8.3 ADAPTIVE SAMPLING

Most of the methods discussed in sampling theory are limited to sampling designs in which the selection of the samples can be done before the survey, so that none of the decisions about sampling depend in any way on what is observed as one gathers the data. A new method of sampling that makes use of the data gathered is called *adaptive sampling.* For example, in doing a survey of a rare plant, a botanist may feel inclined to sample more intensively in an area where one individual is located to see if others occur in a clump. The primary purpose of adaptive sampling designs is to take advantage of spatial pattern in the population to obtain more precise measures of population abundance. In many situations adaptive sampling is much more efficient for a given amount of effort than the conventional random sampling designs discussed above. Thompson (1992) presents a summary of these methods.

8.3.1 Adaptive cluster sampling

When organisms are rare and highly clustered in their geographical distribution, many randomly selected quadrats will contain no animals or plants. In these cases it may be useful to consider sampling clusters in a non-random way. Adaptive cluster sampling begins in the usual way with an initial sample of quadrats selected by simple random sampling with replacement, or simple random sampling without replacement. When one of the selected quadrats contains the organism of interest, additional quadrats in the vicinity of the original quadrat are added to the sample. Adaptive cluster sampling is ideally suited to populations which are highly clumped. Figure 8.2 illustrates a hypothetical example.

To use adaptive cluster sampling we must first make some definitions of the sampling universe:

condition of selection of a quadrat: a quadrat is selected if it contains at least y organisms (often y = 1).

neighborhood of quadrat x: all quadrats having one side in common with quadrat x.

edge quadrats: quadrats that do not satisfy the condition of selection but are next to quadrats that do satisfy the condition (i.e. empty quadrats).

network: a group of quadrats such that the random selection of any one of the quadrats would lead to all of them being included in the sample.

These definitions are shown more clearly in Figure 8.3, which is identical to Figure 8.2 except that the networks and their edge quadrats are all shown as shaded.

It is clear that we cannot simply calculate the mean of the 37 quadrats counted in this example to get an unbiased estimate of mean abundance. To estimate the mean abundance from adaptive cluster sampling without bias we proceed as follows (Thompson 1992):

- Calculate the average abundance of each of the networks:

$$w_i = \frac{\sum_k y_k}{m_i} \tag{8.35}$$

where w_i = Average abundance of the *i*-th network
y_k = Abundance of the organism in each of the k-quadrats in the i-th network
m_i = Number of quadrats in the *i*-th network

- From these values we obtain an estimator of the mean abundance as follows:

$$\bar{x} = \frac{\sum_i w_i}{n} \tag{8.36}$$

where $\bar{x}$ = Unbiased estimate of mean abundance from adaptive cluster sampling
n = Number of initial sampling units selected via random sampling

If the initial sample is selected *with replacement,* the variance of this mean is given by:

$$\hat{\mathrm{var}}(\bar{x}) = \frac{\sum_{i=1}^{n} (w_i - \bar{x})^2}{n(n-1)} \tag{8.37}$$

where $\hat{\mathrm{var}}(\bar{x})$ = estimated variance of mean abundance for sampling with replacement and all other terms are defined above.

If the initial sample is selected *without replacement,* the variance of the mean is given by:

$$\hat{\mathrm{var}}(\bar{x}) = \frac{(N-n)\sum_{i=1}^{n} (w_i - \bar{x})^2}{Nn(n-1)} \tag{8.38}$$

where N = total number of possible sample quadrats in the sampling universe

We can illustrate these calculations with the simple example shown in Figure 8.3. From the initial random sample of $n = 10$ quadrats, three quadrats intersected networks in the lower and right side of the study area. Two of these networks each have 2 plants in them and one network has 5 plants. From these data we obtain from equation (8.36):

$$\bar{x} = \frac{\sum_i w_i}{n} = \frac{\left(\frac{2}{7} + \frac{2}{8} + \frac{5}{15} + \frac{0}{1} + \frac{0}{1} + ...\right)}{10} = 0.08690 \text{ plants per quadrat}$$

Since we were sampling without replacement we use equation (8.38) to estimate the variance of this mean:

$$\hat{\text{var}}(\bar{x}) = \frac{(N-n)\sum_{i=1}^{n}(w_i - \bar{x})^2}{Nn(n-1)}$$

$$= \frac{(400-10)\left[\left(\frac{2}{7} - 0.0869\right)^2 + \left(\frac{2}{8} - 0.0869\right)^2 + ...\right]}{(400)(10)(10-1)}$$

$$= 0.0019470$$

We can obtain confidence limits from these estimates in the usual way:

$$\bar{x} \pm t_\alpha \sqrt{\hat{\text{var}}(\bar{x})}$$

For this example with n = 10, for 95% confidence limits t_α = 2.262 and the confidence limits become:

$$0.0869 \pm (2.262)\left(\sqrt{0.0019470}\right) = 0.0869 \pm 0.0983$$

or from 0.0 to 0.185 plants per quadrat. The confidence limits extend below 0.0 but since this is biologically impossible, the lower limit is set to 0. The wide confidence limits reflect the small sample size in this hypothetical example.

When should one consider using adaptive sampling? Much depends on the abundance and the spatial pattern of the animals or the plants being studied. In general the more clustered the population and the rarer the organism, the more efficient it will be to use adaptive cluster sampling. Thompson (1992) shows, for example, from the data in Figure 8.2 that adaptive sampling is about 12% more efficient than simple random sampling for n = 10 quadrats and nearly 50% more efficient when n = 30 quadrats. In any particular situation it may well pay to conduct a pilot experiment with simple random sampling and adaptive cluster sampling to determine the size of the resulting variances.

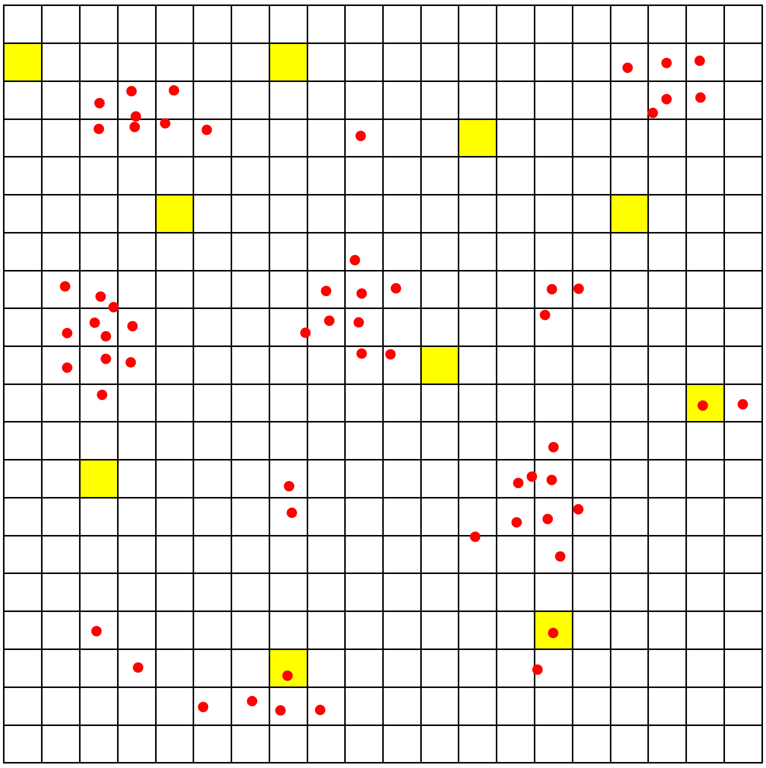

Figure 8.2 A study area with 400 possible quadrats from which a random sample of n = 10 quadrats (shaded) has been selected using simple random sampling without replacement. Of the 10 quadrats, 7 contain no organisms and 3 are occupied by one or more individuals. This hypothetical population of 60 plants is highly clumped.

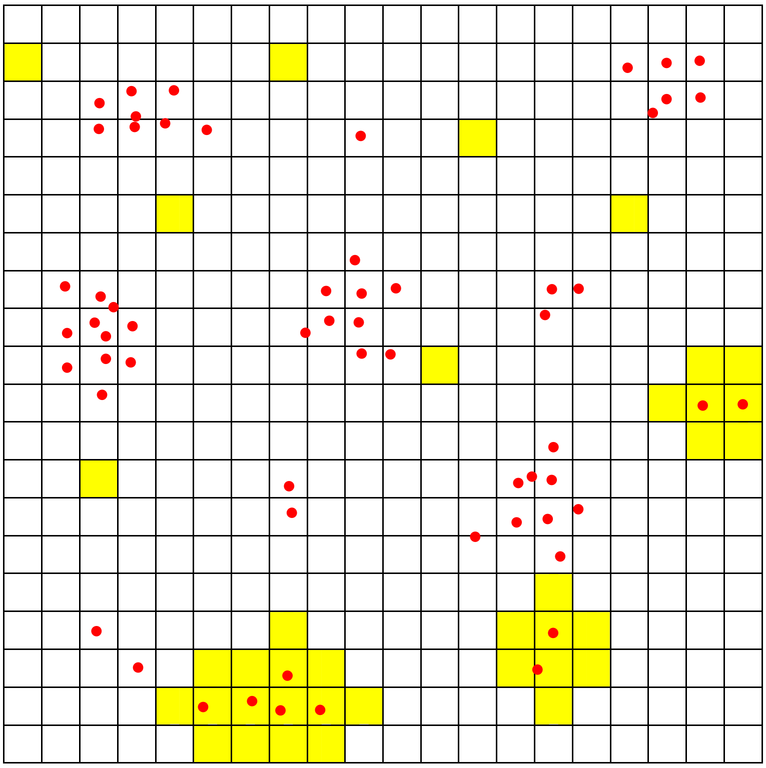

Figure 8.3 The same study area shown in Figure 8.2 with 400 possible quadrats from which a random sample of n = 10 quadrats has been selected. All the clusters and edge quadrats are shaded. The observer would count plants in all of the 37 shaded quadrats.