Group Balanced Repeated Replication Variance Estimation in Stratified Adaptive Cluster Sampling

Nipaporn Pochai[1] and Arthur L. Dryver[2]

[1] Department of Mathematics, Mahasarakham University, Thailand
(j3032024@yahoo.com)
[2] School of Applied Statistics, National Institute of Development Administration
(dryver@gmail.com)

Abstract

The group balanced repeated replication (GBRR) method is used in stratified sampling for variance estimation. This method is applied in stratified adaptive cluster sampling when ignores crossover between strata. The another one method is repeatedly GBRR, which involves independently repeating the random grouping T times and then taking the average of the resulting T of GBRR variance estimators. The network sample in each stratum is divided at random into two groups and then the group balanced repeated replication method is applied to the groups. The modified plus estimator in stratified adaptive cluster sampling is studied and compared with the group balanced repeated replication (GBRR) method for variance estimation by simulation study.

Keywords: Group balanced repeated replication; Stratified sampling; Adaptive cluster sampling; Plus estimator

1. Introduction

Adaptive cluster sampling, proposed by Thompson (1990), is an efficient method for sampling rare and hidden clustered populations. In adaptive cluster sampling, an initial sample of units is selected by simple random sampling. If the value of the variable of interest from a sampled unit satisfies a pre-specified condition C, that is $\{i, y_i \ge c\}$, then the unit's neighborhood will also be added to the sample. If any other units that were "adaptively" added also satisfy the condition C, then their neighborhoods are also added to the sample. This process is continued until no more units that satisfy the condition are found. The set of all units selected and all neighboring units that satisfy the condition is called a network. The adaptive sample units did not satisfy the condition called edge units. A network and its associated edge units are called a cluster. If a unit is selected in the initial sample that does not satisfy the condition C, then there is only one unit in the network.

A neighborhood must be defined such that if unit *i* is in the neighborhood of unit *j* then unit *j* is in the neighborhood of unit *i*. An example of a neighborhood of a unit is defined as the four spatially adjacent units, that is to the left, right, top and bottom (north, south, east and west) of that unit. See Fig 1.

0	0	0	0	0
0	7	0	0	0
0	0	2	0	0
0	2	4	2	0
0	1	5 *	3	0
0	0	0	0	9
0	0	0	0	0

Fig. 1 A unit neighborhood is defined as four spatially adjacent units

Fig.1. illustrates the example of a unit neighborhood. The unit with a star is the initial unit selected. The condition to adaptively add units is a values greater than or equal to 1, units that are to the left, right, top, bottom of one another make up a neighborhood. The unit in the gray shading from a single network. The units in bold numbers are edge units of the network. The network and its edge units make up a cluster

For many populations, prior information exists about the units that are similar. These populations are grouped into strata so that stratification can be used to reduce the variability in the estimators. Adaptive cluster sampling that can be applied to stratified sampling is called stratified adaptive cluster sampling (Thompson, 1991). The estimators of variance for some estimators in stratified adaptive cluster sampling are complicated to compute. The group balanced repeated replication method (Rao, 1996) is easy for computing the estimate variance.

The group balanced repeated replication (GBRR) method or balanced half-sample method of variance estimation proposed by McCarthy (Kish and Frankel ,1970; Valliant, 1987). For this method the sample in each stratum is divided at random into two groups and then the group balanced repeated replication method is applied to the groups. Rao and Shao (1996) proposed the repeatedly group balanced repeated replication method of variance estimation, which involves independently repeating the random grouping T times and then taking the average of the resulting T of GBRR variance estimators.

In this paper, we will apply the GBRR method in stratified adaptive cluster sampling, which the initial sample of units size n_h is selected by simple random sampling without replacement. The network sample in each stratum is divided at random into two groups and then the group balanced repeated replication method is applied to the groups. The unbiased estimator of population total is form by the modified plus estimator (based on Dryver and Thompson 2005) because of the variance and estimator of variance of this estimator are complicated form such that this estimator is described in section 2. The GBRR method and repeatedly GBRR method are described in section 3. In the last section the simulation are studied about comparing the relative bias of variance and variance of variance of GBRR method.

2. Modified Plus Estimator

For stratified adaptive cluster sampling, the population consists of N units is partitioned into L strata based on prior information about units that are to be similar and assume that the population ignores crossover between strata. The population in each stratum consists of N_h units $(h = 1, 2, \ldots, L)$. The total of the y-values in stratum h is τ_h then the total for the whole population is $\tau = \sum_{h=1}^{L} \tau_h$. In each stratum h, the units are selected in the initial sample by simple random sampling without replacement. Define w_{hi} be the total of y-values of the network i in stratum h divided by the network-stratum size, that is

$$w_{hi} = \frac{y_{hi.}}{m_{hi}} \tag{1}$$

where $y_{hi.}$ be the total of y-values in the network i with stratum h

m_{hi} be the number of the units in the network i with stratum h

The unbiased estimator of the population total is

$$\begin{aligned}\hat{\tau}_{st} &= \sum_{h=1}^{L} \hat{\tau}_h \\ &= \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} w_{hi} \end{aligned} \tag{2}$$

A modified plus estimator (based on Dryver and Thompson, 2005) based on Hansen-Hurwitz estimator by applying the Rao-Blackwell theorem are as follows.

The final sample $s = \bigcup_{h=1}^{L} s_h$ can be partitioned into two parts. First, the core part $s_c = \bigcup_{h=1}^{L} s_{hc}$ is the set of all distinct units in the sample for which the condition $y_{hi} \geq c$ is satisfied. The second part $\overline{s}_c = \bigcup_{h=1}^{L} \overline{s}_{hc}$ consists of all the distinct units in the sample for which $y_{hi} < c$.

For unit *i* in stratum h, let f_{hi} be the number of units in the initial sample of stratum *h* that are in the network to which unit i belongs.

Let the statistic d^{+} be defined as

$$d^{+} = \left\{ \left(h, i, y_{hi}, f_{hi}\right) : \left(h, i\right) \in s_{hc}, \left(h, j, y_{hj}\right) : \left(h, j\right) \in \overline{s}_{hc} \right\}$$

Let D^{+} denote a random variable that takes on the possible value of d^{+}. Also let $\mathcal{D}^{+}$ denote the sample space for d^{+}.

For $\left(h, i\right) \in s_{h}$ define the indicator e_{hi} as

$$e_{hi} = \begin{cases} 1 \;\; ;\text{if } \; y_{hi} < c \text{ and } \; i \; \text{ is in the neighborhood of some } \left(h, j\right) \in s_{c} \\ 0 \;\; ;\text{otherwise} \end{cases}$$

Thus in stratum h, e_{hi} =1 if unit *i* is an edge unit and the network that makes it an edge unit is selected in the initial sample.

The number of sample edge units in stratum h is

$$e_{hs} = \sum_{i \in s_{h}} e_{hi} \; . \tag{3}$$

The number of sample edge units picked in the initial sample s_{h0} of stratum h is

$$e_{hs_{0}} = \sum_{i=1}^{n_{h}} e_{hi} = \sum_{i \in s_{h0}} e_{hi} \; . \tag{4}$$

The average y-value for the sample edge units in the final sample of stratum h is

$$\overline{y}_{he} = \frac{\sum\limits_{i \in s_{h}} e_{hi} y_{hi}}{e_{hs}} \quad . \tag{5}$$

For unit i in the sample of stratum h, define a new variable w_{hi}^{+} by

$$w_{hi}^{+} = w_{hi}\left(1 - e_{hi}\right) + \overline{y}_{he} e_{hi} \quad . \tag{6}$$

The new estimator $\hat{\tau}_{st\_DT}^{+}$, applying the Rao-Blackwell theorem, is defined by

$$\hat{\tau}_{st\_DT}^{+} = E\left[\hat{\tau}_{st} \mid D^{+} = d^{+} \right]$$

$$= \sum_{h=1}^{L} \frac{N_{h}}{n_{h}} \sum_{i=1}^{n_{h}} w_{hi}^{+} \quad . \tag{7}$$

Since the initial sample determines the final sample and every value of the statistic d^{+}, let $f\left(s_{0}^{+}\right)$ denote the function that maps an initial sample into a value of d^{+} resulting from its selection. For any two values of s_{0}^{+} and d^{+} let

$$I\left(s_{0}^{+}, d^{+}\right) = \begin{cases} 1 \;\; ;\text{if } \; f\left(s_{0}^{+}\right) = d^{+} \\ 0 \;\; ;\text{otherwise} \end{cases} \tag{8}$$

Let $L\left(d^{+}\right)$ be the number of initial samples compatible with d^{+} and $P\left(d^{+}\right)$ be the probability that $D^{+}=d^{+}$.

Let $\mathcal{S}^{+}$ be the sample space containing all possible initial samples.

The variance of $\hat{\tau}^{+}_{st\_DT}$ (based on Dryver and Thompson, 2005) is

$$V\left(\hat{\tau}^{+}_{st\_DT}\right)=\sum_{h=1}^{L}\left[\frac{N_h\left(N_h-n_h\right)}{n_h\left(N_h-1\right)}\sum_{i=1}^{N_h}\left(w_{hi}-\frac{\tau_h}{N_h}\right)^2\right]$$
$$-\sum_{h=1}^{L}\left[\frac{N_h^2}{n_h^2}\sum_{d^{+}\in\mathfrak{D}^{+}}\frac{P\left(d^{+}\right)}{L\left(d^{+}\right)}\sum_{s_0^{+}\in\mathcal{S}^{+}}I\left(s_0^{+},d^{+}\right)\left(\sum_{(h,i)\in s_0^{+},e_{hi=1}}y_{hi}-e_{hs_0}\overline{y}_{he}\right)^2\right]. \qquad (9)$$

An unbiased estimator of $V\left(\hat{\tau}^{+}_{st\_DT}\right)$ is

$$\hat{V}\left(\hat{\tau}^{+}_{st\_DT}\right)=\sum_{h=1}^{L}\left[\frac{1}{L}\sum_{s_0^{+}\in\mathcal{S}^{+}}I\left(s_0^{+},d^{+}\right)\frac{N_h\left(N_h-n_h\right)}{n_h\left(n_h-1\right)}\sum_{i=1}^{n_h}\left(w_{hi}-\frac{\hat{\tau}_h}{N_h}\right)^2\right]$$
$$-\sum_{h=1}^{L}\left[\frac{N_h^2}{Ln_h^2}\sum_{s_0^{+}\in\mathcal{S}^{+}}I\left(s_0^{+},d^{+}\right)\left(\sum_{(h,i)\in s_0^{+},e_{hi=1}}y_{hi}-e_{hs_0}\overline{y}_{he}\right)^2\right], \qquad (10)$$

where $L=\sum_{s_0^{+}\in\mathcal{S}^{+}}I\left(s_0^{+},d^{+}\right)$.

An unbiased easy-to-compute estimator of the variance of $\hat{\tau}^{+}_{st\_DT}$ is given by

$$\tilde{V}\left(\hat{\tau}^{+}_{st\_DT}\right)=\sum_{h=1}^{L}\left[\frac{N_h\left(N_h-n_h\right)}{n_h\left(n_h-1\right)}\sum_{i=1}^{n_h}\left(w_{hi}-\frac{\hat{\tau}_h}{N_h}\right)^2-\left(\hat{\tau}_{st}-\hat{\tau}^{+}_{st\_DT}\right)^2\right]. \qquad (11)$$

The form of an unbiased estimator of the variance of $\hat{\tau}^{+}_{st\_DT}$ is difficult to compute so the group balanced repeated replication (GBRR) method is very useful for computing the variance estimation in stratified adaptive cluster sampling whereas the estimator of the population total is the same of modified plus estimator.

In the GBRR method, the sample of the network in stratum h is divided at random into two groups containing $g_{h1}=\left[n_h/2\right]$ and $g_{h2}=n_h-\left[n_h/2\right]$ networks. The GBRR method employs the variability among R replicated half –samples that are selected in a balanced way to estimate the variance of $\hat{\tau}^{+}_{st\_DT}$

The half sample r can be defined by a vector $\boldsymbol{\alpha}_r=\left(\alpha_{r1},\alpha_{r2},\ldots,\alpha_{rL}\right)$: let

$$w_h^{+}\left(\boldsymbol{\alpha}_r\right)=\begin{cases}2\sum_{i\in g_{h1}}\dfrac{w_{hi}^{+}}{n_h} & \text{, if } \alpha_{rh}=+1\\ 2\sum_{i\in g_{h2}}\dfrac{w_{hi}^{+}}{n_h} & \text{, if } \alpha_{rh}=-1\end{cases} \qquad (12)$$

The set of R replicate half sample is balanced if

$$\sum_{r=1}^{R}\alpha_{rh}\alpha_{rj}=0 \qquad \text{for all } h\neq j \ ,$$

where $L\leq R\leq L+3$ (Shao and Tu, 1995).

Let $\hat{\tau}^{+}_{st\_DT}(\alpha_r)$ be the estimate of interest, by using only the observation in the half sample selected by α_r.

$$\hat{\tau}^{+}_{st\_DT}(\alpha_r)=\sum_{h=1}^{L}N_h\sqrt{1-(n_h/N_h)}\;w^{+}_{h}(\alpha_r) \qquad (13)$$

The GBRR variance estimator is

$$\hat{V}_{GBBR}\left(\hat{\tau}^{+}_{st}\right)=\frac{1}{R}\sum_{r=1}^{R}\left[\hat{\tau}^{+}_{st}(\alpha_r)-\left(\hat{\tau}^{+}_{st}\right)_m\right]^2, \qquad (14)$$

where $\left(\hat{\tau}^{+}_{st}\right)_m=\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\tau}^{+}_{st}(\alpha_r)\right)_r$.

Another method of GBRR is repeatedly group balanced repeated replication, RGBRR, variance estimator, which involves independently repeating the random grouping T times and then taking the average of the resulting T GBRR variance estimator (Rao and Shao, 1996).

The RGBRR variance estimator is

$$\hat{V}_{RGBRR}\left(\hat{\tau}^{+}_{st}\right)=\frac{1}{T}\sum_{t=1}^{T}\left[v_{GBRR}\left(\hat{\tau}^{+}_{st}\right)_t\right]. \qquad (15)$$

The GBRR variance estimation and the RGBRR variance estimation are bias estimators of the true variance.

3. Simulation Data

In some studies there is more than one variable of interest. In terms of data example from Dryver and Thompson (2005), of primary interest might be the estimation of the one species, duck, but also important could be the estimation of another species, snakes. Fig. 2 consists of real data, the number of ducks at given sectors (Smith et al., 1995), and Fig.3 consists of simulated data representing snakes. The modified plus estimator provides very little improvement for estimation on the ducks, but large improvement on estimation of the snakes (Dryver and Thompson, 2005).

0	0	3	5	0	0	0	0	0	0
0	0	0	24	14	0	0	10	103	0
0	0	0	0	2	3	2	0	13639	1
0	0	0	0	0	0	0	0	14	122
0	0	0	0	0	0	2	0	0	177

Fig. 2 Blue-winged teal data

From the blue-winged teal data size 5 x 10 = 50 square meter, to divide this data into 2 strata and equal size in each stratum that is in each stratum the population size 5 x 5 = 25 units.

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	96
0	0	116	0	0	0	0	0	0	89
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	103	0

Each unit is Poisson distributed with a mean of 100, with a Bernoulli probability 0.1; otherwise it is 0.

Fig. 3 The simulated snake data

From the simulated snake data size 5 x 10 = 50 square meter, to divide this data into 2 strata and equal sizes in each stratum that is in each stratum the population size 5 x 5 = 25 units, so that the total of snake data is 404.

For each estimator 50,000 iterations were performed because the estimate was consistent. The condition is defined by $C = \{y : y \geq 1\}$ for the blue-wing teal data, and the estimate is on snakes. The GBRR and RGBRR variance estimates were computed from a 4×4 matrix; using 2 columns;

		Stratum (h)	
		1	2
Half-sample (r)	α_1	-1	-1
	α_2	+1	-1
	α_3	-1	+1
	α_4	+1	+1

We have use T=20 for the RGBRR variance estimator, lead to 80 half-samples. A varying initial sample size was used. The formula used to estimate the variance of the variance is

$$V\left[\tilde{V}_{\text{Plus}}\right] = \frac{1}{50,000-1}\sum_{i=1}^{50,000}\left(\tilde{V}\left(\hat{\tau}^{+}_{st\_DT}\right)_i - \frac{\sum_{i=1}^{50,000}\tilde{V}\left(\hat{\tau}^{+}_{st\_DT}\right)_i}{50,000}\right)^2, \tag{16}$$

$$V\left[\hat{V}_{\text{GBRR}}\right] = \frac{1}{50,000-1}\sum_{i=1}^{50,000}\left(\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)_i - \frac{\sum_{i=1}^{50,000}\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)_i}{50,000}\right)^2 \tag{17}$$

and

$$V\left[\hat{V}_{\text{RGBRR}}\right] = \frac{1}{50,000-1}\sum_{i=1}^{50,000}\left(\hat{V}_{\text{RGBRR}}\left(\hat{\tau}^{+}_{st}\right)_i - \frac{\sum_{i=1}^{50,000}\hat{V}_{\text{RGBRR}}\left(\hat{\tau}^{+}_{st}\right)_i}{50,000}\right)^2 \tag{18}$$

The formula of relative bias (RB) of the variance estimator is

$$RB_{\text{GBRR}} = \left[\frac{\frac{1}{50,000}\sum_{i=1}^{50,000}\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)_i}{MSE\left(\hat{\tau}^{+}_{st\_DT}\right)}\right] - 1, \tag{19}$$

and

$$RB_{\text{GBRR}} = \left[\frac{\frac{1}{50,000}\sum_{i=1}^{50,000}\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)_i}{MSE\left(\hat{\tau}^{+}_{st\_DT}\right)}\right] - 1, \tag{20}$$

where $\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)$ and $\hat{V}_{\text{RGBRR}}\left(\hat{\tau}^{+}_{st}\right)$ are the biased estimator. $\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)$ and $\hat{V}_{\text{RGBRR}}\left(\hat{\tau}^{+}_{st}\right)$ are the estimate variance by the group balanced repeated replication method from (14) and (15) respectively, and $MSE\left(\hat{\tau}^{+}_{st\_DT}\right) = V\left(\hat{\tau}^{+}_{st\_DT}\right)$ is an unbiased variance estimator of the modified plus estimator.

The formula of coefficient of variation (CV) of the variance estimator is

$$CV_{\text{GBRR}} = \frac{\left[\frac{1}{50,000}\sum_{i=1}^{50,000}\left(\hat{V}_{\text{GBRR}}\left(\hat{\tau}^{+}_{st}\right)_i - MSE\left(\hat{\tau}^{+}_{st\_DT}\right)\right)^2\right]^{1/2}}{MSE\left(\hat{\tau}^{+}_{st\_DT}\right)}, \tag{21}$$

and $$\mathrm{CV}_{\mathrm{RGBRR}} = \frac{\left[\frac{1}{50,000}\sum_{i=1}^{50,000}\left(\hat{V}_{\mathrm{RGBRR}}\left(\hat{\tau}^{+}_{st}\right)_i - \mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right)\right)^2\right]^{1/2}}{\mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right)} , \qquad (22)$$

where $\mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right) = \frac{1}{50,000}\sum_{i=1}^{50,000}\left(\left(\hat{\tau}^{+}_{st\_DT}\right)_i - \tau_y\right)^2 .$

Table 1 Relative bias of variance and coefficient of variation by GBRR method

n_h	$\mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right)$	Group Balanced Repeated Replication (GBRR)		
		$\hat{V}_{\mathrm{GBRR}}\left(\hat{\tau}^{+}_{st\_DT}\right)$	$\mathrm{RB}_{\mathrm{GBRR}}$	$\mathrm{CV}_{\mathrm{GBRR}}$
2	290,758.41	308,998.82	0.063	2.28
5	78,651.66	83,032.76	0.068	1.65
8	33,939.42	36,580.88	0.048	1.34
10	23,935.45	24,053.36	0.005	1.10
12	17,602.06	16,794.01	-0.046	0.90
15	10,278.41	10,197.71	-0.008	0.74

Table 2 Relative bias of variance and coefficient of variation by RGBRR method

n_h	$\mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right)$	Repeatedly GBRR T = 20		
		$\hat{V}_{\mathrm{RGBRR}}\left(\hat{\tau}^{+}_{st\_DT}\right)$	$\mathrm{RB}_{\mathrm{RGBRR}}$	$\mathrm{CV}_{\mathrm{RGBRR}}$
2	290,758.41	308,998.82	0.063	2.28
5	78,651.66	84,047.64	0.069	1.58
8	33,939.42	35,428.24	0.044	1.28
10	23,935.45	24,020.70	0.004	1.06
12	17,602.06	16,804.73	-0.045	0.87
15	10,278.41	10,208.84	-0.007	0.72

From Table 1 and Table 2, it can be seen that the $\mathrm{RB}_{\mathrm{GBRR}}$ and $\mathrm{RB}_{\mathrm{RGBRR}}$ are less than seven percent points; that is, $\hat{V}_{\mathrm{GBRR}}\left(\hat{\tau}^{+}_{st\_DT}\right)$ and $\hat{V}_{\mathrm{RGBRR}}\left(\hat{\tau}^{+}_{st\_DT}\right)$ are close to the $\mathrm{MSE}\left(\hat{\tau}^{+}_{st\_DT}\right)$. The $\mathrm{CV}_{\mathrm{GBRR}}$ is larger than $\mathrm{CV}_{\mathrm{RGBRR}}$ when n_h is more than 2. The CV of both methods will decrease when n_h increases.

Table 3 The variance of the estimate variance of the proposed estimator

n_h	$V\left(\tilde{V}_{\mathrm{Plus}}\right)$	$V\left(\hat{V}_{\mathrm{GBRR}}\right)$	$V\left(\hat{V}_{\mathrm{RGBRR}}\right)$
2	441,402,152,503	437,493,680,223	437,493,680,223
5	21,222,041,691	16,860,785,034	15,440,331,400
8	3,962,967,425	2,079,035,950	1,883,004,003
10	1,685,679,647	687,076,476	646,302,902
12	851,715,306	249,448,570	235,956,492
15	301,480,440	58,161,050	54,476,776

Table 3 shows that the variance of the estimate variance will decrease when n_h increases. The variance of the estimate variance by GBRR and RGBRR method are less than the variance of the estimate variance of the modified plus estimator. The variance of the estimate variance by RGBRR method is less than the variance of the estimate variance by GBRR about 5% to 10% when n_h is more than 2.

4. Conclusions

The group balanced repeated replication (GBRR) variance estimator randomly to assign a sample into 2 groups within each stratum and selects balanced replications from the group. The GBRR method and RGBRR method are applied in stratified adaptive cluster sampling. These methods are useful for computing when the variance estimation is difficult to compute. The numerical study shows that the relative bias of the variance estimator by group balanced repeated replication, GBRR, and repeatedly group balanced repeated replication, RGBRR, is less than seven percent points, from the simulation. The coefficient of variation of the variance estimator by RGBRR is less than the coefficient of variation of the variance by GBRR. The CV of the GBRR method and the RGBRR method will decrease when the initial sample size increase. The dispersion of the estimate variance of the plus estimator is more than the dispersion of the estimate variance by GBRR and RGBRR. The dispersion of the estimate variance by RGBRR is the smallest.

5. Acknowledgements

I would like to profoundly thank Mr. Paveen Chutiman for his programming advice.

References

[1] Dryver, A.L. (1999). Adaptive Sampling Designs and Associated Estimators : Ph.D. Thesis, The Pennsylvania State University.

[2] Dryver, A.L., and Thompson, S.K. (2005). Improved unbiased estimators in adaptive cluster sampling: Journal of the Royal Statistical Society, Ser.B, 67, 157-166.

[3] Kish, L., and Frankel, M.R. (1970). Balanced Repeated Replication in Standard Error: Journal of the American Statistical Association 65, 1071-1094.

[4] Rao, J.N.K. and Shao, J.(1996). On Balanced Half-Sample Variance Estimation in stratified Random Sampling: Journal of the American Statistical Association 91, 343-348.

[5] Sarndal, C.E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling: New York: Springer-Verlag.

[6] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap: New York: Springer-Verlag.

[7] Smith, D.R.; Conroy, M.J. and Brakhage, D.H. 1995. Efficiency of Adaptive Cluster Sampling for Estimating Density Estimating Density Wintering Waterfowl. Biometrics. 51 (June): 777-788.

[8] Thompson, S.K. 1990. Adaptive Cluster Sampling. Journal of the American Statistical Association. 85 (December): 1050 -1059.

[9] Thompson, S.K. 1991. Stratified Adaptive Cluster Sampling. Biometrika. 78 (June): 389-397.

[10] Valliant, R. (1987). Some Prediction Properties of Balanced Half-Sample Variance Estimators in Single-Stage Sampling: Journal of the Royal Statistical Society, Ser.B 49, 68-81.