Day-2 Practical Session, 26 May 2021

Part 1: Incidence Weighting Estimator (IWE) under Bipartite Incidence Graph Sampling (BIGS)

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: BIGS-IWE strategy: two random graphs with different degree-distributions

In this illustration, we will again compare the efficiencies of several IWE estimators including the priority-rule estimators under BIG sampling. This time, two random population graphs with the same total number of edges, but different out-degree distributions, will be generated. These graphs are generated by using the R-function **skthrndBIG**.

Description of the population and sampling strategies

- Population BIG: $\mathcal{B} = (F, \Omega; H)$, H consists of edges between *sampling units* $i \in F$ and *study units* $\kappa \in \Omega$
- Sample BIG: $\mathcal{B}_s = (s_0, \Omega_s; H_s)$ with $s_0 \in F$, $\Omega_s = \alpha(s_0)$, and $s_{ref} = s_0 \times \Omega$ such that $H_s = H \cap s_{ref} = H \cap (s_0 \times \Omega)$
- β_κ: *ancestry* set of $\kappa \in \Omega_s$ and α_i: *successors* of $i \in U$
- s_0 of size n selected with SRSWOR from sampling frame F of size N

Formula sheet

- The parameter of interest: size of Ω:

  $\theta = \sum_{\kappa \in \Omega} y_\kappa$, where $y_\kappa = 1$ for all $\kappa \in \Omega$

- IWE based on $\mathcal{B}_s = (s_0, \Omega_s; H_s)$ by BIGS

  $\hat{\theta} = \sum_{(i\kappa) \in H_s} W_{i\kappa} \frac{y_\kappa}{\pi_i}$

- Hansen-Hurwitz (HH) type estimators: special case of IWE, *constant* weights

  $\hat{\theta} = \sum_{i \in s_0} \frac{z_i}{\pi_i}$, where $z_i = \sum_{\kappa \in \alpha_i} w_{i\kappa} y_\kappa$, with $\sum_{i \in \beta_\kappa} w_{i\kappa} = 1$

  - HH-type estimator with *equal* weights: *multiplicity* estimator (Birnbaum and Sirken 1965)

    $w_{i\kappa} \equiv \frac{1}{|\beta_\kappa|}$

  - HH-type estimator with *unequal* weights: *probability and inverse degree-adjusted (PIDA) weights*

    $w_{i\kappa} \propto \frac{\pi_i}{|\alpha_i|^\gamma}$, $\gamma > 0$

NB. Under SRS of s_0 and when $\gamma = 0$, the HH-type estimator with PIDA weights become equivalent to the multiplicity estimator above

- HTE: a special case of IWE

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

The first-order inclusion probabilies $\pi_{(\kappa)} = \Pr(\kappa \in \Omega_s)$ can be calculated, under SRS of s_0, by

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

- Priority-rule estimators with *priority rule* to the sample edges H_s: $I_{i\kappa} = 1$ if $i = \min(s_0 \cap \beta_\kappa)$, and $I_{i\kappa} = 0$ otherwise

  $\hat{\theta}_p = \sum_{(i\kappa) \in H_s} \left(\frac{I_{i\kappa} \omega_{i\kappa}}{p_{(i\kappa)}} \right) \frac{y_\kappa}{\pi_i}$, where $p_{(i\kappa)} = \Pr(I_{i\kappa} = 1 | \kappa \in \Omega_s)$

The probabilities $p_{(i\kappa)}$ can be calculated, under SRS of s_0, by

$p_{(i\kappa)} = \binom{N-1-d_{i(\kappa)}}{n-1} / \binom{N-1}{n-1}$, where $d_{i(\kappa)}$ the number of nodes with higher probability than i for each $\kappa \in \Omega$ and $i \in \beta_\kappa$ for the priority-rule $\min(s_0 \cap \beta_\kappa)$

NB. R-package **igraph** has to be installed before running R-functions below that generates random graphs.

Description of R-function **skthrndBIG**

1. Function parameters

- **sizeF**: number of sampling units in F; default value 50
- **sizeOmega**: number of study units in Ω; default value 100
- **meanoutdeg**: mean number of out-degrees, $\sum_{i \in F} \alpha_i / \mid F \mid$; default value 10
- **showplot**: Use **TRUE** to get histograms of the *uniform* and *skewed* out-degree distributions; default **FALSE**

2. Main steps of the function

- A random graph generated with exponential degree distribution
- Another random graph with uniform degree distribution generated with the same total number of degrees as in the graph with exponential degree distribution
- Because the number of in- and out-degrees have to be equal in a graph, initial in-degrees are adjusted, so that the total number of in-degrees would become equivalent to the total number of out-degrees. Initial degrees in the graph with uniform degree distribution are also adjusted, so that the total number of degrees in the graph with exponential distribution preserved. Adjustment of degrees is done by compiling R-function **degcorrection** (This function only called in the R-function **skthrndBIG**. Thus user input not needed).

3. Main outputs of the function

- Histograms for uniform and exponential degree distributions shown if **showplot**= **TRUE**
- A list of two random graphs generated: Use $\mathbf{Guniform}$ and $\mathbf{Gskewed}$ to get the graphs with uniform and exponential distributions, respectively

Description of R-function **zFun**

1. Function parameters

- **popgraph**: population graph to be used: outputs of **skthrndBIG**
- **coefgamma**: coefficient to be used in the HH-type estimator with PIDA weights; default value 0. No effect of the choice if **multiplicity**= **TRUE**
- **n**: sample size of initial sample s_0; default value 2
- **multiplicity**: Use **TRUE** to get z_i values based on equal weights, i.e. $w_{i\kappa} = \mid \beta_\kappa \mid^{-1}$; default **FALSE**

2. Main steps of the function

- Edge set derived from the population graph, as well as the labels of the vertices in F and Ω
- $|\alpha_i|$ and $|\beta_\kappa|$ calculated based on the edge set
- z_i values calculated for all $i \in F$ for chosen values of γ

3. Main outputs of the function

- z_i values returned

Description of R-function **mainsimBIGSIWE**

1. Function parameters

- **popgraph**: population graph to be used: use the outputs of the function **skthrndBIG**
- **coefgamma**: coefficient to be used in the HH-type estimator with PIDA weights; default value 0
- **n**: sample size of initial sample s_0; default value 2
- **B**: number of Monte-Carlo replications; default value 50

2. Main steps of the function

- Edge set derived from the population graph, as well as the labels of the vertices in F and Ω
- $|\alpha_i|$ and $|\beta_\kappa|$ calculated based on the edge set
- Inclusion probabilities $\pi_{(\kappa)}$ calculated based on $|\beta_\kappa|$
- B random samples of size n selected with SRS from F
- For each random sample, estimates obtained from the HTE, the HH-type estimator and the priority-rule estimator. For the last one, three random orderings of out-degrees, i.e. α_i, considered: *random*, *ascending* and *descending*

3. Main outputs of the function

- Empirical relative efficiencies the HH-type estimator and the priority-rule estimators against the HTE
