Question: How can you modify this for $K=2,3,\ldots$ time steps?

S_0: $\vec{p}_0$

M

S_1 $\vec{p}_1$

$\Delta t = 1$ time step

$\vec{p}_1 = \vec{p}_0 M$

S_2 $\vec{p}_2$

$\vec{p}_2 = \vec{p}_1 M = \vec{p}_0 M^2$

S_3 p_3

$\vec{p}_3 = \vec{p}_0 M^3$

$\vdots$

S_K

$\vec{p}_K = \vec{p}_0 M^K$

S_0

S_K

The formulation thus far is for is for "DISCRETE TIME"

or modelling evolution from tail to tip of an edge.

Alternatively, there is the CONTINUOUS TIME formulation

$S_0 = S(t=0)$

$S_K = S(t=t_K)$

In this setup, we introduce a rate matrix Q

rates = derivatives!

$$Q = \begin{pmatrix} q_{RR} & q_{RY} \\ q_{YR} & q_{YY} \end{pmatrix}$$

the off-diagonal entries are non-negative $q_{YR}, q_{RY} \geq 0$ and

row sums equal 0

units

$$\frac{\text{substitution per site}}{\text{unit time}}$$

q_{RY} = rate at which Rs are converted to Ys ≥ 0

q_{YR} = " " " Ys are converted to Rs ≥ 0

$q_{RR} \leq 0$ rate at which leaving R state

$q_{YY} \leq 0$

Importantly, $q_{RR} + q_{RY} = 0$ rates balance

added to Y

$\approx$ lost to R class

The root distribution is now a function of time

$$\vec{p}_0 = \vec{p}(t=0) = \big(p_R(0),\ p_Y(0)\big) \quad \text{and at time } t$$

$$\vec{p}_t = \big(p_R(t),\quad p_Y(t)\big)$$

distribution of purines, pyrimidines at time $t \geq 0$.

The distribution of states satisfies the following system of differential equations:

$$(*) \quad \begin{aligned} \frac{d}{dt} p_R(t) &= p_R(t)\, q_{RR} + p_Y(t)\, q_{YR} \\ \frac{d}{dt} p_Y(t) &= \underbrace{p_R(t)\, q_{RY}} + \underbrace{p_Y(t)\, q_{YY}} \end{aligned}$$

(current freq. of R)(rate of conversion R→Y) + (current freq of Y)(rate of loss Y)

In matrix form, the right hand side is:

$$\vec{p}(t)\, Q$$

Check:

$$\big(p_R(t) \quad p_Y(t)\big) \begin{pmatrix} q_{RR} & q_{RY} \\ q_{YR} & q_{YY} \end{pmatrix}$$

Thus, $(*)$ is the differential equation

$$\vec{p}\,'(t) = p(t)\, Q \qquad \text{with} \qquad \vec{p}(0) = \big(p_R(0),\ p_Y(0)\big)$$

Using matrix exponentials, the solution is

$$\vec{p}(t) = \vec{p}(0)\, e^{Qt}$$

with $$e^{Qt} = I + (Qt) + \frac{(Qt)^2}{2!} + \frac{(Qt)^3}{3!} + \frac{(Qt)^4}{4} + \cdots$$

$$\vec{p}(t) = \vec{p}_0 e^{Qt} = \vec{p}_0 M(t)$$

distribution of states at time t

initial state

product

substitution

Markov model at time t $\quad M(t) = e^{Qt}$

Demo: $\quad t=0$ $\quad \vec{p}_0 = (.7, .3) \quad Q = \begin{pmatrix} -.1 & .1 \\ .2 & -.2 \end{pmatrix}$

$t \quad M(t) = e^{Qt}$

Note:

- e^{Qt} is a Markov matrix
- How should the diagonal entries of $\underset{t=1}{M(1)}$ and $\underset{t=2}{M(2)}$ compare?
- MATH: computing matrix exponentials

a) Q is diagonalizable $\quad Q = S \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{pmatrix} S^{-1}$

b) $M = e^{Qt} = I + (Qt) + \frac{(Qt)^2}{2!} + \frac{(Qt)^3}{3!} + \dots$

$$= I + S\Lambda t S^{-1} + S\frac{(\Lambda t)^2}{2!}S^{-1} + S\frac{(\Lambda t)^3}{3!}S^{-1} + \dots$$

$$\Lambda t = \begin{pmatrix} \lambda_1 t & & 0 \\ & \lambda_2 t & \\ 0 & & \lambda_3 t \end{pmatrix}$$

$$= S\left[I + \Lambda t + \frac{(\Lambda t)^2}{2!} + \frac{(\Lambda t)^3}{3!} + \frac{(\Lambda t)^4}{4} + \dots\right] S^{-1}$$

$$= S e^{\Lambda t} S^{-1} = S \begin{pmatrix} 1 & & & \\ & e^{-\lambda_1 t} & & 0 \\ & & e^{-\lambda_2 t} & \\ 0 & & & e^{-\lambda_3 t} \end{pmatrix} S^{-1}$$

i.e. diagonalize Q, exponentiate diagonal entries of Λ, ...

Summary: Continuous-time formulation

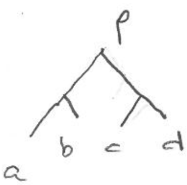
Initial base distribution $\vec{p}_0 = (p_R(0),\ p_Y(0))$ $\qquad t=0$.

At any time $t>0$, the Markov transition matrix is given by

$$M(t) = e^{Qt} \qquad \text{for a fixed rate matrix } Q.$$

III) Markov models on trees

The GENERAL MARKOV model:



parameters: a (rooted) tree T^ρ

and numerical parameters

1) a root dist. vector $\vec{p}_\rho = (p_A, p_G, p_C, p_T)$ $\qquad p_i \geq 0,\ \Sigma p_i = 1$

2) for each edge e of T^ρ directed away from ρ, a

Markov matrix $M_e = \begin{pmatrix} p_{AA} & p_{AG} & p_{AC} & p_{AT} \\ \vdots & & \ddots & \vdots \\ p_{TA} & & & p_{TT} \end{pmatrix}$ with

$$p_{ij} \geq 0, \quad \sum_{j=1}^{4} p_{ij} = 1 \qquad p_{ij} = P(S_{end} = j \mid S_{beg} = i)$$

We make the Markov assumption that the substitution process

along an edge $S_0 \longrightarrow S_1$ depends only on the current state at S_0

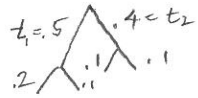
The continuous-time formulation:

parameters T^ρ, $\vec{p}_\rho$ but replace 2) with

2a) a rate matrix $Q = \begin{pmatrix} q_{AA} & \cdots & q_{AT} \\ \vdots & \ddots & \vdots \\ q_{TA} & \cdots & q_{TT} \end{pmatrix}$

with non-negative off diagonal entries and row sums 0

2b) for each edge e in the tree, a non-negative branch length t_e.



The Markov transition matrix $M_e = e^{Qt_e} = \left(e^{Q}\right)^{t_e}$

An important assumption in the continuous time formulation is a <u>Common Rate Matrix</u> <u>Q</u> for all edges of the tree.