Maximum Likelihood

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

September 24, 2009

Principle of Maximum Likelihood

- Given parameters θ and data X
- The function $f(X \mid \theta)$ is the probability of observing data X given parameter θ. (Both X and θ can be multi-dimensional.)
- Keeping θ fixed, and treating f as a function of X, the total probability is one.

Principle of Maximum Likelihood

- The function $L(\theta) = f(X \mid \theta)$ with X fixed and θ unknown is called the *likelihood function*.
- The *principle of maximum likelihood* is to estimate θ with the value $\hat{\theta}$ that maximizes $L(\theta)$.
- In practice, it is common to maximize the log-likelihood, $\ell(\theta) = \ln L(\theta)$.
- This is because X often takes the form of an independent sample so that

$$L(X) = \prod_{i=1}^{n} f(X_i \mid \theta), \qquad \ell(\theta) = \sum_{i=1}^{n} \ln f(X_i \mid \theta)$$

Coin-tossing Example

- A coin has a probability θ of being a head.
- Consider tossing the coin 100 times. The probability of each single sequence with exactly x heads is $f(x \mid \theta) = p^x(1-p)^{100-x}$.
- Say we observe the sequence

$$HHTHTHHT \ldots TTH$$

where heads appear 57 times.
- The maximum likelihood estimate is the value $\hat{\theta}$ that maximizes the function

$$L(\theta) = \theta^{57}(1-\theta)^{43},$$

or, equivalently that maximizes

$$\ell(\theta) = 57(\ln \theta) + 43(\ln(1-\theta)) \ .$$

Simple calculus and common sense lead to the estimate $\hat{\theta} = 0.57$.

Maximum-likelihood edge lengths

- For the Jukes-Cantor model, a pair of sequences have x sites with observed differences and $n - x$ sites with the same base.
- The probability of any given sequence pair is

$$L(d) = \left(\frac{1}{4}\right)^n \times \left(\frac{1}{4} - \frac{1}{4}\mathrm{e}^{-\frac{4}{3}d}\right)^x \times \left(\frac{1}{4} + \frac{3}{4}\mathrm{e}^{-\frac{4}{3}d}\right)^{n-x}$$

which has the form

$$L(\theta) = C \times \theta^x (1 - 3\theta)^{n-x}$$

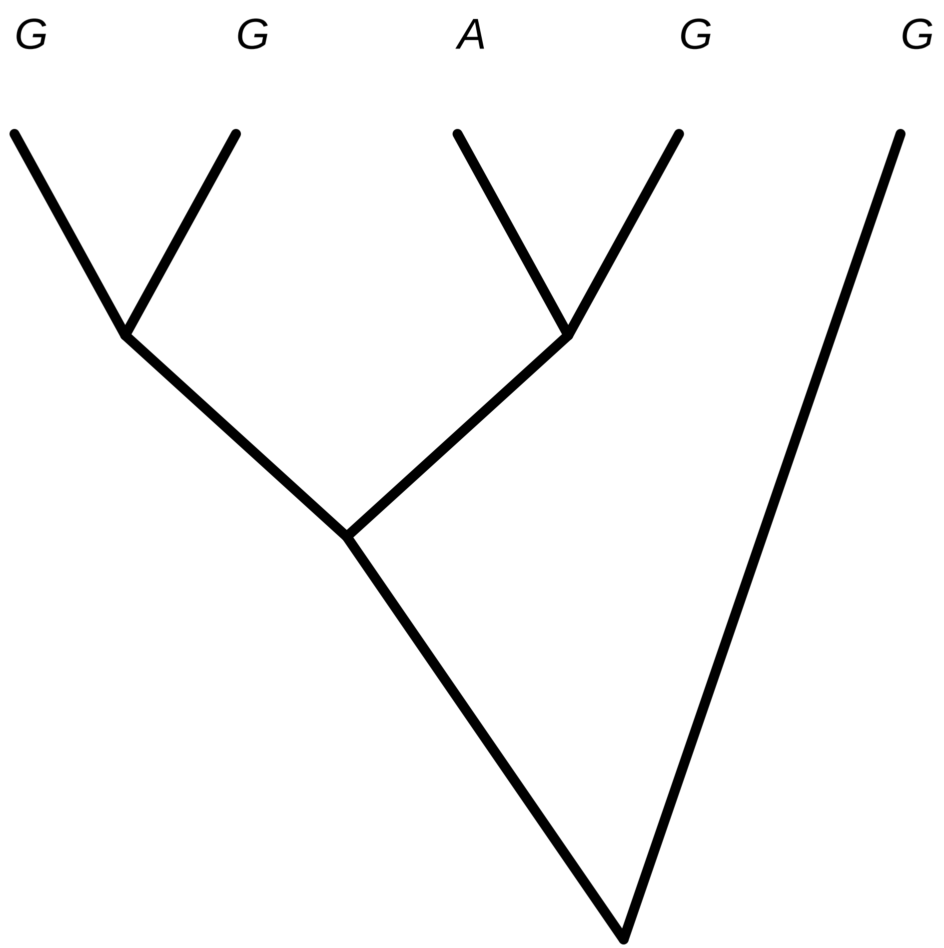
where

$$\theta = \frac{1}{4} - \frac{1}{4}\mathrm{e}^{-\frac{4}{3}d} \; .$$

- Solving the calculus problem yields $\hat{\theta} = \frac{x}{3n}$.
- Plugging in and solving for d gives

$$\hat{d} = -\frac{3}{4} \ln \left(1 - \frac{4}{3}\frac{x}{n}\right)$$

Computing Likelihood on a Tree



Transition Probabilities

$$P(0.1) = \begin{bmatrix} 0.90 & 0.04 & 0.04 & 0.03 \\ 0.03 & 0.91 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.91 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.90 \end{bmatrix} \qquad P(0.2) = \begin{bmatrix} 0.81 & 0.07 & 0.07 & 0.05 \\ 0.05 & 0.83 & 0.07 & 0.05 \\ 0.05 & 0.07 & 0.83 & 0.05 \\ 0.05 & 0.07 & 0.07 & 0.81 \end{bmatrix}$$

$$P(0.4) = \begin{bmatrix} 0.67 & 0.13 & 0.13 & 0.08 \\ 0.08 & 0.71 & 0.13 & 0.08 \\ 0.08 & 0.13 & 0.71 & 0.08 \\ 0.08 & 0.13 & 0.13 & 0.67 \end{bmatrix}$$

Model Selection

12 *rbcL* genes from 12 plant species

Model	p	ℓ
JC69	21	-6262.01
K80	22	-6113.86
HKY85	25	-6101.76
HKY85 + Γ_5	26	-5764.26
HKY85 + C	35	-5624.70