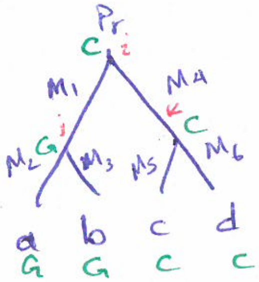
F: "relative" frequency table

has the same information as <u>count data</u> table

In Log-det distance, we use "relative" frequency table.

Model:



single site

	①	②
a	G	A
b	G	A
c	C	A
d	C	A

multiple sites in "data" $\begin{cases} \text{independent} \\ \text{identically distributed} \end{cases}$

each site is an independent trial of the same process

identically distributed $\Leftrightarrow$ using the same model parameters

$$P(AAAA \mid T, M_1, \ldots, M_6, P_r) = \sum_k \sum_j \sum_{i=1}^{4} P_r(i)\, M_1(i,j)\, M_2(j,1)\, M_3(j,1)\, M_4(i,k)\, M_5(k,1)\, M_6(k,1)$$

Expected value of pattern AAAA.

"Expected Pattern frequency Array" with $4 \times 4 \times 4 \times 4$ entries $\underset{256}{\shortparallel}$

$E(AAAA) = .3$

$E(AAAG) = .1$

$E(AAGG) = .005$

$E(GGGG) = .1$

$\boxed{\text{The sum of them is } 1}$.

$\underline{n} = 1,000,000$

the number of sites that have pattern AAAA is $.3 \times 1000000 = 300,000$

$\Rightarrow \hat{P}^{AAAA}_{1000000} = \dfrac{300000}{1000000}$

The parameters of GTR $+\Gamma+I$

w : 1

Q : 5 or 6

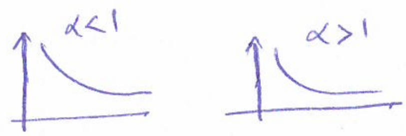
$P_{r,Q}$: 3

t_e : $2n-3$

$P_{r,I}$: 3

Γ : 1 (α)

mean = 1



$$P_{GTR+\Gamma+I} = w\, P_{GTR+\Gamma} + (1-w)\, P_I$$

$P_{GTR+\Gamma+I}$ ↓ 4 classes

$P_{GTR+\Gamma}$ ↓ Expected patterns for variable sites

P_I ↓ Expected patterns for Invariable sites

$2n-3$

3 classes of GTRs

Q : 5 or 6

P_r = 3

$$P_{total} = w_1 P_{GTR,1} + w_2 P_{GTR,2} + (1-w_1-w_2) P_{GTR,3}$$

$$2n-3 + \underset{\substack{\uparrow \\ w_i}}{2} + \underset{\substack{\downarrow \\ \text{3 different classes}}}{3}(8)$$

$8 \rightarrow \underset{Q\ +\ P_r}{5 + 3}$

- eigen values, eigenvectors

$-\frac{4}{3}\alpha$

$$Q = \begin{pmatrix} -0.6 & .1 & .2 & .3 \\ & & 4 & 5 \\ & & & .12 \\ & & & \end{pmatrix} \qquad SQS^{-1} = \begin{pmatrix} 0 & & & \\ & -1 & & 0 \\ & 0 & -2 & \\ & & & -3 \end{pmatrix}$$

eigen values of M : $1, e^{-1}, e^{-2}, e^{-3} > 0$

- show that a stable distribution.

$$M = \begin{pmatrix} .7 & .1 & .1 & .1 \\ & & \vdots & \\ & & & \end{pmatrix}$$ for JC

$S_0 \downarrow S_1$

$\vec{P}_0 = (.25 \quad .25 \quad .25 \quad .25)$

$\vec{P}_0 M = (.25 \quad .25 \quad .25 \quad .25)$ show that it is stable.

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

TIME - REVERSIBILITY

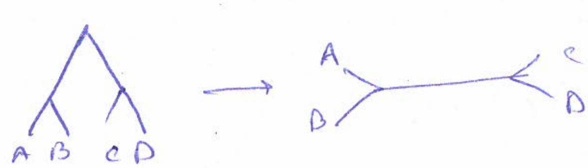
$$P = \mathrm{diag}(P_0)\, M = \begin{pmatrix} P_A & & & 0 \\ & P_G & & \\ & 0 & P_C & \\ & & & P_T \end{pmatrix} \begin{pmatrix} M \end{pmatrix} = \begin{pmatrix} \overset{AA}{p(S_0 = A, S_1 = A)}, & p(AG), & p(AC)\, p(AT) \\ & & \end{pmatrix}$$

$S_0 \downarrow S_1$

$P = P^T$ stable distribution

$\uparrow$

$S_1 \uparrow S_0$



outgroup Can be used

rate $\binom{10}{2}$

waiting time $\frac{1}{\binom{10}{2}}$

- large population $N_1 = 20000$ $\binom{20000}{2}$ rate

$\boxed{WF}$ $N_2 = 300$ $\binom{300}{2}$ = rate → this is faster

choosing ~~the~~ the same parent = $1/300$

$$P(c_2 = 4 \text{ gens}) = \left(1 - \frac{1}{20000}\right)^3 \frac{1}{20000}$$