$  units with associated study values  $y_1, y_2, \dots, y_N$ . An initial selection probability of the  $i$ -th unit is denoted by  $z_i$ . The parameter to be estimated is the population total,

$$\tau = \sum_{i=1}^N y_i. \text{ Assume that population units are divided into two classes according to whether the study values}$$

satisfy a given condition. A common form of the condition is  $\{y : y > c\}$  where  $c$  is a given constant. The class

of units which study values satisfy the condition is defined to be the class ( $C$ ) of interest and  $\bar{C}$  is the class of the remaining units.

In *unequal probability inverse sampling*, we select units one at a time with unequal probability with replacement until we have obtained a given number ( $m$ ) of units from class  $C$ , in the sample. The sample size ( $n_1$ ) is a random variable. The sample ( $s$ ) can be partitioned into two parts: a part  $S_c$  is the set of sample units from the class  $C$  and  $S_{\bar{c}}$  is the set of sample units from  $\bar{C}$  with cardinalities  $m$  and  $n_1 - m$ , respectively.

Let  $k$  and  $g$  be the numbers of distinct units in  $S_c$  and  $S_{\bar{c}}$ , and which indexed by  $i = 1, 2, \dots, k$  and  $i = k + 1, k + 2, \dots, v$ , respectively. If  $r_i$  is the number of times that unit  $i$  appears in the sample. With an ordered

sample ( $S^*$ ), the probability of getting the ordered sample is  $P(S^*) = \prod_{i=1}^v (z_i)^{r_i}$  where the last sampled unit

belongs to the set  $S_c$ . For an unorded sample ( $s$ ), the last sampled unit must belongs to  $S_{\bar{c}}$ , so that after

allocating one  $i$ -th unit of  $k$  sampled unit in  $S_c$ , the rest of sampled units can be ordered in  $\binom{n_1 - 1}{r_1, r_1 - 1, \dots, r_v}$

ways. Therefore, the sample  $s$  can be constructed in  $\sum_{i=1}^k \binom{n_1 - 1}{r_1, r_1 - 1, \dots, r_v}$  ways. The probability of getting a sample  $s$  is

$$P(s) = \sum_{i=1}^k \binom{n_1 - 1}{r_1, r_1 - 1, \dots, r_v} \prod_{i=1}^v (z_i)^{r_i}.$$

**Theorem 1** Under the unequal probability inverse sampling design, an unbiased estimator of the population total is

$$\hat{\tau}_1 = \hat{P} \bar{y}_C + (1 - \hat{P}) \bar{y}_{\bar{C}}, \quad (4)$$

where  $\hat{P} = \frac{m-1}{n_1-1}$ ,  $\bar{y}_C = \frac{1}{m} \sum_{i \in s_C} y_i$  and  $\bar{y}_{\bar{C}} = \frac{1}{n_1-m} \sum_{i \in s_{\bar{C}}} y_i$ . An unbiased estimator of the variance of  $\hat{\tau}_1$  is

$$\hat{V}(\hat{\tau}_1) = (\bar{y}_C - \bar{y}_{\bar{C}})^2 \hat{V}(\hat{P}) + \frac{\hat{\sigma}_C^2}{m} \left[ (m-1) \hat{V}(\hat{P}) - \hat{P}^* \hat{P}^2 \right] + \frac{\hat{\sigma}_C^2 (n_1 - m - 1)}{(n_1 - 1)(n_1 - 2)} \quad (5)$$

**Proof:** Using Murthy's estimator, the  $\frac{P(s|i)}{P(s)}$  is

$$\frac{P(s|i)}{P(s)} = \begin{cases} \frac{r_i(m-1)}{z_i(n_1-1)m} & \text{if } i \in s_C \\ \frac{r_i}{z_i(n_1-1)} & \text{if } i \in s_{\bar{C}} \end{cases}$$

Substituting  $\frac{P(s|i)}{P(s)}$  into (1) and using some algebra, we obtain the expression (4). The  $\frac{P(s|i,j)}{P(s)}$  is given by

$$\frac{P(s|i,j)}{P(s)} = \begin{cases} \frac{r_i r_j (m-2)}{z_i z_j (n_1-1)(n_1-2)m} & \text{if } i, j \in s_C \\ \frac{r_i r_j (m-1)}{z_i z_j (n_1-1)(n_1-2)m} & \text{if } i \in s_C \text{ and } j \in s_{\bar{C}} \\ \frac{r_i r_j}{z_i z_j (n_1-1)(n_1-2)} & \text{if } i, j \in s_{\bar{C}} \end{cases}$$

Substituting  $\frac{P(s|i,j)}{P(s)}$  and  $\frac{P(s|i)P(s|j)}{[P(s)]^2}$  into (3) and using some algebra, we obtain the result as in

expression (5).

Note that using Murthy's estimator, we obtain the same estimator given by Greco and Naddeo (2007).

#### 4. General Unequal Probability Inverse Sampling

In a sampling procedure which avoids sampling large number of sample size, an initial sample of size  $n_0$  is drawn by unequal probability sampling with replacement where  $n_0 \geq m$ . We stop further sampling if the initial sample consists at least  $m$  units from class C. Otherwise, we sequentially continue sampling with unequal probability with replacement until either the sample consists of  $m$  units from the class C or the total sample size is equal to  $n_2$  ( $n_2 \geq n_0$ ) where  $n_2$  is fixed in advance. This sampling scheme is called a *general unequal probability inverse sampling*. Let  $n_1$  denote the sample size of the final sample.

If  $n_0 = m$  and  $n_2$  is not fixed, the sampling design reduces to unequal probability inverse sampling. The sampling design reduces to original unequal probability sampling when  $n_0 = n_2$ . If the probabilities of selection are equal, the sampling design is a general inverse sampling.

**Theorem 2** Under the general unequal probability inverse sampling design, an unbiased estimator of the population total is

$$\hat{\tau}_{II} = \begin{cases} \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{y_i}{z_i} & \text{if } [Q_0] \\ \hat{P}\bar{y}_C + (1 - \hat{P})\bar{y}_{\bar{C}} & \text{if } [Q_1], \\ \frac{1}{n_2} \sum_{i=1}^{n_2} \frac{y_i}{z_i} & \text{if } [Q_2] \end{cases} \tag{6}$$

where  $\hat{P} = \frac{m-1}{n_1-1}$ ,  $\bar{y}_C = \frac{1}{m} \sum_{i \in s_C} \frac{y_i}{z_i}$ ,  $\bar{y}_{\bar{C}} = \frac{1}{n_1-m} \sum_{i \in s_{\bar{C}}} \frac{y_i}{z_i}$ ,  $[Q_0] = \{n_1 = n_0\}$ ,

$[Q_1] = \{(n_0 < n_1 < n_2) \text{ or } (n_1 = n_2 \text{ and } |s_C| = m)\}$  and  $[Q_2] = \{n_1 = n_2 \text{ and } |s_C| < m\}$ . An unbiased estimator of the variance of  $\hat{\tau}_{II}$  is

$$\hat{V}(\hat{\tau}_{II}) = \begin{cases} \frac{1}{n_0(n_0-1)} \sum_{i=1}^{n_0} \left( \frac{y_i}{z_i} - \hat{\tau}_I \right)^2 & \text{if } [Q_0] \\ (\bar{y}_C - \bar{y}_{\bar{C}})^2 \hat{V}(\hat{P}) + \frac{\hat{\sigma}_C^2}{m} [(m-1)\hat{V}(\hat{P}) - \hat{P}^* \hat{P}^2] + \frac{\hat{\sigma}_{\bar{C}}^2 (n_1 - m - 1)}{(n_1 - 1)(n_1 - 2)} & \text{if } [Q_1], \\ \frac{1}{n_2(n_2-1)} \sum_{i=1}^{n_2} \left( \frac{y_i}{z_i} - \hat{\tau}_I \right)^2 & \text{if } [Q_2] \end{cases} \tag{7}$$

where  $\hat{P}^* = \frac{(m-1)(m-2)}{(n_1-1)(n_1-2)}$ ,  $\hat{\sigma}_C^2 = \frac{1}{m-1} \sum_{i \in s_C} \left( \frac{y_i}{z_i} - \bar{y}_C \right)^2$ ,  $\hat{\sigma}_{\bar{C}}^2 = \frac{1}{n_1-m-1} \sum_{i \in s_{\bar{C}}} \left( \frac{y_i}{z_i} - \bar{y}_{\bar{C}} \right)^2$  and  $\hat{V}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{n_1-2}$ .

**Proof:** When  $n_1 = n_0$  and  $n_1 = n_2$ , the  $\frac{P(s|i)}{P(s)}$  is equal to  $\frac{r_i}{n_0 z_i}$  and  $\frac{r_i}{n_2 z_i}$ , respectively. Substituing

$\frac{P(s|i)}{P(s)}$  into (1), we obtain case  $[Q_0]$  and  $[Q_2]$  of expression (6). In case  $[Q_1]$ , the  $\frac{P(s|i)}{P(s)}$  is

$$\frac{P(s|i)}{P(s)} = \begin{cases} \frac{r_i(m-1)}{z_i(n_1-1)m} & \text{if } i \in s_C \\ \frac{r_i}{z_i(n_1-1)} & \text{if } i \in s_{\bar{C}} \end{cases}$$

Substituing  $\frac{P(s|i)}{P(s)}$  into (1), we obtain the case  $[Q_1]$  of expression (6). When  $n_1 = n_0$  and  $n_1 = n_2$ , the

$\frac{P(s|i,j)}{P(s)}$  are equal to  $\frac{r_i r_j}{z_i z_j n_0 (n_0 - 1)}$  and  $\frac{r_i r_j}{z_i z_j n_2 (n_2 - 1)}$ , respectively. In case  $[Q_1]$ , the  $\frac{P(s|i,j)}{P(s)}$  is

given by

$$\frac{P(s|i,j)}{P(s)} = \begin{cases} \frac{r_i r_j (m-2)}{z_i z_j (n_1-1)(n_1-2)m} & \text{if } i, j \in s_C \\ \frac{r_i r_j (m-1)}{z_i z_j (n_1-1)(n_1-2)m} & \text{if } i \in s_C \text{ and } j \in s_{\bar{C}} \\ \frac{r_i r_j}{z_i z_j (n_1-1)(n_1-2)} & \text{if } i, j \in s_{\bar{C}} \end{cases}$$

Substituting  $\frac{P(s|i,j)}{P(s)}$  and  $\frac{P(s|i)P(s|j)}{[P(s)]^2}$  into (3) and using some algebra, we obtain the result as in expression

(7).

### 5. General Unequal Inverse Adaptive Cluster Sampling

We apply the general unequal probability inverse sampling to adaptive cluster sampling. Assume that each unit in the population is defined to have a neighborhood. The neighborhood of a unit is a set of other units associated with the unit. The sampling scheme is as following.

An initial sample is drawn by the general unequal probability inverse sampling with replacement. For the sample units in class  $C$ , their neighborhoods are added to be sampled and observed. The procedure continues until no more units in the class  $C$  are found. This sampling scheme is called a *general unequal probability inverse adaptive cluster sampling* because sampling begins with general unequal probability inverse sampling and incorporates adaptive cluster sampling. The final sample consists of the initial sample and all adaptively sampled units.

The set of units that are adaptively sampled as a result of unit  $i$ -th being sampled and that are also the member of class  $C$  are called the network to which the unit  $i$  belongs. The units that are adaptively sampled but are in the class  $\bar{C}$  are called edge units. By this way, if any unit in the  $i$ -th network is selected in the initial sample, all units in the network are sampled. From definition of network, the population can be divided into  $K$  mutually exclusive networks.

Let  $n$  be the final sample size. Let  $\psi_k$  denote the set of units in the  $k$ -th network and  $m_k$  denote a number of units in the network. The total value of the study variable in a network  $\psi_k$  is  $y_k^* = \sum_{j \in \psi_k} y_j$  and the

probability of selection of that network is  $z_k^* = \sum_{j \in \psi_k} z_j$ . The parameter to be estimated can be written as

$\tau = \sum_{i=1}^N y_i = \sum_{k=1}^K y_k^*$ . Since the probability of any edge unit included in the final sample is not known, some

estimators which included edge units will be biased. So the edge units in the final sample are excluded from the estimation stage. An unbiased estimator for the population total uses the sample units in the class  $\bar{C}$  only when they are drawn to be the initial sample. The estimator is formed by modifying the unbiased estimator given by Theorem 2.

**Theorem 3** Under the general unequal probability inverse adaptive cluster sampling design, an unbiased estimator of the population total is

$$\hat{\tau}_{III} = \begin{cases} \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{y_i^*}{z_i^*} & \text{if } [Q_0] \\ \hat{P} \bar{y}_C^* + (1 - \hat{P}) \bar{y}_{\bar{C}}^* & \text{if } [Q_1] \\ \frac{1}{n_2} \sum_{i=1}^{n_2} \frac{y_i^*}{z_i^*} & \text{if } [Q_2] \end{cases} \quad (8)$$

where  $\hat{P} = \frac{m-1}{n_1-1}$ ,  $\bar{y}_C^* = \frac{1}{m} \sum_{i \in S_C} \frac{y_i^*}{z_i^*}$  and  $\bar{y}_{\bar{C}}^* = \frac{1}{n_1-m} \sum_{i \in S_{\bar{C}}} \frac{y_i^*}{z_i^*}$ . An unbiased estimator of the variance  $V(\hat{\tau}_{III})$  is

$$\hat{V}(\hat{\tau}_{III}) = \begin{cases} \frac{1}{n_0(n_0-1)} \sum_{i=1}^{n_0} \left( \frac{y_i^*}{z_i^*} - \hat{\tau}_{II} \right)^2 & \text{if } [Q_0] \\ \left( \bar{y}_C^* - \bar{y}_{\bar{C}}^* \right)^2 \hat{V}(\hat{P}) + \frac{\hat{\sigma}_C^{*2}}{m} \left[ (m-1) \hat{V}(\hat{P}) - \hat{P}^* \hat{P}^2 \right] + \frac{\hat{\sigma}_C^{*2} (n_1 - m - 1)}{(n_1 - 1)(n_1 - 2)} & \text{if } [Q_1], \\ \frac{1}{n_2(n_2-1)} \sum_{i=1}^{n_2} \left( \frac{y_i^*}{z_i^*} - \hat{\tau}_{II} \right)^2 & \text{if } [Q_2] \end{cases} \quad (9)$$

where  $\hat{P}^* = \frac{(m-1)(m-2)}{(n_1-1)(n_0-2)}$ ,  $\hat{\sigma}_C^{*2} = \frac{1}{m-1} \sum_{i \in S_C} \left( \frac{y_i^*}{z_i^*} - \bar{y}_C^* \right)^2$ ,  $\hat{\sigma}_{\bar{C}}^{*2} = \frac{1}{n_0-m-1} \sum_{i \in S_{\bar{C}}} \left( \frac{y_i^*}{z_i^*} - \bar{y}_{\bar{C}}^* \right)^2$  and

$$\hat{V}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{n_0-2}.$$

**Proof:** Let  $w_i$  represent the new value of a study variable of the  $i$ -th unit in the  $k$ -th network, given by  $w_i = \frac{Z_i Y_k^*}{z_k^*}$ . Under an initial sample with general unequal probability inverse sampling, when observed value  $y_i$  is replaced by the value  $w_i$ , we obtain the results of this theorem.

### 6. Simulation Study

The ring-necked ducks data given by Smith et al. (1995) was used as the study population. The population consists of  $N = 200$  units. In general unequal probability inverse adaptive cluster sampling, the population units are selected by using probability proportional to auxiliary variable. Auxiliary variable ( $x$ ) correlated to the study variable are generated with the coefficients of correlation equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8.

The initial selection probability of the  $i$ -th unit is  $z_i = x_i / \sum_{j=1}^{200} x_j$ . For general inverse adaptive cluster sampling,

the probability of selection a unit is equal to  $1/200$ .

Simulations of sampling from the population were carried out to study the properties of the general unequal probability inverse adaptive cluster sampling (GUIACS) compared to the general inverse adaptive cluster sampling (GIACS) when sampling is with replacement. We chose the condition  $\{y : y > 0\}$  for dividing the

units into class  $C$  or  $\bar{C}$ . The numbers ( $m$ ) of sample units satisfying the condition are 3, 4, 5 and 6. The

initial sample sizes ( $n_0$ ) are given as 10, 15, 20 and 25. The final sample sizes ( $n_2$ ) in the initial sample are

25, 35, 45 and 55. The neighborhood of a unit is defined as the set of the four adjacent units. The simulation consists of 50,000 replications. The population total ( $\tau$ ) was estimated for each sample. In each sampling design,

the values of the estimates ( $\hat{\tau}$ ) are averaged. The averages were interpreted as expected values, i.e.,

$$\tilde{E}(\hat{\tau}) = \frac{1}{50,000} \sum_{i=1}^{50,000} \hat{\tau}_i. \text{ The estimate of the variance is } \tilde{V}(\hat{\tau}) = \frac{1}{50,000-1} \sum_{i=1}^{50,000} (\hat{\tau}_i - \tilde{\tau})^2, \text{ where } \tilde{\tau} = \tilde{E}(\hat{\tau}). \text{ The}$$

estimate of coefficient of variation is given by  $cv(\hat{\tau}) = \sqrt{\tilde{V}(\hat{\tau})} / \tilde{E}(\hat{\tau})$ .

Table 1. The coefficient of variation of estimates under GIACS and GUIACS with  $\rho = 0.1, 0.2$  where  $m = 3, 4, 5, 6$ , and initial sample size  $n_0 = 10, 15, 20, 25$  with  $n_2 = 25, 35, 45, 55$

$n_2$	$m$	GIACS				GUIACS $\rho = 0.1$				GUIACS $\rho = 0.2$			
		$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$
55	6	1.256	1.249	1.238	1.226	1.178	1.166	1.161	1.138	1.096	1.100	1.085	1.064
	5	1.420	1.388	1.364	1.317	1.304	1.294	1.260	1.223	1.227	1.216	1.181	1.142
	4	1.599	1.540	1.465	1.396	1.489	1.433	1.363	1.289	1.412	1.346	1.277	1.212
	3	1.864	1.730	1.582	1.462	1.742	1.604	1.462	1.350	1.641	1.498	1.374	1.257
45	6	1.291	1.281	1.270	1.252	1.197	1.184	1.189	1.161	1.110	1.113	1.105	1.087
	5	1.419	1.388	1.363	1.318	1.305	1.298	1.268	1.223	1.229	1.224	1.188	1.149
	4	1.620	1.549	1.466	1.385	1.485	1.441	1.367	1.279	1.399	1.355	1.273	1.202
	3	1.874	1.726	1.565	1.451	1.741	1.587	1.453	1.342	1.625	1.487	1.358	1.248
35	6	1.367	1.363	1.345	1.320	1.262	1.251	1.246	1.230	1.173	1.174	1.164	1.145
	5	1.455	1.430	1.406	1.366	1.349	1.332	1.308	1.265	1.280	1.250	1.232	1.186
	4	1.620	1.565	1.484	1.409	1.519	1.458	1.382	1.308	1.425	1.372	1.294	1.218
	3	1.866	1.726	1.572	1.460	1.749	1.606	1.462	1.351	1.635	1.506	1.367	1.259
25	6	1.529	1.523	1.521	1.500	1.426	1.416	1.402	1.391	1.331	1.315	1.304	1.298
	5	1.568	1.564	1.544	1.503	1.461	1.453	1.42	1.392	1.374	1.354	1.327	1.300
	4	1.689	1.637	1.582	1.489	1.568	1.525	1.458	1.384	1.479	1.420	1.358	1.290
	3	1.920	1.750	1.611	1.504	1.777	1.634	1.488	1.389	1.662	1.520	1.393	1.301

When  $n_0 = 25$  and  $n_2 = 25$ , the sampling designs reduce to original adaptive cluster sampling. Table 1 shows that the general unequal probability inverse adaptive cluster sampling design outperforms the general inverse adaptive cluster sampling design. Tables 1, 2 and 3 indicate that the coefficients of variation of the estimates under GUIACS decrease when the coefficients of correlation increase with given the numbers  $m$ ,  $n_0$  and  $n_2$ . Under  $m$ ,  $n_0$  and  $\rho$  fixed, the coefficient of variation decreases when the initial sample size ( $n_0$ ) increases. The coefficient of variation decreases when the number of sampled units in class C increases for given sample sizes  $n_0$  and  $n_2$ . With fixed numbers  $m$ ,  $n_0$  and  $\rho$ , the coefficient of variation of the estimates decreases when the truncated sample size ( $n_2$ ) increases. However, if  $n_2$  is too large, the sampling design reduces to inverse adaptive cluster sampling. In this case, although the value  $n_0$  increases, the coefficient of variation may not decrease.

Table 2. The coefficient of variation of estimates under GUIACS with  $\rho = 0.3, 0.4, 0.5$  where  $m = 3, 4, 5, 6$ , and initial sample size  $n_0 = 10, 15, 20, 25$  with  $n_2 = 25, 35, 45, 55$

$n_2$	$m$	$\rho = 0.3$				$\rho = 0.4$				$\rho = 0.5$			
		$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$
55	6	1.041	1.048	1.024	1.011	0.982	0.971	0.962	0.943	0.938	0.930	0.933	0.904
	5	1.154	1.156	1.118	1.082	1.080	1.075	1.047	1.009	1.039	1.030	1.002	0.969
	4	1.334	1.283	1.204	1.139	1.232	1.198	1.123	1.062	1.199	1.145	1.080	1.022
	3	1.546	1.414	1.294	1.185	1.441	1.322	1.199	1.095	1.380	1.263	1.148	1.050
45	6	1.064	1.057	1.041	1.035	0.990	0.985	0.974	0.957	0.950	0.944	0.945	0.920
	5	1.164	1.160	1.118	1.088	1.091	1.085	1.050	1.009	1.051	1.043	1.008	0.969
	4	1.327	1.280	1.207	1.143	1.246	1.196	1.130	1.061	1.193	1.140	1.085	1.014
	3	1.537	1.411	1.288	1.182	1.437	1.311	1.195	1.100	1.377	1.260	1.144	1.048
35	6	1.110	1.111	1.099	1.079	1.032	1.038	1.021	1.004	0.990	0.986	0.979	0.960
	5	1.213	1.191	1.164	1.123	1.125	1.105	1.068	1.042	1.080	1.058	1.026	0.997
	4	1.339	1.300	1.231	1.148	1.250	1.207	1.142	1.063	1.209	1.164	1.093	1.025
	3	1.536	1.422	1.299	1.189	1.447	1.317	1.206	1.105	1.388	1.266	1.156	1.055
25	6	1.251	1.242	1.227	1.217	1.156	1.149	1.138	1.125	1.117	1.105	1.092	1.079
	5	1.289	1.277	1.252	1.221	1.203	1.186	1.169	1.128	1.163	1.138	1.112	1.079
	4	1.403	1.342	1.287	1.213	1.298	1.242	1.188	1.123	1.253	1.189	1.138	1.075
	3	1.576	1.445	1.310	1.222	1.463	1.338	1.222	1.131	1.404	1.278	1.161	1.081

Table 3. The coefficient of variation of estimates under GUIACS with  $\rho = 0.6, 0.7, 0.8$  where  $m = 3, 4, 5, 6$ , and initial sample size  $n_0 = 10, 15, 20, 25$  with  $n_2 = 25, 35, 45, 55$

$n_2$	$m$	$\rho = 0.6$				$\rho = 0.7$				$\rho = 0.8$			
		$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$	$n_0=10$	$n_0=15$	$n_0=20$	$n_0=25$
55	6	0.892	0.878	0.874	0.855	0.839	0.835	0.813	0.808	0.763	0.759	0.753	0.725
	5	1.000	0.971	0.953	0.912	0.940	0.915	0.892	0.854	0.848	0.834	0.808	0.771
	4	1.128	1.086	1.022	0.955	1.068	1.018	0.958	0.894	0.977	0.920	0.861	0.802
	3	1.307	1.200	1.077	0.990	1.225	1.116	1.011	0.922	1.114	0.998	0.902	0.822
45	6	0.895	0.897	0.886	0.867	0.846	0.844	0.834	0.813	0.766	0.767	0.754	0.733
	5	0.994	0.980	0.955	0.912	0.944	0.926	0.897	0.859	0.857	0.841	0.807	0.772
	4	1.133	1.085	1.024	0.961	1.071	1.021	0.959	0.903	0.979	0.921	0.866	0.803
	3	1.316	1.187	1.081	0.992	1.232	1.113	1.013	0.921	1.110	0.997	0.908	0.823
35	6	0.930	0.931	0.924	0.906	0.879	0.879	0.863	0.846	0.794	0.785	0.775	0.758
	5	1.020	1.000	0.975	0.938	0.963	0.948	0.911	0.873	0.867	0.854	0.819	0.784
	4	1.140	1.095	1.033	0.970	1.075	1.024	0.964	0.907	0.972	0.928	0.866	0.807
	3	1.314	1.191	1.091	0.994	1.236	1.122	1.016	0.926	1.121	1.004	0.908	0.824
25	6	1.050	1.042	1.030	1.005	0.975	0.975	0.961	0.939	0.875	0.866	0.860	0.843
	5	1.099	1.071	1.039	1.018	1.031	1.009	0.974	0.943	0.918	0.901	0.866	0.838
	4	1.189	1.125	1.066	1.007	1.101	1.058	0.994	0.942	0.990	0.953	0.886	0.837
	3	1.321	1.206	1.099	1.013	1.242	1.123	1.026	0.943	1.117	1.001	0.918	0.837

## 7. Discussion

An adaptive cluster sampling is an efficient sampling design for rare and clustered population. However, if an initial sample in adaptive cluster sampling is selected by fixed sample size design, a sample may not consist of units of interest. The paper considered a general unequal probability inverse sampling and combined it with adaptive cluster sampling. An unbiased estimator of the population total and an unbiased estimate of its variance were derived using Murthy's method. The simulation study showed that the efficiency of the general unequal probability inverse adaptive sampling design depends on the coefficient of correlation between the study variable and the auxiliary variable, number of sampled units of interest, initial sample size and truncated sample size. When the auxiliary variable is highly correlated with the study variable the general unequal probability inverse adaptive cluster sampling design is fairly efficiency. In addition, the general unequal probability inverse adaptive cluster sampling design is more efficient than general inverse adaptive cluster sampling design. However, when the auxiliary variable is not appropriate for the study variable, an estimator of parameter by using the general unequal probability inverse adaptive cluster sampling design may lose its efficiency.

## Acknowledgements

This work was supported by the Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand. I am grateful to referees and associate editors for their comments.

## References

- Chistman, M. C., & Lan, F. (2001). Inverse adaptive cluster sampling. *Biometrics*, 57, 1096-1105. <http://dx.doi.org/10.1111/j.0006-341X.2001.01096.x>
- Greco, L., & Naddeo, N. (2007). Inverse sampling with unequal selection probabilities. *Communications in Statistics: Theory and Methods*, 36(5), 1039-1048. <http://dx.doi.org/10.1080/03610920601033926>
- Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika*, 33, 222-225. <http://dx.doi.org/10.1093/biomet/33.3.222>
- Murthy, M. N. (1957). Ordered and Unordered Estimators in Sampling without Replacement. *Sankhya*, 18, 379-390.
- Salehi, M. M., & Seber, G. A. F. (2001). A new proof of Murthy's estimator with applies to sequential sampling. *Australian and New Zealand Journal of Statistics*, 43(3), 281-286. <http://dx.doi.org/10.1111/1467-842X.00174>
- Salehi, M. M., & Seber, G. A. F. (2004). A general inverse sampling scheme and its application to adaptive cluster sampling. *Australian and New Zealand Journal of Statistics*, 46(3), 483-494. <http://dx.doi.org/10.1111/j.1467-842X.2004.00345.x>
- Smith, D. R., Conroy, M. J., & Bralhage, D. H. (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, 51, 777-788. <http://dx.doi.org/10.2307/2532964>
- Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85, 1050-1059. <http://dx.doi.org/10.1080/01621459.1990.10474975>