1 Mike Questions

1. ~~Is there a difference between $\vec{c_i}$ and c_i? Should we just use $\vec{c}$ and c_i?~~
2. ~~Why does a $\vec{c_i}$ term appear in the fitness? is $\omega_{\vec{c_i}}$ a function of $\vec{c_i}$ in this notation?~~
3. Where do the mutation terms go? I think I'll include them for completeness. Just turn $\exp\left[-\sum_j \beta_{c_{i,j}}\omega(\vec{c}_{i,j})\phi\right]$ into $\exp\left[-\sum_j (m_{c_j} - \beta_{c_{i,0}}\omega(\vec{c}_{i,j}))\phi\right]$?

2 Notation

$\vec{c}$ is the vector of codons (of length n, the length of the gene)

$\vec{c}_{i,j}$ is the codon in the j^{th} position of the i^{th} synonymous gene

$\vec{c}_{i,j,\ell}$ denotes the ℓ^{th} synonym to codon $\vec{c}_{i,j}$

$\vec{\theta}$ is our given conditions, including β, ϕ, ω, and the effective population size

$\sum_j$ is shorthand for $\sum_{j=1}^{n}$, the sum over the whole gene

$\beta_{c_{i,j}}$ or $\beta_{c_{i,j,k}}$ is the cost of codon translation for codon $c_{i,j}$ or $c_{i,j,k}$. In the NSE model, this is constant. In the ROC model, this is the main variable. Note that the j subscript doesn't matter, β is position independent.

$\omega(\vec{c}_{i,j})$ is the odds of a nonsense error for codon $\vec{c}_{i,j}$, $\dfrac{p_{\vec{c}_{i,j}}}{1 - p_{\vec{c}_{i,j}}}$

ϕ is the expression level of the gene

$\mathbb{C}$ is the set of all possible synonymous sequences

ν is the number of synonymous codons to $\vec{c_i}$

3 Calculation

"The probability of seeing codon $\vec{c_i}$ at position i (given our conditions) is approximately the fitness of that sequence divided by the sum of the fitness of all possible synonymous sequences"

$$P(\vec{c_i}|\vec{\theta}) \approx \frac{\exp\left[-\sum_j \beta_{c_{i,j}}\omega(\vec{c}_{i,j})\phi\right]}{\sum_{K\in\mathbb{C}} \exp\left[-\sum_j \beta_{K_{i,j}}\omega(\vec{c}_{i,j})\phi\right]}$$

Rewrite the sum as the sum across the synonyms. Added notation:

ν is the number of synonyms for the given codon i

$$= \frac{\sum_{\kappa\in\mathbb{C}|\kappa_i = c_i} \exp\left[-\sum_j \beta_{\kappa_{i,j}}\omega(\vec{\kappa}_j)\phi\right]}{\sum_{\ell=1}^{\nu}\sum_{K\in\mathbb{C}|K_{i,j,\ell}=c_{i,j,\ell}} \exp\left[-\sum_j \beta_{K_{i,j}}\omega(\vec{K}_{i,j})\phi\right]}$$

Pull out the term for the current codon in the chain, where $c_j = c_i$

$$= \frac{\left[\sum_{\kappa\in\mathbb{C}} \exp\left[-\sum_{j\neq i} \beta_{\kappa_{i,j}}\omega(\vec{\kappa}_j)\phi\right]\right]\left(\exp\left[-\beta_{c_i}\omega(\vec{c_i})\phi\right]\right)}{\sum_{\ell=1}^{\nu}\sum_{K\in\mathbb{C}|K_{i,j,\ell}=c_{i,j,\ell}} \exp\left[-\sum_j \beta_{K_{i,j}}\omega(\vec{K}_{i,j})\phi\right]}$$

Additional expansion that wasn't on the board. Pull out the $K_{i,j,\ell}$ term, knowing that $K_{i,j,\ell} = c_{i,j,\ell}$

$$= \frac{\left[\sum_{\kappa\in\mathbb{C}} \exp\left[-\sum_{j\neq i}\beta_{\kappa_{i,j}}\omega(\vec{\kappa}_j)\phi\right]\right]\left(\exp\left[-\beta_{c_i}\omega(\vec{c}_i)\phi\right]\right)}{\sum_{\ell=1}^{\nu}\sum_{K\in\mathbb{C}|K_{i,j,\ell}=c_{i,j,\ell}} \exp\left[-\sum_{j\neq i}\beta_{K_{i,j}}\omega(\vec{K}_{i,j})\phi\right]\exp\left[-\beta_{c_{i,j,\ell}}\omega_{c_{i,j,\ell}}(\vec{c}_{i,j,\ell})\phi\right]}$$

The ℓ term in the denominator can be pulled out of the inner sum, leaving

$$= \frac{\left[\sum_{\kappa\in\mathbb{C}} \exp\left[-\sum_{j\neq i}\beta_{\kappa_{i,j}}\omega(\vec{\kappa}_j)\phi\right]\right]\left(\exp\left[-\beta_{c_i}\omega(\vec{c}_i)\phi\right]\right)}{\sum_{\ell=1}^{\nu}\left(\exp\left[-\beta_{c_{i,j,\ell}}\omega_{c_{i,j,\ell}}(\vec{K}_{i,j,\ell})\phi\right]\sum_{K\in\mathbb{C}|K_{i,j,\ell}=c_{i,j,\ell}} \exp\left[-\sum_{j\neq i}\beta_{K_{i,j}}\omega(\vec{K}_{i,j})\phi\right]\right)}$$

Since the inner sum (sum over $\mathbb{C}$) no longer relies on ℓ, we can restate the above as

$$= \frac{\left[\sum_{\kappa\in\mathbb{C}} \exp\left[-\sum_{j\neq i}\beta_{\kappa_{i,j}}\omega(\vec{\kappa}_j)\phi\right]\right]\left(\exp\left[-\beta_{c_i}\omega(\vec{c}_i)\phi\right]\right)}{\sum_{\ell=1}^{\nu}\left(\exp\left[-\beta_{c_{i,j,\ell}}\omega_{c_{i,j,\ell}}(\vec{C}_{i,j,\ell})\phi\right]\right)\left(\sum_{K\in\mathbb{C}} \exp\left[-\sum_{j\neq i}\beta_{K_j}\omega(\vec{K}_{i,j})\phi\right]\right)}$$

Which cancels to

$$P(\vec{c}_i|\vec{\theta}) \approx \frac{\exp\left[-\beta_{c_i}\omega(\vec{c}_i)\phi\right]}{\sum_{\ell=1}^{\nu}\exp\left[-\beta_{c_{i,j,\ell}}\omega_{c_{i,j,\ell}}(\vec{c}_{i,j,\ell})\phi\right]}$$