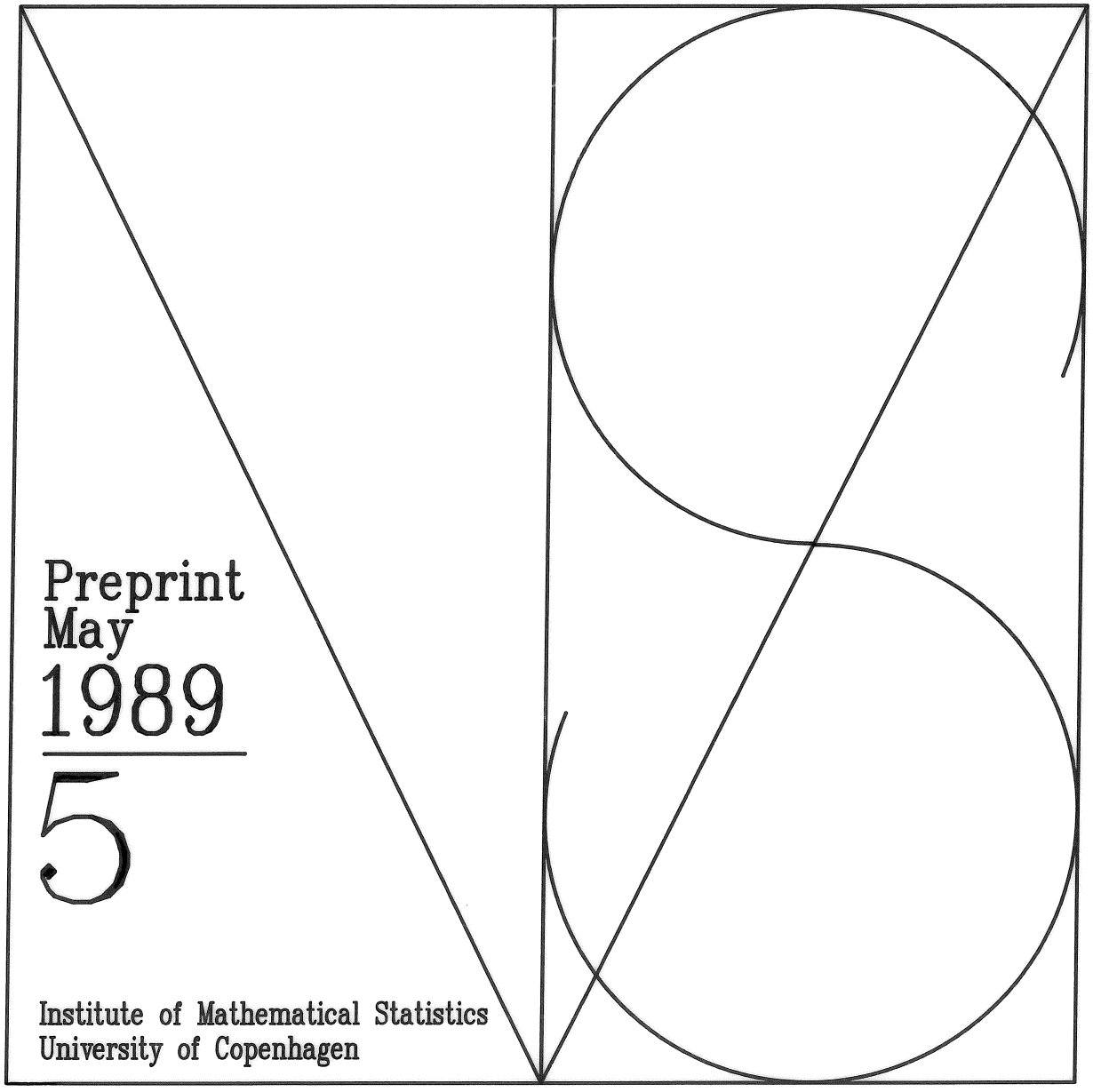
Steven K. Thompson

Adaptive Cluster Sampling



Preprint
May
1989

5

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Steven K. Thompson*

ADAPTIVE CLUSTER SAMPLING

Preprint 1989 No. 5

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

May 1989

*University of Alaska Fairbanks, Fairbanks.

Adaptive Cluster Sampling

STEVEN K. THOMPSON

University of Alaska Fairbanks

Abstract. Sampling designs in which the selection procedure depends on observed values of the variable of interest have been of theoretical interest to statisticians for some time, and, in a variety of real-world sampling situations, researchers would like to be able to adaptively increase sampling effort in the vicinity of observed values that are high or otherwise interesting. This paper describes sampling designs in which, whenever an observed value of a selected unit satisfies a condition of interest, additional units are added to the sample from the neighborhood of that unit. Because such a selection procedure introduces biases into conventional estimators, several estimators are given which are design-unbiased with the adaptive strategy. The Rao-Blackwell Theorem is used to obtain improved unbiased estimators; because of the incompleteness of the minimal sufficient statistic, more than one of these improved estimators are obtained. The results and examples in this paper show that adaptive cluster sampling strategies give lower variance than conventional strategies for certain types of populations and, in particular, provide an extremely effective way of sampling rare, clustered populations.

1. Introduction

In many sampling situations, researchers would like to adaptively increase sampling effort in the vicinity of observed values that are high or otherwise interesting. This paper describes designs in which, whenever the observed value of a selected unit satisfies a condition of interest, additional units are added to the sample from the neighborhood of that unit. Still more units may be added to the sample if any of these additional units also satisfies the condition. Because such selection procedures can introduce biases into conventional estimators, several unbiased estimators

---

April 26, 1989

Research supported by National Science Foundation Grant DMS-8705812 and a grant from the University of Alaska Fairbanks. This paper was written while the author was a sabbatical visitor at the Institute of Mathematical Statistics, University of Copenhagen.

AMS 1980 subject classifications. Primary 62D05, 62L05; secondary 62L12.

Key words and phrases. Adaptive sampling, sequential design, network sampling, informative designs, Rao-Blackwell estimation, cluster sampling.

are given for use with the adaptive designs. Variance formulae and unbiased estimators of variance are also given. For some of the adaptive strategies, simple criteria determine when the adaptive procedure gives lower variance than simple random sampling with equivalent sample size. Based on these results and on the examples evaluated in this paper, adaptive cluster sampling appears to be a highly effective method for sampling populations with natural "clustering" patterns.

The basic idea of the designs in this paper is illustrated in Figure 1, in which the problem is to estimate the mean number of point-objects—which could for example represent locations of animals or mineral deposits—scattered unevenly in a study region partitioned into 400 square sampling units. An initial random sample of 10 units is shown in Figure 1a. Whenever one or more of the objects is observed in a selected unit, the adjacent neighboring units—to the left, right, top or bottom—are added to the sample. When this process is completed, the sample consists of 45 units, shown in Figure 1b. Neighborhoods of units may be defined in many ways other than the spatial proximity system of this example.

A design such as illustrated differs from most classical sampling designs in that the selection procedure depends on observed values of the variable of interest. Motivation for adaptive designs such as this one arises in a number of real-word sampling situations such as the following examples from ecological, geological, and epidemiological studies.

In a survey of a rare and endangered bird species, observers record every bird seen or heard in the vicinity of randomly selected sites. At most of the sites, no birds of the species are detected. When the species is observed at a site, however, subsequent observation of neighboring sites will often reveal additional individuals of the species. Similar patterns have been observed in aerial surveys of polar bears, trawl surveys of fish and shellfish, and surveys of lichen biomass. In studies to assess the mineral or fossil fuel reserves of a region, neighborhoods of initially sampled units with the highest yields may similarly be the most promising for subsequent sampling effort. In an epidemiological study of a sexually transmitted disease, an initial random sample from the population may contain relatively few cases. Subsequent sampling of sexual partners of the infected individuals in the initial sample may reveal considerably higher incidence.

Sampling designs in which the selection procedure depends on observed values of the variable of interest have been of theoretical interest to statisticians for some time. In the paper establishing that the minimal sufficient statistic in finite population sampling is the unordered set of distinct observations together with their unit labels, Basu (1969) expressed the view

that the most efficient designs would be ones in which the selection probabilities were conditional on the observed values. Zacks (1969) described an optimal fixed-sample-size adaptive design from a Bayesian perspective; Soloman and Zacks (1970), while recognizing the theoretical advantage of designs depending on the values of the variable of interest, observed that the optimal design as described would be impractically complex to implement and advocated the development of much simpler sequential designs. Cassel, et. al. (1977) summarized the subsequent literature on sampling designs which make use of observed values ("informative" designs in their terminology), but found little of practical interest there. (Adaptive designs such as described in this paper, being readily implemented and highly efficient for some types of populations, may necessitate a reassessment of the practicality of informative designs in general.)

In the statistical literature on sequential statistical methods [cf. Wald (1947), Chernoff (1972), Woodroofe (1982), Siegmund (1985)] many results are found showing advantages such as increased power, lower expected sample size, or more controllable precision compared to nonsequential methods. Sampling designs which depend on the variable of interest are necessarily sequential, but go somewhat beyond the usual situation considered in sequential statistics in that the unit labels in the sampling data make it possible to choose during a survey not just how much to sample next but which units or group of units to sample next. Although these labels are responsible for many of the complications in the theory of finite population sampling [cf. discussions in Cassel, et. al. (1977), Chaudhuri and Vos (1988)], estimators which use the labels are in some cases better than estimators which do not use the labels. This is certainly the case with the designs in this paper, in which unbiased estimators utilizing information from the labels in the data have lower variance than the unbiased estimator which does not use the unit labels.

The estimators emphasized in this paper are design-unbiased, that is, the unbiasedness is based on the way the sample is selected rather than on assumptions about the population. The concept of unbiased estimation based on the design has had much influence in survey sampling practice since Neyman (1934) and has been the topic of more recent discussions in Särndall (1978), Cassel, et. al. (1979), and Godambe (1982).

By the Rao-Blackwell Theorem, any unbiased estimator that is not a function of the minimal sufficient statistic can be improved upon by taking its conditional expectation given the sufficient statistic. Blackwell's (1947) contribution to the topic was motivated by the problem of obtaining an unbiased estimate of the mean following sequential stopping. The

method has since been used in a sequential context by Ferebee (1983) for estimating the drift of Brownian Motion and by Kremers (1987a), who applied it to sequential estimation of a binomial mean. Kremers (1987b) applied the Rao-Blackwell Theorem to two-stage adaptive sampling of a finite population with sample size depending on the values of initial observations and derived variance and variance estimation expressions for the estimator obtained. In finite population sampling, an unbiased estimator obtained by the Rao-Blackwell method is not in general a unique minimum variance unbiased estimator because the sufficient statistic is not complete, and in this paper, more than one distinct estimator is obtained through the Rao-Blackwell method. (This issue did not arise in Kremers' work because attention was restricted to estimators which ignore sampling unit labels).

Birnbaum and Sirken (1965) described a sampling design, which has subsequently been termed "network" or "multiplicity" sampling, for surveys of patients with rare diseases. In this design, a simple random sample (without replacement) is selected of units which may for example be households or medical institutions. An individual with the disease may be linked to more than one of these units. For example, the individual may be reported not only by his own household but also by households of close relatives; records of a patient with a rare disease may exist at several medical institutions. The probability that the individual is included in the sample is thus related to the number of units to which he/she is linked, called the "multiplicity" of that individual. The additional units to which the individual is linked may or may not be actually observed.

Birnbaum and Sirken derived three unbiased estimators, and variance expressions for the first two, for use with this design. The first estimator divides each observation by its draw-by-draw selection probability, with each observation repeated in the estimator as many times as selected in the sample, and hence is of the Hansen-Hurwitz (1943) type. The second estimator divides each distinct observation by its probability of inclusion in the sample, and hence is of the Horvitz-Thompson (1952) type. (The third estimator depends on the order of selection and may be of less practical interest.) Subsequent papers [Sirken (1970, 1972a, 1972b), Sirken and Levy (1974), Nathan (1976), Levy (1977), Czaja et. al. (1986)] have concentrated on the Hansen-Hurwitz type of estimator. References to many innovative applications are found in Sudman et. al. (1988) and Kalton and Anderson (1986). The network sampling design was first used not to increase efficiency but because it unavoidably arose in the sampling situation (a patient having records at more than one medical institution).

Later papers on the subject recognized its potential for giving lower variance estimates than conventional procedures and for increasing the "yield" of the survey, i.e., the total number of individuals with the disease in the sample.

The designs in this paper are related to network sampling in that selection of certain units may lead to observation of others. Because of the way the decisions to observe additional units depend adaptively on the observed values of the variable of interest, however, the selection and inclusion probabilities are not in general known for all units in the sample. Modifications must therefore be made in estimators of the Hansen-Hurwitz or Horvitz-Thompson types to obtain unbiased estimators. These modified estimators can then be improved if desired by the Rao-Blackwell method.

Seber (1986) and Cormack (1988) have recognized the need for adaptive sampling methods to effectively sample ecological populations, because of the natural clustering tendencies of many such populations or the patchiness of their environments. Some two-stage adaptive designs allowing for increases or decreases in local sampling effort based on observed abundances are described along with some ecological applications in Thompson and Ramsey (1983) and Thompson (1988). The adaptive cluster sampling designs presented in this paper, allowing for much flexibility in the definition of neighborhoods of units, appear to most directly satisfy the inclination of researchers in such fields to increase sampling effort in the vicinity of high observed abundances. In the examples with clustered populations evaluated in this paper, the adaptive strategies give, as anticipated, substantially lower variance than does simple random sampling.

In Section 2 of this paper, adaptive cluster sampling designs are described and some terminology given. Estimators which are unbiased with designs of this type are given in Section 3. Section 4 contains a small example to which the adaptive and conventional strategies are applied and compared. In Section 5, variances and estimators of variance are worked out for the adaptive strategies. Expected sample size and cost are the topics of Section 6. Adaptive sampling and simple random sampling are compared in Section 7. Examples with spatially clustered populations are evaluated in Section 8.

2. Designs

Adaptive cluster sampling refers to designs in which an initial set of units is selected by some probability sampling procedure, and, whenever the

variable of interest of a selected unit satisfies a given criterion, additional units in the neighborhood of that unit are added to the sample. In the designs considered in this paper, the initial sample consists of a simple random sample of n_1 units, selected either with or without replacement.

As in the usual finite population sampling situation, the population consists of N units with labels $1, 2, \ldots, N$ and with associated variables of interest $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$. The sample s is a set or sequence of labels identifying the units selected for observation. The data consist of the observed y-values together with the associated unit labels. The object of interest is to estimate the population mean $\mu = N^{-1} \sum_{i=1}^{N} y_i$ or total $N\mu$ of the y-values. A sampling *design* is a function $p(s|\mathbf{y})$ assigning a probability to every possible sample s. In designs such as those described in this paper, these selection probabilities depend on the population y-values.

It is assumed that for every unit i in the population a neighborhood A_i is defined, consisting of a collection of units including i. These neighborhoods do not depend on the population y-values. Classical (nonadaptive) cluster sampling is a special case in which the neighborhoods form a partition of the population (into clusters), but in general the neighborhoods are overlapping sets of units and do not correspond to clusters. In the spatial sampling examples of this paper, the neighborhood of each unit consists of a set of geographically nearest neighbors, but more elaborate neighborhood patterns are also possible, including a larger contiguous set of units or a noncontiguous set such as a systematic grid pattern around the initial unit. In other sampling situations, neighborhoods may be defined by social or institutional relationships between units. The neighborhood relation is symmetric: if unit j is in the neighborhood of unit i, then unit i is in the neighborhood of unit j.

The condition for additional selection of neighboring units is given by an interval or set C in the range of the variable of interest. The unit i is said to *satisfy the condition* if $y_i \in C$. In the examples of this paper, a unit satisfies the condition if the variable of interest y_i is greater than or equal to some constant c; that is, $C = \{x : x \geq c\}$.

When a selected unit satisfies the condition, all units within its neighborhood are added to the sample and observed. Some of these units may in turn satisfy the condition and some may not. For any of these units that does satisfy the condition, the units in its neighborhood are also included the the sample, and so on.

Consider the collection of all the units that are observed under the design as a result of initial selection of unit i. Such a collection, which

may consist of the union of several neighborhoods, will be termed a *cluster* when it appears in a sample.

Within such a cluster is a subcollection of units, termed a *network*, with the property that selection of any unit within the network would lead to inclusion in the sample of every other unit in the network.

Initial selection of any unit in the network will in fact lead to selection of the entire associated cluster. Any unit in the cluster not satisfying the condition is termed an *edge* unit. While an edge unit does not satisfy the condition, it is in the neighborhood of one or more units that does. Selection of any unit in the network will result in inclusion of all units in the network and all associated edge units, but selection of an edge unit will not result in the inclusion of any other units. These distinctions become important in determining selection probabilities on which unbiased estimators and sample size computations depend.

It will be convenient to use the term network to include isolated (non-edge) units not satisfying the condition in addition to the interconnected groups of one or more units satisfying the condition. Then, given the y-values, the population of units may be partitioned into networks.

A number of sampling designs are possible for the selection of the initial n_1 units in adaptive cluster sampling. Two are considered in this paper:

2.1 Initial sample selected by simple random sampling without replacement. In this design, an initial sample of n_1 units is selected by simple random sampling without replacement. Whenever the y-value of a selected unit satisfies the given criterion, all units in its neighborhood are added to the sample, and the process continues until the neighborhood of every unit satisfying the criterion is included.

Although the n_1 units in the initial sample are distinct due to the without-replacement sampling, the data may contain repeat observations due to selection in the initial sample of more than one unit in a cluster.

The unit i will be included in the sample either if any unit of the network to which it belongs (including itself) is selected as part of the initial sample or if any unit of a network of which unit i is an edge unit is selected. Let m_i denote the number of units in the network to which unit i belongs, and let a_i denote the total number of units in networks of which unit i is an edge unit. Note that if unit i satisfies the criterion C then $a_i = 0$, while if unit i does not satisfy the condition then $m_i = 1$. The probability of selection of unit i on any one of the n_1 draws is $p_i = (m_i + a_i)/N$. The

probability that unit i is included in the sample is

$$\alpha_i = 1 - \binom{N - m_i - a_i}{n_1} \Big/ \binom{N}{n_1}.$$

Neither the draw-by-draw selection probability p_i nor the the inclusion probability α_i can be determined from the data for all units in the sample, because some of the a_i may be unknown.

2.2 Initial sample selected by simple random sampling with replacement. When the initial sample is selected by simple random sampling with replacement, repeat observations in the data may occur due either to repeat selections in the initial sample or to initial selection of more than one unit in a cluster.

With this design, the draw-by-draw selection probability is $p_i = (m_i + a_i)/N$, and the inclusion probability is

$$\alpha_i = 1 - \left(1 - \frac{m_i + a_i}{N}\right)^{n_1}.$$

Under this design, as with the previous one, these probabilities can not be determined from the data for all units.

3. Estimators

Classical estimators such as the sample mean $\bar{y}$, which is an unbiased estimator of the population mean under a nonadaptive design such as simple random sampling, or the mean of the cluster means $\bar{\bar{y}}$, which is unbiased under cluster sampling with selection probabilities proportional to cluster sizes, are biased when used with the adaptive designs described in this paper. (These biases are demonstrated in the small example of the next section.) In this section several estimators are examined which are unbiased for the population mean under the adaptive designs.

The expected value of an estimator t is defined in the design sense, that is, $\mathrm{E}[t] = \sum t_s p(s|\mathbf{y})$, where t_s is the value of the estimate computed when sample s is selected, $p(s|\mathbf{y})$ is the design, and the summation is over all possible samples s.

The sampling strategy—the estimator together with the design—is *design unbiased* for the population mean if $\mathrm{E}[t] = N^{-1} \sum_{i=1}^{N} y_i$ for all population vectors $\mathbf{y} \in \Re^N$. The emphasis on design-unbiased strategies stems from the desire to have estimators whose unbiasedness does not depend on any assumptions about the nature of the population itself.

3.1 The initial sample mean. If the initial sample in the adaptive design is selected by simple random sampling, with or without replacement, the mean $\bar{y}_1$ of the n_1 initial observations is an unbiased estimator of the population mean. This estimator ignores all observations in the sample other than those initially selected.

3.2 A modified Hansen-Hurwitz type of estimator. For sampling designs in which n units are selected with replacement and the probability p_i of selecting unit i on any draw is known for all units, the Hansen-Hurwitz estimator $t_{HH} = (Nn)^{-1} \sum_{k=1}^{n} y_k/p_k$ is an unbiased estimator of the population mean. The n units in the sample may include repeat selections, and an observation is used in the estimator as many times as it is included in the sample.

With the adaptive cluster sampling designs of this paper, the selection probabilities are not known for every unit in the sample. An unbiased estimator can be formed by modifying the Hansen-Hurwitz estimator to make use of observations not satisfying the condition only when they are selected as part of the initial sample. Let Ψ_k denote the network which includes unit k, and let m_k be the number of units in that network. (Recall that a unit not satisfying the criterion is considered a network of size one.) Let $\bar{y}_k^*$ represent the average of the observations in the network which includes the k-th unit of the initial sample, that is, $\bar{y}_k^* = (m_k)^{-1} \sum_{j \in \Psi_k} y_j$. The modified estimator is

$$t_{HH^*} = \frac{1}{n_1} \sum_{k=1}^{n_1} \bar{y}_k^*.$$

To see that t_{HH^*} is unbiased, let z_i indicate the number of times the i-th unit of the population appears in the estimator, which is exactly the number of times the network including unit i is represented in the initial sample. Note that z_i may be less than the number of times unit i appears in the sample, which includes selections of unit i as an edge unit. The random variable z_i has a hypergeometric distribution when the initial sample is selected by simple random sampling without replacement and a binomial distribution when the initial sample is selected by simple random sampling with replacement. With either design, z_i has expected value $E[z_i] = n_1 m_i/N$. Writing the estimator in the form $t_{HH^*} = n_1^{-1} \sum_{i=1}^{N} z_i y_i/m_i$, it follows that $E[t_{HH^*}] = N^{-1} \sum_{i=1}^{N} y_i$, so that t_{HH^*} is a design-unbiased estimator of the population mean.

3.3 Improvement upon $\bar{y}_1$ and t_{HH^*} through the Rao-Blackwell method. Neither $\bar{y}_1$ nor t_{HH^*} is a function of the minimal sufficient

statistic. Therefore, each of these unbiased estimators can be improved upon using the Rao-Blackwell method of taking their conditional expectations given the minimal sufficient statistic.The minimal sufficient statistic D in the finite population sampling setting is the unordered set of distinct, labelled observations; that is, $D = \{(k, y_k) : k \in s\}$, where s denotes the set of distinct units included in the sample. Both $\bar{y}_1$ and t_{HH^*} depend on the order of selection; t_{HH^*} depends on repeat selections, and when the initial sample is selected with replacement, $\bar{y}_1$ also depends on repeat selections.

First consider the estimator $\tilde{y} = E[\bar{y}_1|D]$, the application of the Rao-Blackwell method to the initial sample mean. When the initial sample is selected by simple random sampling, each of the $\binom{N}{n_1}$ possible combinations of n_1 distinct units from the N units in the population has equal probability of being selected as the initial sample. When the initial sample is selected by simple random sampling with replacement, it is easiest to think in terms of the N^{n_1} equally probable sequences, which distinguish order and can include repeat selections, of n_1 units chosen from the N units in the population. Conditional on the minimal sufficient statistic D, all initial samples of n_1 units which give rise through the design to the given value of D have equal selection probability; all other initial samples have conditional probability zero.

Let ν denote the effective sample size, that is, the number of distinct units included in the sample. Since the units of the initial sample are included in the ν distinct units in D, only $\binom{\nu}{n_1}$ combinations or ν^{n_1} sequences need be considered conditional on D. An initial sample which gives rise through the design to the given value of D will be termed *compatible* with D.

A *sample edge unit* is a unit in the sample which does not satisfy the condition but is in the neighborhood of one or more units *in the sample* which does satisfy the condition. Let κ denote the number of distinct networks represented in the sample exclusive of sample edge units. Because of the way the sample is selected, an initial sample of n_1 units gives rise to the given value of D if and only if the initial sample contains at least one unit from each of the κ distinct networks exclusive of sample edge units in D. Letting x_j denote the number of units in the sample from the j-th of these networks, an initial sample of n_1 units from the ν distinct units in D is compatable with D if and only if $x_j \geq 1$ for $j = 1, \ldots, \kappa$.

The conditional expectation of $\bar{y}_1$, given D, is therefore the average of the means of all initial samples which are compatible with D. Indexing the combinations or sequences of n_1 units from ν in any arbitrary way, let $\bar{y}_{1g}$

denote the mean of the y-values of the g-th combination or sequence, and let the indicator variable I_g be one if the g-th combination or sequence is compatible with D and zero otherwise. The number of compatible combinations or sequences is $\xi = \sum_{g=1}^{\eta} I_g$, where $\eta = \binom{\nu}{n_1}$ if the initial sample is selected without replacement and $\eta = \nu^{n_1}$ if it is selected with replacement. The Rao-Blackwell estimator can then be written

$$\tilde{y} = \frac{1}{\xi} \sum_{g=1}^{\eta} \bar{y}_{1g} I_g.$$

Next consider $E[t_{HH*}|D]$, the application of the Rao-Blackwell method to t_{HH*}. The resulting estimator turns out to be precisely $\tilde{y}$, the estimator obtained by application of the method to the initial sample mean. To see that this is the case, it is helpful to consider the statistic D^* consisting of the unordered set of labelled observations together with information about the number of times each unit is included in the sample. That is, $D^* = \{(k, y_k, r_k), k \in s\}$, where s is the set of distinct units included in the sample and r_k is the number of times unit k was included in the sample. The statistic D^* is sufficient but not minimal sufficient.

An initial selection of n_1 units giving rise the statistic D^* determines n_1 networks, some of which may be repeats, contained in D^*. Let κ^* be the number of distinct networks among these. (Note that a sample edge unit forms one of these groups only if it was included in the initial selection; $\kappa^* \geq \kappa$, since no sample edge unit forms one of the κ groups defined earlier.) Because of the way the sample is selected, the same value of the statistic D^* will arise from any initial sample of n_1 units having exactly the given numbers of units in each of the κ^* groups.

Let Ψ_k denote the network which includes unit k, m_k the number of units in it, and $\bar{y}_k^*$ the average of the observations in it. Let w_k be the number of times Ψ_k is represented in the initial sample. (If the unit k in the sample is not a sample edge unit, then $w_k = r_k$. If unit k is a sample edge unit, then w_k equals r_k less the number of times the networks of which it is an edge unit are included in the sample.) Let u_k be the number of times unit k is in the initial sample.

Conditional on D^* (which fixes w_k), the distribution of u_k for any unit k included in the above κ^* networks is Bernoulli with expected value w_k/m_k if the initial sample is selected by simple random sampling without replacement. If the sampling is with replacement, the distribution is binomial with expectation w_k/m_k. For any unit i not included in the κ^* networks with initial representation in D^*, $u_i = 0$.

Writing $\bar{y}_1 = n_1^{-1} \sum_{i=1}^{N} u_i y_i$, the conditional expectation is $E[\bar{y}_1|D^*] = n_1^{-1} \sum_{k=1}^{\kappa^*} \sum_{j \in \Psi_k} w_j y_j / m_j = n_1^{-1} \sum_{i=1}^{n_1} \bar{y}_i^*$, since w_j is constant for $j \in \Psi_k$. Thus, $E[\bar{y}_1|D^*] = t_{HH^*}$. Since D is a function of D^*, $E[\bar{y}_1|D^*] = E[\bar{y}_1|D^*, D]$. Therefore, $E[t_{HH^*}|D] = E\{E[\bar{y}_1|D^*, D]|D\} = E[\bar{y}_1|D] = \tilde{y}$.

Thus, the Rao-Blackwell method applied to either $\bar{y}_1$ or t_{HH^*} leads to the same improved estimator $\tilde{y}$. Notice that observations in the data that were ignored in computing the estimators $\bar{y}_1$ or t_{HH^*} are utilized in forming the estimator $\tilde{y}$.

3.4 A modified Horvitz-Thompson type of estimator. For sampling designs in which the probabilities α_i are known for all units, the Horvitz-Thompson estimator $t_{HT} = N^{-1} \sum_{k=1}^{\nu} y_k / \alpha_k$ is an unbiased estimator of the population mean. The summation is over the ν distinct units in the sample.

With the adaptive designs of this paper, the inclusion probabilities are not known for all units included in the sample. An unbiased estimator can be formed by modifying the Horvitz-Thompson estimator to make use of observations not satisfying the condition only when they are included in the initial sample. Then the probability that a unit is included in the estimator can be computed, even though its actual probability of inclusion in the sample may be unknown. If the initial sample is selected by simple random sampling without replacement, define $\alpha_k^* = 1 - \binom{N-m_k}{n_1} / \binom{N}{n_1}$, where m_k is the number of units in the network which includes unit k. If the initial selection is made with replacement, define $\alpha_k^* = 1 - (1 - m_k/N)^{n_1}$. For any unit not satisfying the condition, $m_k = 1$. Let the indicator variable J_k be zero if the k-th unit in the sample does not satisfy the condition and was not selected in the initial sample; otherwise, $J_k = 1$. The modified estimator is

$$t_{HT^*} = \frac{1}{N} \sum_{k=1}^{\nu} \frac{y_k}{\alpha_k^*} J_k.$$

To see that t_{HT^*} is unbiased, let $z_i = 1$ if unit i is utilized in the estimate and $z_i = 0$ otherwise. For any i, z_i is a Bernoulli random variable with expected value α_i^*. Writing the estimator as $t_{HT^*} = N^{-1} \sum_{i=1}^{N} z_i y_i / \alpha_i^*$ it follows that $E[t_{HT^*}] = N^{-1} \sum_{i=1}^{N} y_i$, the population mean.

3.5 Improvement upon t_{HT^*} through the Rao-Blackwell method. The estimator t_{HT^*} is not a function of the minimal sufficient statistic D because it depends on the order in which the observations are obtained, incorporating an observation which does not satisfy the criterion only if

it is included in the initial part of the sample. The estimator t_{HT^*} can therefore be improved upon using the Rao-Blackwell method to give the unbiased estimator $\hat{y} = E[t_{HT^*}|D]$ having lower variance.

Because of the random sampling of the initial sample, the estimator $\hat{y}$ will, by an argument similar to that of 3.3, be the average of the values of t_{HT^*} associated with the selections, compatible with D, of n_1 units from the ν units in D. Letting $t_{HT_g^*}$ denote the g-th compatible combination or sequence of n_1 units from those in D, and with the notation used in 3.3, the Rao-Blackwell estimator can be written

$$\hat{y} = \frac{1}{\xi} \sum_{g=1}^{\eta} t_{HT_g^*} I_g.$$

The estimator $\hat{y}$, obtained by the Rao-Blackwell method from t_{HT^*}, is not identical with $\tilde{y}$, obtained by the Rao-Blackwell method from either $\bar{y}_1$ or t_{HH^*}. (The difference between the two estimators is demonstrated in the following small example.) The reason for this lack of uniqueness is the lack of completeness of the minimal sufficient statistic D. The incompleteness of D in the finite population sampling situation is due basically to the presence of the unit labels in D. Yet good use is made of these labels in constructing estimators for use with the adaptive designs in this paper—of the five unbiased estimators in this section, all but the initial sample mean depend on the labels in the data.

4. A small example

In this section the sampling strategies are applied to a very small population in order to shed light on the computations and properties of the adaptive strategies in relation to each other and to conventional strategies. The population consists of just five units, the y-values of which are {1,0,2,10,1000}. The neighborhood of each unit includes all adjacent units (of which there are either one or two). The condition is defined by $C = \{x : x \geq 5\}$. The initial sample size is $n_1 = 2$.

With the adaptive design in which the initial sample is selected by simple random sampling without replacement, there are $\binom{5}{2} = 10$ possible samples, each having probability 1/10. The resulting observations and the values of each estimator are listed in Table 1. (If the initial sample is selected with replacement, the corresponding table is most conveniently made in terms of the $5^2 = 25$ possible ordered initial samples, each with equal probability.)

In this population, the 4-th and 5-th units, with the y-values 10 and 1000 respectively, form a network, while the 3-rd, 4-th, and 5-th units, with y-values 2, 10, and 1000, form a cluster. In the fourth row of the table, the 1-st and 5-th units, with y-values 1 and 1000, were selected initially; since $1000 \geq 5$, the single neighbor of the fifth unit, having y-value 10, was added to the sample. Since 10 also exceeds 5, the neighboring unit with y-value 2 is also added to the sample. The computations for the estimators are $t_{HH^*} = (1 + (10 + 1000)/2)/2 = 253$ and $t_{HT^*} = (1/.4 + 10/.7 + 1000/.7)/5 = 289.07$, in which $\alpha_1^* = 1 - \binom{4}{2}/\binom{5}{2} = 0.4$ and $\alpha_2^* = \alpha_3^* = 1 - \binom{3}{2}/\binom{5}{2} = 0.7$. The classical estimator $\bar{y} = 253.25$ is obtained by averaging all four observations in the sample, while $\bar{\bar{y}} = (1 + (10 + 2 + 1000)/3)/2 = 169.67$.

The six distinct values of the minimal sufficient statistic D are indicated by the distinct values of the estimators $\tilde{y}$ and $\hat{y}$, which are obtained by averaging t_{HH^*} and t_{HT^*} respectively over all samples with the same value of D. The seven distinct values of the statistic D^* correspond to the distinct values of the estimator t_{HH^*}.

The population mean is 202.6, and the population variance (defined with $N - 1$ in the denominator) is 198,718. One sees from the table that the unbiased adaptive strategies indeed have expectation 202.6, while the estimators $\bar{y}$ and $\bar{\bar{y}}$, used with the adaptive design, are biased.

With the adaptive design, the effective sample size ν varies from sample to sample, with expected sample size 3.1. For comparison, the sample mean with a simple random sampling design (without replacement) and a sample size of 3.1 (assuming the reader is not disturbed by the artificiality of a noninteger sample size) has, by the standard formula, variance $(198,718)(5 - 3.1)/(5(3.1)) = 24,359$.

From the variances and mean square errors given in the last row of the table, one sees that, for this population, the adaptive design with the estimator $\hat{y}$ has the lowest variance among the unbiased strategies [note, however, the extra digit of reporting precision necessary in the table to show that $var(\hat{y})$ is slightly less than $var(t_{HT^*})$], and that all of the adaptive strategies are more efficient than simple random sampling. Among the five unbiased adaptive strategies, the four which make use of labels in the data have lower variance than the one which does not.

5. Variances and estimators of variance

In this section, variance formulae are given for each of the unbiased adaptive strategies, and unbiased estimators for each of those variances

are given as well. The variance of an estimator is defined in the design sense. If t is an unbiased estimator of the population mean, the variance of t is $var(t) = \sum(t_s - \mu)^2 p(s|y)$, where t_s is the value of the estimate when sample s is selected, $p(s|y)$ is the design, and the summation is over all possible samples.

5.1 The initial mean. Some familiar results about $\bar{y}_1$, the sample mean of a simple random sample of n_1 units, are given here to help establish notation that will be used in succeeding sections. With the population variance defined as $\sigma^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \mu)^2$, the variance of $\bar{y}_1$ is $\sigma^2(N-n_1)/(Nn_1)$ if the sampling is without replacement and σ^2/n_1 if the sampling is with replacement. An unbiased estimator of variance is $\widehat{var}(\bar{y}_1) = \hat{\sigma}_1^2(N-n_1)/(Nn_1)$ in the case of without replacement and $\hat{\sigma}_1^2/n_1$ with replacement, where the initial sample variance is $\hat{\sigma}_1^2 = (n_1 - 1)^{-1}\sum_{k=1}^{n_1}(y_k - \bar{y}_1)^2$.

5.2 The estimator t_{HH^*}. The estimator $t_{HH^*} = n_1^{-1}\sum_{k=1}^{n_1}\bar{y}_k^*$ can be viewed as a sample mean, based on a simple random sample, in which the variable of interest associated with the i-th unit in the population is $\bar{y}_i^*$, the mean of the y-values in the network which includes unit i. The variance of t_{HH^*} is thus

$$var(t_{HH^*}) = \frac{N-n_1}{Nn_1}\sum_{i=1}^{N}\frac{(\bar{y}_i^* - \mu)^2}{(N-1)}$$

if the initial sample is selected without replacement, and $var(t_{HH^*}) = n_1^{-1}\sum_{i=1}^{N}(\bar{y}_i^* - \mu)^2/(N-1)$ if the initial sample is selected with replacement.

An unbiased estimator of this variance is

$$\widehat{var}(t_{HH^*}) = \frac{N-n_1}{Nn_1}\sum_{k=1}^{n_1}\frac{(\bar{y}_k^* - t_{HH^*})^2}{(n_1-1)}$$

if the initial sample is selected without replacement, and $\widehat{var}(t_{HH^*}) = n_1^{-1}\sum_{k=1}^{n_1}(\bar{y}_k^* - t_{HH^*})^2/(n-1)$ if the initial sample is selected with replacement.

5.3 The estimator $\tilde{y}$. Since $\tilde{y} = E[t_{HH^*}]$, where D is the minimal sufficient statistic, the variance of t_{HH^*} can be decomposed as $var(t_{HH^*}) = E[var(t_{HH^*}|D)] + var(E[t_{HH^*}|D])$, so that the variance of $\tilde{y}$ can be written

$$var(\tilde{y}) = var(t_{HH^*}) - E[var(t_{HH^*}|D)].$$

An unbiased estimator of $var(t_{HH^*})$ is $\widehat{var}(t_{HH^*})$, given in the preceding subsection. But by the Rao-Blackwell Theorem, a better unbiased estimator of $var(t_{HH^*})$ is $E[\widehat{var}(t_{HH^*})|D]$. This conditional expectation is obtained as the average, over all compatible selections of n_1 observations from D, of the variance estimates as given in the preceding section. An unbiased estimator of the variance of $\tilde{y}$ is thus provided by

$$\widehat{var}(\tilde{y}) = E[\widehat{var}(t_{HH^*})|D] - var(t_{HH^*}|D).$$

The second term on the right is computed from the sample as $var(t_{HH^*}|D) = \xi^{-1}\sum(t_{HH_g^*} - \tilde{y})^2$, where $t_{HH_g^*}$ denotes the modified Hansen-Hurwitz estimate obtained from the g-th compatible selection, and the summation is over the ξ compatible selections of n_1 observations from D.

Although unbiased, this estimator of variance can, with some sets of data, take on negative values.

5.4 The estimator t_{HT^*}. To obtain the variance of t_{HT^*}, it will be most convenient to change notation to deal with the networks into which the population is partitioned, rather than individual units. Let ζ denote the number of networks in the population and let Ψ_j be the set of units comprising the j-th network. Let m_j be the number of units in network j. The total of the y-values in network j will be denoted $y_{j\cdot} = \sum_{i\in\Psi_j} y_i$. The probability α_i^* (from 3.4) that the unit i is utilized in the estimator is the same for all units within a given network j; this common probability will be denoted π_j. Define the indicator variable v_i to be one if the initial sample contains one or more units from the j-th network and zero otherwise. For any network j, v_j is a Bernoulli random variable with expected value $E[v_j] = \pi_j$ and $var(v_j) = \pi_j(1-\pi_j)$. For two networks j and h, the covariance of the indicator variables is $cov(v_j, v_h) = E[v_j v_h] - E[v_j]E[v_h] = \pi_{jh} - \pi_j\pi_h$, where π_{jh} is the probability that the initial sample contains at least one unit in each of the networks j and h. This joint inclusion probability is $\pi_{jh} = 1 - \left\{\binom{N-m_j}{n_1} + \binom{N-m_h}{n_1} - \binom{N-m_j-m_h}{n_1}\right\} \Big/ \binom{N}{n_1}$ when the initial sample is selected without replacement and $\pi_{jh} = 1 - \{[1 - m_j/N]^{n_1} + [1 - m_h/N]^{n_1} - [1 - (m_j + m_h)/N]^{n_1}\}$ when the initial selection is with replacement.

With the above notation, the estimator t_{HT^*} can be written $t_{HT^*} = N^{-1}\sum_{j=1}^{\zeta} v_j y_{j\cdot}/\pi_j$. The variance of the estimator is, with the convention

that $\pi_{jj} = \pi_j$,

$$var(t_{HT^*}) = \frac{1}{N^2}\sum_{j=1}^{\varsigma}\sum_{h=1}^{\varsigma}\left(\frac{y_{j\cdot}y_{h\cdot}}{\pi_j\pi_h}cov(v_j, v_h)\right)$$
$$= \frac{1}{N^2}\sum_{j=1}^{\varsigma}\sum_{h=1}^{\varsigma}\frac{\pi_{jh} - \pi_j\pi_h}{\pi_j\pi_h}y_{j\cdot}y_{h\cdot}.$$

An unbiased estimator of the variance of t_{HT^*} is

$$\widehat{var}(t_{HT^*}) = \frac{1}{N^2}\sum_{k=1}^{\kappa^*}\sum_{m=1}^{\kappa^*}\frac{(\pi_{km} - \pi_k\pi_m)}{\pi_k\pi_m}\frac{y_{k\cdot}y_{m\cdot}}{\pi_k m},$$

where summation is over the κ^* distinct networks represented in the initial sample.

To see that $\widehat{var}(t_{HT^*})$ is unbiased, let v_{jh} be one if units from both networks i and j are selected in the initial sample and zero otherwise. Then

$$\widehat{var}(t_{HT^*}) = \frac{1}{N^2}\sum_{j=1}^{\varsigma}\sum_{h=1}^{\varsigma}\frac{(\pi_{jh} - \pi_j\pi_h)}{\pi_j\pi_h}\frac{y_{j\cdot}y_{h\cdot}}{\pi_{jh}}v_{jh},$$

and unbiasedness follows since $E[v_{jh}] = \pi_{jh}$.

Just as the estimator t_{HT^*} differs from the usual Horvitz-Thompson estimator in that some observations in the data are not used in t_{HT^*}, the above variance expressions differ from the usual ones for the Horvitz-Thompson estimator in that the variables $y_{j\cdot}$ in the above expressions are network totals, ignoring sample edge units not in the initial part of the sample.

5.5 The estimator $\hat{y}$. Since $\hat{y} = E[t_{HT^*}|D]$, where D is the minimal sufficient statistic, the variance of $\hat{y}$ can, by decomposition of the variance of t_{HT^*}, be written

$$var(\hat{y}) = var(t_{HT^*}) - E[var(t_{HT^*})|D].$$

An unbiased estimator of $var(t_{HT^*})$ is $\widehat{var}(t_{HT^*})$, given in the preceding subsection. By the Rao-Blackwell Theorem, a better unbiased estimator is $E[\widehat{var}(t_{HT^*})|D]$, which is the average, over all compatible selections of n_1 observations from D, of the variance estimates as given in the preceding subsection. An unbiased estimator of the variance of $\hat{y}$ is

$$\widehat{var}(\hat{y}) = E[\widehat{var}(t_{HT^*})|D] - var(t_{HT^*}|D).$$

The second term on the right is the variance of the values of t_{HT} over all compatible selections of n_1 initial observations from D. That is, $var(t_{HT^*}) = \xi^{-1} \sum (t_{HT_g^*} - \hat{y})^2$, where $t_{HT_g^*}$ denotes the value of the modified Horvitz-Thompson type estimator obtained from the g-th compatible selection, and the summation is over the ξ compatible selections of n_1 observations from D.

This estimator of variance, although unbiased, can sometimes take on negative values.

6. Expected sample size and cost

Under any sampling design, the expected value of the effective sample size ν is the sum of the inclusion probabilities: $E[\nu] = \sum_{i=1}^{N} \alpha_i$. For the adaptive cluster sampling designs of this paper, the inclusion probabilities α_i are given in Section 2.

In the examples in this paper, comparisons of adaptive strategies with simple random sampling are made on the basis of expected (effective) sample size. In classical cluster sampling, comparisons are often made on the basis of cost, since it is often less expensive, in terms of time or money, to sample units within a cluster than to select a new cluster. The same may be true in applications of adaptive cluster sampling. A reasonable cost equation might then be $c = n_1 c_1 + n_2 c_2$, where c is total cost, n_1 and n_2 are the initial and subsequent sample sizes respectively, and c_1 and c_2 are constants. In addition, there may in many applications be lower cost associated with observing a unit which does not satisfy the criterion than one that does, in which case the cost equation can be modified accordingly. (For example, if the y-variable is biomass of a plant species on sample plots, the measurement is easier on plots with zero.) When the above conditions apply, the relative advantage of the adaptive to the nonadaptive strategy would tend to be greater than in comparisons based on expected sample size.

7. Adaptive vs. nonadaptive sampling

It was pointed out earlier that the unbiasedness of the adaptive designs in this paper does not depend on the type of population being sampled, because the unbiasedness is design-based. Whether an adaptive design is more efficient or less efficient than a nonadaptive design such as simple random sampling does, however, depend on the type of population being sampled.

Consider adaptive cluster sampling with the initial sample of n_1 units selected by simple random sampling without replacement and with the estimator t_{HH^*}. Since $t_{HH^*} = E[\bar{y}_1|D^*]$, where D^* is the unordered collection of labelled observations with repeat frequencies, the variance of t_{HH^*} can be written

$$var(t_{HH^*}) = var(\bar{y}_1) - E[var(\bar{y}_1|D^*)].$$

Thus the variance of t_{HH^*} will always be less than or equal to the variance of $\bar{y}_1$, which is $\sigma^2(N-n_1)/(Nn_1)$.

The variance of the sample mean of a simple random sample of fixed size n will, by comparison, have variance $\sigma^2(N-n)/(Nn)$. Comparing this quantity with the above expression for the variance of t_{HH^*}, gives the following result: The adaptive strategy will have lower variance than the sample mean of a simple random sample of size n if and only if

$$\left(\frac{1}{n_1} - \frac{1}{n}\right)\sigma^2 < E[var(\bar{y}_1|D^*)].$$

The expression for the variance of t_{HH^*} given in §5 can be rewritten in terms of the ζ distinct networks in the population as follows: $var(t_{HH^*}) = b\sum_{i=1}^{N}(\bar{y}_i^* - \mu)^2 = b\sum_{j=1}^{\zeta}\sum_{k\in\Psi_j}(\bar{y}_k^* - \mu)^2$, where b is the constant term $(N-n_1)/(Nn_1(N-1))$ and Ψ_j is the j-th network in the population. Similarly, the variance of $\bar{y}_1$ can be written $var(\bar{y}_1) = b\sum_{j=1}^{\zeta}\sum_{k\in\Psi_j}(y_k-\mu)^2$. Decomposition of the total sum of squares into terms between and within networks then shows that $E[var(\bar{y}_1|D^*)]$ is the within-network variance, that is,

$$E[var(\bar{y}_1|D^*)] = \frac{N-n_1}{Nn_1(N-1)}\sum_{j=1}^{\zeta}\sum_{k\in\Psi_j}(y_k - \bar{y}_j^*)^2.$$

Thus, adaptive cluster sampling with the estimator t_{HH^*} will be more efficient than simple random sampling if the within-network variance of the population is sufficiently high. This principle is similar to the one that holds in classical cluster sampling, but the adaptive design seeks to find the high variance areas during the survey to most efficiently sample populations in which the locations and shapes of such clusters are not known or can not be predicted ahead of time.

With adaptive cluster sampling, the improved estimator $\tilde{y}$ is more efficient than t_{HH^*}. Since $\tilde{y} = E[\bar{y}_1|D]$, the variance of $\tilde{y}$ can be written

$$var(\tilde{y}) = var(\bar{y}_1) - E[var(\bar{y}_1|D)],$$

and a corresponding result obtained: The adaptive strategy with $\tilde{y}$ will have lower variance than simple random sampling with $\bar{y}$ if and only if

$$\left(\frac{1}{n_1} + \frac{1}{n}\right)\sigma^2 < E[var(\bar{y}_1|D)].$$

Comparing the decompositions of $var(\tilde{y})$ and $var(t_{HH^*})$ and using the relation $var(\tilde{y}) \leq var(t_{HH^*})$, the expected conditional variances satisfy $E[var(\bar{y}_1|D)] \geq E[var(\bar{y}_1|D^*)]$.

8. Examples

In this section, adaptive cluster sampling is examined using three examples, in a spatial setting, of populations exibiting "clustered" patterns. The first example is the population illustrated in Figure 1 of the Introduction. The second example, based on the first, is a population in which the y-values are either zero or one, and hence the within-network variance is zero. The third example, having few units that satisfy the condition and high within-network variability, is what one might call a "rare, clustered population."

The population of each example is contained in a square region partitioned into $N = 20 \times 20 = 400$ units. The neighborhood of each unit consists, in addition to itself, of all adjacent units (i.e., that share a common boundary line). A unit satisfies the condition for additional sampling if the y-value associated with the unit is greater than or equal to one. Because of the two dimensional arrangement of the units, it is convenient to label them with two index variables, corresponding to the position of each unit in a two-dimensional array, and to display the population y-values in array form.

For each example, variances are computed for the estimators t_{HH^*} and t_{HT^*} under the design adaptive cluster sampling with the initial sample of n_1 units selected by simple random sampling without replacement. (In fact, with the populations of the examples, the estimator $\hat{y}$ is identical to t_{HT^*} as a result of all edge units having y-value zero.) Results are listed for a selection of initial sample sizes, from $n_1 = 1$ to $n_1 = 200$.

For comparison, the variance is also computed for the sample mean of a simple random sample (without replacement) with sample size equal to the expected (effective) sample size $E[\nu]$ under the adaptive design. For each adaptive strategy, the relative variance—the variance of the adaptive strategy divided by the variance of the nonadaptive strategy—is also listed.

8.1 Example 1. The population of point objects illustrated in Figure 1 was produced as a realization in the unit square of a Poisson cluster process (cf. Diggle, 1983) with five "parent locations" from a uniform distribution and random Poisson (mean=40) numbers of point objects dispersed in relation to the parent locations with a symmetric Gaussian distribution having standard deviation 0.02. The population mean is 190/400=0.475. The following array contains the y-values giving the number of point-objects in each of the units of the population:

$$
\begin{pmatrix}
0&0&0&0&5&13&3&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&2&11&2&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&3&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&5&39&10&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&5&13&4&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&2&22&3&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&10&8&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&7&22&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{pmatrix}
$$

Table 2 lists the expected sample sizes, variances and relative variances for the different sampling strategies for a selection of initial sample sizes. With an initial sample size of 1, the variances of the adaptive strategies are about equal to that obtained with simple random sampling. The relative advantage of the adaptive strategies increases with increasing n_1. With an initial sample size of 200, the adaptive design leads on average to observing a total of about 224 units. The simple random sampling strategy with that sample size has a variance $1/0.065 = 15.4$ times as large as that obtained with the adaptive strategy and the estimator t_{HT^*}.

8.2 Example 2. The y-values for this population are either zero or one. The population was obtained from that of Example 1 (Figure 1), letting the y-value of each unit indicate the presence or absence of point-objects in that unit. Thus, the pattern of the "1's" in the population is identical to the pattern of nonzero entries in the array above for Example 1.

For such a population, the within-network variance is low—zero, in fact—since every network in the population consists either of a unit with

$y_i = 0$ or a group of one or more units each with $y_i = 1$. Therefore, by the results of Section 7, the adaptive strategy with the estimator t_{HH^*} can not do better than simple random sampling in this situation; the additional observations made after the initial n_1 add to the expected sample size but do not change the value of t_{HH^*}, which is based on the averages of the observations in each of the n_1 networks associated with the initial sample.

The variance computations in Table 3 reveal, however, that the estimator t_{HT^*} used with the adaptive design does turn out to be more efficient than simple random sampling for initial sample sizes of 50 or larger.

8.3 Example 3. This example has approximately the same population mean ($\mu = 189/400 = .4725$) as Example 1, and was produced by the same type of process, but with parameters leading to tighter clustering of the population (2 parent locations, Poisson means 100 and dispersion standard deviation 0.01). The population exemplifies a "rare" population in the sense that few units in the population satisfy the condition—only 9 out of the 400 units, or 2.25 percent, satisfy the condition by having y-value greater than zero. The y-values are given in the array below:

$$
\begin{pmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&2&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&82&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&2&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&86&4&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&10&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{pmatrix}
$$

As listed in Table 4, the relative variance of t_{HT^*} with the adaptive design to that obtained with simple random sampling ranges from 0.357 to 0.072. In other words, the adaptive cluster sampling strategy is 2.8 to 13.9 times as efficient as simple random sampling with this population.

References

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā A* **31** 441-454.

Birnbaum, Z.W., and Sirken, M.G. (1965). Design of sample surveys to estimate the prevelence of rare diseases: Three unbiased estimates. *Vital and Health Statistics* Series 2, No.11. Government Printing Office, Washington.

Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* **18** 105-110.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling.* Wiley, New York.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1979). Prediction theory for finite populations when model-based and design-based principles are combined. *Scand. J. Statist.* **6** 97-106.

Chaudhuri, A. and Vos, J.W.E. (1988). *Unified Theory and Strategies of Survey Sampling.* North-Holland, Amsterdam.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design.* SIAM, Philadelphia.

Cormack, R.M. (1988). Statistical challenges in the environmental sciences: a personal view. *J. Roy. Statist. Soc. Ser. A* **151** 201-210.

Czaja, R.F., Snowdon, C.B., and Casady, R.J. (1986). Reporting bias and sampling errors in a survey of a rare population using multiplicity counting rules. *J. Amer. Statist. Assoc.* **81** 411-419.

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press, New York.

Ferebee, B. (1983). An unbiased estimator for the drift of a stopped Wiener process. *J. Appl. Prob.* **20** 94-102.

Godambe, V.P. (1982). Estimation in survey sampling: Robustness and optimality. *J. Amer. Statist. Assoc.* **77** 393-403.

Hansen, M.M. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* **14** 333-362.

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.

Kalton, G. and Anderson, D.W. (1986). Sampling rare populations. *J. Roy. Statist. Soc. Ser. A* **149** 65-82.

Kremers, W.K. (1987a). An improved estimator of the mean for a sequential binomial sampling plan. *Technometrics* **29** 109-112.

Kremers, W.K. (1987b). Adaptive sampling to account for unknown variability among strata. *Preprint No. 128.* Institut für Mathematik, Universitat Augsburg, Federal Republic of Germany.

Levy, P.S. (1977). Optimum allocation in stratified random network sampling for estimating the prevelence of attributes in rare populations. *J. Amer. Statist. Assoc.* **72** 758-763.

Nathan, G. (1976). An empirical study of response and sampling errors for multiplicity estimates with different counting rules. *J. Amer. Statist. Assoc.* **71** 808-815.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc. Ser A* **97** 558-606.

Särndall, C.E. (1978). Design-based and model-based inference in survey sampling. *Scand. J. Statist.* **5** 27-52.

Seber, G.A.F. (1986). A review of estimating animal abundance. *Biometrics* **42** 267-292.

Siegmund, D. (1985). *Sequential Analysis; Tests and Confidence Intervals.* Springer, New York.

Sirken, M.G. (1970). Household surveys with multiplicity. *J. Amer. Statist. Assoc.* **63** 257-266.

Sirken, M.G. (1972a). Stratified sample surveys with multiplicity. *J. Amer. Statist. Assoc.* **67** 224-227.

Sirken, M.G. (1972b). Variance components of multiplicity estimators. *Biometrics* **28** 869-873.

Sirken, M.G. and Levy, P.S. (1974). Multiplicity estimation of proportions based on ratios of random variables. *J. Amer. Statist. Assoc.* **69** 68-73.

Solomon H. and Zacks, S. (1970). Optimal design of sampling from finite populations: A critical review and indication of new research areas. *J. Amer. Statist. Assoc.* **65** 653-677.

Sudman, S., Sirken, M.G., and Cowan, C.D. (1988). Sampling rare and elusive populations. *Science* **240** 991-996.

Thompson, S.K. (1988). Adaptive sampling. *Proc. Section Survey Research Methods Amer. Statist. Assoc.* 784-786.

Thompson, S.K. and Ramsey, F.L. (1983). Adaptive sampling of amimal populations. *Technical Report 82.* Department of Statistics, Oregon State University, Corvallis.

Wald, A. (1947). *Sequential Analysis.* Wiley, New York.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* SIAM, Philadelphia.

Zacks, S. (1969). Bayes sequential designs of fixed size samples from finite populations. *J. Amer. Statist. Assoc.* **64** 1342-1349.

Department of Mathematical Sciences
University of Alaska Fairbanks
Fairbanks, Alaska 99775

Figure 1. Adaptive cluster sampling to estimate the number of point objects in a study region of 400 units. An initial random sample of 10 units is shown in (a). Adjacent neighboring units are added to the sample whenever one or more of the objects of the population is observed in a selected unit. The resulting sample is shown in (b).

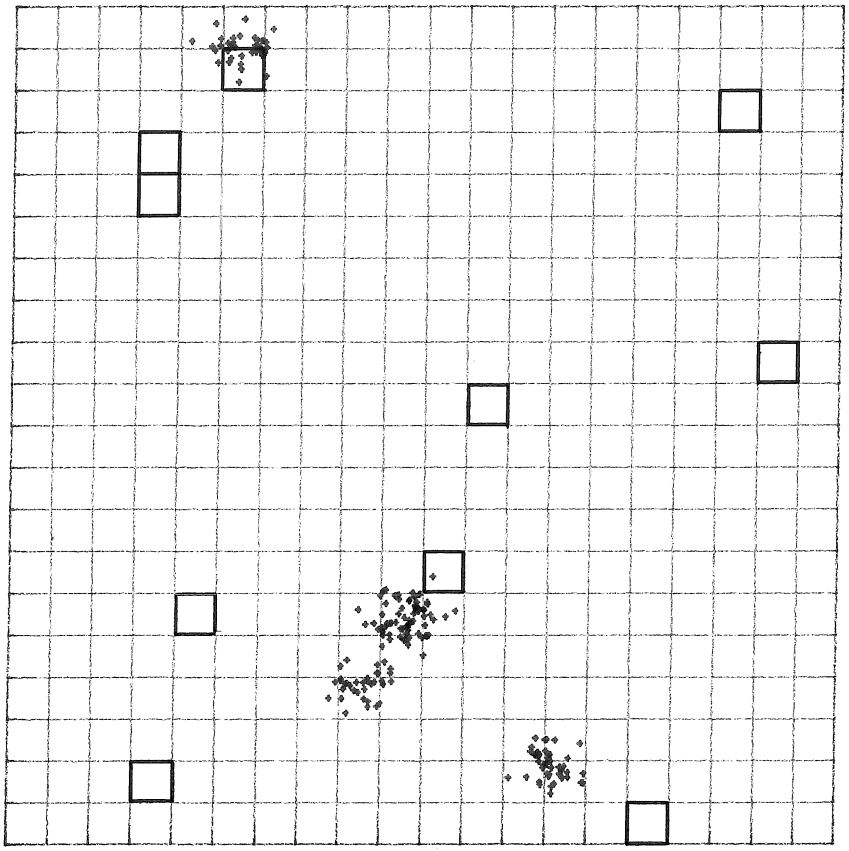


Figure 1a

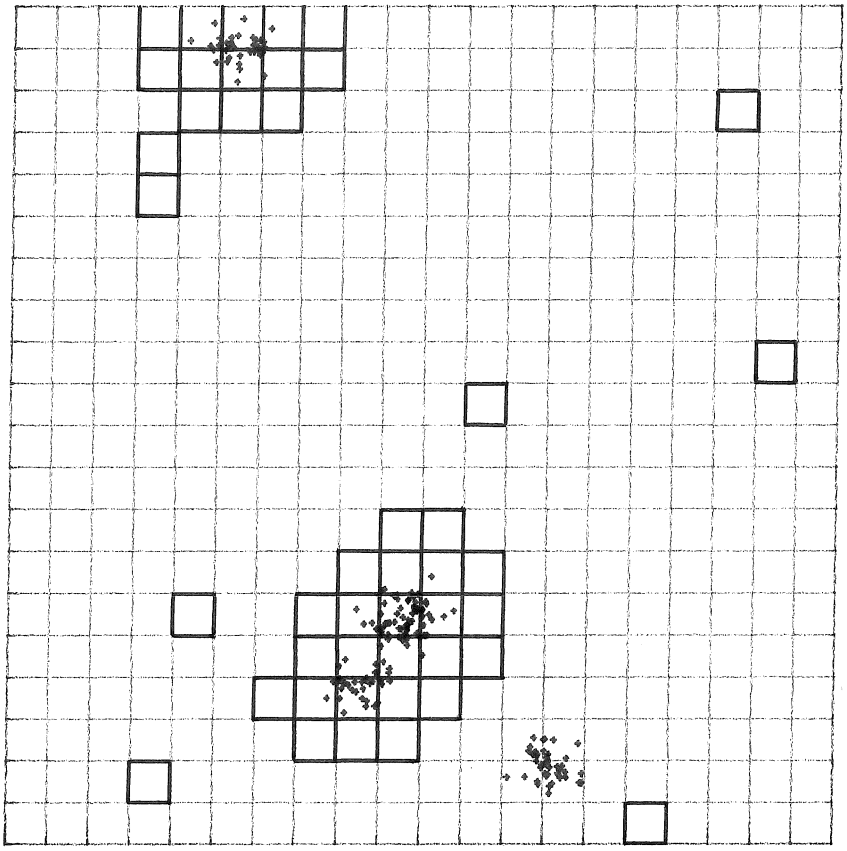


Figure 1b

Table 1

observations	$\bar{y}_1$	t_{HH^*}	$\tilde{y}$	t_{HT^*}	$\hat{y}$	$\bar{y}$	$\bar{\bar{y}}$
1, 0	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1, 2	1.50	1.50	1.50	1.50	1.50	1.50	1.50
1, 10; 2, 1000	5.50	253.00	253.00	289.07	289.07	253.25	169.67
1, 1000; 10, 2	500.50	253.00	253.00	289.07	289.07	253.25	169.67
0, 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0, 10; 2, 1000	5.00	252.50	252.50	288.57	288.57	253.00	168.67
0, 1000; 10, 2	500.00	252.50	252.50	288.57	288.57	253.00	168.67
2, 10; 1000	6.00	253.50	337.33	289.57	289.24	337.33	337.33
2, 1000; 10	501.00	253.50	337.33	289.57	289.24	337.33	337.33
10, 1000; 2	505.00	505.00	337.33	288.57	289.24	337.33	337.33
–	–	–	–	–	–	–	–
Mean :	202.6	202.6	202.6	202.60	202.60	202.75	169.17
Bias :	0	0	0	0	0	0.15	−33.43
MSE :	59,615	22,862	18,645	17,418.4	17,418.3	18,660	18,086

Table 1. All possible outcomes of adaptive cluster sampling for a population of five units with y-values 1,0,2,10,1000, in which the neighborhood of each unit consists of itself plus adjacent units. The initial sample of two units is selected by simple random sampling without replacement. Whenever an observed y-value exceeds 5, the neighboring units are added to the sample. Initial observations are separated from subsequent observations in the table by a semicolon. For each possible sample, the value of each estimator is given. The bottom line of the table gives the mean square error for each estimator. The sample mean of a simple random sample of equivalent sample size has variance 24,359.

Table 2

n_1	$E[\nu]$	$var(t_{HH^*})$	$var(t_{HT^*})$	$var(\bar{y}; srs)$	$\frac{var(t_{HH^*})}{var(\bar{y}; srs)}$	$\frac{var(t_{HT^*})}{var(\bar{y}; srs)}$
1	1.92	4.29705	4.29705	4.28364	1.0031	1.0031
2	3.82	2.14314	2.12386	2.14420	0.9995	0.9905
10	18.26	0.42001	0.38655	0.43240	0.9714	0.8940
20	34.66	0.20462	0.17097	0.21805	0.9384	0.7841
30	49.56	0.13282	0.10030	0.14627	0.9081	0.6857
40	63.26	0.09693	0.06587	0.11012	0.8802	0.5982
50	76.00	0.07539	0.04593	0.08819	0.8548	0.5208
60	87.97	0.06103	0.03322	0.07338	0.8317	0.4528
100	130.80	0.03231	0.01096	0.04258	0.7588	0.2575
200	223.86	0.01077	0.00106	0.01628	0.6616	0.0650

Table 2. Example 1: Variances of t_{HH^*} and t_{HT^*} with adaptive cluster sampling and initial sample size n_1 for the population illustrated in Figure 1. The variance of $\bar{y}$ with simple random sampling is calculated for sample size $E[\nu]$, the expected sample size with the adaptive design. Relative variances of the adaptive to nonadaptive strategies are in the last two columns.

Table 3

n_1	$E[\nu]$	$var(t_{HH^*})$	$var(t_{HT^*})$	$var(\bar{y}; srs)$	$\frac{var(t_{HH^*})}{var(\bar{y}; srs)}$	$\frac{var(t_{HT^*})}{var(\bar{y}; srs)}$
1	1.92	0.04974	0.04974	0.02581	1.9270	1.9270
2	3.82	0.02481	0.02459	0.01292	1.9200	1.9028
10	18.26	0.00486	0.00448	0.00261	1.8659	1.7181
20	34.66	0.00237	0.00198	0.00131	1.8027	1.5074
30	49.56	0.00154	0.00116	0.00088	1.7443	1.3183
40	63.26	0.00112	0.00076	0.00066	1.6909	1.1495
50	76.00	0.00087	0.00053	0.00053	1.6420	0.9997
60	87.97	0.00071	0.00038	0.00044	1.5976	0.8675
100	130.80	0.00037	0.00012	0.00026	1.4577	0.4843
200	223.86	0.00012	0.00001	0.00010	1.2709	0.1050

Table 3. Example 2: Variance comparisons with the y-variable indicating presence or absence of objects in the population of Figure 1.

Table 4

n_1	$E[\nu]$	$var(t_{HH^*})$	$var(t_{HT^*})$	$var(\bar{y}; srs)$	$\frac{var(t_{HH^*})}{var(\bar{y}; srs)}$	$\frac{var(t_{HT^*})}{var(\bar{y}; srs)}$
1	1.26	10.02437	10.02437	28.07236	0.3571	0.3571
2	2.52	4.99962	4.97852	14.01104	0.3568	0.3553
10	12.45	0.97983	0.94253	2.76158	0.3548	0.3413
20	24.58	0.47735	0.43892	1.35488	0.3523	0.3240
30	36.42	0.30986	0.27172	0.88559	0.3499	0.3068
40	47.99	0.22611	0.18865	0.65065	0.3475	0.2899
50	59.32	0.17587	0.13924	0.50945	0.3452	0.2733
60	70.43	0.14237	0.10668	0.41511	0.3430	0.2570
100	113.03	0.07537	0.04392	0.22523	0.3346	0.1950
200	211.74	0.02512	0.00569	0.07887	0.3185	0.0722

Table 4. Example 3: Variance comparisons with a highly clustered population.

PREPRINTS 1988

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and Is Unique.

No. 2 Johansen, Søren and Juselius, Katarina: Hypothesis Testing for Cointegration Vectors - with an Application to the Demand for Money in Denmark and Finland.

No. 3 Jensen, Søren Tolver, Johansen, Søren and Lauritzen, Steffen L.: An Algorithm for Maximizing a Likelihood Function.

No. 4 Bertelsen, Aksel: On Non-Null Distributions Connected with Testing that a Real Normal Distribution Is Complex.

No. 5 Tjur, Tue: Statistical Tables for Personal Computer Users.

No. 6 Tjur, Tue: A New Upper Bound for the Efficiency of a Block Design.

No. 7 Bunzel, Henning, Høst, Viggo and Johansen, Søren: Some Simple Non-Parametric Tests for Misspecification of Regression Models Using Sign Changes of Residuals.

No. 8 Brøns, Hans and Jensen, Søren Tolver: Maximum Likelihood Estimation in the Negative Binomial Distribution.

No. 9 Andersson, S.A. and Perlman, M.D.: Lattice Models for Conditional Independence in a Multivariate Normal Distribution.

PREPRINTS 1989

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33 .

No. 1 Bertelsen, Aksel: Asymptotic Expansion of a Complex Hypergeometric Function.

No. 2 Davidsen, Michael and Jacobsen, Martin: Weak Convergence of Twosided Stochastic Integrals, with an Application to Models for Left Truncated Survival Data.

No. 3 Johansen, Søren: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.

No. 4 Johansen, Søren and Juselius, Katarina: The Full Information Maximum Likelihood Procedure for Inference on Cointegration - with Applications.

No. 5 Thompson, Steven K.: Adaptive Cluster Sampling.