Untitled

A set of n genes $N \equiv \{i \in \{1, \dots, n\}\}$ is partially order accordignly to its relations as parent and offsprings such that for any $i, j \in N$ for which a directed pathlength exists, $i < j$ iff i is a parent node and j a descendant.

For leaf nodes we use missclassification probabilities ψ, for which we have 2^p different states, hence we obtain an $N \times 2^P$ array. Each row is filled with a particular combination of $\{0,1\}^P$, e.g. $s = \{0,0,0\}$.

Each leaf is defined by a vector $z_l \equiv \{z_{lp}\}_{p=1}^P$ with

$$z_{lp} = \begin{cases} 1 & \text{if the function } p \text{ is active} \\ 0 & \text{if the function } p \text{ is not active} \\ 9 & \text{if we don't have information} \end{cases} \tag{1}$$

This way, given that the true state is $s \equiv \{s_p\}_{p=1}^P$, the probability of having correctly classified Z_l is:

$$\Psi_{ss'} = \Pr\{Z = s' \mid S = s\} = \begin{cases} \prod_p \{\psi_p^{\mathbf{1}\{s'_{lp} = s_{lp}\}} (1 - \psi_p)^{\mathbf{1}\{s'_{lp} \neq s_{lp}\}}\} & \text{if } s'_{lp} \neq 9 \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

For internal node i we use gain and loss functions, also defined by a $N \times 2^P$ array. This is conditional on i's O_i offsprings $o_i \subset \{j \in N : j < i\}$ and the true state s. Let $x_i \equiv \{x_{ip}\}_{p=1}^P$ denote the vector of functional states of i, then, the probability that we observe $X = x_i$ conditional on the true state of its offsprings $j \in o_i$ is:

$$\Pr\{\mu, \psi\}_{n,s} = \prod_{o_n} \sum_{s'_n} \Psi_{s'_n s'} \prod_p \left(\left[\underbrace{\mu_0^{\mathbf{1}\{s'_{np}\}}}_{\text{Gain}} \underbrace{(1 - \mu_0)^{\mathbf{1}\{\neg s'_{np}\}}}_{\text{No gain}} \right]^{\mathbf{1}\{\neg s_p\}} \times \left[\underbrace{\mu_1^{\mathbf{1}\{\neg s'_{np}\}}}_{\text{Loss}} \underbrace{(1 - \mu_1)^{\mathbf{1}\{s'_{np}\}}}_{\text{No Loss}} \right]^{\mathbf{1}\{s_p\}} \right) \tag{3}$$

Computationally, observe that the larger parenthesis can be computed only once and then retrieved depending on the values of $\{s'_{np}, s_p\}$. Let $M \equiv \{m_{s_n,s}\}$ to be an array of size 2×2 holding the Gain/Loss probabilities, then, the previous equation reduces to:

$$\Pr\{\mu, \psi\}_{n,s} = \prod_{o_n} \sum_{s'_n} \Psi_{s'_n s'} \prod_p m_{s'_{np}, s} \tag{4}$$

Finally, define $\pi \equiv \{\pi_0, \pi_1\}$ to be the root node state probabilities, then, the likelihood function can be computed as

$$L(\pi, \mu, \Psi) = \sum_s \pi_s \Pr\{\mu, \psi\}_{n,s} \tag{5}$$