A Brief History of Outcome Dependent Sampling and Two Phase Design: Notation and Abbreviation

Chiara Di Gravio

November 24, 2020

Below is some of the common notation used in the report: boldface characters denotes matrices or vector, while non boldface characters represents scalars. Unless otherwise stated, for each variable of interest, the single subscript i (or j) denotes the i^{th} (or j^{th}) subject, while the double subscript ij denotes the j^{th} observations for the i^{th} subject. If no subscript are present, then the variable refers to the whole cohort.

Variables and Parameters of Interest

- $\boldsymbol{Y}$: outcome(s)
- $\boldsymbol{X}$: expensive covariate(s)
- $\boldsymbol{Z}$: inexpensive covariate(s)
- $\boldsymbol{\theta}$: coefficients linking $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{Z}$
- $f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})$: model of interest
- $dG(\boldsymbol{X}|\boldsymbol{Z})$: conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Z}$

Two Phase Design

- N: phase one sample size
- n_V: phase two sample size ($n_V < N$)
- R_i: indicator of whether $\boldsymbol{X}$ is measured for subject i
- V: index set of all subjects measured in phase one and two
- $\bar{V}$: index set of all subjects measured in phase one only
- $\pi(\boldsymbol{Y_i}, \boldsymbol{Z_i})$: probability that subject i is sampled for phase two
- $\mathcal{S}_k$: stratum k defined by the observed outcome or by a combination of the outcome and inexpensive covariates

Abbreviations

- ODS: outcome dependent sampling
- SELE: semiparametric empirical likelihood estimator
- SPMLE: semiparametric maximum likelihood estimator

- MELE: maximum estimated likelihood estimator
- SMLE: semiparametric maximum likelihood estimator
- BLUP: best linear unbiased predictors
- ACML: ascertainment corrected maximum likelihood