FRAME FREE ADAPTIVE DESIGNS

Paul L. Mosquin, Steven K. Thompson, The Pennsylvania State University
Paul L. Mosquin, Department of Statistics,
326 Thomas Building ,The Pennsylvania State University,
University Park, PA, 16802 (mosquin@stat.psu.edu)

Key Words: Adaptive sampling, Quadrat sampling, Point process, Point pattern, Rao-Blackwell.

Abstract:

Sampling, both adaptive and non-adaptive, is usually conducted after dividing the study region into sample units by placement of a sampling frame. For cases where the population can be represented on a finer scale, it may be desirable to collect the sample without the use of a frame. In this paper, we study populations that can be considered to be realizations of point processes. For these populations, an initial sample can be collected by uniformly choosing sample point locations from within the study region, and then sampling point objects close to these point locations. Additional point objects can be added to the sample during an adaptive stage, ultimately allowing for the calculation of various unbiased adaptive estimates of the population mean or total. This paper will present and discuss estimators for these "frame free" designs. Like the frame based designs, these designs allow Hansen-Hurvitz and Horwitz-Thompson type estimation, as well as appropriate Rao-Blackwellization.

1. Introduction

The last decade has seen the introduction of the adaptive sampling designs, which allow for efficient sampling from rare and clustered populations [2]. These designs are known as adaptive since they proceed in two stages: an initial random sampling stage, and a secondary adaptive stage in which additional sampling effort is applied near "informative" units revealed in the first stage. These methods were originally developed for adaptive cluster sampling after a first stage simple random sample, and have since been extended to other classical designs such as systematic [3], and stratified sampling [4]. This work, as well as other more recent work has been summarized in a book by Thompson and Seber [5].

Thanks to George Chao and Arthur Dryver for comments on an earlier version.

For the sampling of spatial observations, most research on adaptive designs has assumed that samples are collected from a sampling frame, which is here considered to be a square or rectangular grid that contains the units of the population. In the case of spatial sampling of a marked point process $\{\boldsymbol{y}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{D}\}$ where countable index set $\mathcal{D} \subset \Re^d$ (we will often assume $d = 2$), the sampling frame would be superimposed on a realization of the process, and for the N units of the frame, unit i would provide an associated response y_i. Typically this response would simply equal a count of the number of point objects in that frame unit.

Alternatively, a sample from the realized point process can be collected without the placement of a sampling frame. By this approach, the point objects (and not frame units) are labeled from $1, ..., N$, where point object i located at $\boldsymbol{x}_i$ has an associated response of interest $y_i = f(\boldsymbol{y}(\boldsymbol{x}_i))$ where $\boldsymbol{x}_i$ is the location of point object i, and $\boldsymbol{y}(\boldsymbol{x}_i)$ is a vector of attributes of point object i. A sample of n point locations $z_1, ..., z_n$ are uniformly chosen from within the study region (a volume $\mathcal{A}$ of $\Re^d$ where $\mathcal{D} \subset \mathcal{A}$) and point objects are then included in the sample depending on whether they are close in some sense to these point locations.

Estimators for designs of this type were first provided by Roesch [1] who considered designs with unequal probability sampling in the first stage. Additional discussion can be found in Thompson and Seber [5]. In this paper, the earlier work will be extended to provide estimators for a variety of other frame free designs.

2. The frame based adaptive sample

A standard frame based adaptive cluster design has two sampling stages. In the first stage, a simple random sample of n units is collected from the population. Units from the first stage that satisfy some condition then have neighboring regions sampled in a second, adaptive stage. This adaptive stage may reveal other regions which contain units that satisfy the condition, and these units in turn have their neighboring regions sampled. Sampling continues until all neighboring regions of units that satisfy the condition have been sampled. The algorithmic nature of an adaptive sampling design allows the construction of unbiased estimates of parameters of interest since once a group of units satisfying the condition has been discovered, it is then known *a posteriori* what the probability of selecting that group was given the initial simple random sample.

3. The no frame initial sample

When a sampling frame is not present, the initial sample will be obtained by uniformly choosing n point coordinates $z_1, ..., z_n$, and associating with these points the label sets $s_1, ..., s_n$ where point object label $j \in s_i$ if $I(x_i \in A_j) = 1$. The sets $A_1, ..., A_N$ give inclusion regions of $\mathcal{A}$ in which placement of a sample point location leads to the sampling of a given point object. Often the sets can be specified as the region in which $f(z, x_j, y(x_j)) < r_1$, where function $f(z, x_j, y(x_j))$ will be refered to as an inclusion function. Some examples of the sets are:

1. If $f(z, x_j, y(x_j)) = ||z - x_j||^{\frac{1}{2}}$ then A_j is a circular/spherical region of radius r_1 centered on point object j.
2. If $f(z, x_j, y(x_j)) = min(|z_k - x_k|)$ where $x_j = (x_1, x_2, ..., x_d)$ then A_j is a square/cubic region of width $2r_1$ centered on point object j.
3. If $f(z, x_j, y(x_j)) = d_j||z - x_j||^{\frac{1}{2}}$ then there is probability proportional to size sampling. In Roesch [1], point objects (trees) were included depending on their diameters, and so in this case d_j is a scaling factor for the diameter of tree j.
4. If point locations are allocated to disjoint sets that partition $\mathcal{A}$, and A_j equals the disjoint set to which point object j belongs, then the disjoint sets will describe "units" for a sampling frame based design.

An example of an initial sample for a square inclusion function is given in Figure 1. In the figure, a sample of three point locations has been taken, and around each the "sample unit" region outlined. This is the region that would be physically sampled in search of point objects. Of the three sample point locations, the first lies within the sampling regions of two point objects. The third sample point was closer than r_1 to the study region boundary, and if desired, point objects could be sampled from the other side of the study region if the inclusion function was allowed to 'wrap' around to the other side (dotted lines).

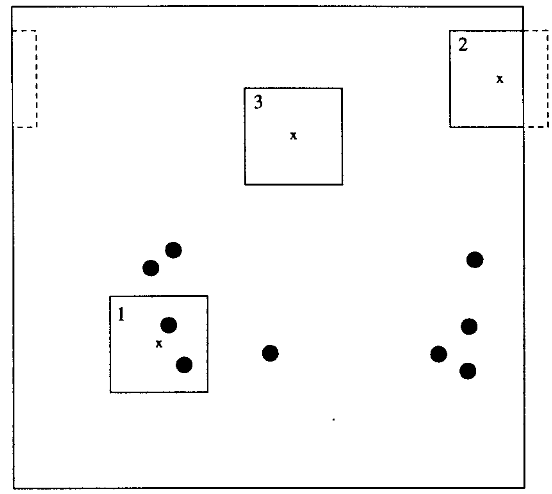

Figure 1: Initial no frame sample

4. The no frame adaptive sample

There are two methods that can be used to decide if point satisfy the condition: a single point object condition and a multiple point object condition. Note that without a sampling frame, the adaptive condition is satisfied by point objects, and not units. Once adaptive sampling is finished, the final sample data D will consist of the locations and responses of all point objects discovered, as well as locations sampled in search of point objects.

4.1 The single point object condition

By this method, sample point objects satisfy the condition if their response exceeds a constant value c, that is, point object j satisfies the condition if $y_j > c$. Adaptive sampling is initiated by allocating each sample point object to its own network set, creating network sets $n_1, ..., n_K$ where there were K point objects in the initial sample. If any network sets are equal, then they are collapsed into a single network set, and the remaining sets are relabelled. The adaptive sampling then iterates over all point objects that satisfy the condition, applying adaptive inclusion function $g(x_i, x_j, y(x_j))$ from point object $i \in n_k$ to $j \notin n_k$ where point object i satisfies the condition. Include point object j in n_k if $g(x_i, x_j, y(x_j)) < r_2$ and $g(x_j, x_i, y(x_i)) < r_2$ and j satisfies condition. Whenever $n_k = n_{k'}$ $k \neq k'$ remove $n_{k'}$ and relabel. The adaptive inclusion function is not necessarily the same as the initial sample inclusion function, and is applied in a way that ensures that point objects i and j will include each other regardless of which one was sampled first.

4.2 The multiple point object condition

In the multiple point object method, sample point objects satisfy the condition if the sum of the responses within a given label set exceeds the specified value, in this case point object j satisfies the condition if $\sum_{i \in s_k} y_i > c$ where $j \in s_k$. This method will require that the same inclusion function be used for the initial and adaptive stages, where the inclusion function does not depend on point object attributes. Adaptive sampling begins by setting network sets $n_k = s_k$ $k = 1, ..., n$ and then iterates over point objects that satisfy the condition, to determine if any point objects not in a given network set should be added to that set. A point object is added to a network set if it is sufficiently close, and if an initial sample point could have been placed such that the point object would satisfy the condition. That is, if point object j is not in network set n_k then include point object j in n_k if there is a location z such that $I(z \in A_j) = 1$ and $I(z \in A_i) = 1$ for some $i \in n_k$ and $\sum_{i=1}^{N} y_i I(z \in A_i) > c$.

By either the single point object or multiple point object approach, the population of point objects will be uniquely partitioned into networks. As well, during the adaptive stage, for some networks there will be point objects which are close in the sense that they are within r_1 (multiple point object case) or r_2 (single point object case) of the network and yet do not satisfy the condition. These point objects behave as "edge units" do for a frame based sample by not contributing to later estimators, unless they are in the initial sample or there is Rao-Blackwellization.

5. Initial sample unbiased estimators

It is useful to first consider the estimators that would be used if there was no adaptive stage in the sample design. Two well known unequal probability sampling unbiased estimators that can be used are the Horvitz-Thompson estimator and the Hansen-Hurwitz estimator.

5.1 Horvitz-Thompson estimator

The Horvitz-Thompson estimator of the population total $\hat{\tau} = \sum_{i=1}^{N} y_i$ is

$$\hat{\tau} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i$$

where I_i indicates whether point object i was included in the sample, and π_i is its inclusion probability. This probability can be calculated as

$$\pi_i = 1 - \left(1 - \frac{a_i}{A}\right)^n$$

which is one minus the probability that the point object is not present in the sample since A is the area of the study region and a_i is the area in which placement of a sample point location leads to inclusion of point object i in the sample. The area a_i can be found according to $a_i = \int I(z \in A_i) dz$. These a_i can be considered in some cases to provide the area of the sampling "units".

The variance of the Horvitz-Thompson estimator is

$$Var[\hat{\tau}] = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j$$

with the π_{ij} being joint inclusion probabilities for point objects i and j in the sample. They are given by

$$\pi_{ij} = 1 - \left[\left(1 - \frac{a_i}{A}\right)^n + \left(1 - \frac{a_j}{A}\right)^n - \left(1 - \frac{a_i + a_j - a_{ij}}{A}\right)^n\right]$$

An unbiased estimate of the variance is then

$$\widehat{Var}[\hat{\tau}] = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j I_i I_j$$

where I_i and I_j indicate inclusion of point objects i and j in the sample, respectively. Note that $\pi_{ij} = \pi_i$ if $i = j$.

5.2 Hansen-Hurwitz estimator

A second unbiased estimator which can be used on the initial sample data is the Hansen-Hurwitz estimator. For an initial random sample without an adaptive stage the Hansen-Hurwitz estimator of the population total τ is

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

where f_i is the count or frequency of point object i in the sample and $E[f_i]$ is its expectation. This expectation is equal to $E[f_i] = np_i$ where $p_i = a_i / A$.

The variance of $\tilde{\tau}$ is given as

$$Var[\tilde{\tau}] = \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \left(\frac{A a_{ij}}{a_i a_j} - 1 \right)$$

where a_{ij} is the area in which placement of a point location leads to both point objects i and j being included in the sample. This variance is similar to that of the classic Hansen-Hurwitz estimator, with a modification made for the possiblility of including more than one point object from a given sample location.

Let $w_i = \sum_{j=1}^{N} y_j I_j / p_j$ be the value of the Hansen-Hurwitz estimator for the i^{th} sample draw, and $\bar{w}$ be the mean of the w_i . An unbiased estimate of the variance is then

$$\widehat{Var}[\tilde{\tau}] = \frac{1}{n(n-1)} \sum_{i=1}^{n} (w_i - \bar{w})^2 \qquad (1)$$

6. Adaptive estimators for single point object condition

For the case where the condition is satisfied by properties of single point objects, the non-Rao-Blackwell estimators were presented in Roesch [1]. In his paper, the point objects were trees, and there was probability proportonal to size sampling in the initial stage according to tree trunk diameter.

Two of the unbiased estimators proposed in his paper were a Hansen-Hurwitz estimator, and a Horvitz-Thompson estimator. These estimators are the same as the non-adaptive estimators given earlier except that inclusion areas associated with point objects are now determined by the networks to which they belong.

6.1 Rao-Blackwell estimators

A single point object condition estimator not considered by Roesch is the Rao-Blackwellization of either the Hansen-Hurvitz or the Horwitz-Thompson non-adaptive estimator. Suppose that for the n coordinates placed in the initial sample, there are a total of G different compatible collections of the label sets $s_1, ..., s_n$ such that the same final adaptive sample D would have been obtained. For each of these compatable collections there will be a corresponding no adaptive stage sample estimate $\hat{\tau}_g$ of the either the Hansen-Hurwitz or Horwitz-Thompson type. The Rao-Blackwell estimator $\hat{\tau}^{RB}$ of τ is then

$$\begin{aligned} \hat{\tau}^{RB} &= E[\hat{\tau}|D] \\ &= \sum_{g=1}^{G} \hat{\tau}_g f_{\hat{\tau}|D}(\hat{\tau}_g) \end{aligned}$$

where $f_{\hat{\tau}|D}$ is the density of a non-adaptive frame free estimator $\hat{\tau}$ conditional on obtaining sample D. By the Rao-Blackwell Theorem $\hat{\tau}^{RB}$ will have variance less than or equal to that of the given initial sample estimate.

7. Adaptive estimators for multiple point object condition

With a multiple point object conditon, the difficulty in constructing estimators is that it is not possible to know the true inclusion probabilities for point objects that do not satisfy the condition in the initial sample. This contrasts with the single point object condition approach where point objects satisfy the condition depending only on their associated responses, in which case point object inclusion probabilities are more readily found. It is also contrasts with the frame based designs where all initial sample units have a known initial sample inclusion probability.

However, it is still possible to construct unbiased estimators once it is realized that point objects that can be either adaptively or non-adaptively sampled will have known initial inclusion probabilities *if they are adaptively sampled.* As a consequence, when such point objects are adaptively included in a sample these known inclusion probabilities can be used to weight their contribution to an adaptive estimator so that unbiasedness of the estimator is maintained. The resulting estimators can be viewed as generalizations of the Horvitz-Thompson and Hansen-Hurwitz estimators.

7.1 Horvitz-Thompson type estimator

The Horvitz-Thompson type unbiased estimator of the population total is

$$\hat{\tau} = \sum_{i=1}^{N} \left(\frac{1}{\pi_i} I_i + \frac{c_i}{\pi_i^*} J_i \right) y_i$$

which is a sum over all point objects, of the point object responses times the sum of two functions of indicator variables.

The first indicator variable I_i indicates if point object i is in the initial sample but does not satisfy the condition at the end of this stage. This indicator is multiplied by the factor $1/\pi_i$ where π_i is the inclusion probability of point object i *if there were no adaptive stage.* This inclusion probability is equal to

$$\pi_i = 1 - \left(1 - \frac{a_i}{A}\right)^n$$

where a_i is the area of the region in which placement of an initial point location leads to inclusion of point object i in the sample.

The second indicator variable J_i indicates whether point object i satisfies the condition in the final sample. It is multiplied by c_i/π_i^* where π_i^* is the inclusion probability for point object i given that it satisfies the condition. This inclusion probability is equal to

$$\pi_i^* = 1 - \left(1 - \frac{a_i^*}{A}\right)^n$$

where a_i^* is the area in which placement of an initial sample point leads to point object i satisfying the condition in the final sample.

The constant c_i is equal to

$$c_i = 1 - \frac{\pi_i^{**}}{\pi_i}$$

where the probability π_i^{**} gives the probability that point object i is in the initial sample and did not satisfy the condition while in the inital sample. This probability is equal to

$$\pi_i^{**} = 1 - \left(1 - \frac{a_i^{**}}{A}\right)^n$$

where a_i^{**} is the area in which placement of an initial sample point includes point object i in the initial sample without satisfying the condition. The c_i can be thought of as compensating within the estimator for those point objects that can either satisfy or not satisfy the condition in the initial sample, depending on location of the initial sample point.

Note that the estimator does not contain an indicator variable for point objects that are found in the adaptive sample but do not satisfy the condition. These point objects will not contribute to the estimator and so behave as edge units do in a frame based design.

If there is no adaptive stage, the J_i will all equal zero, and so the estimator reduces to the no adaptive stage Horvitz-Thompson estimator. If there are no point objects that can be both adaptively or non-adaptively sampled then the c_i will equal zero when the J_i equal one, and the estimator reduces to the Horvitz-Thompson estimator for single point object conditions.

The variance of this estimator is a rather involved expression which for brevities sake will not be included here. Unfortunately, an estimate of this variance cannot be obtained by the usual approach because the sample information does not provide all of the necessary inclusion probabilities.

7.2 Hansen-Hurwitz type estimator

The Hansen-Hurwitz type estimator can be given as

$$\tilde{\tau} = \sum_{i=1}^{N} \left(\frac{1}{E[f_i^{na}]} f_i + \frac{c_i}{E[f_i^*]} f_i^* \right) y_i$$

where f_i is the frequency of point object i where it is in the initial sample but does not satisfy the condition at the end of that stage. The expectation $E[f_i^{na}]$ is the expectation of this frequency if there were no adaptive stage. This expectation is equal to np_i, where $p_i = a_i/A$ and the a_i are defined as above.

Similarly, f_i^* is the frequency that point object i satisfies the condition in the final sample, and $E[f_i^*]$ its expectation. This expectation is equal to np_i^* where $p_i^* = a_i^*/A$ and the a_i^* are as previously defined.

The constant c_i is equal to

$$c_i = 1 - \frac{a_i^{**}}{a_i}$$

where a_i^{**} is as given above.

This estimator reduces to the non-adaptive Hansen-Hurwitz or the single point object Hansen-Hurwitz estimator in the same manner as the Horvitz-Thompson type estimator does. It is easily seen to be unbiased once it is realized that for the sample of size one it equals the (unbiased) Horvitz-Thompson type estimator. Since the Hansen-Hurwitz is the average of n such estimators, unbiasedness follows as a property of the sample mean.

As with the Horvitz-Thompson type estimator, the variance of the Hansen-Hurwitz type estimator is a lengthy expression, and so not included here. It does yield an unbiased estimator, however, and this estimator is the same as (1), the variance estimator for the non-adaptive Hansen-Hurwitz estimator where the w_i are now

$$w_i = \sum_{j=1}^{N} \left(\frac{y_j}{p_j} I_j + \frac{c_j}{p_j^*} J_j \right)$$

7.3 Rao-Blackwell Estimator

As in the single point object case, the Rao-Blackwellized estimator averages one of the no adaptive stage estimators over all initial samples compatible with the final adaptive sample. For a sample of size one, the Horvitz-Thompson, Hansen-Hurwitz, and Rao-Blackwell estimators are identical. Another view of the Hansen-Hurwitz estimator then is that it is a draw by draw average of Rao-Blackwell estimators.

8. Comparison of frame and frame free approaches

Besides issues of practicality for a specific study, the decision of whether or not to use a sampling frame can be considered in terms of amount of area sampled during the initial and adaptive stages. For the frame free designs, this area is found by integrating over all point locations within r_1 and r_2 (if single point object method) of the inclusion functions applied to point locations from the initial sample and point objects that satisfied the condition in the final sample.

During the initial stage, a frame based design can have sampling either with or without replacement of study region. An advantage of without replacement sampling is that there is no repetition of effort by the resampling of formerly sampled regions. With a frame free design, however, sampling would generally be with replacement since the inclusion probabilities are then much easier to compute. A consequence of with replacement sampling, though, is that regions can be sampled more than once during the initial stage. This duplication of effort suggests that the without replacement frame based designs are desirable during the initial stage (although adaptive designs are often used in situations where the chance of sampling the same initial region is small).

During the adaptive stage, though, it would be expected that the frame free designs sample less region since the "edge" of a sampled region will more closely follow the shape of a network of point objects. Using square units for ease of comparison, Figure 2 presents an example of possible gains. This figure shows the region sampled around a point object that satisfies the condition at the end of the adaptive stage for both frame and frame free designs. Some study reveals that the frame based aproach sampled one additional unit of area than the frame free approach.

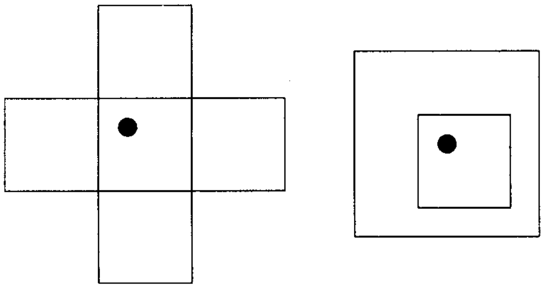

Figure 2: Comparison of sample region.

Finally, an additional efficiency advantage to the adaptive frame free approach is that by more closely following the shape of the network, point objects can be part of a network which would otherwise have been in a frame edge unit. That is, point objects that are part of units that satisfy the condition in a frame based design will satisfy the condition in a frame free design, however the converse is not generally true.

9. Combined approaches

Since a frame based approach may be prefered for the initial stage, while a frame free approach prefered during the adaptive stage, it seems reasonable to ask if a combined approach can be used. In fact, it can, although the initial inclusion probabilities change in the Horvitz-Thompson type estimator to accomodate sampling without replacement. To distinguish the number of frame units from the number of point objects, let N' be the number of units in the sampling frame. For a sampling frame of regularly sized units, the adaptive inclusion probability for point object i becomes

$$\pi_i^* = 1 - \binom{N' - m_i}{n} / \binom{N'}{n}$$

where m_i is the number of frame units that satisfy the condition in the network to which point object i belongs. Note that for the multiple point object condition the units of the frame must correspond to the units determined by the inclusion function.

Interestingly, with this combined approach point objects that are in network edge units according to the frame still do not contribute to the final estimate since for these point objects $a_i = a_i^{**}$ and so $c_i = 0$.

References

[1] Roesch F.A. Jr. Adaptive cluster sampling for forest inventories. *Forest Science*, 39:655–669, 1993.

[2] Thompson S.K. Adaptive cluster sampling. *Journal of the American Statistical Association*, 85:1050–1059, 1990.

[3] Thompson S.K. Adaptive cluster sampling : Designs with primary and secondary units. *Biometrics*, 47:1103–1115, 1991.

[4] Thompson S.K. Stratified adaptive cluster sampling. *Biometrika*, 78:389–397, 1991.

[5] Thompson S.K and Seber G.A.F. *Adaptive Sampling.* John Wiley and Sons, 1996.