ADAPTIVE CLUSTER DOUBLE SAMPLING

Martin H. Felix-Medina
Pennsylvania State University and Universidad Autonoma de Sinaloa
Steven K. Thompson
Pennsylvania State University
Martin H. Felix-Medina, Department of Statistics, 326 Thomas Building,
University Park, PA 16802-2111 (martin@stat.psu.edu)

Key Words: clustered population; finite population; regression estimator; Horvitz-Thompson estimator.

1 Introduction

Adaptive cluster sampling (ACS) was introduced by Thompson (1990) as an efficient sampling procedure to estimate totals and means of rare and clustered populations. The idea behind this method is to take an initial sample by some ordinary sampling procedure, and then to increase the sample size by adding elements in the vicinity of the sampled units that satisfy a previously specified condition. Despite the appropriateness of this sampling design for sampling clustered populations, one of its main drawbacks is the lack of control of the final sample size. Several suggestions have been proposed to limit the final sample size. For instance, Brown (1994) has proposed an ACS version in which the initial sample is sequentially selected until the final sample size reaches a specified value or first exceeds that value. Although this design controls well the final sample size, efficient estimators have not been developed under this design. Salehi and Seber (1997) have proposed a two-stage ACS version in which the whole vicinity of a secondary sampling unit is restricted to lie within the primary sampling unit that contains the secondary unit. Despite the fact that efficient estimators of the population mean have been developed, the problem with this design is that large-size clusters are not completely sampled, but only their portion or portions that are intersected by the primary units in the initial sample are sampled. Therefore, in the case of populations with large-size clusters (with respect to the size of the primary units), a reduction in the efficiency of this sampling design should be expected.

In the context of conventional designs, double sampling designs have been used in situations where the survey variable is expensive or difficult-to-measure, and it is cheaper or easier-to-measure an auxiliary variable correlated with the survey variable. The strategy used in double sampling is to take a large-size initial sample and measure the auxiliary-variable values associated with the elements in that sample. Then, a small-size subsample is taken from the initial sample and the survey-variable values associated with the elements in the subsample are measured. Regression type of estimators are used to estimate the population mean because they allow one to take advantage of the relationship between the survey variable and the auxiliary one. One of the advantages of this design is that even with a few measurements of the survey variable it is possible to get good estimates of the population mean.

The goal of this paper is to develop a sampling design which combines the ideas used in ACS and those used in double sampling. This new sampling design, which we have called Adaptive Cluster Double Sampling (ACDS), is appropriate when the condition for additional sampling can be based on an easy-to-measure auxiliary variable x correlated with the survey variable y. For example, in some situations, even without measuring the actual y-value associated with a unit, it might be possible to decide whether or not that value satisfies the condition for additional sampling; hence, a potential auxiliary variable could be a binary variable which takes the value 1 if the y-value satisfies the condition for additional sampling and the value 0 otherwise. In other situations, the auxiliary variable could be defined as that whose values are "eyeball" estimates of the actual y-values, or that whose values are measurements of the y-values made with a not very accurate instrument. The idea behind ACDS is to model the relationship between y and x through a regression model and to take an ordinary adaptive cluster sample based on the auxiliary variable. Notice that the x-value associated with each element in this sample

Research funded by the NIH/NIDA, Grant R01 DA09872. The first author also acknowledges the support from CONACYT, Mexico.

is measured, and because the auxiliary variable is inexpensive and easy-to-measure, a relatively large-size initial sample could be employed to select the adaptive sample, which would increase the probability of detecting clusters with elements satisfying the condition. Then, using the network structure of the adaptive cluster sample, a subsample of networks is taken by means of a conventional design. Finally, from each network in the subsample, a sample of units is taken, using a conventional design, and their y-values are measured. Despite the fact that this design would not allow one to control the number of measurements of the auxiliary variable, it would allow one to take a previously specified number of measurements of the survey variable. Regression type of estimators could be used to estimate the population mean, which can be based on one of the estimators proposed by Thompson (1990), a Horvitz-Thompson type of estimator (HTE) and a Hansen-Hurwitz type of estimator (HHE).

In this paper, we will focus on the construction of a regression estimator based on the HTE, and on another regression estimator constructed by applying the estimating function approach to the HTE. We will also present asymptotic expressions for the variances of the regression estimators, and estimators of those variances obtained by using the Delta method. In the last part of this paper we will present the results of a simulation study carried out to compare the ACS design with the ACDS procedure.

2 Notation and design

Let $U = \{u_1, ..., u_N\}$ be a finite population of size N. Let us denote by y and x the survey variable and the auxiliary variable, respectively. Similarly, let y_i and x_i be the values of y and x associated with the element u_i, $i = 1, \ldots, N$. We will assume that the relationship between y and x can be modeled through a stochastic regression model ξ with mean $\mathbf{E}_\xi(y_i|\mathbf{x}_i) = \mathbf{x}_i^t\beta$ and variance $\mathbf{V}_\xi(y_i|\mathbf{x}_i) = v_i\sigma^2$, $i = 1, \ldots, N$, where $\mathbf{x}_i$ is a vector in $\mathbf{R}^p$, whose elements are functions of the auxiliary variable x_i [for example if $x_i \in \mathbf{R}$, then we might have $\mathbf{x}_i = (1, x_i, x_i^2)^t$], and $v_i = \varphi(\mathbf{x}_i)$, where the function φ is assumed to be known. Let $\mathbf{y_U} = (\mathbf{y_1}, ..., \mathbf{y_N}) \in \mathbf{R^N}$ and $\mathbf{x_U} \in \mathbf{R^N} \times \mathbf{R^p}$ be the population vector of the y-values and the population matrix of the $\mathbf{x}$-values, respectively. Let $\mu_y = \sum_{i=1}^N y_i/N$ and $\boldsymbol{\mu}_\mathbf{x} = \sum_{i=1}^N \mathbf{x}_i/N$ be the population means of the y-values and $\mathbf{x}$-values, respectively. Throughout this paper, the goal is to estimate μ_y; however, because of the assumed regression model, we will also require to estimate the finite population regression parameter $\mathbf{B_U} = (\mathbf{x_U}^t\mathbf{v_U}^{-1}\mathbf{x_U})^{-1}\mathbf{x_U}^t\mathbf{v_U}^{-1}\mathbf{y_U}$, where $\mathbf{v_U}$ is a diagonal matrix whose elements are the variances v_i, $i = 1, \ldots, N$. Notice that $\mathbf{B_U}$ can be thought as the finite-population version of the regression parameter β.

The first phase of an ACDS procedure consists of taking an ordinary adaptive cluster sample S_1 based on the values of the auxiliary variable. To do this, we first define both a condition C_x for additional sampling and a set of neighboring units for each unit $u_i \in U$. Next, by using an ordinary sampling procedure, we select an initial sample S_0 of n units from U. We observe the x-values associated with the units in S_0, and we add to the sample the neighboring units of every unit in S_0 that satisfies C_x. We repeat this procedure with the new added units, and we stop when no new added unit satisfies C_x. The set formed by an original unit $u_i \in S_0$ and the units added as a consequence of including u_i in S_0 is called a cluster. A cluster minus its edge units (units which do not satisfy C) is called a network, as does any set formed by a single unit which does not satisfy C_x. The definition of C_x and that of neighborhood give rise to a partition of U into K networks $A_1, ..., A_K$, and from among these networks, the initial sample S_0 intersects k different of them.

The second phase of an ACDS procedure consists of selecting a conventional sample S_2 of k_1 ($k_1 \leq k$) networks, $A_1, \ldots, A_{k_1}$, from the k different networks intersected by S_0. Finally, the third phase consists of taking a conventional subsample of units from each network in S_2, and recording the y-value associated with every unit in those subsamples. Here, we will assume that the k_1 subsamples are independently selected.

The previous description of the ACDS procedure suggests that the second and third phases are carried out once the first phase has been completed, and therefore, that the subsampled networks are visited two times (one on the first phase, and another one on the second phase). However, if we decided to subsample every network intersected by S_0, then we could subsample a network as soon as we know which units belong to that network, and consequently, each network would be visited only one time. Nevertheless, the procedure consisting of subsampling after the first phase has been completed allows the researcher a better control of the size of the subsample.

To conclude this section we will introduce the following notation. We will denote by m_i the size of the network A_i, and by Y_i and $\mathbf{X}_i$ the sums of the y-values and x-values in A_i, that is, $Y_i = \sum_{u_j \in A_i} y_j$ and $\mathbf{X}_i = \sum_{u_j \in A_i} \mathbf{x}_j$, $i = 1, \ldots, K$. The subsample taken from A_i will be denoted by S_{3i}, and by m_i' ($m_i' \leq m_i$) we will denote the size of S_{3i}, $i = 1, \ldots, k_1$. Finally, the set of units whose y-values are measured will be denoted by S_3, that is, $S_3 = \cup_{i=1}^{k_1} S_{3i}$

3 Regression estimators of the population mean

To estimate the population mean μ_y, we propose to use regression-type estimators. As was indicated previously, different types of regression estimators can be constructed depending on whether we want to use HTEs or HHEs. In both cases, the general form of the regression estimator is

$$\hat{\mu}_R = \hat{\mu}_y^{'''} + (\hat{\boldsymbol{\mu}}_{\mathbf{x}} - \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''})^t \hat{\mathbf{B}}_{\mathbf{s_3}}, \qquad (1)$$

where $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ is an estimator of $\boldsymbol{\mu}_{\mathbf{x}}$ computed from the elements in the adaptive cluster sample S_1; $\hat{\mu}_y^{'''}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''}$ are estimators of μ_y and $\boldsymbol{\mu}_{\mathbf{x}}$, respectively, computed from the elements in S_3; and $\hat{\mathbf{B}}_{\mathbf{S_3}}$ is an estimator of $\mathbf{B}_{\mathrm{U}}$ computed from the elements in S_3.

In the case of the regression estimator based on the HTE, $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ is given by

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{k}\frac{1}{\pi_i}\mathbf{X}_i,$$

where $\mathbf{X}_i = \sum_{u_j \in A_i} \mathbf{x}_j$, and π_i is the probability of intersecting A_i by S_0, $i = 1, \ldots, K$. For example, if S_0 were a simple random sample without replacement (SRSWOR), π_i would be given by $\pi_i = 1 - \binom{N-m_i}{n}/\binom{N}{n}$. To construct the estimators $\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''}$ and $\hat{\mu}_y^{'''}$, we first need to estimate $\mathbf{X}_i$ and Y_i from the observations in S_{3i}, $i = 1, \ldots, k_1$. To estimate these quantities, we need to take into account the sampling design used to select S_{3i}. For instance, if S_{3i} were a simple random sample without replacement (SRSWOR), an estimator of Y_i could be $\hat{Y}_i = (m_i/m_i')\sum_{u_j \in S_{3i}} y_j$. Here, we will assume that appropriate estimators $\hat{Y}_i$ and $\hat{\mathbf{X}}_i$ of Y_i and $\mathbf{X}_i$ have been constructed. Then, HTEs of $\hat{\mu}_y^{'''}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''}$, based on the sample S_3, are

$$\hat{\mu}_y^{'''} = \frac{1}{N}\sum_{i=1}^{k_1}\frac{\hat{Y}_i}{\pi_i\pi_{i|s_1}} \quad \text{and} \quad \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''} = \frac{1}{N}\sum_{i=1}^{k_1}\frac{1}{\pi_i\pi_{i|s_1}}\hat{\mathbf{X}}_i,$$

where $\pi_{i|s_1}$ is the conditional probability, given S_1, of including the network A_i, $i = 1, \ldots, k_1$, in the second stage sample S_2. Finally, an estimator $\hat{\mathbf{B}}_{\mathbf{S_3}}$ of $\mathbf{B}_{\mathrm{U}}$, constructed from the elements in S_3, is

$$\hat{\mathbf{B}}_{\mathbf{S_3}} = \left[\sum_{i=1}^{k_1}\frac{1}{\pi_i\pi_{i|s_1}}(\widehat{\mathbf{XX}^t})_i\right]^{-1}\sum_{i=1}^{k_1}\frac{1}{\pi_i\pi_{i|s_1}}(\widehat{\mathbf{X}Y})_i,$$

where $(\widehat{\mathbf{XX}^t})_i$ and $(\widehat{\mathbf{X}Y})_i$ are estimators of $\sum_{u_j \in A_i}\frac{1}{v_j}\mathbf{x}_j\mathbf{x}_j^t$ and $\sum_{u_j \in A_i}\frac{1}{v_j}\mathbf{x}_j\mathbf{y}_j$, respectively.

By using the formula for the variance of a three-phase-sampling estimator (see Cochran, 1977, p. 276), we have that

$$\begin{aligned}\mathbf{V}(\hat{\mu}_R) &= \mathbf{V}_1\{\mathbf{E}_2[\mathbf{E}_3(\hat{\mu}_R - \mu_y)]\} \\ &\quad + \mathbf{E}_1\{\mathbf{V}_2[\mathbf{E}_3(\hat{\mu}_R - \mu_y)]\} \\ &\quad + \mathbf{E}_1\{\mathbf{E}_2[\mathbf{V}_3(\hat{\mu}_R - \mu_y)]\} \\ &= \mathbf{V}_{FP} + \mathbf{V}_{SP} + \mathbf{V}_{TP},\end{aligned}$$

where the subscript i of the expectation (variance) operator indicates that the expectation (variance) is taken over all the i-phase sample selections, and $\mathbf{V}_{FP}$, $\mathbf{V}_{SP}$, and $\mathbf{V}_{TP}$ are measures of the variability corresponding to the first, second and third-phase sample selections, respectively.

By using the Delta method, and a similar strategy to that used by Särndal et al. (1992, Section 9.7) we get the following asymptotic approximation to $\mathbf{V}_{FP}$:

$$\mathbf{V}_{FP} \approx \frac{1}{N^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}Y_iY_j,$$

where $\pi_{ii} = \pi_i$, $i = 1, \ldots, K$. Similarly, an asymptotic approximation to $\mathbf{V}_{SP}$ is

$$\mathbf{V}_{SP} \approx \mathbf{E}_1\left[\frac{1}{N^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\pi_{ij|s_1} - \pi_{i|s_1}\pi_{j|s_1}}{\pi_{i|s_1}\pi_{j|s_1}}\frac{E_i}{\pi_i}\frac{E_j}{\pi_j}\right],$$

where $E_i = \sum_{u_j \in A_i} e_j$, $e_j = y_j - \mathbf{x}_j^t\mathbf{B}_{\mathrm{U}}$ is the population regression-error associated with $u_j \in A_i$, $\pi_{ij|s_1}$ is the conditional probability, given S_1, of including the networks A_i and A_j, $i \neq j$, in the second-stage sample S_2, and $\pi_{ii|s_1} = \pi_{i|s_1}$, $i = 1, \ldots, k$. Finally, if the estimators $\hat{Y}_i$ and $\hat{\mathbf{X}}_i$ used in the third-stage sampling are HTEs, then an approximation to $\mathbf{V}_{TP}$ is

$$\mathbf{V}_{TP} \approx \mathbf{E}_1\mathbf{E}_2\left[\frac{1}{N^2}\sum_{i=1}^{k_1}\frac{1}{(\pi_i\pi_{i|s_1})^2} \times \sum_{u_j \in A_i}\sum_{u_{j'} \in A_i}\frac{\pi_{jj'|i} - \pi_{j|i}\pi_{j'|i}}{\pi_{j|i}\pi_{j'|i}}e_je_{j'}\right],$$

where $\pi_{j|i}$ is the conditional probability, given S_2, that u_j is including in S_{3i}; $\pi_{jj'|i}$, $j \neq j'$, is the conditional probability, given S_2, that both u_j and $u_{j'}$ are including in S_{3i}; and $\pi_{jj|i} = \pi_{j|i}$, $j = 1, \ldots, k_1$.

An estimator $\hat{\mathbf{V}}(\hat{\mu}_R)$ of $\mathbf{V}(\hat{\mu}_R)$ can be obtained by constructing estimators of $\mathbf{V}_{FP}$, $\mathbf{V}_{SP}$, and $\mathbf{V}_{TP}$.

An estimator of $\mathbf{V}_{FP}$, obtained by using the Delta method, is

$$\hat{\mathbf{V}}_{FP} = \frac{1}{N^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_{ij|s_1}}\frac{\hat{Y}_i}{\pi_i}\frac{\hat{Y}_j}{\pi_j} - \sum_{i=1}^{k_1}\frac{1-\pi_i}{\pi_i^2\pi_{i|s_1}} \times \sum_{u_j\in S_{3i}}\sum_{u_{j'}\in S_{3i}}\frac{\pi_{jj'|i}-\pi_{j|i}\pi_{j'|i}}{\pi_{jj'|i}\pi_{j|i}\pi_{j'|i}}y_jy_{j'}\right].$$

Similarly, an estimator $\hat{\mathbf{V}}_{SP}$ of $\mathbf{V}_{SP}$ is

$$\hat{\mathbf{V}}_{SP} = \frac{1}{N^2}\sum_{i=1}^{k_1}\sum_{j=1}^{k_1}\frac{\pi_{ij|s_1}-\pi_{i|s_1}\pi_{j|s_1}}{\pi_{ij|s_1}\pi_{i|s_1}\pi_{j|s_1}} \times\left(\frac{1}{\pi_i}\sum_{u_t\in S_{3i}}\frac{\hat{e}_t}{\pi_{t|i}}\right)\left(\frac{1}{\pi_j}\sum_{u_t\in S_{3j}}\frac{\hat{e}_t}{\pi_{t|j}}\right) - \frac{1}{N^2}\sum_{i=1}^{k_1}\frac{1-\pi_{i|s_1}}{\pi_{i|s_1}^2}\frac{\hat{\mathbf{V}}_{S_3i}}{\pi_i^2}, \tag{2}$$

where

$$\hat{\mathbf{V}}_{S_3i} = \sum_{u_j\in S_{3i}}\sum_{u_{j'}\in S_{3i}}\frac{\pi_{jj'|i}-\pi_{j|i}\pi_{j'|i}}{\pi_{jj'|i}\pi_{j|i}\pi_{j'|i}}\hat{e}_j\hat{e}_{j'}$$

and $\hat{e}_j = y_j - \mathbf{x}_j^t\hat{\mathbf{B}}_{\mathbf{s_3}}$. Finally, an estimator $\hat{\mathbf{V}}_{TP}$ of $\mathbf{V}_{TP}$ is

$$\hat{\mathbf{V}}_{TP} = \frac{1}{N^2}\sum_{i=1}^{k_1}\frac{1}{(\pi_i\pi_{i|s_1})^2}\hat{\mathbf{V}}_{S_3i}.$$

Therefore, $\hat{\mathbf{V}}(\hat{\mu}_R)$ is given by $\hat{\mathbf{V}}(\hat{\mu}_R) = \hat{\mathbf{V}}_{FP}+\hat{\mathbf{V}}_{SP}+\hat{\mathbf{V}}_{TP}$.

A slightly different regression estimator ($\hat{\mu}_R^*$) can be obtained by solving the following set of estimating equations:

$$\hat{\mathbf{G}}_{s_1}(\mathbf{x}_{s_1},\boldsymbol{\mu}_{\mathbf{x}}) = \sum_{i=1}^{k}\sum_{u_j\in A_i}\frac{1}{\pi_i}(\mathbf{x}_j-\boldsymbol{\mu}_{\mathbf{x}}) = \mathbf{0},$$

$$\hat{\mathbf{G}}_{s_3}(y_{s_3},\mathbf{x}_{s_3},\boldsymbol{\beta}) = \sum_{i=1}^{k_1}\frac{1}{\pi_i\pi_{i|s_1}}\left[(\widehat{\mathbf{X}Y})_i - (\widehat{\mathbf{X}\mathbf{X}^t})_i\boldsymbol{\beta}\right] = \mathbf{0},$$

and

$$\hat{\mathbf{G}}_{s_3}(y_{s_3},\mathbf{x}_{s_3},\mu_y) = \sum_{i=1}^{k_1}\frac{(\hat{Y}_i-\hat{m}_i\mu_y)+(\hat{m}_i\hat{\boldsymbol{\mu}}_{\mathbf{x}}-\hat{\mathbf{x}}_i)^t\hat{\mathbf{B}}_{\mathbf{s_3}}}{\pi_i\pi_{i|s_1}} = 0,$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is the solution of the first equation, $\hat{\mathbf{B}}_{\mathbf{s_3}}$ is the solution of the second one, and $\hat{m}_i$ is an estimator of the size m_i of the network A_i. This set of equations gives rise to an estimator of the same form as that given by (1), but the value of N that appears in the estimator $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is replaced by $\hat{N}_{s_1} = \sum_{i=1}^{k} m_i/\pi_i$, and the value of N appearing in $\hat{\mu}_y^{'''}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{'''}$ is replaced by $\hat{N}_{s_3} = \sum_{i=1}^{k_1}\hat{m}_i/(\pi_i\pi_{i|s_1})$.

The asymptotic variance of $\hat{\mu}_R^*$ is the same as that of $\hat{\mu}_R$. The variance of $\hat{\mu}_R^*$ might be estimated by the same estimator used in the case of $\hat{\mu}_R$; however, an alternative variance estimator is obtained by replacing the N's appearing in the expression for $\hat{\mathbf{V}}(\hat{\mu}_R)$ by $\hat{N}_{s_1}$.

4 Monte Carlo studies

To compare the performance of ACDS with that of ACS two simulation studies were carried out. Next, we will describe both studies.

4.1 A simulated population

A population of point-objects was generated from a Poisson Cluster process, conditioning on the number of "parents". The number of parents was set to 4, and they were randomly located in the study region. The study area was a 25 × 25 unit square divided into 625 equal size 1 unit^2 quadrats. Poisson distributions with means 100, 70, 30 and 20 were used to generate the number of offspring associated with parents 1, 2, 3 and 4, respectively. The offspring were placed around their parents using bivariate normal distributions with means centered on the parents' locations, and compound symmetric variance-covariance matrices. The values of the variances were set to 4, 4, 1, and 1 (listed in the same order as that of the means of the Poisson distributions), and those of the correlations were set to 0.3, 0.3, 0 and 0. If a coordinate of a child's location was outside the study region, that coordinate was set to a value on the border of the region. The y-value associated with each quadrat was the number of point-objects located in the quadrat. The finite population mean was $\mu = 0.414$.

Three sampling designs were compared: the ordinary ACS design, and two variants of ACDS. In the three cases, the initial sample of quadrats was selected by a SRSWOR design, and the neighborhood of a quadrat was defined as the set consisting of that quadrat plus the quadrats placed directly to the north, south, east, and west of that quadrat. In the case of ACS, the condition for additional sampling

Table 1: Simulation results based on 2000 replicated samples

Sampling design	Initial sample size	Marginal costs	Expected cost	Exp. numb. of y-values	MSE
ACS	10	$c_0 = 1000$, $c_1 = 100$, $c_2 = 20$, $c_3 = 90$, $c_4 = 15$	3307.41	21.1	.1206
ACDS I	16	$c'_0 = 1800$, $c'_1 = 15$, $c'_2 = 20$, $c'_3 = 10$, $c'_4 = 15$, $c'_5 = 100$ $c'_6 = 90$	3303.78	6.3	.0818
ACDS I	30	$c'_0 = 1200$, $c'_1 = 15$, $c'_2 = 20$, $c'_3 = 10$, $c'_4 = 15$, $c'_5 = 100$ $c'_6 = 90$	3289.96	6.3	.0413
ACDS II	23	$c''_0 = 1000$, $c''_1 = 15$, $c''_2 = 20$, $c''_3 = 10$, $c''_4 = 15$, $c''_5 = 100$	3286.27	11.8	.0459

was $C_y = \{y : y \geq 1\}$. In the case of the two variants of ACDS, the auxiliary variable was defined as $x_j = 1$ if unit u_j satisfies C_y, and $x_j = 0$ otherwise, and the condition for additional sampling (based on x) was $C_x = \{x : x \geq 1\}$. The previous definitions of neighborhood and criteria for additional sampling gave rise to a partition of the study region into 556 networks. Only 5 of those networks had y-values greater than zero. The sizes and y-values of those 5 networks were the following: $m_1 = 25$, $m_2 = 25$, $m_3 = 15$, $m_4 = 8$, $m_5 = 1$; $Y_1 = 101$, $Y_2 = 80$, $Y_3 = 61$, $Y_4 = 16$, and $Y_5 = 1$.

In both variants of the ACDS procedure, every network intersected by the initial sample was subsampled ($k_1 = k$), and the subsamples were taken by SRSWOR designs with sizes proportional to the network sizes. The difference between the two variants was that in one, which will be labelled ACDS I, the maximum number of y-measurements to be recorded was previously specified, and the subsamples were assumed to be selected once the adaptive sample based on x had been completed. As was indicated previously, this would imply that every network that contains quadrats satisfying C_y was visited two times. In the other variant, which will be labeled ACDS II, the number of recorded y-measurements was not controlled, and the subsample S_{3i} was selected as soon as the x-values associated with quadrats in the network A_i were registered. This would imply that every sampled network was visited only one time.

To make a fair comparison of the three designs, a cost function was defined for each design (except for the ACDS I design, which had two cost functions associated with it), and the initial samples were determined so that the expected costs of the three designs were almost the same. In the case of ACS, the initial sample size was set to $n = 10$ quadrats, and the cost function was defined as $C_T = c_0 + c_1 n_c + c_2(n - n_c) + c_3\nu_c + c_4\nu_{edge}$, where C_T and c_0 were the total and the fixed costs; c_1, c_2, c_3 and c_4 were the cost of measuring the y-value of an initially sampled quadrat satisfying C_y, an initially sampled quadrat which does not satisfy C_y, an adaptively added quadrat satisfying C_y, and an adaptively added quadrat which does not satisfy C_y; and n_c, ν_c and ν_{edge} were the number of initially selected quadrats satisfying C_y, the number of adaptively added quadrats satisfying C_y, and the number of adaptively added edge units. In the case of the ACDS I design, the maximum number of recorded y-values was set to 7, and the cost function was defined as $C'_T = c'_0 + c'_1 n'_c + c'_2(n' - n'_c) + c'_3\nu'_c + c'_4\nu'_{edge} + c'_5 n'_c + c'_6(m' - n'_c)$, where the costs C'_T, c'_0, c'_1, c'_2, c'_3 and c'_4 were similarly defined as those in the cost function for the ACS design, but they refer to costs of measuring x-values instead of y-values; the costs c'_5 and c'_6 were the costs of measuring the y-values of an initially selected unit satisfying C_y and an adaptively added unit satisfying C_y; n', n'_c, ν'_c and ν'_{edge} were defined as in the previous case; and m' was the total number of y-measurements. Finally, in the case of the ACDS II design, the size of the subsample S_{3i} was set to 30% of the size of the network A_i, $i = 1, \ldots, k$, and the cost function was defined as $C''_T = c''_0 + c''_1 n''_c + c''_2(n'' - n''_c) + c''_3\nu''_c + c''_4\nu''_{edge} + c''_5 m''$, where the costs C''_T, c''_0, c''_1, c''_2, c''_3 and c''_4 were similarly defined as those in the cost function for the ACS I design; c''_5 was the cost of measuring the y-value of a unit satisfying C_y; and the numbers n'', n''_c, ν''_c, ν''_{edge} and m'' were similarly defined as those in the cost function for the ACDS I design.

In the case of the ACS design, the population mean was estimated by the Horvitz-Thompson type of estimator. In the case of the ACDS designs, the population mean was estimated by (1), which in the

Table 2: Simulation results based on 10000 replicated samples

Sampling design	Initial sample size	Marginal costs	Expected cost	Exp. numb. of y-values	MSE
ACS	5	$c_0 = 1000$, $c_1 = 100$, $c_2 = 20$, $c_3 = 90$, $c_4 = 15$	1904.64	7.7	65909.29
ACDS II	15	$c_0'' = 1000$, $c_1'' = 15$, $c_2'' = 20$, $c_3'' = 10$, $c_4'' = 15$, $c_5'' = 100$	1901.198	4.2	199710.9
ACDS III	15	$c_0'' = 1000$, $c_1'' = 15$, $c_2'' = 20$, $c_3'' = 10$, $c_4'' = 15$, $c_5'' = 100$	1901.198	4.2	5451.344

case of our auxiliary binary variable x, it reduces to the unbiased estimator $\hat{\mu}_y'''$. The simulation mean square errors (MSE) of the estimators, based on 2000 replicated samples, are shown in Table 1. From the results in this table, we can see that all the considered ACDS designs gave better results (in terms of the MSE) than those given by the ACS design. The minimum MSE was achieved by the design that used the largest initial sample size. In fact, for the particular sampled population there is a decreasing relationship between the MSE and the initial sample size. The reason for this is that the variability within each network is not so large ($S_1^2 = 11.29$, $S_2^2 = 5.91$, $S_3^2 = 6.92$, $S_4^2 = 1.71$, and $S_5^2 = 0$); therefore small subsamples are enough to yield good estimates of the network-totals Y_i, and the best strategy is the one that samples, in the average, more number of networks.

4.2 A real population

Here, we considered a population of blue-winged teals reported in Smith et al. (1995). The study region was a 5000 km^2 rectangle divided into $5 \times 10 = 50$ quadrats of 100 km^2. The y-value associated with a quadrat was the number of birds in that quadrat, and the population mean was $\mu = 282.42$. In this study, we considered the same definitions of x, C_y, C_x, and neighborhood of a quadrat as those given in the previous study. Those definitions of neighborhood and criteria for additional sampling gave rise to a partition of the study region into 38 networks. Only 3 of those networks had y-values greater than zero. The sizes and y-values of those 3 networks were the following: $m_1 = 7$, $m_2 = 7$, $m_3 = 1$, $Y_1 = 53$, $Y_2 = 14066$, and $Y_3 = 2$. It is important to indicate that the y-value associated with one of the quadrats in the second network was equal to 13639, and that this y-value was responsible for the high value of Y_2.

We compared the ACS design, the ACDS II design, and another variant of ACDS, which will be labelled ACDS III. This variant was identical to the ACDS II design but whenever the second network was intersected by the initial sample, the quadrat with y-value equal to 13639 was included in the subsample S_{3i} with probability 1. The results of this study, based on 10000 replicated samples are shown in Table 2. From this results, we can see that the ACDS II design performed very badly. The problem with this design is that the variability within the second network was 2.6×10^7. On the other hand, the very good performance of the ACDS III design is because the inclusion of the quadrat with the highest y-value in S_{3i} with probability 1 reduced the within variability of the second network to 5346.2, which is not so large.

References

Brown, J.A. (1994). The application of adaptive cluster sampling to ecological studies. In: D.J. Fletcher and B.F.J. Manly (eds.) *Statistics in Ecology and Environmental Monitoring, 2*, pp. 86-97. Dunedin, New Zealand: University of Otago Press.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

Salehi, M.M., and G.A.F. Seber (1997). Two-stage adaptive cluster sampling. *Biometrics* **53**, 959-970.

Särndal, C.E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling.* Springer, New York.

Smith, D.R., M.J. Conroy, and D.H. Brakhage (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics* **51**, 777-788.

Thompson, S.K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association* **85**, 1050-1059.