June 16

To-Do

- ☐ Writeup current research.
- ☐ Create bibliography
- ☐ Make current document of all definitions
- ☐ Writeup dirichlet properties
- ☐ Citations
- ☐ Biological definitions
- ☐ GEE notation
- ☐ Writeup of identifiability issue

June 17 - Meeting

This week I focused on taking a step back and writing up current research, make a place for all definitions used in one place.

Todo: Literature search on dirichlet, look at link function, see if used logit,

Implement penalty first.

Look up logit idea, remember

Continue

Start career search

Wednesday Jun 22

- ☐ Work on writeup
  - Create bibliography
  - Make current document of all definitions
  - Writeup dirichlet properties
  - Citations
  - Biological definitions
  - GEE notation
  - Writeup of identifiability issue
- ☐ Literature search on dirichlet logit link function
- ☐ Implemement penalty

Literature search on dirichlet link

Douma2019a Use the log link like we were under normal parametrization: $\log(\alpha_c) = \eta_c = X_c\beta_c$.

Alternatively,

$$E(p_c) = \mu_c$$

and $var[p_c] = \frac{\mu_c(1-\mu_c)}{\phi+1}$, where $\phi = \alpha_0$.

Then the link function for μ is the multinomial logit function.

Somewhat reccomends alternate parametrization. Then only $C-1$ values fitted. treat C as baseline category

ϕ as the precision parameter, can model with log function.

So model both μ and ϕ? seperately?

Zero inflation: offers different transformation

$$p^* = \frac{p(n-1) + \frac{1}{C}}{n}$$

Also reccomended by Maier 2014 This compresses the data symmetrically around .5 so extreme values are more affected.

Reccomended R packages: DIRMULT, BRMS, DirichletReg

Notes zero-augmented Dirichlet regression Tsagris and Stewart 2017

Uses the logit link? or the one with a 1 on the bottom.

Could be interesting to look further into this method

June 23

- ☐ Work on writeup for first pomofocus
- ☐ Add penalty.

Change in beta order wouldnt change the GEE equations, but wondering how current formulation outputs old formulation? Explore code more.

Start on the penalty

I dont quite understand the derivatives parts...

We want to heavily penalize

- Add penalty $\lambda \sum_{j=1}^{p} \beta_j^2$ to GEE equation
- Add penalty

Meeting

:

Messed up on penalty, shoudl be matrix of 1s! Try this again.

Lambda larger,

Why does it have noise in only one clump?

Try larger sample size? Is it variation? Seems like.

Add diagonal back to hessian.

Try continuous X

Something with intercept. TODO: Constraint has to be for each beta. Ie the intercept betas have to sum to 0, same for all other covariate betas.

Interpretability on alphas. Cant interpret on betas since

Start trying the logit link function. Will need to re-derive the matrix of partial derivatives.

June 27th

- ☐ First & Second pomofocus on the writeup
- ☐ Third pomofocus on literature review

References

Maier, Marco J. (Jan. 2014). *DirichletReg: Dirichlet Regression for Compositional Data in R*. https://epub.wu.ac.at/4077/. Paper. Vienna.

Tsagris, Michail and Connie Stewart (June 2017). *A Dirichlet Regression Model for Compositional Data with Zeros*. DOI: 10.48550/arXiv.1410.5011. arXiv: 1410.5011 [stat].