Day-1 Practical Session, 25 May 2021

Part 1: Epidemic Prevalence Estimation

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: Estimation of change in prevalence over time under panel without tracing, panel ACS and iterative ACS designs

In this illustration, we will estimate the relative efficiencies of the change estimators under panel ACS and iterative ACS designs in comparison to those under panel design without adaptive tracing. The relative efficiency will be evaluated on the basis of the structure of the case-networks and the type of change in the structure over time.

We will run R-functions **gen.pop** and **mainEpiPanel** to generate the population and to calculate the relative efficiencies for different scenarios, respectively.

Description of R-function **gen.pop**

1. Function parameters

- **N**: population size
- **theta**: a 2×1 vector of prevalences at time points t and $t+1$
- **M1**: number of case-networks at time t
- **M12**: number of new case-networks as such that **M12**<**M1**
- **Y12**: number of new cases. If $Y12 > M1$, use a value divisible by **M1** to ensure $Y1 = Y2$ when the prevalences same at both time points
- **emerge**: **TRUE** if *emerging* and **FALSE** if *evolving*

2. Main steps of the function

- Equal case-network size at time point t, **K**, is calculated based on **N**, **theta[1]** and **M1**, and the total number of cases at t is calculated by **Y1**=**M1*K**
- Equal case-network size at time point $t+1$ for new case-networks, **Kny**, is calculated based on **Y12** and **M12**, and the eventual number of cases at $t+1$ is calculated by **Y12**=**M12*Kny**
- $N \times 2$ array of index for case-networks, **kidx**, is created for all cases and set to 0 for all non-cases
  - The first column of **kidx** refers to index for t: the first **Y1** units in the population take a network id number while the last $N - Y1$ units take value 0, that is,
    $$\underbrace{1,1,\ldots,1}_{K \text{ times}},\underbrace{2,2,\ldots,2}_{K \text{ times}},\ldots,\underbrace{M1,M1,\ldots,M1}_{K \text{ times}},\underbrace{0,0,\ldots,0}_{(N-Y1) \text{ times}}$$
  - The second column of **kidx** refers to index for $t+1$:
    - If **emerging**=**TRUE**: the first **Y1** units take the same ids as at t, the last **N-(Y1+Y12)** units in the population take value 0, and the rest takes network id number existing only at $t+1$, that is,
      $$\underbrace{1,1,\ldots,1}_{K \text{ times}},\underbrace{2,2,\ldots,2}_{K \text{ times}},\ldots,\underbrace{M1,M1,\ldots,M1}_{K \text{ times}},\underbrace{M1+1,M1+1,\ldots,M1+1}_{Kny \text{ times}},$$
      $$\underbrace{M1+2,M1+2,\ldots,M1+2}_{Kny \text{ times}},\ldots,\underbrace{M1+M12,M1+M12,\ldots,M1+M12}_{Kny \text{ times}},$$
      $$\underbrace{0,0,\ldots,0}_{(N-Y1-Y12) \text{ times}}$$
    - If **emerging**=**FALSE**: the first **Y1** units take the same ids as at t, the last **N-(Y1+Y12)** units in the population take value 0, and the rest takes network id number existing both at t and $t+1$. For time point $t+1$, a random sample of size **M12** is selected without-replacement among the **M1** ids at time point t, that is,
      $$\underbrace{1,1,\ldots,1}_{K \text{ times}},\underbrace{2,2,\ldots,2}_{K \text{ times}},\ldots,\underbrace{M1,M1,\ldots,M1}_{K \text{ times}},\underbrace{i_1,i_1,\ldots,i_1}_{Kny \text{ times}},\underbrace{i_2,i_2,\ldots,i_2}_{Kny \text{ times}},\ldots,$$
      $$\underbrace{i_{M12},i_{M12},\ldots,i_{M12}}_{Kny \text{ times}},\quad\underbrace{0,0,\ldots,0}_{(N-Y1-Y12) \text{ times}}$$
      , where the $i_{m'}$ with $m \in 1\{1,2,\ldots,M12\}$, denote the sample ids that could take values in $\{1,2,\ldots,M1\}$
- Finished number of cases from t to $t+1$, **Y21**, is calculated by $Y21 = N * theta[1] - (N * theta[2] - Y12)$
- Finished case-network size, **K21**, is obtained based on **Y21** and **M1**
- Identification of the ids for finished cases by either selecting the last **K21** units in each network at t if $K21 > 0$ or selecting a random sample of size **Y21** without replacement from **Y1** units in the population, otherwise
- The **Y21** elements corresponding to the selected finished cases among the first **Y1** elements in the second column of the index array **kidx** is replaced with 0
  - For example, in case $K21 = 2$, the first **Y1** indexes of **kidx[,2]** will be
    $$\underbrace{1,1,\ldots,1,0,0}_{K \text{ times}},\underbrace{2,2,\ldots,2,0,0}_{K \text{ times}},\ldots,\underbrace{M1,M1,\ldots,M1,0,0}_{K \text{ times}}$$

3. Main outputs of the function

- A list of index of case- and non-case networks for time points t and $t+1$

Description of R-function **mainEpiPanel**

1. Function parameters

- **N**: population size
- **theta**: a 2×1 vector of prevalences at time points t and $t+1$
- **M1**: number of case-networks at time t
- **M12**: number of new case-networks as such that **M12**<**M1**
- **Y12**: number of new cases. If $Y12 > M1$, use a value divisible by **M1** to ensure $Y1 = Y2$ when the prevalences same at both time points
- **emerge**: **TRUE** if *emerging* and **FALSE** if *evolving*
- **f**: sampling fraction
- **lift**: the odds-ratio between cases and non-cases, denoted by η (see Illustration I: Size-biased sampling and adaptive network tracing)
- **B**: number of replications
- **go**: use **TRUE** to get analytic standard error of the change estimator under panel ACS and the relative efficiency against the change estimator under panel design and use **FALSE** otherwise

2. Main steps of the function

- R-function gen.pop is complied to get indexes for case- and non-case networks at time points t and $t+1$
- A 2×1 vector of total cases is obtained based on **kidx**, denoted by **Y** inside the function
- An array of dimension $N \times 2$ of y dummy variables being equal to 1 for cases and 0 for non-cases
- A 2×1 vector of maximum case-network sizes at each time point, $K = (K1, K2)^\top$, is created, denoted by **K** inside the function
- Sample size n_0, **n0** inside the function, is calculated based on **N** and **f**
- Inclusion probabilities at t, $\pi_i = \Pr(i \in s_0)$, denoted by **p** inside the function, are calculated as proportional to **lift** if $y_i = 1$, and 1 otherwise
- Inclusion probabilites of case-networks, denoted by **pr.k** inside the function, are calculated by
  - $\pi_{(\kappa)} = 1 - (1 - p.1)^{m_\kappa}$, where p.1=max(p), under Poisson sampling of s_0, where m_κ is the size of the case-network κ
- Two arrays, **n1idx** and **n2idx**, with dimensions $N \times 2$ are created. The first columns are replaced with number of cases in the corresponding networks. Non-cases have value 0. The second columns are replaced with the inclusion probabilities of the corresponding networks. Non-cases have value n_0/N.
- Sampling variances of the HTE of change in prevalence from t to $t+1$ under panel design are calculated by
  - Under SRS of s_0: **v.srs**
    $$V_{srs}(\hat{\Delta}^{panel}_{t,t+1}) = \big(1 - \frac{n_0}{N}\big)\frac{1}{N-1}\frac{\sum_{i \in N}(z_i - \Delta^{panel}_{t,t+1})^2}{n_0}, \quad z_i = y_2 - y_1, \quad \Delta^{panel}_{t,t+1} = \theta_2 - \theta_1$$
  - Under poisson sampling of s_0: **v.pois**
    $$V_{pois}(\hat{\Delta}^{panel}_{t,t+1}) = \frac{1}{N^2}\sum_{i \in N}\big(\frac{1}{\pi_i} - 1\big)(y_2 - y_1)^2$$
- Sampling variances of the HTE of prevalences at t and $t+1$ under panel ACS design are calculated by implementing the formula (4.2) in the Lecture Notes. For simplicity in coding, the covariance term is ignored for variance at $t+1$ as the variance term dominates the variance.
- If **go**=**TRUE**: The sampling variance of the HTE of the change in prevalence from t to $t+1$ is calculated by
  $$\mathrm{V}(\hat{\Delta}^{pACS}_{t,t+1}) = \mathrm{V}(\hat{\theta}_t) + \mathrm{V}(\hat{\theta}_{t+1}) - 2\mathrm{Cov}(\hat{\Delta}^{pACS}_{t,t+1}),$$
  where the covariance term is obtained by implementing the covariance formula in Section 4.4.2 in the Lecture notes
- Simulation study
  - **B** random samples selected with *Sequential Poisson Sampling* (SPS) from the population for time point t
  - Sample units in $s(t)$ and $s(t+1)$ are obtained under panel, panel ACS and iterated ACS designs
  - For each pair of random samples at time points t and $t+1$, the change of prevalence is estimated under (see formulas for the estimators in Section 4.4.2 in Lecture Notes)
    - panel design without tracing, based on $s(t) = s_0$ over time
    - panel ACS: $s(t)$ based on s_0 and A_t; $s(t+1)$ based on s_0 and A_{t+1}
    - iterated ACS design: $s(t)$ based on s_0 and A_t; $s(t+1)$ based on $s(t)$ and A_{t+1}

2. Main outputs of the function

- Monte-Carlo expectations of the change estimators under three designs
- Monte-Carlo standard errors of the change estimators under three designs
- Relative efficiencies of panel ACS and iterated ACS designs against panel design without adaptive tracing

---

Possible choices of function parameters (M1, M12, Y12, emerge) for populations with different dynamics over time

- L1, Large, Quickly Evolving: (10, 2, 200, F)
- L2, Large, Quickly Emerging: (10, 2, 200, T)
- L3, Large, Slowly Emerging: (10, 5, 100, T)
- M1, Medium, Quickly Evolving: (100, 10, 400, F)
- M2, Medium, Quickly Emerging: (100, 10, 400, T)
- M3, Medium, Slowly Emerging: (100, 10, 100, T)
- S1, Small, Quickly Evolving: (500, 10, 400, F)
- S2, Small, Quickly Emerging: (500, 10, 400, T)
- S3, Small, Slowly Emerging: (500, 50, 100, T)
