Likelihood

$$\mathbb{P}\left(\tilde{D}_n \;\middle|\; x_n\right) = \sum_{\mathbf{x}^n} \mathbb{P}\left(\mathbf{x}^n \mid x_n\right) \prod_{i \in \mathbf{O}(n)} \mathbb{P}\left(\tilde{D}_i \;\middle|\; x_i\right); \tag{1}$$

$$\mathbb{P}\left(\mathbf{x}^n = \mathbf{x} \mid x_n\right) = \frac{\exp\left\{\theta^{\mathbf{t}} s(\mathbf{x}, x_n)\right\}}{\sum_{\mathbf{x}^n} \exp\left\{\theta^{\mathbf{t}} s(\mathbf{x}^n, x_n)\right\}} \tag{2}$$

where $\mathbf{x}^n \equiv \{x_i^n\}_{i \in \mathbf{O}(n)}$ is an array of size P (functions) $\times$ $|\mathbf{O}(n)|$ (offspring) representing the state of node n's offspring, x_n is a binary vector representing the state of node n, θ is a column vector of parameters, and $s(\cdot)$ is a column vector of sufficient statistics which may include terms such as: the total number of functional gains, the number of subfunctionalization or neofunctionalization events, etc

Prediction

$$\mathbb{P}\left(\mathbf{x}^p = \mathbf{x} \,\middle|\, \tilde{D}\right) = \underbrace{\left\{ \prod_{m \in \mathbf{O}(p)} \mathbb{P}\left(\tilde{D}_m \,\middle|\, x_m\right) \right\}}_{\text{Everything below } \mathbf{x}^p} \underbrace{\sum_{x_p} \mathbb{P}\left(x_p \,\middle|\, \tilde{D}\right) \frac{\mathbb{P}\left(\mathbf{x}^p = \mathbf{x} \mid x_p\right)}{\mathbb{P}\left(\tilde{D}_p \,\middle|\, x_p\right)}}_{\text{Everything above } \mathbf{x}^p}$$