

Day-1 Practical Session, 25 May 2021

Part 2: Bipartite Incidence Graph (BIG) for Adaptive Cluster Sampling (ACS)

Li-Chun Zhang^{1,2,3} and *Melike Oguz-Alper*²

¹*University of Southampton (L.Zhang@soton.ac.uk)*, ²*Statistics Norway*, ³*University of Oslo*

Illustration II: Two-Stage Adaptive Cluster Sampling (Thompson 1991)

In this illustration, a modified BIG strategy will be applied to the two-stage ACS considered by Thompson (1991). Variances of the HTEs of the population mean under SRS and two-stage ACS will be compared for chosen sample sizes.

Description of the population and sampling strategies

- A spatial field divided into 20×20 grids.
- The parameter of interest: the mean number of rare species per grid
- A two-stage sampling applied: vertical strips the *primary sampling units* and grids the *secondary sampling units*
- Strips selected with SRSWOR
- All grids in sample strips surveyed
- Adaptive tracing implemented at the second stage: once species observed all neighbouring grids surveyed. This continues until reaching empty neighbour grids in all four directions
- Modified BIG, \mathcal{B}^* applied due to existence of *edge* grids
- Use functions ****skthPSUACS**** and ****skthPSUACS_BIG**** to visualise two-stage ACS and its BIG representation with modified BIG

Formula sheet

- F : strips, Ω : grids
- Population mean

$\theta = \frac{\sum_{i \in F} y_i}{\bar{M}N} = 326/400 = 0.815$, where \bar{M} number of grids in each vertical strip and N number of strips

- Unbiased estimator of the population mean based on initial sample, s_0

$\hat{\theta}_{s_0} = \frac{\sum_{i \in s_0} y_i}{\bar{M}n}$, where n number of sample strips

- HTE of the population mean under ACS:

$\hat{\theta}_{HT} = \frac{1}{\bar{M}N} \sum_{\kappa \in \Omega_s} \frac{y_{\kappa}}{\pi_{(\kappa)}}$

- Variance of the HTE under SRS based on s_0 :

$V(\hat{\theta}_{s_0}) = \frac{1}{\bar{M}^2} \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\sum_{i \in s_0} (y_i - \bar{y}_i)^2}{n-1}$

- Variance of the HTE under ACS:

$V(\hat{\theta}_{HT}) = \frac{1}{\bar{M}^2 N^2} \sum_{\kappa \in \Omega} \sum_{\ell \in \Omega} \left(\frac{\pi_{(\kappa\ell)}}{\pi_{(\kappa)}\pi_{(\ell)}} - 1 \right) y_{\kappa} y_{\ell}$

- Calculation of the first-order inclusion probabilities $\pi_{(\kappa)} = \Pr(\kappa \in \Omega_s)$ under SRS of s_0 :

$\pi_{(\kappa)} = 1 - \pi_{\beta_{\kappa}} = 1 - \binom{N-|\beta_{\kappa}|}{n} / \binom{N}{n}$, where $|\beta_{\kappa}|$ is the size of the ancestor network of κ . We have $|\beta_{\kappa}| = 4$ for all $\kappa \in \Omega$.

- Calculation of the second-order inclusion probabilities $\pi_{(\kappa\ell)} = \Pr(\kappa, \ell \in \Omega_s)$ under SRS of s_0 :

$\pi_{(\kappa\ell)} = \pi_{(\kappa)} + \pi_{(\ell)} - (1 - \pi_{\beta_{\kappa} \cup \beta_{\ell}}) = 1 - \binom{N-|\beta_{\kappa} \cup \beta_{\ell}|}{n} / \binom{N}{n}$. We have $|\beta_{\kappa} \cup \beta_{\ell}| = 6$ for the first two specie-networks and $|\beta_{\kappa} \cup \beta_{\ell}| = 8$ for $\kappa \in \{1, 2\}$ and $\ell = 3$.

N.B. R-package **igraph** has to be installed

Description of R-function ****mainPSUACS****

1. Function parameters

- **n**: Sample size. Choose among 1, 2, . . . 20. Default value 1.

2. Main steps of the function

- Population data constructed
- Variances under SRS and two-stage ACS calculated

3. Main outputs of the function

- Variances of the HTE of the population mean under SRS and two-stage ACS