Supplementary Material for "Generating heterogeneous data on gene trees"

Martí Cortada Garcia, Adrià Diéguez Moscardó, Marta Casanellas

Lemma 1. *Given a time-reversible Markov matrix M with stationary distribution π, let $\tilde{M} = (1-a)M + aI$. If $a \in (0,1)$, then $\tilde{M}$ is a Markov matrix with stationary distribution π.*

Proof. As the set of Markov matrices is a convex space, $\tilde{M}$ is also a Markov matrix if $a \in (0,1)$.

The matrix M is time-reversible with stationary distribution π if and only if

$$\pi = \pi M \quad \text{and } D_\pi M = M^t D_\pi, \tag{1}$$

where D_π is the diagonal matrix with π on its diagonal elements (t denotes transpose here).

To show that these conditions hold for $\tilde{M}$ is immediate: $\pi\tilde{M} = (1-a)\pi M + a\pi = \pi$ and $D_\pi \tilde{M} = (1-a)D_\pi M + aD_\pi = (1-a)M^t D_\pi + aD_\pi = \left((1-a)M^t + aI\right) D_\pi = \tilde{M}^t D_\pi$. □

Remark 2. Note that if $M = ADA^{-1}$ is an eigendecomposition of M with eigenvalues 1, λ_2, λ_3, and λ_4, we have $\tilde{M}_2 = A((1-a)D + aI)A^{-1}$ and its determinant can be computed as

$$p(a) = ((1-a)\lambda_2 + a)((1-a)\lambda_3 + a)((1-a)\lambda_4 + a).$$

We compute the roots of the polynomial $p(a) - d_2$ (where $d_2 = \det(M_2)$) and choose the root a in $(0,1)$ if exists (it may happen that the real root of this polynomial is not in this interval, and then we repeat the process in Algorithm 2 starting with another M).