Special Proceedings: ISBN #: 978-81-950383-5-0
26th Annual Conference, 26-28 February 2024; pp 181-195

Two-stage Adaptive Cluster Sampling: A Prediction Approach

Sanghamitra Pal[1] **and Dipika Patra**[2]
[1]*Department of Statistics, West Bengal State University, North 24 Parganas, West Bengal - 700126*
[2]*Department of Statistics, Seth Anandram Jaipuria College, Kolkata, West Bengal -700005*

Received: 05 August 2024; Revised: 16 August 2024; Accepted: 20 August 2024

Abstract

Adaptive Cluster Sampling (ACS) due to Thompson's (1990) is a useful tool to survey rare and clustered population. Salehi and Seber (1997) described a two-stage ACS design that used simple random sampling without replacement (SRSWOR) of primary units and then the ACS of secondary units within each of the selected primary unit. Two variations on this design were proposed in their paper depending on whether networks in secondary units are allowed to cross primary unit boundaries or not.

In executing the adaptive sampling design, it is observed that the collection of information from all neighbouring rare units becomes challenging due to various hazards. Pal and Patra (2021) duly addressed the issue and proposed predictors of the population total considering appropriate superpopulation models with suitable assumptions in single stage ACS. The current work is an attempt to find predictors for two-stage ACS under same situation. To illustrate the findings, a numerical example has been carried out.

Key words: Adaptive cluster sampling; Horvitz-Thompson estimator; Prediction approach; Superpopulation; Two-stage designs; Unequal probability sampling.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

Let $U = (1, 2, \ldots, N)$ be a finite population and $y = (y_1, y_2, \ldots, y_N)$ be the variable of interest bearing rarity and clustered characteristics. It is challenging to survey such population through any traditional sampling methods due to the absence of such units in sample with enough number. Thus, the estimation procedures related to the traditional sampling methods such as simple random sampling, stratified sampling may underestimate the population parameters. Thompson's (1990) adaptive cluster sampling reduces the effort to adapt enough number of rare units in the sample and increases the precision. This design has been recently gaining attention because of its greater efficiency. Thompson (1991a) introduced the idea of primary units and secondary units in ACS. The design was further extended by

Corresponding Author: Dipika Patra
Email: dipika.patra1988@gmail.com

Thompson (1991b, 1992). Chaudhuri (2000), Pal and Patra (2021, 2023); Patra and Pal (2023) developed this design under unequal probability sampling designs. A monograph of Seber and Salehi (2013) covers many advents of this design.

In estimating $\tau = \sum_{i=1}^{n} y_i$ by ACS design, an initial sample s of size n is drawn by a probability sampling design and the y-values are observed. Wherever the observed unit satisfies the pre-considered condition of rarity say, $y_i > c$, the uniquely defined neighbouring units (for example - South, North, East, and West) are observed for further detection of rarity. Now, if some of them are found to meet the rarity condition, their neighbouring units are also observed and such procedure continues until a unit is detected with no rarity. It is worth noting that the neighbourhood relation is symmetric. Now, to proceed further in details, one need to know few related terminologies like cluster, edge units, network etc. All neighbouring units corresponding to an initial sampling unit form a cluster. Edge unit is the neighbouring unit that does not satisfy the rarity condition. Thus, each cluster is bounded by edge units. Eliminating all edge units from a cluster, the remaining units that meet the pre-considered rarity condition belong to the network of that particular initial sampling unit. It is also noteworthy, if a unit in s does not satisfy the rarity condition, its network consists of that unit only.

Salehi and Seber (1997), Rocco (2008) and many others, strengthened the literature of Two-stage ACS design. In their proposal, a sample of primary units (PSU) is selected first by simple random sampling without replacement (SRSWOR). Then, an initial sample is taken from secondary units within each selected primary unit, to carry out the ACS design. Surveyors then have two possibilities to stop the adaptively adding procedure of secondary units. Either they can stop at the boundary of PSU (non-overlapping scheme) or allow overlapping into neighbouring PSUs (overlapping scheme). However, in execution stage, surveyors may be unable to observe all the neighbouring secondary units due to hazardous conditions. This deficiency was highlighted in Pal and Patra (2021) for single stage ACS design under unequal probability sampling. Appropriate superpopulation models were adopted there to employ Royall (1970) prediction approach. Implementation of Pal and Patra (2021) approach in two-stage ACS design becomes critical for overlapping scheme. Thus, some modifications are needed following Royall (1976), Valliant *et al.* (2000). In this paper, the main contribution is to develop prediction approach for two-stage ACS-overlapping scheme.

Section 2 elaborately describes the estimation procedure of two-stage ACS design. The next section describes how a superpopulation approach can be used in two-stage ACS to predict the population total or mean in presence of various hazards. Suitable predictors and mean square errors (MSEs) are derived in Section 4. Section 5 illustrates our contribution with a numerical example. Finally, it is concluded in Section 6.

2. Two-stage ACS

Suppose the population U of size N can be partitioned into M primary units of sizes $N_i, i = 1, 2, \ldots, M$ and y_{ij} denotes the y-value of the j^{th} ($j = 1, 2, \ldots, N_i$) secondary unit of the i^{th} primary unit. Also let $\tau = \sum_{j=1}^{N_i} y_{ij}$ be the sum of the y-value in the i^{th} primary unit and $\tau = \sum_{i=1}^{M} \tau_i$ be the population total. A rarity condition is defined as $y_{ij} > c$ and neighbouring units might be observed only if this rarity condition is satisfied for a given unit.

According to Salehi and Seber (1997)s two stage ACS design, at first, simple random sample (SRS) of size m is drawn from a M primary stage units (PSU). Next, an initial sample s_i of size n_i $(i = 1, 2, \ldots, m)$ is drawn from secondary stage units (SSU) of i^{th} selected PSUs by SRS, such that $n = \sum_{i=1}^{n} n_i$ - total initial sample size. Then, the neighbourhoods may be added adaptively to build up a cluster as well as network.

Now in two-stage ACS, two design-based situations arise. In the first-design, the clusters are truncated at selected PSU's boundaries so that each PSU can be treated separately and it is termed as Non-overlapping scheme. The other one, called overlapping scheme, ignores the PSU boundary so that total population units N can be partitioned into distinct networks. We narrate these two schemes below in details, in the subsections 2.1 - 2.2, with Horvitz-Thompson estimation procedure only. However, Salehi and Seber (1997) described the estimation procedures for Hansen and Hurwitz (1943) and Horvitz and Thompson (1952) both.

2.1. Non-overlapping scheme

In the non-overlapping scheme, the modified Horvitz-Thompson estimator for the population mean $(\mu = \frac{\tau}{N})$ is

$$\widehat{\mu}_{HT}^{N} = \frac{1}{N}(M \sum_{i=1}^{m} \frac{\widehat{\tau}_i}{m} \quad)$$

where $\widehat{\tau}_i = \sum_{k=1}^{K_i} y_{ik}^{*}(\frac{I_{ik}}{\alpha_{ik}})$ is the unbiased estimate of i^{th} primary units total having variance $var\,(\widehat{\tau}_i) = \sum_{r=1}^{K_i}\sum_{s=1}^{K_i} y_{ir}^{*} y_{is}^{*}(\frac{\alpha_{irs}-\alpha_{ir}\alpha_{is}}{\alpha_{ir}\alpha_{is}})$.

To the above equations, K_i denotes the number of networks of the i^{th} primary unit and $\alpha_{ik} = 1 - \frac{\binom{N_i - m_{ik}}{n_i}}{\binom{N_i}{n_i}}$ is the probability that the initial sample of unit in i^{th} primary unit intersect the network k. Also, $\alpha_{ikk^T} = \alpha_{ik} + \alpha_{ik^T} - \left(1 - \frac{\binom{N_i - m_{ik} - m_{ikT}}{n_i}}{\binom{N_i}{n_i}}\right)$ is the probability that the initial sample of unit in i^{th} primary unit intersect both the networks k and k^T. The sum of y-value associate with the network k is denoted here by y_{ik}^{*}.

The variance estimator of $\widehat{\mu}_{HT}^{N}$ is

$$V(\widehat{\mu}_{HT}^{N}) = \frac{1}{N^2} M(M-m)\frac{\sigma_M^2}{m} + \frac{1}{N^2}\frac{M}{m}\sum_{i=1}^{M} var(\widehat{\tau}_i)$$

taking $\sigma_M^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\tau_i - \overline{\tau})^2$ and $\overline{\tau} = \frac{1}{M}\sum_{i=1}^{M}\tau_i$.

An unbiased estimate of $V(\widehat{\mu}_{HT}^{N})$ is

$$v(\widehat{\mu}_{HT}^{N}) = \frac{1}{N^2} M(M-m)\frac{s_M^2}{m} + \frac{1}{N^2}\frac{M}{m}\sum_{i=1}^{m} \widehat{var(\widehat{\tau}_i)}$$

where $\widehat{var(\widehat{\tau}_i)} = \sum_{r=1}^{\chi_i}\sum_{s=1}^{\chi_i} y_{ir}^{*} y_{is}^{*}(\frac{\alpha_{irs}-\alpha_{ir}\alpha_{is}}{\alpha_{irs}\alpha_{ir}\alpha_{is}})$.

Here, χ_i denotes the number of distinct networks intersected in the i^{th} primary unit.

2.2. Overlapping scheme

Here, all population units can be partitioned into K number of distinct networks, ignoring the PSU boundaries.

Thus, the modified Horvitz-Thompson estimator is

$$\widehat{\mu}_{HT}^{O} = \frac{1}{N}\left(\sum_{k=1}^{K}\frac{y_k^* J_k}{\alpha_k}\right) \quad .$$

In the above equation, J_k is the indicator function with the value 1 or 0 if the initial sample of size $n = \sum_{i=1}^{m} n_i$ intersects network k or not and y_k^* is the sum of $y-$values for the network k. Salehi and Seber (1997) derived the variance of $\widehat{\mu}_{HT}^{O}$ $(V\left(\widehat{\mu}_{HT}^{O}\right))$ and an unbiased variance estimate $\left(v\left(\widehat{\mu}_{HT}^{O}\right)\right)$ as follows,

$$V\left(\widehat{\mu}_{HT}^{O}\right) = \frac{1}{N^2}\sum_{k=1}^{K}\sum_{k^T=1}^{K}\frac{y_k^* y_{k^T}^*(\alpha_{kk^T} - \alpha_k\alpha_{k^T})}{\alpha_k\alpha_{k^T}}$$

$$v\left(\widehat{\mu}_{HT}^{O}\right) = \frac{1}{N^2}\sum_{k=1}^{\chi}\sum_{k^T=1}^{\chi}\frac{y_k^* y_{k^T}^*(\alpha_{kk^T} - \alpha_k\alpha_{k^T})}{\alpha_{kk^T}\alpha_k\alpha_{k^T}} \quad .$$

Here, χ denotes the number of distinct networks in the sample and α_k is the inclusion probability for the network k and α_{kk^T} is the probability that the initial sample intersects both networks k and k^T. In order to evaluate $V\left(\widehat{\mu}_{HT}^{O}\right)$ and $v\left(\widehat{\mu}_{HT}^{O}\right)$, one needs to know the expressions for α_{kk^T} and α_k which are derived in the Appendix of Salehi and Seber (1997). Here, we have just written the formulas.

$$\alpha_k = P\left[J_k = 1\right]$$

$$= \sum_{i\in B_k}\frac{m}{M}\left(1 - \frac{\binom{N_i - m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right) - \sum_{i}\sum_{i^T<i}\frac{m(m-1)}{M(M-1)}\left(1 - \frac{\binom{N_i - m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right)\left(1 - \frac{\binom{N_{i^T} - m_{i^T k}}{n_i}}{\binom{N_{i^T}}{n_{i^T}}}\right) + \ldots$$

$$+ (-1)^{g_k+1}\frac{m(m-1)\ldots(m-g_k+1)}{M(M-1)\ldots(M-g_k+1)}\prod_{i\in B_k}\left(1 - \frac{\binom{N_i - m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right)$$

and

$$\alpha_{kk^T} = P\left[J_k = 1,\ J_{k^T} = 1\right]$$

where B_k is the set of PSUs intersected by the network k having g_k number of elements and m_{ik} is the number of units of network k located in i^{th} PSU.

3. Prediction approach in two stage sampling scheme

A finite population problem can be formulated as prediction problem and can be solved using Bayesian approach. A more classical superpopulation approach is also possible using Royall (1976)s theorem of best linear unbiased estimator.

Suppose, the objective is to estimate the population total

$$\tau = \sum_{i=1}^{M} \sum_{j=1}^{N_i} y_{ij} = \sum_{i=1}^{M} \tau_i$$

by two-stage design which can be expressed as

$$\tau = \sum_{i \in s} \sum_{j \in s_i} y_{ij} + \sum_{i \in s} \sum_{j \in s_i^c} y_{ij} + \sum_{i \in s^c} \sum_{j=1}^{N_i} y_{ij}. \tag{1}$$

Here, s is the PSU sample of size m and s^c is the set of PSU units not in s. Similarly, s_i is the SSU sample of $i^{th}(i \in s)$ PSU and s_i^c is the complementary of s_i.

In the above expression, it is obvious that the first term is known from the sample. However the second and third terms are unknown and it should be estimated.

The prediction approach of finite population theory considers the total τ is a realization of a random vector T. For a given sample,

$$T = \sum_{i \in s} \sum_{j \in s_i} y_{ij} + Z \tag{2}$$

with $Z = \sum_{i \in s} \sum_{j \in s_i^c} y_{ij} + \sum_{i \in s^c} \sum_{j=1}^{N_i} y_{ij}$.

Now, expressing T as (2), the problem of estimating T is equivalent to the prediction of Z.

Mathematically,

$$\widehat{T} = \sum_{i \in s} \sum_{j \in s_i} y_{ij} + \widehat{Z} \tag{3}$$

clarifies the matter.

The following probability model is adopted here to establish the relationship among N random variable $Y_{ij}\ ; i = 1, 2 \ldots . M \quad j = 1, 2 \ldots N_i$:

$$\begin{cases} E\left(Y_{ij}\right) = \theta & \\ Cov\ \ (Y_{ij},\ Y_{lm}) & = \sigma_i^2, \quad i = l, j = m \\ & = \rho_i \sigma_i^2, i = l, j \neq m \\ & = 0, i \neq l \end{cases} \tag{4}$$

It is assumed here that the random variables within cluster i have common mean θ_i and variance σ_i^{T2} and covariance $\rho_i^T \sigma_i^{T2}$ and the $\{\theta_i^T\}$ are the realizations of uncorrelated random variables with common mean θ and variance φ^2. Then the model (4) applies with $\sigma_i^2 = \varphi^2 + \sigma_i^{T2}$ and $\rho_i = \frac{\varphi^2 + \rho_i^T \sigma_i^{T2}}{\varphi^2 + \sigma_i^{T2}}$.

Royall (1976) suggested an optimal (BLU) estimator in such case and this can be expressed as

$$\widehat{T}^* = \sum_{i \in s} \sum_{j \in s_i} y_{ij} + \sum_{i \in s} (N_i - n_i) \left[\omega_i \overline{y}_{si} + (1 - \omega_i)\, \widehat{\theta}\right] + \sum_{i \notin s} N_i \widehat{\theta} \tag{5}$$

where $\omega_i = \frac{\rho_i n_i}{(1-\rho_i+n_i\rho_i})$ and $\widehat{\theta} = \sum_{i\in s}\theta_i\overline{y}_{si}$ is the weighted average of sample means with weights $\theta_i = \left[\frac{n_i\sigma_i^2}{(1-\rho_i+n_i\rho_i)}\right]/[\sum_{i\in s}\frac{n_i\sigma_i^2}{(1-\rho_i+n_i\rho_i)]}$.

Here, in $\widehat{T}^*$, non-sampled units in sample cluster i can be estimated by $\omega_i\overline{y}_{si} + (1-\omega_i)\,\widehat{\theta}$ and all the units in non-sampled clusters are estimated by $\widehat{\theta}$.

This $\widehat{T}^*$ further can be written as $\widehat{T}^* = \sum_{i\in s}(1+g_i)n_i\overline{y}_{si}$ taking $\sum_{i\in U}N_i = N$ and $\sum_{i\in s}n_i = n$, and $g_i = [\omega_i\frac{(N_i-n_i)}{n_i} + \{N-n-\sum_{i\in s}\omega_i\,(N_i-n_i)\}\frac{\theta_i}{n_i}]$.

The error variance of $\widehat{T}^*$ can be written as

$$
\begin{aligned}
Var\left(\widehat{T}^*-T\right) &= Var\ \left(\sum_{i\in s}g_in_i\overline{y}_{si} - \sum_{i\in s}\sum_{j\in s_i^c}y_{ij} - \sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right)\\
&= Var\left(\sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right) + Var\left(\sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right) + Var\left(\sum_{i\in s}g_in_i\overline{y}_{si}\right) + 2cov\left(\sum_{i\in s}\sum_{j\in s_i^c}y_{ij},\sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right)\\
&\qquad - 2cov\left(\sum_{i\in s}g_in_i\overline{y}_{si},\sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right) - 2cov\left(\sum_{i\in s}g_in_i\overline{y}_{si},\ \sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right)\\
&= Var\left(\sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right) + Var\left(\sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right) + Var\left(\sum_{i\in s}g_in_i\overline{y}_{si}\right) - 2cov\left(\sum_{i\in s}g_in_i\overline{y}_{si},\ \sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right)\\
&= v + Var\left(\sum_{i\in s}g_in_i\overline{y}_{si}\right) - 2cov\left(\sum_{i\in s}g_in_i\overline{y}_{si},\ \sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right)\\
&= v + \left(\sum_{i\in s}\rho_i\sigma_i^2n_i^2g_i^2 + \sum_{i\in s}(1-\rho_i)\,\sigma_i^2n_ig_i^2\right) - 2\sum_{i\in s}\rho_i\sigma_i^2g_in_i(N_i-n_i)\\
&= v - \sum_{i\in s}\rho_i\sigma_i^2(N_i-n_i)^2 + \sum_{i\in s}\rho_i\sigma_i^2[n_ig_i-(N_i-n_i)]^2 + \sum_{i\in s}(1-\rho_i)\sigma_i^2n_ig_i^2
\end{aligned}
$$

where

$$
\begin{aligned}
v &= Var\left(\sum_{i\in s}\sum_{j\in s_i^c}y_{ij}\right) + Var\left(\sum_{i\notin s}\sum_{j=1}^{N_i}y_{ij}\right)\\
&= \sum_{i\in s}(N_i-n_i)\,\sigma_i^2[1-\rho_i+(N_i-n_i)\,\rho_i] + \sum_{i\notin s}N_i\sigma_i^2[1-\rho_i+N_i\rho_i].
\end{aligned}
$$

4. Proposed predictors for two-stage ACS

Simple random sampling without replacement scheme is frequently used in ACS design to draw an initial sample. Chaudhuri (2000) clarified that any sampling method admitting an unbiased estimator for a population total may be extended to adaptive sampling design yielding unbiased estimator. This work insisted us to select PSUs adapting an unequal probability sampling, say PPSWOR instead of SRSWOR in case of Two-stage ACS design, as discussed in Section 2.

4.1. Non-overlapping scheme

Therefore, an unbiased estimator of population total $\tau = \sum_{i=1}^{M} \tau_i = \sum_{i=1}^{M} \sum_{j=1}^{N_i} y_{ij}$ is

$$e = \sum_{i=1}^{m} \frac{\widehat{\tau}_i}{\pi_i} \tag{6}$$

taking $\widehat{\tau}_i$ as the estimate of i^{th} PSU total. This

$$\widehat{\tau}_i = \sum_{k=1}^{\chi_i} \frac{y_{ik}^*}{\alpha_{ik}}, \tag{7}$$

if networks are truncated at selected PSU (Non-overlapping scheme). Here, $y_{ik}^* = \sum_{j \in A(i,k)} y_{ij}$ is the sum of the $y-$values present in $A(i,k)$, the k^{th} network of i^{th} PSU. This network $A(i,k)$ can be partitioned into two parts captured $A_c(i,k)$ and uncaptured $A_{uc}(i,k)$. Obviously, $A(i,k) = A_c(i,k) \cup A_{uc}(i,k)$.

It is obvious that $E(e) = E_1 E_2(e) = E_1 \left(\sum_{i=1}^{m} \frac{\tau_i}{\pi_i} \right) = \sum_{i=1}^{M} \tau_i = \tau$. E_1 denotes here the expectation due to first stage unit selection and E_2 , the expectation due to second stage.

The variance of e can be written as,

$$V(e) = E_1 V_2(e) + V_1 E_2(e)$$

where $E_1(V_2(e)) = E_1(V_2 \left(\sum_{i=1}^{m} \frac{1}{\pi_i} \sum_{k=1}^{\chi_i} \frac{y_{ik}^*}{\alpha_{ik}} \right)) = E_1(\sum_{i=1}^{m} \frac{1}{\pi_i^2} V_2 \left(\sum_{k=1}^{\chi_i} \frac{y_{ik}^*}{\alpha_{ik}} \right))$

$$= E_1 \left(\sum_{i=1}^{m} \frac{1}{\pi_i^2} \left(\sum_{k=1}^{K_i} \sum_{k^T=1}^{K_i} y_{ik}^* y_{ik^T}^* \left(\frac{\alpha_{ikk^T} - \alpha_{ik}\alpha_{ik^T}}{\alpha_{ik}\alpha_{ik^T}} \right) \right) \right)$$

$$= \sum_{i=1}^{M} \frac{1}{\pi_i} \left(\sum_{k=1}^{K_i} \sum_{k^T=1}^{K_i} y_{ik}^* y_{ik^T}^* \left(\frac{\alpha_{ikk^T} - \alpha_{ik}\alpha_{ik^T}}{\alpha_{ik}\alpha_{ik^T}} \right) \right)$$

and $V_1(E_2(e)) = V_1 \left(\sum_{i=1}^{m} \frac{\tau_i}{\pi_i} \right) = \sum_{i<j} \sum_{=1}^{M} (\pi_i \pi_j - \pi_{ij})(\frac{\tau_i}{\pi_i} - \frac{\tau_j}{\pi_j})^2$.

To compute unbiased estimate of $V(e)$, let assume

$$v_1(e) = \sum_{i=1}^{m} \frac{1}{\pi_i^2} \left(\sum_{k=1}^{\chi_i} \sum_{k^T=1}^{\chi_i} y_{ik}^* y_{ik^T}^* \left(\frac{\alpha_{ikk^T} - \alpha_{ik}\alpha_{ik^T}}{\alpha_{ikkT}\alpha_{ik}\alpha_{ik^T}} \right) \right) + \sum_{i<j} \sum_{=1}^{m} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) (\frac{\widehat{\tau}_i}{\pi_i} - \frac{\widehat{\tau}_j}{\pi_j})^2$$

Therefore, $E\left(v_1(e)\right)=E_1E_2\left(v_1(e)\right)$

$$=E_1\left(\sum_{i=1}^{m}\frac{1}{\pi_i^2}\left(\sum_{k=1}^{K_i}\sum_{k^T=1}^{K_i}y_{ik}^*y_{ik^T}^*\left(\frac{\alpha_{ikk^T}-\alpha_{ik}\alpha_{ik^T}}{\alpha_{ik}\alpha_{ik^T}}\right)\right)\right)$$
$$+E_1\left(\sum_{i<j}\sum_{=1}^{m}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\right)\left(\left(\left(\frac{\tau_i}{\pi_i}-\frac{\tau_j}{\pi_j}\right)^2+\frac{V_2\left(\widehat{\tau}_i\right)}{\pi_i^2}+\frac{V_2\left(\widehat{\tau}_j\right)}{\pi_j^2}\right)\right)\right)$$
$$=\sum_{i=1}^{m}\frac{1}{\pi_i}\left(\sum_{k=1}^{K_i}\sum_{k^T=1}^{K_i}y_{ik}^*y_{ik^T}^*\left(\frac{\alpha_{ikk^T}-\alpha_{ik}\alpha_{ik^T}}{\alpha_{ik}\alpha_{ik^T}}\right)\right)+\sum_{i<j}\sum_{=1}^{M}(\pi_i\pi_j-\pi_{ij})\left(\frac{\tau_i}{\pi_i}-\frac{\tau_j}{\pi_j}\right)^2$$
$$+\sum_{i<j}\sum_{=1}^{M}(\pi_i\pi_j-\pi_{ij})\left(\frac{V_2\left(\widehat{\tau}_i\right)}{\pi_i^2}+\frac{V_2\left(\widehat{\tau}_j\right)}{\pi_j^2}\right)$$
$$=V(e)+\sum_{i<j}\sum_{=1}^{M}(\pi_i\pi_j-\pi_{ij})\left(\frac{V_2\left(\widehat{\tau}_i\right)}{\pi_i^2}+\frac{V_2\left(\widehat{\tau}_j\right)}{\pi_j^2}\right)$$

where $V_2\left(\widehat{\tau}_i\right)=\sum_{k=1}^{K_i}\sum_{k^T=1}^{K_i}y_{ik}^*y_{ik^T}^*\left(\frac{\alpha_{ikk^T}-\alpha_{ik}\alpha_{ik^T}}{\alpha_{ik}\alpha_{ik^T}}\right)$.

Thus,

$$v(e)=v_1(e)-\sum_{i<j}\sum_{=1}^{m}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{\widehat{V_2\left(\widehat{\tau}_i\right)}}{\pi_i^2}+\frac{\widehat{V_2\left(\widehat{\tau}_j\right)}}{\pi_j^2}\right) \tag{8}$$

is an unbiased estimator of $V(e)$ where $\widehat{V_2\left(\widehat{\tau}_i\right)}=\sum_{k=1}^{\chi_i}\sum_{k^T=1}^{\chi_i}y_{ik}^*y_{ik^T}^*\left(\frac{\alpha_{ikk^T}-\alpha_{ik}\alpha_{ik^T}}{\alpha_{ikk^T}\alpha_{ik}\alpha_{ik^T}}\right)$.

Now, in case surveyors are unable to gather information from all units belonging to a network, then mathematically it can be express as

$$y_{ik}^*=\sum_{j\in A(i,k)}y_{ij}=\sum_{j\in A_c(i,k)}y_{ij}+\sum_{j\in A_{uc}(i,k)}y_{ij}.$$

Undoubtedly, second term of this expression is unknown and can be predicted easily following Section 3.1 and 3.2 of Pal and Patra (2021). We avoid here the unnecessary repetition.

However, the complication arises if the surveyor decided to ignore PSU boundaries for network construction. Below we describe prediction steps in this case, in details.

4.2. Overlapping scheme

In this case, we need to consider the distinct networks included in two-stage. Thus, an unbiased estimator of the population total τ may be written as

$$e^*=\sum_{k=1}^{\chi}\frac{y_k^*}{\alpha_k^*} \tag{9}$$

where $\alpha_k^*=\sum_{i\in B_k}\pi_i\left(1-\frac{\binom{N_i-m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right)-\sum_i\sum_{i^T<i}\pi_{ii^T}\left(1-\frac{\binom{N_i-m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right)\left(1-\frac{\binom{N_{i^T}-m_{i^T k}}{n_i}}{\binom{N_{i^T}}{n_{i^T}}}\right)+\ldots+(-1)^{g_k+1}\pi_{ii^T.....l}\prod_{i\in B_k}\left(1-\frac{\binom{N_i-m_{ik}}{n_i}}{\binom{N_i}{n_i}}\right)$ and $\alpha_{kk^T}^*=P\left[J_k=1,\ J_{k^T}=1\right]$. Here B_k, with g_k number of elements, is the set of those primary units intersected by k^{th} network.

The variance is $V(e^*) = \left(\sum_{k=1}^{\chi}\sum_{k^T=1}^{\chi} y_k^* y_{k^T}^* \left(\frac{\alpha_{kk^T}^* - \alpha_k^*\alpha_{k^T}^*}{\alpha_k^*\alpha_{k^T}^*}\right)\right)$ and an unbiased estimate of variance is $v(e^*) = \left(\sum_{k=1}^{\chi}\sum_{k^T=1}^{\chi} y_k^* y_{k^T}^* \left(\frac{\alpha_{kk^T}^* - \alpha_k^*\alpha_{k^T}^*}{\alpha_{kk^T}^*\alpha_k^*\alpha_{k^T}^*}\right)\right)$. However, the computation of α_k^* and $\alpha_{kk^T}^*$ are not an easy task. Thus, a modification is needed.

We take Chaudhuri (2000)s approach here to propose an unbiased estimator of τ as

$$e^{T*} = \sum_{i=1}^{m} \frac{\widehat{\tau}_i}{\pi_i} = \sum_{i=1}^{m} \frac{1}{\pi_i}\left(\frac{N_i}{n_i}\sum_{j=1}^{n_i} t_{ij}\right) \tag{10}$$

where $t_{ij} = \frac{1}{d_{ij}}\sum_{i=1}^{M}\sum_{j\in A(i,j)} y_{ij}$ is the average of $y-values$ of the units belong to the network $A(i,j)$, ignoring the PSU boundaries. It is much easier to compute than the previous one (equation 9).

Taking expectation, we get

$$\begin{aligned}
E\left(e^{T*}\right) &= E_2E_1\left(\sum_{i=1}^{m}\frac{1}{\pi_i}\left(\frac{N_i}{n_i}\sum_{j=1}^{n_i}t_{ij}\right)\right) = E_2(\sum_{i=1}^{M}\frac{1}{\pi_i}(\frac{N_i}{n_i}\sum_{j=1}^{n_i}t_{ij})\pi_i) \\
&= E_2\left(\sum_{i=1}^{M}\frac{N_i}{n_i}\sum_{j=1}^{n_i}t_{ij}\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N_i}t_{ij} = \sum_{i=1}^{M}\sum_{j=1}^{N_i}y_{ij} \;\; \text{(see Thompson's (1990) and Chaudhuri (2000))} \\
&= \tau, \text{ the population total.}
\end{aligned}$$

Table 1: Two-stage ACS structure for population

PSU	SSU	Networks of SSU	Cardinality of Networks	Statistic based on SSU
1	$y_{11}, y_{12}, \ldots, y_{1N_1}$	$A(1;1),\ A(1;2)\ldots A(1,N_1)$	$d_{11}, d_{12}, \ldots, d_{1N_1}$	$t_{11}, t_{12}, \ldots, t_{1N_1}$
2	$y_{21}, y_{22}, \ldots, y_{2N_2}$	$A(2;1),\ A(2;2)\ldots A(2,N_2)$	$d_{21}, d_{22}, \ldots, d_{2N_2}$	$t_{21}, t_{22}, \ldots, t_{2N_2}$
…	…	…	…	…
	$y_{i1}, y_{i2}, \ldots, y_{iN_i}$	$A(i;1),\ A(i;2)\ldots, A(i,N_i)$	$d_{i1}, d_{i2}, \ldots, d_{iN_i}$	$t_{i1}, t_{i2}, \ldots, t_{iN_i}$
M	$y_{M1}, y_{M2}, \ldots, y_{MN_M}$	$A(M;1),\ A(M;2)\ldots, A(M,N_M)$	$d_{M1}, d_{M2}, \ldots, d_{MN_M}$	$t_{M1}, t_{M2}, \ldots, t_{MN_M}$

The variance can be written as

$V\left(e^{T*}\right) = E_2V_1\left(e^{T*}\right) + V_2E_1\left(e^{T*}\right) = \sum_{i<j}\sum_{=1}^{M}(\pi_i\pi_j - \pi_{ij})(\frac{\tau_i}{\pi_i} - \frac{\tau_j}{\pi_j})^2 + \sum_{i=1}^{M}\frac{N_i^2}{n_i}(1-f_i)S_i^2$

where $S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(t_{ij} - \bar{t}_i)^2$ and $\bar{t}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}t_{ij}$

An unbiased estimator of $V\left(e^{T*}\right)$ is

$$v\left(e^{T*}\right) = \sum_{i<j}\sum_{=1}^{m}(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}})(\frac{\widehat{\tau}_i}{\pi_i} - \frac{\widehat{\tau}_j}{\pi_j})^2 + \sum_{i=1}^{m}\frac{N_i^2}{n_i}(1-f_i)\,s_i^2 \tag{11}$$

where s_i^2 is an unbiased estimator of S_i^2.

Let $A(i;j)$ can be written as $A_c(i;j) \cup A_{uc}(i;j)$ where $A_c(i;j)$ is the observed units and $A_{uc}(i;j)$ is a set of unobserved units of the network of j^{th} unit of i^{th} PSU. Also, let the cardinality of each set is known and it is possible due to satellite imagery or previous records.

Then, $t_{ij} = \frac{1}{d_{ij}} \sum_{i=1}^{M} \sum_{j \in A(i,j)} y_{ij}$ may be treated as

$$t_{ij} = \frac{1}{d_{ij}} \left[\left(\sum_{i \in s} \sum_{j \in A_c(i;j)} y_{ij} + \sum_{i \notin s} \sum_{j \in A_c(i;j)} y_{ij} \right) + \left(\sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij} \right) + \left(\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij} \right) \right] \quad (12)$$

$$= \frac{1}{d_{ij}} [(\sum_{i \in s} \text{sum of observed units from } i^{th} \text{ PSU } + \sum_{i \notin s} \text{sum of observed units from } i^{th} \text{ PSU})$$

$$+ \sum_{i \in s} \text{sum of unobserved from } i^{th} \text{ PSU} + \sum_{i \notin s} \text{sum of unobserved from } i^{th} \text{ PSU}]$$

for a network $A(i;j)$

Similarly, $d_{ij}-$ cardinality of the network $A(i;j)$ can be partitioned as

$$d_{ij} = \left(\sum_{i \in s} d_{i1}(ij) + \sum_{i \notin s} d_{i2}(ij) \right) + \sum_{i \in s} d_{i3}(ij) + \sum_{i \notin s} d_{i4}(ij) \quad (13)$$

$$= (d_1(ij) + d_2(ij)) + d_3(ij) + d_4(ij) \quad (14)$$

where $d_{i1}(ij)$ is the number of observed units belongs to $i^{th}(i \in s)$ PSU from $A(i;j)$ network and $d_{i2}(ij)$ is the number of observed units belongs to $i^{th}(i \notin s)$ PSU but from the network $A(i;j)$. Similarly, $d_{i3}(ij)$ and $d_{i4}(ij)$ are the numbers of unobserved units belongs to sampled and non-sampled PSU from $A(i;j)$ network, respectively.

Thus to estimate t_{ij}, we need to predict the terms $\sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij}$ and $\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}$.

Now, following Royall (1976)s prediction approach and adopting the model (4) with restrictions $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$, we get

$$\left\{ \begin{array}{rl} E(Y_{ij}) = \delta & \\ Cov\ (Y_{ij},\ Y_{lm}) = \sigma^2, & i = l, j = m \\ = \rho\sigma^2, & i = l, j \neq m \\ = 0, & i \neq l \end{array} \right. \quad (15)$$

With the model (15), we may get

$$est \left(\sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij} \right) = \sum_{i \in s} d_{i3}(ij) \left[w_i \overline{y}_{si} + (1 - w_i) \widehat{\delta} \right] \qquad \ldots using\ (5)$$

where $w_i = \frac{\rho d_{i1}(ij)}{1-\rho-\rho d_{i1}(ij)}$ and $\widehat{\delta} = \sum_{i \in s} \delta_i \overline{y}_{si}$ with $\delta_i = \frac{[\frac{d_{i1}(ij)}{(1-\rho-\rho d_{i1}(ij))}]}{[\sum_{i \in s} \frac{d_{i1}(ij)}{(1-\rho-\rho d_{i1}(ij))}]}$.

Here, $\overline{y}_{si} = \frac{1}{d_{i1}(ij)} \sum_{j \in A_c(i;j)} y_{ij} \ \ \forall i \in s$ is the average of those observed units from a sampled PSU i, belongs to the network $A_c(i;j)$.

It is noteworthy that ρ is generally unknown to us. It can be estimated by analysis of variance (ANOVA) technique (see Valliant *et al.* (2000) chapter 8), if prior information is not given.

With the assumption $\overline{\overline{y}}_* = \frac{\sum_{i \in s} d_{i1}(ij)\overline{y}_{si}}{\sum_{i \in s} d_{i1}(ij)} = \frac{\sum_{i \in s} d_{i1}(ij)\overline{y}_{si}}{d_1(ij)}$, sum of squares of the ANOVA is derived based on the following relation:

$$\begin{aligned}
\sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{\overline{y}}_*)^2 &= \sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si} + \overline{y}_{si} - \overline{\overline{y}}_*)^2 \\
&= \sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si})^2 + \sum_{i \in s} \sum_{j \in A_c(i;j)} (\overline{y}_{si} - \overline{\overline{y}}_*)^2 + 2\sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si})(\overline{y}_{si} - \overline{\overline{y}}_*) \\
&= \sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si})^2 + \sum_{i \in s} \sum_{j \in A_c(i;j)} (\overline{y}_{si} - \overline{\overline{y}}_*)^2 \\
&= \sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si})^2 + \sum_{i \in s} d_{i1}(ij)(\overline{y}_{si} - \overline{\overline{y}}_*)^2
\end{aligned}$$

Table 2: ANOVA table for a sample taken by two-stage adaptive cluster sampling

Source	Sum of squares	Degrees of freedom	Expected mean squares
Between Clusters	$\sum_{i \in s} d_{i1}(ij)(\overline{y}_{si} - \overline{\overline{y}}_*)^2$	$m^* - 1$	$\sigma^2(1-\rho) + \frac{\rho\sigma^2}{m^*-1}\{d_1(ij) - \sum_{i \in s} \frac{d_{i1}^2(ij)}{d_1(ij)}\}$
Within Clusters	$\sum_{i \in s} \sum_{j \in A_c(i;j)} (y_{ij} - \overline{y}_{si})^2$	$d_1(ij) - m^*$	$\sigma^2(1-\rho)$

m^* =Number of sampled PSUs in the cluster

Now, for the term $\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}$,

$$est\left(\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right) = \sum_{i \notin s} d_{i4}(i;j)\widehat{\delta} \qquad \ldots using\ (5)$$

Thus, our suggested optimal (BLU) predictor is

$$\widehat{t_{ij}} = \frac{1}{d_{ij}}\left[\left(\sum_{i \in s} \sum_{j \in A_c(i;j)} y_{ij} + \sum_{i \notin s} \sum_{j \in A_c(i;j)} y_{ij}\right) + \sum_{i \in s} d_{i3}(ij)\left[w_i\overline{y}_{si} + (1-w_i)\,\widehat{\delta}\right] + \sum_{i \notin s} d_{i4}(ij)\widehat{\delta}\right] \tag{16}$$

and the above can be written as

$$\widehat{t_{ij}} = \left(\frac{1}{d_{ij}} \sum_{i \in s} \sum_{j \in A_c(i;j)} y_{ij}\right) + \frac{1}{d_{ij}} \sum_{i \in s} (1 + g_i^*) d_{i1}(ij)\overline{y}_{si} \tag{17}$$

where

$$g_i^* = \left\{ \frac{d_{i3}(ij)}{d_{i1}(ij)} w_i + \frac{\delta_i}{d_{i1}(ij)} \sum_{i \in s} d_{i3}(ij)\,(1-w_i) + \frac{\delta_i}{d_{i1}(ij)} d_4(ij) \right\}.$$

Now, the error variance of $\widehat{t_{ij}}$ can be derived as below.

$$\begin{aligned} MSE\left(\widehat{t_{ij}}\right) &= Var\left(\widehat{t_{ij}} - t_{ij}\right) \qquad (18)\\ &= Var\left(\frac{1}{d_{ij}} \sum_{i \in s} g_i^* d_{i1}(ij) \overline{y}_{si} - \frac{1}{d_{ij}} \sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij} - \frac{1}{d_{ij}} \sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right)\\ &= \frac{1}{d_{ij}^2}\left\{Var\left(\sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right) + Var\left(\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right) + Var\left(\sum_{i \in s} g_i^* d_{i1}(ij) \overline{y}_{si}\right)\right\}\\ &\qquad - \frac{1}{d_{ij}^2}\left\{2\ cov\left(\sum_{i \in s} g_i^* d_{i1}(ij) \overline{y}_{si}\ ,\ \sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right)\right\}\\ &= \frac{1}{d_{ij}^2}\left[v^* + \left(\sum_{i \in s} \rho\sigma^2 d_{i1}^2(ij) g_i^{*2} + \sum_{i \in s} (1-\rho)\sigma^2 d_{i1}(ij) g_i^{*2}\right) - 2\sum_{i \in s} \rho\sigma^2 g_i^* d_{i1}(ij) d_{i3}(ij)\right]\\ &= \frac{1}{d_{ij}^2}\left[v^* - \rho\sigma^2 \sum_{i \in s} d_{i3}^2(ij) + \rho\sigma^2 \sum_{i \in s} (d_{i1}(ij) g_i^* - d_{i3}(ij))^2 + \sum_{i \in s} (1-\rho)\sigma^2 d_{i1}(ij) g_i^{*2}\right] \end{aligned}$$

where $v^* = Var\left(\sum_{i \in s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right) + Var\left(\sum_{i \notin s} \sum_{j \in A_{uc}(i;j)} y_{ij}\right)$

$= \sum_{i \in s} d_{i3}(ij)\sigma^2\,(1 - \rho + d_{i3}(ij)\rho) + \sum_{i \notin s} d_{i4}(ij)\sigma^2\,(1 - \rho + d_{i4}(ij)\rho).$

Thus,

$$\widehat{e}^{T*} = \sum_{i=1}^{m} \frac{\widehat{\tau}_i^*}{\pi_i} = \sum_{i=1}^{m} \frac{1}{\pi_i}\left(\frac{N_i}{n_i} \sum_{j=1}^{n_i} \widehat{t_{ij}}\right) \qquad (19)$$

becomes our final estimator of population total with variance estimator

$$v\left(\widehat{e}^{T*}\right) = \sum_{i<j} \sum_{=1}^{m} (\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}})(\frac{\widehat{\tau}_i^*}{\pi_i} - \frac{\widehat{\tau}_j^*}{\pi_j})^2 + \sum_{i=1}^{m} \frac{N_i^2}{n_i}\,(1 - f_i)\, s_i^{*2} + \sum_{i=1}^{m} \sum_{j=1}^{n_i} MSE(\widehat{t_{ij}}) \qquad (20)$$

where $s_i^{*2} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\widehat{t_{ij}} - \widehat{\overline{t}}_i)^2$ and $\widehat{\overline{t}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{t_{ij}}$.

Note that $\widehat{t_{ij}} = t_{ij}$ if values of all units in the network $A(i,j)$is known.

5. Numerical Example

To illustrate our proposed methodology in prediction approach for Two-stage ACS-overlapping scheme numerically, we consider here Population 1 - the point-objects population of Thompson's (1990) which is further reproduced in Rocco (2008)-page-319 as Figure 1. The population contains $N = 400$ units and it is partitioned into $M = 20$ primary units each of $N_i = 20$ $(\forall i = 1, 2, \ldots, M)$ secondary units. From Population 1, it can be seen that very few units having $y-$ values greater than 0 and the population total is $\tau = \sum_{i=1}^{20} \sum_{j=1}^{20} y_{ij} = 190$. Now, we assume the rarity condition for ACS design is $y > 0$.

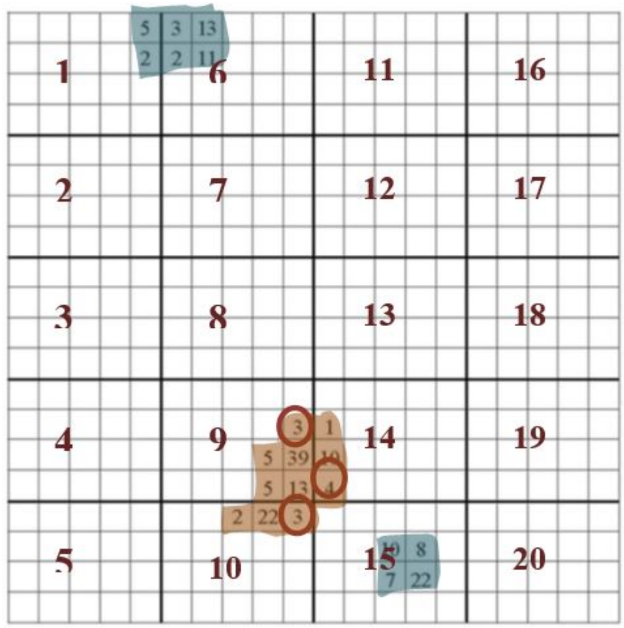

Figure 1: Two-stage ACS data

Table 3 represents the information gathered from a sample with six PSUs each with five SSUs. This sample is selected by Lahiri (1951)-Midzuno (1952)-Sen (1953) sampling scheme for first stage units and SRSWOR for second stage units. With this sampled data, we mainly illustrate the proposed methodology numerically for overlapping scheme, step-by-step. Here, $m = 6$ and $n_i = 5 \ \forall \ i = 1, 2, \ldots, m$. The 1$^{\text{st}}$ column of Table 3 shows the selected PSUs. Inclusion probabilities of the selected PSUs are computed for Lahiri-Midzuno-Sen scheme and mentioned in 2$^{\text{nd}}$ column. Also, $y - values$ of selected SSUs are shown in the table (4$^{\text{th}}$ column) along with the network size (5$^{\text{th}}$ column), if two-stage ACS-overlapping scheme is performed. It can be seen that in this case there are only 2 SSUs having non-zero $y - value$ and based on these SSUs, we can capture more rare units through ACS design. In this way, we found a network of 11 units ignoring the PSU boundaries, of which some are unobserved. Here the unobserved units are marked by red circle.

Table 3: Sampled data

Selected PSU	**Inclusion Probability** ($\pi_\mathbf{i}$)	**Selected SSU(j) for a particular PSU(i)**	**y-values** ($\mathbf{y_{ij}}$)	**Cardinality of Network**($\mathbf{d_{ij}}$)
1	0.323	11,17,14, 07,10	0,**5**,0,0,0	01,**06**,01,01,01
5	0.309	05,09,01,13,18	0,0,0,0,0	01,01,01,01,01
9	0.325	04,11,17,20,03	0,0,0,**13,**0	01,01,01,**11,**01
10	0.270	02,11,10,18,12	0,0,0,0,0	01,01,01,01,01
12	0.268	19,16,08,05,10	0,0,0,0,0	01,01,01,01,01
17	0.276	05,03,18,10,15	0,0,0,0,0	01,01,01,01,01

Now, let us consider the network of **20$^{\text{th}}$ SSU** of **9$^{\text{th}}$ PSU**s ($A(9; 20)$-*orange shaded area*) which can be treated as $A_c(9; 20) \cup A_{uc}(9; 20)$. The set of observed units, $A_c(9; 20)$ contains 15$^{\text{th}}$, 16$^{\text{th}}$, 19$^{\text{th}}$, 20$^{\text{th}}$ units from **9$^{\text{th}}$ PSU** and 9$^{\text{th}}$, 13$^{\text{th}}$ units from **10$^{\text{th}}$ PSU** and also 2^{nd} , 3^{rd} units from **14^{th} PSU**. The set of unobserved units, $A_{uc}(9;20)$ contains 18^{th} unit from **9^{th} PSU**, 17^{th} unit from **10^{th} PSU** and 4^{th} unit from **14^{th} PSU**.

Thus, the cardinality of the network $A(9;20)$ can be partitioned as $11 = (4+2)+(1+1)+(1+1)+1$, according to equations (13) and (14).

Now from equation (12) we get,

$$\sum_{i\in s}\sum_{j\in A_c(i,j)} y_{ij} = (5+5+39+13)+(2+22\ ) = 62+24 = 86 \text{ and}$$

$$\sum_{i\notin s}\sum_{j\in A_c(i,j)} y_{ij} = (1+10) = 11.$$

However, $\sum_{i\in s}\sum_{j\in A_{uc}(i,j)} y_{ij}$ and $\sum_{i\notin s}\sum_{j\in A_{uc}(i,j)} y_{ij}$ are unknown to us and can be predicted through equation (4.2.7) and ANOVA with ρ and σ^2 , two unknown again.

Now, to predict ρ and σ^2, let us first compute sum of squares for between cluster (SSB) and within cluster (SSW).

Here, $SSB = \sum_{i\in s} d_{i1}(ij)(\overline{y}_{si} - \overline{\overline{y}}_{*})^2 = 4(\frac{62}{4} - \frac{86}{6})^2 + 2(\frac{24}{2} - \frac{86}{6})^2 = 16.33$ and $SSW = \sum_{i\in s}\sum_{j\in A_c(i;j)}(y_{ij} - \overline{y}_{si})^2 = 979$ and $\rho = -0.538,\ \sigma^2 = 159.103$.

It is noteworthy that under model (16), ρ can be negative however there is a lower bound. In this case the lower bound is -0.599. To get better idea of this, readers may consider Valliant *et al.* (2000, page 261). The above mentioned two unknown sums can be predicted by $\sum_{i\in s} d_{i3}(ij)\left[w_i\overline{y}_{si} + (1-w_i)\,\widehat{\delta}\right]$, $\sum_{i\notin s} d_{i4}(ij)\widehat{\delta}$ respectively and the predicted values are 28.103, 14.051. The values of w_i and $\widehat{\delta}$ are computed as per given formulas in Section 4. Thus, $\widehat{t_{ij}} = \frac{1}{11}(86+11+28.103+14.051) = 12.65$. Note that, the actual t_{ij} is 9.727 if all units of this network ($A(9;20)$) are observed.

Therefore, based on the sampled data (see Table 3) the final estimate of population total is $229.9957 \approx 230$ (using equation 19) and the estimated variance is 16074.43 (using equation 20). It is worth noting that if all units from the sampled networks are observed, then the estimated population total and estimated variance are 194.0203 and 10722.75 respectively. In other words, if all units from a sampled network are observed, one may get better result. It is obvious condition. However, these two situations are incomparable.

6. Conclusion

Two-stage sampling has several advantages over ordinary single stage (one-stage) sampling. In application of two-stage sampling in ACS, we add many units stopping at the PSU boundary or crossing across the PSU boundary. It is quite obvious that the surveyors may be unable to gather information from one or more rare units. Under such a situation, prediction approach under linear regression model considering correlation structure within network in two-stage ACS is satisfactory. Thus, to achieve a practical solution in two-stage ACS, we have employed Royalls prediction approach. In practice, it brings a novelty in prediction of the population total involving rare units under two stage sampling.

References

Chaudhuri, A. (2000). Network and adaptive sampling with unequal probabilities. *Calcutta Statistical Association Bulletin*, **50**, 237–253.

Hansen, M. M. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, **14**, 333–362.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association*, **47**, 663–685.

Lahiri, D. B. (1951). A method of sample selection providing unbiased ratio estimates. *Bulletin of International Statistical Institute*, **3**, 133–140.

Midzuno, H. (1952). On the sampling system with probability proportional to the sum of the sizes. *Annals of Institute of Statistical Mathematics*, **3**, 99–107.

Pal, S. and Patra, D. (2021). A prediction approach in adaptive sampling. *Metron*, **79**, 93–108.

Pal, S. and Patra, D. (2023). Modifications on re-scaling bootstrap for adaptive sampling. *Communications in Statistics - Simulation and Computation*, **52**, 1607–1620.

Patra, D. and Pal, S. (2023). Application of kalman filtering with bayesian formulation in adaptive sampling. *Communications in Statistics - Simulation and Computation*, **1**. https://doi.org/10.1080/03610918.2023.2265084.

Rocco, E. (2008). Two-stage restricted adaptive cluster sampling. *Metron*, **66**, 313–327.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377–389.

Royall, R. M. (1976). The linear least-squares prediction approach to two-stage sampling. *Journal of American Statistical Association*, **71**, 657–664.

Salehi, M. M. and Seber, G. A. (1997). Adaptive cluster sampling with networks selected without replacement. *Biometrika*, **84**, 209–219.

Seber, G. A. and Salehi, M. M. (2013). *Adaptive Sampling Designs: Inference for Sparse and Clustered Populations*. Springer.

Sen, A. R. (1953). On the estimator of the variance in sampling with varying probabilities. *Journal of Indian Society of Agricultural Statistics*, **5**, 119–127.

Thompson, S. K. (1991a). Adaptive cluster sampling: designs with primary and secondary units. *Biometrics*, **47**, 1103–1115.

Thompson, S. K. (1991b). Stratified adaptive cluster sampling. *Biometrika*, **78**, 389–397.

Thompson, S. K. (1992). *Sampling*. John Wiley & Sons.

Thompson's, S. K. (1990). Adaptive cluster sampling. *Journal of American Statistical Association*, **85**, 1050–1059.

Valliant, R., Dorfman, A. H., and Royall, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley New York.