

# Application of Adaptive Sampling in Fishery Part 1: Adaptive Cluster Sampling and Its Strip Designs

**M. Salehi M.**

School of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran

Email: salehi\_m@cc.iut.ac.ir

**Abstract:** The precision of conventional sampling designs is not usually satisfactory for estimating parameters of clump and rare populations. Many of fish species live in school and disperse all over a vast area like a sea so that they are rare compare to their habitats. Theory of a class of sampling designs called adaptive sampling designs has rapidly grown during last decade which solved the mentioned problem. These sampling designs are successfully used to assess fish life stocks in number of cases. The first aim of this article is to develop some versions of adaptive sampling which are more useful for fishery study. The author will introduce Adaptive Cluster Sampling (ACS) and its strip version which is more useful for fishery and marine studies. Since the parameter of interest is usually the density in fishery study, estimators of the density are developed in this article. In many biological studies, it has been shown that non-overlap two stage adaptive sampling and strip adaptive sampling are very efficient sampling designs so that we also develop non-overlap two stage adaptive cluster sampling of strips. The second aim of this article is to introduce adaptive cluster sampling to our fishery researchers' community and demonstrate the idea and computation to those who know statistics in some level.

**Key Words:** Sampling design, Clump fish population, Abundance estimation

## Introduction

In situations we are dealing with, sampling from an area (e.g. sea or ocean), partitioned into units, each unit contains one or more fish(es), and with each fish we usually are concerned about measurements of one or more parameters called the variables of study (e.g. their weight or length). Our goal is to obtain information about unknown population characteristics or parameters of interest (e.g. the total number of fishes, total of fishes' weight or length). Parameters are functions of some of the study variable values, which are, called the variables of interest. From such population, a sample is selected. The procedure for selecting sampling units is called "the sampling design". The sampling units are observed, i.e. the variables of study are measured and recorded. The recorded variable values are used to calculate estimates of the finite population parameters of interest (totals, means, ratio, etc.). Thompson and Seber (1996) divided the set of all possible probability sampling designs into two types: conventional and adaptive.

In conventional design, the procedure for selecting sample does not depend on any observations of the variables of study, so that the entire sample may be selected prior to the survey. This can help us assign the facilities in advance to get some idea about the precision of the estimators using a pilot survey. However we do not have any control over the procedure for selecting the sampling units during the survey. All the familiar probability designs such as simple random sampling, stratified random sampling, sampling with probability proportional to size, systematic sampling, cluster sampling, multistage sampling etc. are conventional designs. Conventional sampling methods are limited to sampling design in which the selection of the samples (units) can be done before the survey, so that none of the decisions about sampling depend on what is observed as one gathering the data.

In adaptive sampling, the procedure for selecting the sample may depend on the values of the variable of study observed in the sample. Examples include simple random sampling with a stopping rule based on observed values, "inverse" sampling (keep sampling until you have ten units containing at least one fish), fixed cost sequential sampling (keep sampling until you run out of time, money or resources), adaptive cluster sampling (if fishing is good, keep fishing in that area), restricted adaptive cluster sampling (if fishing are good in some areas keep fishing in those areas until you run out of the resources) and adaptive allocation of sampling effort in stratified sampling (Francis, 1984) ; Jolly & Hampton (1990),

allocate more samples to those strata, which are cheaper to observe and have more fish, which become known for us while we are gathering the data). Simply, adaptive sampling makes better use of the data gathered.

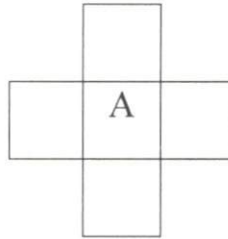
Suppose that we are able to perfectly detect and measure the variable of study in each selected unit. Otherwise we can use distance sampling technique (Pollard & Buckland, 1997). Thompson (1990) introduced the adaptive cluster sampling design. Adaptive cluster sampling was motivated by the problem of sampling rare, clustered populations. This sampling design indicates that wherever you find, something that you are looking for, keep sampling its vicinity to find more. During last decade, this sampling procedure draw researcher's attention and has considerably developed in different ways. Thompson (1991a,b) developed stratified, and design with primary and secondary adaptive cluster sampling, Munholland & Borkowski (1993) developed Adaptive Latin square sampling +1, Roesch (1993) developed adaptive cluster sampling for forest inventories, Thompson & Seber (1994) discussed delectability in adaptive sampling, Brown (1994) developed restricted adaptive cluster sampling, Smith *et al.* (1995) used it for estimating density of wintering waterfowl, Thompson (1996) developed adaptive cluster sampling based on ordered statistics, Salehi & Seber (1997b) developed two-stage adaptive cluster sampling, Salehi & Seber (1997a) developed adaptive cluster sampling with networks selected without replacement, Lo *et al.* (1997) used restricted adaptive sampling to estimate Pacific Hake larval abundance, Pollard and Buckland (1997) introduced an adaptive sampling for line transect survey, Brown and Manly (1998) discussed problem regarding restricted adaptive cluster sampling, Salehi (1999) solved the problem of selected edge units for adaptive cluster sampling, Clausen *et al.* (1999) used adaptive cluster sampling for estimating rockfish abundance in two study areas of Kodiak Island, Alaska. There are many other papers on adaptive sampling published in recent years. The interested readers can find them in the Internet. All the above mentioned papers used this principle that "If fishing good keep fishing in that area" and related to the pioneer work adaptive cluster sampling introduced by Thompson (1990). Therefore, we first introduce adaptive cluster sampling.

The presented materials are mainly revisions of the materials prepared by the author for Bengulia Environment Fisheries Interaction and Training Program

Survey design Workshop hold in Swakopmund, Namibia, 7-10 August 2000. A recent review of adaptive sampling is given by Seber & Salehi (2002).

### Adaptive Cluster Sampling

In adaptive cluster sampling the population is partitioned into units (quadrate Fig.1).



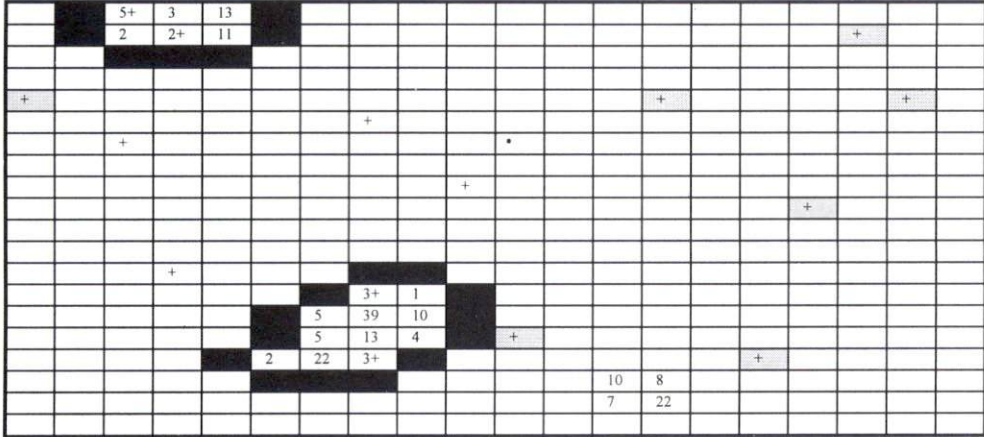
**Figure 1:** The type of neighborhood for unit A used in the example

A set of neighboring units is defined for each unit. The neighborhood relationship is assumed to be symmetric in the sense that if unit  $j$  is in the neighborhood of unit  $i$  then  $i$  is also in the neighborhood of  $j$ . The neighborhood of a unit consists of the unit itself and the four adjacent units sharing a common boundary (Fig 2). An initial simple random sample (SRS) of units is selected and, whenever the value of the variable of interest in a selected unit satisfies a specified condition, neighboring units are added to the sample. This process continues until a cluster of units is formed with a boundary of units, called edge units, which do not satisfy  $C$ . A cluster without its edge units forms a network and a unit not satisfying the condition  $C$  forms a network of size one. The networks are disjoint and form a partition of population units. The condition for extra sampling might be the presence of a fish or plant species, high abundance of a spatially clustered species.

### Notations and estimators

The site of study is partitioned into  $N$  units  $(u_1, u_2, \dots, u_N)$  labeled by  $(1, 2, \dots, N)$ . With  $u_i$  is associated a variable of interest  $y_i$  (can be the number of

fish in quadrat  $i$ ), for  $i = 1, 2, \dots, N$ . A simple random sample of size  $n_1$  is taken without replacement. Further units are then added adaptively using Condition C.



**Figure 2:** The number in each unit denotes the number of fishes and those without number denote the unit without fish. An initial sample of 15 units is selected, unit with a +. The condition for extra sample is "observing at least one fish". Adaptively adding the extra units, 13 networks are selected (shaded region). The black region are selected edge units.

Let  $A_i$  denote the network containing unit  $i$  and with  $m_i$  units in it ( $m_i$  is the size of  $A_i$ ). The probability that the initial sample intersect  $A_i$  is an unbiased estimator of the mean based on Horwitz-Thompson estimator is given by

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

Where  $y_k^*$  is the sum of the  $y$ -values for the  $k$ th network, and  $\kappa$  is the number of distinct networks intersected by the initial sample. We have  $\alpha_\kappa = \alpha_i$  for every unit  $i$  in network  $k$ .

An unbiased estimator of the variance is

$$\widehat{\text{Var}}[\hat{\mu}_{HT}] = \frac{1}{N^2} \sum_{k=1}^K \sum_{k'=1}^K \frac{y_k^* y_{k'}^*}{\alpha_{kk'}} \left( \frac{\alpha_{kk'}}{\alpha_k \alpha_{k'}} - 1 \right), \quad (2)$$

Where;

$$\alpha_k = 1 - \frac{\binom{N - m_k}{n_1}}{\binom{N}{n_1}}, \quad (3)$$

and  $\alpha_{kk} = \alpha_k$

$$\alpha_{kk'} = 1 - \frac{\left[ \binom{N - m_k}{n_1} + \binom{N - m_{k'}}{n_1} - \binom{N - m_k - m_{k'} - m_{kk'}}{n_1} \right]}{\binom{N}{n_1}} \quad (4)$$

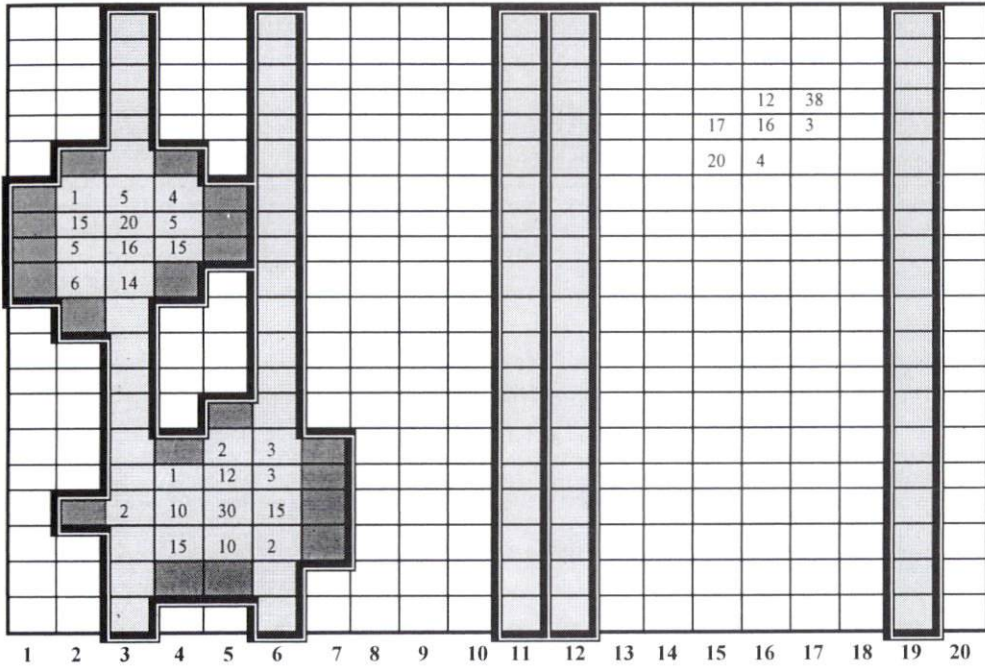
Another unbiased estimator based on Hansen-Horvitz estimator is

$$\hat{\mu}_{HH} = \frac{1}{n_1} \sum_{i=1}^{n_1} w_i \quad (5)$$

where  $w_i = y_i^* / m_i$  is the mean of the  $m_i$  observations in  $A_i$ . Here  $\hat{\mu}_{HH}$  can be recognized as the sample mean obtained by taking a simple random sample of size  $n_1$  from a population of  $w_i$  values (Thompson & Seber 1996). with unbiased estimate

$$\widehat{\text{var}}[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1 (n_1 - 1)} \sum_{i=1}^{n_1} (w_i - \hat{\mu}_{HH})^2 \quad (6)$$

**Example 1:** (*Computation of estimates.*) Consider a region of a sea which is partitioned into 400 quadrates (Fig. 3). This example is introduced to demonstrate the computations for  $\hat{\mu}_{HT}$  and  $\tilde{\mu}_{HT}$ . The initial sample consists of  $n_1 = 15$  units selected by simple random sampling from  $N = 400$  units. The neighborhood of a unit consists of the unit itself and the four adjacent units sharing a common boundary line (i.e. forming a cross).



**Figure 3:** A population of 400 units is partitioned into 20 strip of size 20. An initial sample of size 5 from strips is selected. The gray units the selected units which uses in the estimating procedure. The dark units are edge units.

Adding adaptively to those initial units which contain at least one individual leads to the shaded (networks) regions plus the black regions (edge units). Two of these units, near the top of the study region, intersected a network of  $m_1 = 6$  units containing a total of  $y_1^* = 36$  fishes. Another two units intersected a network of  $m_2 = 11$  units containing  $y_2^* = 107$  fishes. For the other eleven units of the initial sample  $y_i = 0$  and  $m_i = 1$ . There were also 20 edge units which are not used in the calculation of the estimates. The intersection probability for the first network is

$$\alpha_1 = 1 - \frac{\binom{400-6}{15}}{\binom{400}{15}} = 1 - 0.793896551 = 0.206103449$$

For the other large network intersected by the initial sample,

$$\alpha_1 = 1 - \frac{\binom{400-11}{15}}{\binom{400}{15}} = 1 - 0.6531889484 = 0.346810515$$

For the networks of size one the probability  $s\alpha_k = 15/400 = 0.0375$ . Hence the estimate using intersection probabilities is, from (5),

$$\hat{\mu}_{HT} = \frac{1}{400} \left( \frac{36}{0.206103449} + \frac{107}{0.346810515} + \frac{0}{0.0375} + \dots + \frac{0}{0.0375} \right) = 1.207988487$$

fishes per unit or  $400(1.207988487) = 483.195$  total fishes in the population.

To compute  $\text{var}[\hat{\mu}_{HT}]$  we need (4).

$$\begin{aligned} \alpha_{12} &= \alpha_1 + \alpha_2 - \left( 1 - \frac{\binom{\binom{40-17}{15}}{\binom{400}{15}}}{\binom{400}{15}} \right) \\ &= 0.206103449 + 0.346810515 - 1 + \frac{369 \cdot 370 \dots 383}{386 \cdot 387 \dots 400} \\ &= 0.20610344 + 346810515 - 1 + 0.515101729 = 0.0680 \end{aligned}$$



Then,

$$\begin{aligned} \hat{\text{var}}[\hat{\mu}_{HT}] &= \frac{1}{N^2} \left[ \frac{y_1^{*2}}{\alpha_1} \left( \frac{1}{\alpha_1} - 1 \right) + \frac{y_2^{*2}}{\alpha_2} \left( \frac{1}{\alpha_2} - 1 \right) + \frac{2y_1^* y_2^*}{\alpha_{21}} \left( \frac{\alpha_{12}}{\alpha_1 \alpha_2} - 1 \right) \right] \\ &= \frac{1}{400^2} \left[ \frac{36^2}{0.2061} \left( \frac{1}{0.2061} - 1 \right) + \frac{107^2}{0.3468} \left( \frac{1}{0.3468} - 1 \right) \right. \\ &\quad \left. + \frac{2(36)(107)}{0.0680} \left( \frac{0.0680}{(0.2061)(0.3468)} - 1 \right) \right] = 0.5055 \end{aligned}$$

giving an estimated standard error of  $\sqrt{0.5055} = 0.711$

For the first network there was an average of  $36/6 = 6$  fishes per unit, for the second  $107/11 = 9.727$ , and for the rest 0.

Hence,  $w_1 = w_2 = 6$ ,  $w_3 = w_4 = 69.727$  and  $w_i = 0$  for  $i = 5 \text{ to } 15$ . From (5),

$$\begin{aligned} \hat{\mu}_{HH} &= \frac{1}{15} \left( \frac{36}{6} + \frac{36}{6} + \frac{107}{11} + \frac{107}{11} + \frac{0}{1} + \dots + \frac{0}{1} \right) \\ &= \frac{1}{15} (2 \times 6 + 2 \times 9.727 + 0 + \dots + 0) = 2.097 \end{aligned}$$

fishes per unit.

From (6), the estimated variance is

$$\begin{aligned} \hat{\text{var}}[\hat{\mu}_{HH}] &= \frac{(400-15)}{400(15)(15-1)} \left[ 2 \times (6-2.097)^2 + 2 \times (9.727-2.097)^2 + (0-2.097)^2 \right. \\ &\quad \left. + (0-2.097)^2 + \dots + (0-2.097)^2 \right] = 0.865 \end{aligned}$$

The variance estimators indicate that the HT estimator may be more precise than the HH estimator. Salehi (2002) showed that there are several reasons indicating the HT is the superior estimator.

## Discussion

For using adaptive cluster sampling, it is crucial that population should be clustered and reasonably rare, otherwise there is no point to use it. Almost all researches done on adaptive cluster sampling confirm this fact. When a population consists of some kind of fish which live in school and disperse all over a vast area like a sea one can be convinced to use adaptive sampling.

In such a situation, the following assumptions are needed to be made: i) Probability of detection in units is 1. ii) There is no responsive movement of fish prior to detection, furthermore no detected fish moves to another selected unit in the sample.

Adaptive cluster sampling is more efficient (in the sense that its estimators have lower variances) than simple random sample for rare and clumped population. Its traveling cost is much less than simple random sample. Its disadvantages are: 1) the final sample size is random and therefore unknown. 2) If an inappropriate criterion  $C$  for adding neighborhoods is used, we may end up sampling too many units. On the other hand we might not get enough units. 3) A lot of effort expended in locating initial sample as we travel to the site of each such unit. Each of these disadvantages can potentially be a serious obstacle practice.

To deal with first and second disadvantages, one may use restricted adaptive sampling (Brown, 1994 ; Brown & Manly, 1998 ; Salehi & Seber, 2002) or non-overlap two-stage adaptive cluster sampling (Salehi & Seber, 1997b) which can reasonably solve the problem. For both methods unbiased estimator is available. There are some other methods which can deal with these disadvantages but an unbiased estimator is not available (Quinn *et al.*, 1999).

To deal with third disadvantage, strip adaptive cluster sampling (Thompson, 1991) can be used for which an unbiased estimator is available. Line transect adaptive sampling is also appropriate method (Pollard and Buckland, 1997) for which an unbiased estimator is not available. In this method, distance sampling is used.

### Strip adaptive cluster sampling

It has been shown that Adaptive cluster sampling is more efficient when the initial sample has a good coverage of the area of study. Systematic and Latin square sampling design have such property but reaching to the units selected in the initial sampling is usually costly. In the fishery study, since traveling cost may be more than observation and measuring cost so that initial systematic and Latin square type of sampling design may not be desirable. Line transect sampling design type probably is more suitable for fishery study. In literature, most of fishery and marine studies are associated with line transect sampling.

### Notations and Estimators

Suppose that the area of study is partitioned into  $N$  strip and each strip is divided into  $M$  units (Fig 3). Instead of labeling the units by a single integer we now use a double integer notation. Thus unit  $(i, j)$  represents the  $j$ th secondary unit in the  $i$ th strip, and  $y_{ij}$  is the associated  $y$ -value. Unit  $(i, j)$  is said to satisfy the condition of interest  $C$  if  $y_{ij}$  is in a specified set, defined for example by  $y_{ij} > c$ . In Fig 3 we have  $c = 0$ . The initial sample consists of a selection of  $n_1$  strip and we add further secondary units using the adaptive method described above. Our aim may be to estimate the population total;

$$\tau = \sum_{i=1}^N \sum_{j=1}^M y_{ij} \quad (7)$$

The mean  $\mu = \frac{\tau}{MN}$  or the density  $\rho = \frac{\tau}{A}$  where  $A$  is the area of study.

The modified HT estimate of  $\mu$  is,

$$\hat{\mu} = \frac{1}{MN} \sum_{k=1}^K \frac{y_k^* J_k}{\alpha_k} = \frac{1}{MN} \sum_{k=1}^K \frac{y_k^*}{\alpha_k} \quad (8)$$

with the symbols appropriately redefined below. As before, we can divide up the population of  $MN$  secondary units into  $K$  distinct networks using clusters without

their edge units. Let  $K$  denote the number of networks in the sample. Let  $y_k^*$  denote the sum of the  $y$ -values in the  $k$ th network ( $k = 1, 2, \dots, K$ ) and let  $m_k$  now denote the number of strip in the population that intersect the  $k$ th network. [For the special case when the strip is just a single secondary unit, then  $m_k$  becomes the number of secondary units in the  $k$ th network. We define  $J_k$  to take the value of 1 with (Intersection) probability  $\alpha_k$  if the initial sample of strip intersects the  $k$ th network and 0 otherwise. Unbiased estimator of variance is given by;

$$\hat{\text{var}}[\hat{\rho}] = \frac{1}{\hat{A}^2} \sum_{k=1}^K \sum_{k'=1}^K \frac{(y_k^* - \hat{\rho}a_k^*)(y_{k'}^* - \hat{\rho}a_{k'}^*)}{\alpha_{kk'}} \left( \frac{\alpha_{kk'}}{\alpha_k \alpha_{k'}} - 1 \right) \quad (9)$$

where  $\alpha_k$  and  $\alpha_{kk'}$  are respectively given in formula (3) and (4).

In fishery study, we are usually interested in estimating the density ( $\rho$ ). If the sizes of the secondary units and strips are the same in the population the estimator of  $\rho$  would be  $\hat{\rho} = \hat{\mu}/a$ , where  $a$  is the area of each secondary units. If the sizes of the secondary units are different, say  $a_{ij}$  size of the unit  $j$  in the strip  $i$ , an estimator of  $\rho$  would be,

$$\hat{\rho} = \frac{\sum_{k=1}^K y_k^* / \alpha_k}{\sum_{k=1}^K a_k^* / \alpha_k} = \frac{\hat{\tau}}{\hat{A}} \quad (10)$$

where  $a_k^*$  is the area of network  $k$ . This estimator is ratio type estimator and approximately unbiased. Its approximate of variance is;

$$\text{var}[\hat{\rho}] = \frac{1}{\hat{A}^2} \sum_{k=1}^K \sum_{k'=1}^K \frac{(y_k^* - \rho a_k^*)(y_{k'}^* - \rho a_{k'}^*)}{\alpha_{kk'}} \left( \frac{\alpha_{kk'}}{\alpha_k \alpha_{k'}} - 1 \right), \quad (11)$$

The variance estimator is;

$$\hat{\text{var}}[\hat{\rho}] = \frac{1}{\hat{A}^2} \sum_{k=1}^{\kappa} \sum_{k'=1}^{\kappa} \frac{(y_k^* - \hat{\rho}a_k^*)(y_{k'}^* - \hat{\rho}a_{k'}^*)}{\alpha_{kk'}} \left( \frac{\alpha_{kk'}}{\alpha_k \alpha_{k'}} - 1 \right) \quad (12)$$

The HH estimator is;

$$\tilde{\mu} = \frac{1}{n_1} \sum_{i=1}^{n_1} w_i = \bar{w}, \quad (13)$$

where,

$$w_i = \frac{1}{M} \sum_{k=1}^{\kappa_i} \frac{y_k^*}{m_k} \quad (14)$$

and  $\kappa_i$  is the number of networks that intersect the  $i$ th strip. Since  $\tilde{\mu}$  is the mean of a simple random sample of  $w_i$  values, and  $E[\bar{w}] = \mu$ , we have;

$$\text{var}[\tilde{\mu}] = \frac{\sigma_w^2}{n_1} \left( 1 - \frac{n_1}{N} \right) \quad (15)$$

$$\sigma_w^2 = \frac{1}{N-1} \sum_{i=1}^N (w_i - \mu)^2 \quad (16)$$

An unbiased estimate of the variance of  $\tilde{\mu}$  is;

$$\hat{\text{var}}[\tilde{\mu}] = \frac{s_w^2}{n_1} \left( 1 - \frac{n_1}{N} \right), \quad (17)$$

where,

$$s_w^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (w_i - \tilde{\mu})^2 \quad (18)$$

**Example 2.** In Fig 3, the area of study is partitioned into 400 secondary units. Each strip contains 20 secondary units. The initial sample consists of five randomly selected strips. The secondary units are square units whenever a unit in the sample contains one or more fishes, adjacent units are added adaptively to the sample, the final sample from this procedure is shown in the figure. In this sample, the first (extreme left) of the five strips in the sample intersects two collections of units which satisfy the condition, leading to the additional units added to the sample. The network of units satisfying the condition within the extreme top of these clusters has a total  $y$ -value of  $y_1^* = 106$ , from the sample it can be determined that this network intersects four of the strips in the population. The lower cluster has total  $y$ -value of  $y_2^* = 105$  and also intersects four strips. All other observations in the sample give zero  $y$ -values.

Since  $n_1 = 5$ , the intersection probability for each of the two "non-zero" networks in the sample is;

$$1 - \frac{\binom{20-4}{5}}{\binom{20}{5}} = 0.7183$$

Hence,

$$\hat{\mu} = \frac{1}{400} \left( \frac{106}{0.7183} + \frac{105}{0.7183} \right) = 0.7344 \quad (19)$$

Since two strips intersect both networks, the joint intersection probability for the two sample networks is

$$\alpha_{12} = 1 - \left[ \frac{\binom{20-4}{5} + \binom{20-4}{5} - \binom{20-4-4+2}{5}}{\binom{20}{5}} \right] = 0.5657$$

The sample estimate of variance is;

$$\begin{aligned} \hat{\text{var}}(\hat{\mu}) = & \frac{1}{400} \left\{ \frac{106^2}{0.7183} \left[ \frac{1}{0.7183} - 1 \right] + \frac{105^2}{0.7183} \left[ \frac{1}{0.7183} - 1 \right] \right. \\ & \left. + 2(106)(105) \frac{1}{0.5678} \left[ \frac{0.5678}{0.7183^2} - 1 \right] \right\} = 0.09963 \end{aligned}$$

The variable  $w_1$  associated with the first sample strip is  $w_1 = (1/20)[(106/4) + (105/4)] = 2.6375$ . For the second sample strip the term is;  $w_1 = (1/20)(105/4) = 1.3125$ , and, for the other three strips in the sample,  $w_j = 0$ . Then based on formula 13:

$$\tilde{\mu} = \frac{1}{5}(2.6375 + 1.3125 + 0 + 0 + 0) = 0.79$$

The variance estimate by (16) is:

$$\hat{\text{var}}(\tilde{\mu}) = \frac{(20 - 5)}{20(5)}(1.38966) = 0.2084,$$

in which 1.38966 is the sample variance of the  $w$ -values 2.6375, 1.3125, 0, 0, and 0.

### Advantages and disadvantages of this design

In line transect type sampling design, we can perform our sampling from shore and we do not have to travel to the units which are selected at random and do nothing until we reach the site of selected unit. In this way we save time and budget. In acoustic survey, we may turn on the equipment and register all the way through. In such a case, traveling to the site of initial sample units (in adaptive cluster sampling) has no cost. For a trawl survey, we might not want to tow a net all time that we are traveling with the vessel but we would still prefer to have less travel to reach the site of towing. Line transect (strip) adaptive cluster sampling would be

desirable when we would like not to move around the area of study without measuring the variable of interest e.g. biomass, abundance of a marine life; ect.

Initial sample strips may encounter numbers of large networks and we end up sampling large number of units such that we run out of available resources. Three ways are available to deal with this problem:

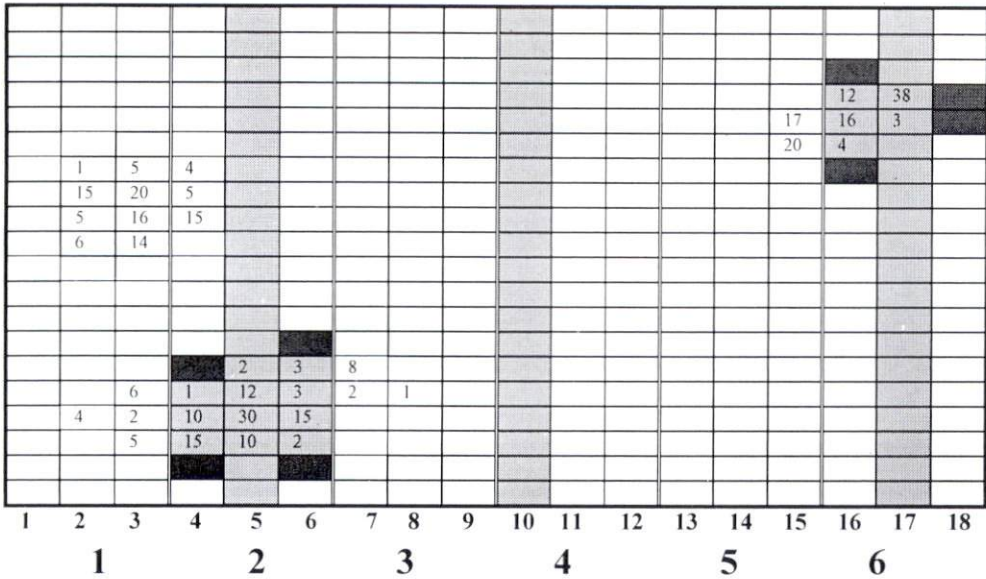
- 1) Using non-overlap two-stage adaptive sampling of strips (Salehi & Seber 1997),
- 2) Using restricted adaptive cluster sampling (Brown 1998),
- 3) Curtailing sampling after a predetermined number of additional steps sampling were conducted (Quinn II, *et.al.* 1999).

We now discuss the first solution in the next section and leave the second and third solutions for the next article.

### **Non-overlap two-stage adaptive cluster sampling of strip**

The area of study is partitioned into Primary Sample Strips (PSS) and a simple random sample of PSS's is selected, then a strip adaptive sampling is carried out in the selected PSS's. In Fig 4, the 18 strips are partitioned into 6 PSS's. A sample of size 3 is selected from the 6 PSS's, then a strip adaptive sampling with initial sample of size one are carried on in each of the selected PSS's. In PSS 2, second strip is selected which is encountered a cluster (network plus edge units) speared over 8 strips of the population. In the adaptively adding stage we will stop at PSS's boundaries. We are now sure that we move at most two strips from the initial selected strip. Therefore, we can limit the potential size of clusters.





**Figure 4:** A population of 360 units is considered as a population of 18 strips. The strips are partitioned into 6 Primary Sample Strips (PSS). The PSS's 2,4 and 6 are selected in the first stage. In the second stage, the second strip of the PSS 2, the first strip of the PSS 4 and the second strip of the PSS 6 are selected as the initial samples. The gray units are selected units and the black units are edge units.

**Notations and Estimators**

Suppose that the population is partitioned into  $L$  PSS's indexed  $l = 1, 2, \dots, L$ . Let  $\lambda$  be the sample size of PSS's. The HT estimator of the mean is,

$$\hat{\mu}_{HT} = \frac{L}{MN} \sum_{l=1}^L \frac{\hat{\tau}_l}{\lambda} \tag{17}$$

where,

$$\hat{\tau}_l = \sum_{k=1}^{\kappa_l} y_{lk}^* / \alpha_{lk}$$

Here  $\kappa_l$ ,  $y_{kl}^*$  and  $\alpha_{lk}$  are respectively the number of networks in the primary sample strip  $l$ , the sum of the  $y$ -values associated with network  $k$ , and the

probability that the initial sample of units in primary sample strip  $l$  intersects network  $k$ . If  $\alpha_{lkk'}$  is the probability that the initial sample of units in primary unit  $l$  intersects both the  $k$  and  $k'$  networks there, then;

$$\alpha_{lk} = 1 - \frac{\binom{N_l - m_{lk}}{n_l}}{\binom{N_l}{n_l}} \quad (18)$$

and

$$\alpha_{lkk'} = \alpha_{lk} + \alpha_{lk'} - \left( 1 - \frac{\binom{N_l - m_{lk} - m_{lk'}}{n_l}}{\binom{N_l}{n_l}} \right) \quad (19)$$

Its variance is again given by;

$$\text{var}[\hat{\mu}_{HT}] = \frac{1}{N^2 M^2} L(L - \lambda) \frac{\sigma_L^2}{\lambda} + \frac{1}{N^2 M^2} \frac{L}{\lambda} \sum_{l=1}^L V_l \quad (20)$$

where  $\sigma_L^2 = \sum_{l=1}^L (\tau_l - \bar{\tau})^2 / (L - 1)$ ,  $\bar{\tau} = \sum_{l=1}^L \tau_l / L$  and

$$V_l = \frac{\sum_{k=1}^{K_l} \sum_{k'=1}^{K_l} y_{lk}^* y_{lk'}^* \frac{\alpha_{lkk'} - \alpha_{lk} \alpha_{lk'}}{\alpha_{lk} \alpha_{lk'}}}{\alpha_{lk} \alpha_{lk'}}$$

The variance of  $\hat{\tau}_l$ . Here  $V_l = 0$  if  $K_l = 0$ . An unbiased estimator of  $\text{var}[\hat{\mu}_{HT}]$  is given by;

$$\hat{\text{var}}[\hat{\mu}_{HT}] = \frac{1}{N^2 M^2} L(L - \lambda) \frac{s_L^2}{\lambda} + \frac{1}{N^2 M^2} \frac{L}{\lambda} \sum_{l=1}^L \hat{V}_l \quad (21)$$

where,

$$s_L^2 = \sum_{l=1}^L (\tau_l - \sum_{i=1}^{\lambda} \hat{\tau}_i / \lambda)^2 / (\lambda - 1)$$

$$V_l = \sum_{k=1}^{\kappa_l} \sum_{k'=1}^{\kappa_l} y_{lk}^* y_{lk'}^* \frac{\alpha_{lkk'} - \alpha_{lk} \alpha_{lk'}}{\alpha_{lk} \alpha_{lk'}}, \quad (22)$$

where  $\kappa_l$  is the number of distinct networks intersected in primary sample strip  $l$ . The HH estimator can be derived similar to the HT estimators (Salehi & Seber, 1997b).

We are now approaching a situation in which adaptive cluster sampling can be used in practice for fishery and marine study. We therefore explain a situation in which researchers come into conclusion to use and investigate the properties of some sort of line transect adaptive cluster sampling for Fishery science survey.

Acoustic surveys of pelagic fish in Namibia, South Africa and Angola (e.g. on sardine, anchovy, sardinellas and juvenile horse mackerel) are imprecise because of the low encounter rate of targets. The problem is particularly severe for Namibian sardine, which is a classic example of a rare, highly-clustered population, difficult to sample using conventional sampling designs. Not only the estimates of biomass for these populations are inaccurate because of the high sampling variance, but also the estimates of variance themselves are often poor because of the skewed sampling distribution of the mean. Attempts to increase the precision by increasing the encounter rate through ad hoc adaptations in survey design once targets are encountered (an approach which has been used extensively on Namibian sardine, using information from purse-seiners acting as scouts) are likely to give biased estimates of abundance and invalid estimates of variance. Such adaptations need to be made according to statistically sound principles and according to strict rules to ensure unbiasedness.

Based on the above problem marine and Coastal Management institution, South Africa and National Marine Information and Research Center Namibia, are doing a three-year project to find a reasonable sampling method for dealing with the above problem. They design a sampling method including a stratified sampling for which in each stratum a strip adaptive cluster sampling either performs with single vessel (ship) or a survey vessel plus scouting vessel(s). In the second

method the strips are measured by survey vessel and the adaptively added units measured by scouting vessel(s).

Another type of line transect sampling was developed by (Pollard and Buckland (1997). In their methods they use only one ship and whenever they come across a school of fish they zigzag for a period and then return to normal route (straight line).

### **Acknowledgment**

I would like to express my sincere thanks to Dr. B. Mojazi Amiri and three anonymous referees for their comments and suggestions. I also would like to acknowledge financial support provided for this research by the BENEFIT institution, South Africa.

### **References**

- Brown, J.A. , 1994. The application of adaptive cluster sampling to ecological studies. *In: Statistics in Ecology and Environmental Monitoring.* (eds. D.J. Fletcher and B.F.J. Manly), Dunedin, New Zealand, University of Otago Press. **2:** 86-97,
- Brown, J.A. and Manly, B.F.J. , 1998. Restricted adaptive cluster sampling. *Environmental and Ecological Statistics.* **5:** 49-63.
- Clausen, D.M. ; Hanselman, D. ; Lunsford, C. ; Quinn II, T. and Heifetz, T. , 1999. Rockfish adaptive sampling experiment in the central Gulf of Alaska, 1998: chartered fishing vessel Unimak Enterprise Cruis 98-01. U.S. Dept. Commer, NOAA, NMFS AFSC Proc. Rept. 99-04. 49 P.
- Francis, R.I.C.C. , 1984. An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research,* **18:**59-71.
- Jolly, G.M. and Hampton, I. , 1990. A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences,* **47:** 1282-1291.
- Lo, C.H. ; Griffith, D. and Hunter, J.R. , 1997. Using a restricted adaptive sampling to estimate pacific hake larval abundance. *CalCOFI Rep.* **38:**103-113.

- Pollard, J.H. and Buckland, S.T. , 1997. A strategy for adaptive sampling in shipboard line transects surveys. Reps. of International Whaling Commission. **47**:921-931.
- Quinn II, T.J. ; Hanselman, D.H. ; Clausen, D.M. ; Heifetz, J. and Lunsford, C. , 1999. Adaptive cluster sampling of rockfish populations, Proceedings of the American Statistical Association, Joint Statistical Meetings, Biometrics Section, pp.11-20.
- Roesch, F.A., Jr. , 1993. Adaptive cluster sampling for forest inventories, Forest Science. **39**:655-669.
- Salehi M., M. and Seber, G.A.F. , 1997a. Adaptive cluster sampling with networks selected without replacement. *Biometrika* **84**:209-219.
- Salehi M., M. and Seber, G.A.F. , 1997b. Two-Stage adaptive cluster sampling. *Biometrics*. **53**:959-970.
- Salehi M., M. , 1999. Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling. *Environmental and Ecological Statistics*. **6**:183-195.
- Salehi M., M. , 2002. Comparison between and Hansen-Hurwitz and Horvitz-Thompson for adaptive cluster sampling. *Environmental and Ecological Statistics* in press.
- Salehi M., M. and Seber, G.A.F., 2002. Unbiased Estimators for Restricted Adaptive Cluster Sampling with Networks Selected With and Without replacement, *Australian and New Zealand Journal of Statistics*. **44**:63-74.
- Seber, G.A.F and Salehi M., M. , 2002. Adaptive Sampling, *Encyclopedia of Biostatistics Volume 1*. 2nd edition. (eds. P. Armitage and T. Colton). John Wiley & Sons, Ltd, Chichester, in press.
- 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.

- Thompson, S.K. , 1990. Adaptive cluster sampling. *Journal of the American Statistical Association.* **85**:1050-1059.
- Thompson, S.K. , 1991. Adaptive cluster sampling: Designs with primary and secondary units, *Biometrics.* **47**: 1103-1115.
- Thompson, S.K. , 1996. Adaptive cluster sampling based on order statistics, *Environmetrics.* **7**:123-133.
- Thompson, S.K. and Seber, G.A.F. , 1996. *Adaptive Sampling.* New York, Wiley.