15 August 15th

15.1 Goals for the day

Goals from Last Time

1. *Continue analyzing cubfits-master/R files, especially mycubappr.r*
   (a) *Especially study the MCMC from mycubappr.r 124-196*
   (b) *nu.Phi and sigma.Phi?*
2. *Read REU Paper on Nonsense Model*

Additional Goals

15.2 Progress/Notes

15.2.1 Continue analyzing cubfits-master/R files, especially mycubappr.r

I want to analyze the MCMC that happens in my.cubappr.r, which is only about 50 lines of code in my.cubappr.r, but it expands more.

This function calls my.drawBConditionalAll... but that function is gotten from .cubfitsEnv. It could be .current, .random, or .RW_Norm. Only .random uses VGAM, so that's relevant. I can't tell from my.cubfitsappr.r or roc.appr.r which is being called. The best next step is to probably put in print statements for each of these splits and see which is called.

my.drawPhiConditionalAll is from .cubfitsEnv as well, but it looks like there is only one .drawPhiConditionalAll. Will put in a print statement just to double check.

Understanding the workflow is key to understanding the code.

15.2.2 Read REU Paper on Nonsense Model

Equation 2 needed explanation.

$$\eta(\vec{c}) = \frac{\sum_{i=1}^{n+1} \beta_{i-1} p_i \sigma_{i-1}}{\sigma_n}$$

Where

β_i = cost of a codon $= a_1 + a_2 i$

p_i = probability of an error at codon i

σ_i = probablity of succesfulling reaching codon $i = \prod_{j=1}^{i}(1 - p_j)$

We peel off the n^{th} term.

$$\eta(\vec{c}) = \beta_n + \frac{\sum_{i=1}^{n} \beta_{i-1} p_i \sigma_{i-1}}{\sigma_n}$$

Substitute σ

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} p_i \frac{\prod_{k=1}^{i-1} 1 - p_i}{\prod_{k=1}^{n} 1 - p_k}$$

Divide out the $\prod$ term

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} p_i \frac{1}{\prod_{k=i}^{n} 1 - p_k}$$

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} \frac{p_i}{1 - p_i} \prod_{k=i+1}^{n} \frac{1}{1 - p_k}$$

A simplification on the $\prod$

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} \frac{p_i}{1 - p_i} \prod_{k=i+1}^{n} \frac{1 - p_k + p_k}{1 - p_k}$$

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} \frac{p_i}{1 - p_i} \prod_{k=i+1}^{n} (1 + \frac{p_k}{1 - p_k})$$

The version in the REU paper, expanding β

$$\eta(\vec{c}) = a_1 + a_2 n + \sum_{i=1}^{n} (a_1 + a_2(i-1)) \frac{p_i}{1 - p_i} \prod_{k=i+1}^{n} (1 + \frac{p_k}{1 - p_k})$$

Easy to read version using $\beta_i = a_1 + a_2 i$ and $\omega_i = \frac{p_i}{1 - p_k}$

$$\eta(\vec{c}) = \beta_n + \sum_{i=1}^{n} \beta_{i-1} \omega_i \prod_{k=i+1}^{n} (1 + \omega_k)$$

Take the first order taylor approximation, about $p_i = 0$, then apply that to the fitness function ((1) in the paper), and we get the main function

$$Pr(\vec{c}|\phi) \propto \prod_{i=1}^{n} \exp[\ln \mu_i + \omega_i (a_1 + a_2) y_1 + \omega_i a_2 y_1 i]$$

The rest of the paper has their results, which might be important to look at, in the future. For now, the important part is to understand how the equations work, and then how the code works.

15.3 Future Goals

1. Analyze my.cubappr.r
   (a) Add in print statements, then rerun the example
   (b) my.drawBConditionalAll.??????
2. (Optional) Study R user manual more?