1 Adaptive Sampling

George A. F. Seber

1.1 Introduction

Adaptive sampling is a method of unequal probability sampling whereby the selection of sampling units at any stage of the sampling process depends on information from the units already selected. In general terms, it means that if you find what you are looking for at a particular location, you sample in the vicinity of that location with the hope of obtaining even more information.

Methods of estimation were initially developed in the three pioneering papers of Thompson [23–25] and the sampling book by Thompson [26]. The material considered in this review is described briefly by Seber and Thompson [20], while full details are given in the book by Thompson and Seber [31].

1.2 Adaptive Cluster Sampling

Suppose we have a population spread over a large area that is highly clumped but is generally sparse or empty between clumps. If one selects a simple random sample (without replacement) of units, then most of the units selected will be empty. Density estimation based on this meager information will then have poor precision. Furthermore, if the population species is rare, we will get little physiological information about individuals. It would be better to begin with an initial sample and, if individuals are detected on one of the selected units, then sample the neighboring units of that unit as well. If further individuals are encountered on a unit in the neighborhood, then the neighborhood of that unit is also added to the sample, and so on, thus building up a cluster of units. We call this *adaptive cluster sampling*. If the initial sample includes a unit from a clump, then the rest of the clump will generally be sampled. Such an approach will give us a greater number of individuals.

As well as counting individuals, we may wish to measure some other characteristic of the unit, for example, plant biomass or pollution level, or even just note the presence or absence of some characteristic using an indicator variable. In addition to rare-species and pollution studies, we can envisage a wide range of populations that would benefit from adaptive sampling, for example, populations that form large aggregations such as fish, marine mammals, and shrimp. We can also add mineral deposits and rare infectious diseases in human populations (e.g., AIDS) to our list. Recently the method has been used in sampling houses for a rare characteristic [5] and

in sampling animal habitats [15].

To set out the steps involved in adaptive cluster sampling, we begin with a finite population of N units indexed by their "labels" $(1, 2, \ldots, N)$. With unit i is associated a variable of interest y_i for $i = 1, 2, \ldots, N$. The object is to select a sample, observe the y-values for the units in the sample, and then estimate some function of the population y-values such as the population total $\sum_{i=1}^{N} y_i = \tau$ or the population mean $\mu = \tau/N$.

The first step is to define, for each unit i, a neighborhood consisting of that unit and a set of "neighboring" units. For example, we could choose all the adjacent units with a common boundary, which, together with unit i, form a cross. Neighborhoods can be defined to have a variety of patterns; the units (plots) in a neighborhood do not have to be contiguous. However, they must have a *symmetry* property; that is, if unit j is in the neighborhood of unit i, then unit i is in the neighborhood of unit j. We assume, for the moment, that these neighborhoods do not depend on y_i.

The next step is to specify a condition C (for instance, $y > c$, where c is a specified constant). We now take an initial random sample of n_1 units selected with or without replacement from the N units in the population. Whenever the y-value of a unit i in the initial sample satisfies C, all units in the neighborhood of unit i are added to the sample. If in turn any of the added units satisfies the condition, still more units are added. The process is continued until a cluster of units is obtained that contains a "boundary" of units called *edge* units that do not satisfy C. If a unit selected in the initial sample does not satisfy C, then there is no augmentation and we have a cluster of size one. The process is demonstrated in Figure 1, where the units are plots and the neighborhoods form a cross. Here y_i is the number of animals on plot i, and $c = 0$, so that a neighborhood is added every time animals are found. In Figure 1a we see one of the initial plots that happens to contain one animal. As it is on the edge of a "clump," we see that the adaptive process leads to the cluster of plots in Figure 1b.

We note that even if the units in the initial sample are distinct, as in sampling without replacement, repeats can occur in the final sample, as clusters may overlap on their edge units or even coincide. For example, if two non edge units in the same cluster are selected in the initial sample, then that whole cluster occurs twice in the final sample. The final sample then consists of n_1 (not necessarily distinct) clusters, one for each unit selected in the initial sample.

1.3 Applications and Extensions

In applications, other methods are sometimes used for obtaining the initial sample. For instance, in forestry, the units are trees and these are usually selected by a method of unequal probability sampling, where the probability of selecting a tree is proportional to the basal area of a tree (the cross-sectional area of a tree at the basal height—usually 4.5 feet in the USA). Roesch [16] described a number of estimators for this situation.

In ecology, larger sample units other than single plots are often used. For example, a common sampling unit is the strip transect, which we might call the primary unit. In its adaptive modification, the strip would be divided up into smaller secondary units, and if we found animals in a secondary unit, we would sample units on either side of that unit, with still further searching if additional animals are sighted while on this search. Strips are widely used in both aerial and ship surveys of animals and marine mammals. Here the aircraft or vessel travels down a line (called a *line*

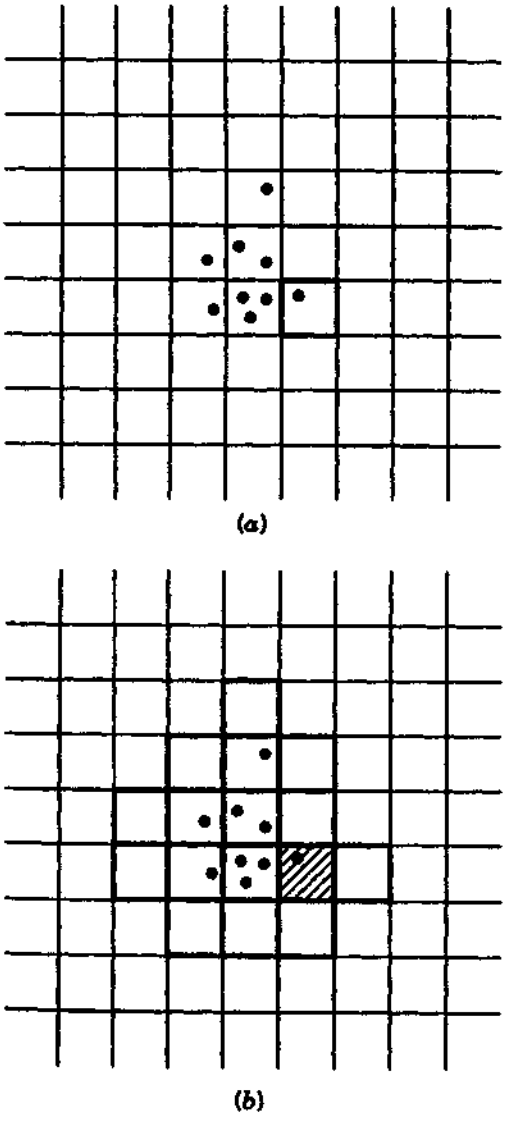

Figure 1: (a) Initial sample plot. (b) Cluster obtained by adding adaptively.

transect), and the area is surveyed on either side out to a given distance. Thompson [24] showed how the above theory can be applied to this sampling situation. He pointed out that a primary unit need not be a contiguous set of secondary units. For example, in some wildlife surveys, the selection of sites chosen for observation is done systematically (with a random starting point), and a single systematic selection then forms the primary unit. We can then select several such primary units without replacement and add adaptively as before. Such a selection of secondary units will tend to give better coverage of the population and then a simple random sample.

Clearly other ways of choosing a primary unit to give better coverage are possible. Munholland and Borkowski [13, 14] suggest using a Latin square + 1 design selected from a square grid of secondary units (plots). The Latin square gives a secondary unit in every row and column of the grid, and the extra (i.e., +1) unit ensures that any pair of units has a positive probability of being included in the initial sample. The latter requirement is needed for unbiased variance estimation.

In some situations, it is hard to know what c should be for the condition $y > c$. If we choose c too low or too high, we end up with a feast or famine of extra plots. Thompson [28] suggested using the data themselves, in fact, the order statistics. For example, c could be the rth largest y-value in the initial sample statistic, so that the neighborhoods are now determined by the y-values. This method would be particularly useful in pollution studies, where the location of "hot spots" is important.

Another problem, regularly encountered with animal population studies, is that not all animals are detected. Thompson and Seber [30] developed tools for handling incomplete detectability for a wide variety of designs, including adaptive designs, thus extending the work of Steinhorst and

Samuel [22].

Often we are in a multivariate situation where one needs to record several characteristics or measurements on each unit, e.g. the numbers of different species. Thompson [27] pointed out that any function of the variables can be used to define the criterion C, and he obtained unbiased estimates of the mean vector and covariance matrix for these variables.

We can use any of the above methods in conjunction with stratification. If we do not allow the clusters to cross stratum boundaries, then individual stratum estimates are independent and can be combined in the usual fashion. Thompson [25] extended this theory to allow for the case where clusters do overlap. Such an approach makes more efficient use of sample information.

Finally, there are two further developments relating to design, namely, selecting networks without replacement and a two-stage sampling procedure [17, 18].

1.4 Unbiased Estimation

Although the cluster is the natural sample group, it is not a convenient entity to use for theoretical developments because of the double role that edge units can play. If an edge unit is selected in the initial sample, then it forms a cluster of size 1. If it is not selected in the initial sample, then it can still be selected by being a member of any cluster for which it is an edge unit. We therefore introduce the idea of the network A_i for unit i, defined to be the cluster generated by unit i but with its edge units removed. In Figure 1b we get the sampled network by omitting the empty units from the sampled cluster. Here the selection of *any* unit in the network leads to the selection of *all* of the network. If unit i is the only unit in a cluster satisfying C, then A_i consists of just unit i and forms a network of size 1. We also define any unit that does not satisfy C to be a network of size 1, as its selection does not lead to the inclusion of any other units. This means that all clusters of size 1 are also networks of size 1. Thus, any cluster consisting of more than one unit can be split into a network and further networks of size 1 (one for each edge unit). In contrast to having clusters that may overlap on their edge units, the distinct networks are *disjoint* and form a *partition* of the N units.

Since the probability of selecting a unit will depend on the size of the network it is in, we are in the situation of unequal-probability sampling and the usual estimates based on equal-probability sampling will be biased. However, we have the well-known Horvitz–Thompson (HT) and Hansen–Hurwitz (HH) estimators (cf. Refs. 8 and 9) for this situation, the latter being used in sampling with replacement. These estimators, however, require knowing the probability of selection of each unit in the final sample. Unfortunately these probabilities are only known for units in networks selected by the initial sample and not for the edge units attached to these networks. Therefore, in what follows, we ignore all edge units that are not in the initial sample and use only network information when it comes to computing the final estimators.

Motivated by the HT estimator for the population mean μ, we consider

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} y_i \frac{I_i}{E[I_i]},$$

where I_i takes the value 1 if the initial sample intersects network A_i, and 0 otherwise; $\hat{\mu}$ is an unbiased estimator for sampling with or without replacement.

Another possible estimator (motivated by the HH estimator) that is also obviously unbiased for sampling with or without re-

placement, is

$$\tilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{E[f_i]},$$

where f_i is the number of times that the ith unit in the final sample appears in the estimator, that is, the number of units in the initial sample that fall in (intersect) A_i determined by unit i; $f_i = 0$ if no units in the initial sample intersect A_i. It can be shown that

$$\tilde{\mu} = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \bar{w}, \quad \text{say},$$

where w_i is the mean of the observations in A_i; i.e., $\bar{w}$ is the mean of the n_1 (not necessarily distinct) network means.

1.5 Adaptive Allocation

There are other ways of adaptively adding to an initial sample. For instance, suppose the population is divided up into strata or primary units each consisting of secondary units. An initial sample of secondary units is taken in each primary unit. If some criterion is satisfied such as $\bar{y} > c$, then a further sample of units is taken from the *same* primary unit. Kremers [12] developed an unbiased estimator for this situation.

If the clumps tend to be big enough so that they are spread over several primary units, we could use what is found in a particular primary unit to determine the level of the sampling in the next. This is the basis for the theory developed by Thompson et al. [29]. Other forms of augmenting the initial sample that give biased estimates are described by Francis [6,7] and Jolly and Hampton [10,11]. This kind of adaptive sampling based on allocating more units rather than adding more neighborhoods is called *adaptive allocation*.

1.6 Rao–Blackwell Modification

An adaptive sample can be defined as one for which the probability of obtaining the sample depends only on the distinct unordered y-observations in the sample, and not on the y-values outside the sample. In this case d, the set of distinct unordered labels in the sample together with their associated y-values, is minimal sufficient for μ. This is proved for "conventional designs" by Cassel et al. [3] and Chaudhuri and Stenger [4], and their proofs readily extend to the case of adaptive designs. (This extension is implicit in Basu [1].) This means that an unbiased estimator that is not a function of d can be "improved" by taking the expectation of the estimator conditional on d to give an estimator with smaller variance. For example, consider three unbiased estimators of μ, namely $\bar{y}_1$ (the mean of the initial sample of n_1 units), $\hat{\mu}$, and $\tilde{\mu}$. Each of these depends on the order of selection, as they depend on which n_1 units are in the initial sample; $\tilde{\mu}$ also depends on repeat selections; and when the initial sample is selected with replacement, all three estimators depend on repeat selections. Since none of the three estimators is a function of the minimal sufficient statistic d, we can apply the Rao–Blackwell theorem. If T is any one of the three estimators, then $E[T|d]$ will give a better unbiased estimate, i.e., one with smaller variance. We find that this estimator now uses all the units including the edge units.

Finally we mention the "model-based" or "superpopulation" approach (cf. Särndal et al. [19], for example). Here the population vector $\mathbf{y}$ of y-values is considered to be a realization of a random vector $\mathbf{Y}$ with some joint distribution F, which may depend on an unknown parameter ϕ. In a Bayesian framework ϕ will have a known prior distribution. For

this model-based approach, Thompson and Seber [31] indicate which of the results for conventional designs carry over to adaptive designs and which do not. They also show in their Chapter 10 that optimal designs tend to be adaptive.

1.7 Relative Efficiency

An important question one might ask about adaptive sampling is "How does it compare with, say, simple random sampling?" This question is discussed by Thompson and Seber [31, Chapter 5], and some guidelines are given. Cost considerations are also important. Simple examples given by them throughout their book suggest that there are large gains in efficiency to be had with clustered populations. Two simulation studies which shed light on this are by Brown [2] and Smith et al. [21].

References

1. Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā A*, **31**, 441–454.
2. Brown, J. A. (1994). The application of adaptive cluster sampling to ecological studies. In *Statistics in Ecology and Environmental Monitoring*, D. J., Fletcher and B. F. J., Manly. eds., University of Otago Press, Dunedin, New Zealand, pp. 86–97.
3. Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley, New York.
4. Chaudhuri, A. and Stenger, H. (1992). *Survey Sampling: Theory and Methods*. Marcel Dekker, New York.
5. Danaher, P. J. and King, M. (1994). Estimating rare household characteristics using adaptive sampling. *New Zealand Statist.*, **29**, 14–23.
6. Francis, R. I. C. C. (1984). An adaptive strategy for stratified random trawl surveys. *New Zealand J. Marine and Freshwater Res.*, **18**, 59–71.
7. Francis, R. I. C. C. (1991). Statistical properties of two-phase surveys: Comment. *Can. J. Fish. Aquat. Sci.*, **48**, 1128.
8. Hansen, M. M. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.*, **14**, 333–362.
9. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663–685.
10. Jolly, G. M. and Hampton, I. (1990). A stratified random transect design for acoustic surveys of fish stocks. *Can. J. Fish. Aquat. Sci.*, **47**, 1282–1291.
11. Jolly, G. M. and Hampton, I. (1991). Reply of comment by R. I. C. C. Francis. *Can. J. Fish. Aquat. Sci.*, **48**, 1128–1129.
12. Kremers, W. K. (1987). Adaptive Sampling to Account for Unknown Variability Among Strata. *Preprint No. 128*, Institut für Mathematik, Universität Augsburg, Germany.
13. Munholland, P. L. and Borkowski, J. J. (1993). Adaptive Latin Square Sampling +1 Designs. *Technical Report No. 3-23-93*, Department of Mathematical Sciences, Montana State University, Bozeman.
14. Munholland, P. L. and Borkowski, J. J. (1996). Latin square sampling +1 designs. *Biometrics*, **52**, 125–132.
15. Ramsey, F. L. and Sjamsoe'oed, R. (1994). Habitat association studies in conjunction with adaptive cluster samples. *J. Environmental Ecol. Statist.*, **1**, 121–132.
16. Roesch, F. A., Jr. (1993). Adaptive cluster sampling for forest inventories. *Forest Sci.*, **39**, 655–669.
17. Salehi, M. M. and Seber, G. A. F. (1997). Adaptive cluster sampling with networks selected without replacement. *Biometrika*, **84**, 209–219.

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

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

20. Seber, G. A. F. and Thompson, S. K. (1994). Environmental adaptive sampling. In *Handbook of Statistics, Vol. 12 (Environmental Sampling)*, G. P. Patil and C. R. Rao, eds. North Holland/Elsevier Science, New York, pp. 201–220.

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

22. Steinhorst, R. K. and Samuel, M. D. (1989). Sightability adjustment methods for aerial surveys of wildlife populations. *Biometrics*, **45**, 415–425.

23. Thompson, S. K. (1990). Adaptive cluster sampling. *J. Amer. Statist. Ass.*, **85**, 1050–1059.

24. Thompson, S. K. (1991). Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, **47**, 1103–1115.

25. Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika*, **78**, 389–397.

26. Thompson, S. K. (1992). *Sampling*. Wiley, New York.

27. Thompson, S. K. (1993). Multivariate aspects of adaptive cluster sampling. In *Multivariate Environmental Statistics*, G. P. Patil and C. R. Rao, eds. North Holland/Elsevier Science, New York, pp. 561–572.

28. Thompson, S. K. (1995). Adaptive cluster sampling based on order statistics. *Environmetrics*, **7**, 123–133.

29. Thompson, S. K., Ramsey, F. L., and Seber, G. A. F. (1992). An adaptive procedure for sampling animal populations. *Biometrics*, **48**, 1195–1199.

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

31. Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*. Wiley, New York.