Environmental Study by GCTA method

Xuelong Wang

March 25, 2019

Environmental Data and Goal

Data

- Covariates are contamination levels of different toxicants e.g. PCBs

  - Continuous
  - The number of predictors are around 30 to 50
  - There is a high correlation among those covariates
  - Observed levels are very low

- Response are health outcomes, e.g. boold pressure, disease status,etc.

Goal

The overarching goal is to understand the the relationship of among checimal exposures and health outcomes.

Issuse

Epi issue

- lack of traditional epidemiology methodology: e.g. The biological pathways of chemicals and mixtures in human's body are not clear
- Food consumption

Statistical issue

- Many weak sginals, hard to identify and esitmate a signal chemicals coefficfient
- Lasso-type of approach is not working because the sparity assumption is not held in this case

What is the GCTA method

- GCTA: Genome-wide complex trait analysis
- Instead of estimating the signals of each covariate, GCTA estimates the variance of total signals related to all the covariates.
- Using a working linear mixed effects model (Miss-specificed model)
- Using Restriced MLE to estimate the total variance

Advantage of using GCTA on enviromental data

- deal with weak signal problem

Limitations

Assumption

Covariates have to be independent to each other

Real world

Each covariates are more likely to be correlated to each other

Decorrelation

- SVD method is used to get uncorrelated data
- Linear transformation will not change the total variance
- Simulation study shows that using the uncorrelated data, though not truely independent, imporve the accuracy of variance estimation

Higher order approximation

Assume that the true model is more complicted than just a linear model, one strenghtforward idea to use higher order taylor expansion to approximate the true model.

Adding the second order

combined estimation

- Since interaction term is dependent on the main effect, it not easy to directly estimate the variance of interaction effect
- However, by the property of variance, we can combine the main and interaction as the total effect and estimate the total effect intead.