Contents

Motivation

Start writing what might be a good introduction for any paper or chapter.

Include citations!

Definitions

- ASV
- OTU
- Zero-Inflation
- Phylogenetic tree
- Compositional
- Covariates/variables/ coefficients

Notation

- Number of samples: n
- Sample index: $i : i = 1, \ldots, n$
- Number of ASVs: k
- ASV index: $j : j = 1, \ldots, k$
- Number of covariates: q. Note that q does not include the intercept.
- Covariate index: $k : k = 1, \ldots, q$
- Responses: Transformed counts into relative abundances.
  Note that an alternative transformation to relative abundance might be good.

$$y_{ij}$$

$$\mathbf{y}_{np\times 1} = \begin{pmatrix} y_{i=1,j=1} \\ \vdots \\ y_{i=1,j=p} \\ \vdots \\ y_{i=n,j=1} \\ \vdots \\ y_{i=n,j=p} \end{pmatrix}$$

- Design matrix $\mathbf{x}$ ('little x').

$$\mathbf{x} = \begin{pmatrix} x_{i=1,k=1} & \cdots & x_{i=1,k=q} \\ \vdots & \ddots & \\ x_{i=n,k=1} & \cdots & x_{i=n,k=q} \end{pmatrix}$$

Since we don't have coefficient values measured on each ASV, they are only measured on each sample. Thus $x_{ij} = x_i$ for all $j = 1, \dots, p$. Then we convert to design matrix big X:

Design matrix X:

$$\begin{aligned} X &= x \otimes I_p \\ &= \begin{pmatrix} x_{i=1,k=1} & 0 & \cdots & 0\overbrace{\cdots}^{p} & 0 & x_{i=1,k=q} & 0 & \cdots & 0 \\ 0 & x_{i=1,k=1} & & & & & & & \\ x_{i=2,k=1} & 0 & \cdots & x_{i=2,k=2} & \cdots & x_{i=2,k=q} & 0 & \cdots & \end{pmatrix} \end{aligned}$$

- Parameter vector β with entries β_{jk}

$$\boldsymbol{\beta}_{pq\times 1} = \begin{pmatrix} \beta_{j=1,k=1} \\ \vdots \\ \beta_{j=p,k=1} \\ \vdots \\ \beta_{j=1,k=q} \\ \vdots \\ \beta_{j=p,k=q} \end{pmatrix}$$

- Link:

  Link covariates to response.

  First assume that $\mathbf{y}$ has the same mean and covariance structure as if $\mathbf{y}_i \sim Dir(\alpha_{1p}, \dots, \alpha_{ip})$ for all $i = 1, \dots, n$.

  Then,

$$\begin{aligned} g(\alpha) = \log(\alpha) &= X\beta \\ \log(\alpha_{ij}) &= x_i\beta_{ij} \end{aligned}$$

Dirichlet ideas

Assume $\eta_i = (\eta_{i1}, \dots, \eta_{ip}) \sim Dir(\alpha_{i1}, \dots, \alpha_{ip})$, and $\sum_{j=1}^{p} \alpha_{ij} = \alpha_{i0}$

Dirichlet Expected value

$$E(\eta_{ij}) = \frac{\alpha_{ij}}{\alpha_{i0}}$$

Dirichlet variance

$$Var(\eta_{ij}) = \frac{\alpha_{ij}(\alpha_{i0} - \alpha_{ij})}{\alpha_{i0}^2(\alpha_{i0} + 1)}$$

Note: $0 \leq Var(\eta_{ij}) \leq 1$

Dirichlet covariance

For $i \neq j$

$$Cov(\eta_{is}, \eta_{it}) = -\frac{\alpha_{is}\alpha_{it}}{\alpha_{i0}^2(\alpha_{i0}+1)}$$

Note this matrix is singular. Note this quantity is always negative

Dirichlet correlation

For $i \neq j$

$$\begin{aligned} Cor(\eta_{is}, \eta_{it}) &= \frac{Cov(\eta_{is}, \eta_{it})}{\sqrt{Var(\eta_{is})Var(\eta_{it})}} \\ &= -\frac{\alpha_{is}\alpha_{it}}{\alpha_{i0}^2(\alpha_{i0}+1)} \cdot \sqrt{\frac{\alpha_{i0}^2(\alpha_{i0}+1)}{\alpha_{is}(\alpha_{i0}-\alpha_{is})}}\sqrt{\frac{\alpha_{i0}^2(\alpha_{i0}+1)}{\alpha_{it}(\alpha_{i0}-\alpha_{it})}} \\ &= -\sqrt{\frac{\alpha_{is}\alpha_{it}}{(\alpha_{i0}-\alpha_{is})(\alpha_{i0}-\alpha_{it})}} \end{aligned}$$

Note this is always negative

GEE ideas

(include definitions of all)

Identifiability issues