

Likelihood and models of evolution

Emily Jane McTavish

Life and Environmental Sciences
University of California, Merced

`ejmctavish@ucmerced.edu`, `twitter:snacktavish`

(With thanks to Paul Lewis for slides)

How can we deal with character conflict?

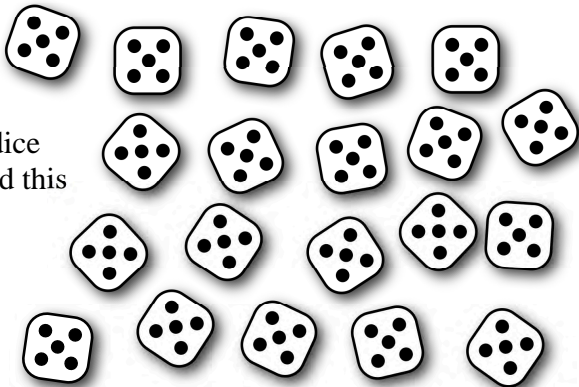
- ▶ We need to apply an error model
- ▶ Likelihood provides a measure of surprise under different models

The Likelihood Criterion

The probability of the observations computed using a model tells us how surprised we should be.

The preferred model is the one that surprises us least.

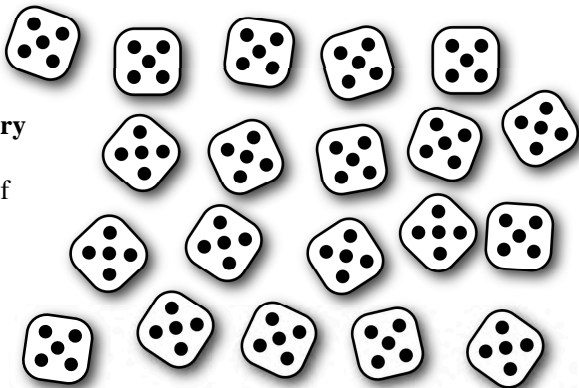
Suppose I threw 20 dice down on the table and this was the result...



The Fair Dice model

$$\Pr(\text{obs.} | \text{fair dice model}) = \left(\frac{1}{6}\right)^{20} = \frac{1}{3,656,158,440,062,976}$$

You should have been **very surprised** at this result because the probability of this event is **very small**: only 1 in 3.6 quadrillion!

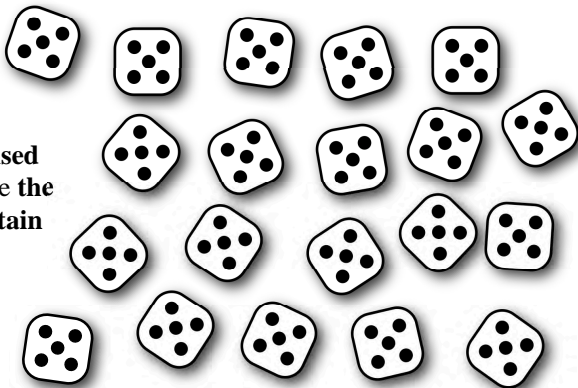


The Trick Dice model

(assumes dice each have 5 on every side)

$$\Pr(\text{obs.} | \text{trick dice model}) = 1^{20} = 1$$

You should **not be surprised at all** at this result because **the observed outcome is certain** under this model



Results

Model	Likelihood	Surprise level
Fair Dice	$\frac{1}{3,656,158,440,062,976}$	Very, <i>very</i> , very surprised
Trick Dice	1	Not surprised at all

winning model maximizes likelihood
(and thus minimizes surprise)

Likelihood: why a new term?

Outcome	Fair coin model	Two-heads model
H	0.5	1
T	0.5	0
	1	1

Likelihoods of models given one particular data outcome are *not* expected to sum to 1.0

Probabilities of data outcomes given one particular model sum to 1.0

Likelihood and model comparison

- Analyses using likelihoods ultimately involve **model comparison**
- The models compared can be **discrete** (as in the fair vs. trick dice example)
- More often the models compared differ **continuously**:
 - Model 1: branch length is 0.01
 - Model 2: branch length is 0.02
 - Model 3: branch length is 0.03

Rather than having an infinity of models, we instead think of the branch length as a **parameter** within one model

Likelihood calculated from a single sequence

$$\Pr(A) = \pi_A$$

$$\Pr(C) = \pi_C$$

$$\Pr(G) = \pi_G$$

$$\Pr(T) = \pi_T$$

First 32 nucleotides of the $\psi\eta$ -globin gene of gorilla:

GAAGTCCTTGAGAAATAAACTGCACACACTGG

$$\begin{aligned} L &= \pi_G \pi_A \pi_A \pi_G \pi_T \pi_C \pi_C \pi_T \pi_T \pi_G \pi_A \pi_G \pi_A \pi_A \pi_A \pi_T \pi_A \pi_A \pi_A \pi_C \pi_T \pi_G \pi_C \pi_A \pi_C \pi_A \pi_C \pi_A \pi_C \pi_T \pi_G \pi_G \\ &= \pi_A^{12} \pi_C^7 \pi_G^7 \pi_T^6 \end{aligned}$$

Note that we are assuming independence among sites here

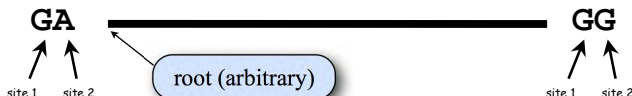
$$\log L = 12 \log(\pi_A) + 7 \log(\pi_C) + 7 \log(\pi_G) + 6 \log(\pi_T)$$

We can already see by eye-balling this that a model allowing **unequal** base frequencies will **fit better** than a model that assumes **equal** base frequencies because there are about twice as many As as there are Cs, Gs and Ts.

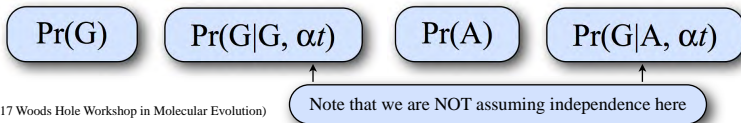
Likelihood of the simplest tree

sequence 1 sequence 2

To keep things simple, assume that the sequences are only 2 nucleotides long:



$$\begin{aligned}
 L &= L_1 L_2 \\
 &= \left[\begin{pmatrix} 1 \\ 4 \end{pmatrix} \left(\frac{1}{4} + \frac{3}{4} e^{-4\alpha t} \right) \right] \left[\begin{pmatrix} 1 \\ 4 \end{pmatrix} \left(\frac{1}{4} - \frac{1}{4} e^{-4\alpha t} \right) \right]
 \end{aligned}$$



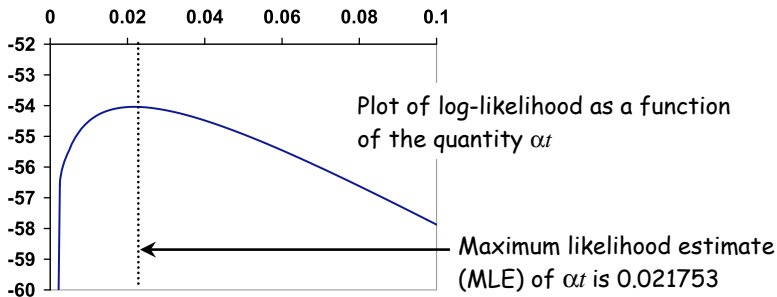
Maximum likelihood estimation

First 32 nucleotides of the $\psi\eta$ -globin gene of gorilla and orangutan:

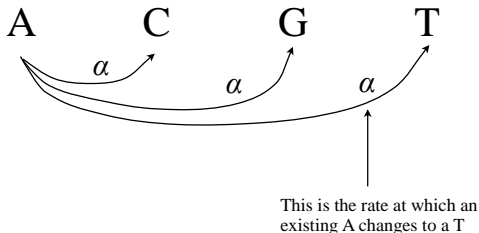
gorilla **G****A****A**GTCCTTGAGAAATAAACTGCACACACTGG

orangutan **G****G****A**CTCCTTGAGAAATAAACTGCACACACTGG

$$L = \left[\left(\frac{1}{4} \right) \left(\frac{1}{4} + \frac{3}{4} e^{-4\alpha t} \right) \right]^{30} \left[\left(\frac{1}{4} \right) \left(\frac{1}{4} - \frac{1}{4} e^{-4\alpha t} \right) \right]^2$$



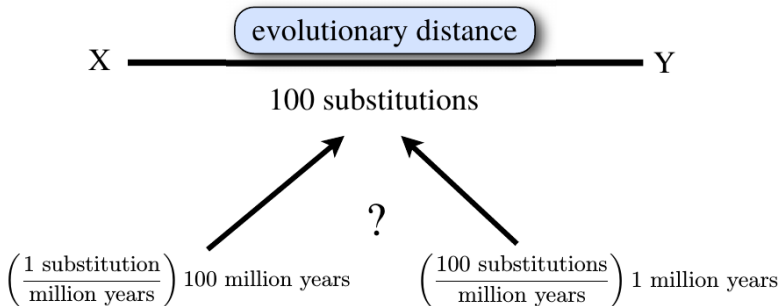
number of substitutions = rate \times time



Overall substitution rate is 3α , so the expected number of substitutions (v) is

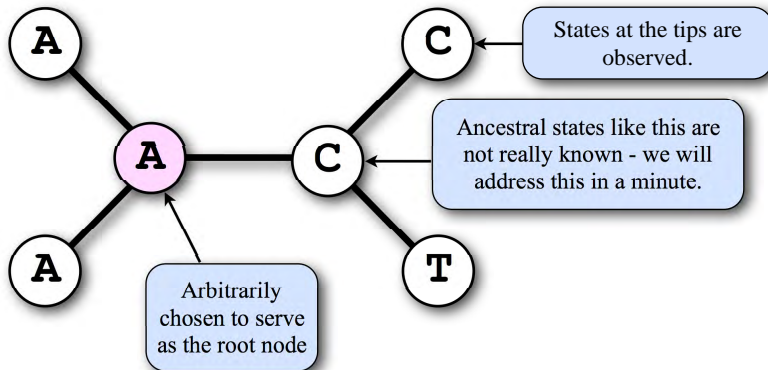
$$v = 3\alpha t$$

Rate and time are confounded

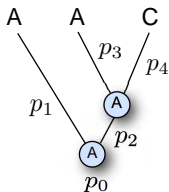


Likelihood of an unrooted tree

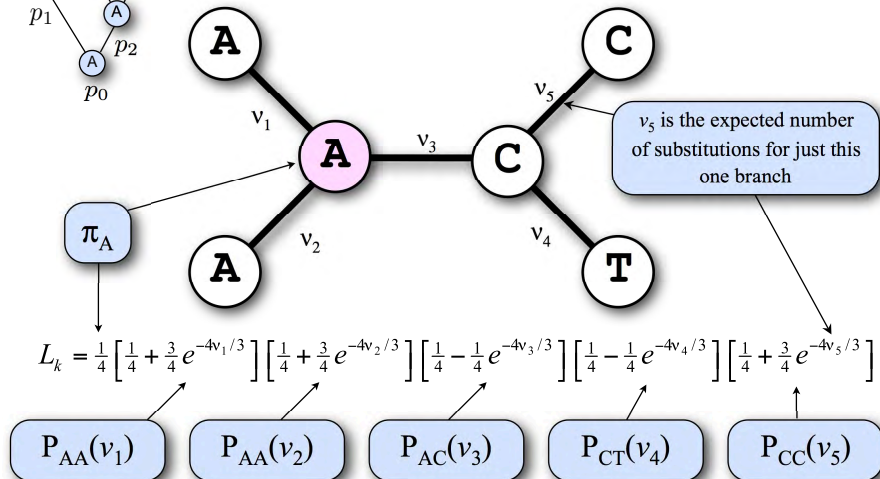
(data shown for only one site)



From slide 6



Likelihood for site k

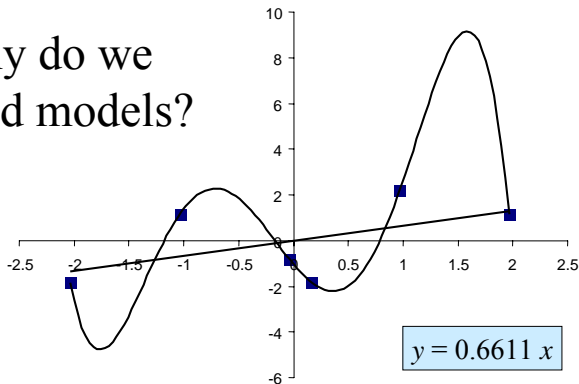


Note use of the AND probability rule

Substitution Models

$$y = -1.5972 x^5 + 23.167 x^4 - 126.18 x^3 + 319.17 x^2 - 369.22 x + 155.67$$

Why do we
need models?



Models

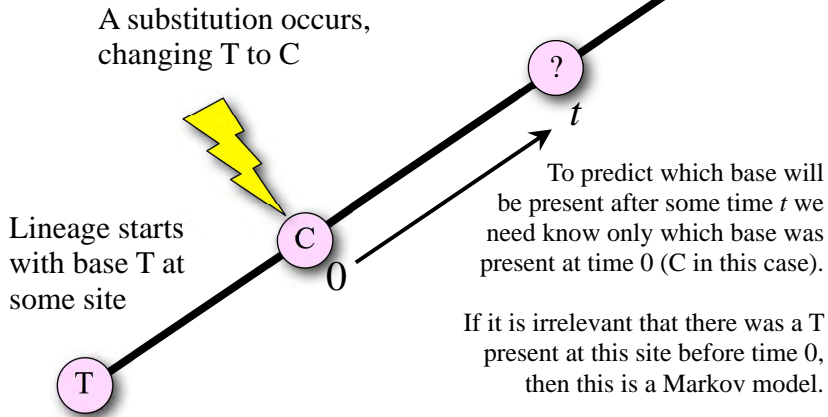
- Models help us intelligently **interpolate between our observations** for purposes of **making predictions**
- **Adding parameters** to a model generally increases its fit to the data
- **Underparameterized** models lead to poor fit to observed data points
- **Overparameterized** models lead to poor prediction of future observations
- Criteria for choosing models include likelihood ratio tests, AIC, BIC, Bayes Factors, etc.
 - all provide a way to choose a model that is neither underparameterized nor overparameterized

Jukes-Cantor (JC69) model

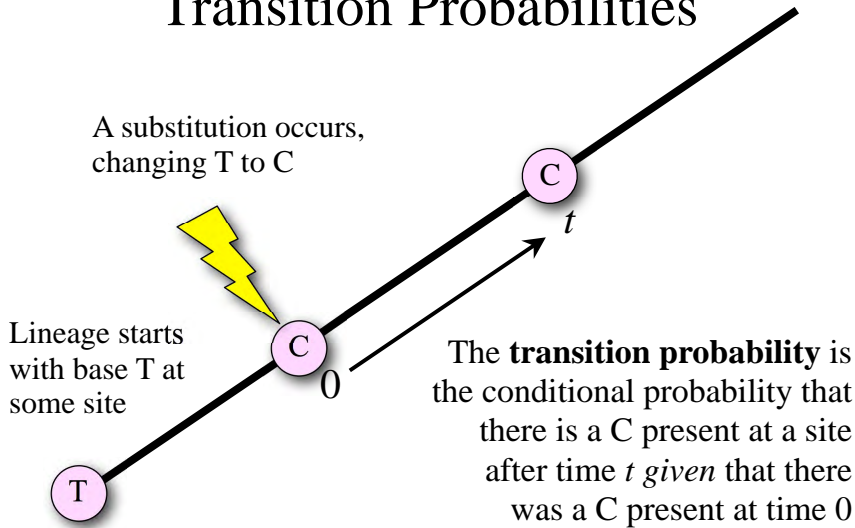
- The four bases (A, C, G, T) are expected to be **equally frequent** in sequences ($\pi_A = \pi_C = \pi_G = \pi_T = 0.25$)
- Assumes **same rate** for all types of substitution
($r_{A \leftrightarrow C} = r_{A \leftrightarrow G} = r_{A \leftrightarrow T} = r_{C \leftrightarrow G} = r_{C \leftrightarrow T} = r_{G \leftrightarrow T} = \alpha$)
- Usually described as a **1-parameter** model (the parameter being the edge length)
 - Remember, however, that each edge in a tree can have its own length, so there are really as many parameters in the model as there are edges in the tree!
- Assumes substitution is a **Markov** process...

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21-132 in H. N. Munro (ed.), *Mammalian Protein Metabolism*. Academic Press, New York.

What is a Markov process?



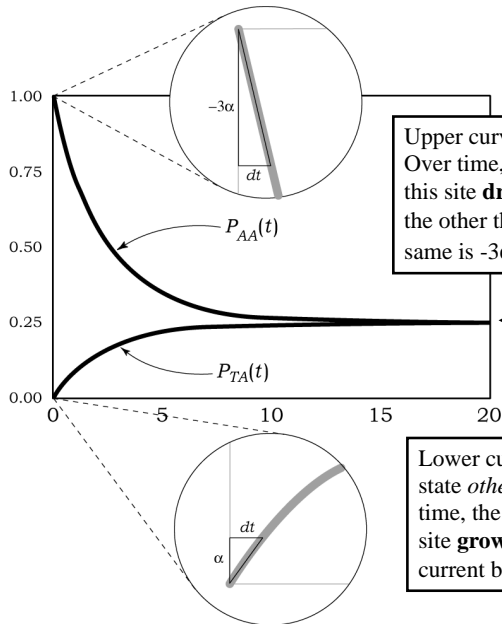
Transition Probabilities



Equilibrium frequencies

- The JC69 model assumes that the frequencies of the four bases (A, C, G, T) are equal
- The equilibrium relative frequency of each base is thus 0.25
- Why are they called *equilibrium* frequencies?

Pr(A|A) and Pr(A|T) as a function of time



Upper curve assumes we started with A at time 0. Over time, the probability of still seeing an A at this site **drops** because rate of changing to one of the other three bases is 3α (so rate of staying the same is -3α).

The equilibrium relative frequency of A is 0.25

Lower curve assumes we started with some state *other* than A (T is used here). Over time, the probability of seeing an A at this site **grows** because the rate at which the current base will change into an A is α .