

## Week 1

- Broad goal: Fix code.
  - ☒ Use less filtered dataset
    - Now at 30% with 39 taxa.
  - ☒ Make graphs of parameter values at each iteration
    - Did a lot of graphs, histograms and changes of betas between loops. Need to do final after changes are made.
  - ☒ Try stepwise update
    - Using .1 as a constant. Still doesn't work, so likely a more complicated approach would not be helpful.
  - ☒ Fix problem where R inverse and others not updating in beta loop
  - ☒ Change R-inv function to depend on alpha instead of X and beta since already calculated.
  - ☒ Flip order of beta and phi loops.
  - ☒ Look at residuals in each iteration. (in phi step)
  - ☐ Keep something constant through loop and/or try using identity for R.
  - ☒ Try different dataset? Maybe simulated one from previous paper.
    - Same troubles
  - ☐ Distance matrix argument to dist?
    - Need both forms, so may or may not be useful
  - ☐ Check if phi is correct now, and see how is used
  - ☒ Figure out large alpha problem
    - Seems the + and one loop beta fixed this?
  - ☐ Try a different covariate?
  - ☐ Figure out why geem code uses + instead of - on update
    - I think + is correct. At least it converges.
  - ☐ Identify which covariate is causing the trouble.
  - ☐ Try removing the single OTU that seems to be contributing to problem.
  - ☐ Save everything to list correctly to look at later.
    - Saving rho, omega, phi, beta, diff (abs), num iterations
    - Also do residuals. Which residuals? Cross product ones? Squared ones? Standardized squared ones? Probably standardized but not squared. Should be distributed around zero. ?
  - ☒ Plot lists
    - Plotted rhos, omegas, phis, diffs, not sure what else to plot.
  - ☒ Remove scalar update.
    - Doesn't work. see journal
  - ☐ Write up algorithm for Yuan to go through? Ask to check.
  - ☐ Final run through before meeting of graphs and numbers
  - ☒ Move timing into function

## Week 2

Main tasks:

- To-do from sheet
- Prepare presentation
- ✓ Wednesday email research update.
- ✓ Include more taxa: more zero-inflation
  - Not great results. see journal  $\omega = .99$ ,  $\rho = .1$
  - runtime of 20 min.
- ✓ Add computation time in model.
- ☐ Try different covariates (parasite burden/antibiotics/exposure)
- ☐ Think about genus level analysis/ phylogenetic tree for taxonomic units
- ☐ Choose  $\alpha$  automatically (simple line search)
- ☐ Write up line-search steps and goals
- ☐ Make presentation
- ☐ Make sure subtraction in  $\phi$  is correct
- ☐ Figure out why adding the minus sign changes the code.
- ☐ Understand how Hessian is calculated and expected values equal zero.
- ☐ Run on previous dataset from original paper.
- ☐ Rename file with algorithm
- ☐ Compare to independence GEE written already geeglm?
- ☐ Calculate sandwich variance estimators for asymptotic significance tests.
- ☐ See if new shared tree that Tom had is the same?
- ☐ Add residuals to output and look at plots
- ★ Change initial value to something different?
  - change initial  $\rho$  to 10. Changes results (of at least ests of  $\rho$  and  $\omega$ )
  - Try previous iterations  $\rho$  and  $\omega$ .
  - See how beta estimates change
  - Run code with initial value of 1 again to compare.
- ☐ Read numerical optimization book.
- ✓ Try changing the minus sign on larger dataset.
  - Either immediate convergence or infinite alpha values on 1st iteration.....
- ✓ Set up automatic save outputs along with description
- ✓ Try with a medium sized dataset, like prevalence of .2?
- ✓ Save test datasets to files for easy loading
- ✓ Only include last beta output?
- ☐ Save convergence results of nls optim