

Parsimony & Probability

Under what circumstances is the character vector [011100] more probable given tree A than given tree B?

I.e., under what circumstances is tree A more likely than tree B given [011100]?

Parsimony & Probability

- P[change] is the same on each branch;
 - Branch length unimportant;
 - No rate shifts on tree;
 - Other characters do not affect probability of change;
 - P[gain] = P[loss].
- Only a single ancestral reconstruction is considered per node.

Parsimony & Probability

Tree A requires only one change.

Parsimony & Probability

The probability of the character vector is:

$$P[\text{change}]^{\text{changes}} \times (1-P[\text{change}])^{\text{static branches}}$$

Log-likelihood of tree is:

$$\text{changes} \times \ln(P[\text{change}]) + \text{statics} \times \ln(1-P[\text{change}])$$

Parsimony & Probability

If P[change] = 0.1, then:

$$P[\text{character} \mid \text{tree}] = 0.1^1 \times 0.9^9 = \mathbf{3.87 \times 10^{-2}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(0.1) + (9 \times \ln[0.9]) = \mathbf{-3.25}$$

Parsimony & Probability

If P[change] = 0.1, then:

$$P[\text{character} \mid \text{tree}] = 0.1^2 \times 0.9^8 = \mathbf{4.30 \times 10^{-3}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[0.1]) + (8 \times \ln[0.9]) = \mathbf{-5.45}$$

Parsimony & Probability

If P[change] = 0.01, then:

$$P[\text{character} \mid \text{tree}] = 0.01^1 \times 0.99^9 = \mathbf{9.14 \times 10^{-3}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(0.01) + (9 \times \ln[0.99]) = \mathbf{-4.70}$$

Parsimony & Probability

If P[change] = 0.01, then:

$$P[\text{character} \mid \text{tree}] = 0.01^2 \times 0.99^8 = \mathbf{9.23 \times 10^{-5}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[0.01]) + (8 \times \ln[0.99]) = \mathbf{-9.29}$$

Parsimony & Probability

If $P[\text{change}] = 0.001$, then:

$$P[\text{character} \mid \text{tree}] = 0.001^1 \times 0.999^9 = \mathbf{9.91 \times 10^{-4}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(10^{-3}) + (9 \times \ln[0.999]) = \mathbf{-6.91}$$

Parsimony & Probability

If $P[\text{change}] = 0.001$, then:

$$P[\text{character} \mid \text{tree}] = 0.001^2 \times 0.999^8 = \mathbf{9.92 \times 10^{-7}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[10^{-3}]) + (8 \times \ln[0.999]) = \mathbf{-13.82}$$

Infinity and beyond.....

| $P[\text{change}]$ | $\ln L[\text{tree A}]$ | $\ln L[\text{tree B}]$ | Difference |
|--------------------|------------------------|------------------------|-------------|
| 10^{-1} | -3.25 | -5.45 | 2.20 |
| 10^{-2} | -4.70 | -9.29 | 4.60 |
| 10^{-3} | -6.92 | -13.82 | 6.91 |
| $10^{-\infty}$ | $-\infty$ | $-2 \times \infty$ | ∞ |

Shorter tree is more likely while $P[\text{change}] < 0.5$

| $P[\text{change}]$ | $\ln L[\text{tree A}]$ | $\ln L[\text{tree B}]$ | Difference |
|--------------------|------------------------|------------------------|--------------|
| 0.2 | -3.62 | -5.00 | 1.62 |
| 0.4 | -5.51 | -5.92 | 0.41 |
| 0.5 | -6.93 | -6.93 | 0.00 |
| 0.6 | -8.76 | -8.35 | -0.41 |

Shorter tree is more likely while $P[\text{change}] < 0.5$

| $P[\text{change}]$ | $\ln L[\text{tree A}]$ | $\ln L[\text{tree B}]$ | Difference |
|--------------------|------------------------|------------------------|-------------|
| 0.5 | -6.93 | -6.93 | 0.00 |

Shift does not occur at $P[\text{change}] > 0.15$ because only a single way of generating one or two changes is considered.

Relaxing assumptions of parsimony

- Low vs. high rates of change.
- Homogeneous vs. heterogeneous rates.
- Unit vs. variable branch lengths.
- Certain vs. uncertainty in ancestral reconstructions.
- Independent vs. correlated character change.

Likelihood & phylogeny: some important concepts

- Conditional probability: $P[X | Z]$ when X is affected by “intermediate” parameter Y:

$$P[X | Z] = \sum P[X | Y] \times P[Y | Z]$$
- Parameters: some variables can have multiple parameters (e.g., chars. 1-10 have rate π_i whereas chars. 11-20 have rate π_k)
 - Increasing number of differing parameters will increase likelihoods;
- Likelihood Ratio tests: used to evaluate whether additional parameters are justified
 - χ^2 distribution if *hierarchical*.
 - Information theory criteria if non-hierarchical.

Advantages of Likelihood

- Allows assumptions of parsimony (or other likelihood) analyses to be dissected;
 - Allows testing of parameters other than phylogeny;
 - Assumptions plainly stated.
- Allows different data types (e.g., morphology / stratigraphy / molecules) to weigh in against hypothesis.

How Likelihood can Reject General Phylogenetic Hypotheses

Disadvantages of Likelihood

- Much slower than parsimony analyses!
- Some hold that parameters assumed to be important in likelihood analyses (e.g., rates, branch lengths, etc.), should be discovered by learning phylogeny, not part of the test for phylogeny.

Effect of Branch Lengths: Felsenstein 1973

- Given a rate π and a branch duration of time b , the expected number of changes is πb .
 - Probability of Δ changes modeled as a Poisson process (i.e., change can occur at any time).
- $$P[\Delta | \pi b] = \frac{(\pi b)^\Delta \times e^{-(\pi b)}}{\Delta!}$$

Effect of Branch Lengths: Example

$$\begin{aligned}
 &L[\tau, \pi=0.95 | \text{char}] \\
 &= P[0 \rightarrow 2 | b=0.96] = ([0.95 \times 0.96]^2 \times e^{-(0.95 \times 0.96)})/2! \\
 &\times P[0 \rightarrow 0 | b=0.96] = e^{-(0.95 \times 0.96)} \\
 &\times P[0 \rightarrow 0 | b=0.14] = e^{-(0.95 \times 0.14)} \\
 &\times P[0 \rightarrow 2 | b=1.10] = ([0.95 \times 1.10]^2 \times e^{-(0.95 \times 1.10)})/2! \\
 &\times P[0 \rightarrow 0 | b=1.10] = e^{-(0.95 \times 1.10)} \\
 &= 3.97 \times 10^{-3}
 \end{aligned}$$

Effect of Branch Lengths: Example

$$\begin{aligned}
 &\ln L[\tau, \pi=0.95 | \text{char}] \\
 &= \ln P[0 \rightarrow 2 | b=0.96] = (2 \times \ln[0.95 \times 0.96]) - (0.95 \times 0.96) - \ln(2) \\
 &+ \ln P[0 \rightarrow 0 | b=0.96] = -(0.95 \times 0.96) \\
 &+ \ln P[0 \rightarrow 0 | b=0.14] = -(0.95 \times 0.14) \\
 &+ \ln P[0 \rightarrow 2 | b=1.10] = (2 \times \ln[0.95 \times 1.10]) - (0.95 \times 1.10) - \ln(2) \\
 &+ \ln P[0 \rightarrow 0 | b=1.10] = -(0.95 \times 1.10) \\
 &= -5.53
 \end{aligned}$$

Effect of Branch Lengths: Example

$$\begin{aligned}
 &L[\tau, \pi=0.95 | \text{char}] \\
 &= P[0 \rightarrow 2 | b=0.96] = ([0.95 \times 0.96]^2 \times e^{-(0.95 \times 0.96)})/2! \\
 &\times P[0 \rightarrow 0 | b=0.96] = e^{-(0.95 \times 0.96)} \\
 &\times P[0 \rightarrow 0 | b=0.14] = e^{-(0.95 \times 0.14)} \\
 &\times P[0 \rightarrow 2 | b=1.10] = ([0.95 \times 1.10]^2 \times e^{-(0.95 \times 1.10)})/2! \\
 &\times P[0 \rightarrow 0 | b=1.10] = e^{-(0.95 \times 1.10)} \\
 &= e^{-(0.95 \times 4.26)} \times [0.95 \times 0.96]^2 \times [0.95 \times 1.10]^2 / (2! \times 2!)
 \end{aligned}$$

Tree Likelihood Rephrased

- $e^{-(0.95 \times 4.26)} \times [0.95 \times 0.96]^2 \times [0.95 \times 1.10]^2 / (2! \times 2!)$
- $e^{-(\text{rate} \times \sum \text{branches durations})}$
 - $\times \prod [\text{rate} \times \text{branch durations}]^{\text{changes}_i} \div \prod \text{changes}_i!$ for all branches showing change in character.
- Log-likelihood there is just:
 - Rate $\times \sum \text{static branch durations}$
 - + $\sum \text{changes} \times \ln (\text{rate} \times \text{branch duration})$
 - $\ln (\text{changes}!)$
 - for all branches showing change in the character.
- Can it be this easy???

What is the likelihood of the 2nd nodes states?

$$\begin{aligned}
 & \underline{L[\text{node } 2 = 0 \mid \text{taxa } 1, 2] = 0.067} \\
 & = P[0 \rightarrow 2 \mid b=0.96] = ([0.95 \times 0.96]^2 \times e^{-(0.95 \times 0.96)})/2! \\
 & \times P[0 \rightarrow 0 \mid b=0.96] = e^{-(0.95 \times 0.96)} \\
 & \underline{L[\text{node } 2 = 1 \mid \text{taxa } 1, 2] = 0.134} \\
 & = P[1 \rightarrow 2 \mid b=0.96] = ([0.95 \times 0.96]^1 \times e^{-(0.95 \times 0.96)})/1! \\
 & \times P[0 \rightarrow 0 \mid b=0.96] = ([0.95 \times 0.96]^1 \times e^{-(0.95 \times 0.96)})/1! \\
 & \underline{L[\text{node } 2 = 2 \mid \text{taxa } 1, 2] = 0.067}
 \end{aligned}$$

What is the likelihood of the basal nodes states?

$$\begin{aligned}
 & L[\text{node}_1 = X \mid 2, 0, 2, 0] \\
 & = P[0 \mid \text{node}_1 = X] \\
 & \times P[2 \mid \text{node}_1 = X] \\
 & \times (P[0 \mid \text{node}_1 = X] \times P[0 \mid \text{node}_2=0] \times P[0 \mid \text{node}_2=0] \\
 & \quad + P[1 \mid \text{node}_1 = X] \times P[0 \mid \text{node}_2=1] \times P[0 \mid \text{node}_2=1] \\
 & \quad + P[2 \mid \text{node}_1 = X] \times P[0 \mid \text{node}_2=2] \times P[0 \mid \text{node}_2=2])
 \end{aligned}$$

Note: final terms are the likelihoods of node 2 states times the conditional probabilities of those states given node 1.

Ancestral conditions as conditional probability

$$\begin{aligned}
 & L[\tau, \pi=0.95 \mid 2, 0, 2, 0] \\
 & = \omega_0 \times L[\text{node}_1 = 0 \mid 2, 0, 2, 0] \\
 & + \omega_1 \times L[\text{node}_1 = 1 \mid 2, 0, 2, 0] \\
 & + \omega_2 \times L[\text{node}_1 = 2 \mid 2, 0, 2, 0]
 \end{aligned}$$

Where ω_i is the probability of beginning with state i.
Tree likelihood obviously modified.

Phylogeny Likelihood

- $$L[\tau, \pi \mid \mathbf{C}] = \prod_{i=1}^{\text{characters}} \sum_{k=0}^{\text{states}} \prod_{j=1}^{\text{branches}} P[\Delta_{ijk} \mid b_j, \pi]$$
 - π : rate;
 - b : branch j on tree τ
 - \mathbf{C} : character matrix
 - Δ_{ijk} : number of changes in character i on branch j given ancestral state k.
- Different phylogenies matching the same cladogram will have different likelihoods!

Changing Branch Durations Changes Likelihood

Likelihood of upper node as well as P[0], P[1] or P[2] red, yellow and orange branches now altered.

Sum of potentially static lineages AND lineages over which change accrued also differ

on the two trees.

Upshot: cladogram does not have likelihood *unless* you sum over all possible phylogenies!

With fossils, adding unsampled ancestors increases likelihood.

Hypothetical ancestors decrease necessary transitions and increase possible pathways.

$$\text{If } P[\Delta] = 0.05, \text{ then } L[\tau_1] = 0.05^2 \\ = 2.5 \times 10^{-3}$$

$$L[\tau_2] = \omega_0 \times (L[\tau_{2A}] + L[\tau_{2B}]) + \omega_1 \times (L[\tau_{2A}] + L[\tau_{2B}])$$

where ω_i is the probability of state i at the base.

$$L[\tau_{2A}] = 0.95 \times 0.05 \times 0.95 \times 0.95 \\ = 4.29 \times 10^{-2}$$

$$L[\tau_{2B}] = 0.05 \times 0.95 \times 0.95 \times 0.05 \\ = 2.26 \times 10^{-3}$$

$$L[\tau_{2C}] = 0.05 \times 0.05 \times 0.05 \times 0.95 \\ = 1.19 \times 10^{-4}$$

$$L[\tau_{2D}] = 0.05 \times 0.95 \times 0.95 \times 0.05 \\ = 2.26 \times 10^{-3}$$

$$\text{If } P[\Delta] = 0.05, \text{ then } L[\tau_1] = 2.5 \times 10^{-3} \\ L[\tau_2] \approx 2.4 \times 10^{-2}$$

Changing Rate Changes Likelihood

First tree's likelihood maximized at $\pi \approx 0.95$;

Second tree's likelihood maximized at $\pi \approx 1.20$;

Same number of changes favored, but less time:
($t = 4.26$ vs. $t = 3.30$)

Upshot: cladogram does not have likelihood *unless* you sum over all possible rates!

Continuous vs. Pulsed Change

- Equations presented above assume continuous change.
 - What if change is pulsed? (speciational, punctuated, etc.);
 - If so, then change should have a binomial distribution at each pulse;
 - However, pulses themselves might have a Poisson distribution
 - e.g., based on speciation rate.
 - This gives a Poisson distribution of binomial events!
- $P[\Delta | t] = \sum_{i=1}^{anc} P[\Delta | i \text{ species}, \pi] \times P[i \text{ species} | \mu, t]$,
 - μ = speciation rate,
 - t = time;
 - anc = unsampled ancestral species.

“Weights” and likelihood

Doubling a character's weight invokes two step matrices:

| From\To: | 0 | 1 | From\To: | 0 | 1 |
|----------|---|---|----------|---|---|
| 0 | 0 | 1 | 0 | 0 | 2 |
| 1 | 1 | 0 | 1 | 2 | 0 |

This assumes that $P[\text{change char. B}] = P[\text{change char. A}]^2$, *not* $P[\text{change char. B}] = 2 \times P[\text{change char. A}]$.

| From\To: | 0 | 1 | From\To: | 0 | 1 |
|----------|---------|---------|----------|-------------|-------------|
| 0 | $1-p_a$ | p_a | 0 | $1-(p_a)^2$ | $(p_a)^2$ |
| 1 | p_a | $1-p_a$ | 1 | $(p_a)^2$ | $1-(p_a)^2$ |

Thus, weights reflect exponents of “base” rate.

“Ordered states” and likelihood

Doubling a character's weight invokes two step matrices:

| From\To: | 0 | 1 | 2 |
|----------|---|---|---|
| 0 | 0 | 1 | 2 |
| 1 | 1 | 0 | 1 |
| 2 | 2 | 1 | 0 |

Instead of implying that 1 must evolve between 0 and 2, it now implies that $P[0 \leftrightarrow 1] = P[0 \leftrightarrow 2]^2$.

| From\To: | 0 | 1 | 2 |
|----------|-------------|--------|-------------|
| 0 | $1-(p+p^2)$ | p | p^2 |
| 1 | p | $1-2p$ | p |
| 2 | p^2 | p | $1-(p+p^2)$ |

Note: Each row must sum to 1.0.

“Unordered states” and likelihood

Doubling a character's weight invokes two step matrices:

| From\To: | 0 | 1 | 2 |
|----------|---|---|---|
| 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 |
| 2 | 1 | 1 | 0 |

The probability of changing to any one state is simply one divided by the number of options (e.g., 2 if 3 states).

| From\To: | 0 | 1 | 2 |
|----------|-----|-----|-----|
| 0 | 1-p | p/2 | p/2 |
| 1 | p/2 | 1-p | p/2 |
| 2 | p/2 | p/2 | 1-p |

Changing Branch Durations Changes Likelihood

Likelihood of upper node as well as P[0], P[1] or P[2] red, yellow and orange branches now altered.

Sum of potentially static lineages AND lineages over which change accrued also differ on the two trees.

Upshot: cladogram does not have likelihood *unless* you sum over all possible phylogenies!

Bayesian Probability

- Bayesian probability: $P[\text{hypothesis} \mid \text{data}]$
 - Classical probability is $P[d \mid H]$;
 - where d is data & H is hypothesis
 - only good for rejecting hypotheses.
 - Likelihood: $L[H \mid d] = P[d \mid H]$
 - Good for inference (ML)
 - Also for hypothesis testing (e.g., ratio tests).
 - It is possible for $L[H \mid d] = 1.0$ for many hypotheses.
- $P[H \mid d] = P[h] \times L[H \mid d] / P[d]$
 - $P[H]$: *prior* probability.
 - Only one hypothesis can have $P[H \mid d] > 0.5$
- Given that a bird is black, what is the probability that it belongs to a given species?
 - Crow:
 - 1% of all birds ($P[H] = 0.01$)
 - All of them are black ($P[d \mid H] = 1.00$)
 - New Zealand All Black Cuckoo:
 - $10^{-5}\%$ of all birds ($P[H] = 10^{-7}$)

- All of them are black ($P[d|H] = 1.00$)
- Pigeon:
 - 5% of all birds ($P[H] = 0.05$)
 - 1% of them are black ($P[d|H] = 0.01$).
- Birds that are black are 4% of birds ($P[d] = 0.04$)
- $P[\text{spec.} | \text{black}] = (P[\text{spec.}] \times P[\text{black}|\text{spec.}]) / P[\text{black}]$
- $P[\text{Crow} | \text{black}] = (0.01 \times 1.00) / 0.04 = 0.25$
- $P[\text{NZ Cuckoo} | \text{black}] = (10^{-7} \times 1.00) / 0.04 = 2.5 \times 10^{-6}$
- $P[\text{Pigeon} | \text{black}] = (0.05 \times 0.01) / 0.04 = 0.0125$
- NZ All black cuckoo is more likely than pigeon because a greater frequency of cuckoos are black;
- Pigeon is more probable because a greater frequency of black birds are pigeons.

Bayesian Probability of General Phylogeny

- $P[\text{cladogram} | \text{data}] = \sum P[\text{tree}] \times L[\text{tree} | \text{data}]$
- $P[\text{tree}]$ assumed to be $1/(\text{total trees})$ for each tree;
 - I.e., *flat priors*.
- $P[\text{data}]$ assumed to be $1/\text{possible matrices}$;
- Approach basically sums tree likelihoods and divides by the number of trees examined.
- Bayesian or conditional likelihood?