Adaptive Sampling in Real Life: Large Objects and Stopping Rules

Clive Orton

UCL Institute of Archaeology,
31-34 Gordon Square,
London, WC1H 0PY, U.K.
c.orton@ucl.ac.uk

Abstract. A new approach to spatial sampling, adaptive cluster sampling, differs from conventional sampling in that the procedure for selecting the sample depends on values observed during the survey. This paper considers some problems that may arise if this approach is used in archaeological fieldwork, at either regional or site level. Computer simulations based on two case studies are used to examine possible solutions to two problems: large 'objects' and a highly variable final sample size. They suggest that the former can be dealt with in a straightforward way. The latter problem seems to be less tractable; even the best of the approaches used here, restricted adaptive sampling with poststratification, reduces rather than overcomes the problem. Nevertheless, results obtained so far are sufficiently encouraging to suggest directions for further research.
Keywords: adaptive cluster sampling, archaeological field survey, archaeological excavation.

1 Background

A need has long been felt in 'field' sciences, such as ecology, for an approach to sampling that could, under certain circumstances, be more efficient than conventional approaches such as simple or stratified random sampling. The circumstances envisaged were those in which the population being studied tended to have a very 'patchy' distribution, for example shoals of fish. The theory for such an approach, called *adaptive sampling*, was built up through the late 1980s and the 1990s (for example, Thompson 1990), culminating in a comprehensive treatment of the subject (Thompson and Seber 1996).

The aim of this paper is to look at one aspect of this approach, *adaptive cluster sampling*, to see if it might have any value for archaeological sampling, at either a regional or a site level. Since the dangers of adopting techniques uncritically from other disciplines are well known (Aldenderfer 1987: 90), a cautious approach is suggested: first thinking about the nature of archaeological data and the problems it might bring, then using computer simulation to assess the importance of these problems and to examine ways in which they might be overcome, and finally trying an adaptive approach in the field. This paper deals only with the first two stages; potential test-beds for field trials are being sought.

2 Adaptive Cluster Sampling

The basic idea behind adaptive sampling is that "the procedure for selecting the sample may depend on values of the variable of interest observed during the survey" (Thompson and Seber 1996: 1). In adaptive cluster sampling the sampling units are spatial ones (e.g. quadrats) and the variable of interest is the quantity of relatively rare 'objects', such as sites in a region or features in a site. This has an obvious archaeological appeal (Shennan 1997: 385-390).

The theory of this approach has already been summarised for an archaeological audience (Orton 2000: 34-38), but will be repeated here for convenience. In adaptive cluster sampling (Thompson and Seber 1996: 94-5), we first define a neighbourhood of units belonging to each unit: it might consist of every adjacent unit, every unit with a common edge, or some more complicated pattern (Fig. 1). The only restriction on the pattern is that if unit B is in the neighbourhood of unit A, then unit A must be in the neighbourhood of unit B (the *reflexive* property).

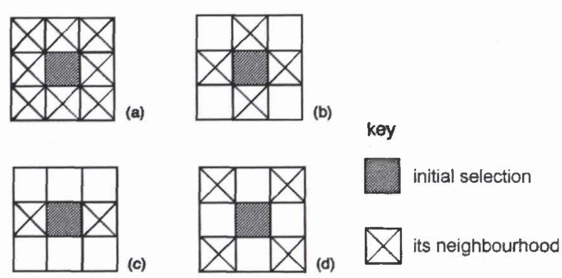

Fig. 1. Examples of neighbourhoods in adaptive sampling

We then survey an initial sample of units and, for every unit that meets a certain pre-assigned sampling condition (e.g. that it contains more than a certain number of objects), we additionally sample all the units in its neighbourhood. If any of these units meet the condition, we sample all the units in their neighbourhood, and so on until the process stops. If a unit does not meet the condition, we do not sample any additional units. The final sample consists of one or more clusters of units, each of which is bounded by a series of edge units that do not meet the condition (Fig. 2). Because edge units can belong to more than one cluster, Thompson and Seber (*ibid.*) define the *network* of a unit as all those units in the same cluster as it, excluding the edge units; any unit that does not meet the condition is defined as a network of size 1. Thus every unit belongs to just one network, and the selection of any unit in a network leads to the selection of all the units in that network.

The approach can be quite flexible; for example, one could combine an initial transect sample with additional quadrats added at points where the condition is met (*ibid.*:121-3) (Fig. 3).

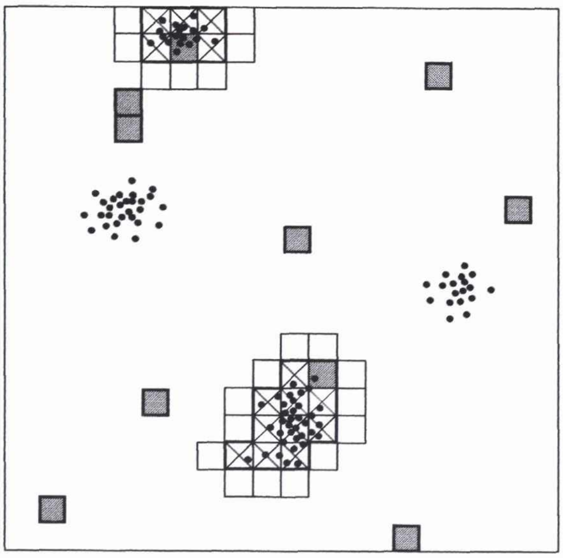

Fig. 2. Example of an adaptive sample

Having selected the sample, how do we estimate the number of objects in our study area? and what is the precision of this estimate? Formulae for an unbiased estimator and its variance are given below. We need some extra notation (*ibid.*:96):
the number of networks in the sample is κ,
the number of units in the *k*th network is x_k,
the value of the variable of interest (e.g. number of objects) in a unit is denoted by y,
and the sum of the y-values in the *k*th network is y_k*.
The number of units in the initial sample is n_1.
Then an unbiased estimate of the average per unit μ is

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{\kappa}\frac{y_k^*}{\alpha_k}$$

where $\alpha_k = 1 - [((N - x_k)!(N - x_k - n_1)!)/(N!(N - n_1)!)]$.

There is also a formula for its variance (*ibid.*:97).

We are interested in whether adaptive sampling can give greater precision for the same sample size (or better, for the same overall cost) as *conventional sampling* (for example, simple random sampling). The comparison depends on the characteristics of the population being studied, but some general criteria, indicating circumstances in which adaptive sampling is likely to be the more efficient, have been suggested (*ibid.*:159):
1. the population is clustered or tends to aggregate.
2. the number of units in the population is large compared to the number that satisfy the condition, i.e. the study area is big relative to the area in which sites (or other archaeological objects) are likely to be encountered.
3. the expected final sample size is not much larger than the initial sample size.
4. the costs of observing units in clusters or networks is less than the cost of observing the same number selected at random throughout the region.
5. the cost of observing units not satisfying the condition is less than the cost of observing units satisfying the condition.

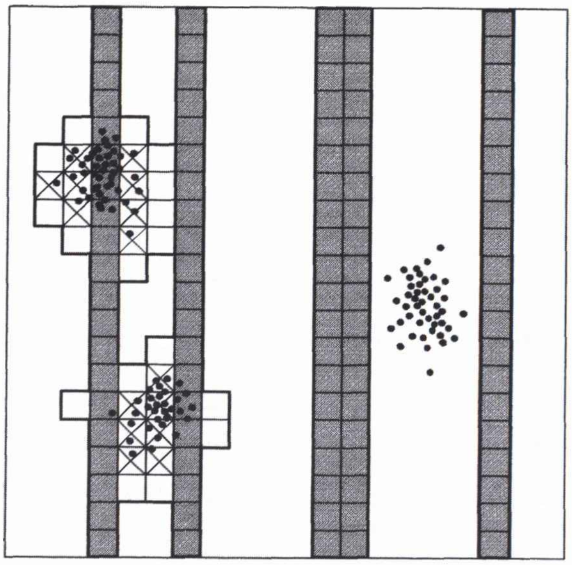

Fig. 3. Example of adaptive sampling, using initial transects

Since it is difficult to weigh up these criteria *a priori*, it is useful to undertake some simulation experiments to assess the relative performance of adaptive and conventional sampling strategies in real archaeological situations. Data on criteria 4 and 5 would be valuable, but are difficult to find. One might expect both criteria to hold, but the relative costs are impossible to assess without practical experiments.

3 Transfer to Archaeology: Potential Problems

Three problems come to mind immediately when one starts to think about applying this approach in archaeology:

1. *false negatives:* because the initial sample is smaller than the corresponding conventional sample, there is a greater chance of not finding *any* objects at all,
2. *large objects:* the theory assumes that each object is located entirely within a unit. In archaeology, this may not always be so. For example, a large site or a large feature may easily overlap two or more sampling units,
3. *open-ended sampling:* the final sample size depends on the initial selection, and cannot be predicted. This makes it difficult if not impossible to predict in detail the cost of, or the time needed for, a particular fieldwork task.

One can also think of possible solutions to these problems, but they need to be tested:

(a) *false negatives:* we need to specify acceptable chances of not finding any objects when certain proportions of the units contain them, using the binomial distribution,
(b) *large objects:* can be tackled by varying the sampling condition (see Valdeflores),
(c) *open-ended sampling:* we need a 'stopping rule' to limit the size of the final sample. This leads to *restricted adaptive sampling* (see Lasley Vore), in contrast to *unrestricted adaptive sampling.*

4 Case Studies

Two case studies were chosen, one for each of problems (b) and (c). They also represent regional survey and site excavation respectively. Computer simulation, using the spreadsheet program Excel, was used to compare various forms of adaptive sampling with simple random sampling in each case. Details of the simulation methods follow the case studies. Computer simulation was not needed for problem (a) because it can be approached algebraically.

4.1 Large Objects (Valdeflores Survey)

The Valdeflores Survey (Plog 1976) covered an area of 50.5 sq. km containing 33 known sites of various sizes (Orton 2000: Fig. 4.9). Plog experimented with simple random samples of 24 units of 0.5 by 0.5 km, and of 6 units of 1 by 1 km, using them to estimate the number of sites in the region, and calculating the standard deviations of such estimates.

Three adaptive sampling experiments were made (*ibid*:93-97), all with 0.5 km units, and with the neighbourhood of a unit defined as the four contiguous units (see Fig. 1(b)), but with different sampling conditions:

(a) initial sample size = 10, condition was to sample further if a unit contains at least part of a site. Ten simulations were run.
(b) initial sample size = 10, condition was to sample further if a unit contains at least one *complete* site. Ten simulations were run.
(c) initial sample size = 16, condition was to sample further if a unit contains at least one *complete* site. Twenty-five simulations were run.

In experiments 2 and 3, partially-sampled sites were counted as fractions. Sites which on the basis of evidence from the initial sample were likely to occupy two units were given a value of $y = ½$ for each sampled unit in which they occurred; ones likely to occupy four were given a value of $y = ¼$ for each sampled unit in which they occurred, and so on.

The outcomes of the three experiments are shown in Tables 1-3 respectively. The outcome of the first experiment was not satisfactory; it combined a large final sample size (an average of 39 units compared to Plog's 24) with a large s.d. (21, compared to Plog's best values of 13-14). The poor performance appears to be due to the large sites located in networks A and B (see Orton 2000:Fig. 4.9), which require a great deal of surveying but contribute relatively little to the number of sites.

Table 1. Valdeflores Survey, outcome of first adaptive sampling experiment

	min.	mean	max.	s.d.	'target'
estimated number of sites	6.4	25.9	70.4	21.4	33
no. of units in final sample	22		51	-	24
number of sites located	2		15		

Table 2. Valdeflores Survey, outcome of second adaptive sampling experiment

	min.	mean	max.	s.d.	'target'
estimated number of sites	5.1	25.9	61.1	20.1	33
no. of units in final sample	10	14.6	24	-	24
number of sites located	1	3.0	6		

Table 3. Valdeflores Survey, outcome of third adaptive sampling experiment

	min.	mean	max.	s.d.	'target'
estimated number of sites	7.4	33.1	70.4	15.3	33
no. of units in final sample	16	24.7	33	-	24
number of sites located	2	4.7	8		

The outcome of the second experiment was more satisfactory; it had a slightly smaller s.d. (20) on a sample less than 40% of the original final size (an average of 15 units). The main problem was that the final sample was too small.

The third experiment was scaled up from the second to give an appropriate final sample size for comparison with Plog's work. Its outcome was the most satisfactory of the three experiments, with a s.d. of 15 (slightly greater than Plog's best figures) and an average final sample size of 25.

The outcome of the third experiment is encouraging, since:

1. the average final sample size of about 25 units is probably cheaper than the target simple random sample of 24 units, because of saving in the costs of locating and travelling between units. This saving could be used to 'purchase' extra units and reduce the s.d. to perhaps about the same level as that of Plog's best simple random samples.
2. adaptive sampling shows to its best advantage when the 'objects' (sites) are highly clustered; in this example, the degree of clustering is relatively low, but nevertheless adaptive sampling seems to perform about as well as simple random sampling when an appropriate sampling condition is used. This suggests that gains could be expected if patterns that were more highly clustered were sampled.
3. the adaptive sample locates, on average, more sites per survey than the conventional method. This may be an important archaeological consideration.

4.2 Stopping Rules (Lasley Vore Site)

The Lasley Vore site (Odell 1992), has been modified, for the purposes of this and an earlier experiment (Orton 2000:133-135), to comprise an area of 120 m by 120 m, divided into 576 quadrats, each 5 m by 5 m (*ibid*:Fig. 5.12). Of the quadrats, 55 contain a total of 82 archaeological features; features have been moved slightly to lie inside single units. A neighbourhood was defined as the four units contiguous to a selected unit (Fig. 1(b)); the sampling condition was to continue sampling if a unit contained at least one feature.

Four sets of simulations were run:

1. simple random samples, with sample sizes of 40 and 30. Eighty and 106 simulations respectively.
2. unrestricted adaptive samples, initial sample size = 20. One hundred and sixty simulations.
3. restricted adaptive samples (Brown 1994). Select the initial sample one by one, sample it and its network before selecting the next. Stop (but only at the end of a network) as soon as the total exceeds a limit, here chosen to be 40. One hundred and sixty simulations.
4. *Poststratification* (Thompson and Seber 1996: 160). Select all the initial sample, but don't sample any yet. Design a sample route across the site. Survey each selected unit and its network in this order. When the number of units sampled plus the number of initial units remaining exceeds the limit (40 was chosen here), stop adaptive sampling and sample the rest of the initial sample conventionally. Divide both population and sample into two strata; stratum A sampled adaptively and stratum B sampled conventionally. Use appropriate formulae to estimate the numbers of objects in each stratum, and combine the two. One hundred and sixty simulations.

The outcomes of the four experiments are shown in Tables 4-7 respectively. The outcome of unrestricted adaptive sampling (Table 5) compares unfavourably with that of simple random sampling with $n = 40$ (Table 4); the s.d. is about 20% greater. A particular problem is the very large maximum final sample size (at 88, over twice the target figure).

Table 4. Lasley Vore site, outcome of simple random sampling

$n = 40$	min.	mean	max.	s.d.	'target'
estimated no. of features	0	82.8	230.4	51.3	82
number of features found	0	5.75	16		

proportion of samples in which no features were found = 0.0125
expected proportion (using binomial distribution) = 0.018

$n = 30$	min.	mean	max.	s.d.	'target'
estimated no. of features	0	85.1	288	66.6	82
number of features found	0	4.4	15		

proportion of samples in which no features were found = 0.0566
expected proportion (using binomial distribution) = 0.049

Table 5. Lasley Vore site, outcome of unrestricted adaptive sampling

	min.	mean	max.	s.d.	'target'
estimated no. of features	0	81.9	308.5	61.5	82
no. of units in final sample	20	38.9	88	14.8	40
number of features found	0	13.7	50	13.3	

proportion of samples in which no features were found = 0.144
expected proportion (using binomial distribution) = 0.134

Table 6. Lasley Vore site, outcome of restricted adaptive sampling (Brown's method)

	min.	mean	max.	s.d.	'target'
estimated no. of features	0	96.1	406.7	85.9	82
no. of units in final sample	20	34.6	62	9.4	40
number of features found	0	12.4	50	11.6	

proportion of samples in which no features were found = 0.144
expected proportion (using binomial distribution) = 0.134

Table 7. Lasley Vore site, outcome of restricted adaptive sampling (poststratification)

	min.	mean	max.	s.d.	'target'
estimated no. of features	0	82.5	276.4	66.0	82
no. of units in final sample	20	34.9	57	9.4	40
number of features found	0	12.9	40	12.0	

proportion of samples in which no features were found = 0.144
expected proportion (using binomial distribution) = 0.134

The first restricted method (Table 6) appears biased (mean of 96 compared to target of 82) and has a very large maximum value (over 400) and s.d. (86). This seems to arise because there are occasions when only a very small part of the initial sample is actually sampled (the fewest was 6 units). The maximum final sample size (62), although much smaller than in the unrestricted experiment, is still over 50% greater than the target.

The second restricted method (post-stratification, see Table 7) is the best of the three adaptive methods, with a mean close to the target and a s.d. of 66. The average final sample size is 35, but the maximum is still rather high at 57. In terms of the estimated number of features, this method is comparable to the simple random sample with $n = 30$.

From these experiments, it appears that some form of restriction on the final sample size is needed. Of the two methods used here, post-stratification is the better.

4.3 Note on Simulation Methods

Each experiment at Lasley Vore consisted of a separate Excel 5 workbook. It would have been possible to combine them, but this would have created a file too large to back up easily, and felt like too many eggs in one basket. Each workbook consisted of four sheets:

1. the site: one row for each unit, giving its location (easting and northing), unit code, number of features in the unit, network to which it belongs, and any networks (maximum of two) of which it is an edge unit.
2. the simulated excavations: one row for each unit sampled in each run, giving run number, unit code, number of features in the unit, network to which it belongs (both derived from Sheet 1 by a look-up function), the number of features in that network and the number of units in its cluster (both derived from Sheet 3 by a look-up function), the 'overlap' (see below), the contribution of that unit to the estimated number of features (also derived from Sheet 3 by a look-up function), and the final sample size. There are sub-total rows for each run. In the restricted samples, adjustments had to be made when the initial sample size was reduced to limit the final sample size.
3. the networks: one row for each network, giving network code, number of units, number of features, number of edge cells, initial sample size, alpha-value and the contribution to the estimated number of features if that network is sampled. For restricted adaptive sampling the rows are repeated for different initial samples sizes, since in some runs not all the initial sample is used.
4. summary results: one row for each run, giving run number, number of features found, estimated number of features and final sample size, extracted from Sheet 2.

The operation was relatively straightforward, if rather tedious. The main general problem encountered was that of 'overlaps' – units that belonged both to the initial sample and to the cluster of another unit in the initial sample. They were identified manually and removed from the calculations.

5 Discussion

Although the spatial patterns at Valdeflores and Lasley Vore are only weakly clustered, adaptive sampling seems to perform about as efficiently as simple random sampling in estimating the total number of objects present. It performs better in that it locates a higher proportion of the objects, especially when they are all 'small', as at Lasley Vore. Its drawbacks are that it results in a higher proportion of false negatives, and has a variable final sample size, even when attempts are made to limit it.

6 Future Work

This is a very preliminary study. Much more work needs to be done to establish whether adaptive sampling is a practical tool for archaeologists, and how it should be used. Further experiments are needed in the following areas:

1. alternative neighbourhoods; the use of diagonal neighbourhoods (Fig. 1 (d)) and neighbourhoods two units distant from the initial unit have both been suggested,
2. false negatives; the problem of the increased probability of a false negative outcome due to the smaller initial sample size needs to be faced,
3. more than one problem at a time; what happens when we have large objects and restricted adaptive sampling?
4. performance over a wider range of different types of sites and regions; the benefits of adaptive sampling over conventional sampling are site- (or region-) specific, and we need to be able to recognise the sorts of situations in which adaptive sampling is likely to be beneficial.

Finally, the approach will have to be tested in the field.

Acknowledgements

I am grateful to all those who contributed to the discussion at the Quantitative Methods session of CAA99.

References

Aldenderfer, M.S. 1987. On the structure of archaeological data. In: Aldenderfer, M.S. (ed.), *Quantitative Research in Archaeology*, 89-113, Newbury Park: Sage Publications.

Brown, J.A. 1994. The application of adaptive cluster sampling to ecological studies. In: Fletcher, D.J. and Manly, B.F.J. (eds) *Statistics in Ecology and Environmental Monitoring*, 86-97, Otago Conference Series No. 2, Dunedin, New Zealand: University of Otago.

Odell, G.H. 1992. Bewitched by mechanical site-testing devices, *American Antiquity* 57 (4), 692-703.

Orton, C.R. 2000. *Sampling in Archaeology*. Cambridge: Cambridge University Press.

Plog, S. 1976. Relative efficiencies of sampling techniques for archaeological surveys. In: Flannery, K.V. (ed.), *The Early Mesoamerican Village*, 136-158, New York: Academic Press.

Shennan, S.J. 1997. *Quantifying Archaeology* (2nd edn.). Edinburgh: Edinburgh University Press.

Thompson, S.K. 1990. Adaptive cluster sampling, *Journal American Statistical Association* 85, 1050-9.

Thompson, S.K. and Seber, G.A.F. 1996. *Adaptive Sampling*. New York: John Wiley and Sons.