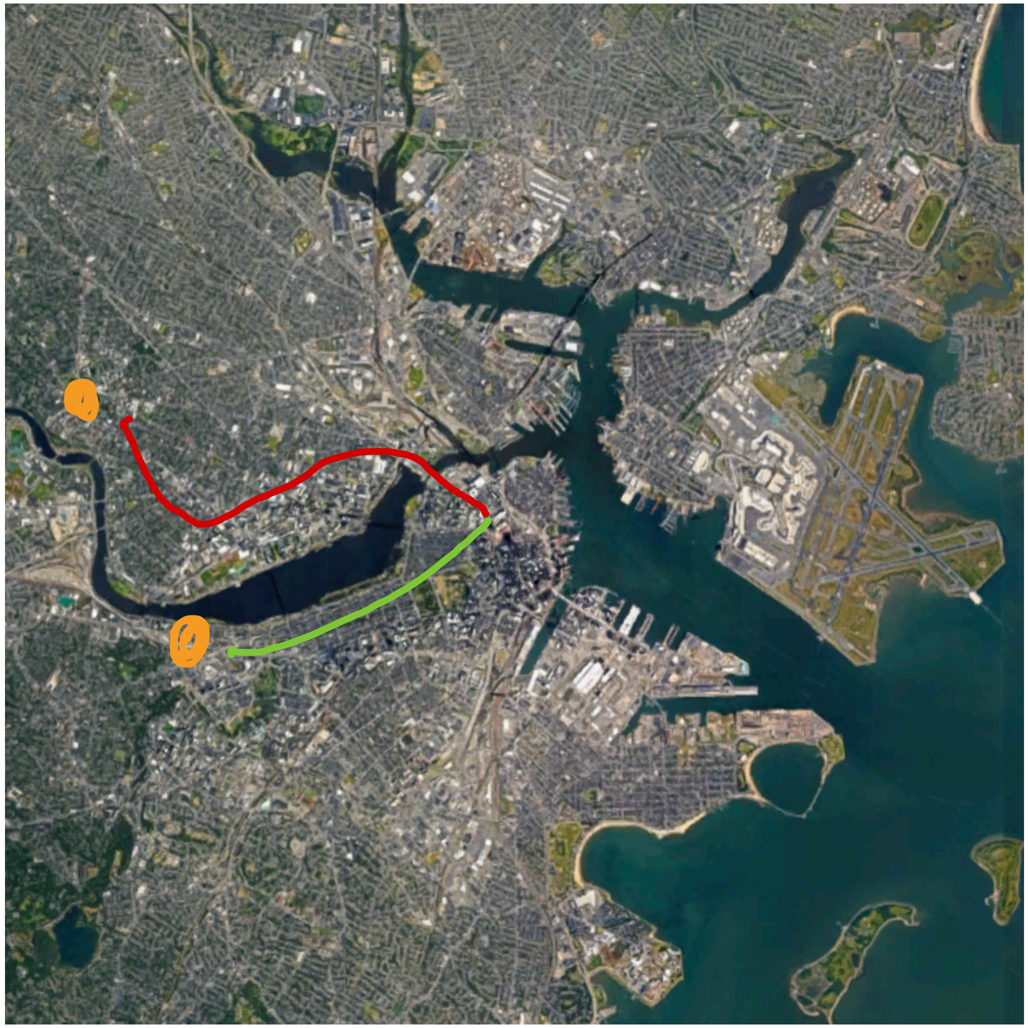
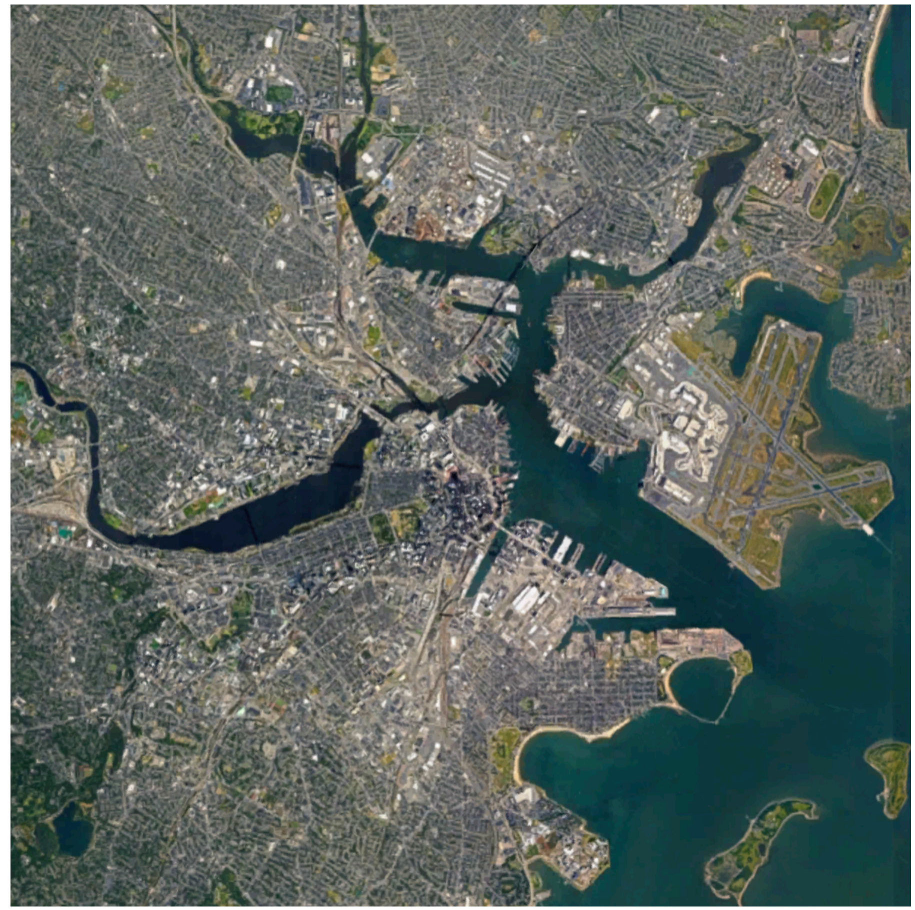
A very *practical* MBTA subway map

A very *realistic* MBTA subway map



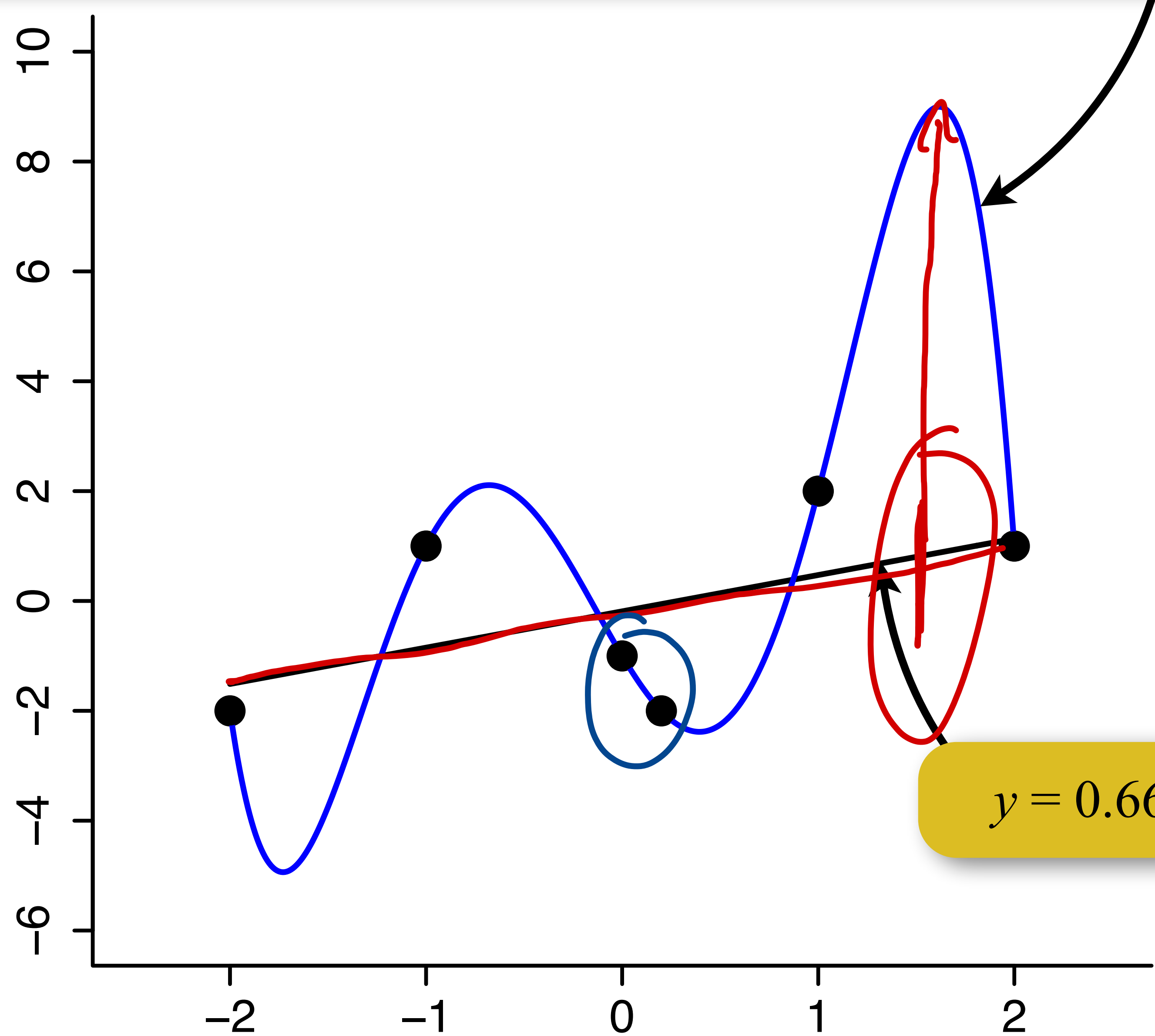
MBTA
MASSACHUSETTS BAY TRANSPORTATION AUTHORITY

Which is more useful?



polynomial

$$y = -1.5972\,x^5 + -0.7917\,x^4 + 8.0694\,x^3 + 3.2917\,x^2 + -5.9722\,x + -1.0$$



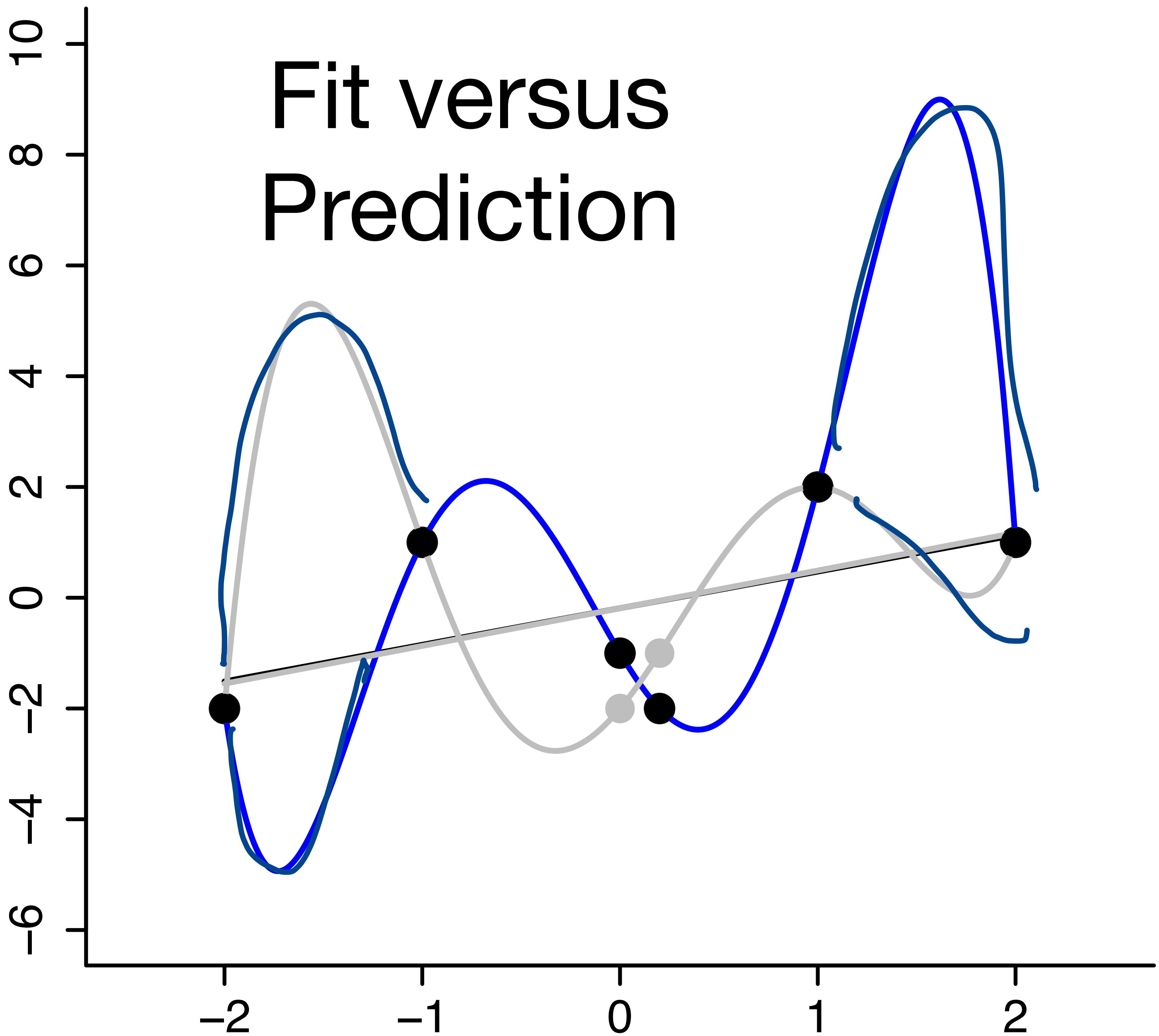
linear

$$y = 0.6611x + -0.1887$$

goodness-of-fit

prediction

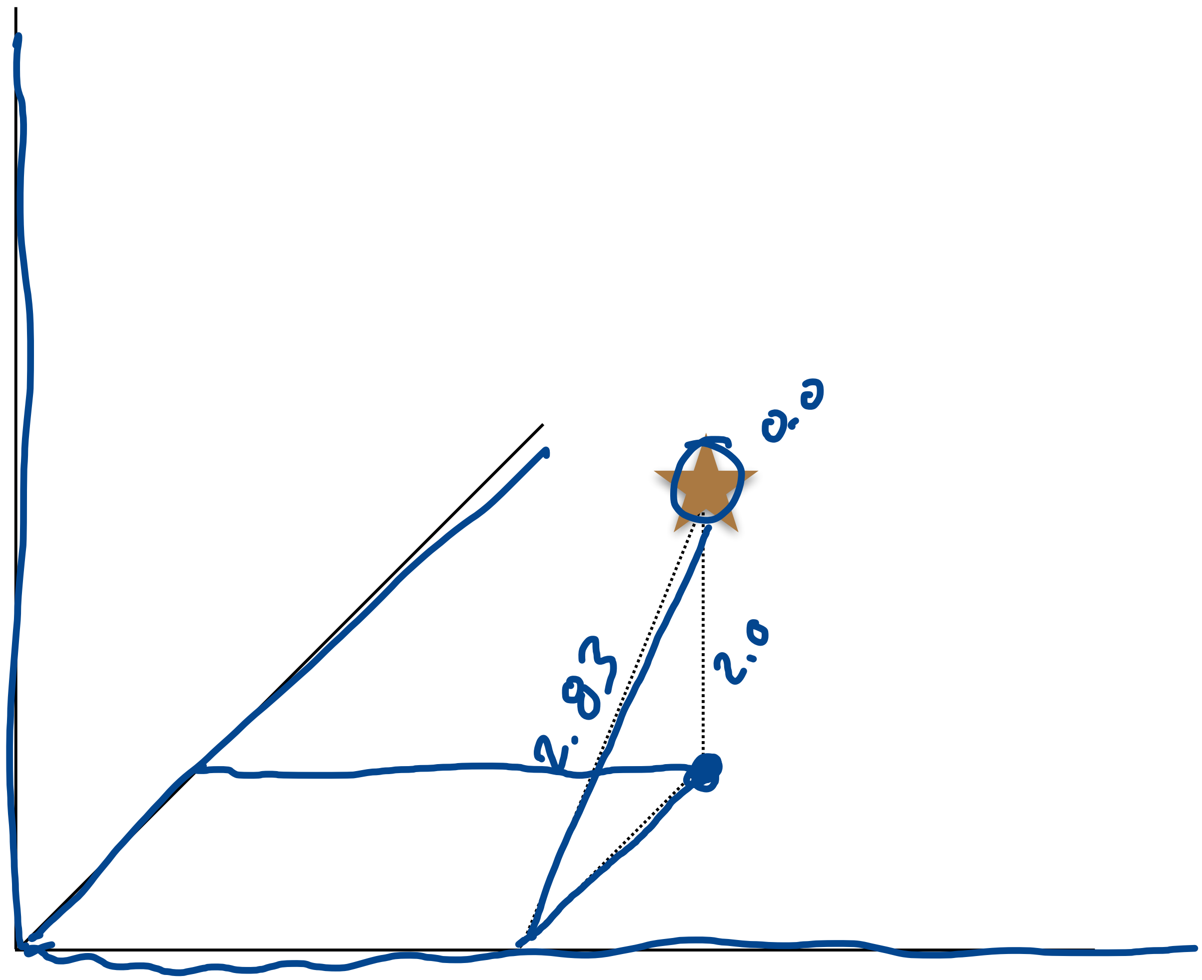
(Thanks to David Swofford for this great example)



complex models

simple models

Model dimensions

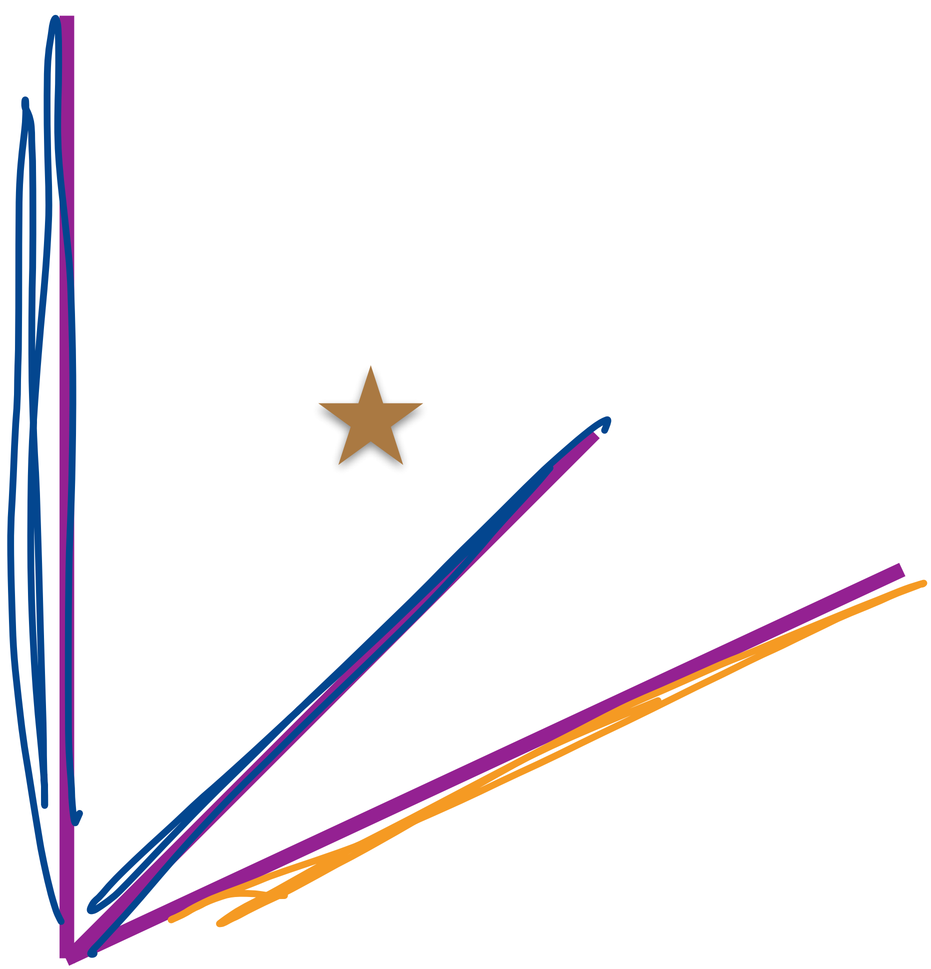


1-parameter model: 2.83

2-parameter model: 2.00

3-parameter model: 0.00

Model dimensions



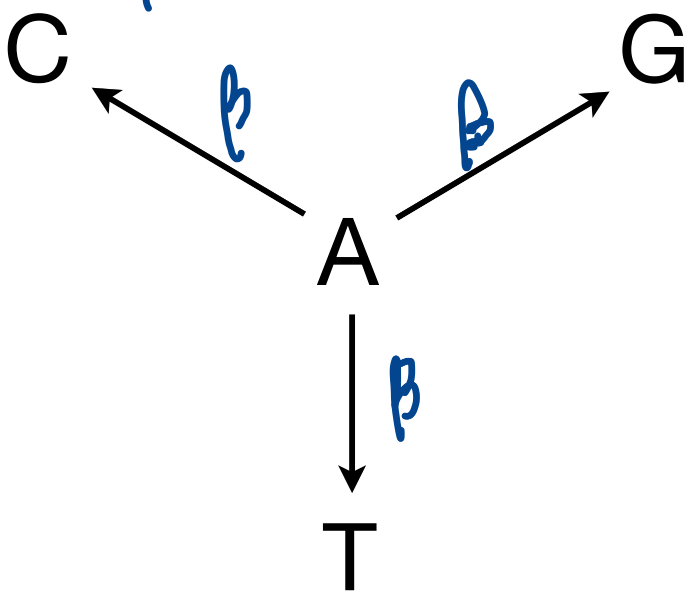
gratuitous complexity
colinearity

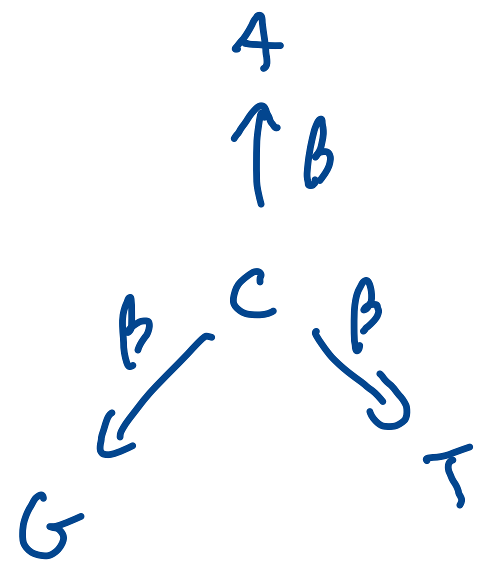
JC69 model

$$\left.\begin{aligned} \pi_A &= \tfrac{1}{4} \\ \pi_G &= \tfrac{1}{4} \\ \pi_C &= \tfrac{1}{4} \\ \pi_T &= \tfrac{1}{4} \end{aligned}\right\}$$

equilibrium relative frequencies

total rate $= 3\beta$





Jukes & Cantor (1969)

number of substitutions
substitution rate
total substitution rate

Edge lengths

$$\text{number} = (\text{rate})(\text{time})$$

$$100 \text{ miles} = \left(50 \frac{\text{miles}}{\cancel{\text{hour}}}\right)\left(2 \cancel{\text{hours}}\right)$$

$$\text{expected no. subst.} = \left(\text{rate of subst.}\right)(\text{time})$$

↑

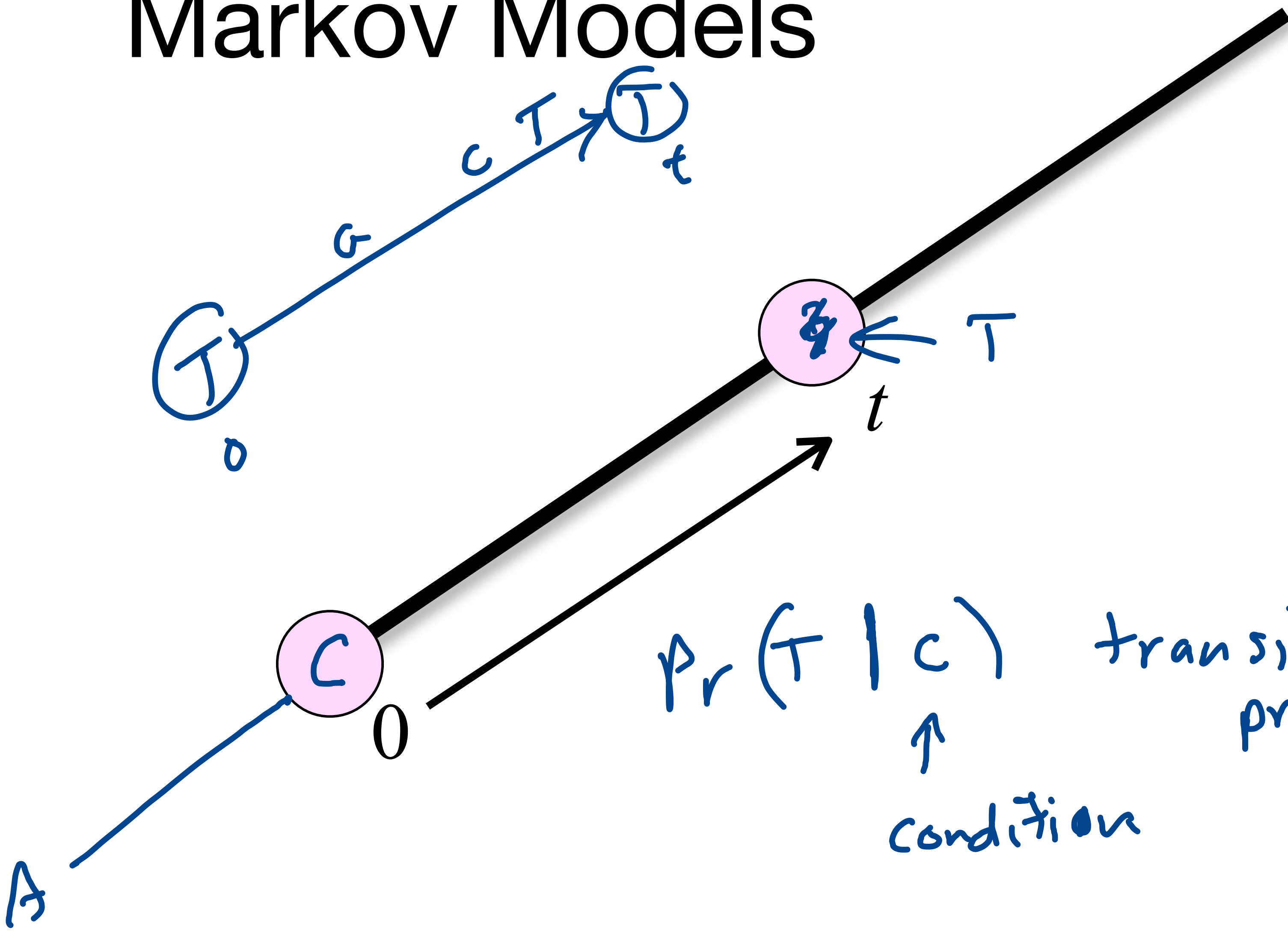
sequence data

$$v = (3\beta)\, t \quad \leftarrow \text{JC69}$$

edge length parameters

long edge lengths means...

Markov Models



$Pr(T \mid c)$ transition probability

condition

Markov property

transition probability

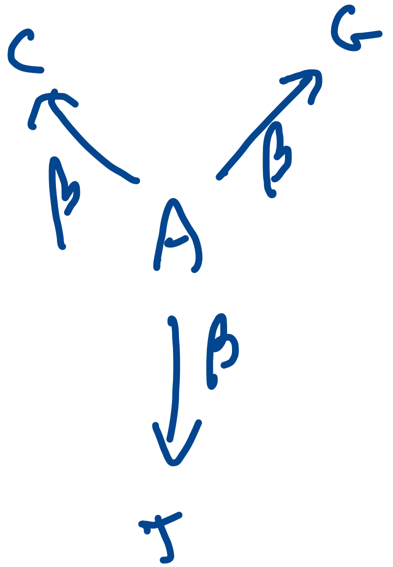
conditional probability

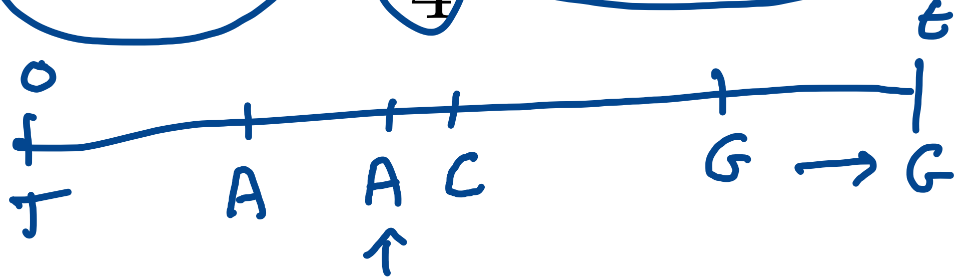
multiple hits

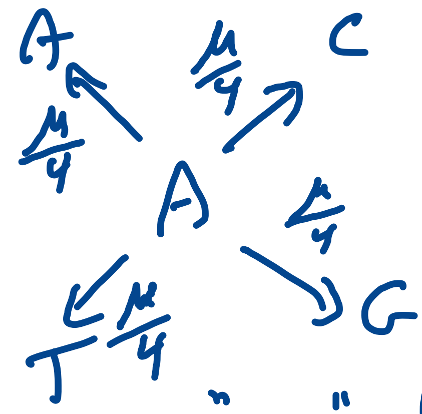
$\beta = \frac{\mu}{4}$

JC69 Transition Probability

$$P(G|T) = P_{TG}(t) = \frac{1}{4}\left(1 - e^{-4\beta t}\right)$$







μ = "tick" rate

→ prob. at least one tick

prob. last tick resulted in G → $\frac{1}{4}$

Poisson $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ $\lambda = \mu t$

$$= \frac{(\mu t)^k e^{-\mu t}}{k!}$$

$0! = 1$

$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

$$p(0) = \frac{(\mu t)^0 e^{-\mu t}}{0!}$$

e = 2.718281828459045...

"perturbation" rate vs. substitution rate

poisson distribution

$e^{-\mu t}$ = prob. no tick marks from 0 to t

$1 - e^{-\mu t}$ = prob. at least one tick

$\frac{\mu}{4} = \beta \quad \mu = 4\beta$

$1 - e^{-4\beta t}$

$$P(G|T) = \frac{1}{4}\left(1 - e^{-4\beta t}\right)$$

JC69 Transition Probability

$$P_{TA}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t}$$

$$P_{TC}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t}$$

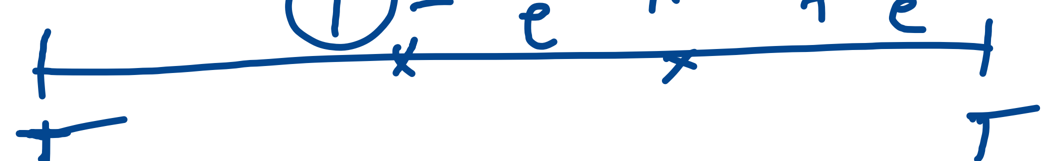
$$P_{TG}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t}$$

$$P_{TT}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t}$$

$$\frac{1}{4} + \frac{3}{4}e^{-4\beta t}$$

$$+ \frac{4}{4}e^{-4\beta t}$$

$$1 - e^{-4\beta t} + e^{-4\beta t}$$

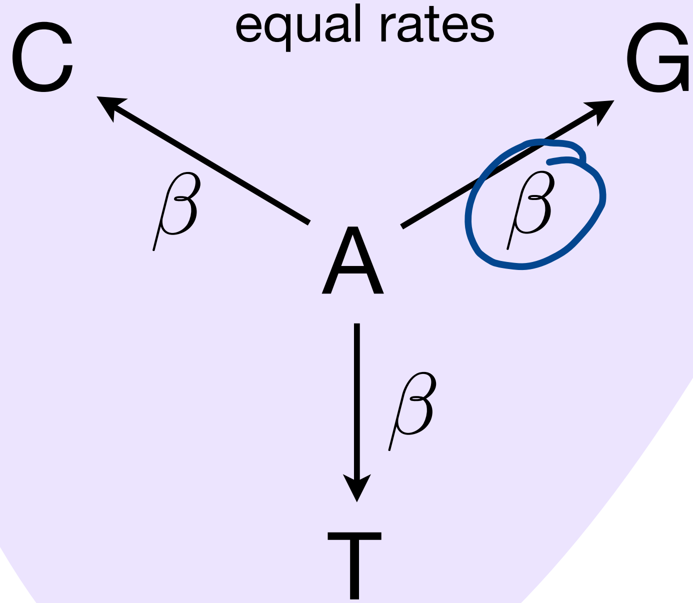


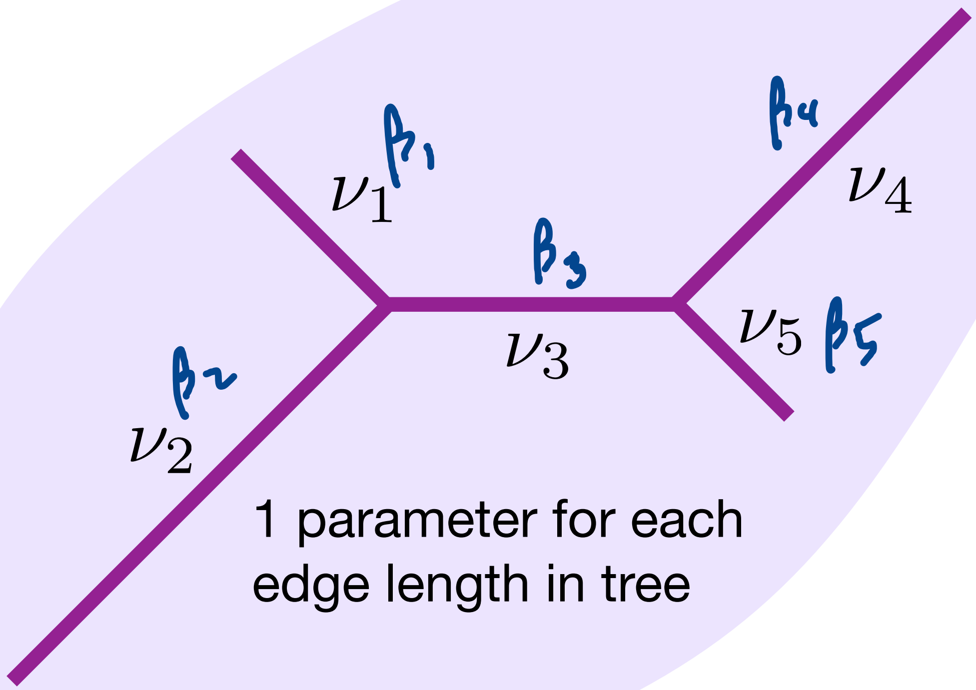
JC69 model assumptions

equal frequencies

$$\pi_A = \pi_C = \pi_G = \pi_T = \frac{1}{4}$$

equal rates

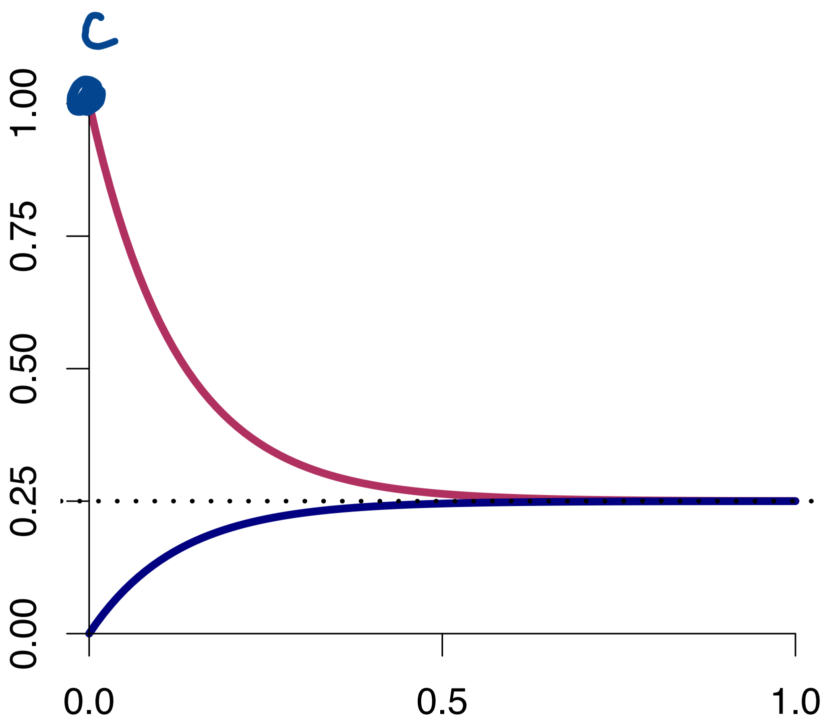


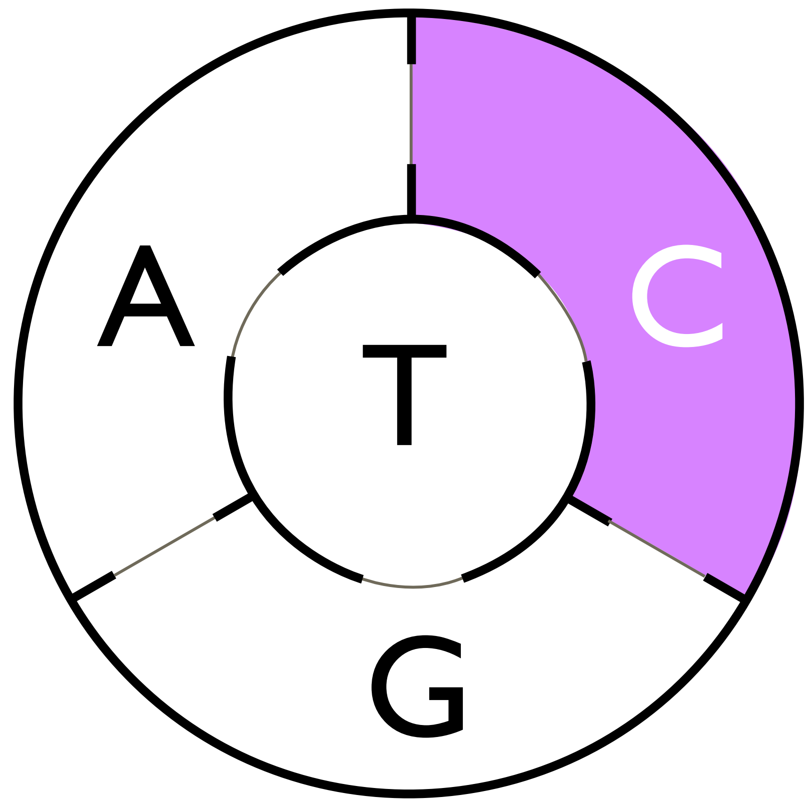


1 parameter for each edge length in tree

Jukes & Cantor (1969)

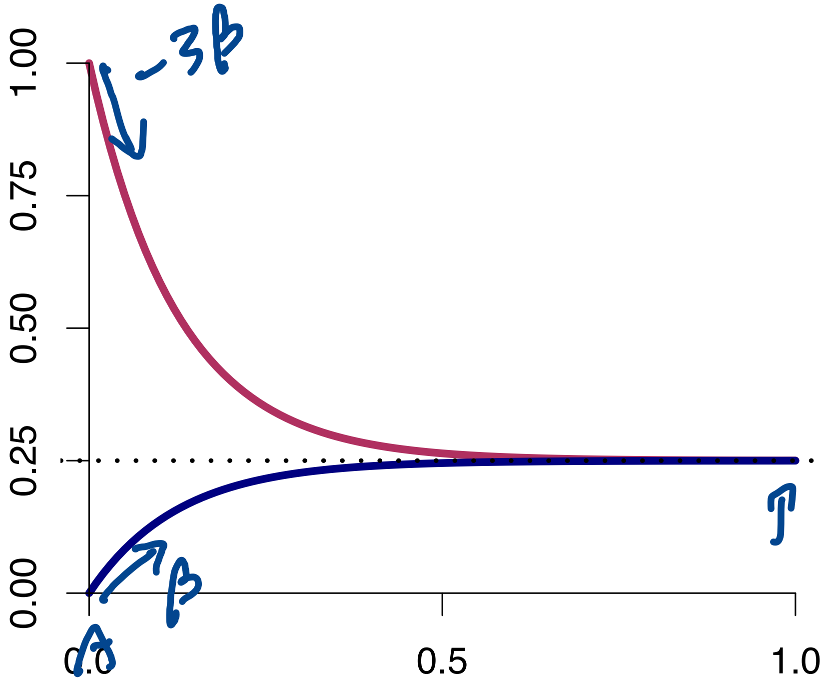
Equilibrium Frequencies

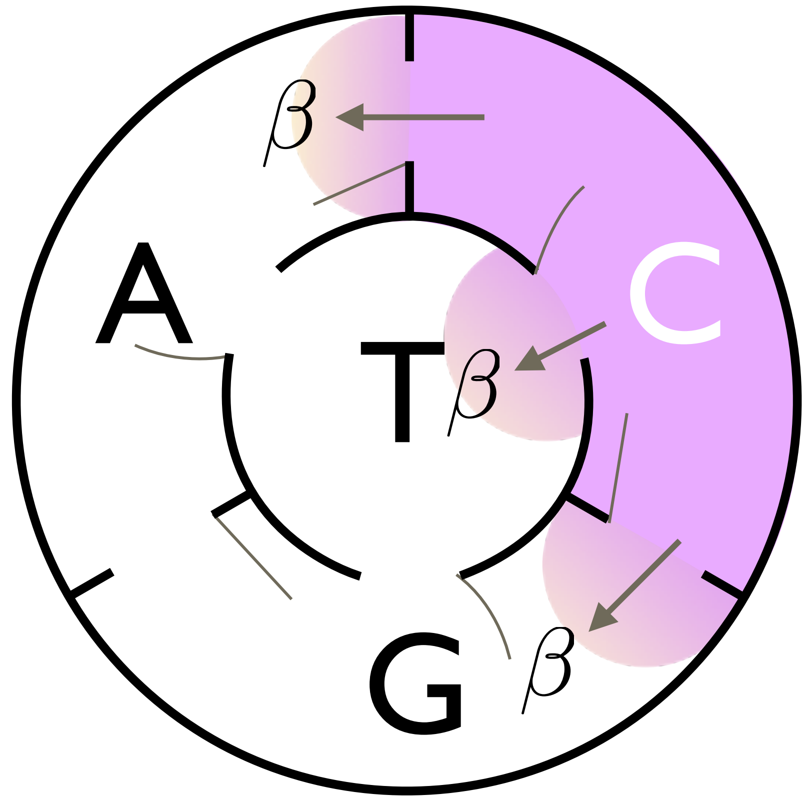




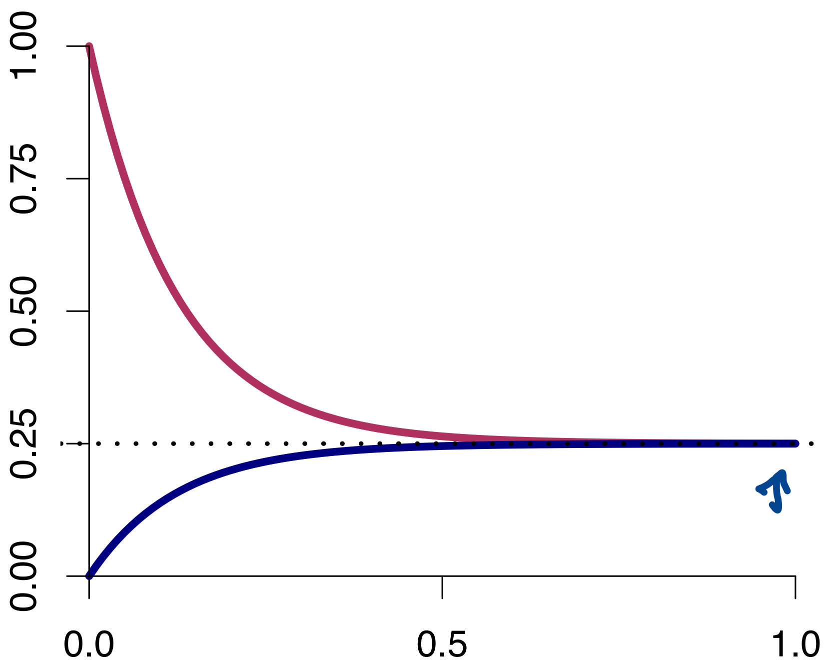
(architect: Joe Bielawski)

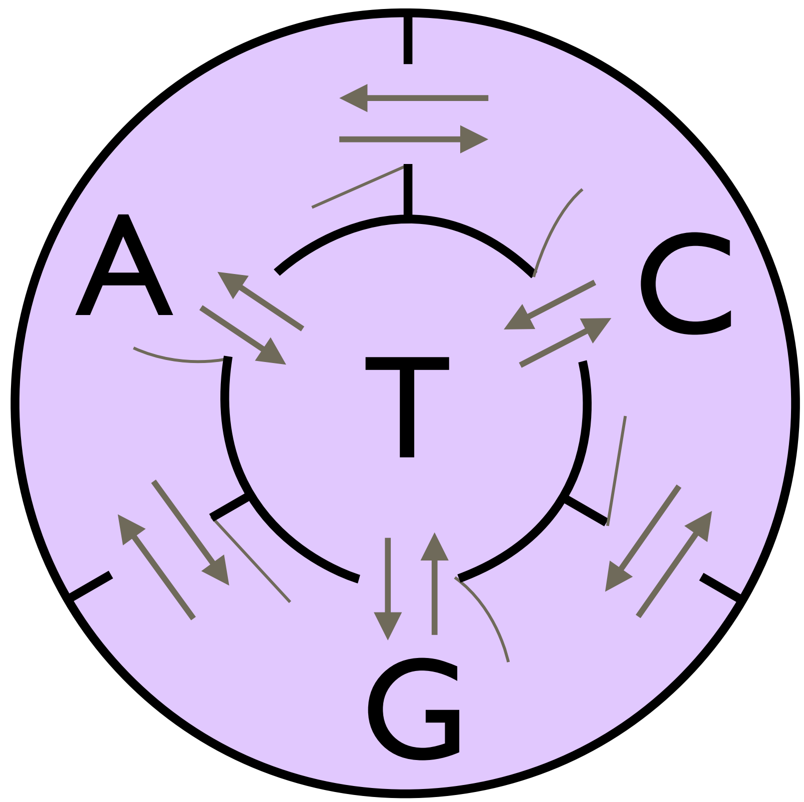
Equilibrium Frequencies





Equilibrium Frequencies





STOPPED HERE 2024-01-30

$$\pi_A = \pi_C = \pi_G = \pi_T = \frac{1}{4}$$

JC69 Distance Formula

$$p = \frac{3}{4}\left(1 - e^{-4\beta t}\right)$$

$$\log e^{a} = a$$

$$e^{\log a} = a$$

$$\frac{4}{3} p = \cancel{\frac{4}{3}}\,\cancel{\frac{3}{4}}\left(1 - e^{-4\beta t}\right)$$

$$0 = \cancel{\frac{4}{3}p - \frac{4}{3}p} = 1 - e^{-4\beta t} - \frac{4}{3}p$$

$$e^{-4\beta t} = 1 - \cancel{e^{-4\beta t}} + \cancel{e^{-4\beta t}} - \frac{4}{3}p$$

$$\cancel{\log}\ \cancel{e}^{-4\beta t} = \log\left(1 - \frac{4}{3}p\right)$$

$$\frac{-4\beta t}{-4} = \frac{\log\left(1 - \frac{4}{3}p\right)}{-4}$$

$$v = -\frac{3}{4}\log\left(1 - \frac{4}{3}p\right)$$

.2326

.2

$$v = 3\beta t = -\frac{3}{4}\log\left(1 - \frac{4}{3}p\right)$$

JC69 rate matrix

1-parameter model

β

$$
\begin{array}{cc}
 & \text{“To” state} \\
 & \begin{array}{cccc} A & C & G & T \end{array} \\
\text{“From” state} \begin{array}{c} A \\ C \\ G \\ T \end{array} &
\begin{pmatrix}
-3\beta & \beta & \beta & \beta \\
\beta & -3\beta & \beta & \beta \\
\beta & \beta & -3\beta & \beta \\
\beta & \beta & \beta & -3\beta
\end{pmatrix}
\end{array}
$$

Jukes & Cantor (1969)

no. parameters

K80 (K2P) rate matrix

2 parameters: α, β

$$
\begin{array}{c|cccc}
 & A & C & G & T \\
\hline
A & -2\beta-\alpha & \beta & \alpha \;\; \kappa\beta & \beta \\
C & \beta & -2\beta-\alpha & \beta & \kappa\beta \;\; \alpha \\
G & \alpha \;\; \kappa\beta & \beta & -2\beta-\alpha & \beta \\
T & \beta & \alpha \;\; \kappa\beta & \beta & -2\beta-\alpha
\end{array}
$$

$\alpha = \beta$
$= JC69$

$\kappa = \alpha/\beta$
kappa

rate of transitions $= \alpha$
rate of transversions $= \beta$
transition/transversion rate ratio α/β
no. parameters
equivalence to JC69

Kimura (1980)

“Transition/transversion ratio” vs. “transition/transversion *rate* ratio”

Possible transitions:

Possible transversions:

$$\frac{\text{E[No. transitions]}}{\text{E[No. transversions]}} = \frac{}{} =$$

F81 rate matrix

→ Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376.

4 parameters

μ, π_A, π_C, π_G

$\pi_T = 1 - \pi_A - \pi_C - \pi_G$

JC69: $\pi_A = \pi_C = \pi_G = \pi_T = \frac{1}{4}$

$$\begin{array}{c|cccc} & A & C & G & T \\ \hline A & - & \pi_C\mu & \pi_G\mu & \pi_T\mu \\ C & \pi_A\mu & - & \pi_G\mu & \pi_T\mu \\ G & \pi_A\mu & \pi_C\mu & - & \pi_T\mu \\ T & \pi_A\mu & \pi_C\mu & \pi_G\mu & - \end{array}$$

$\frac{1}{4}\mu = \beta$

no. parameters

equivalence to JC69

HKY85 rate matrix

5 parameters
$\beta, \alpha, \pi_A, \pi_C, \pi_G$

$$\begin{array}{c|c} & \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{array}{c} A \\ C \\ G \\ T \end{array} & \begin{pmatrix} - & \pi_C \beta & \pi_G \alpha & \pi_T \beta \\ \pi_A \beta & - & \pi_G \beta & \pi_T \alpha \\ \pi_A \alpha & \pi_C \beta & - & \pi_T \beta \\ \pi_A \beta & \pi_C \alpha & \pi_G \beta & - \end{pmatrix} \end{array}$$

Hasegawa, Kishino, & Yano (1985)

no. parameters

equivalence to JC69, F81

F84 vs. HKY85

F84 model:

μ rate of process generating *all types of substitutions*

$k\mu$ rate of process generating *only transitions*

Becomes F81 model if $k = 0$

HKY85 model:

β rate of process generating *only transversions*

$\kappa\beta$ rate of process generating *only transitions*

Becomes F81 model if $\kappa = 1$

F84 first used in Joe Felsenstein's PHYLIP in 1984

F84 published by Kishino & Hasegawa (1989)

GTR rate matrix

9 parameters

μ, π_A, π_C, π_G

a, b, c, d, e

$$
\begin{array}{c}
\\
A \\
C \\
G \\
T
\end{array}
\begin{array}{c}
\begin{array}{cccc} \quad A \quad & \quad C \quad & \quad G \quad & \quad T \quad \end{array} \\
\begin{pmatrix}
- & a\mu\pi_C & b\mu\pi_G & c\mu\pi_T \\
\mu\pi_A a & - & d\mu\pi_G & e\mu\pi_T \\
\mu\pi_A b & \mu\pi_C d & - & f\mu\pi_T \\
\mu\pi_A c & \mu\pi_C e & \mu\pi_G f & -
\end{pmatrix}
\end{array}
$$

Lanave et al. (1984)

no. parameters

equivalence to JC69, F81, K80, HKY

Other kinds of models
(we'll get to some of these later)

- Amino acid substitution models
- Codon models
- Secondary structure models
- Insertion/deletion models
- Relaxed molecular clock models
- Correlated evolution models
- Discrete morphological character models
- Brownian motion models for continuous traits

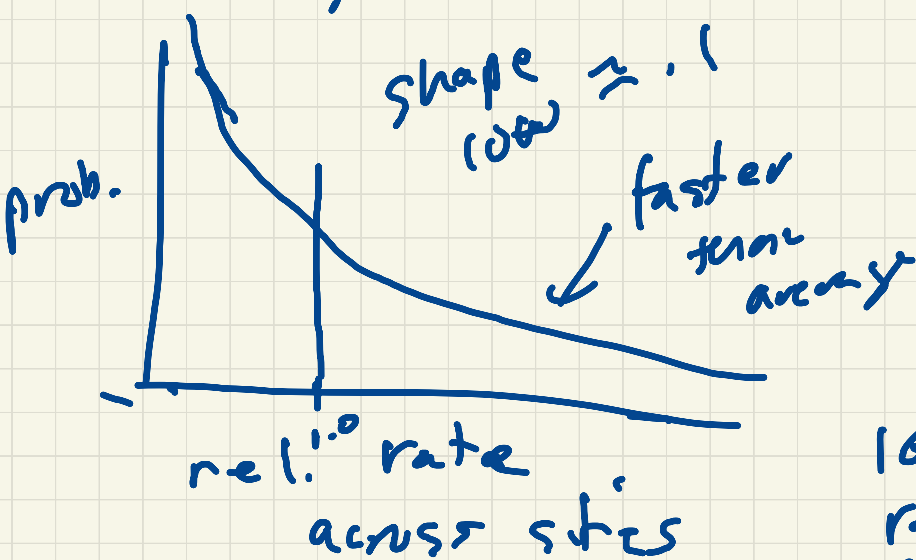
HKY85

I model
G model

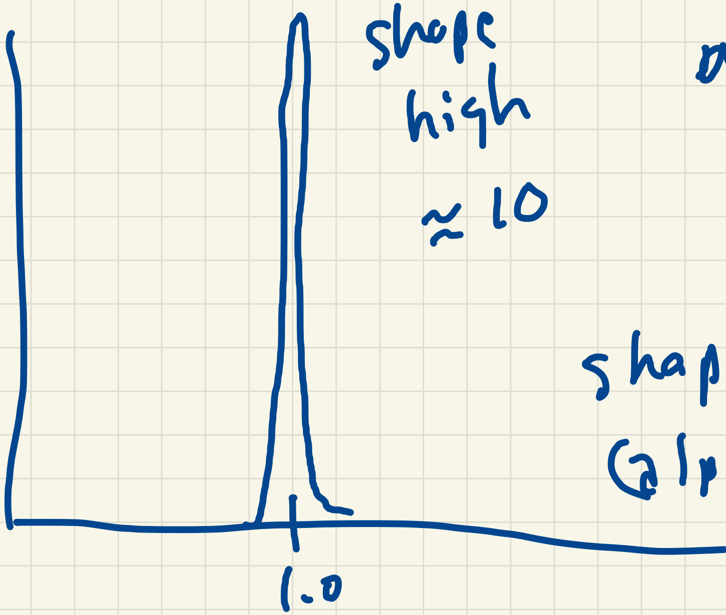
I = invariable sites

Pinvar = proportion of sites that are invariable (rate = 0)

G = gamma model



lots of rate het.



shape high ≈ 10

almost no rate het.

shape parameter (alpha)

---

in lab

K80 rate matrix

trs:trv rate ratio $= \kappa = \dfrac{\alpha}{\beta}$

$$\begin{pmatrix} - & \beta & \kappa\beta & \beta \\ \beta & - & \beta & \kappa\beta \\ \kappa\beta & \beta & - & \beta \\ \beta & \kappa\beta & \beta & - \end{pmatrix}$$

$$\text{trs:trv ratio} = \frac{\text{Expected no. trs.}}{\text{Expected no. trv}}$$

$$= \frac{\kappa\cancel{\beta}\cancel{t}}{2\cancel{\beta}\cancel{t}} = \frac{\kappa}{2}$$ } true for each row of rate matrix

Thus, <u>rate</u> <u>ratio</u> not same as <u>ratio</u>