Erica Swanson

Department of Mathematical Sciences
Montana State University

May 3, 2013

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Erica Swanson

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	YOUR ADVISOR'S NAME HERE Writing Project Advisor
Date	Megan D. Higgs Writing Projects Coordinator

Introduction

Precise and accurate estimates of a population total or population mean along with the associated variance estimates are essential when drawing inference about a population. Conventional sampling designs are most efficient when the population is evenly distributed over the study area. Many populations, however, are rare, hidden, and/or hard to reach. When conventional sampling designs such as simple random sampling are used on these types of populations, standard errors are inflated. A new design-based method known as adaptive cluster sampling (ACS) was introduced by Thompson in 1990 to improve estimation for these types of populations. Applications of ACS are primarily ecological in which the definition of neighbor is in terms of spatial adjacency; however, it can be applied to other types of populations with rare characteristics such as people who use illicit drugs or who are infected by HIV as long as there is an appropriate neighborhood definition. For this type of population a social linkage would be used to adaptively sample such as social contact or kinship.

In adaptive cluster sampling, the sampling is "adapted" to the data. An initial sample of size n_1 is taken from the population. When the observed value of an initially sampled unit satisfies some condition of interest C, additional units in a pre-defined neighborhood are also selected. This process is iterated until no units satisfying C are encountered. An example of a pre-determined choice of C could be if at least 1 member of the population is observed in the selected area. This type of sampling is designed to give the researcher a good idea of the location of where a rare population resides and allow sampling in the vicinity to gather as much accurate information about the population.

Terminology

For the rest of this paper, I will be talking about Adaptive Cluster Sampling in terms of spatial adjacency. Many low-abundance plants and animals occur in clusters; therefore, once a quadrat is found to have a member of the population of interest, quadrats in the neighborhood of that occupied quadrat are more likely to have additional members of that population. A criterion for sampling subsequent quadrats could be observing more than some threshold number of members such as the presence of a single member of the population of interest. The only constraint is that the definition of neighbor must be transitive. This means if i is a neighbor j, then j is also a neighbor of i and vice versa. In order for adaptive cluster sampling to work more efficiently than simple random sampling, neighbors must have positive covariation in the sampled attribute.

In spatial clustering, a *neighborhood* could be defined as 2, 4, or 8 adjacent quadrats. If an initially sampled unit meets the criterion, then each surrounding quadrat defined in the neighborhood is also sampled. This procedure results with networks of sampled quadrats until each network is surrounded by a ring of empty quadrats that do not satisfy condition C. The final result consists of a sample with 3 different types of quadrats. The first type is quadrats sampled initially whether they satisfy C or not. The second type are referred to as *secondary units* which are the quadrats sampled, not because they are part of the initial sample, but because they are in the recursive neighborhood of a quadrat in the initial sample that satisfied C. The third type is *edge quadrats* which are quadrats that do not satisfy C and are not in the initial sample, but are neighbors of quadrats with members of either of the first 2 types. A *network* is defined as a set of quadrats such that if any quadrat in the network is sampled, all quadrats in that particular network are sampled. A network can be an initially sampled quadrat satisfying C plus all neighboring quadrats that satisfy C, but do not include the edge units. A network can also be of size 1 which occurs when an initially sampled quadrat does not meet the criterion C.

Estimators

The sample mean $\bar{y}$ of the sampled quadrats is a biased estimate of the population mean because more quadrats satisfying C were disproportionately included and quadrats within larger networks have a increased probability of being added to the sample (Philippi 2005). Estimators have been modified that are design-unbiased for the population mean along with design-unbiased estimators of their variances. Additionally, the Rao-Blackwell method has been used to obtain smaller variance design-unbiased estimators for ACS.

The following description and formulas for the ACS estimators are taken from Thompson (1990). The Horvitz-Thompson and Hansen-Hurwitz estimators are the two types of estimators used in adaptive cluster sampling. In order to apply the Horvitz-Thompson estimator, the inclusion probabilities π_i must be known for each sampled quadrat. This is why it is necessary the neighborhood definition is pre-defined and transitive. Quadrats may be included via more than one process either in the initial sample, or an edge quadrat adjacent to a network with at least one quadrat in the initial sample.

Suppose an initial sample n is taken via SRSWOR. Because of the multiple ways for a quadrat to be included in the sample, the probability of its inclusion is 1 minus the fraction of possible sample draws that would not have included that quadrat.

$$\pi_i = 1 - \frac{\binom{N-m_i-b_i}{n}}{\binom{N}{n}}$$

In the formula above, π_i is the probability a quadrat is selected, n is the number of quadrats in the initial sample, m_i is the number of quadrats in the network that includes quadrat i, b_i is the number of quadrats in networks for which quadrat i is an edge. N and n are known initially and m_i can be tallied for each sampled quadrat as long as the sampling continues until the entire network is sampled. If the sampling procedure does not detect the existence of a cluster, potential edge units will not be sampled and so b_i can not be determined because the sampling design does not provide enough information to calculate the probability of each quadrat included in the sample. The solution to this defect is to exclude edge quadrats from the estimation (Thompson 1990). This allows for the calculation of the probability of quadrat i's inclusion in the sample using the information obtained in the sampling.

The probability of inclusion is equal for all quadrats in a network; therefore, the probability of inclusion for network a_k can be calculated.

$$a_k = 1 - \frac{\binom{N-x_k}{n}}{\binom{N}{n}}$$ x_k is the number of quadrats found in network k

The Horvitz-Thompson estimator for this adaptive design is $\hat{\mu}_{HT} = \frac{1}{N}\sum_{k=1}^{n}\frac{y_k}{a_k}$ where y_k is the number of individuals found in the entire network y_k.

To estimate the variance of the Horvitz-Thompson estimator, the probabilities that both network j and network k are included in the sample are needed.

$$a_{jk} = 1 - \frac{\binom{N-x_j}{n} + \binom{N-x_k}{n} - \binom{N-x_j-x_k}{n}}{\binom{N}{n}}$$

$$\widehat{Var}(\hat{\mu}_{HT}) = \frac{1}{N^2}\sum\sum\frac{y_j y_k}{a_{jk}}(\frac{a_{jk}}{a_j a_k} - 1)$$

If the initial sample is chosen with replacement, then the inclusion probabilities would change:

$$a_k = 1 - (1 - \frac{x_k}{N})^n$$
$$a_{jk} = 1 - [(1 - \frac{x_k}{N})^n + (1 - \frac{x_j}{N})^n - (1 - \frac{(x_k + x_h)}{N})^n]$$

The summations for both $\hat{\mu}_{HT}$ and the estimated variance of $\hat{\mu}_{HT}$ are taken over the unique networks v ($\leq n$) observed in the sample. A given network may be intersected by more than one quadrat in the initial sample. For the H-T estimators, that network is only included once. Essentially, the estimators are based on the probabilities of inclusion or intersection of each sampled network.

A second unbiased estimator is based on the Hansen-Hurwitz estimator used in ACS. This estimator is based on the number of times each network is intersected.

$$\hat{\mu}_{HH} = \frac{1}{n}\sum_{i=1}^{n} w_i$$
$$\widehat{Var}(\hat{\mu}_{HH}) = \frac{N-n}{Nn(n-1)}\sum_{i=1}^{n}(w_i - \hat{\mu}_{HH})^2$$

where w_i is the mean quadrat abundance for quadrats in cluster i. In general,

$$\widehat{Var}(\hat{\mu}_{HT}) < \widehat{Var}(\hat{\mu}_{HH})$$

Improvement of the Estimators Through the Rao-Blackwell Method

Thompson (1990) discusses the improvement of the estimators used in ACS through the Rao-Blackwell Method which is unique in that it utilizes edge units even if they are not selected in the initial sample. None of the estimators discussed earlier is a function of the minimal sufficient statistic which means each can be improved by using the Rao-Blackwell method of taking their conditional expectations given the minimal sufficient statistic.

The three unbiased estimators $\bar{y}_1$, t_{HH}, and t_{HT} depend on the order of selection. In addition, t_{HH} depends on repeat selections; and if the initial sample is selected with replacement, $\bar{y}_1$ also depends on repeat selections. Define t as any of the three unbiased estimators mentioned earlier and consider the estimator t_{RB}=E(t|D) where D is the minimal sufficient statistic that is the unordered set of distinct, labeled observations in the finite population. D=$\{(k, y_k)\colon k \in s\}$ where s denotes the set of distinct units included in the sample. Define v

as the effective sample which is the number of units in the final sample. There are $G = \binom{v}{n_1}$ combinations of the initial sample when units are selected without replacement in a simple random sample. Let t_g be the value of the estimator t obtained when the initial sample consists of combination g. Similarly, $\widehat{var}(t_g)$ is the value of the variance estimator obtained with initial sample g.

Define k^* as the number of distinct networks represented in the sample excluding the sample edge units. An initial sample of n_1 units gives rise to the given value of D iff the initial sample contains at least one unit from each k^* unique networks minus the sample edge units. Any initial sample that gives rise through the design to the given value D of the minimal sufficient statistic is *compatible* with D. If x_j is the number of units in the initial sample from the jth network, then an initial sample of n_1 units from the v distinct units in D is compatible with D iff $x_j \geq 1$ for j=1,...,k^*. The indicator variable I_g is 1 if the gth combination of n_1 units from the sample is compatible with D and 0 otherwise; thus, the total number of compatible combinations is $\zeta = \sum_{g=1}^{G} I_g$.

The Rao-Blackwell estimator $t_{RB} = \zeta^{-1} \sum_{g=1}^{G} t_g I_g$ and

$$\widehat{var}(t_{RB}) = \zeta^{-1} \sum_{g=1}^{G} [\widehat{var}(t_g) - (t_g - t_{RB})^2] I_g.$$

The variance estimate is unbiased; however, it can result in negative values with certain sets of data. Overall, the Rao-Blackwell method is a recent and useful development in ACS in terms of obtaining smaller variance design-unbiased estimators. On the other hand, the method requires further research in the calculations of Rao-Blackwell estimators because the numbers of terms in the preceding expressions are potentially large.

An Example

In the 10×10 grid below, a hypothetical population of 100 members is distributed over 100 quadrats. To estimate the true population total, an initial simple random sample of 15 quadrats without replacement is implemented in which the quadrat with 2 members of the population, the quadrat with 9 members of the population, and 13 quadrats without members of the population are selected. If the criterion C is pre-set to 1 member of the population and the neighborhood is defined as the 4 surrounding quadrats, then the adaptive sample consists of 13 networks of size 0, one of size 21, and the other of size 57. The Y values, sum of Y values, and size of each network are given in Table 1 and the Horvitz-Thompson and Hansen-Hurwitz estimates are given in Table 2. The R-code used to calculate these estimates can be found in the Appendix. The Horvitz-Thompson estimate of the true population total of 107 is more accurate than the Hansen-Hurwitz estimate of 64; however, in this example, the standard error of the Horvitz-Thompson total estimate of 37 is larger than the standard error of the Hansen-Hurwitz total estimate of 16.

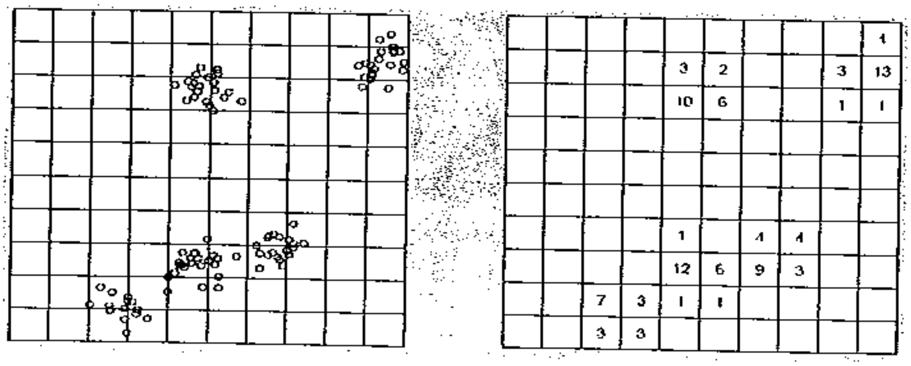

Table 1

Y Values	Sum of Y Values	Network Size
3, 2,10, 6	21	4
1, 4, 4, 12, 6, 9, 3, 7, 3, 1, 1, 3, 3	57	13
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1

Table 2

Parameter	Horvitz-Thompson	Hansen-Hurwitz
Total	107.	64
SE of estimated total	37	16
Approximate 95% CI for total	(34,180)	(33,96)
Mean	1.07	0.64
SE of estimated mean	0.37	0.16

Comparisons Between Classical Designs and ACS Designs

The objective of both classical and ACS designs is to select a sample of units, observe the corresponding y-values, and then estimate some function z(y) (e.g., the population total $z(y) = \sum_{i=1}^{N} y_i = \tau$, the population mean $z(y) = \frac{1}{N}\sum_{i=1}^{N} = \mu$, or the population variance $z(y) = \sum_{i=1}^{N} \frac{(y_i-\mu)^2}{N-1} = \sigma^2$). With classical designs, the entire selection of sample units is made prior to observing the y_i values; whereas, in ACS, the selection of y_i values is adaptive such that the sample depends on the observed y_i values (Turk & Borkowski 2005).

Advantages

ACS has many advantages over classical designs. ACS can be more efficient such that the variances of estimators will be smaller given an equivalent sampling effort. The locations and shape of clusters of individuals are oftentimes not known prior to sampling; thus, ACS can be implemented where stratification may not be possible (Thompson, 1991b). ACS is designed to effectively detect areas of high abundance which increases the information provided by the sample. This is particularly useful in gathering as much information as possible about individuals in populations that are rare and hard to detect (Seber & Thompson 1994). ACS is flexible in that the researcher determines the initial sample size, the sampling condition C, the unit size, and the definition of the neighborhood. Subsequently, there exists a multitude of methods for choosing the initial sample size (simple random sampling with replacement (SRSWR), simple random sampling without replacement (SRSWOR), strip sampling (example in *A Simulation Study*), systematic sampling, stratified sampling, sampling using probabilities proportional to size, and simple Latin square sampling. Two other advantages in terms of cost are the average distances between sampled units are smaller and the quadrat locations are easier to find (Salehi & Seber, 1997; Brown & Manly, 1998).

Disadvantages

Because ACS is adaptive, the final sample size is random which makes it difficult to determine and control the sampling effort as well as the cost of the survey in advance. Researchers have proposed methods to deal with these issues which will be discussed in *Methods to Ease the Problem of a Random Final Sample Size.* The flexibility of ACS in terms of the selection of C, the initial sample size, the unit size, and the neighborhood definition can also be a disadvantage because it makes finding variance-optimal designs complicated and are, at the same time, critical to the efficiency of the ACS design. Brown (2003) discusses how these factors should be determined in order to get a more efficient ACS sampling design which is detailed later in *Designing an Efficient Adaptive Cluster Sample.* Another difference in ACS relative to classical designs is that not all information from sampled units is used in ACS. In particular, edge units are only used if they are part of the initial sample. The development of Rao-Blackwell estimators addresses this concern by incorporating information from edge units.

Increase in Efficiency of ACS

Thompson (1997) gives the following characteristics that tend to increase the efficiency of adaptive cluster sampling relative to conventional random sampling: (1) the within-network variance makes up a large proportion of the total population variance (i.e., the population is highly aggregated with large variability within those aggregations) (2) The population is rare (3) The expected final sample size is not much larger than the initial sample size (4) The cost of observing units in networks is not more than the cost of observing the same number in a random sample (5) The cost of observing units not satisfying the criterion C is less than the cost of observing units satisfying the criterion (6) The condition for extra sampling may be based on an auxiliary variable that is easy to measure (7) A Rao-Blackwell estimator or other efficient estimator is used with ACS.

A Simulation Study

Relative efficiency $\frac{Var(\hat{y})}{Var(\hat{\mu})}$, is the ratio of the sample variance of the estimated density from simple random sampling and the sample variance of the estimated density from adaptive cluster sampling; thus, the higher the relative efficiency, the more efficient ACS is compared to the classical design. Conners and Schwager (2002) evaluate the relative efficiency of ACS and traditional sampling designs in a hydroacoustic survey setting. They use simulations to show that high relative efficiency of ACS is associated with distributions that are strongly skewed, have high kurtosis, and have a large proportion of units with zero over very low densities. The researchers use data from a hydroacoustic survey of rainbow smelt from the eastern basin of Lake Erie. The Lake Erie Fisheries unit would like to optimize a survey design for estimating the total stock size of smelt, with an accurate estimate of the associated variance. In order to test the efficiency of ACS, a simulation study is conducted using fish stocks similar to Lake Erie smelt. The simulated test stocks are created with known true total size and different levels of spatial aggregation. Selected stocks are then sampled repeatedly using both traditional and ACS designs.

The selected models consist of stocks with no local correlation ("Random"), with strong local correlation over a large range ("Big Patches"), and with strong local correlation over a small range ("Small Patches). The fourth stock ("Rare Patches") represents strong local correlation with relatively high background noise which is most similar to the Lake Erie smelt data.

Sampling and estimation for the traditional designs use one-stage cluster sampling formulas. Traditional designs include random selection of 10 transects, systematic selection of 10 transects with a random start and three stratified, random sampling designs in which equal allocation of transects to strata is used. Results are found to be similar for the 3 types of stratification so only one of the stratum designs is presented.

The ACS sampling uses the transects as primary units and the individual grid cells as secondary units. A "neighborhood" definition of 4 adjacent cells is used and the critical value defining networks is set at the 80th percentile for the true distribution of the grid points. Initial sample sizes for the ACS designs are selected to give expected final sample sizes (total number of secondary units sampled) as close to the fixed size for the traditional designs.

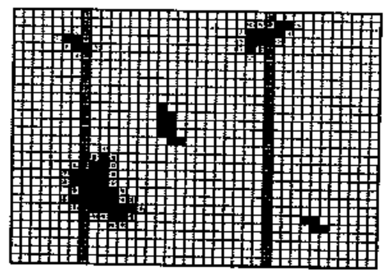

Figure 1. An example of strip adaptive cluster sampling for a patchy population. Black squares represent fish densities above the critical value. The initial sample consists of two transects which detect three patches. Units adjacent to high densities are added adaptively until densities drop below the critical value. The final sample size includes the initial transects plus adaptively added units.

The results of the simulation show that fixed-size cluster designs are more efficient than SRS only when systematic and stratified cluster designs are applied to the stock with the big patches. ACS proves to be more efficient than SRS when the target stock is rare or highly aggregated or both. ACS is 3 times that of SRS, in terms of efficiency, for the "rare patches" stock. ACS is not efficient for the "big patches" stock because the large final sample size makes the equivalent SRS variance small. Overall, ACS estimators exhibit an unbiased, symmetric distribution with a consistently lower variance than traditional designs when the population is spatially patchy or rare or both.

Final Sample Size

As mentioned before, one of the greatest limitations in the implementation of ACS is the random nature of the final sample size and the possibility that the sample size will grow too large to be feasible. The efficiency of ACS can be assessed by comparing the initial sample size to the final sample size. For the stock with no spatial correlation, the average final sample sizes are only 11-13% larger. None of the final sample sizes exceed 1.25 times the initial sample size for this stock. For the "small" and "rare" stocks, the final sample size using stratified random ACS is 1.7-2.1 times the initial sample size when starting from a systematic sample. The stock with a few "big clusters" saw the greatest increase in final sample size with the average final sample size being 2-4 times the size of the initial sample. Because the "small" and "rare" stocks consistently have a lower variance through ACS than the other populations and because the final sample size is between 1.7 and 2.1 times the initial sample size which is "not too large", ACS is most efficient for these populations. This efficiency is discussed further in *Designing an Efficient Adaptive Cluster Sample.*

Methods to Ease the Problem of a Random Final Sample Size

Restricted Adaptive Cluster Sampling

Brown and Many (1998) propose restricted adaptive cluster sampling as a method to deal with the limitation of the random nature in the final sample size. In this modification, a limit is placed on the final sample size prior to sampling resulting in less variation of the final sample size and allowing sampling effort to be predicted with some certainty. In restricted adaptive cluster sampling, if the cumulating total sample size is below a predefined limit on the number of quadrats then another "initial" quadrat is selected. If the selected quadrat and associated quadrats in the network result in a cumulative sample size greater than the predetermined final sample size, then the network is included, but then sampling terminates. This results with a final sample size either equivalent to the predetermined value or just slightly above it.

Stratification

Thompson and Seber (1996, p 134) offer another solution which involves stratifying the study area prior to sampling and decisions regarding the continuation of the sampling effort is determined after the completion of adaptive sampling of each stratum. Thompson and Seber state that termination of a network at the stratum boundary is slightly less efficient than using complete networks but the ACS estimators will still be design-unbiased for the stratum totals. The stratum totals can be combined into an overall estimate assuming independence of the strata. Essentially, this "partitioning of boundaries" limits the potential size of any network which is one of the strategies used to control the final sample size (Conners 1322).

Inverse Adaptive Cluster Sampling

Christman and Lan (2001) discuss inverse adaptive cluster sampling as another way to decrease excessively large final sample sizes. The method involves selecting initial sampling units until some pre-specified number (k) of nonzero y-values are observed.

Sampling is random (with or without replacement) and stops when $1 < k \leq M$ units are sampled, k given. An alternative, unbiased estimator for the population total τ under this stopping rule is $\hat{\tau}_I = M\bar{y}_M + (N - M)\bar{y}_{N-M}$. Since M is rarely known, $\hat{M} = \frac{N(k-1)}{n_1-1}$; thus, $\hat{\tau}_I = \hat{M}\bar{y}_M + (N - \hat{M})\bar{y}_{N-M}$. The estimated variance is given below.

$$\widehat{Var}(\hat{\tau}_I) = [\frac{\sigma_M^2}{k} + (\mu_M - \mu_{N-M})^2]var_{n_1}(\hat{M}) + \frac{\sigma_M^2 M^2}{k} + \sigma_{N-M}^2 E_{n_1}[\frac{(N-\hat{M})^2}{n_1-k}]$$

Application of Inverse Adaptive Cluster Sampling

Inverse Adaptive Cluster Sampling is applied to Y={0, 0, 0, 0, 10, 40, 0, 0, 5, 20, 60, 0, 0, 0, 0, 0, 0, 0, 0, 0}. M=5 nonzero y-values, N=20, Population total τ=135, and C={y: y > 0}. The adaptive neighborhood is defined as the units to the immediate left and right. The initial random sample n_o=5 and k is set to 2.

Suppose $n_0 = \{y_1 = 0, y_3 = 0, y_7 = 0, y_9 = 5, y_{15} = 0\}$. One unit satisfies the adaptive sampling condition C in this case. It's network means are $\{\bar{y}_1 = 0, \bar{y}_3 = 0, \bar{y}_7 = 0, \bar{y}_9 = 28.33, \bar{y}_{15} = 0\}$. This sample does not satisfy the requirement k=2; therefore, an additional unit is sampled, say $y_4 = 0$. This does not satisfy k=2 so a sample of one unit is taken again, say $y_{11} = 60$. The stopping rule is now met and the final adaptive sample is $\{\bar{y}_1 = 0, \bar{y}_3 = 0, \bar{y}_7 = 0, \bar{y}_9 = 28.33, \bar{y}_{15} = 0, \bar{y}_4 = 0, \bar{y}_{11} = 28.33\}$. The estimate of M is $\hat{M} = \frac{N(k-1)}{n_1-1} = \frac{20(1)}{6} = 3.33$; thus, the estimate of the population total is $\hat{M}\bar{y}_M = 3.33(28.33) = 94.34$.

A Simulation Study on Inverse Adaptive Cluster Sampling

In a simulation study, Christman and Lan (2001) demonstrate how a modified stopping rule that incorporates an adaptive sampling component and utilizes an initial random sample of fixed size is the best in term of minimizing variance. The sequential sampling strategies are applied to 4 small (N=200) populations, 3 real and 1 artificial. The green-winged teal and the blue-winged teal populations represent typical rare populations for which adaptive cluster sampling is suited. There are few quadrats that contain nonzero y-values and the within-network variance is high compared with the between network variance. The ring-necked duck population contains a small number of quadrats with nonzero y-values, but the within-network variance is small relative to the between-network variance. ACS will prove to not be suitable for this population. The artificial population consists of quadrats with low abundances; however, there are a large number of quadrats (29% of population) that have nonzero y-values. C={y: y>0} for the simulation study which utilizes two stopping rules: (1) sequentially sample with replacement until k units from P_M are observed and (2) take a fixed sample of size n_0 with replacement and, if at least k units from P_M are not observed, continue to sample until k units are observed. Both methods are sampled with and without an adaptive component and the sampling is done with replacement. Results will be very similar to sampling without replacement except the sample sizes and variances will be smaller.

25,000 Monte Carlo samples are taken from the above populations using the two stopping rules for k=2, 3, 5, or 6 and for n_0=10, 20, 35, and 50. Overall, the three estimators based on ACS almost always have smaller variability than the 3 without the adaptive sampling component and the best estimator for all populations and any choice of k is the unbiased estimator based on inverse sampling $\hat{\tau}_{IA}$. The estimators based on sequential adaptive clustering have lower $\frac{\sqrt{MSE}}{\tau}$ than the nonadaptive samples as shown below in Figure 1. This decrease is not because of higher sample sizes. There is only a slight increase in the final sample sizes when adaptively sampled for the rare, clustered green-winged teal and blue-winged teal populations (Figure 2). This is not true for the artificial population at larger values of k since adaptive cluster sampling is not appropriate for this higher density population. The results show that the initial sample size is important in determining both the coefficient of variation as well as the final sample size. The green-winged teal population is the most rare (2.5% of units are in P_M) so additional sampling will most likely occur. In contrast, 30% of the artificial population are members of P_M so it is likely that nonzero networks will be sampled initially with reasonable probability.

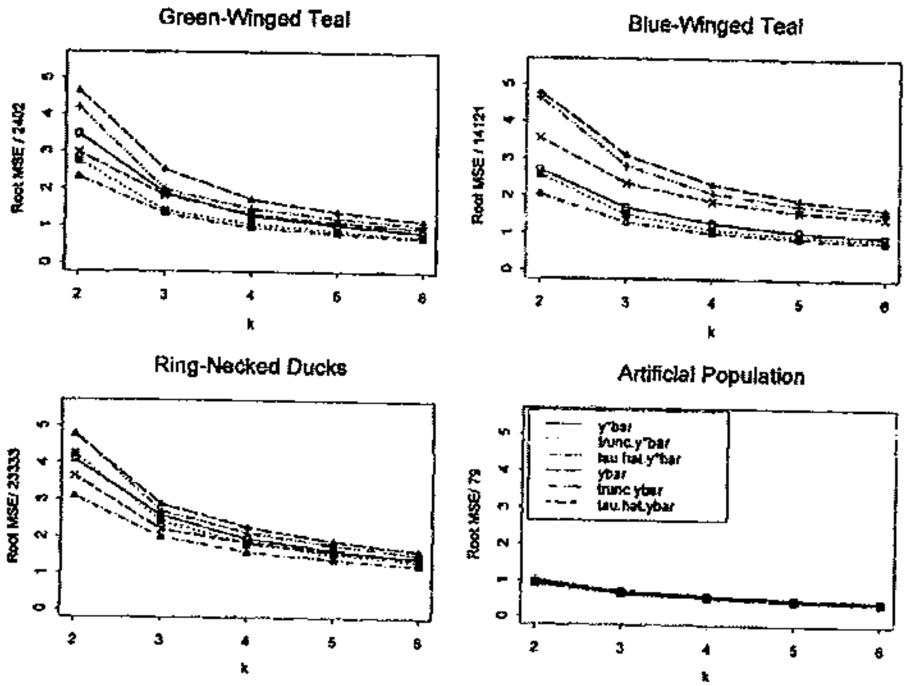

Figure 1. The ratio of the square root of MSE to the true total, τ, for six estimators of τ using inverse sampling as a function of k. The estimators are denoted as follows in the legend: $\hat{\tau}_I$ is tau.hat.ybar, $\hat{\tau}_{IA}$ is tau.hat.y*bar, $N\bar{y} = Nn^{-1}\sum_{i=1}^{n} y_i$ is ybar, $N\bar{y}^* = Nn^{-1}\sum_{i=1}^{n} y_i^*$ is y*bar, $N\bar{y}_{\text{trunc}} = N(n-1)^{-1}\sum_{i=1}^{n-1} y_i$ is trunc.ybar, $N\bar{y}^*_{\text{trunc}} = N(n-1)^{-1}\sum_{i=1}^{n-1} y_i^*$ is trunc.y*bar.

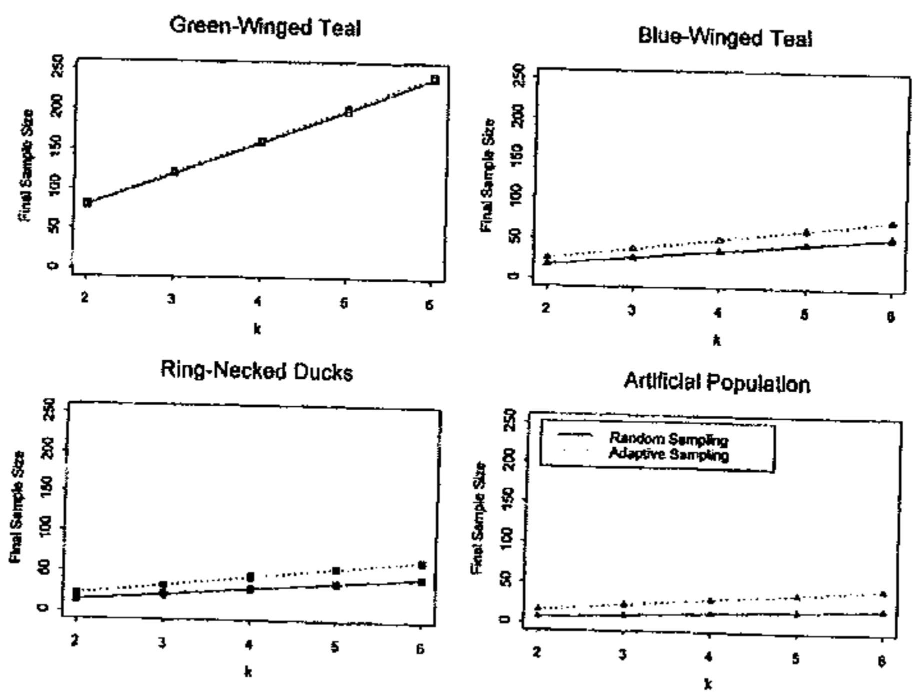

Figure 2. Expected sample sizes for sequential adaptive cluster sampling and for sequential random selections in the four populations. In both cases, sampling stops after the kth unit from P_M is observed.

Designing an Efficient Adaptive Cluster Sample

Brown (2003) explains how to design an efficient adaptive cluster sample. She discusses the importance in the choice of neighborhood and critical value in adaptive cluster sampling and that the aim should be for a small difference between the initial and final sample size along with a small difference between the within-network and population variances. These two aims can work against each other; thus, the overall aim should be networks small enough to ensure the final sample size is not excessively large relative to the initial sample size, but large enough to ensure within-network variance is a large proportion of the population variance.

Critical Value and Neighborhood Definition

The neighborhood definition and choice of critical value is important in designing an efficient sampling design. If the critical value is too small, the final sample size will be excessively large. If the critical value is too large, there will be no adaptive selection of units because no units in the initial sample will meet the criterion C. The same can be said for how neighborhood is defined. If the neighborhood is too large, the final sample size will be excessively large. If it is too small, not enough units will be adaptively sampled. More work still needs to be done in determining the choice of neighborhood and critical value. Techniques such as running a pilot study or sampling in a n-stage manner are ways to get around this. A pilot study involves surveying all the initial sampled units and based on the observed y-values, the critical value C is set to, say, the top 10th percentile. A disadvantage to this is that it could be costly to resample.

Brown's study discusses how differing neighborhood definitions, initial sample sizes, and critical values affect the efficiency of ACS. A Poisson cluster process is used to create 120 population-generating models in a $20 \times 2 \times 3$ factorial design. The mean number of clusters, λ_1, is 5, 10, 15, up to 100; the mean number of individuals associated with each cluster is 10 or 20; and the mean distance of individuals from the cluster center is θ= 0.5, 1.5, or 3.5 units. Each model is used to simulate 200 populations while the variety of models is used to cover a range of spatial patterns. In particular, 2 types of spatial patterns are used.

1. The same number of clusters, and individuals in a cluster, but with differences in compactness of clusters are compared. Compactness of clusters is how close individuals in a cluster are to each other.

2. Different numbers of individuals in a cluster along with a variety of clusters are compared.

Three neighborhood definitions are used for each simulated population: 8, 4, and 2 surrounding units and 2 critical values, 1 and 2. For each population, the variance σ^2 and the total T for the central N=400 units are calculated. Averages are taken over the 200 simulations for number of units that are network units, number of networks, average network size, network variances, and final sample size for each population in each adaptive cluster design.

Results indicate as the number of networks increase, relative efficiency decreases below 1 as shown in in the five figures below. Recall, relative efficiency is the ratio of the sample variance of the estimated density from simple random sampling and the sample variance of the estimated density from adaptive cluster sampling $\frac{Var(\hat{y})}{Var(\hat{\mu})}$ and so the higher the relative efficiency, the more efficient ACS is compared to the classical design. The number of networks or clusters depends on the underlying spatial pattern of the population and on the way the sample is designed (i.e. neighborhood and critical value definition).

The plots given below display the results from Brown's simulation study of how differing neighborhood definitions, critical values, and initial sample sizes as well as cluster compactness and size affect relative efficiency as a function of the true population total. Each plot also shows how as number of networks increases (solid line), relative efficiency (dashed line) decreases. The plot in Figure 2 below shows how the effect of neighborhood on relative efficiency depends on the true population total. A larger neighborhood definition (nd4 and nd8) results in a higher relative efficiency for true population totals below $\approx$ 200; whereas, a small neighborhood definition (nd2) results in a higher relative efficiency for true population totals above $\approx$ 200. Figure 3 shows how more compact clusters (θ=1.5 vs θ=3.5) consistently results in a higher relative efficiency as a function of the true population total. Similarly, in Figure 4, relative efficiency is consistently higher for clusters with 20 individuals than with 10 individuals. Figure 5 is similar to Figure 2 in that the effect of criterion value on relative efficiency depends on the true population total. The smaller criterion value of 1 results in a higher relative efficiency for true population totals below $\approx$ 200; however, relative efficiency is higher for a larger criterion value of 2 for true population totals above $\approx$ 200. Initial sample size does not appear to make a difference as relative efficiency only improves slightly with a larger initial sample size of 40 compared to 10 (Figure 6).

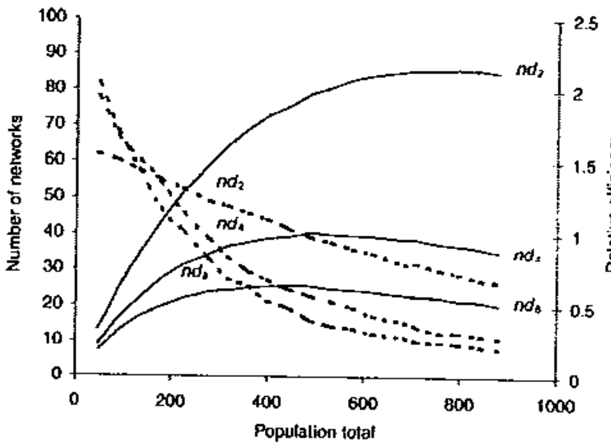

Figure 2. Change in the number of networks (solid line) and the relative efficiency (dashed line three neighborhood designs, with increasing T with λ_1 ranging from 5 to 100, $\lambda_2 = 20$, $\theta = 1$

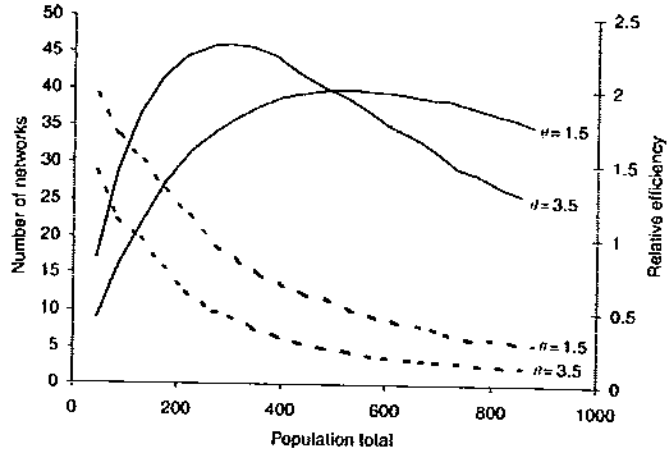

Figure 3. Change in the number of networks (solid line) and the relative efficiency (dashed line with nd_4 and cv_1, for two levels of networks "compactness" with increasing T with λ_1 ranging fro 5 to 100, $\lambda_2 = 20$, $\theta = 1.5$ and 3.5.

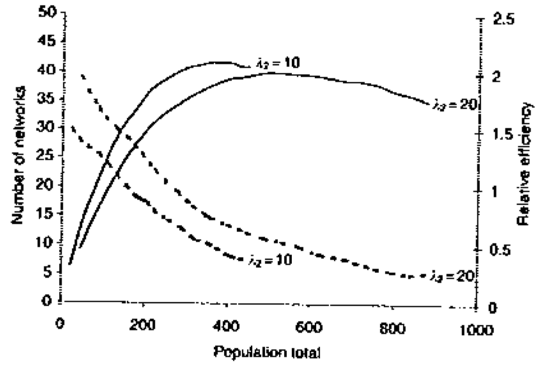

Figure 4. Change in the number of networks (solid line) and the relative efficiency (dashed line), with nd_4 and cv_1, for two levels of networks sizes with increasing T with λ_1 ranging from 5 to 100, $\lambda_2 = 10$ and 20, $\theta = 1.5$.

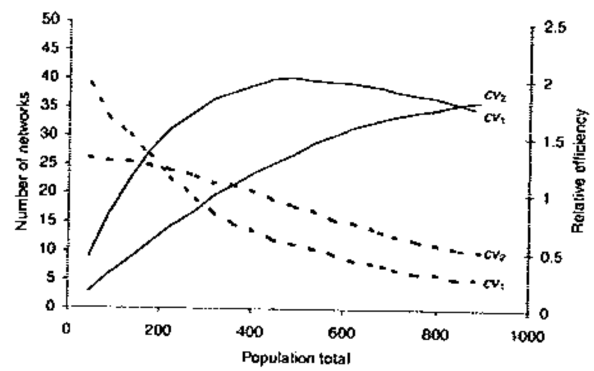

Figure 5. Change in the number of networks (solid line) and the relative efficiency (dashed line), for two levels of critical value, cv_1 and cv_2, with nd_4 and increasing T with λ_1 ranging from 5 to 100, $\lambda_2 = 20$, $\theta = 1.5$.

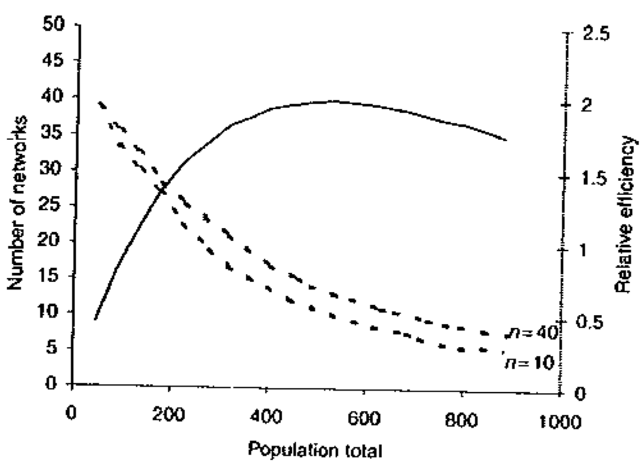

Figure 6. Change in the relative efficiency (dashed line) for two levels of initial sample size, $n = 10$ and $n = 40$, with increasing T with nd_4, cv_1, and λ_1 ranging from 5 to 100, $\lambda_2 = 20$, $\theta = 1.5$. Also shown is the number of networks (solid line).

Equation to Demonstrate when ACS is More Efficient than SRS

Thompson (1990) uses the following equation to demonstrate when adaptive cluster sampling is more efficient than SRS. ACS is more efficient than SRS when:

$$Var(\tilde{y}|v) > Var(\hat{\mu}|n)$$

Note: v is the expected final sample size and n is the initial sample size; thus,

$$\sum_{k=1}^{k}\sum_{i\in\Psi}(\frac{N-n}{Nn}) > \sigma^2(\frac{1}{n}-\frac{1}{v})$$

$$\frac{\sum_{k=1}^{k}\sum_{i\in\Psi}(y_i-w_i)^2}{N-1}(\frac{1-\frac{n}{N}}{1-\frac{n}{v}}) > \sigma^2$$

where $\frac{\sum_{k=1}^{k}\sum_{i\in\Psi}(y_i-w_i)^2}{N-1}$ is a measure of within-network variance σ^2_{wn}. Then

$$\sigma^2_{wn}(\frac{1-\frac{n}{N}}{1-\frac{n}{v}}) > \sigma^2$$

Recall, Brown (2003) explains the aim of an adaptive cluster sample should be small networks so the final sample size is not excessively greater than the initial sample size, but networks should not be so small that the within-network variance is too low. As discussed earlier, one way to achieve small networks is with a smaller neighborhood definition; however, if the networks are very small the disadvantage of the relative small sample size of σ^2_{wn} to σ^2 outweighs the advantage of a final sample size approximately equal to the initial sample size. The same can be said for a large critical value. A large critical value that results with very small network sizes may lead to σ^2_{wn} being too small relative to σ^2 even though $\frac{n}{v} \approx 1$.

Application of ACS to Lake Erie Smelt

The following application of ACS to Lake Erie smelt, taken from Conners (2002), highlights real problems researchers face when implementing ACS as well as what can be done to overcome these issues. In 1998, ACS was used to sample Lake Erie smelt. The trial demonstrated the feasibility of ACS; however, problems did arise. The design did not include a four-adjacent cell neighborhood definition due to its impracticality in a hydroacoustic survey.

Instead, the neighborhood consisted of parallel transect segments by using Loran navigation lines as approximate parallels. Adaptive units for ACS were segments of parallel transects over the same latitudes that meet the criterion C. Out of 24 units in the initial sampled segments, 4 met the ACS criterion of a density greater than 5000 smelt/ha. Adaptive transect segments were surveyed on either side of the initial transect over the latitude range of both "patches". Efficiency of this ACS design was improved upon by using a survey vessel with a high traveling speed when not sampling. There were locations where addition of adaptive units was halted due to boundary on one of the sampled stratums and because of approaching daylight.

The greatest concern for this trial arose when the detection of a large patch resulted in a large final sample size. The researchers acknowledge the need for a "best" neighborhood definition which would help ease the problem of a large final sample size. Large final sample sizes can substantially increase the cost of a survey and potentially surpass a budget or schedule. Methods to limit final sample sizes, mentioned earlier, would be beneficial for future designs like this one.

Conclusion

Many populations exist in few, high-density areas such as the Lake Erie smelt. Increasing sampling in these areas through ACS sampling can increase sampling effort and provide efficient estimation through appropriate choice in design type, estimation, and design factors such as critical value, neighborhood choice, unit size, and sample size. The decision on when to use ACS is more challenging especially if there is no prior knowledge about the population of interest. If there is knowledge, the researcher should first make sure the population is sufficiently rare and clustered for ACS to be efficient and practical.

Appendix

R Code for Example

```
N<-100; n<-15; v<-15
x<-c(4,1,1,1,1,1,1,1,1,1,1,1,1,1,13); y<-c(21,0,0,0,0,0,0,0,0,0,0,0,0,0,57)

w<-y/x #mean quadrat abundance for cluster i

#Hansen-Hurwitz
meanhh<-sum(w)/n
var.mean.hh<-((N-n)/(N*n*(n-1)))*(sum((w-meanhh)^2))
meanhh
tauhh<-N*meanhh
tauhh
var.mean.hh
var.total.hh<-N*(var.mean.hh)
var.total.hh

#Horvitz-Thompson
alphak<-1-(choose(N-x,n)/choose(N,n)) #inclusion probabilities without replacement
%alphak<-1-((1-(x[k]/N))^n) #inclusion with replacement
alphakh<-matrix(0,nrow=v,ncol=v) #joint inclusion probabilities

for(k in 1:(v-1)){
for(h in (k+1):v){
alphakh[k,h]<-alphakh[h,k]<-1-(((choose(N-x[k],n)+choose(N-x[h],n)-
choose(N-x[h]-x[k],n)))/(choose(N,n)))#without replacement
}}
%alphakh[k,h]<-alphakh[h,k]<-1-(((1-(x[k]/N))^n)+((1-(x[h]/N))^n)-
%((1-(x[k]+x[h]/N))^n))#with replacement
%alphakh[k,h]<-alphakh[h,k]<-1-(((choose(N-x[k],n)+choose(N-x[h],n)-
%choose(N-x[h]-x[k],n)))/(choose(N,n)))#without replacement

tauhat<-sum(y/alphak)
muhat.2<-tauhat/N
```

```
term1<-sum((1-alphak)*y^2/alphak^2)#first term of variance of tauhat
term2<-0 #second term of variance
for(k in 1:(v-1)){
for(h in (k+1):v){
term2<-term2+2*((alphakh[k,h]-alphak[k]*alphak[h])/(alphak[k]*alphak[h]))
}}

var.tauhat<-term1+term2
var.muhat.ht<-(1/N^2)*var.tauhat #variance of mu for HT
c(tauhat,sqrt(var.tauhat)) #estimate of total and SE
c(muhat.2,sqrt(var.muhat.ht)) # estimate of mean and SE
```

References

Blair, Johnny. (1999) A Probability Sample of Gay Urban Males: The Use of Two-Phase Adaptive Sampling. *Journal of Sex Research.* 36:1, 39-44.

Brown, J.A. (2003) Designing an Efficient Adaptive Cluster Sample. *Enironmental and Ecological Statistics.* 10: 95-105.

Brown, J.A. and Manly, B.J.F. (1998) Restricted Adaptive Cluster Sampling. *Environmental and Ecological Statistics.* 5: 49-63.

Christman, M.C. and Lan, F. (2001) Inverse Adaptive Cluster Sampling. *Biometrics.* 57: 1096-1105.

Conners, E.M. and Schwager, S.J. (2002) The Use of Adaptive Cluster Sampling for Hydroacoustic Surveys. *ICES Journal of Marine Science.* 59: 1314-1325.

Philippi, Thomas. (2005) Adaptive Cluster Sampling For Estimation of Abundances Within Local Populations of Low-Abundance Plants. *Ecology.* pp. 1091-1100.

Seber, G.A.F. and Thompson, S.K. (1994) Environmental adaptive sampling. *Handbook of Statistics.* Volume 12: *Environmental Statistics*, pp. 201-220.

Salehi, M.M. and Seber, G.A.F. (1997) Adaptive cluster sampling with networks selected without replacement. *Biometrika.* 84: 209-219.

Smith, D.R., Villella, R.F. and Lemarié. (2003) Application of Adaptive Cluster Sampling to Low-Density Populations of Freshwater Mussels. *Environmental and Ecological Statistics.* 10: 7-15.

Thompson, Steven. (1997) Adaptive Sampling in Behavioral Surveys. *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates.* pp. 296-311.

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

Thompson, S.K. (1991b) Stratified Adaptive Cluster Sampling. *Biometrika.* 78: 389-397.

Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive Sampling.* John Wiley & Sons.

Thompson, S.K. and Seber, G.A.F. (1994) Detectability in conventional and adaptive sampling. *Biometrics.* 50: 712-724.

Turk, P. and Borkowski, J.J. (2005) A Review of Adaptive Cluster Sampling. *Environmental and Ecological Statistics.* 12: 55-94.