METRON
https://doi.org/10.1007/s40300-020-00195-1

A prediction approach in adaptive sampling

Sanghamitra Pal[1] · **Dipika Patra**[2]

Received: 26 March 2020 / Accepted: 30 December 2020

Abstract
Adaptive cluster sampling (ACS) due to Thompson (J Am Stat Assoc 85(412):1050–1059, 1990) is a tool to survey rare and hidden elements in a population as an improvement over traditional survey procedures. In ACS, if an observed sampled unit satisfies the given criterion of rarity, its neighboring units are added to the sample and this is continued until one is detected with no rarity. Chaudhuri (Calcutta Stat Assoc Bull 50(3–4):238–253, 2000) extended the above to unequal probability sampling as Adaptive sampling. In practice, often network sizes turn out too big demanding high cost and time. So, Chaudhuri et al. (J Stat Plan Inference 121: 175–189, 2004) gave a subsampling technique to restrict the sample size in Adaptive Sampling. Towards this end, Chaudhuri et al. (J Stat Plan Inference 134: 254–267, 2005) developed a sample size restriction technique. But in Adaptive sampling, capturing neighboring rare units turn out difficult because of various hazards. We propose to try Royall's (Biometrika 57: 377–389, 1970) prediction approach here to model features of uncaptured network units. We employ Brewer's (J Am Stat Assoc 74 (368): 911–915, 1979) model-assisted approach to derive a predictor with asymptotic design unbiasedness based on unequal probability samples and examine its efficacy by simulations.

Keywords Adaptive sampling · Asymptotic · Model assisted approach · Prediction approach · Unequal probability

Mathematics Subject Classification 62D05

1 Introduction

Thompson [24] introduced the Adaptive Cluster Sampling (ACS) technique for estimating the population total or mean of the rare and hidden clustered population. Later Thompson [25–27], Thompson and Seber [28], Seber and Salehi [22] elaborated their

✉ Sanghamitra Pal
mitrapal2013@gmail.com

Dipika Patra
dipika.patra1988@gmail.com

[1] Department of Statistics, West Bengal State University, Kolkata, India

[2] West Bengal State University, Kolkata, India

ideas where original samples were drawn by simple random sampling (SRS). But in a survey, we frequently face a situation where an unequal probability sampling scheme has been used. Chaudhuri [5] discussed the necessary changes of ACS when the original sample was drawn by an unequal probability sampling scheme and mentioned the design as adaptive sampling.

Let $U = (1, 2, ..., N)$ be a finite population of N units and $y = \left(y_1, y_2, ..., y_N\right)$ be a study variable relating to the rare and clustered characteristics. We are interested in estimating the population total $\tau_Y = \sum_{i=1}^{N} y_i$. In adaptive sampling, first, an initial sample of size n is chosen by any sampling design, and the value of y is observed on them. A rarity criterion is defined as $y_i > C$, with the predetermined value C. Whenever the observed unit satisfies the pre-defined condition of rarity, well-defined neighboring units are observed for further detection of the rare units. The process is continued until the network is bounded by edge units with no rarity condition. According to Thompson [24], the neighborhood unit's relation is symmetric: if the unit i is a neighbor of the unit j, then j is also a neighbor of the unit $i, i \neq j$. Those neighboring units do not satisfy the given rarity condition, are called the "edge units". A collection of several neighborhoods corresponding to the ith initial unit satisfying the rarity criterion is termed as "cluster". Dropping all the edge units from the cluster, the rest of the units which satisfy the pre-fixed condition belong to the "network" of that particular initial unit. In socio-economic surveys and geographical surveys, network sizes may end in exorbitantly large for pre-assigned initial sample size. Keeping this in mind, Brown and Manly [3] presented a way of limiting the final sample size, called restricted adaptive cluster sampling but their estimators are positively biased though, the bias was successfully estimated by bootstrapping. More recently, Gattone et al. [12] provided a stopping rule criterion to overcome the sample size issue. That practical difficulty was also encountered by Chaudhuri et al. [7] where the initial sample was drawn by unequal probability sampling. They prescribed a design-based method, adding a constraint on the network sizes after ascertaining the networks of the initial sampling units. Earlier Chaudhuri et al. [8] elaborated another design-based method, the subsampling technique approach tackling the large network size problem.

But it does not solve the problem of handling unobserved units in the network. It may be difficult to capture all the neighboring units in a particular network $A(i)$ of the ith sampled unit due to the various hazardous conditions. The investigators may be unable to cover the whole part of the network due to hill or rocky area or bad weather or due to deadly animals etc. as hazardous conditions. Highlighting this deficiency, a model-assisted approach (Särndal et al. [21]) on unequal probability sampling has been proposed here by employing Royall's [20] prediction approach to model the uncaptured part of the network. Brewer's [1] model-assisted approach has been employed to derive the predictor of the population total.

This article is organized as follows: In Sect. 2.1, we briefly review the adaptive sampling design with unequal probability. In Sect. 3, we describe the proposed model-assisted approach along with the unbiased estimators of the population totals and the related estimators of their variances. Using the district wise published data of the National Wasteland and Identification project (https://dolr.gov.in/district-and-category-wise-wastelands-year-2000), Sect. 4 presents numerical illustrations to predict the area of salinity wasteland in Southern India–Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Goa, and Maharashtra. The concluding remarks are finally incorporated.

2 Adaptive sampling design

According to Thompson [24–27], an initial sample of size n is drawn from the population U by simple random sampling without replacement (SRSWOR) and the y values of the sampled units are observed then. If a rare unit bearing the criterion $y_i > C$ is found, its neighborhood units are searched to get more units with rarity condition. This procedure is continued until a network $A(i)$ is formed corresponding to the ith initial unit. Let m_i denotes the number of units in the ith$(i = 1, 2, ..., n)$ network. Thompson [24] suggested two modified estimators of the population total,$\tau_Y = \sum_{i=1}^{N} y_i$ or mean $\overline{\tau}_Y = \frac{1}{N}\sum_{i=1}^{N} y_i$ following Hansen–Hurwitz (HH) [13] and Horvitz–Thompson (HT) [14].

Suppose, K is the number of distinct networks present in the population U of which k are in the final sample while initial sampling begins with SRSWOR. Also, let α_i be the probability of selecting the ith network in the sample and y_i^* be the sum of all y values present in the ith network and $t_{HT} = \sum_{i=1}^{k} \frac{y_i^*}{\alpha_i}$ be the modified HT estimator [13] of population total τ_Y. Here, $\alpha_i = 1 - \binom{N - m_i}{n} \Big/ \binom{N}{n} \forall i = 1, 2, ..., k.$

The variance of the above estimator is $V(t_{HT}) = \sum_{i=1}^{K}\sum_{j=1}^{K} \frac{\alpha_{ij} - \alpha_i\alpha_j}{\alpha_i\alpha_j} y_i^* y_j^*$ where $\alpha_{ij} = 1 - \left\{ \frac{\binom{N-m_i}{n} + \binom{N-m_j}{n} - \binom{N-m_i-m_j}{n}}{\binom{N}{n}} \right\}$ is the probability of selecting the ith and jth networks.

An unbiased estimator of the above variance is given by $v(t_{HT}) = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\alpha_{ij} - \alpha_i\alpha_j}{\alpha_{ij}\alpha_i\alpha_j} y_i^* y_j^*$.

Another unbiased estimator for population total called modified HH estimator [13] is defined as $t_{HH} = \sum_{i=1}^{n} \frac{y_i^*}{m_i} = \sum_{i=1}^{n} \overline{y_i^*}$ whose variance can be written as follows, $V(t_{HH}) = \frac{n(N-n)}{N(N-1)} \sum_{i=1}^{N} \left(\overline{y_i^*} - N^{-1}\tau_Y\right)^2$. An unbiased variance estimator of the modified HH estimator is $v(t_{HH}) = \frac{n(N-n)}{N(n-1)} \sum_{i=1}^{n} \left(y_i^* - t_{HH}\right)^2$.

2.1 Adaptive sampling with unequal probability

Chaudhuri [5] extended the ACS design under the unequal probability sampling scheme. According to his approach, an initial sample s of size n is drawn from the population U by a general sampling design say, $p(s)$ instead of SRSWOR. To do so, a size measure variable z (auxiliary variable may be) has been considered which is known for the entire population. It is obvious that the variable z is not rare. So, it may be available for the entire population. After the selection of the initial sample by unequal probability sampling design, the rest of the adaptive sampling procedure is the same as Thompson [24].

Let the ith sampled unit be observed as rare and starting from it, a network $A(i)$ is formed with cardinality $m_i \ \forall i = 1, 2, \ldots, n$. Then the average of y values say t_i is given by $t_i = \frac{1}{m_i}\sum_{j \in A(i)} y_j$. If an unit not satisfying the rarity criterion is present in the initial sample, it will form a network of size $m_i = 1$. Chaudhuri [5] also observed that $\tau_T = \sum_{i \in U} t_i = \sum_{i \in U} y_i = \tau_Y$. So it may be concluded that the estimation of τ_Y is equivalent to the estimation of τ_T.

Then the HT estimator [14] to estimate $\tau_T = \sum_{i=1}^{N} t_i$, equivalently $\tau_Y = \sum_{i=1}^{N} y_i$ is defined as $t_{HT} = \sum_{i \in s} \frac{t_i}{\pi_i}$ where the first order inclusion probability of the ith unit in the sample is denoted as $\pi_i = \sum_{s \ni i} p(s)$. The distinct network concept of Thompson [24]'s approach has not been used here. For each initial sampled unit, a network is formed.

Following Chaudhuri [5], an unbiased estimator of the related variance $V(t_{HT}) = \sum_{i<j=1}^{N} \sum (\pi_i \pi_j - \pi_{ij})(\frac{t_i}{\pi_i} - \frac{t_j}{\pi_j})^2$ is given by $v(t_{HT}) = \sum_{i<j \in s} \sum \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) (\frac{t_i}{\pi_i} - \frac{t_j}{\pi_j})^2$ denoting π_{ij} as the second order inclusion probability of ith and jth units in the sample.

In the usual sampling process, we use a general sampling design termed as conventional sampling design. In the conventional sampling design, the unbiased estimator of the population total $\tau_Y = \sum_{i=1}^{N} y_i$ will be the same as t_{HT}, only replacing t_i by y_i i.e. the HT estimator of the population total may be termed as,$e_{TS} = \sum_{i \in s} \frac{y_i}{\pi_i}$. The related unbiased variance estimator will be the same as $v(t_{HT})$, just replacing t_i by y_i. The Yates-Grundy [30] form of the related unbiased variance estimator is given by $v(e_{TS}) = \sum_{i<j \in s} \sum \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) (\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2$.

But in practice, network size (m_i) may turn out exorbitantly large demanding high resources even if the initial sample size is moderate in size. Outlining that limitation, Chaudhuri et al. [8] prescribed a design-based approach where initial samples were drawn by unequal probability scheme. Assuming the known structure of the ith network $A(i) \forall i = 1, 2, \ldots, n$, a suitable subsample of the network $A(i)$, say $B(i)$ has been suggested. This work was extended in Chaudhuri et al. [7] in which the subsample is drawn by SRSWOR with cardinality l_i such that $\sum_{i \in s} l_i$ may not exceed the predefined final sample size L. Such a technique is called "sample-size-restrictive adaptive sampling".

3 Proposed model assisted prediction approach

This article is an attempt to lay bare a salient feature of adaptive design in a real survey. Sometimes surveyor may be unable to capture all the specified units due to the crucial causes as already mentioned in the introduction. If the network size is unmanageably large, Chaudhuri et al. [7, 8]'s subsampling approach is required for each network of the sampled units. But this approach can't be implemented in the presence of unobserved units.

Let, the ith sampled unit's network $A(i)$ has a part which is unobserved due to hazardous condition. Denoting $A_s(i)$ as the observed part of the network $A(i)$ and $R_s(i)$ as unobserved part, it can be written as,

$$t_i = \frac{1}{m_i} \sum_{j \in A(i)} y_j = \frac{1}{m_i} \left(\sum_{j \in A_s(i)} y_j + \sum_{j \in R_s(i)} y_j \right).$$

It is obvious that, $A_s(i) \cup R_s(i) = A(i) \forall i \in s$.

Undoubtedly, the second part of the above expression is $\sum_{j \in R_s(i)} y_j$ unknown as these are unobserved. It follows immediately that t_i is also unknown.

Now the main task is to predict the unknown part $R_s(i)$ of the ith unit's network.

A predictor of t_i can be written as,

$$\hat{t}_i = \frac{1}{m_i} \left(\sum_{j \in A_s(i)} y_j + est\left(\sum_{j \in R_s(i)} y_j \right) \right);$$

denoting $est(\sum_{j \in R_s(i)} y_j)$ the estimate of unknown $\sum_{j \in R_s(i)} y_j$.

At this point, the question arises how the network $A(i) \forall i = 1, 2, \ldots, n$ is known in the presence of unobserved units. Let us consider the following two situations:

(a) In estimating the area (in km^2) under soil degradation, high-resolution satellite and airborne images have been primarily used for mapping. The image characteristics, such as color, tone, texture, pattern, and shape identify the different types of wasteland. But the accuracy can be achieved by correlating them with the ground survey.
(b) Traditional methods for monitoring endangered or wild animals mainly rely on ground surveys. But survey regions are often difficult to access on the ground. For this reason, the modern survey considers thermal and infrared imagery to improve the counting of endangered animals like white-tailed deer. A ground survey along with GPS data has been used. So, the network of ith unit i.e.$A(i)$ $\forall i = 1, 2, \ldots, n$ may be known before starting a ground survey.

In this section, we develop here two models using the features of unobserved units. The basic approach in such a problem is to predict the unobserved part of $A(i)$ employing linear regression models with Brewer's [2] and Royall's [20] prediction approach. A well correlated (positively) auxiliary variable is required for such a model. The use of an auxiliary variable in an adaptive sampling design is nothing new. Lee [18] used an auxiliary variable to develop the theory of an ACS two-phase design where in the first phase a sample was based on an auxiliary variable and in the second phase, a sample was selected from the first phase using probability proportional to size with replacement (PPSWR) sampling. Later Felix-Medina and Thompson [10] also employed an inexpensive and easy to measure auxiliary variable in their adaptive cluster double sampling literature. Gattone et al. [11] used a negatively correlated auxiliary variable with the variable of interest to provide a precise estimate. Latpate and Kshirsagar [16, 17] in their literature also considered highly negatively correlated variables to develop negative adaptive cluster sampling and negative adaptive cluster double sampling. In Chaudhuri's unequal probability sampling approach, Chaudhuri et al. [7] employed an easily available auxiliary variable that was not surveyed during the study of the rare variable. They had taken the auxiliary variable from the Indian Population Census (1991) to perform with generalized regression (Greg) estimator (Cassel et al. [4], Särndal et al. [21]).

In this study x is an easily available auxiliary variable whose values are known. We provide here, two different models described in the following Sects. 3.1 and 3.2. By these models the unobserved part of t_i may be predicted.

3.1 Measure without intraclass correlation

Suppose the ith unit is selected in the initial sample and the network $A(i)$ is formed with m_i number of units i.e. $\{y_j : j \in A(i)\}, j = 1, 2, ..., m_i$. Some of the units belonging to this network are unobserved and predicted by the following models.

Model 1: The model M_1 for $A(i)$ may be defined as follows:

$M_1 : y_j = \beta_i x_j + \varepsilon_j$, β_i be the unknown constant and ε_j's are random variables having the following features

$$E_{M_1}(\varepsilon_j) = 0$$

$$V_{M_1}(\varepsilon_j) = \sigma_j^2 \ (unknown)$$

$$E_{M_1}(\varepsilon_j, \varepsilon_k) = 0, \quad i = 1, 2...N;\ j, k \in A(i);\ j \neq, k.$$

Any predictor of t_i under the above model can be expressed as,
$\hat{t}_i = \frac{1}{m_i}\left(\sum_{j\in A_s(i)} y_j + \hat{\beta}_i \sum_{j\in R_s(i)} x_j\right)$.

For simplicity, we impose linearity condition on $\hat{\beta}_i$ i.e.$\hat{\beta}_i = \sum_{j\in A_s(i)} l_j y_j$ where l_i's are constant such that $E_m(\hat{\beta}_i) = \beta_i$. Now, this $\hat{\beta}_i$ is so chosen that $E_m(\hat{\beta}_i - \beta_i)^2$ is least which is equivalent to minimize the model-based variance $E_m(\sum_{j\in A_s(i)} l_j y_j - \beta_i)^2$ subject to the model unbiasedness condition $\sum_{j\in A_s(i)} l_j x_j = 1$.

That implies,

$$\hat{\beta}_{i0} = \left(\sum_{j\in A_s(i)} \frac{y_j x_j}{\sigma_j^2}\right) / \left(\sum_{j\in A_s(i)} \frac{x_j^2}{\sigma_j^2}\right),$$

and it is the best linear unbiased predictor (BLUP) for $\hat{\beta}_i$.

Hence a predictor,

$$\hat{t}_{i0} = \frac{1}{m_i}\left[\sum_{j\in A_s(i)} y_j + \left(\sum_{j\in R_s(i)} x_j\right)\left(\sum_{j\in A_s(i)} \frac{y_j x_j}{\sigma_j^2}\right) / \left(\sum_{j\in A_s(i)} \frac{x_j^2}{\sigma_j^2}\right)\right],$$

of t_i can be derived but this predictor $\hat{t}_{i0}$ is unusable as σ_j^2's $\forall j \in A(i)$ are unknown.

Following Brewer's [1] prediction approach the model-assisted predictor for t_i may be proposed as

$$t_{gi} = \frac{1}{m_i}\left[\sum_{j\in A_s(i)} \frac{y_j}{\pi_j} + b_{Q_i}\left(\sum_{j\in A(i)} x_j - \sum_{j\in A_s(i)} \frac{x_j}{\pi_j}\right)\right],$$

where $b_{Q_i} = \frac{\sum_{j\in A_s(i)} y_j x_j w_j}{\sum_{j\in A_s(i)} x_j^2 w_j}$;$w_j$'s are assignable constants. Following Brewer [1], w_j may be taken as $\frac{1-\pi_j}{\pi_j x_j}$.

The predictor t_{gi} is also called the "Generalized regression predictor" or Greg predictor. (Cassel et al. [4], Särndal et al. [21]). It may be noted that the total of all the x values of the network $A(i)$ is known.

The asymptotic model unbiased expectation of t_{gi} can be written as follows,

$$limE_{M_1}(t_{gi}) = \frac{1}{m_i}\left[E_{M_1}\left(\sum_{j\in A_s(i)} \frac{y_j}{\pi_j}\right) + \frac{limE_{M_1}\left(\sum_{j\in A_s(i)} y_j x_j w_j\right)}{limE_{M_1}\left(\sum_{j\in A_s(i)} x_j^2 w_j\right)}\left(\sum_{j\in A(i)} x_j - E_{M_1}\left(\sum_{j\in A_s(i)} \frac{x_j}{\pi_j}\right)\right)\right]$$

$$= \frac{1}{m_i}\left[\sum_{j\in A(i)} y_j + \frac{\sum_{j\in A(i)} y_j x_j w_j \pi_j}{\sum_{j\in A(i)} x_j^2 w_j \pi_j}\left(\sum_{j\in A(i)} x_j - \sum_{j\in A(i)} x_j\right)\right] = \frac{1}{m_i}\sum_{j\in A(i)} y_j = t_i.$$

Hence, t_{gi} be an unbiased predictor of t_i and $e = \sum_{i\in s} \frac{t_{gi}}{\pi_i}$ be an unbiased predictor of $\sum_{i=1}^{N} t_i = \tau_T$. So for $\tau_Y = \sum_{i=1}^{N} y_i$. It has to be noted that $t_{gj} = t_j$ if all the units belonging to the jth($j \in s$) unit's network are observed.

The recommended predictor for the whole population total $\tau_Y = \sum_{i=1}^{N} y_i$ is

$$e = \sum_{i\in s} \frac{t_{gi}}{\pi_i},$$

and the corresponding variance is

$$\begin{aligned} V(e) &= E_P V_{M_1}(e) + V_P E_{M_1}(e) = E_P V_{M_1}\left(\sum_{i\in s} \frac{t_{gi}}{\pi_i}\right) + V_P E_{M_1}\left(\sum_{i\in s} \frac{t_{gi}}{\pi_i}\right) \\ &= \sum_{i=1}^{N} V_{M_1}\left(\sum_{i\in s} \frac{t_{gi}}{\pi_i}\right) + V_P\left(\sum_{i\in s} \frac{t_i}{\pi_i}\right) = \sum_{i=1}^{N} V_{M_1}\left(\sum_{i\in s} \frac{t_{gi}}{\pi_i}\right) \\ &\quad + \sum_{i<j=1}^{N} \sum \left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{t_i}{\pi_i} - \frac{t_j}{\pi_j}\right)^2, \end{aligned}$$

denoting E_p, V_p as the expectation, variance respectively using the design p and E_{M_1}, V_{M_1} stand for the expectation, variance respectively for the model M_1.

The unbiased variance estimator $v(e)$ can be computed as follows.

Considering $v_1(e) = \sum_{i<j\in s} \sum \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{t_{gi}}{\pi_i} - \frac{t_{gj}}{\pi_j}\right)^2 + \sum_{i\in s} \frac{v_{M_1}(t_{gi})}{\pi_i^2}$ as the unbiased variance estimator of $V(e)$, we may write

$$\begin{aligned} E\left(v_1(e)\right) &= E_P E_{M_1}\left(\sum_{i<j\in s} \sum \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{t_{gi}}{\pi_i} - \frac{t_{gj}}{\pi_j}\right)^2\right) + E_P E_{M_1}\left(\sum_{i\in s} \frac{v_{M_1}(t_{gi})}{\pi_i^2}\right) \\ &= E_P\left(\sum_{i<j\in s} \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right) E_{M_1}\left(\frac{t_{gi}^2}{\pi_i} + \frac{t_{gj}^2}{\pi_j} - 2\frac{t_{gi}}{\pi_i}\frac{t_{gj}}{\pi_j}\right)\right) + E_P\left(\sum_{i\in s} \frac{V_{M_1}(t_{gi})}{\pi_i^2}\right) \\ &= E_P\left(\sum_{i<j\in s} \left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\left(\frac{t_i}{\pi_i} - \frac{t_j}{\pi_j}\right)^2 + \frac{V_{M_1}(t_{gi})}{\pi_i^2} + \frac{V_{M_1}(t_{gj})}{\pi_j^2}\right)\right) \\ &\quad + \sum_{i=1}^{N} \frac{V_{M_1}(t_{gi})}{\pi_i} \\ &= \sum_{i<j=1}^{N} \sum \left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{t_i}{\pi_i} - \frac{t_j}{\pi_j}\right)^2 + \sum_{i=1}^{N} \frac{V_{M_1}(t_{gi})}{\pi_i} \\ &\quad + \sum_{i<j=1}^{N} \sum \left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{V_{M_1}(t_{gi})}{\pi_i^2} + \frac{V_{M_1}(t_{gj})}{\pi_j^2}\right) \\ &= V(e) + \sum_{i<j} \sum \left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{V_{M_1}(t_{gi})}{\pi_i^2} + \frac{V_{M_1}(t_{gj})}{\pi_j^2}\right). \end{aligned}$$

So, $v_1(e) - \sum_{i<j\in s} \sum \frac{(\pi_i\pi_j - \pi_{ij})}{\pi_{ij}}\left(\frac{v_{M_1}(t_{gi})}{\pi_i^2} + \frac{v_{M_1}(t_{gj})}{\pi_j^2}\right)$ is the unbiased estimator of $V(e)$.

Thus, the unbiased variance estimator $v(e)$ is given as,

$$v(e)=\sum_{i<j\in s}\sum\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)(\frac{t_{gi}}{\pi_i}-\frac{t_{gj}}{\pi_j})^2+\sum_{i\in s}\frac{v_{M_1}(t_{gi})}{\pi_i^2}-\sum_{i<j\in s}\sum\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{v_{M_1}(t_{gi})}{\pi_i^2}+\frac{v_{M_1}(t_{gj})}{\pi_j^2}\right)$$

where $v_{M_1}(t_{gi})=\frac{1}{m_i^2}\sum_{k<l\in A_s(i)}\sum\left(\frac{\pi_k\pi_l-\pi_{kl}}{\pi_{kl}}\right)\left(\frac{e_k}{\pi_k}-\frac{e_l}{\pi_l}\right)^2$ and $e_j=y_j-b_{Qi}x_j$.

3.2 Measure with intraclass correlation

In adaptive sampling, a correlation between the units of the same network defined as intraclass correlation may be present and it is denoted by ρ. The following model considers such kind of feature to develop a predictor.

Model 2: $M_2: y_j=\beta_i x_j+\varepsilon_j$, β_i be the unknown constant and ε_j's are random variables having the following features

$$E_{M_2}(\varepsilon_j)=0, V_{M_2}(\varepsilon_j)=\sigma_j^2(Unknown),\ E_{M_2}(\varepsilon_j,\varepsilon_k)=\rho\sigma_j\sigma_k \text{ where } \begin{matrix} i=1,2...N \\ j,k\in A(i) j\neq k \end{matrix}.$$

Then t_i can be estimated through the best linear unbiased predictor.

Now the model unbiasedness condition $E_{M_2}(\hat{t}_i)=t_i$ implies $E_{M_2}\left((\hat{\beta}_i-\beta_i)\sum_{j\in R_s(i)}x_j\right)=0$. It is possible if $E_{M_2}(\hat{\beta}_i)=\beta_i$ with the linearity condition $\hat{\beta}_i=\sum_{j\in A_s(i)}B_jy_j$ which is equivalent to $\sum_{j\in A_s(i)}B_jx_j=1$.

Thus, we have to minimize the mean square error (MSE) $MSE_i=E_{M_2}(\hat{t}_i-t_i)^2$ subject to the condition $\sum_{j\in A_s(i)}B_jx_j=1$.

Mean square error for the predictor of t_i is given as

$$\begin{aligned}
MSE_i &= E_{M_2}(\widehat{t}_i-t_i)^2=E_{M_2}(\widehat{\beta}_i\sum_{j\in R_s(i)}\frac{x_j}{m_i}-\sum_{j\in R_s(i)}\frac{y_j}{m_i})^2\\
&=\frac{1}{m_i^2}E_{M_2}((\widehat{\beta}_i-\beta_i)\sum_{j\in R_s(i)}x_j-\sum_{j\in R_s(i)}(y_j-\beta_ix_j))^2=\frac{1}{m_i^2}[(\sum_{j\in R_s(i)}x_j)^2E_{M_2}(\widehat{\beta}_i-\beta_i)^2\\
&\quad+E_{M_2}(\sum_{j\in R_s(i)}\left(y_j-\beta_ix_j\right))^2-2E_{M_2}[((\widehat{\beta}_i-\beta_i)\sum_{j\in R_s(i)}x_j)(\sum_{j\in R_s(i)}(y_j-\beta_ix_j))]]\\
&=\tfrac{1}{m_i^2}[(\textstyle\sum_{j\in R_s(i)}x_j)^2E_{M_2}(\sum_{j\in A_s(i)}(B_j\left(y_j-\beta_ix_j\right))^2)+E_{M_2}(\sum_{j\in R_s(i)}\varepsilon_j)^2\\
&\qquad-2E_{M_2}[(\textstyle\sum_{j\in R_s(i)}x_j\sum_{j\in A_s(i)}B_j(y_j-\beta_ix_j))\sum_{j\in R_s(i)}\varepsilon_j]]\\
&=\frac{1}{m_i^2}[(\sum_{j\in R_s(i)}x_j)^2E_{M_2}(\sum_{j\in A_s(i)}B_j\varepsilon_j)^2\\
&\quad+E_{M_2}(\sum_{j\in R_s(i)}\varepsilon_j)^2-2(\sum_{j\in R_s(i)}x_j)E_{M_2}[(\sum_{j\in A_s(i)}B_j\varepsilon_j)(\sum_{k\in R_s(i)}\varepsilon_k)]]\\
&=\frac{1}{m_i^2}[(\sum_{j\in R_s(i)}x_j)^2(\sum_{j\in A_s(i)}B_j^2E_{M_2}(\varepsilon_j^2)+\sum_{\substack{k,j\in A_s(i)\\k\neq j}}\sum B_kB_jE_{M_2}(\varepsilon_k\varepsilon_j))\\
&\quad+E_{M_2}(\sum_{j\in R_s(i)}\varepsilon_j^2+\sum_{\substack{j,k\in R_s(i)\\j\neq k}}\sum\varepsilon_j\varepsilon_k)-2(\sum_{j\in R_s(i)}x_j)E_{M_2}(\sum_{j\in A_s(i)}\sum_{k\in R_s(i)}B_j\varepsilon_j\varepsilon_k)
\end{aligned}$$

For the simplicity of derivation, we have followed Chaudhuri and Stenger [6], Särndal et al. [21], to approximate the unknown σ_j where $\sigma_j = \sigma_e x_j$, with $\sigma_e (> 0)$ as unknown and $x_j's$ known. Then, we get

$$MSE_i = \frac{1}{m_i^2}\sigma_e^2\left[(\sum_{j\in R_s(i)} x_j)^2(\sum_{j\in A_s(i)} B_j^2 x_j^2 + \sum_{\substack{k,j\in A_s(i)\\ k\neq j}}\sum \rho B_k B_j x_k x_j)\right.$$

$$\left. + \left(\sum_{j\in R_s(i)} x_j^2 + \sum_{\substack{j,k\in R_s(i)\\ j\neq k}}\sum \rho x_j x_k\right) - 2\rho\left(\sum_{j\in R_s(i)} x_j\right)^2\left(\sum_{j\in A_s(i)} B_j x_j\right)\right]$$

$$= \frac{1}{m_i^2}\sigma_e^2\left[\left(\sum_{j\in R_s(i)} x_j\right)^2\left(\rho + (1-\rho)\sum_{j\in A_s(i)} B_j^2 x_j^2\right) + (1-\rho)\sum_{j\in R_s(i)} x_j^2 - \rho\left(\sum_{j\in R_s(i)} x_j\right)^2\right]$$

Taking λ as a Lagrangian multiplier and solving the following equation,

$$\frac{\partial}{\partial B_j}\left[\frac{\sigma_e^2}{m_i^2}\{(\sum_{j\in R_s(i)} x_j)^2(\rho + (1-\rho)\sum_{j\in A_s(i)} B_j^2 x_j^2 + (1-\rho)\sum_{j\in R_s(i)} x_j^2 - \rho\left(\sum_{j\in R_s(i)} x_j\right)^2 - \lambda\left\{\sum_{j\in A_s(i)} B_j x_j - 1\right\}\right] = 0,$$

we get, $\lambda = 2\frac{\sigma_e^2}{m_i^2}(1-\rho)\frac{(\sum_{j\in R_s(i)} x_j)^2}{(\sum_{j\in A_s(i)} 1)}$.

Consequently, $B_j = \frac{1}{x_j\left(\sum_{j\in A_s(i)} 1\right)}$; and $\hat{\beta}_i = \sum_{j\in A_s(i)} \frac{y_j}{x_j}\left(\frac{1}{\sum_{j\in A_s(i)} 1}\right)$.

Hence, the predictor of t_i is

$$\hat{t}_i = \frac{1}{m_i}\left[\sum_{\mathrm{j\in A_s(i)}} y_j + \left(\sum_{\mathrm{j\in A_s(i)}} \frac{y_j}{x_j}\left(\frac{1}{\sum_{j\in A_s(i)} 1}\right)\right)\left(\sum_{\mathrm{j\in R_s(i)}} x_j\right)\right].$$

Also, the resulting value of the MSE_i is $MSE_{io} = \frac{\sigma_e^2}{m_i^2}(1-\rho)\left\{\sum_{j\in R_s(i)} x_j^2 + \frac{\left(\sum_{j\in R_s(i)} x_j\right)^2}{\left(\sum_{j\in A_s(i)} 1\right)}\right\}$.

Under the above model,

$$e' = \sum_{i\in s}\frac{\hat{t}_i}{\pi_i}$$

is an unbiased predictor for $\sum_{i=1}^N t_i = \tau_T$ and also for $\tau_Y = \sum_{i=1}^N y_i$.
Then the unbiased estimator of variance $v(e')$ is

Table 1 ANOVA table for a sample by adaptive sampling

Source	Sum of squares	Degrees of freedom	Expected mean square
Between Cluster	$\sum_{i=1}^{n}(\hat{t}_i-\hat{\bar{t}})^2$	$n-1$	$\sigma_e^2(1-\rho)+\frac{\rho\sigma_e^2}{n-1}\left(m-\sum_{i\in s}\frac{m_i^{'2}}{m}\right)$
Within Cluster	$\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j\in A_s(i)}\left(y_j-\hat{t}_i\right)^2$	$\sum_{i=1}^{n}m_i^{'}-1$	$\sigma_e^2(1-\rho)$

$^{*}m_i^{'}$ = number of observed units in the ith unit's network

$^{**}m=\sum_{i\in s}m_i^{'}$

$$v(e')=\sum_{i<j\in s}\sum\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{\hat{t}_i}{\pi_i}-\frac{\hat{t}_j}{\pi_j}\right)^2+\sum_{i\in s}\frac{v_{M_2}(\hat{t}_i)}{\pi_i^2}-\sum_{i<j\in s}\sum\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\frac{v_{M_2}(\hat{t}_i)}{\pi_i^2}+\frac{v_{M_2}(\hat{t}_j)}{\pi_j^2}\right)$$

and

$v_{M_2}(\hat{t}_i)=MSE_{io}$, the function of two unknowns σ_e^2 and ρ. These unknowns can be computed by the ANOVA procedure (see Valliant et al. [29]) given in Table 1.

The analysis of variance is based on the following relation

$$\begin{aligned}\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j\in A(i)}\left(y_j-\bar{y}\right)^2&=\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j\in A(i)}\left(y_j-\bar{t}\right)^2\\&=\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j\in A(i)}(y_j-t_i)^2+\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j\in A(i)}(t_i-\bar{t})^2\\&=\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j\in A(i)}(y_j-t_i)^2+\sum_{i=1}^{N}(t_i-\bar{t})^2.\end{aligned}$$

i.e. we may write,

Total sum of squares = within cluster sum of squares + between cluster sum of squares.

The ANOVA table provides two equations with the two unknowns that can be easily solved.

4 Simulation study

In this section, we conduct a simulation to evaluate the performance of the proposed models. Considering two populations naming *Population 1* and *Population 2*, we have shown the efficacy of our models. In *Population 1*, we have considered data published by the National Afforestation and Eco-Development Board. The district-wise different types of wastelands (like Gullied land, scrubland, Marshy land, Salinity/ alkalinity land, etc.) with their area (in km^2) were published in https://dolr.gov.in/district-and-category-wise-wastelands-year-2000. In *Population 2*, we have considered the data containing the study variable along with a highly correlated auxiliary variable from Chutiman and Chiangpradit [9] (Sect. 5.2). The study area is a square grid of $20\times 20=400$ plots. The correlation between the study variable and the auxiliary variable is $r=0.91$.

In Population 1, we are interested in estimating the salinity wasteland areas (in km^2). Districts are considered here as units. The population consists of six states (Andhra

Pradesh, Karnataka, Kerala, Tamil Nadu, Goa, and Maharashtra) of India including 86 districts. 23 districts contain wastelands, out of those 86 districts under study.

As we are interested in estimating the salinity affected areas, the variable of interest $y = (y_1, y_2, ..., y_{86})$ is defined as the area of "Salinity wasteland of the district (in km^2)" and $y_i > 0$ be the "predetermined value of rarity criterion" $i = 1, 2, ..., 86$. The neighborhood of a unit is defined here by the four surrounding districts in the east, west, north, and south direction, ignoring the states' boundaries. For this purpose of study, two auxiliary variables x and z (strictly non-rare) are considered. "Area of the district (in km^2)" is defined as the auxiliary variable x which is used in prediction purpose and partially known. "Total wasteland of the district (in km^2)" is used as a size measure variable to draw an initial sample from the population with an unequal probability scheme. The correlation between the study variable y and the auxiliary variable x is $r = 0.608$. We assume this data as the population and the population total is $\tau_Y = \sum_{i=1}^{N} y_i = 1070.72$ km^2. In the population, we find that the districts Chittoor, Kurnool, Nellore, Prakasam, Anantapur, Cuddapah, Chitradurga, and Chikmagalur are in the same network with its size 8. It can be said more clearly that if any of those districts appear in the initial sample, we need to survey all those eight districts. A similar finding is that the five districts Khammam, Srikakulam, Nalgonda, Vizianagaram and East Godavari are in the same network. All the districts of Tamil Nadu except Kanniyakumari have salinity wasteland and they are also in the same network of size 9. It may possible to have more than one unit from the same network in the initial sample. Following the arguments of Chaudhuri [5], the networks are considered repeatedly in the estimation step.

Without loss of generality, in this study, we are assuming that the districts Prakasam, Anantapur, Nalgonda Chitradurga, Salem, Tiruchirapalli, North Arcot and Dharmapuri are unobserved due to some hazardous conditions.

An initial sample of size 19 is taken from those 86 districts by Lahiri-Midzuno-Sen [15, 19, 23] sampling strategy using the size measure variable z. Whenever an initial unit (district) is found with the criterion of the presence of salinity, its network is selected. Suppose the investigator is unable to visit all the districts belonging to the network. Then the unobserved part of the network is predicted using the proposed models as discussed in Sects. 3.1 and 3.2.

For example, in a particular sample of size 19, two rare units may be found and automatically using adaptive sampling, two networks are selected. So, we need to observe those networks but due to some hazardous conditions, some districts remain unobserved. In that situation, our proposed models work effectively.

In order to estimate the population total of the study variable $y = \left(y_1, y_2, \ldots, y_{400}\right)$ from *Population 2*, we have defined the rarity condition as $y_i > 0\, i = 1, 2, \ldots, 400$. In this population, 57 plots contain a value of y greater than zero and form five distinct networks of size 12, 28, 7, 5, and 2. Here the simulation study is carried out by selecting an initial sample of size $n = 60, 80, 100, 120$ according to Lahiri-Midzuno-Sen [15, 19, 23] sampling strategy. For this purpose, a size measure variable z is required and we have simulated the values of z as $z_i = 20 + 4x_i + \epsilon_i$ where $\epsilon_i \sim N(30, 22)$. Without loss of generality, we are assuming 5 plots are unobserved in the network of size 28 and 2 plots are unobserved in the network of size 12. Those unobserved parts of the networks are predicted by the proposed methods.

The proposed predictors for the two models are compared with each other through the average length (AL), average coefficient of variance (ACV) and average coverage probability (ACP). The proposed predictors are also compared with Chaudhuri [5]'s adaptive sampling and general sampling or conventional sampling (as described in Sect. 2) while the unobserved situation is not turned up. For better comparison, we

Table 2 Estimate, estimated standard deviation, confidence interval limits of 512th simulation considering Population 1 (Here $\tau_Y = 1070.72$ km^2)

	Sample size = 19, Sampling estimator: Horvitz Thompson			
	Proposed model-assisted prediction approach		Adaptive Sampling	Conventional sampling
	Model 1 $\rho = 0$	Model 2 $\rho \neq 0^*$		
Estimate $(\hat{t})$	1027.26	1075.24	1050.05	1130.73
Estimated St. dev. $\left(\sqrt{v(\hat{t})}\right)$	310.519	207.753	167.521	395.091
$t_L(s)$	418.642	668.044	721.709	356.351
$t_U(s)$	1635.878	1482.355	1378.391	1905.108
CI length $t_U(s) - t_L(s)$	1217.244	814.391	656.682	1548.756

* Estimated Intraclass correlation $(\hat{\rho}) = 0.823224$

consider the same initial sample for all the methods and the simulation is done 10,000 times from the above population to judge the efficacy. Denoting $\hat{t}$ as the unbiased estimator of the population total τ_Y with its unbiased variance estimator as $v(\hat{t})$, the confidence interval $CI(s)$ is calculated for each simulated sample and we also check whether the $CI(s)$ covers τ_Y or not. The simple difference between the two endpoints of $CI(s)$ *i.e.* $t_U(s) - t_L(s)$ is called the length of the confidence interval ("CI Length $t_U(s) - t_L(s)$" in Table 2). The probability that the unknown total τ_Y is contained in the confidence interval $CI(s) = \left[t_L(s), t_U(s)\right]$ for every s is called coverage probability (CP) of the interval. Suppose k intervals are found to cover the parameter τ_Y, the proportion $\frac{k}{10000}$ should attain the desired confidence level $1 - \alpha$ (here 95%). The average of the estimated

Table 3 Comparison of different models considering Population 1 (Here $\tau_Y = 1070.72$ km^2)

	Sample size = 19, sampling estimator: Horvitz Thompson			
	Proposed model-assisted prediction approach		Adaptive Sampling	Conventional Sampling
	Model 1 $\rho = 0$	Model 2 $\rho \neq 0^*$		
Estimate $(\hat{t})$ (515th simulation)	1067.09	1090.164	1028.472	1147.86
Estimated st. dev. $\left(\sqrt{v(\hat{t})}\right)$ (515th simulation)	325.537	208.482	169.259	410.067
ACV	30.761	19.631	16.517	34.309
AL	1290.645	820.712	656.831	1610.262
ACP	97.1	62.7	73.4	100

* Estimated Intraclass correlation$(\hat{\rho}) = 0.823224$

Table 4 Comparison of different models considering Population 2 (Here $\tau_Y = 487$)

		Sample Size	Estimate $(\hat{t})$ (515th simulation)	Estimated St. Dev. $\left(\sqrt{v(\hat{t})}\right)$ (515th simulation)	ACV	AL	ACP
Conventional sampling		60	408.9594	191.7882	47.3156	751.8098	73.0
		80	417.6743	170.259	40.8249	667.4153	76.7
		100	421.7926	155.0044	36.855	607.6171	78.4
		120	413.3557	134.4735	32.505	527.136	76.9
Adaptive sampling		60	461.3855	165.7853	37.5349	649.8789	87
		80	461.8556	138.6597	30.975	543.546	87.5
		100	453.7186	120.228	27.0603	471.2939	89
		120	458.9649	106.8164	23.6921	418.7204	90.8
Proposed model-assisted prediction approach	Model 1 $\rho = 0$	60	423.6545	156.7549	38.499	614.4791	82.3
		80	434.5506	135.2009	32.0245	529.9875	85.3
		100	443.9651	120.0361	27.6919	470.5414	87.3
		120	446.409	106.2455	24.2204	416.4825	88.1
	Model 2 $\rho \neq 0^*$	60	472.8275	180.2433	40.0416	706.5536	90.6
		80	478.0443	160.5983	34.7839	629.5482	93.5
		100	470.5861	147.7994	32.1869	579.3737	94.1
		120	477.8636	138.8306	29.1236	552.0561	96.8

* Estimated Intraclass correlation $(\hat{\rho}) = 0.72$

coefficient of variation $\left(cv = 100\frac{\left(\sqrt{v(\hat{t})}\right)}{\hat{t}} \right)$ is called ACV. The value $\left(\sqrt{v(\hat{t})}\right)$ represents the estimated standard deviation of the estimator $\hat{t}$ which is tabulated in Table 2, Table 3 and Table 4 as "Estimated St. Dev.$\left(\sqrt{v(\hat{t})}\right)$". The average length of confidence intervals is called AL. Usually, ACV which is less than 10% is excellent and at most 30% is acceptable. For better understanding, a particular simulation (512th) result is shown in Table 2 for *Population 1*. Table 3 reports the simulation result across the 10,000 samples for *Population 1*. Table 4 represents the simulation result with varying sample sizes for *Population 2*.

Considering *Population 1* with 26.74% rare units and Table 3, the estimate of the salinity wasteland for the 515th simulation is 1147.86 km^2 by the conventional sampling (method without adaptive sampling). In that case, the ACV turns out 34.309% which is very high. However, for Chaudhuri [5]'s adaptive sampling design ACV turns out 16.517%. In the presence of unobserved units, we can't perform Chaudhuri [5]'s adaptive sampling design. In that case, the performances of the proposed models give us satisfactory results.

In Table 4, we have compared the results for different initial sample sizes. For each sample size, Chaudhuri [5]'s adaptive sampling performs better than the conventional sampling in terms of ACV, AL and ACP where all the sampled units are observed. In case of unobserved units in the network, the performances of both the proposed models are demonstated. Population 2 contains 14.25% rare units. From the simulation results of Tables 3 and 4, it is observed that Model 1 is performing better in Population 1 than Population 2 in terms of coverage probabilities. In terms of AL and ACV, both models perform well. The ACV values (in %)may be accepted upto 30 in any practical survey situation. In Population 2, ACP value (in %) is close to 95 which is preferred. In terms of ACP, Model 2 performs well. ACV and AL are decreasing but ACP is increasing as the sample size increases.

5 Concluding remarks

Adaptive sampling may be an effective tool with a rare and clustered population. As the study variable is rare and clustered, Chaudhuri [5]'s adaptive sampling design performs well than the conventional sampling design (see Sect. 2) which is obvious. But these can be implemented only if the surveyor does not face the unobserved situation of the network $A(i)$. If the network size is exorbitantly large and surveyors are unable to capture all the units in the network due to some hazardous conditions, the proposed prediction models give a contribution to overcome such a problem. In particular, Model 2 takes into account the intra-class correlation present among the units of the network of a particular sampled unit. Through the comparison criterion ACP, ACV and AL, we have shown the efficacy of Model 1 and Model 2. For a clustered population, it is obvious that the units in the same cluster are correlated and it is termed as intraclass correlation (ρ). Model 1 may be suitable where intraclass correlation estimate values are low. In other cases Model 2 may be recommended.

But the computation of the variance estimate for Model 2 is the complicated one. In that case, Model 1 may be a good alternative. So, we may conclude that the proposed two models are equally efficient and can be used in estimating rare and clustered population parameters.

Acknowledgements The authors are grateful to the editor and the referees for their valuable comments and suggestions.

References

[1] Brewer, K.R.: Ratio estimation an ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process. Austral. J. Stat. **5**(3), 95–105 (1963)
[2] Brewer, K.R.: A class of robust sampling designs for large-scale surveys. J. Am. Stat. Assoc. **74**, 911–915 (1979)
[3] Brown, J.A., Manly, B.J.: Restricted adaptive cluster sampling. Environ. Ecol. Stat. **5**, 49–63 (1998). https://doi.org/10.1023/A:1009607403647
[4] Cassel, C.M., Sarndal, C.E., Wretman, J.H.: Some results on generalized difference estimation and generalized regression estimation for finite population. Biometrika **63**, 615–620 (1976)
[5] Chaudhuri, A.: Network and adaptive sampling for unequal probabilities. Calcutta Stat. Assoc. Bull. **50**, 237–253 (2000)
[6] Chaudhuri, A., Stenger, H.: Survey Sampling: Theory and Methods, 2nd edn. Chapman and Hall, CRC Press, Taylor & Francis Group, New York (2005)
[7] Chaudhuri, A., Bose, M., Dihidar, K.: Sample size-restrictive adaptive sampling: an application in estimating localized elements. J. Stat. Plan. Inference **134**, 254–267 (2005)
[8] Chaudhuri, A., Bose, M., Ghosh, J.K.: An application of adaptive sampling to estimate highly localized population segments. J. Stat. Plan. Inference **121**, 175–189 (2004)
[9] Chutiman, N., Chiangpradit, M.: Ratio estimator in adaptive cluster sampling without replacement of networks. J. Probab. Stat. **14**(2), 1–6 (2014)
[10] Felix-Medina, M.H., Thompson, S.K.: Adaptive cluster double sampling. Biometrika **91**, 877–891 (2004)
[11] Gattone, S.A., Mohamed, E., Dryver, A.L., Munnich, R.T.: Adaptive cluster sampling for negatively correlated data. Environmetrics **27**(2), 103–113 (2016)
[12] Gattone, S., Mohamed, E., Di, B.T.: Adaptive cluster sampling with clusters selected without replacement and stopping rule. Environ. Ecol. Stat. **23**, 453–468 (2016)
[13] Hansen, M.M., Hurwitz, W.N.: On the theory of sampling from finite populations. Ann. Math. Stat. **14**, 333–362 (1943)
[14] Horvitz, D.G., Thompson, D.J.: An generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. **47**, 663–685 (1952)
[15] Lahiri, D.B.: A method of sample selection providing unbiased ratio estimates. Bull. Int. Stat. Inst. **3**(2), 133–140 (1951)
[16] Latpate, R.V., Kishirsagar, J.: Negative adaptive cluster sampling. Model Assist. Stat. Appl. **14**, 65–81 (2019)
[17] Latpate, R.V., Kshirsagar, J.K.: Negative adaptive cluster double sampling. Stat. Appl. **16**, 115–131 (2018)
[18] Lee, K.: Two phase adaptive cluster sampling with unequal probabilities selection. J. Korean Stat. Soc. **27**, 265–278 (1998)
[19] Midzuno, H.: On the sampling system with probability proportional to the sum of the sizes. Ann. Inst. Stat. Math. **3**, 99–107 (1952)
[20] Royall, R.M.: On finite population sampling theory under certain linear regression models. Biometrika **57**, 377–389 (1970)
[21] Sarndal, C.E., Swensson, D.E., Wretman, J.H.: Model Assisted Survey Sampling. Springer, New York (1992)
[22] Seber, G.A., Salehi, M.M.: Adaptive sampling designs inference for sparse and clustered populations. Springer (2013). https://doi.org/10.1007/978-3-642-33657-7
[23] Sen, A.R.: On the estimator of the variance in sampling with varying probabilities. J. Indian Soc. Agric. Stat. **5**, 119–127 (1952)
[24] Thompson, S.K.: Adaptive cluster sampling. J. Am. Stat. Assoc. **85**, 1050–1059 (1990). https://doi.org/10.1080/01621459.1990.10474975
[25] Thompson, S.K.: Adaptive cluster sampling: designs with primary and secondary units. Biometrics **47**, 1103–1115 (1991). https://doi.org/10.2307/2532662
[26] Thompson, S.K.: Stratified adapative cluster sampling. Biometrika **78**, 389–397 (1991). https://doi.org/10.2307/2337266

[27] Thompson, S.K.: Sampling. Wiley (1992)
[28] Thompson, S.K., Seber, G.A.: Adaptive Sampling. Wiley, New York (1996)
[29] Valliant, R., Dorfman, A.H., Royall, R.M.: Finite Population Sampling and Inference: A Prediction Approach. Wiley, New York (2000)
[30] Yates, F., Grundy, P.: Selection without replacement from within strata with probabilities proportion to size. J. R. Stat. Soc. Ser. B **15**, 235–261 (1953)