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]

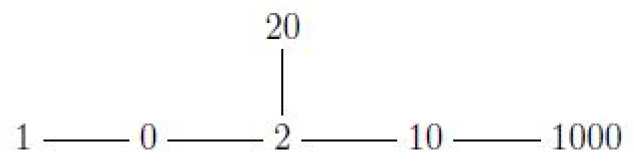
[1]*University of Southampton (L.Zhang@soton.ac.uk)*, [2]*Statistics Norway*, [3]*University of Oslo*

Illustration I: Adaptive Cluster Sampling, a small data set extended from Thompson (1990)

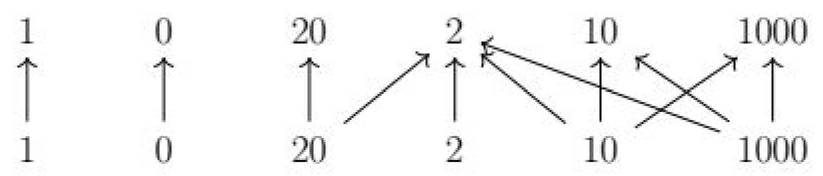
In this illustration, the efficiency of the unbiased estimators of the population mean will be examined under three strategies described below using BIG sampling (BIGS).

Description of the population and sampling strategies

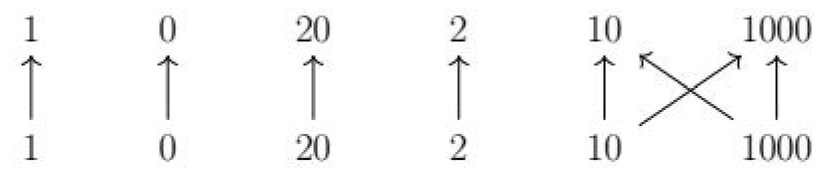
- Data set



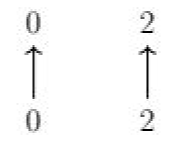
- Threshold for adaptive tracing: $y_i \geq 5$
- s_0 of size: $|s_0| = 2$ selected with SRS
- Strategy I: $\mathcal{B}$, *edge* node 2 is not used in the estimation if it is observed via its ancestor network; *modified* HTE:



- Strategy II: $\mathcal{B}^*$, HTE

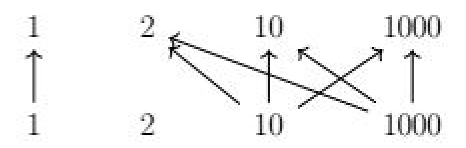


- Strategy III: *Sample-dependent BIGS strategy*, HTE, apply either strategy $\mathcal{B}^*$ or $\mathcal{B}^\dagger$, depending on which nodes selected first, over repeated sampling given the first selection
    - If $s_0 \cap \{20, 2, 10, 1000\} = \{2\}$, e.g. $s_0 = \{0, 2\}$, apply strategy $\mathcal{B}^*$:

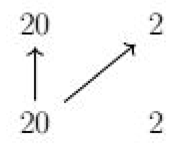


    - If $s_0 \cap \{20, 10, 1000\} \subseteq \{10, 1000\}$ or $s_0 \cap \{20, 10, 1000\} = \{20\}$), apply strategy $\mathcal{B}^\dagger$:

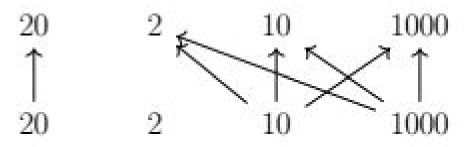
For e.g. $s_0 = \{1, 10\}$:



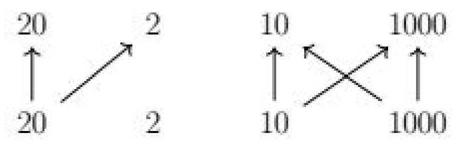
For e.g. $s_0 = \{2, 20\}$:



    - If $s_0 \cap \{20, 10, 1000\} \subseteq \{20, 10\}$ or $s_0 \cap \{20, 10, 1000\} = \{20, 1000\}$, choose either $\{10, 1000\}$ or $\{20\}$ as the ancestor network of edge node 2 and apply strategy $\mathcal{B}^\dagger$:



or



Formula sheet

- Population mean

$\theta = \frac{\sum_{i \in F} y_i}{N}, \quad y_i = \{1, 0, 2, 20, 10, 1000\}, \quad N = 6$

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

$\hat{\theta}_{s0} = \frac{\sum_{i \in s_0} y_i}{n}$

- HTE of the population mean under ACS:

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

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

$\pi_{(\kappa)} = 1 - \bar{\pi}_{\beta_\kappa} = 1 - \binom{N - |\beta_\kappa|}{n} / \binom{N}{n}$, where $|\beta_\kappa|$ is the size of the ancestor network of κ

- Modified HTE, $\hat{\theta}^*_{HT}$, and the HTE provides the same estimates. The former differ with respect to the Rao-Blackwell (RB) method:

$\hat{\theta}_{RB} = E(\hat{\theta}^*_{HT} \mid \Omega_s)$

Description of R-function **mainACS**

1. Function parameters

- **first.sample**: The first sample selected in repeated sampling. A vector of size two with y_i values. Choose values such taht $s_0 \cap \{2, 10, 1000, 20\} \neq \emptyset$. Default value $c(0, 2)$
- **choose20**: Use 20 as the ancestor network of edge-node 2 in case **first.sample** consists of $\{20, 10\}$ or $\{20, 1000\}$. No effect for other choices of **first.sample**

2. Main steps of the Function

- Edges, *observational links*, between sampling units in U and the study units in Ω are constructed under each strategy
- The size of the ancestors of $\kappa \in \Omega$ are obtained based on the observational links
- All possible samples of sizes $n = 2$ selected from the population with SRSWOR and Ω_s obtained via observational links. The first sample in the repeated sampling is provided by **first.sample**
- For each random sample, the population mean is estimated with the unbiased estimators mentioned above
- RB method is applied to the modified HTE under strategy I

3. Main outputs of the function

- Expected values and sampling variances of the estimators under SRS and under ACS with different BIGS strategies

---