Variance Estimation for Adaptive Cluster Sampling with a Single Primary Unit

Urairat Netharn

National Institute of Development Administration, School of Applied Statistics.
(ffisurnh@ku.ac.th)

Abstract

In this paper, an initial sample of the adaptive cluster sampling design is considered in terms of primary and secondary units. The primary unit contains units that are called secondary units and all these secondary units are arranged in systematic order. Although in some situations in field work it is necessary to use a single primary, the variance estimator of the estimate of the population mean or total in this design is not available now. Two new variance estimators of the estimate of the population mean were found based on the method of splitting the sample. The preliminary study in terms of bias, MSE and percentage of confidence interval containing the population mean for each variance estimator was carried out with a small population.

1. Introduction

Adaptive cluster sampling, a design with primary and secondary units, was proposed by Thompson [1] and is a suitable design for rare and cluster populations, especially for a biological population. In this design a primary unit contains units that are called secondary units. All of these secondary units are arranged in systematic order. A primary unit, sometimes called a systematic sample, is selected at random. The selected primary unit is called the initial sample. All secondary units within the initial sample will be automatically consisted in the sample. The condition for adding units is that y-values must be equal to or greater than *c*, a pre-specified value. Whenever a secondary unit within an initial sample satisfies the condition, its neighborhood will be added to the sample. Note that if a u_{ij} is in the neighborhood of unit $u_{i'j'}$; then unit $u_{i'j'}$ is also in the neighborhood of unit u_{ij}. A secondary unit that is not in the initial sample is called an adaptively added unit. An adaptively added unit may or may not satisfy the condition, but if it does not, it will be called an *edge unit*. A set of secondary units that satisfies the condition is called a *network*. A network and its associated edge units make up a *cluster*. For example, the study area is divided into 36 units of equal size. The y-value within a unit is the count of black dots. The condition of interest is $C = \{y_{ij} : y_{ij} \geq 1\}$. By using systematic sampling method there are 4 possible primary units. Figure 1A is one of them that the selected primary unit is a set of 9 dark squares. Starting with the gray area in figure 1A which is a secondary unit within the selected primary unit, its neighborhood is a set of units in the North, South, East and West direction as shown in figure 1B. The set of units that is shaded with the gray color in figure 1C is called a network. Edge units and all units within the dash lines are a cluster as shown in figure 1D.

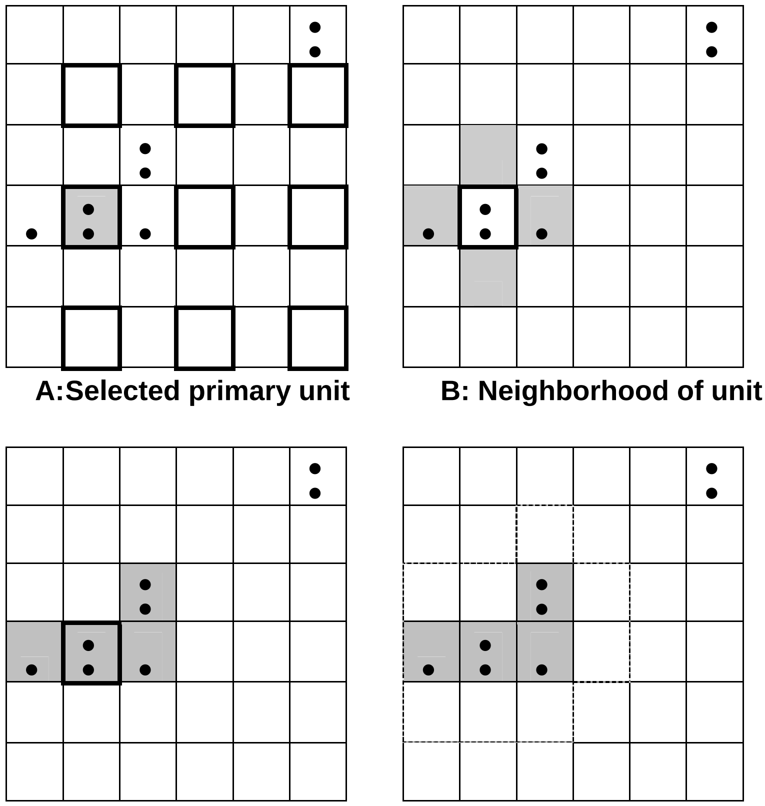

Figure 1: An example of a selected sample from adaptive clustering sampling with a single primary unit.

In real work, an initial sample with a single primary unit is necessary in some situations. However, as stated in Thompson [1] "For an initial systematic sample with only one starting point (i.e., only one primary unit is selected), some of the joint inclusion probabilities are zero, underscoring the fact that an unbiased estimator of variance is not available for such a design." This problem is likely to occur with conventional sampling when a conventional systematic sample is used. In such a case, eight biased estimators of the variance are reviewed and compared in Wolter [2]. Two of them are based on the method of splitting the sample in to sub-sample. The first estimator though is not recommended to be used in practice because of large value of bias when the number of sub-samples equals two, is simple for application and its bias may be reduced when the number of sub-samples is increased which make it interesting. The second estimator often gives a small bias compared to the other estimators in many situations, such as the population that has a linear trend and stratification effects, etc. Therefore, these two methods of estimation are of interest and will be adaptive cluster sampling with a single primary unit.

In this paper, two methods of variance estimation will be proposed based on the above two methods. The properties of these estimators that will be considered are bias, MSE, and the proportion of confidence interval that contain the population mean.

2. Methodology

In conventional systematic sampling from a finite population of size $N = nq$, where n is a sample size, $q = \frac{N}{n}$ is the sampling interval. Let Y_{ij} be the y-value of the jth units in the ith systematic sample, where $i = 1,2,...,q$ and $j = 1,2,...,n$. Suppose the sth systematic sample is selected, let $y_j^* = Y_{sj}$, where $j = 1,2,...,n$, y_j be the total y-value of units within the network associated with y_j^* and x_j be the number of primary units in the network associated with

y_j. The problem can be considered in two methods. First, each sample of size $n = mp$, say, is treated as if it is divided into p sub-samples with size $m = \frac{n}{p}$ units. Second, for $m = 2$, the sample is treated as if it is divided into $p = \frac{n}{2}$ groups (assume that n is an even number) and each group of size $m = 2$. Both methods give the biased variance estimators [2] as follows

Method I: $n = mp, p \geq 2$ and $m \geq 2$.

$$\hat{v}_1(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{1}{p}\frac{\sum_{t=1}^{p}(\bar{y}_t - \bar{y})^2}{p-1} \tag{1}$$

where $\bar{y}_t$ represents the sample mean of the t[th] sub-sample of size $m = \frac{n}{p}$ and $\bar{y}$ represents the systematic sampling mean.

Method II: $n = 2p, p \geq 2$ and $m = 2$.

$$\hat{v}_2(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{1}{n}\frac{\sum_{j=1}^{n/2}\left(y_{2j}^{*} - y_{2j-1}^{*}\right)^2}{n} \tag{2}$$

where y_{2j}^{*} denotes the y-value of the (2j)[th] unit.

In Adaptive Sampling Design, based on Thompson [1], an unbiased estimator of the population mean and its true variance when only one primary unit selected are

$$\hat{\mu} = \frac{1}{n}\sum_{k \in \kappa_s}\frac{y_k}{x_k} \tag{3}$$

$$V(\hat{\mu}) = \left(\frac{1}{q}\right)\sum_{s=1}^{q}(\hat{\mu}_s - \mu)^2 \tag{4}$$

where y_k is the total y-values of the k[th] network, x_k is the number of primary units associated with the k[th] network, $\hat{\mu}_s$ represents the sample mean if the s[th] primary unit is selected and κ_s is the number of networks associated with the selected primary unit.

In order to find the estimator of variance of $\hat{\mu}$, the selected sample will be split at random into p

sub-samples, where $p \geq 2$. Each sub-sample contains m secondary units, where $m = \frac{n}{p}$ and $m \geq 2$. There are $\frac{n!}{(m!)^p \, p!}$ possible sub-samples. After that, all units within the network corresponding to the selected sample will be included. Two methods of estimation are investigated: First, the number sub-samples is $p \geq 2$ and each contains $m \geq 2$ secondary units. Second, the number of sub-samples is $p \geq 2$ and $m = 2$. Then the estimator of the true variance $V(\hat{\mu})$ can be found. Details can be seen in the following section.

3. Variance Estimation: The proposed variance estimators

Let $\hat{\mu}_t = \frac{1}{m} \sum_{k \in \kappa_t} \frac{y_k}{h_k x_k}$ be the sample mean of the t[th] sub-sample, t represent the sub-samples, where m is the number of secondary units in each sub-sample, h_k is the number of sub-samples in which the k[th] network appears, κ_t is the number of network associated with the t[th] sub-sample and $t = 1,2,.., p$.

An unbiased estimator in equation (3) and its variance can be rewritten in terms of $\hat{\mu}_t$ as follows

$$\hat{\mu} = \frac{1}{p} \sum_{t=1}^{p} \hat{\mu}_t \tag{5}$$

$$V(\hat{\mu}) = \left(\frac{1}{q}\right) \frac{1}{p^2} \sum_{s=1}^{q} \left[\left(\sum_{t=1}^{p} \hat{\mu}_t \right)_s - p\mu \right]^2 \tag{6}$$

Method I: $p \geq 2$ and $m \geq 2$.

Based on equation (1), an estimator of variance is suggested to be

$$v_1(\hat{\mu}) = \left(\frac{q-1}{q} \right) \frac{1}{p} \frac{\sum_{t=1}^{p} (\hat{\mu}_t - \hat{\mu})^2}{p-1} \tag{7}$$

This estimator is biased (see appendix A).

Method II: $p \geq 2$ and $m = 2$.

Let $U_{2t} = \sum_{k \in \kappa_{2t}} \frac{y_k}{v_k x_k}$, where v_k is the number of times that the k[th] network appears and κ_{2t} is the number of networks associated with the (2t)th secondary unit.

Based on equation (2), an estimator of variance is suggested the to be

$$v_2(\hat{\mu}) = \left(\frac{q-1}{q}\right)\frac{1}{n}\sum_{t=1}^{p}\frac{(U_{2t}-U_{2t-1})^2}{n} \quad (8)$$

Note: $p = \frac{m}{2}$. This estimator is biased (see appendix B). The computation of estimators will be shown in the next part.

4. Examples

In this part, the bias and mean squared error (MSE) of each estimator were computed as well as the proportion of confidence intervals that contained the population mean. For confidence interval, consider here is 95% confidence interval. The computation of these was applied to a small population. The artificial populations of size $N = 12$ secondary units is used. The case of $q = 3$ and $q = 2$ primary units are considered. There are shown in figure 2A and 2B, respectively. The condition of interest for both cases is defined by $C = \{y_{ij} \geq 2\}$. In addition, a number appear in cell $i-j$ is the value of y_{ij}.

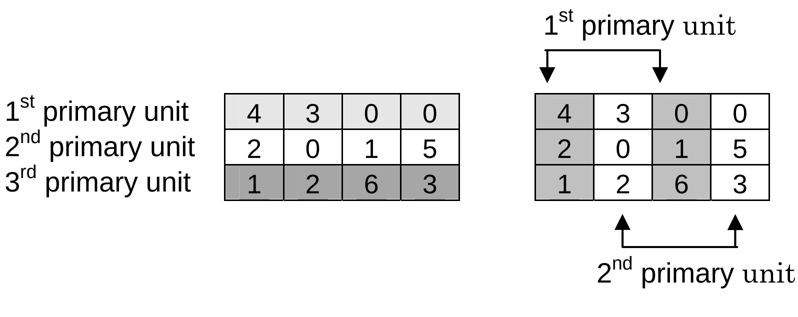

Figure 2: represent primary units and secondary units of both cases.

The population mean is μ = (4+3+0+0+2+0+1+5+1+2+6+3)/12 = 2.25.

For $q = 3$, table 1 shows the 3 possible sample mean $\hat{\mu}$ = 1.125, 3.375 and 2.250, respectively, and hence its expectation is 2.25 and the actual variance is 0.84375. Since an initial sample consists of 4 secondary units, the sub-sample can take on only one value, that is, $p = 2 = m$. All possible sub-samples and their statistics based on equation (5), (7) and (8) are shown in table 2. In addition, 95% confidence interval of $\hat{\mu}$ are also calculated from all cases under the assumption that the population is normally. When $q = 3$ and $p = 2$, estimator $v_1(\hat{\mu})$ seems to be the best choice in term of minimum bias, but the percentage of CI containing the population mean of $v_2(\hat{\mu})$ is higher than to the percentage of CI containing the population mean of $v_1(\hat{\mu})$.

Table 1: Sample means of all possible samples for adaptive cluster sampling with a single primary unit from the population in figure 2A ($q = 3$).

Sample	1				2				3			
Unit labels: (i, j)	**1,1**	**1,2**	**1,3**	**1,4**	**2,1**	**2,2**	**2,3**	**2,4**	**3,1**	**3,2**	**3,3**	**3,4**
Y_{ij}	4	3	0	0	2	0	1	5	1	2	6	3
y_k	9	9	0	0	9	0	1	16	1	16	16	16
x_k	2	2	1	1	2	1	1	2	1	2	2	2
network number	*1*	*1*	*2*	*3*	*1*	*4*	*5*	*6*	*7*	*6*	*6*	*6*
y_k / x_k	4.5	4.5	0	0	4.5	0	1	8	1	8	8	8
$\hat{\mu}$	**1.125**				**3.375**				**2.250**			

Table 2: All possible samples for $q = 3$ and $p = 2$ and the calculation of $v_1(\hat{\mu}_1)$ and $v_2(\hat{\mu}_1)$.

Primary unit	Sub-sample	$\hat{\mu}$	Method I			Method II		
			$v_1(\hat{\mu}_1)$	95% CI		$v_2(\hat{\mu}_1)$	95% CI	
				Lower	Upper		Lower	Upper
1	(4, 3) (0, 0)	1.125	0.8438	-0.675	2.926	0	na	na
(4, 3,0, 0)	(4, 0) (3, 0)	1.125	0.000	na	na	0.4219	-0.148	2.399
	(4, 0) (3, 0)	1.125	0.000	na	na	0.4219	-0.148	2.399
2	(2, 0) (1, 5)	3.375	0.8438	1.574	5.176	2.8854	0.045	6.705
(2, 0, 1, 5)	(2, 1) (0, 5)	3.375	0.2604	2.374	4..376	3.1771	-0.118	6.869
	(2, 5) (0, 1)	3.375	5.5104	-1.225	7.976	0.5521	1.918	4.833
3	(1, 2) (6, 3)	2.250	0.0417	1.849	2.651	0.1157	1.583	2.917
(1, 2, 6, 3)	(1, 6) (2, 3)	2.250	0.0417	1.849	2.651	0.1157	1.583	2.917
	(1, 3) (2, 3)	2.250	0.0417	1.849	2.651	0.1157	1.583	2.917
Expectation		**2.250**	**0.8426**			**0.8673**		
Bias		**0.000**	**-0.0012**			**0.0235**		
MSE		**0.8438**	**2.8302**			**1.3723**		
% of CI contains the true mean				**85.71**			**100.00**	

Note: na means is not available.

For $q = 2$, table 3 shows the 2 possible sample mean $\hat{\mu}$ = 2.417 and 2.083 and there are two possible values of p, that is, $p = 2,3$. When $p = 2$, it is impossible to obtain the second method variance since m are not be equal to 2. That is, only method I is considered and the result are shown in table 4. It can be seen that the true variance of the unbiased estimator of $\hat{\mu}$, is 0.0278. The bias of $v_1(\hat{\mu})$ is large relatively to its expectation. However, the MSE of $v_1(\hat{\mu})$ in this case is not too large compare with the case when $q = 3, p = 2$. For $q = 2$ and $p = 3$, both methods can be considered. The true variance of $\hat{\mu}$ is 0.0278 which is the same as in the case $q = 2, p = 2$. All possible sub-samples within the selected primary unit and their statistics are shown in tables 5. When $q = 2, p = 2$, both estimators have a large bias. However, the estimator $v_2(\hat{\mu})$ has a smaller bias than the estimator $v_1(\hat{\mu})$. Additionally, the percentage of CI containing the population mean of $v_2(\hat{\mu})$ is higher than the percentage of CI containing the population mean of $v_1(\hat{\mu})$. In additional, the MSE of estimators of both methods will decrease when q increases.

Table 3: Sample means of all possible samples for adaptive cluster sampling with a single primary unit from the population in figure 2B ($q = 2$).

Sample	1						2					
Unit labels: (i, j)	1,1	1,2	1,3	1,4	1,5	1,6	2,1	2,2	2,3	2,4	2,5	2,6
Y_{ij}	4	0	2	1	1	6	3	0	0	5	2	3
y_k	9	0	9	1	1	16	9	0	0	16	16	16
x_k	2	1	2	1	1	2	2	1	1	2	2	2
network number	*1*	*2*	*1*	*5*	*7*	*6*	*1*	*3*	*4*	*6*	*6*	*6*
y_k / x_k	4.5	0	4.5	1	1	8	4.5	0	0	8	8	8
$\hat{\mu}$	2.417						2.083					

Table 4: All possible samples for $q = 2, p = 2$ and the calculation of $v_1(\hat{\mu}_1)$.

Primary unit	Sub-sample	$\hat{\mu}$	Method I		
			$v_1(\hat{\mu}_1)$	95% CI	
				Lower	Upper
1 (4, 0, 2, 1, 1, 6)	(4, 0, 2) (1, 1, 6)	2.417	0.4201	1.146	3.688
	(4, 0, 1) (2, 1, 6)	2.417	0.8889	0.568	4.265
	(4, 0, 1) (2, 0, 6)	2.417	0.8889	0.568	4.265
	(4, 0, 6) (2, 1, 1)	2.417	0.5000	1.030	3.803
	(4, 2, 1) (0, 1, 6)	2.417	0.1701	1.608	3.226
	(4, 2, 1) (0, 1, 6)	2.417	0.1701	1.608	3.226
	(4, 2, 6) (0, 1, 1)	2.417	0.1531	-0.008	4.843
	(4, 1, 1) (0, 2, 6)	2.417	0.5000	1.030	3.803
	(4, 1, 6) (0, 2, 1)	2.417	0.8889	0.568	4.265
	(4, 1, 6) (0, 2, 1)	2.417	0.8889	0.568	4.265

Table 4: All possible samples for q=2 and p=2 and the calculation of $v_1(\hat{\mu}_1)$ (continued).

Primary unit	Sub-sample	$\hat{\mu}$	Method I		
			$v_1(\hat{\mu}_1)$	95% CI	
				Lower	Upper
2 (3, 0, 0, 5, 2, 3)	(3, 0, 0) (5, 2, 3)	2.083	0.1701	1.274	2.891
	(3, 0, 5) (0, 2, 3)	2.083	0.2813	1.043	3.123
	(3, 0, 2) (0, 5, 3)	2.083	0.2813	1.043	3.123
	(3, 0, 3) (0, 5, 2)	2.083	0.2813	1.043	3.123
	(3, 0, 5) (0, 2, 3)	2.083	0.2813	1.043	3.123
	(3, 0, 2) (0, 5, 3)	2.083	0.2813	1.043	3.123
	(3, 0, 3) (0, 5, 2)	2.083	0.2813	1.043	3.123
	(3, 5, 2) (0, 0, 3)	2.083	0.2813	1.043	3.123
	(3, 5, 3) (0, 0, 2)	2.083	0.2813	1.043	3.123
	(3, 2, 3) (0, 0, 5)	2.083	0.2813	1.043	3.123
Expectation		**2.250**	**0.4774**		
Bias		**0.000**	**0.4497**		
MSE		**0.0278**	**0.3233**		
% of CI contains the true mean				**100.00**	

Note: method II is not negligible.

Table 5: All possible samples for q=2 and p=3 and the calculation of $v_1(\hat{\mu}_1)$ and $v_2(\hat{\mu}_1)$.

Primary unit	Sub-sample	$\hat{\mu}$	Method I			Method II		
			$v_1(\hat{\mu}_1)$	95% CI		$v_2(\hat{\mu}_1)$	95% CI	
				Lower	Upper		Lower	Upper
1 (4,0,2,1,1,6)	(4, 0) (2, 1) (1, 6)	2.417	0.5530	0.959	3.875	0.7726	0.693	4.140
	(4, 0) (2, 1) (1, 6)	2.417	0.5530	0.959	3.875	0.7726	0.693	4.140
	(4, 0) (2, 6) (1, 1)	2.417	0.9175	0.539	4.295	0.5265	0.990	3.843
	(4, 2) (0, 1) (1, 6)	2.417	0.6701	0.812	4.022	0.6944	0.783	4.051
	(4, 2) (0, 1) (1, 6)	2.417	0.6701	0.812	4.022	0.6944	0.783	4.051

Table 5: All possible samples for q=2 and p=3 and the calculation of $v_1(\hat{\mu}_1)$ and $v_2(\hat{\mu}_1)$ (continued).

Primary unit	Sub-sample	$\hat{\mu}$	Method I			Method II		
			$v_1(\hat{\mu}_1)$	95% CI Lower	95% CI Upper	$v_2(\hat{\mu}_1)$	95% CI Lower	95% CI Upper
1	(4, 1) (0, 1) (2, 6)	2.417	0.9700	0.486	4.347	0.4948	1.038	3.800
	(4, 1) (0, 6) (2, 1)	2.417	0.3134	1.319	3.514	0.9323	0.524	4.310
	(4, 1) (0, 2) (1, 6)	2.417	0.5530	0.959	3.875	0.7726	0.693	4.140
	(4, 1) (0, 1) (2, 6)	2.417	0.9700	0.486	4.347	0.4948	1.038	3.796
	(4, 1) (0, 6) (2, 1)	2.417	0.3134	1.319	3.514	0.9323	0.524	4.310
	(4, 6) (0, 2) (1, 1)	2.417	0.9175	0.539	4.295	0.5295	0.990	3.843
	(4, 6) (0, 1) (2, 1)	2.417	0.9700	0.486	4.347	0.4948	1.038	3.800
	(4, 6) (0, 1) (2, 1)	2.417	0.9700	0.486	4.347	0.4948	1.038	3.800
2	(3, 0) (0, 5) (2, 3)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
(3,0,0,5,2,3)	(3, 0) (0, 2) (5, 3)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
	(3, 0) (0, 3) (5, 2)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
	(3, 0) (0, 5) (2, 3)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
	(3, 0) (0, 2) (5, 3)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
	(3, 0) (0, 3) (5, 2)	2.083	0.0035	1.967	2.200	0.3800	0.785	3.292
	(3, 5) (0, 0) (2, 3)	2.083	0.7535	0.382	3.785	0.0467	1.659	2.507
	(3, 5) (0, 2) (0, 3)	2.083	0.2813	1.043	3.123	0.2442	1.114	3.052

Table 5: All possible samples for q=2 and p=3 and the calculation of $v_1(\hat{\mu}_1)$ and $v_2(\hat{\mu}_1)$ (continued).

Primary unit	Sub-sample	$\hat{\mu}$	Method I			Method II		
			$v_1(\hat{\mu}_1)$	95% CI		$v_2(\hat{\mu}_1)$	95% CI	
				Lower	Upper		Lower	Upper
	(3, 2) (0, 5) (0, 3)	2.083	0.2813	1.043	3.123	0.2442	1.114	3.052
	(3, 2) (0, 3) (0, 5)	2.083	0.2813	1.043	3.123	0.2442	1.114	3.052
	(3, 3) (0, 0) (5, 2)	2.083	0.7535	0.382	3.785	0.0467	1.659	2.507
	(3, 3) (0, 5) (0, 2)	2.083	0.2813	1.043	3.123	0.2442	1.114	3.052
	(3, 3) (0, 2) (0, 5)	2.083	0.2813	1.043	3.123	0.2442	1.114	3.052
Expectation		**2.250**	**0.4747**			**0.4719**		
Bias		**0.000**	**0.4467**			**0.4441**		
MSE		**0.0278**	**0.2668**			**0.2429**		
% of CI contains the true mean				**80.00**			**100.00**	

5. Conclusion

From a numerical example, the mean of the sub-sample mean of both methods is equal to $\hat{\mu}$, as in equation (5) , which is an unbiased estimator of the population mean. The proposed variance estimators though are biased, the numerical example shows that the bias of the variance estimators of both methods may reduce when q and p increase. However, the bias, MSE and the proportion of confidence intervals that contain the population mean of these new variance estimators in this paper are preliminarily studied. Further study is needed the point of increasing q and p. Moreover, to investigate properties of these estimators, Monte-Carlo method can be used so that various situations can be compared. The other issue such as the “partially systematic sampling” [3] and [4] will be studied in the future.

Acknowledgements

I would like to thank Assistant Prof Dr Dryver for his suggestions.

References

[1] Thompson, S. K. 1991. Adaptive cluster sampling: Designs with primary and Secondary units. **Biometrics**. 47: 1103-1105.

[2] Wolter, K. M. 1984. An Investigation of Some Estimators of Variance for Systematic Sampling. **Journal of the American Statistical Association**. 79: 781-790.

[3] Leu and Tsui. 1996. New Partially Systematic Sampling. **Statistica Sinica**.6: 617-630.

[4] Zinger, A. 1980. Variance Estimation in Partially Systematic Sampling. **Journal of the American Statistical Association**. 75: 206-211.

Appendix A

$E(v_1(\hat{\mu}))$

Consider $\hat{\mu}-\mu \quad = \dfrac{1}{p}\sum_{t=1}^{p}(\hat{\mu}_t-\hat{\mu}+\hat{\mu}-\mu)$

$$(\hat{\mu}-\mu)^2 = \frac{1}{p^2}\left\{\sum_{t=1}^{p}\left((\hat{\mu}_t-\hat{\mu})^2+(\hat{\mu}-\mu)^2\right)+\sum_{t=1}^{p}\sum_{t\neq t'}^{p}(\hat{\mu}_t-\hat{\mu}+\hat{\mu}-\mu)(\hat{\mu}_{t'}-\hat{\mu}+\hat{\mu}-\mu)\right\}$$

$$= \frac{1}{p^2}\left\{\sum_{t=1}^{p}\left((\hat{\mu}_t-\hat{\mu})^2+(\hat{\mu}-\mu)^2\right)+\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)\right\} \quad (\because \sum_{t=1}^{p}(\hat{\mu}_t-\hat{\mu})(\hat{\mu}-\mu)=0)$$

$$\left(\frac{p-1}{p}\right)(\hat{\mu}-\mu)^2 = \frac{\sum_{t=1}^{p}(\hat{\mu}_t-\hat{\mu})^2}{p^2}+\frac{\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)}{p^2} \text{ (Subtract with } \frac{(\hat{\mu}-\mu)^2}{p}\text{)}$$

$$\left(\frac{q-1}{q}\right)(\hat{\mu}-\mu)^2 = \left(\frac{q-1}{q}\right)\frac{\sum_{t=1}^{p}(\hat{\mu}_t-\hat{\mu})^2}{p(p-1)}+\left(\frac{q-1}{q}\right)\frac{\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)}{p(p-1)} \text{ (Multiply with} \left(\frac{q-1}{q}\right)\Big/\left(\frac{p-1}{p}\right)\text{)}$$

Since $v_1(\hat{\mu})=\left(\dfrac{q-1}{q}\right)\dfrac{\sum_{t=1}^{p}(\hat{\mu}_t-\hat{\mu})^2}{p(p-1)}$ So $v_1(\hat{\mu})=\left(\dfrac{q-1}{q}\right)(\hat{\mu}-\mu)^2-\left(\dfrac{q-1}{q}\right)\dfrac{\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)}{p(p-1)}$

$$E(v_1(\hat{\mu}))=\left(\frac{q-1}{q}\right)E(\hat{\mu}-\mu)^2-\left(\frac{q-1}{q}\right)E\left[\frac{\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)}{p(p-1)}\right]$$

Since $E(\hat{\mu})=\mu$ so $E(\hat{\mu}-\mu)^2=V(\hat{\mu})$

$$\therefore\ E(v_1(\hat{\mu}))=\left(\frac{q-1}{q}\right)V(\hat{\mu})-\left(\frac{q-1}{q}\right)\sum_{s=1}^{q}\left(\frac{\sum_{t=1}^{p}\sum_{t'\neq t}^{p}(\hat{\mu}_t-\mu)(\hat{\mu}_{t'}-\mu)}{p(p-1)}\right)_s\left(\frac{1}{q}\right) \neq V(\hat{\mu})$$

Appendix B

$E(v_2(\hat{\mu}))$.

$\because\ \hat{\mu}$ can be rewritten in terms of $\overline{U}_t$ as follows $\hat{\mu}=\dfrac{1}{n/2}\sum_{t=1}^{p}\left(\dfrac{U_{2t}+U_{2t-1}}{2}\right)=\dfrac{1}{n/2}\sum_{t=1}^{p}\overline{U}_t$, where $p=n/2$.

Consider $\hat{\mu}-\mu \quad = \dfrac{1}{n/2}\sum_{t=1}^{p}\left(\left(\dfrac{U_{2t}-U_{2t-1}}{2}\right)-\hat{\mu}+\hat{\mu}-\mu\right)$

$$(\hat{\mu}-\mu)^2=\left\{\frac{1}{n/2}\sum_{t=1}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\hat{\mu}+\hat{\mu}-\mu\right)\right\}^2.$$

$$=\frac{4}{n^2}\left\{\sum_{t=1}^{p}\left(\left[\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\hat{\mu}\right]+(\hat{\mu}-\mu)\right)^2+2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\hat{\mu}+\hat{\mu}-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\hat{\mu}+\hat{\mu}-\mu\right)\right\}$$

$$=\frac{4}{n^2}\left\{\sum_{t=1}^{p}\left(\frac{U_{2t}-U_{2t-1}}{2}\right)^2+\sum_{t=1}^{p}U_{2t}U_{2t-1}-2\mu\left(\sum_{t=1}^{p}\left(\frac{U_{2t}+U_{2t-1}}{2}\right)\right)+\frac{n}{2}\mu^2+2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right\}$$

$$=\frac{4}{n^2}\left\{\sum_{t=1}^{p}\left(\frac{U_{2t}-U_{2t-1}}{2}\right)^2+\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)+2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right\}$$

$$=\frac{1}{n^2}\sum_{t=1}^{p}(U_{2t}-U_{2t-1})^2+\frac{4}{n^2}\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)+\frac{4}{n^2}2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)$$

$$\left(\frac{q-1}{q}\right)(\hat{\mu}-\mu)^2=\left(\frac{q-1}{q}\right)\frac{1}{n^2}\sum_{t=1}^{p}(U_{2t}-U_{2t-1})^2$$
$$+\left(\frac{q-1}{q}\right)\frac{4}{n^2}\left\{\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)+2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right\}$$

Since $v_2(\hat{\mu})=\left(\frac{q-1}{q}\right)\frac{\sum_{t=1}^{p}(U_{2t}-U_{2t-1})^2}{n^2}$

$$v_2(\hat{\mu})=\left(\frac{q-1}{q}\right)(\hat{\mu}-\mu)^2-\left(\frac{q-1}{q}\right)\frac{4}{n^2}\left\{\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)+2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right\}$$

$$E(v_2(\hat{\mu}))=\left(\frac{q-1}{q}\right)E(\hat{\mu}-\mu)^2-E\left\{\left(\frac{q-1}{q}\right)\frac{4}{n^2}\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)\right\}$$
$$+E\left\{\left(\frac{q-1}{q}\right)\frac{4}{n^2}2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right\}$$

. Since $E(\hat{\mu})=\mu$ so $E(\hat{\mu}-\mu)^2=V(\hat{\mu})$

$$\therefore E(v_2(\hat{\mu}))=\left(\frac{q-1}{q}\right)V(\hat{\mu})-\sum_{s=1}^{q}\left(\left(\frac{q-1}{q}\right)\frac{4}{n^2}\sum_{t=1}^{p}(U_{2t}-\mu)(U_{2t-1}-\mu)\right)_s\frac{1}{q}$$
$$+\sum_{s=1}^{q}\left(\left(\frac{q-1}{q}\right)\frac{4}{n^2}2\sum_{t<t'}^{p}\left(\left(\frac{U_{2t}+U_{2t-1}}{2}\right)-\mu\right)\left(\left(\frac{U_{2t'}+U_{2t'-1}}{2}\right)-\mu\right)\right)_s\frac{1}{q}$$
$$\neq V(\hat{\mu})$$