Genotype imputation

Define hidden stage as $\mathbf{x}$, which has K different reference haplotype stages; define $\mathbf{o}_{1,S}$ as the observational sequence. Then the genotype imputation is the argmax of the posterior probability $p(x_j = k|h_{k,\leq M}) \propto P(x_j = k, h_{k,\leq j})P(h_{k,[j+1,M]}|x_j = k)$

Forward path:

$$\begin{aligned}\alpha_j(x) \equiv P(x_j = x, h_{k+1,\leq j}) &= \sum_{k' \in K} P(h_{k+1,j}|x_j = x)P(x_j = x|x_{j-1} = k')P(x_{j-1} = k', h_{k+1,\leq j}) \\ &= \gamma_j(x)(p_{j-1}\alpha_{j-1}(k) + (1 - p_{j-1})\frac{1}{k}\sum_{k'} \alpha_{j-1}(k'))\end{aligned}$$

where $\gamma_{j+1}(x) = Pr(h_{k+1,j+1}|X_{j+1} = x, h_1, ..., h_k)$

Backward path:

$$\begin{aligned}\boldsymbol{\beta}_j(x) \equiv P(h_{k+1,[j,M]}|X_{j-1} = x) &= \sum_{k' \in K} P(h_{k+1,[j+1,M]}|X_j = k')P(h_{k+1,j}|x_j = k')P(x_j = k'|X_{j-1} = x) \\ &= \sum_{k' \in K} \boldsymbol{\beta}_{j+1}(k')\gamma_j(k')P(X_j = k'|X_{j-1} = x) \\ &= p_{j-1}\beta_{j+1}(k)\gamma_j(k) + (1 - p_{j-1})\frac{1}{K}\sum_{k'} \beta_{j+1}(k')\gamma_j(k')\end{aligned}$$

where $p_j = exp(-\rho_j d_j/K)$ and $\gamma_j(x) = Pr(h_{k+1,j+1}|X_{j+1} = x, h_1, ..., h_k)$

Log forward path:

$$\begin{aligned}\log(\alpha_j(k)) &= \log\gamma_j(k) + \log\left(p_{j-1}\alpha_{j-1}(k) + (1 - p_{j-1})\frac{1}{K}\sum_{k'} \alpha_{j-1}(k')\right) \\ &= \log\gamma_i(k) + \log\left(exp(-\rho_j d_j/K)exp\log(\alpha_{j-1}(k)) + \frac{1}{K}\sum_{k'} exp\log\alpha_{j-1}(k') - exp(-\rho_j d_j/K)\sum_{k'} exp\log\alpha_{j-1}(k')/K\right) \\ &= \log\gamma_i(k) + \log\left(exp(-\rho_j d_j/K + \log\alpha_{j-1}(k)) + \sum_{k'} exp\log\alpha_{j-1}(k')/K - \sum_{k'} exp(\log\alpha_j(k') - \rho_j d_j/K)/K\right)\end{aligned}$$

Posterior calculation:

$$\begin{aligned}p(h_j = 1|h_{1,M}) &= \sum_{k \in K} p(h_j = 1, x_j = k|h_{1,M}, h_{train}) \\ &= \sum_{k \in K} p(h_j = 1|x_j = k, h_{1,M}, h_{train})p(x_j = k|h_{1,M}) \\ &= \sum_{k \in K} \mathbf{1}[h_{x,j} = 1]\left(K/(K + \tilde{\theta}) + (1/2)\tilde{\theta}/(K + \tilde{\theta})\right)p(x_j = k|h_{1,M})\end{aligned}$$