

Dissertation meeting

April 13, 2023

Talking points

- Literature review update
 - How should I get paper if we do not have institution access?
- Approaches to modify QIC proposed by Joe
- Update Joe on dissertation hours, see 2023_04_06_notes.md

Literature review

The current collection of papers are a good start, but we need to continue going. Joe indicated that he wanted to read the the C_p (sato_2020.pdf) paper.

Approaches to modify QIC

We were NOT able to discuss Joe's idea for model selection for GEEs; however, we did discuss an idea I had based on bootstrapping that came from BIOS:7240 lecture on Bootstrapping for inference in high-dimensional settings. See the PNG screenshot in the Miscellaneous folder.

Bootstrapping model selection approach

There are three forms of bootstrapping, which are listed below with their definitions obtained from: <https://online.stat.psu.edu/stat555/node/119/>.

- Parametric (Simulation): Assuming the data comes from a known distribution with unknown parameters, we are able to estimate the parameters from the data then use the estimated distributions to simulate the bootstrap samples.
- Non-parametric (Resampling): Samples of the same size as the data are taken from the data with replacement to provide a bootstrap sampling distribution.
- Semi-parametric (Adding noise): Non-parametric bootstrapping can only reproduce the items that were in the original sample. The semi-parametric bootstrap assumes that the population includes other items that are similar to the observed sample by sampling from a *smoothed* version of the sample histogram. It turns out this can be done very simply by first taking a sample with replacement from the observed sample and then adding noise. For regression, the noise comes from randomly selected residuals produced by an initial model fit.

Generally, non-parametric or semi-parametric bootstrapping is employed in model selection. The idea that is had is based on the semi-parametric bootstrap and is outlined below.

- Fit initial model assuming independence within clusters and collect Pearson residuals.
- With the independence assumption, we are NOT adjusting for correlation within clusters, so the residuals should have this information.
- Using semi-parametric bootstrap, generate many residuals.
- Use the bootstrap residuals to obtain some type of *information* about the underlining mean and correlation structures.

Note for Bernoulli outcomes, Pearson residuals would NOT work. So, this is a limitation of this method.

Hours update

This was taken care of. Joe likes the plan to take 2 hours of dissertation credits for the Fall 2023, then 1 hour for the remaining semester until the end. If I need more dissertation hours at the end, then enroll in the number of hours needed.

Action items

- Continue literature review
- Add files to P: drive for Joe to access
 - I have also decided to upload my dissertation folder to GitHub, so Joe and I can access everything through the internet.
- Next meeting make time to discuss Joe's idea for model selection for GEEs