Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Adaptive cluster sampling, a quasi Bayesian approach

by Glen Meeden and Muhammad Nouman Qureshi

Release date: December 20, 2024

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2024
Vol. 50, No. 2, pp. 209-234
Statistics Canada, Catalogue No. 12-001-X

Adaptive cluster sampling, a quasi Bayesian approach

Glen Meeden and Muhammad Nouman Qureshi[1]

Abstract

Adaptive cluster sampling designs were proposed as a method that could be used when sampling rare populations whose units tend to appear in clusters. The resulting estimator is not based on any model assumptions and is design unbiased. It can have smaller variance than the standard estimator which does not incorporate the fact that one is dealing with a rare population. Here we will demonstrate that, when adaptive cluster sampling is appropriate, its estimator does not take into account all the available information in the design. We present a quasi Bayesian approach which incorporates the information which is now ignored. We will see that the resulting estimator is a significant improvement over the current methods.

Key Words: Adaptive cluster sampling; Bayesian inference; Finite population sampling; Prior information.

1. Introduction

Consider the problem of estimating the total number of a plant or animal species that live in a specified geographical area where the area has been partitioned into a collection of equally sized squares. Furthermore, assume that the species of interest is rare in this area so that most of the squares will contain none of the species. In addition assume that the few squares that do contain some of the species tend to cluster together in just a few neighborhoods of adjacent squares.

For this problem Thompson (1990) introduced the notion of adaptive cluster sampling, ACS. An initial simple random sample of squares is taken and the number of the species in each selected square is observed. Most of these observed counts will typically be zero. But whenever the count in a square is greater than zero, the adjacent squares, those to the left, right, above and below are added to the sample. When any of these squares have a count greater than zero then all of its unobserved adjacent squares are also observed. This process is continued until we obtain a set of contiguous nonempty squares surrounded by empty squares. A set of contiguous nonempty squares is called a network and the surrounding empty squares its edges. By definition an empty square is a network of size one. For this adaptive cluster sampling plan the usual estimator of the population total based on the counts in all the observed squares will be biased upwards. Thompson (1990) developed an unbiased estimator for the population total along with an estimator of its variance. More detail along with more references can be found in Thompson (2012). Many field researchers have adopted various versions of adaptive cluster sampling and it is used in a variety of disciplines. Turk and Borkowski (2005) describes several such examples. Latpate, Kshirsagar, Gupta and Chandra (2021) discuss some modifications of the standard ACS approach.

In the Bayesian approach to survey sampling prior information about the population of interest is incorporated in a prior distribution. After units in a sample have been observed inferences about the population are based on the posterior distribution of the unobserved units given the observed units. Moreover, this posterior distribution does not depend on how the units in the sample were selected. The

1. Glen Meeden, Emeritus, School of Statistics, University of Minnesota, Minneapolis, MN 55455. E-mail: gmeeden@umn.edu; Muhammad Nouman Qureshi, School of Statistics, University of Minnesota, Minneapolis, MN 55455. E-mail: qures089@umn.edu.

details of this approach have been described by Basu (Ghosh, 1988). Three Bayesian approaches to adaptive cluster sampling are given in Rapley and Welsh (2008), Pacifici, Reich, Dorazio and Conroy (2016) and Goncalves and Moura (2016). In these approaches the authors construct a Bayesian model for possible populations consistent with the assumptions underlying adaptive cluster sampling. Nolau, Goncalves and Pereira (2022) introduce a Bayesian model which includes auxiliary variables which could include additional information about the count in a square.

Given an ACS sample, our goal is to find a point estimator and an upper bound for the total number of the species in the population which have good frequentist properties. Since we are considering rare species we will assume that the ratio of the number of units with counts greater than zero to the total number of units in the population is small. We will break the problem into two parts. First we specify a prior distribution for the number of squares or units in the population with counts greater than zero, say θ an unknown parameter. This prior will reflect our assumption that the number of such units is small. Given the sample and our prior we have *a posterior* for θ. So our first step is to simulate a possible value for θ, say $\hat{\theta}$. Then conditional on $\hat{\theta}$, we find an estimate for the total of all counts greater than zero by using a distribution which assumes exchangeability among all the observed and unobserved counts greater than zero. This distribution does not arise from a prior distribution but is specified after the ACS sample has been observed. This is not a standard Bayesian procedure because this second "posterior" distribution does not follow from any prior distribution defined on the unknown finite population. This explains the terminology "quasi Bayesian" in our title. Two other recent examples where inferences are based on pseudo posterior distributions are given in Si, Pliiai and Gelman (2015) and Savitsky and Toth (2014). This makes sense when there is information in the sampling design which cannot be incorporated into a prior distribution. But we then combine these two distributions to simulate complete copies of the unknown population. We will see that the resulting point and interval estimators of the population total have better frequentist properties than the standard ACS estimators.

In Section 2 we briefly review the adaptive cluster sampling approach and outline our approach to the problem. In Section 3 we explain our approach in detail and present our estimators. We developed our estimator by doing simulations on a set of six populations where ACS sampling would be appropriate. In Section 4 we describe these six populations. In Section 5 we present simulations to compare our approach to the standard ACS approach. This is done for the six populations in Section 4 and six new populations which were not used in the development of our method. In Section 6 we discuss some possible extensions when more prior information is available about the population of interest. Section 7 contains some concluding remarks.

2. Adaptive cluster sampling

2.1 The basic setup

We begin by introducing some notation. We assume that the unknown population is a rectangular area consisting of N_r by N_c squares or units. So $N = N_r \times N_c$ is the population size. For integers (i, j) where $1 \le i \le N_r$ and $1 \le j \le N_c$ let $y_{i,j}$, denote the number of the species in the $(i\text{th}, j\text{th})$ square. Note $y_{i,j}$ is a

non negative integer. Let Y denote the matrix of the $y_{i,j}$ values. Given a square, the neighbors of a square are those squares just above and below it and those squares just to the right and left of it, with obvious modifications for squares on the boundary of the population.

In adaptive cluster sampling for each square in the initial random sample with a y value greater than zero all its neighbors are also observed and if any of them have a y value greater than zero then their neighbors are also observed and so on. This process is continued until only zero values are observed. For a given square the resulting set of squares found this way with y values greater than zero is called a **network**. Hence a network consists of a set of non zero squares with the property that if one of them appears in the sample then all of them will. The set of squares with a y value of zero that were observed in the process are called the edges of a network. As we noted in the introduction the network for a square whose y value is zero is just itself. For square (i, j) let $\Psi_{i,j}$ be all squares in its network.

Suppose now an initial simple random sample without replacement of size n_1 is taken. For $k = 1, \ldots, n_1$ let Ψ_{i_k,j_k} denote the network of the square which appears on the k th draw of the sample. Note that if two squares, which are in the same network, are in the first sample then their networks are identical and they each will be included when estimating the population total. For each k we let m_k be the number of squares in Ψ_{i_k,j_k} and $\bar{y}_k^*$ the mean of the counts for the units that appear in Ψ_{i_k,j_k}. For ACS (adapted cluster) sampling Thompson (1990) gave an unbiased estimator of the population total and an unbiased estimator of its variance. Given an ACS sample this estimate, $\hat{T}_{ac}$ and its estimated variance, $\hat{v}_{ac}$ are given by

$$\hat{T}_{ac} = N\frac{\sum_{k=1}^{n_1}\bar{y}_k^*}{n_1} \quad \text{and} \quad \hat{v}_{ac} = \frac{N(N-n_1)}{n_1}\frac{\sum_{i=k}^{n_1}(\bar{y}_k^* - \hat{T}_{ac}/N)^2}{n_1 - 1}. \tag{2.1}$$

We see that the ACS estimators, point and interval, depend only on the network means. This means that the variability of counts within a given network plays no role. In effect it is assuming that all the counts within a given network are the same. In addition, the fact that there were just a few squares with counts greater than zero in the population seems to play no explicit role in the inference stage after the ACS sample has been selected.

There is an alternative way to think about the ACS estimator in the above equation which is described in Dryver and Chao (2007). They consider a second but related version of the population. To form this second population they proceed as follows. For each network in the population we replace each y value by the average of all the y values in that network. If a network contains just one square then its y value is unchanged. But if a network contains more than one square then each of its y values are changed to the mean of all the y values of the squares making up the network.

Clearly this alternative population has the same total as the original population. Moreover, the observed y_k values in the second population, for the units in the initial sample, are identical to the $\bar{y}_k^*$ in equation (2.1). This makes it clear that the ACS estimator is an unbiased estimator but it also makes it clear that at the inferential stage the ACS estimator does not make use of the fact that we are sampling from a rare population.

Finally, some might consider it surprising that $\hat{T}_{ac}$ depends on N and that it includes the networks with a mean of zero. But a similar thing happens when estimating a domain total when the domain size is unknown and one has a random sample from the entire population. See for example the discussion on domain estimation in Cochran (1977).

2.2 A new approach

The ACS approach is based on two basic assumptions; there are only a few squares with counts greater than zero and these squares tend to be grouped in clusters. Although the sampling design is based on these two assumptions, as we have just noted at the end of previous section, the ACS estimator never seems to make use of the information that the proportion of squares with counts greater than zero is small at the inference stage. One should be able to do better if one incorporates this information when constructing an estimator.

Let D_b be the set of all the units in the population with counts greater than zero and let θ be the number units in D_b. For us θ is an unknown parameter and it will play an important role in what follows. Let T_b be the total of all the units in D_b. Of course T_b is also the total of all the units in the population. But we introduce this notation to emphasize the fact that our approach focuses on D_b. We break the problem into two parts. In the first part we find an estimate for θ using the information contained in the initial random sample of size n_1. Given this estimate and all the counts in the observed networks we then find an estimate for T_b.

As far as we know the notion of rare has never been explicitly defined in the literature. In some sense it is the analog for a few items of the Sorites paradox for many items. One version of which is "How many stones are needed to be considered a pile?". If we remove a single stone from a pile it should still be considered a pile but if we repeat this enough times we will no longer have a pile. Similarly if a species is rare in our population adding one positive count to a zero square would not change our perception that it is rare. But if we add enough positive counts then the species would no longer be rare. Let K_θ be the largest integer less than or equal to $N/10$. We will begin by assuming that our parameter space for θ is the set of integers

$$\Theta = \{i\colon i = 0, i = 1, \ldots, i = K_\theta\}. \tag{2.2}$$

One could argue that this choice is somewhat arbitrary but it is consistent with the notion of rarity and it holds for many of the examples considered in the literature. Later we will see that we can relax this assumption.

We do not make any other assumptions about the population except that the squares or units with counts greater than zero tend to appear in clumps and form networks. Where the networks are located in the population plays no role when ACS sampling is used. We make no explicit assumption about the range of possible counts. Anyone who wants to use ACS sampling and assumes that the number of positive counts is rare, as we described in the previous paragraph, could use our approach.

To develop our estimator we considered six possible populations. Three of them have appeared in the literature and the other three we constructed. They will be described in detail in Section 4 when we discuss our simulation results. Our estimator was found by trial and error by doing simulations from these six fixed populations.

It needs to be emphasized that finding sensible point and interval estimators for T_b is not an easy problem when θ is small. In such cases an ACS sample may contain zero or one or two counts greater than zero. In such cases we believe, finding sensible estimators is only possible when one has additional prior information. But here we are assuming that no such information is in hand. In the spirit of our objective approach we decided to ignore such samples in our simulation studies.

In the next section we will present our estimators and will explain the underlying logic and intuition that led to them.

3. A new estimator

Recall that D_b is the subset of Y consisting of all the squares with a y value greater than zero. If T is the population total of Y then T is equal to T_b, the total of the units belonging to D_b. Our goal is to estimate T_b. Remember that θ denotes the size of D_b, the number of $y_{i,j}$'s in the population which are greater than zero. For us θ is an unknown parameter and we take as its parameter space, Θ, the set of integers which are greater than or equal to zero and less than or equal to K_θ.

Let X be the number of units that belong to D_b in the first simple random sample without replacement of size n_1. Then X has a hypergeometric distribution which depends on N, n_1 and the unknown parameter θ. Given $X = x$ we can use the resulting likelihood function when estimating θ. Let y_b be the values of all the units in the ACS sample with values greater than zero. Let n_b be the number of units in y_b. Note that $n_b \geq x$. Let $y_{b'}$ be the remaining members of D_b which were not observed in the ACS sample. Note that if $\theta = n_b$ then $y_{b'}$ is the empty set.

From the Bayesian perspective one can break the problem of estimating T_b into two stages. First one simulates a possible value for θ from its posterior. Then given this value, one simulates possible counts for the $\theta - n_b \geq 1$ set of possible values for the unobserved values making up $y_{b'}$. Combining this simulated set of values with the observed y_b we have generated one complete copy of D_b and its corresponding value of T_b. Repeating these two steps many times we can use the resulting totals, in a Bayesian way, to find a point estimate and upper bound for the total of the units belonging to D_b.

We will now present our prior distribution for θ and explain how we simulate complete copies of D_b after the ACS sample has been observed.

3.1 Estimating the number of counts greater than zero

Given $X = x$ one natural estimate of θ is the maximum likelihood estimate. But one sees in simulations that the likelihood function tends to give too much weight to larger values of θ for the smaller values of x

we are likely to observe with ACS sampling. An alternative would be the standard Bayesian approach where one could specify a prior distribution over Θ. One possible non informative prior would be the uniform distribution over Θ, defined in equation (2.2). Its prior expectation is approximately $N/20$ which could be a good default guess for the value of θ. However, simulations demonstrate that it has the same weakness as the maximum likelihood estimator. They both tend to overestimate θ for ACS samples.

Our goal was to find a good default distribution for θ that will work well for a variety of populations where ACS sampling would be used. Rather than putting our distribution on Θ we considered distributions on Θ/N, the population proportion of $y_{i,j}$'s greater than zero. Any such distribution can then be thought of as a distribution on θ.

Recall that the beta distribution with parameters $\alpha > 0$ and $\beta > 0$ is defined on the unit interval and its density function is given by

$$f_{\alpha,\beta}(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1} \quad \text{for} \quad 0 \le z \le 1. \tag{3.1}$$

If for any choice of the parameters α and β we take the values of the above function, for all the members of θ in Θ, and then divided them by their sum the resulting set of numbers is a probability distribution defined on Θ. After much experimentation on our six test populations we chose as our prior distribution on Θ the re-normalized beta density with $\alpha = 1$ and $\beta = 90$.

To help understand this choice consider the case when $N = 400$. The prior mean for our choice is 3.92 while the choice of $\alpha = \beta = 1$ has a prior mean of 20. When the true value of θ/N is around 0.02 or 0.03 the later choice will give an over estimate of the size of θ. Our choice will help to decrease the upward bias of the likelihood function for the initial random sample. This will be discussed in more detail when we present the simulations in Section 5.

This is related to a similar problem with the ACS estimator $\hat{T}_{ac}$ (defined in equation (2.1)) that we saw in our simulations. Under the ACS design an important property of $\hat{T}_{ac}$ is that it is an unbiased estimator of the population total. If n_1, the size of the original random sample, is small and θ is small then there is a non-trivial probability that one observes samples where each count is zero. For such samples, $\hat{T}_{ac}$ gives an estimate of zero for the population total. This will be an underestimate except when the population total is in fact zero. To compensate for this underestimate, we saw that $\hat{T}_{ac}$ overestimates when the sample contains "lots" of units with counts greater than zero. It gets closest to the true value of θ when the number of units in the sample, with counts greater than zero, is close to $\theta/2$.

We need to emphasize that our estimator of θ does not depend on the fact that our basic units are squares and that the population is a rectangle. In addition it is not based on any assumptions about the form or shape of the networks or the values of the counts within a network. It depends only on the fact that we have a random sample from the basic units of the population. It should work whenever the basic units are approximately of the same size and the notion of neighbor can be defined in a sensible fashion. The overall form of the configuration of the population is immaterial. In addition it is not based on any assumptions about the values in y_b, the counts in the ACS sample greater than zero.

3.2 Estimating the population total

Given a value for θ, say $\hat{\theta}$, generated from our posterior distribution, and n_b the number of observed counts in y_b we need to define a distribution which simulates possible values for the remaining $\hat{\theta} - n_b \geq 1$ units whose counts must be greater than zero. We are assuming that little is known about the shape of the networks and how the size of a network may affect its y values. In this case a simple assumption is to assume that y_b, all the counts greater than zero in the full ACS sample, is approximately a "representative" sample of the values in D_b. Under this assumption a sensible way to simulate the unobserved members of D_b is to use Polya sampling.

Specifically, place n_b balls into an urn where each ball represents one of the counts belonging to y_b. Give each of the balls the weight $w > 0$. Pick a ball at random from the urn. Return it to the urn along with another ball which is assigned the count of the selected ball. This new ball is given the weight one. Now another ball is selected from the urn, which now has $n_b + 1$ balls, with probability proportional to their weights. The selected ball is returned to the urn along with another ball whose count is equal to that of the selected ball. This new ball is given a weight of one and the urn now contains $n_b + 2$ balls. This process is continued until the urn contains $\hat{\theta}$ balls. Under this distribution the expected value to the total of all the counts in the simulated urn does not depend on w and is $\hat{\theta}\bar{y}_b$, where $\bar{y}_b$ is the mean of the counts in y_b. The variance, however, does depend on w and decreases as w increases. It is demonstrated in Meeden (1999), (equation 2.5), that the variance is given by

$$\text{Var}(T_b \mid y_b, n_b, \hat{\theta}, w) = \hat{\theta}(\hat{\theta} - n_b)\frac{\text{var}(y_b)}{n_b}\frac{n_b - 1}{1 + n_b w}\frac{\hat{\theta} + n_b w - n_b}{\hat{\theta}} \tag{3.2}$$

where $\text{var}(y_b)$ is the sample variance of the counts in y_b.

Now we need to specify a value for w, the weight of each ball in the urn at the beginning of this process. Note that w only appears in the last two fractions in equation (3.2). It is easy to check that the value of the product to these two fractions is one if we take as our value of w

$$w^* = \frac{n_b(\hat{\theta} - n_b + 1) - 2\hat{\theta}}{n_b(\hat{\theta} - n_b + 1)}. \tag{3.3}$$

It is easy to check that $w^* > 0$ when $n_b > 2$.

With this choice of w^* and for a fixed y_b and fixed value of $\hat{\theta}$ our conditional variance is given by

$$\text{Var}(T_b \mid y_b, n_b, \hat{\theta}, w^*) = \hat{\theta}(\hat{\theta} - n_b)\frac{\text{var}(y_b)}{n_b}. \tag{3.4}$$

This reflects our assumption that we are viewing the values in y_b as exchangeable and arising from something approximating a random sample. Now if n_b is reasonably large, say around 20, and y_b is approximately a random sample then the above should be a reasonably good estimate of variance. But for smaller values of n_b, say less than five, it will not work well. In the following we will describe a way to handle this problem.

But first let $\hat{T}_{ab}$ denote our quasi or approximate Bayes estimator of T_b, the total number of the species in the population, that is based on our two stage simulation procedure. Let $\overline{y}_b$ be the mean of all the units in y_b. Then our estimate of T_b is

$$\hat{T}_{ab} = E(T_b) = E(E(T_b \mid \theta)) = E(\theta \overline{y}_b) = \hat{\theta}_{1.90} \overline{y}_b \tag{3.5}$$

where $\hat{\theta}_{1.90}$ is the mean of our posterior distribution for θ.

To find the variance of $\hat{T}_{ab}$ we use the well known formula that a variance can be written as the variance of a conditional expectation plus the expectation of a conditional variance. So for a fixed ACS sample we have that

$$\begin{aligned} \mathrm{Var}(\hat{T}_{ab}) &= \mathrm{Var}\Big(E(\hat{T}_{ab} \mid \theta)\Big) + E\Big(\mathrm{Var}(\hat{T}_{ab} \mid \theta)\Big) \\ &= \mathrm{Var}(\theta \overline{y}_b) + E\left(\theta(\theta - n_b)\frac{\mathrm{var}(y_b)}{n_b}\right) \\ &= \overline{y}_b^2 V(\theta) + \Big(E(\theta^2) - n_b E(\theta)\Big)\frac{\mathrm{var}(y_b)}{n_b}. \end{aligned} \tag{3.6}$$

We still need to find a way to use the variance of our estimator to produce a good interval estimator of the total number of the species in the population. In ACS sampling the initial random sample will consist of mostly zero counts. Recall x is the number of counts greater than zero in the initial sample of size n_1 and n_b is the number of units in the full ACS sample with counts greater than zero. Note that x is less than or equal to n_b. A very naive upper bound would be our point estimate plus the product of 1.96 and the square root of $\mathrm{Var}(\hat{T}_{ab})$. But, as we have already noted, it is no surprise that this works poorly because in ACS sampling the values of x and n_b can be quite small.

Consider a case where a population contains one or two networks of size one or two with counts much larger than the rest of the counts in the population. For such populations, when x is small, these networks are unlikely to be observed and our estimate of variance will be too small. To help protect against this possibility we need to increase the naive upper bound given just above, especially when n_b is small. To this end we let

$$\lambda = 10^{(2.5/n_b)}. \tag{3.7}$$

Note that λ will decrease as n_b increases. We are not claiming that this is an optimal choice for adjusting upwards our upper bound. We just found that it seemed to work well for our test populations.

When calculating upper bounds for our estimate we will use

$$\hat{T}_{ab} + \lambda \times 1.96 \sqrt{\mathrm{Var}(\hat{T}_{ab})} \tag{3.8}$$

as our approximate 95% upper confidence bound. Because of the small sample sizes in ACS sampling assuming that $\hat{T}_{ab}$ has, approximately, a normal distribution might seem surprising. But we will see in the

simulations that this seems to work reasonably well for most cases where ACS sampling is appropriate because of our choice for λ.

The second part of our two stage simulation process is clearly not Bayesian since our choices for w^* and λ depend on n_b which comes from the observed data . It does not arise from some prior distribution although the fact that the initial ACS was a random sample is important information for us. As we noted in the Introduction two other examples where "posterior" distributions are defined without a prior distribution and which are based in part on the sampling design can be found in Si et al. (2015) and Savitsky and Toth (2014). This seems to make sense when there is available prior information that cannot be incorporated into a prior distribution.

A sensible lower bound for T_b is just the sum of all the ACS sample counts greater than zero. For the ACS estimator we will use the same lower bound and for its upper bound we will use its estimate plus the product of 1.96 and its estimated variance. Again assuming that the ACS estimator has, approximately, a normal distribution.

One might object that our assumption that y_b is approximately a "representative" sample of the values in D_b is too strong. But recall that at the end of Section 2.1 we saw that the only information used in the ACS estimator is the mean of the counts in a network and by considering the alternative population of Dryver and Chao (2007), we saw that the ACS estimator is essentially assuming exchangeability among the network means. We believe that all the observed counts can contain some additional information and that our assumption that the observed sample is approximately a representative sample is no stronger than assuming the exchangeability among the network means.

The way we found our estimator was by doing simulations on a set of six populations where ACS sampling would be appropriate. Three of them had appeared in the literature and the other three we constructed. We then used simulation studies to see how different choices of a prior and an estimated variance would work. What we have described above is the best that we have found. We found others that worked almost as well but these are our best choices at this time. In the next section we describe these six populations.

4. The populations

In our simulations we used three different populations that have appeared in the literature. One is the first example discussed in Thompson (1990). He presents it as a typical example where ACS sampling could be used. He gives no further details so presumable it was constructed by him. The population has 400 units with three networks whose sizes are 6, 11 and 4 and whose means are 6.0, 9.73 and 11.75 respectively. The mean of all the counts greater than zero is 9.4 and the largest count is 39 which appears in the network of size 11. We denote this population by thmp. Gattone, Mohamed and Di Battista (2016) describe two samples

of African buffalo and African hartebeest taken in 2010. We denote these two populations by afrbuf and afrhart.

We also constructed three more populations. To construct a population we considered a grid of points on the surface of the earth. Assume that their latitudes and longitudes are equally spaced, although the successive differences in the two directions need not be the same. Depending on the topography of the location of the points, their altitudes, measured in meters, can exhibit the clumping behavior which is the underlying assumption of ACS sampling.

To find the altitude on a grid we used the function, *elevation*, in the R (R Core Team, 2023) package rgbif (Chamberlain, Ram, Mcglinn and Barve, 2019). To get the final set of "counts" we did three things. First we rounded every altitude in the set to its closest integer value. Next we choose an $\epsilon > 0,$ but close to zero, and found the $1-\epsilon$ percent quantile of our set of integer values, say q_ϵ. We set every count less than q_ϵ to zero. Finally, we subtracted some integer, less than $q_\epsilon,$ from every count greater than zero. The number of resulting counts greater than zero would then be small enough to represent counts of a rare species. For example, if we set $\epsilon = 0.05$ the resulting set of counts will have five percent of it's values greater than zero. So in the resulting population, the counts or the $y_{i,j}$ values, are either zero if their rounded altitude falls below a certain level or the difference between their rounded altitude and some constant. Depending on the topography of an area, the resulting grid of "counts" can exhibit the clumping behavior that makes ACS sampling a sensible choice. This is a flexible method that allows one to find many realistic populations to use in simulation studies of ACS sampling.

Next we describe the three grids we used to construct the three populations used in our simulation studies. For the first we chose a grid in Paris, France, the second a grid at Niagara Falls, on the border between Canada and the United States and the third a grid near Devil's Tower in the western United States. For the first two we used a 23 by 23 grid while for the third we used a 26 by 45 grid. We denote these three populations by paris, nfalls and devt. Summary information for the six populations are give in Table 4.1. Note the proportion of counts greater than zero range from 0.038 to 0.084. Three dimensional plots of the populations are given in the six figures on the next two pages.

Table 4.1
Summary information for the populations used in the simulations.

Population	N	N_{ntw}	θ	θ/N	$y_{\max}$	T_b	T_b/θ
afrbuf	391	5	15	0.038	99	334	22.3
nfalls	529	5	20	0.038	59	368	18.4
thmp	400	3	21	0.053	39	190	9.4
devt	1,170	10	85	0.073	34	868	10.2
paris	529	6	43	0.081	63	1,112	25.9
afrhart	391	9	33	0.084	20	171	5.18

Notes: Recall that N is the population size, θ is the number of units greater than zero and T_b is their sum. We let N_{ntw} be the number of networks in the population and $y_{\max}$ be the maximum y value in the population.

Figure 4.1 A three dimensional plot of the population afrbuf.

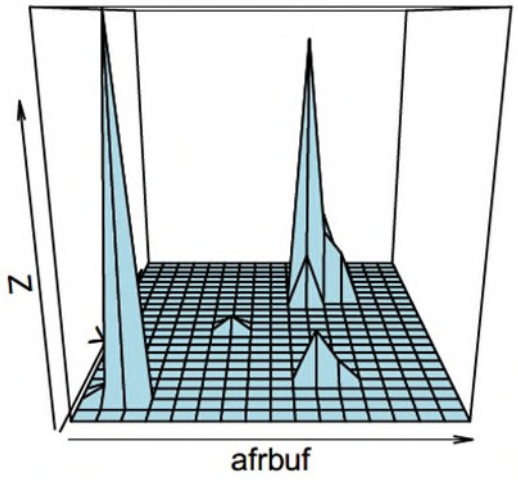


Figure 4.2 A three dimensional plot of the population nfalls.

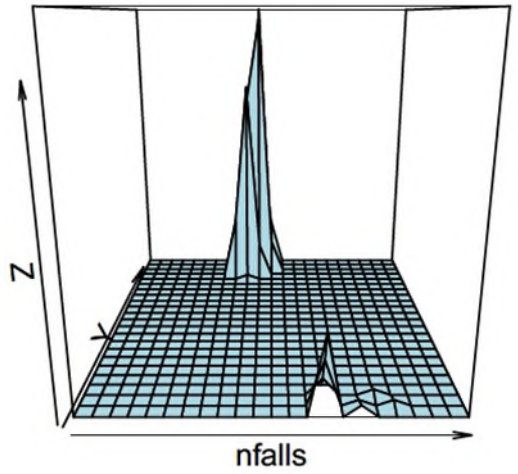


Figure 4.3 A three dimensional plot of the population thmp.

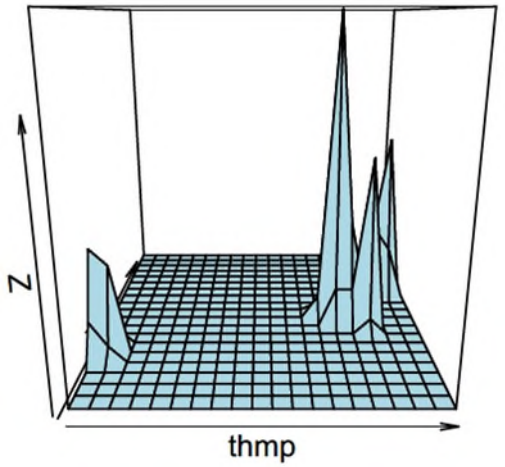


Figure 4.4 A three dimensional plot of the population devt.

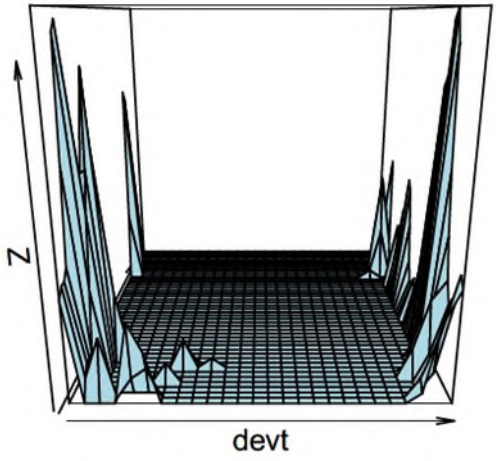


Figure 4.5 A three dimensional plot of the population paris.

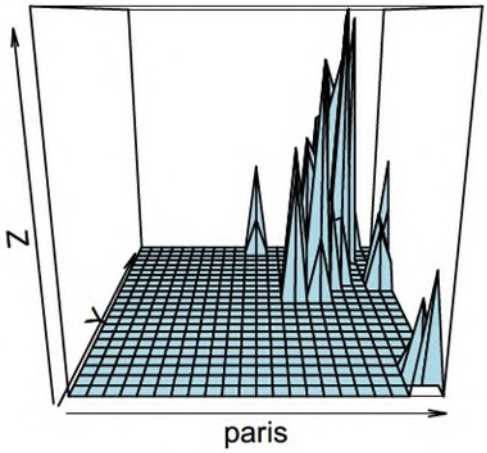
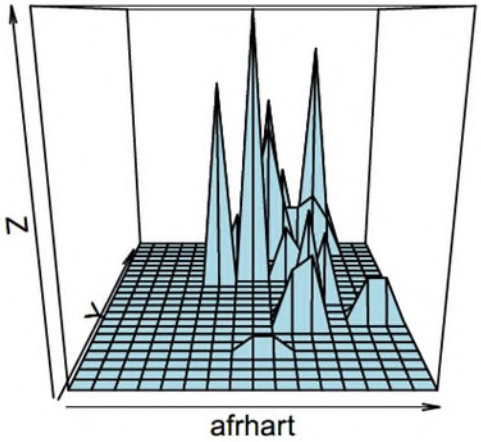


Figure 4.6 A three dimensional plot of the population afrhart.



5. The simulations

5.1 Doing the simulations

Near the end of Section 2.1 we explained why we would only consider ACS samples which had at least three counts greater than zero. For this reason, in our simulations, we will only consider samples where $n_b > 2$. In other words, our frequency of coverage is a conditional one; it is conditional on seeing at least three counts greater than zero. So the results given in the Tables 5.1 through 5.8 and 5.10 through 5.13 for the adaptive cluster sampling estimator and our quasi Bayes estimator, ACS and BAY respectively, are conditional. For each method the tables give the average value of an estimator, Est, its average relative bias, Rbias, its average absolute error, ABerr, the average lower bound of its interval estimate, Lowbd, the average length of its interval, Len, and the frequency which its upper bound was larger than T_b.

In our adaptive cluster sampling simulations we used two different initial sample sizes; ten percent and twenty percent of the population size. Here we will present just the ten percent results. How the two methods compare is essentially the same for the two different initial sample sizes but of course they both do better for the larger initial sample size.

Finally, someone might be concerned about what would happen if the true size of D_b was slightly bigger than $N/10$ and hence violating our definition of a rare species. In practice it is unlikely that the observed n_b would be bigger than this bound. But to allow for this possibility, when doing the simulations, we defined our prior on the integers between $0.01\,N$ and $0.15\,N$. The results hardly differ from what happens if the set of possible integers are the non negative integers less than or equal to $0.1\,N$. This happens because as we move away from n_b the posterior decreases very quickly and points far out in the tail contribute little probability. So even though we developed our "posterior" with the smaller upper bound in mind it works almost as well with the much larger and imprecise upper bound. Hence to use our method one does not need a good guess for the upper bound of θ.

5.2 Results for the six populations

There are several things to note about the simulation results given in Tables 5.1 through 5.6. For all the populations the BAY estimator has smaller average absolute error than the ACS estimator, sometimes dramatically so. Overall the BAY upper bounds perform better than the ACS upper bounds. The population afrbuf is the only case where the ACS upper bound is superior. The BAY upper bound is too large. For population nfalls the two methods behave about the same. For populations thmp, devt and paris the BAY upper bound is clearly the best. The only population where its frequency of coverage falls below 0.90 is population afrhart. In this case its frequency of coverage is only 0.892 but its upper bound was much smaller than the ACS bound which had a frequency of coverage of 0.861. Some of the information in the tables is presented graphically in Figure 5.1.

Table 5.1
Population afrbuf.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	489	0.464	187.2	195	974	0.98
BAY	334	-0.000	69.3	195	1,644	0.98

Notes: For 1,000 simple random samples of size 39, there were 660 samples with at least three counts greater than zero. The average total number of positive values observed was 6.96.

Table 5.2
Population nfalls.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	448	0.217	198.9	319	706	0.924
BAY	447	0.214	126.5	319	724	0.998

Notes: For 1,000 simple random samples of size 53, there were 819 samples with at least three counts greater than zero. The average total number of positive values observed was 13.57.

Table 5.3
Population thmp.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	211	0.112	87.9	118	341	0.921
BAY	157	-0.173	45.7	118	274	1.000

Notes: For 1,000 simple random samples of size 40, there were 887 samples with at least three counts greater than zero. The average total number of positive values observed was 12.98.

Table 5.4
Population devt.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	874	0.007	224	723	708	0.942
BAY	857	-0.012	89	723	548	0.972

Notes: For 1,000 simple random samples of size 117, there were 1,000 samples with at least three counts greater than zero. The average total number of positive values observed was 65.98.

Table 5.5
Population paris.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	1,119	0.006	389.5	970	1,119	0.909
BAY	1,114	0.002	69.2	969	603	0.984

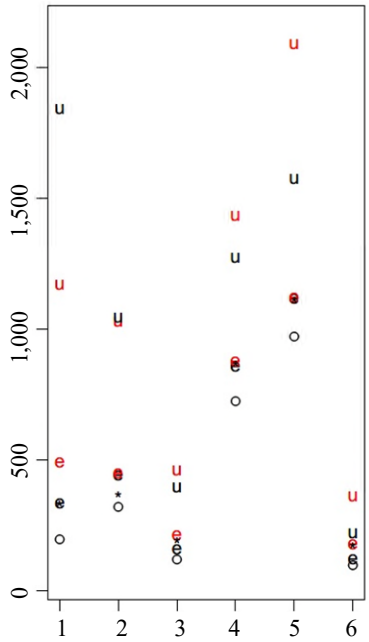
Notes: For 1,000 simple random samples of size 53, there were 983 samples with at least three counts greater than zero. The average total number of positive values observed was 34.19.

Table 5.6
Population afrhart.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	175	0.021	78.4	96	261	0.861
BAY	119	-0.305	52.4	96	120	0.892

Notes: For 951 simple random samples of size 39, there were 968 samples with at least three counts greater than zero. The average total number of positive values observed was 18.74.

Figure 5.1 The above represents, graphically, some of the information in Tables 5.1 through 5.6 for the six populations afrbuf(1), nfalls(2), thmp(3), devt(4), paris(5) and aftrhart(6).



Notes: The number in parentheses is their location on the horizontal axis. The true population total is denoted by $\star$. The ACS estimator and its upper bound are denoted by e and u. For the BAY estimator these quantities are in the color black and along with the common lower bound, o.

For the six populations in Table 4.1 the one for which the BAY estimator had the most extreme negative bias, -0.305, was population afrhart. Looking more closely at afrhart we see it has 33 units greater than zero in 9 networks with an average value of 5.18. The three largest values in the population are 17, 15 and 20. These latter two each appear in a network of size one while the first appears in the largest network which has 16 units. Most of the rest of the values in the population are 5 or less. The mean of these remaining 30 units is 3.97. With the choice of $\alpha = 1$ and $\beta = 90$ the average value of our estimator of $\theta = 33$ was 23.5. This helps to explain the large negative bias of our estimator for this population. One way to improve its performance would be to increase the estimate of θ by a different choice of the prior distribution. More generally, having good prior information about the size of θ can lead to improved results.

It is not surprising that the ACS estimator is biased upwards. This is because we are ignoring all ACS samples with less than three counts greater than zero. If they were included then the ACS estimator would always be unbiased. On the other hand our BAY estimator can be both biased upwards or downwards. It can be biased upwards when we over estimate the size of D_b. It can be biased downwards when we underestimate the size of D_b or when there is a very small network whose counts are much larger than the counts in the remaining networks. But this latter case is also a problem for the ACS estimator. We will discuss this in more detail in Section 5.3.

We see from Table 4.1 that population paris has 43 units greater than zero in 6 networks with an average value of 25.9. Checking the networks we find that just one network contains 31 units whose average value

is 30.0. We see from Table 5.5 that on the average we observed 39.4 units. This means the ACS sample almost always contained the units in this network. This helps to explain our excellent results for this population.

A similar thing happens with population devt. In this population there are 10 networks containing 85 units with counts greater than zero. The total of these units is 868. The two biggest networks contain 31 and 23 units respectively and their respective means are 11.3 and 12.8. The next biggest network contains 13 units and all the rest have 5 or less units. From Table 5.4 we see that the average number of units in the final ACS samples was 65. This means that most of the ACS samples contained the two largest networks. This is not surprising, but the fact that these two largest networks are good representative samples of D_b explains why our estimators perform so well for this population.

From Table 4.1 we see that for our six populations the ratio θ/N, the proportion of units with positive counts in the population, ranges from 0.038 to 0.084. Recall that given an ACS sample, the adjusted factor λ, defined in equation (3.7), was introduced to increase our estimate of variance when the number of counts greater than zero in the ACS sample was quite small. For populations afrbuf and nfalls, the populations with the smallest values of the ratio θ/N, the average value of λ was 1.42 and 1.20 respectively while for the two largest populations, devt and paris, these averages were 1.01 and 1.02 respectively. This shows that our choice of the λ is working as expected.

5.3 Six more populations

As we have mentioned, we settled on our estimator by simulation studies on these six populations. They were chosen because they represent a variety of situations where ACS sampling would be used. A reader might be worried that our estimator was too dependent on the populations we used in our study even though they are quite different. In an attempt to show that this is not the case we now present six new populations that were not used in the development of our estimator. We constructed the first two populations on a grid of 400 squares with just two networks. The first network had three members with values 50, 60 and 70. The second network had twelve members with the values 14, 15 and 16 each appearing four times. Let ref1 denote this population. Let ref2 denote the population where 14, 15 and 16 appear in the smaller network and 50, 60 and 70 each appear four times in the larger network. Tables 5.7 and 5.8 give the results for 1,000 samples of size 40 for the two populations. The same set of samples was generated for both populations.

We see that in the 789 samples with at least two counts greater than 0 the average number of observed counts was 11.67. This means that in the majority of these samples only the larger network was observed. Note however that our results for population ref1 are quite good, even though such samples only contain the small counts. Our average estimate of θ was 15.93, a slight overestimate. Because we have observed the larger network this helps to compensate for the fact that only the units with the smaller count size are in the sample. While for population ref2, having all the larger counts in the larger network, means that our estimator is biased upwards but less so than the ACS estimate.

Table 5.7
Population ref1.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	464	0.290	269.6	223	892	1
BAY	330	-0.084	129.8	223	704	1

Notes: For 1,000 simple random samples of size 40, there were 789 samples with at least three counts greater than zero. The average total number of positive values observed was 11.67.

Table 5.8
Population ref2.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	957	0.251	444	653	1,574	0.885
BAY	865	0.131	244	653	962	1.000

Notes: For 1,000 simple random samples of size 40, there were 789 samples with at least three counts greater than zero. The average total number of positive values observed was 11.67.

Reasoning similarly to the discussion of population ref2 just given we can explain the upward bias of our estimator for population nfalls in Table 5.2. This population has seven networks containing a total of $\theta = 20$ units. Here one network contains 13 of the units and it contains all the largest units as well. The average number of observed counts greater than zero was 13.57 so that the network containing 13 units was almost always in the sample. The average value of our estimator for θ was 19.5, which is quite good. But it cannot compensate for the fact that the ACS samples almost always include the larger counts.

Remember that $\hat{T}_{ac}$ would be unbiased if we used all possible samples and not just those with at least three counts greater than zero. When all the observed counts are zero then it will estimate zero. This event will happen with positive probability and will be an underestimate except when the population total is in fact zero. To compensate for this underestimate, $\hat{T}_{ac}$ overestimates when the sample contains "lots" of neighborhoods with counts greater than zero. This also helps to explain why BAY does better than ACS.

To see this we will look more closely at the two populations ref1 and ref2. First consider the three cases where only the second network was either observed once or twice or three times. For population ref1, whose total is 360, the corresponding estimates BAY(ACS) ranged from 226(150) to 257(450). Next consider the three cases where the first network was observed just once and the second network was observed either once or twice or three times. For these three cases the estimates ranged 446(750) to 500(1,050). For population ref2 whose population total is 765 the corresponding values of the estimates for the first three cases ranged from 903(600) to 1,029(1,800) and for the second set of cases the corresponding values of the estimates ranged from were 949(750) to 1,063(1,950). Note that the range of values for the ACS estimator is much larger than the range for the BAY estimator. This also explains why in all our simulations $\hat{T}_{ac}$ is biased upwards. It is because we are ignoring all the samples with less than three counts greater than zero.

We will now describe the remaining four new populations, mich, rome, afrdeer and sunspt. Summary information about the four populations are given in Table 5.9. Plots of the four populations are given in Figures 5.2 through 5.5.

Table 5.9
Summary information for the new populations rome, mich, afrdeer and sunspt.

Population	N	N_{ntw}	θ	θ/N	$y_{\max}$	T_b	T_b/θ
rome	1,600	16	48	0.03	495	7,927	165.1
mich	400	10	20	0.05	3,577	24,740	1,237
afrdeer	391	13	76	0.19	140	1,309	16.6
sunspt	500	14	27	0.05	98	434	16.1

Notes: Recall that N is the population size, θ is the number of units greater than zero and T_b is their sum. We let N_{ntw} be the number of networks in the population and $y_{\max}$ be the maximum y value in the population.

Figure 5.2 A three dimensional plot of the population rome.

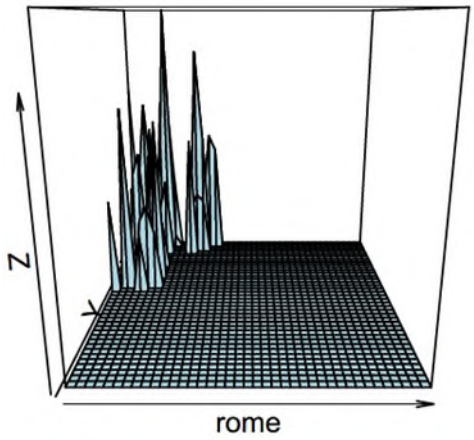


Figure 5.3 A three dimensional plot of the population mich.

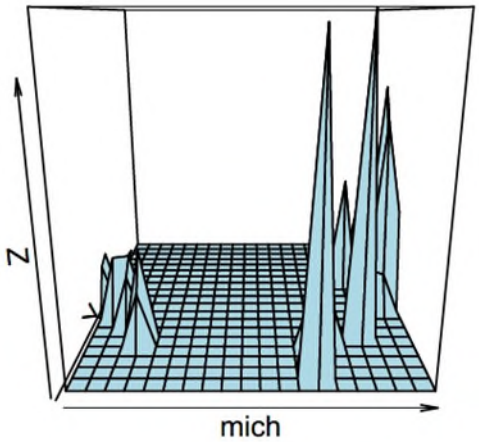


Figure 5.4 A three dimensional plot of the population afrdeer.

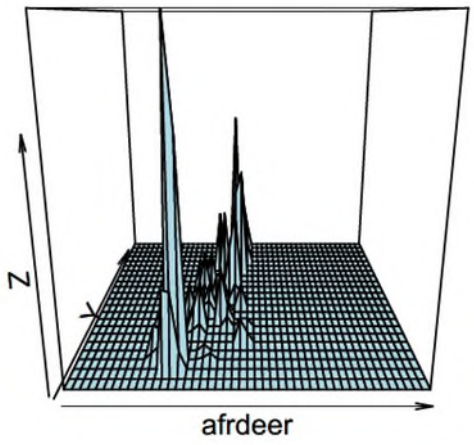
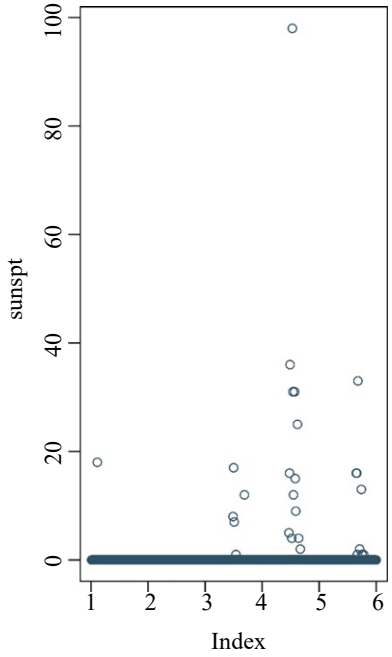


Figure 5.5 A plot of the population sunspt.



The first population is based on a 40 by 40 grid in Rome, Italy. The second is based on a 20 by 20 grid in the state of Michigan in the United States. We denote these two populations by rome and mich respectively

For population rome we see that there are 16 networks which contain 48 counts greater than zero. Four of the networks, that range in size from 6 to 11, contain 34 of the 48 counts greater than zero. The means of these networks range from 139.3 to 219. The mean of these four means is 176.8 which is quite close to 165.1, the mean of all the counts greater than zero. The remaining 12 networks contain either one or two values and most of them quite small. The two largest are 335 and 301. This explains the excellent behavior

of our estimator BAY in Table 5.10. With high probability the ACS sample will contain one of the four larger networks.

Table 5.10
Population rome.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	8,206	0.035	2,793	4,044	11,107	0.923
BAY	7,459	-0.059	1,731	4,044	12,673	1.000

Notes: For 1,000 simple random samples of size 160, there were 986 samples with at least three counts greater than zero. The average total number of positive values observed was 22.81.

For population mich we see that there are 10 networks containing 20 counts greater than zero. Of these counts, 11 are in 6 networks where the maximum count is 695 and the six smallest range from 105 to 190. The largest network contains 5 counts and its mean is 969.2. The largest count is 3,077 and is in a network of size one. So although there are just a few units with counts greater than zero the actual counts can be quite large. We include this example to see what could happen when this is the case. The large negative bias of the BAY estimator in Table 5.11 occurs because the largest count is a network of size one. Even with this bias it still has significantly smaller average absolute error than the ACS estimator. Both of their upper bounds perform poorly however. It is difficult to get a sensible upper bound when there is one network with one or two values which are significantly larger than the rest of the counts in the population. Both will underestimate if the big counts are not included in the sample and overestimate when they are. Unless there is additional prior information we believe that it is very difficult to find sensible estimates without approximate exchangeability across the counts in the networks when we only observe a very small number of counts greater than zero.

Table 5.11
Population mich.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	32,657	0.320	14,333	8,065	63,114	0.964
BAY	16,805	-0.321	8,602	8,065	64,281	1.000

Notes: For 1,000 simple random samples of size 40, there were 686 samples with at least three counts greater than zero. The average total number of positive values observed was 6.49.

For the next new population we consider a third population given in Gattone et al. (2016). We let afrdeer denote this population and summary information is given in Table 5.9. The majority of the units belong to the three largest networks of sizes 13, 17 and 21. Their respective means are 29.3, 8.8 and 25.8. The next largest remaining network is of size 5 and the remaining values are mostly small. The largest value in these other networks is 34. Note that for this population the true value of θ is 76 so that the ratio $\theta/N = 76/391 = 0.19$ is greater than 0.15 the upper bound used in defining our prior.

So the natural question is how does our estimator work in this case? The results of this simulation are given in Table 5.12. We see that as in the other examples we do reasonably well and better than the ACS estimator. This happens because the ACS samples tend to include the larger networks and the smaller networks tend to have smaller counts.

Table 5.12
Population afrdeer.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	1,307	-0.001	406	962	1,304	0.917
BAY	1,213	-0.073	214	962	1,269	0.970

Notes: For 1,000 simple random samples of size 39, there were 1,000 samples with at least three counts greater than zero. The average total number of positive values observed was 49.76.

At the end of Section 3.1 we noted that our posterior distribution for θ does not depend on the fact that our basic units are squares and that the population is a rectangle. But it is important that the basic units be approximately the same size and that neighbors can be defined in a sensible manner. To demonstrate this point we now consider a population defined on a set of successive intervals of equal length. The neighbors of an interval are just the two adjacent intervals to the left and to the right except the end points which have only one neighbor. To get the counts we use the population of sunspots that is available in R (R Core Team, 2023). These are the monthly mean relative sunspot numbers from 1,749 to 1,983 and they have the clumping behavior which should make ACS sampling appropriate. We took the first 500 values of sunspots and subtracted 141 from each one. The resulting negative numbers were set to zero. Summary information for this new population, sunspt, is given in Table 5.9. In sunspt there are 14 networks: 8 of size 1, 1 of size 2, 3 of size 3 and 2 of size 4. so theta = 27. The range of the 27 counts greater than zero range from 1 to 98. The next largest count is 36. The largest count appears in a network of size four with a mean of 36.25 which is quite a bit larger than $T_b/\theta =$ 16.1. We see in Table 5.13 that the behavior of the two estimators is very similar to what we have seen in our two dimensional examples. This example demonstrates that our approach could be useful when studying longitudinal data which satisfies our assumptions of rareness and clumping.

Table 5.13
Population sunspt.

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ACS	497	0.144	217.8	140	930	0.952
BAY	304	-0.299	159.3	140	1,369	0.999

Notes: For 1,000 simple random samples of size 50, there were 874 samples with at least three counts greater than zero. The average total number of positive values observed was 7.29.

Three dimensional plots of the populations mich, rome and afrdeer are given in Figures 5.2, 5.3 and 5.4. A plot of sunspt is given in Figure 5.5.

6. R code for calculating our estimator

Here we argued for a particular quasi Bayes estimator when little is known about the population of interest except that the species of interest is rare. But in fact we have defined a whole family of possible estimators by considering different choices of the parameters in the beta density and different choices for the definition of the parameter space Θ. Any of these possible estimators is very simple to compute in R (R Core Team, 2023). What follows below is an R function which calculates one of these estimates for the population total, along with an upper bound and the lower bound. One just needs to specify the parameter space Θ, the beta distribution which one uses to define their prior distribution on Θ and some of the information from a full ACS sample. One does not need to know the edges of the observed networks.

More formally here is what you need to specify;

- n, the size of the initial random sample,
- nbiginrs, the number of counts greater than zero in the initial random sample,
- bigsmp, all the counts greater than zero in the full ACS sample,
- bds, used to define the parameter space Θ,
- alp and bet, the parameters for the beta distribution used to define our prior distribution,
- N, the population size.

Next we give the code for the function

```
qbay<-function(n,nbiginrs,bigsmp,bds,alp,bet,N)
  {
     klw<-floor(bds[1]*N)
     kup<-ceiling(bds[2]*N)
     theta<-klw:kup
     nbigsmp<-length(bigsmp)
     dtheta<-theta[theta>=nbigsmp]
     ntheta<-length(dtheta)
     llike<-lchoose(dtheta,nbiginrs) + lchoose(N-dtheta,n-nbiginrs)
     lprior<-log(dbeta(dtheta/N,alp,bet))
     dum<-lprior + llike
     post<-rep(0,ntheta)
     for(i in 1:ntheta){
       post[i]<-1/sum(exp(dum-dum[i]))
      }
     pstmnth<-sum(dtheta*post)
     est<-pstmnth*mean(bigsmp)#the point estimate in equation (5.6)
     pst2ndmnth<-sum(dtheta^2*post)
     pstvrth<-pst2ndmnth - pstmnth^2
```

```
    nbig<-length(bigsmp)
    mnbg<-mean(bigsmp)
    lwbd<-nbig*mnbg
    vr1<-(mnbg)^2*pstvrth
    nb<-length(bigsmp)
    d1<-var(bigsmp)
    d2<-sum(post*(dtheta-nb)^2)
    vr2<-d1*d2
    vr<-vr1+vr2              #the variance in equation (5.7)
    lam<-10^(2.5/nbig)       #the adjustment factor in equation (5.8)
    upbd<-est+sqrt(vr)*lam*1.96
    ans<-c(est,lwbd,upbd)
    return(ans)
}
```

This function allows one to explore a better choice of a quasi Bayes estimator when one has some additional prior information about the population. Note that one can also easily change our adjustment factor, λ, that appears in equation (3.7). In the next section we will briefly discuss some possible extensions.

7. Possible extensions

Thompson (1990) introduced ACS sampling for biological applications where one was interested in a rare species that tended to appear in clumps. As we noted in Section 2.2 we could not find in the literature any definition of rare. But as our simulations demonstrated our approach can work well for a fairly wide range of rareness. Our prior for the number of counts greater than zero does not depend on how large they are. From one point of view this is sensible because the notion of rareness can depend on the species under study. In some cases one might have good prior information about the number of counts greater than zero. In this case one could select different values of α and β in the prior distribution that better reflects this information.

As far as we know there is no formal definition of clumping in the literature. An extreme form of clumping would be if only one unit contains the total T_b elements of the species of interest. But this is clearly not in the spirit of ACS sampling. It seems to us that ACS sampling presumes that such a large concentrated count is not possible. Instead it assumes that such a large count would tend to spread out into neighboring squares forming a network. As we saw in our description of population thmp at the beginning of Section 4 its three networks are of this form. When the size of the networks tend to be larger and counts within a network are representative of all the counts in the population we will have good results.

When this is not true and there is a large count, compared to the rest, in a small network then both the ACS and BAY upper bounds perform poorly. We saw this for populations afrbuf and mich. Our choice of

the adjustment factor λ in equation (3.7) is based on an implicit clumping assumption about the range of possible values for the counts and on how likely it is to have a small network with extremely large counts. In some cases one could have information about what is a "big" count for the population and the proportion on such "big" counts in the population. Depending on the size of the counts in the ACS sample one could let the adjustment factor, λ, depend on the value of the counts in the sample and the prior information. One could also weaken our assumption of approximate exchangeability and use prior information to replace the mean of the observed counts by a different estimator. Prior information could improve our estimator but will result in poorer estimators when it is incorrect. How to make use of additional prior knowledge in a more formal manner needs further study.

Rapley and Welsh (2008) developed interesting models for the study of the type of populations where ACS sampling is used. In particular they modeled how networks and edges could be formed. Our approach is much simpler because we are ignoring the networks and only consider the number of nonempty units in the population. Our suggested prior distribution for the size of the D_b is very similar to a prior that is used in a slightly different context in their model. Here we have been interested in showing that one can improve on the standard ACS approach without making any model assumptions. It would be interesting to compare our approach to Bayesian approaches to the problem. How much better a Bayesian approach can do than what we have presented here when one has in hand good prior information is a question that needs further study.

Rapley and Welsh also noted that the sampling could be done sequentially. That is, a unit is selected at random from the population. If its count is greater than zero then one observes all the units in its network and its edges. At each step we only selected one unit at random from the remaining unobserved units. We continue in this way until our stopping rule tells us the sampling is finished. Our approach can be extended to such sequential sampling plans. Assuming that the order in which the sample was taken is known then the form of the likelihood function would change but one could use the same prior distribution for θ. From a theoretical point of view sequential sampling makes a lot of sense but it is not clear if it would be practical in many of the problems where ACS sampling would be used.

8. Concluding remarks

Adaptive cluster sampling was introduced to improve sampling efficiency for populations with a special type of structure. It is appropriate when the statistician knows that the population has just a few cells with a y value greater than zero and that they tend to appear in clusters. It is an interesting approach which has been widely adopted when studying biological populations in the field. We showed however that in its focus on finding an estimator which is design unbiased it has ignored some of the available information. Here we introduced a quasi Bayesian approach to the problem that makes use of this information. We showed that our point estimator and our 95% upper confidence bound for the population total gave much better results than the standard approach.

Acknowledgements

We wish to thank Gabriel Mersy for his help with using the *R* package rgbif.

References

Chamberlain, S., Ram, K., Mcglinn, D. and Barve, V. (2019). *rgbif: Interface to the Global Biodiversity Information Facility API*. https://CRAN.R-project.org/package=rgbif.

Cochran, W. (1977). *Sampling Techniques* (third ed.). New York: John Wiley & Sons, Inc.

Dryver, A., and Chao, C. (2007). Ratio estimators in adaptive cluster sampling. *Environmetrics*, 18, 607-620.

Gattone, S., Mohamed, E. and Di Battista, T. (2016). Adaptive cluster sampling with clusters selected without replacement and stopping rule. *Environmental and Ecological Statistics*, 23, 453-468.

Ghosh, J.K. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays by D. Basu*. New York: Springer-Verlag.

Goncalves, K.C.M., and Moura, F.A.S. (2016). A mixture model for rare and clustered populations under adaptive cluster sampling. *Bayesian Analysis*, 11, 519-544.

Latpate, R., Kshirsagar, J., Gupta, V.K. and Chandra, G. (2021). Adaptive cluster sampling. *Advanced Samping Methods*, 125-156. Springer Singapore.

Meeden, G. (1999). Interval estimators for the population mean for skewed distributions with a small sample size. *Journal of Applied Statistics*, 26, 81-96.

Nolau, I., Goncalves, K.C.M. and Pereira, J.B.M. (2022). Model-based inference for rare and clustered populations from adaptive cluster sampling using auxiliary variables. *Journal of Survey Statistics and Methodology*, 10, 439-465.

Pacifici, K., Reich, B., Dorazio, R. and Conroy, M. (2016). Occupancy estimation for rare species, using a spatially-adaptive design. *Methods in Ecology and Evolution*, 7, 285-293.

R Core Team (2023). *R: A Language and Evironment for Statistical Computing*. R Foundation for Statistical Computing. http://R-project.org.

Rapley, V.E., and Welsh, A.H. (2008). Model-based inferences from adaptive cluster sampling. *Bayesian Analysis*, 3, 717-736.

Savitsky, T., and Toth, D. (2014). Bayesian estimation under informative sampling. *Electronic Journal of Statistics*, 10, 1677-1708.

Si, Y., Pliiai, N. and Gelman, A. (2015). Bayesian nonparametric weighted sampling inference. *Bayesian Analysis*, 10, 605-625.

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

Thompson, S. (2012). *Sampling* (third ed.). Hoboken, New Jersey: Wiley.

Turk, P., and Borkowski, J. (2005). A review of adaptive cluster sampling: 1990-2003. *Ecological and Environmental Statistics*, 12, 55-94.