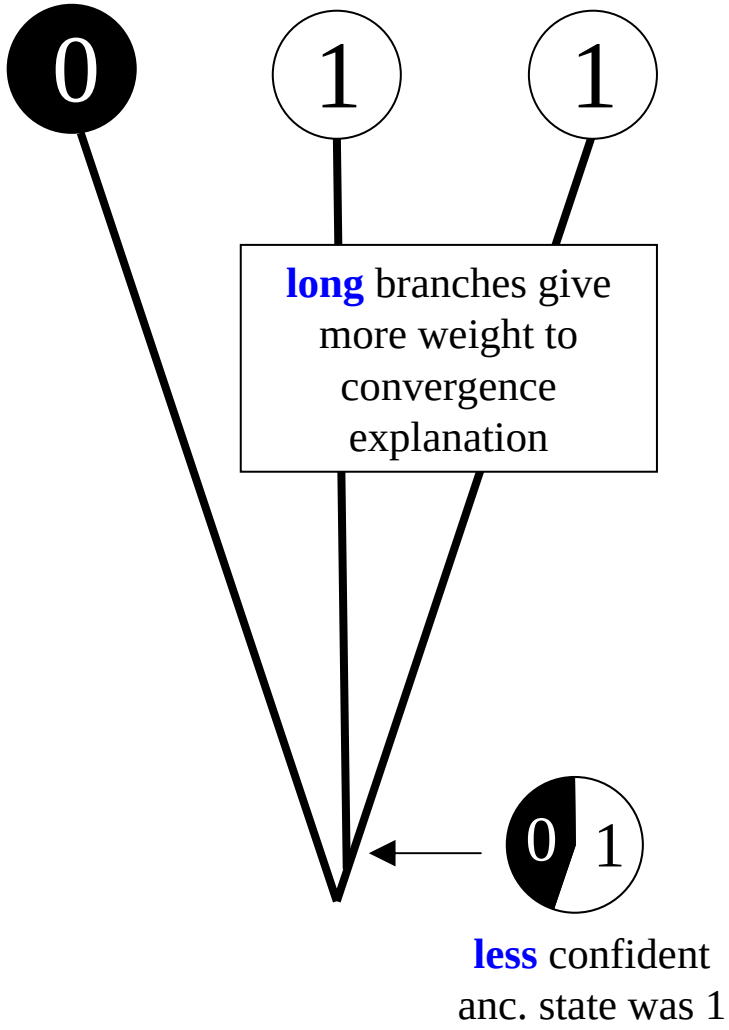
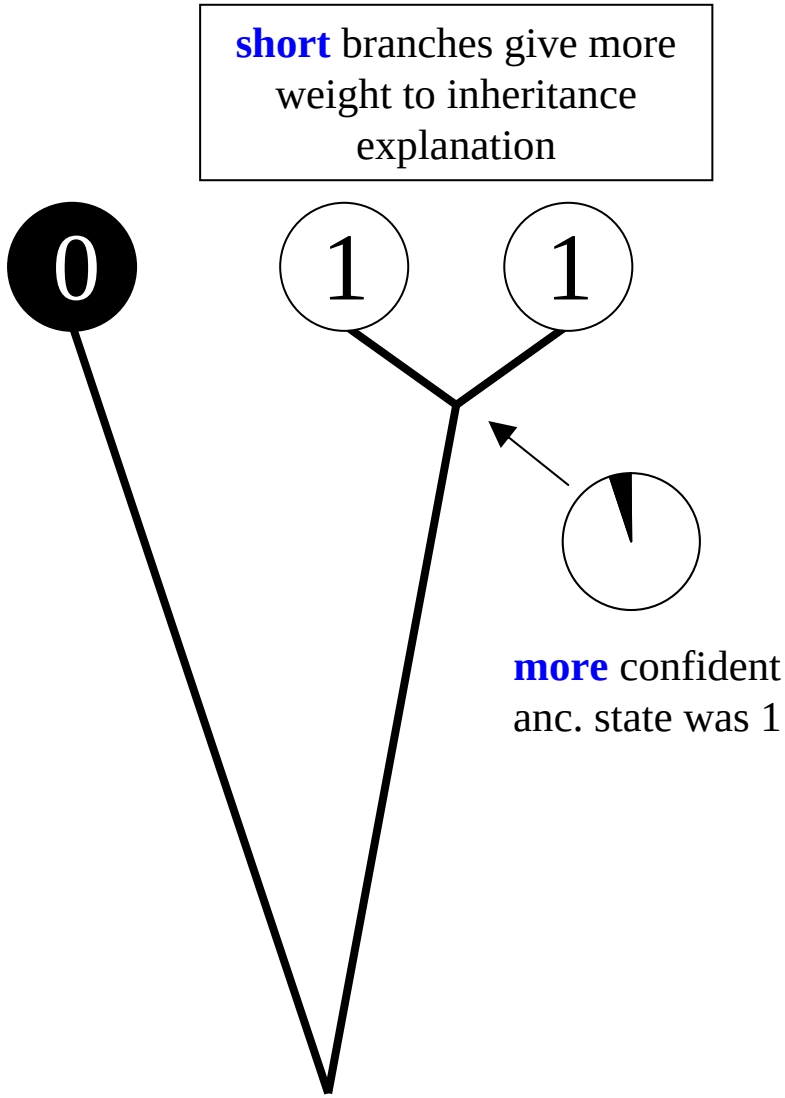
Some of these slides have been borrowed from Dr. Paul Lewis, Dr. Joe Felsenstein. Thanks!

Paul has many great tools for teaching phylogenetics at his web site:

http://hydrodictyon.eeb.uconn.edu/people/plewis

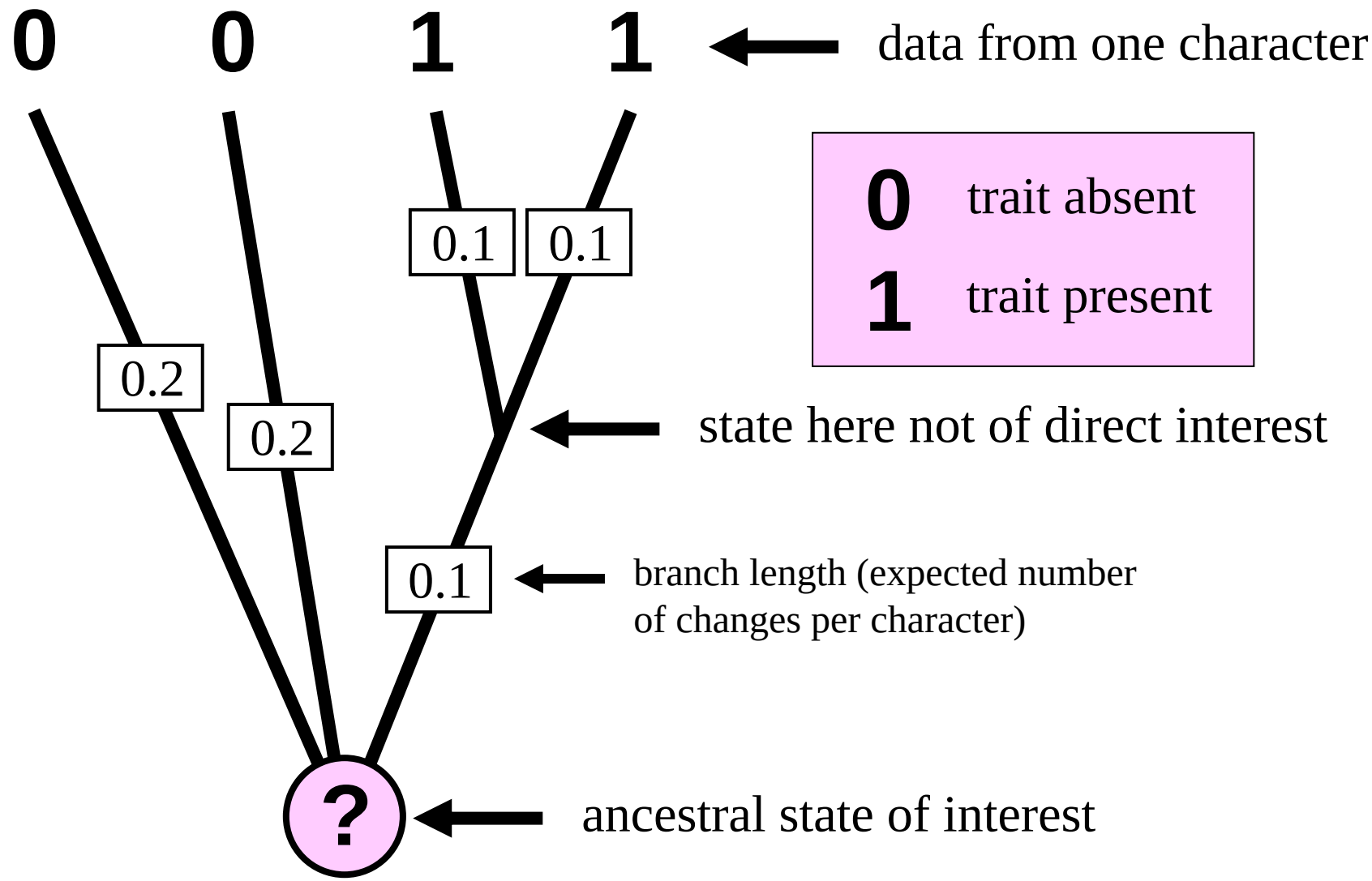
A matter of confidence





Copyright © 2007 by Paul O. Lewis

Maximum likelihood ancestral state estimation



Copyright © 2007 by Paul O. Lewis

"Mk" models

"Cavender-Farris" model:

2 possible states (k = 2)

expected no. changes = μt

$P_{ii}(t) = 0.5 + 0.5\, e^{-2\mu t}$

$P_{ij}(t) = 0.5 - 0.5\, e^{-2\mu t}$

Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. Journal of Genetics 8:299-309.

Cavender, J. A. 1978. Taxonomy with confidence. Mathematical Biosciences 40:271-280.

Farris, J. S. 1973. A probability model for inferring evolutionary trees. Systematic Zoology 22:250-256.

Felsenstein, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biological Journal of the Linnean Society 16:183-196.

Jukes-Cantor (1969) model:

4 possible states (k = 4)

expected no. changes = $3\, \mu t$

$P_{ii}(t) = 0.25 + 0.75\, e^{-4\mu t}$

$P_{ij}(t) = 0.25 - 0.25\, e^{-4\mu t}$

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

Copyright © 2007 by Paul O. Lewis

For branch of length 0.1

$\mu t = 0.1$

$P_{00}(0.1) = 0.5 + 0.5\, e^{-2(0.1)} = 0.90936$

$P_{01}(0.1) = 0.5 - 0.5\, e^{-2(0.1)} = 0.\ 09064$

$P_{10}(0.1) = 0.5 - 0.5\, e^{-2(0.1)} = 0.\ 09064$

$P_{11}(0.1) = 0.5 + 0.5\, e^{-2(0.1)} = 0.90936$

Copyright © 2007 by Paul O. Lewis

For branch of length 0.2

$\mu t = 0.2$

$P_{00}(0.2) = 0.5 + 0.5\, e^{-2(0.2)} = 0.83516$

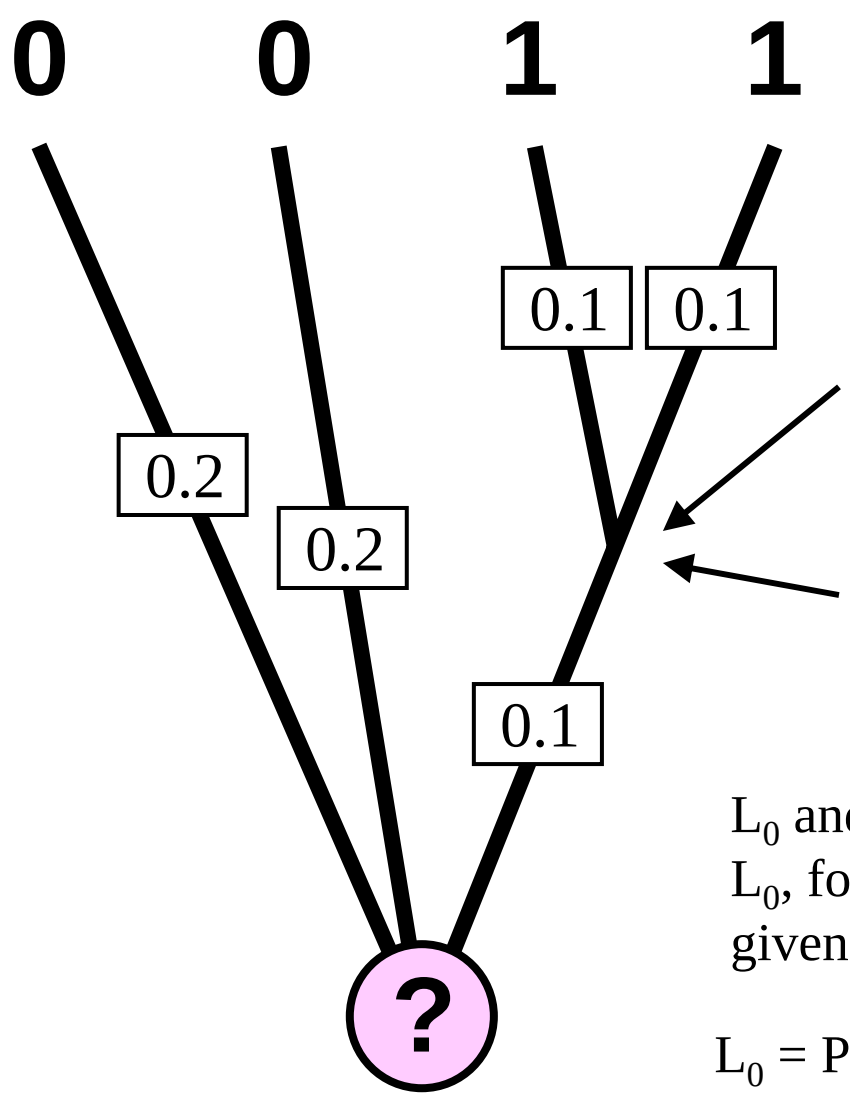
$P_{01}(0.2) = 0.5 - 0.5\, e^{-2(0.2)} = 0.\ 16484$

$P_{10}(0.2) = 0.5 - 0.5\, e^{-2(0.2)} = 0.\ 16484$

$P_{11}(0.2) = 0.5 + 0.5\, e^{-2(0.2)} = 0.\ 83516$

Copyright © 2007 by Paul O. Lewis

Calculating conditional likelihoods at upper node



$L_0 = P_{01}(.1)\ P_{01}(.1)$
$= [.09064]^2 = 0.00822$

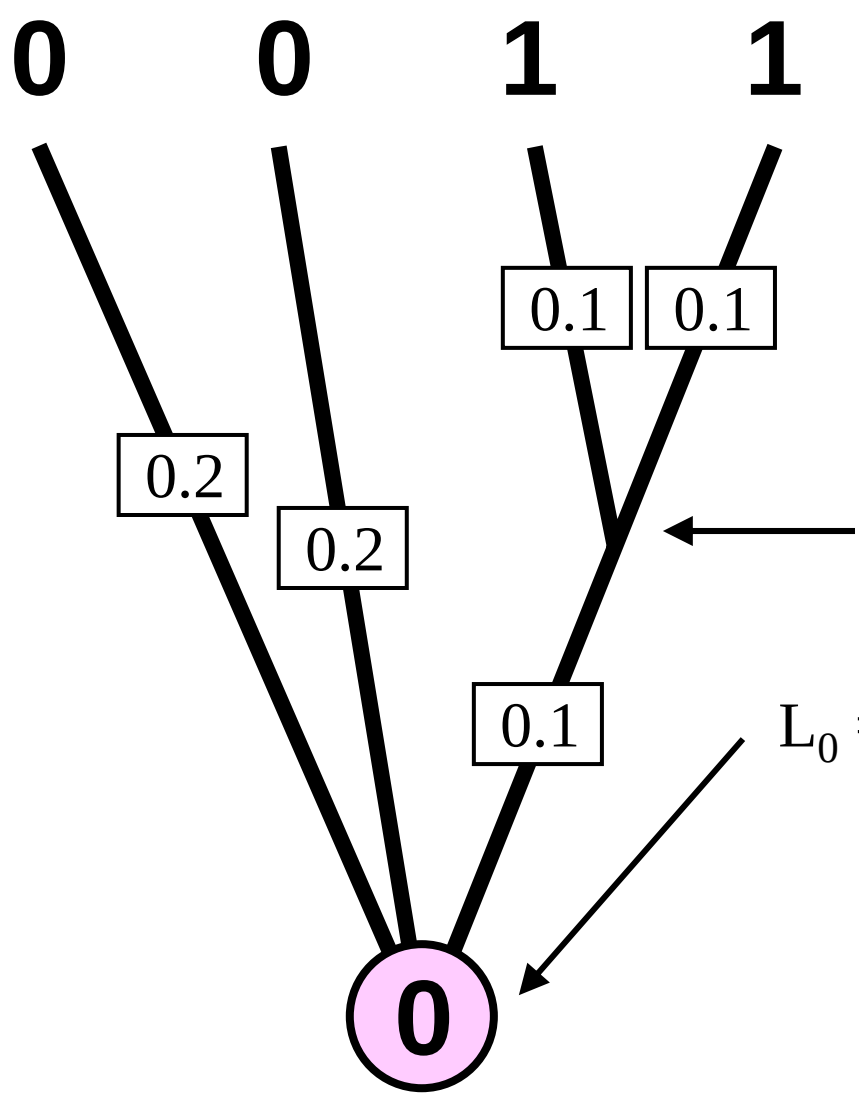
$L_1 = P_{11}(.1)\ P_{11}(.1)$
$= [0.90936]^2 = 0.82694$

L_0 and L_1 are known as *conditional likelihoods*. L_0, for example, is the likelihood of the subtree given that the node has state 0.

$L_0 = \Pr(\text{data above node} \mid \text{node state} = 0)$

Copyright © 2007 by Paul O. Lew

Calculating likelihood at basal node conditional on basal node having state 0

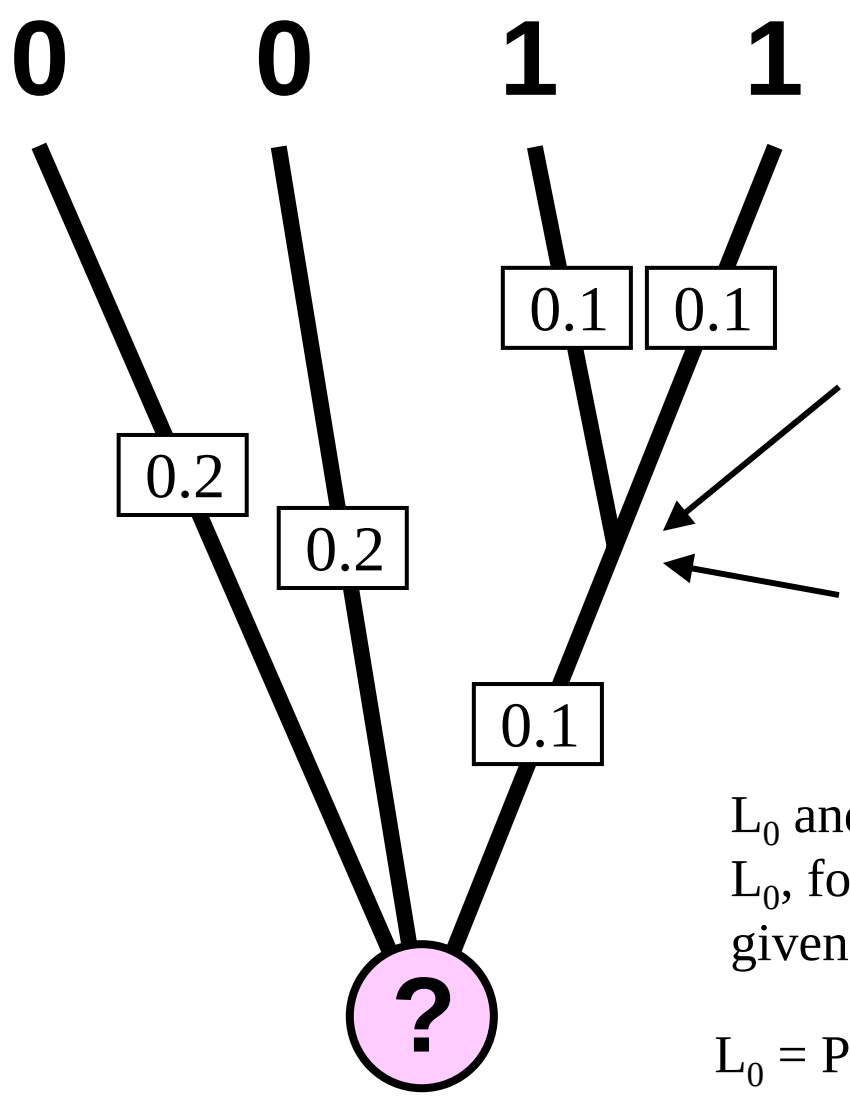


$L_0 = 0.00822 \quad L_1 = 0.82694$

$$
\begin{aligned}
L_0 &= P_{00}(.2)\, P_{00}(.2)\, P_{00}(.1)\, L_0 \\
&\quad + P_{00}(.2)\, P_{00}(.2)\, P_{01}(.1)\, L_1 \\
&= (.83516)^2\, (.90936)\, (.00822) \\
&\quad + (.83516)^2\, (.09064)\, (.82694) \\
&= 0.05749
\end{aligned}
$$

Copyright © 2007 by Paul O. Lewis

Calculating conditional likelihoods at upper node



$L_0 = P_{01}(.1)\ P_{01}(.1)$
$= [.09064]^2 = 0.00822$

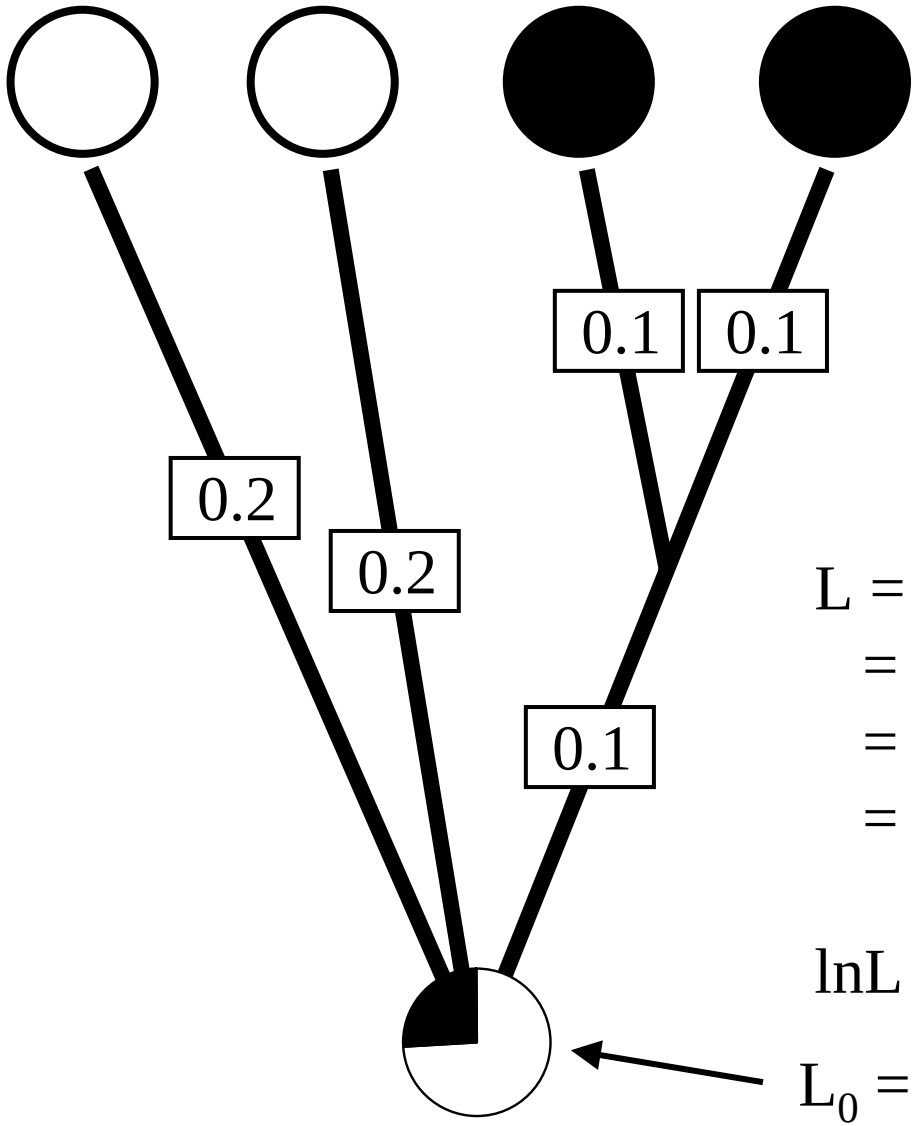
$L_1 = P_{11}(.1)\ P_{11}(.1)$
$= [0.90936]^2 = 0.82694$

L_0 and L_1 are known as *conditional likelihoods*. L_0, for example, is the likelihood of the subtree given that the node has state 0.

$L_0 = \Pr(\text{data above node} \mid \text{node state} = 0)$

Copyright © 2007 by Paul O. Lewis

Calculating likelihood of the tree using conditional likelihoods at the basal node



D here refers to the observed data at the tips

0 here refers to the state at the basal node

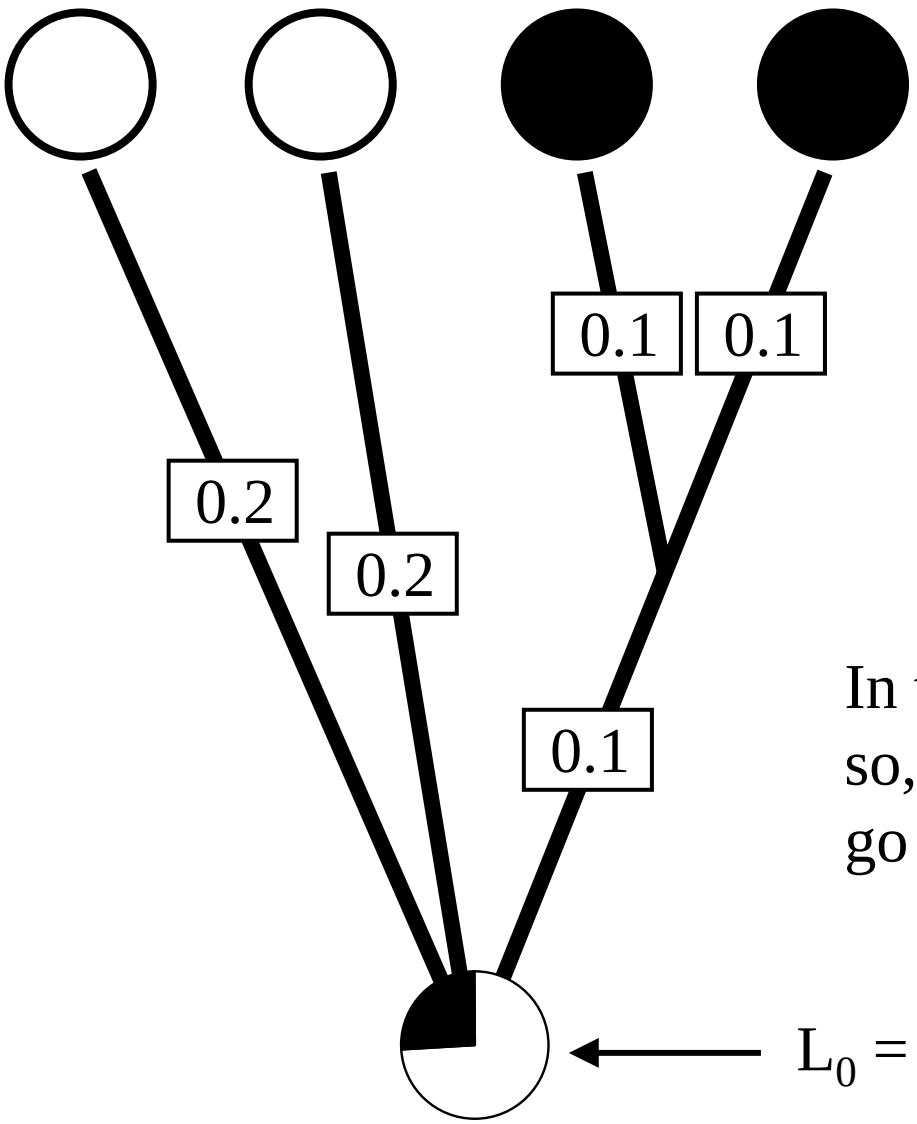
$$L = Pr(0)\ Pr(D|0) + Pr(1)\ Pr(D|1)$$
$$= (0.5)\ L_0 + (0.5)\ L_1$$
$$= (0.5)\ (0.05749) + (0.5)\ (0.02045)$$
$$= (0.5)(0.07794) = 0.03897$$

$$\ln L = -3.24496$$

$L_0 = 0.05749 \quad L_1 = 0.02045$

Copyright © 2007 by Paul O. Lewis

Choosing the maximum likelihood ancestral state is simply a matter of seeing which conditional likelihood is the largest



In this case, L_0 is larger of the two so, if forced to decide, you would go with 0 as the ancestral state.

$L_0 = 0.05749$ (73.8%) $L_1 = 0.02045$ (26.2%)

Copyright © 2007 by Paul O. Lewis

Is this a ML or Bayesian approach?

$L_0 = 0.05749$ (73.8%) $L_1 = 0.02045$ (26.2%)

$$0.738 = \frac{\overset{\text{prior}}{(0.5)}\ \overset{\text{likelihood}}{(0.05749)}}{(0.5)\,(0.05749) + (0.5)\,(0.02045)}$$

posterior probability of state 0

marginal probability of the data

Converting likelihoods to percentages produces Bayesian posterior probabilities (and implicitly assumes a flat prior over ancestral states)

Yang, Z., S. Kumar, and M. Nei. 1995. A new method of inference of ancestral nucleotide and amino-acid sequences. Genetics 141:1641-1650.

Copyright © 2007 by Paul O. Lewis

Empirical Bayes vs. Full Bayes

The likelihood used in the previous slide is the maximum likelihood, i.e. the likelihood at one particular combination of branch length values (and values of other model parameters).

A **fully Bayesian** approach would approximate the posterior probability associated with state 0 and state 1 at the node of interest using an MCMC analysis that allowed all the model parameters to vary, including the branch lengths.

Because some parameters are fixed at their maximum likelihood estimates, this approach is often called an **empirical Bayesian** method.

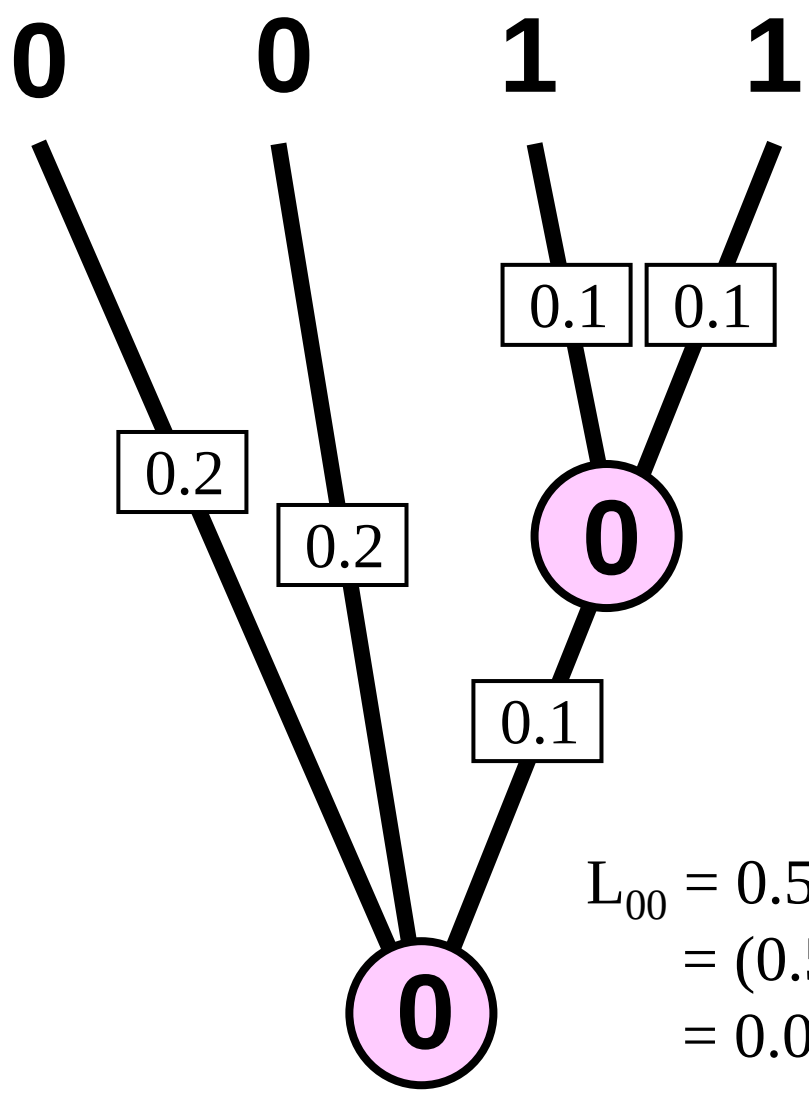
Copyright © 2007 by Paul O. Lewis

Marginal vs. Joint Estimation

- marginal estimation involves summing likelihood over all states at all nodes except one of interest (most common)
- joint estimation involves finding the combination of states at all nodes that *together* account for the highest likelihood (seldom used)

Copyright © 2007 by Paul O. Lewis

Joint Estimation

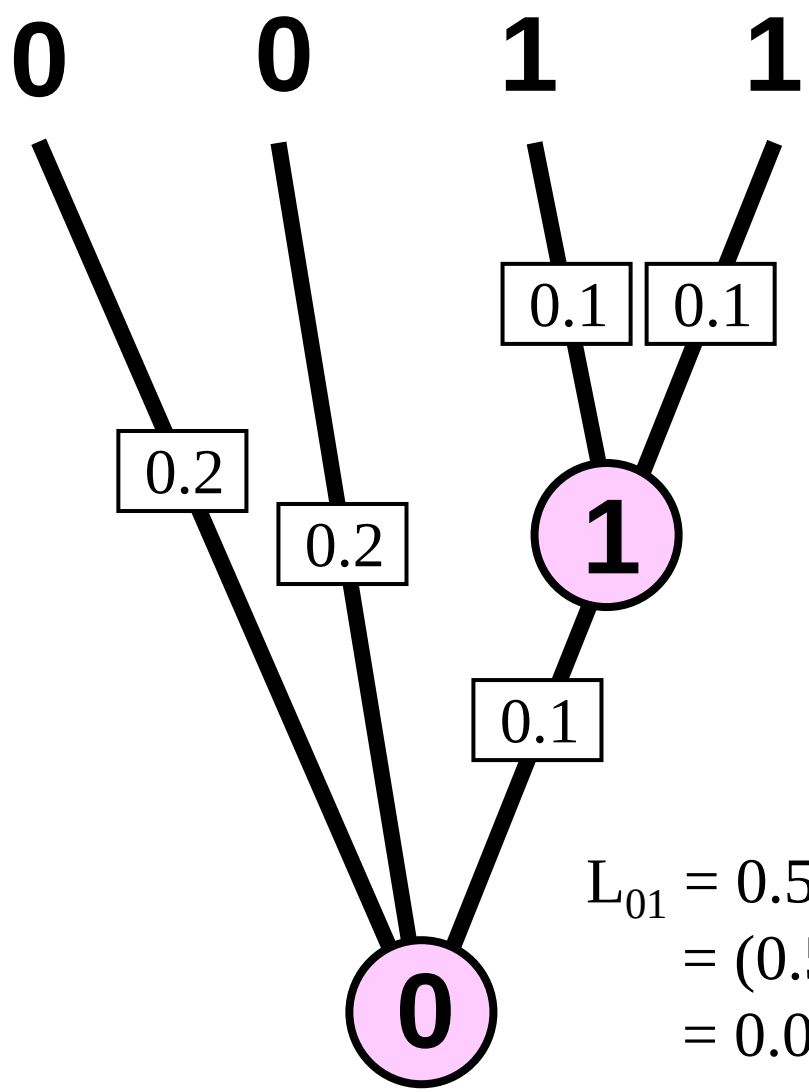


Find likelihood of this particular combination of ancestral states (call it L_{00})

$$L_{00} = 0.5\ P_{00}(.2)\ P_{00}(.2)\ P_{00}(.1)\ P_{01}(.1)\ P_{01}(.1)$$
$$= (0.5)\ (.83516)^2\ (.90936)\ (.09064)^2$$
$$= 0.00261$$

Copyright © 2007 by Paul O. Lewis

Joint Estimation

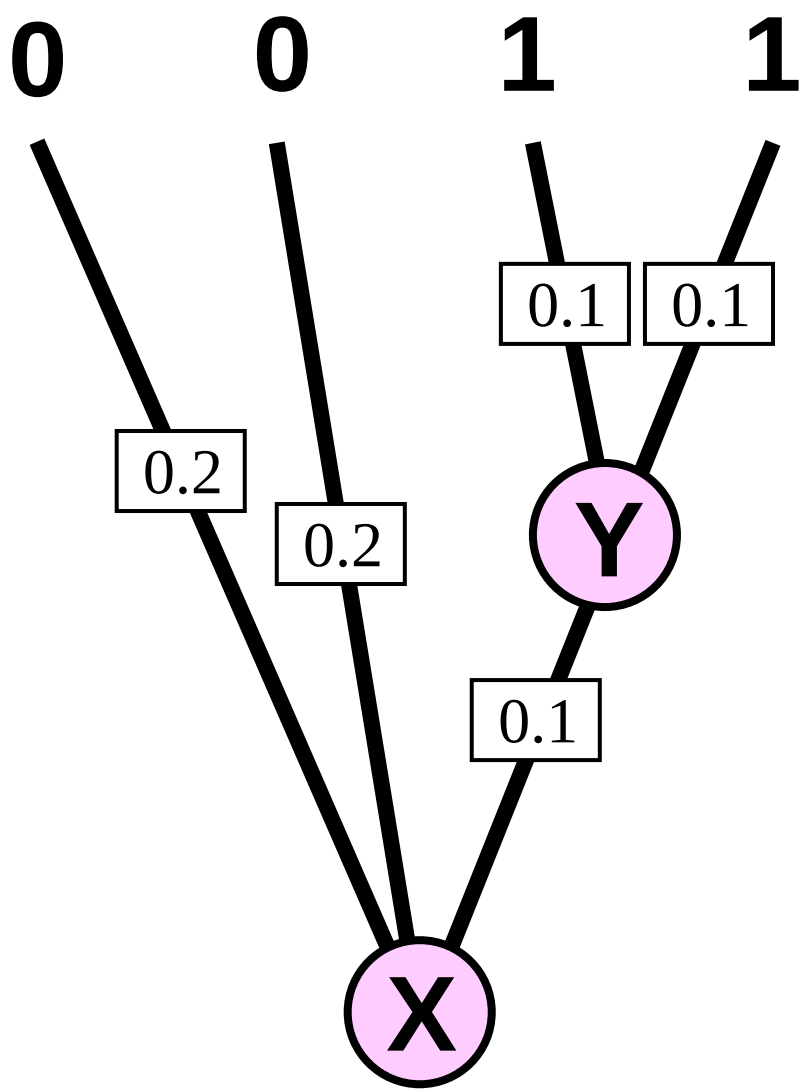


Now we are calculating the likelihood of a different combination of ancestral states (call this one L_{01})

$$L_{01} = 0.5\ P_{00}(.2)\ P_{00}(.2)\ P_{01}(.1)\ P_{11}(.1)\ P_{11}(.1)$$
$$= (0.5)\ (.83516)^2\ (.09064)\ (.90936)^2$$
$$= 0.02614$$

Copyright © 2007 by Paul O. Lewis

Joint Estimation



Summary of the four possibilities:

$L_{00} = 0.00261$

$L_{01} = 0.02614$ ← best

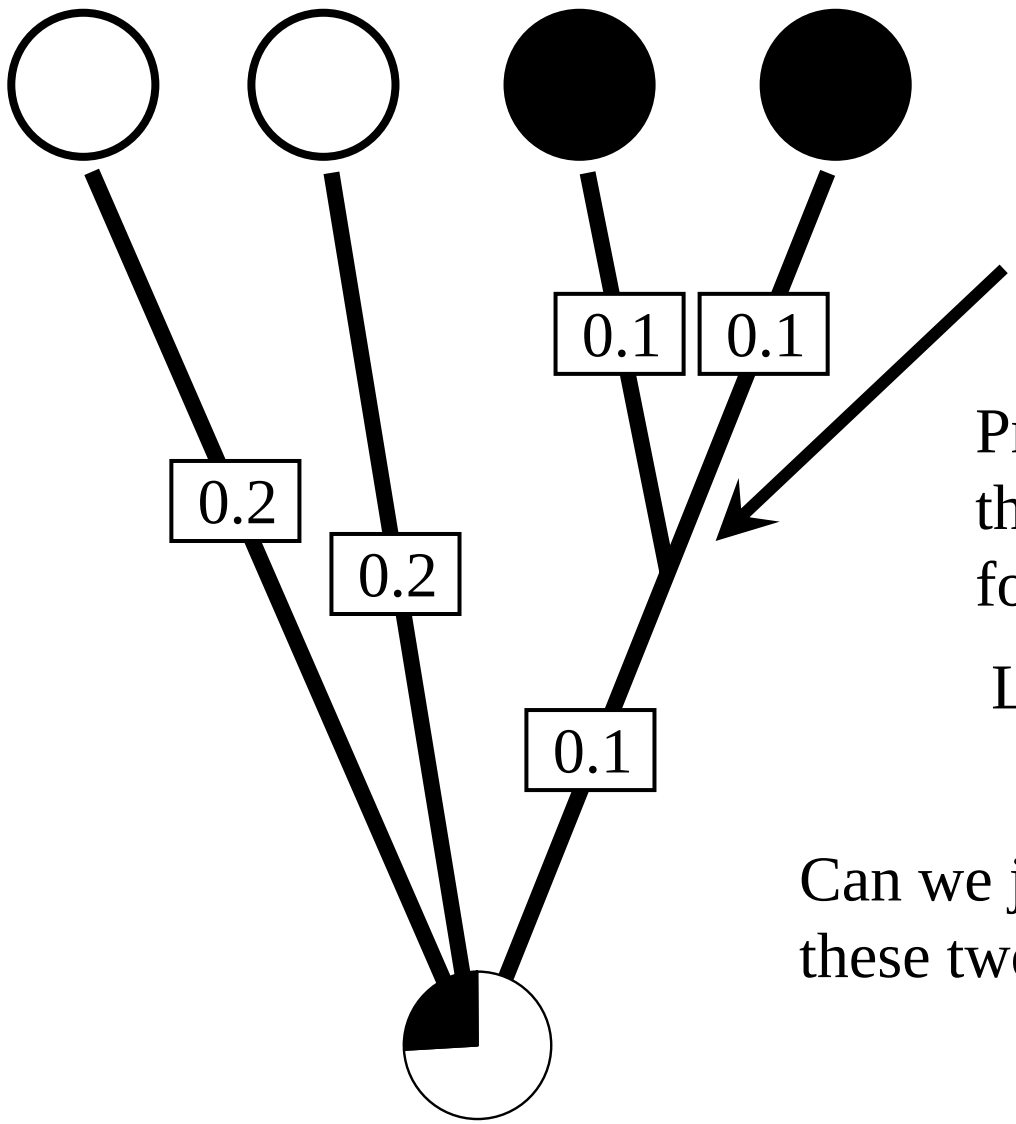
$L_{10} = 0.00003$

$L_{11} = 0.01022$

X Y

Copyright © 2007 by Paul O. Lewis

Returning to marginal estimation...



What about a pie diagram for this node?

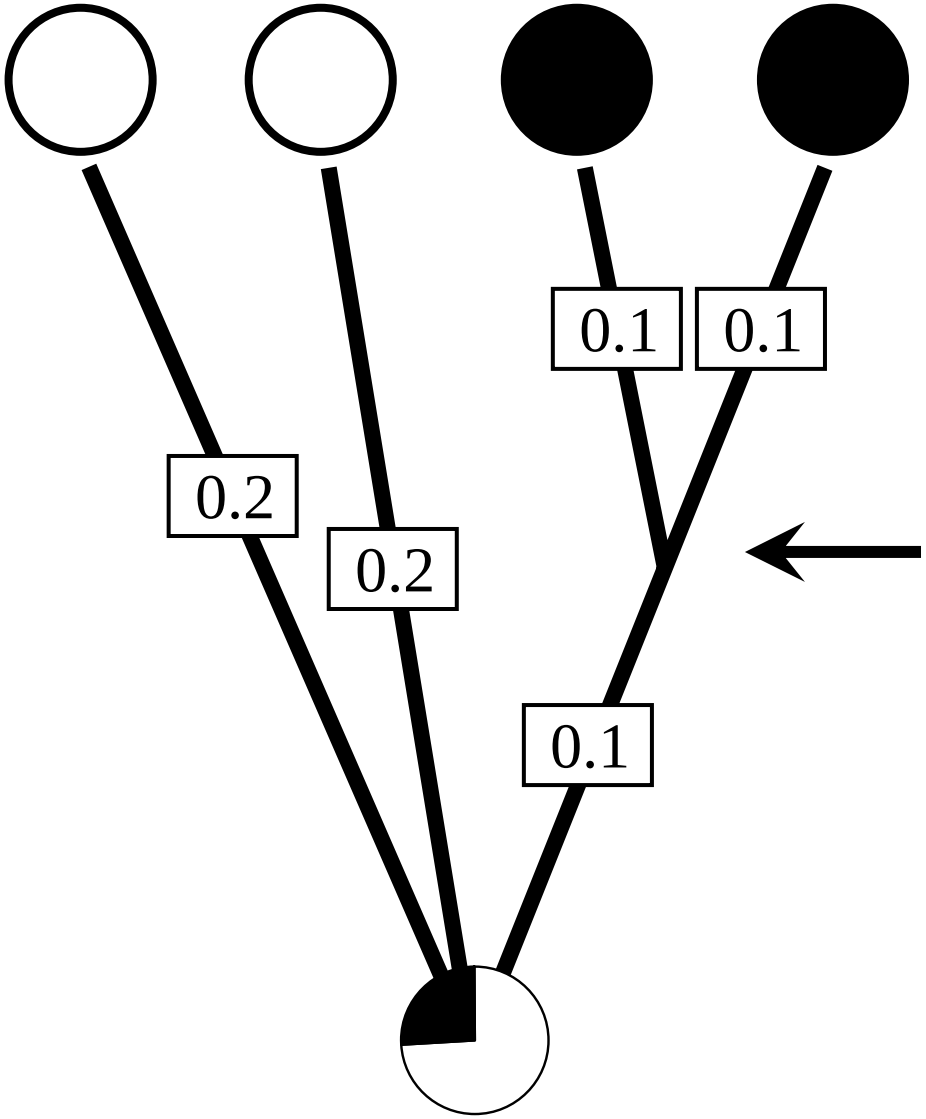
Previously, we computed these conditional likelihoods for this node:

$L_0 = 0.00822 \quad L_1 = 0.82694$

Can we just make a pie diagram from these two values?

Copyright © 2007 by Paul O. Lewis

Returning to marginal estimation...



The conditional likelihoods we computed only included data above this node (i.e. only half the observed states)

Copyright © 2007 by Paul O. Lewis

Estimation of where changes take place

1. What is the probability that a change $0 \rightarrow 1$ took place on a particular branch in the tree?
2. What is the probability that the character changed state exactly x times on the tree ?

Should we answer these questions with a joint or marginal ancestral character state reconstruction?

No

We should figure out the probability of explaining the character state distribution with exactly the set of changes we are interested in.

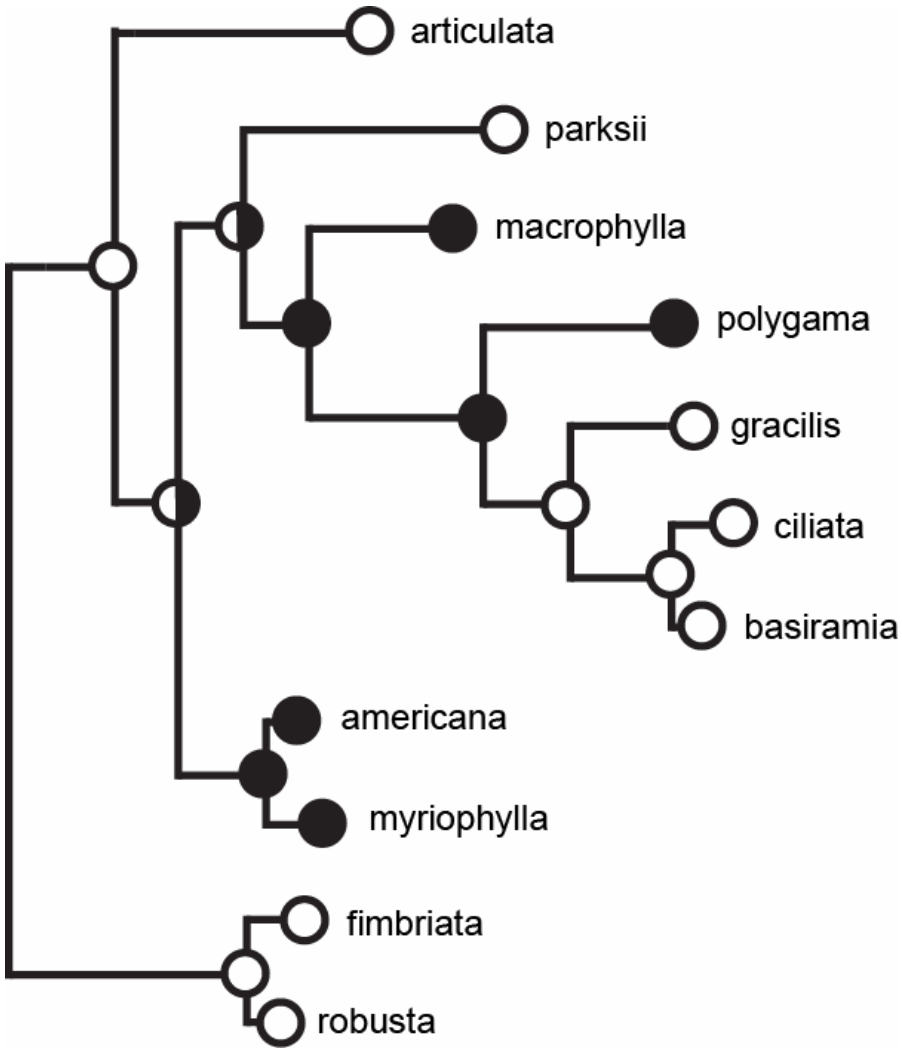
Marginalize over all states that are not crucial to the question

1. What is the probability that a change $0 \rightarrow 1$ took place on a particular branch in the tree? – fix the ancestral node to 0 and the descendant to 1, calculate the likelihood and divide that by the total likelihood.
2. What is the probability that the character changed state exactly x times on the tree ?

1. What is the probability that a change $0 \rightarrow 1$ took place on a particular branch in the tree?
2. What is the probability that the character changed state exactly x times on the tree ? – Sum the likelihoods over all reconstructions that have exactly x changes. This is not easy to do! Recall that $\Pr(0 \rightarrow 1|\nu, \theta)$ from our models accounts for multiple hits. A solution is to sample histories according to their posterior probabilities and summarize these samples.

Ancestral states are one thing...

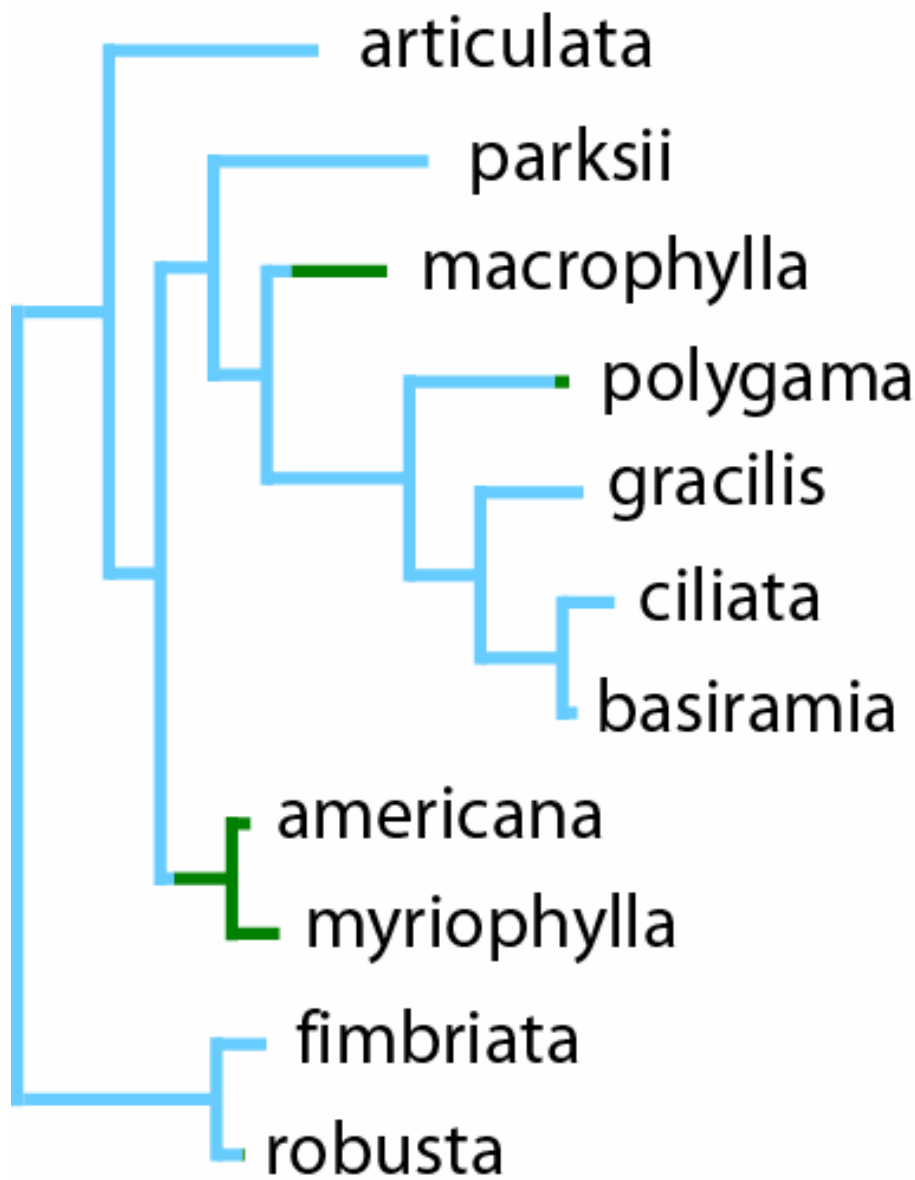
But wouldn't it be nice to estimate not only what happened at the nodes, but also what happened along the edges of the tree?



© 2007 by Paul O. Lewis

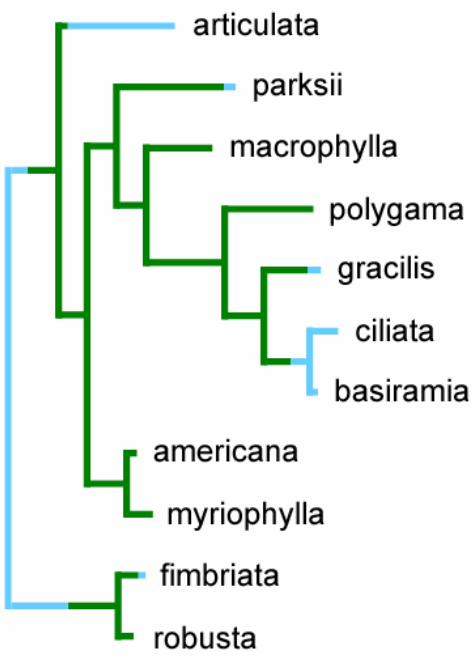
Stochastic mapping simulates an entire character history *consistent with the observed character data*.

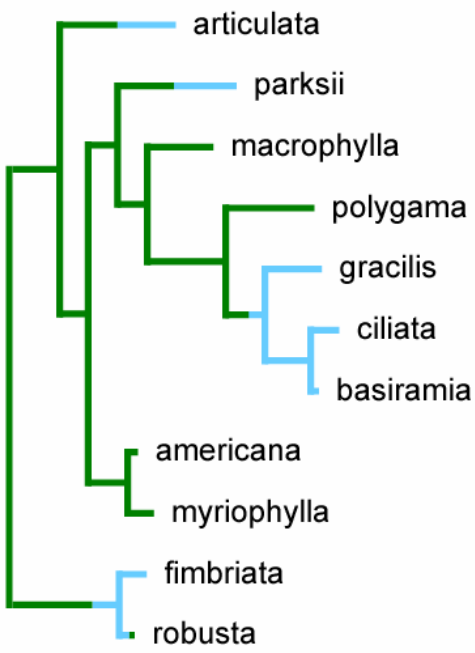
In this case, green means perennial and blue means annual in the flowering plant genus *Polygonella*

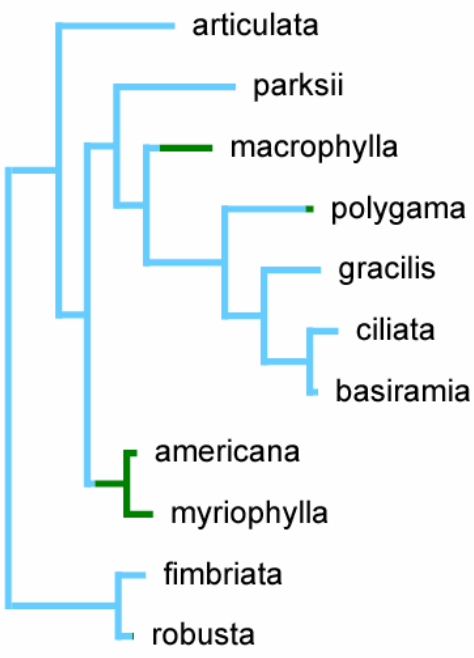


© 2007 by Paul O. Lewis

Three Mappings for Lifespan







Blue = annual
Green = perennial

Each simulation produces a different history (or mapping) consistent with the observed states.

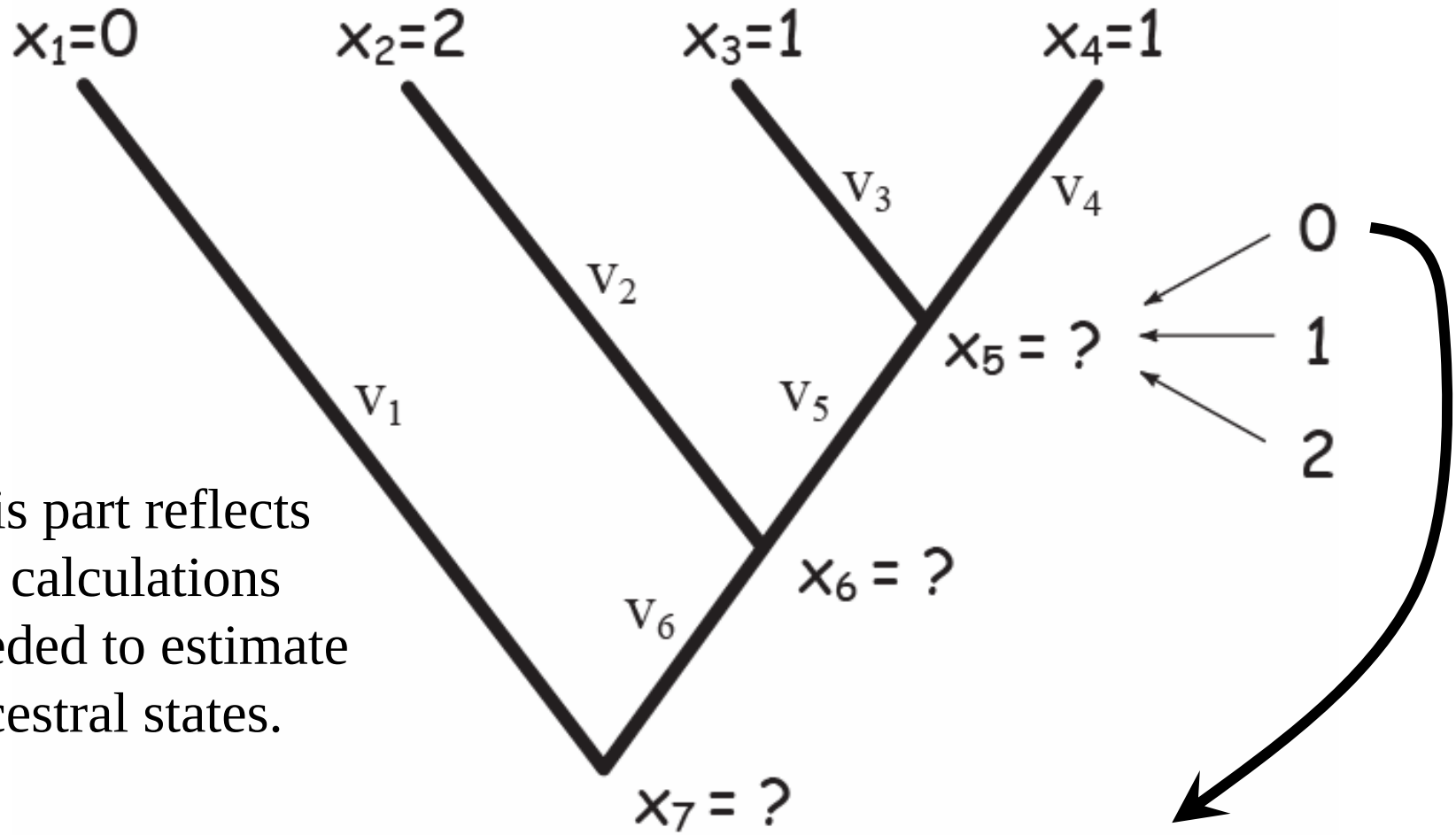
© 2007 by Paul O. Lewis

Kinds of questions you can answer using stochastic mapping

- How long, on average, does a character spend in state 0? In state 1?
- Is the total time in which *two* characters are *both* in state 1 more than would be expected if they were evolving independently?
- What is the number of switches from state 0 to state 1 for a character (on average)?

© 2007 by Paul O. Lewis

Step 1: Downpass to obtain conditional likelihoods

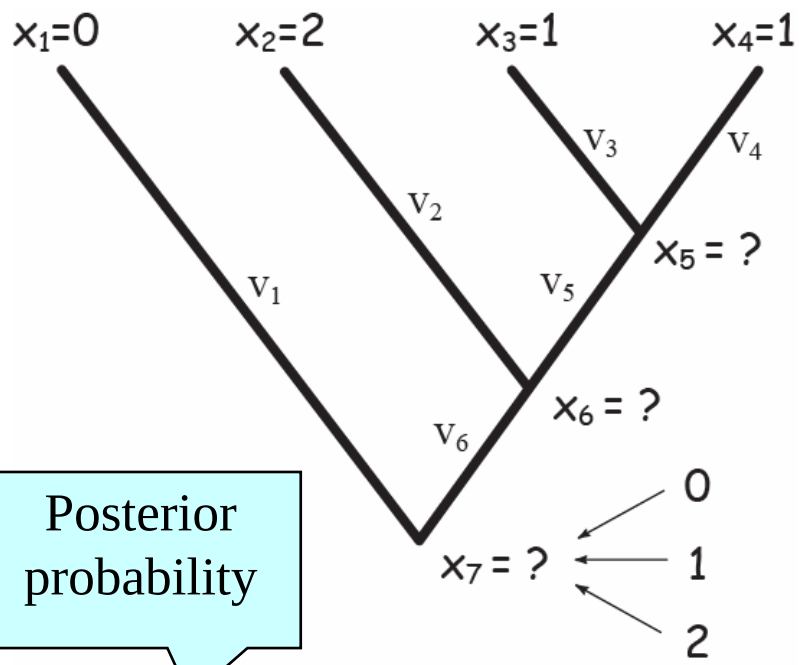


This part reflects the calculations needed to estimate ancestral states.

$$L_0 = \Pr(x_3 = 1 | x_5 = 0, v_3)\ \Pr(x_4 = 1 | x_5 = 0, v_4)$$

© 2007 by Paul O. Lewis

Step 2: Up-pass to sample states at the interior nodes



This part resembles a standard simulation, with the exception that we must conform to the observed states at the tips

Posterior probability

Likelihood

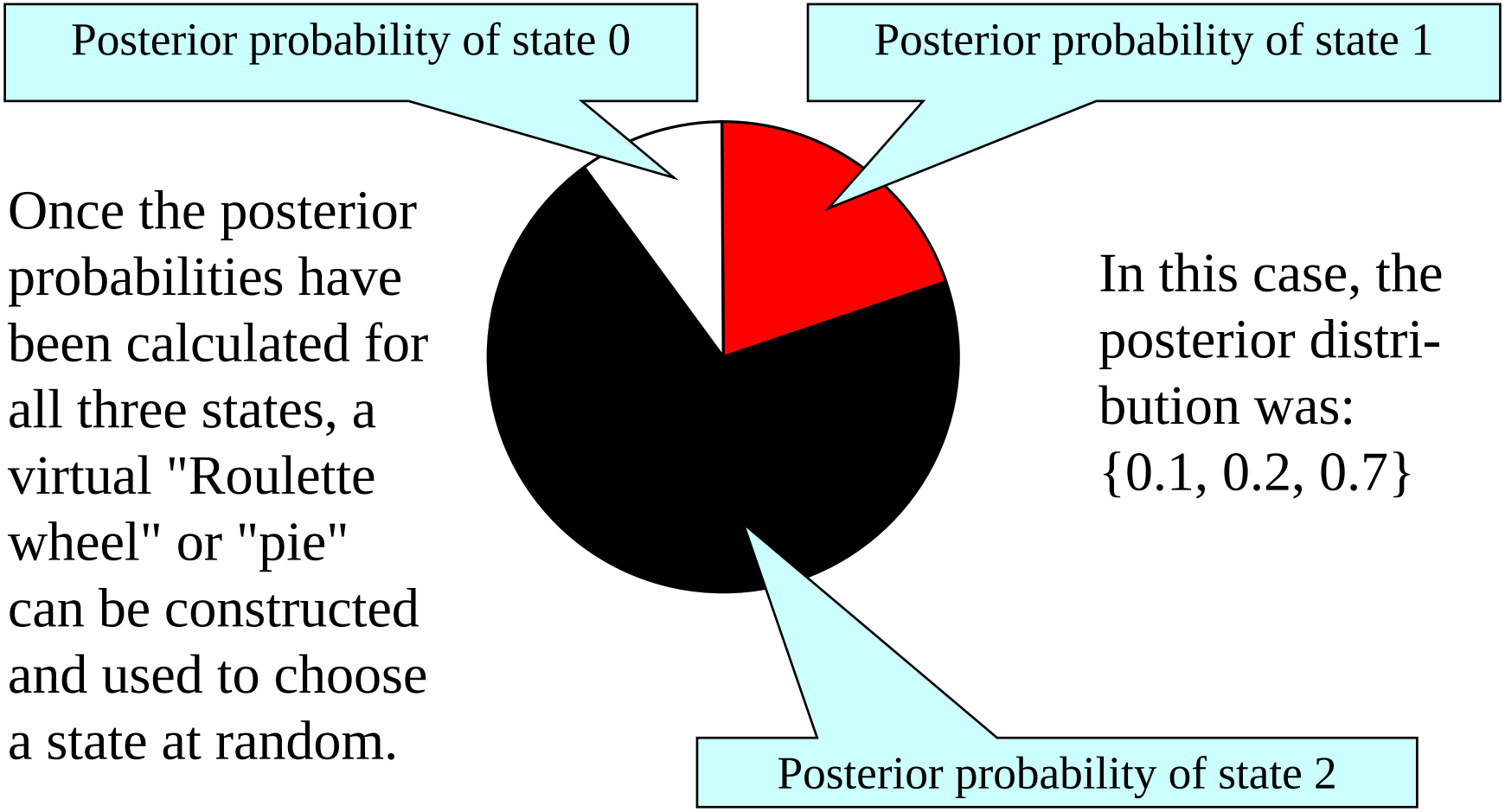
Prior probability

$$\Pr(x_7 = 0|x_1, x_2, x_3, x_4) \propto \Pr(x_1, x_2, x_3, x_4|x_7 = 0)\ \Pr(x_7 = 0)$$
$$\Pr(x_7 = 1|x_1, x_2, x_3, x_4) \propto \Pr(x_1, x_2, x_3, x_4|x_7 = 1)\ \Pr(x_7 = 1)$$
$$\Pr(x_7 = 2|x_1, x_2, x_3, x_4) \propto \Pr(x_1, x_2, x_3, x_4|x_7 = 2)\ \Pr(x_7 = 2)$$

© 2007 by Paul O. Lewis

Choosing a state at an interior node

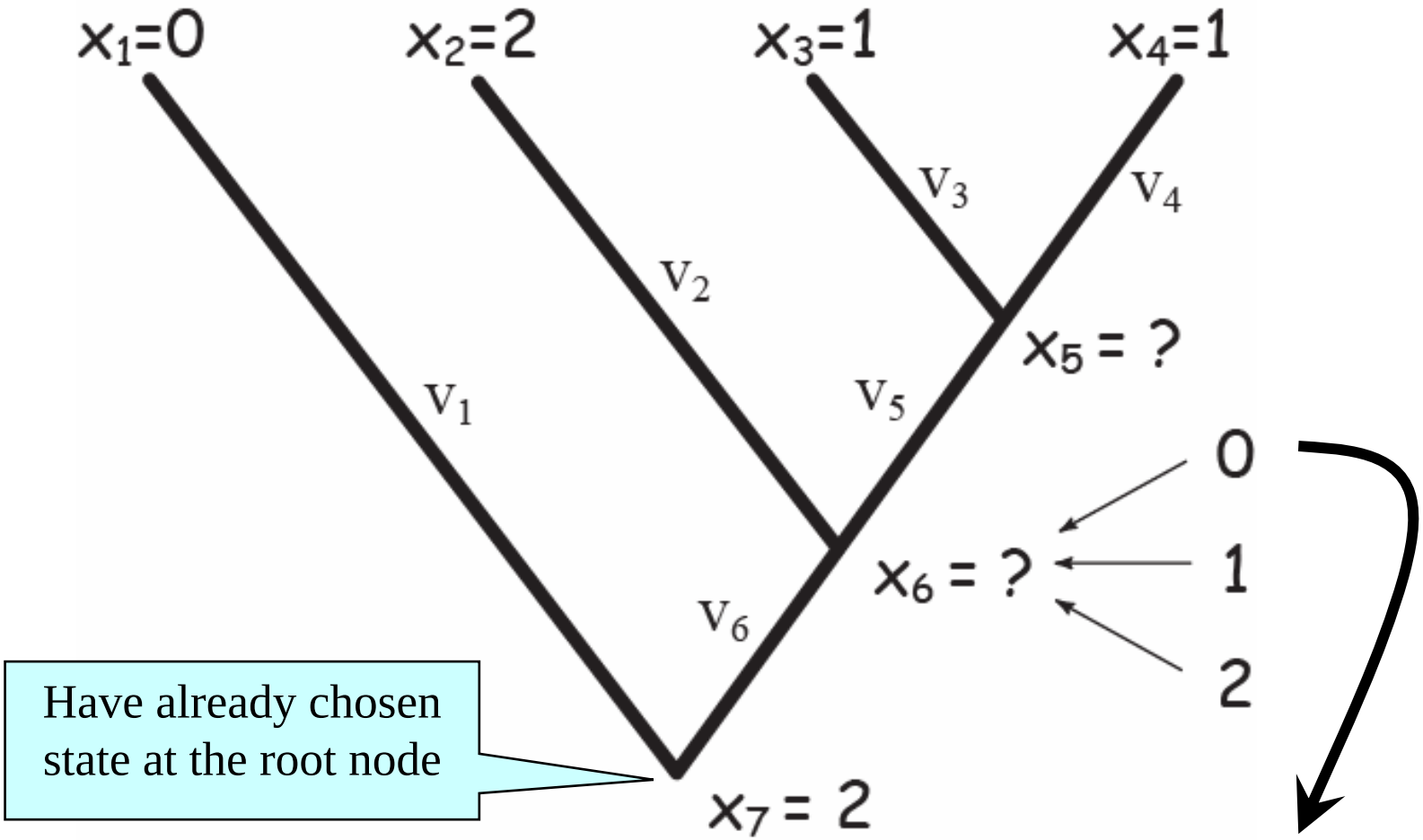
Once the posterior probabilities have been calculated for all three states, a virtual "Roulette wheel" or "pie" can be constructed and used to choose a state at random.



In this case, the posterior distribution was: {0.1, 0.2, 0.7}

© 2007 by Paul O. Lewis

Sampling state at node 6

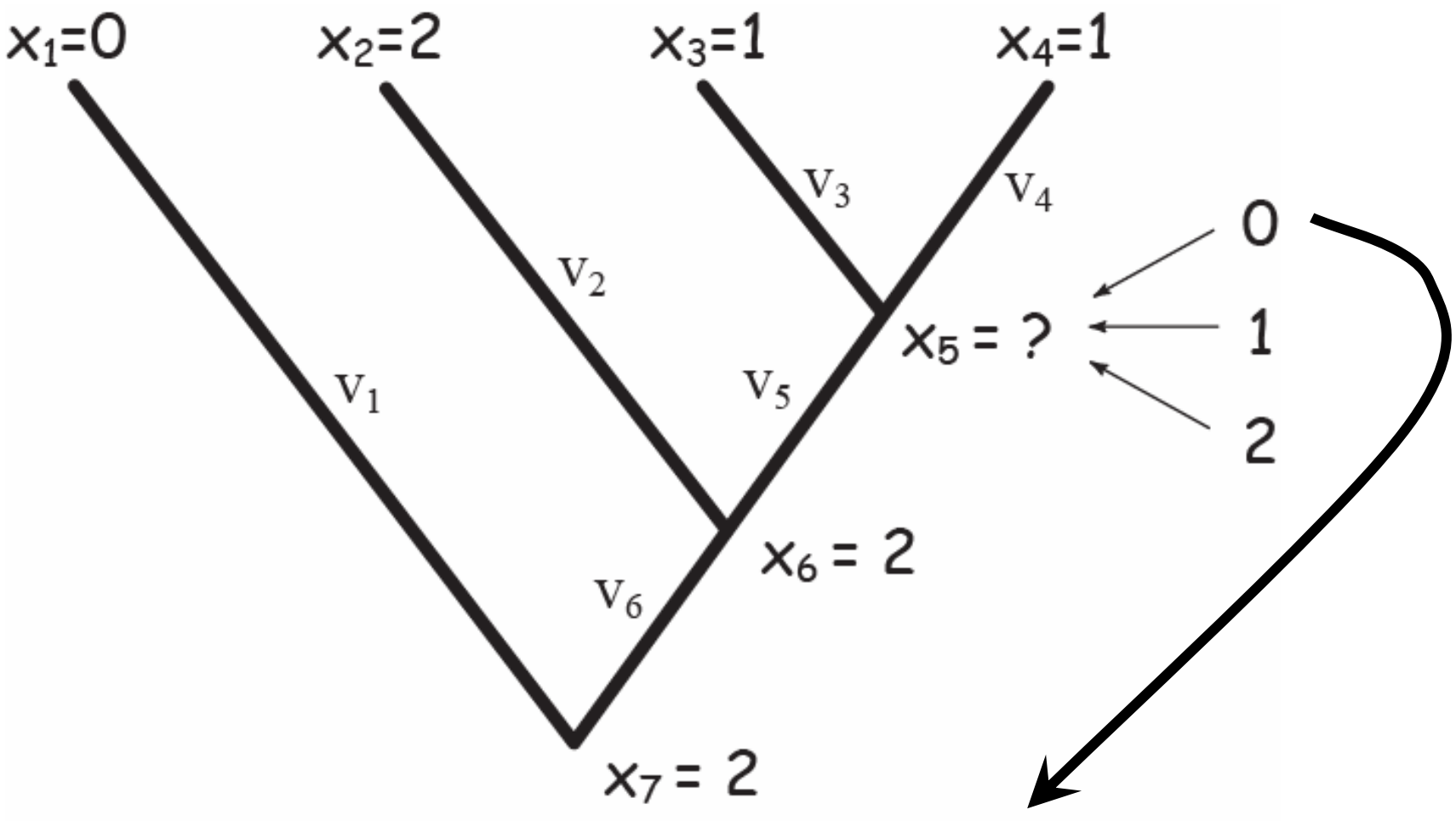


$$\Pr(x_6 = 0|x_2, x_3, x_4) \propto \Pr(x_2, x_3, x_4|x_6 = 0)\ \Pr(x_6 = 0|x_7 = 2)$$

© 2007 by Paul O. Lewis

Note: only the calculation for x_6=0 is presented (but you must do a similar calculation for all three states)

Sampling state at node 5



$$\Pr(x_5 = 0|x_3, x_4) \propto \Pr(x_3, x_4|x_5 = 0)\ \Pr(x_5 = 0|x_6 = 2)$$

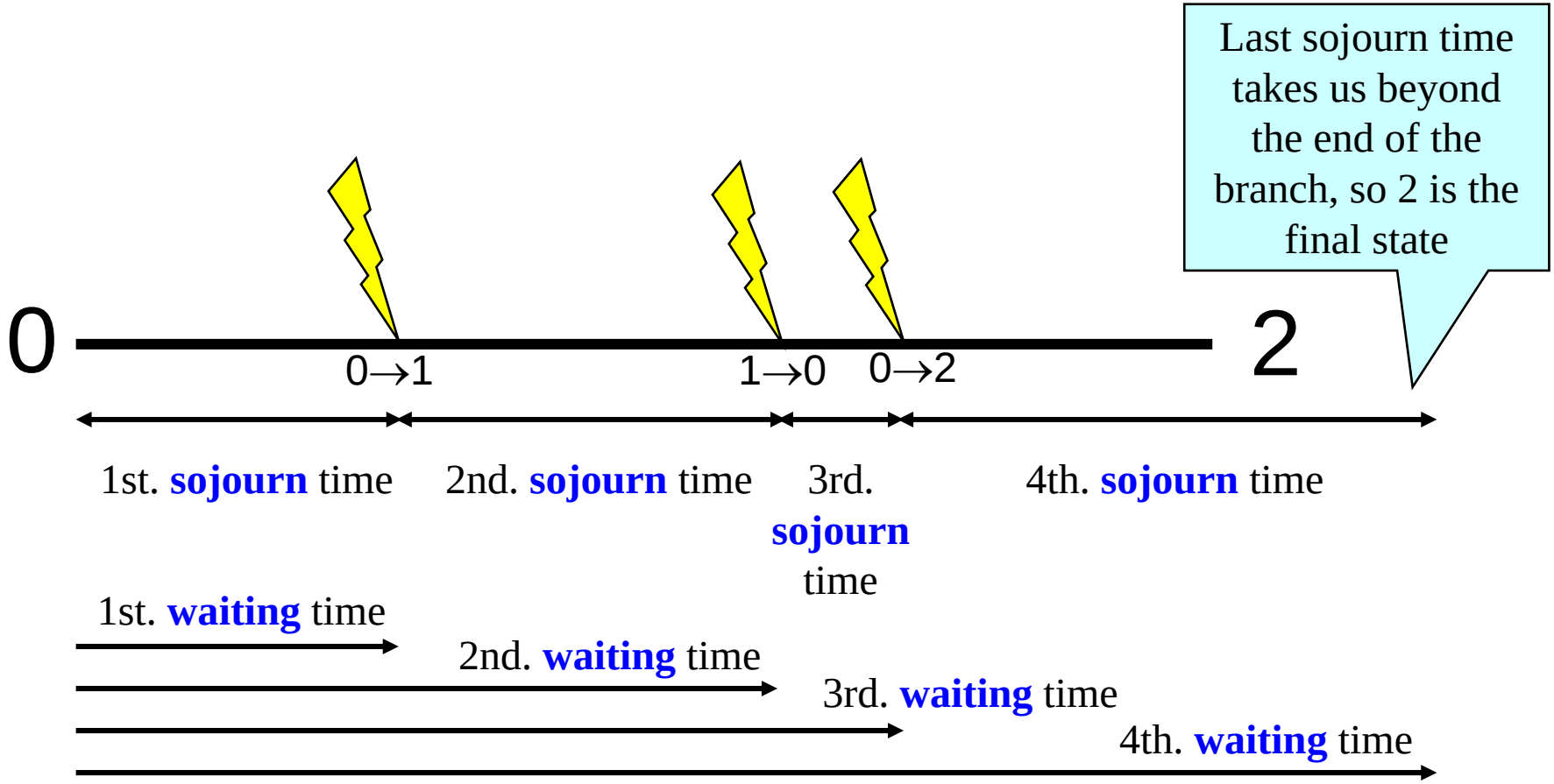
© 2007 by Paul O. Lewis

Note: only the calculation for x_5=0 is presented

"Mapping" edges

- **Mapping an edge** means choosing an evolutionary history for a character along a particular edge that is consistent with the **observed states** at the tips, the **chosen states** at the interior nodes, and the **model of character change** being assumed
- This is done by choosing random **sojourn times** and "causing" a change in state at the end of each of these sojourns
- What is a sojourn time? It is the time between substitution events, or the time one has to wait until the next change of character state occurs

© 2007 by Paul O. Lewis



Sojourn times are related to substitution rates. High substitution rates mean short sojourn times, on average. A **dwell** time is the total amount of time spent in one particular state. For example, the dwell time associated with state 0 is the sum of the 1st. and 3rd. sojourn times.

© 2007 by Paul O. Lewis

Substitution rates and sojourn times

Symmetric model for character with 3 states: 0, 1 and 2

	0	1	2
0	-2 β	β	β
1	β	-2 β	β
2	β	β	-2 β

The rate at which state 0 changes to something else (either state 1 or state 2) is 2β

Note that 2βt equals the branch length. For these symmetric models, the branch length is (k-1)βt where k is the number of states.

© 2007 by Paul O. Lewis

Sojourn times are exponentially distributed

If λ is the total rate of leaving state 0, then you can simulate a soujourn by:

- drawing a time according to the cumulative distribution function of the exponential distribution $(1 - e^{-\lambda t})$.
- pick the "destination" state by normalizing the rates away from state 0 such that the rates sum to one.

This is picking j and t by:

$$\Pr(0 \to j, t) = \Pr(\text{mut. at } t) \Pr(j|\text{mut.})$$

But we want quantities like:

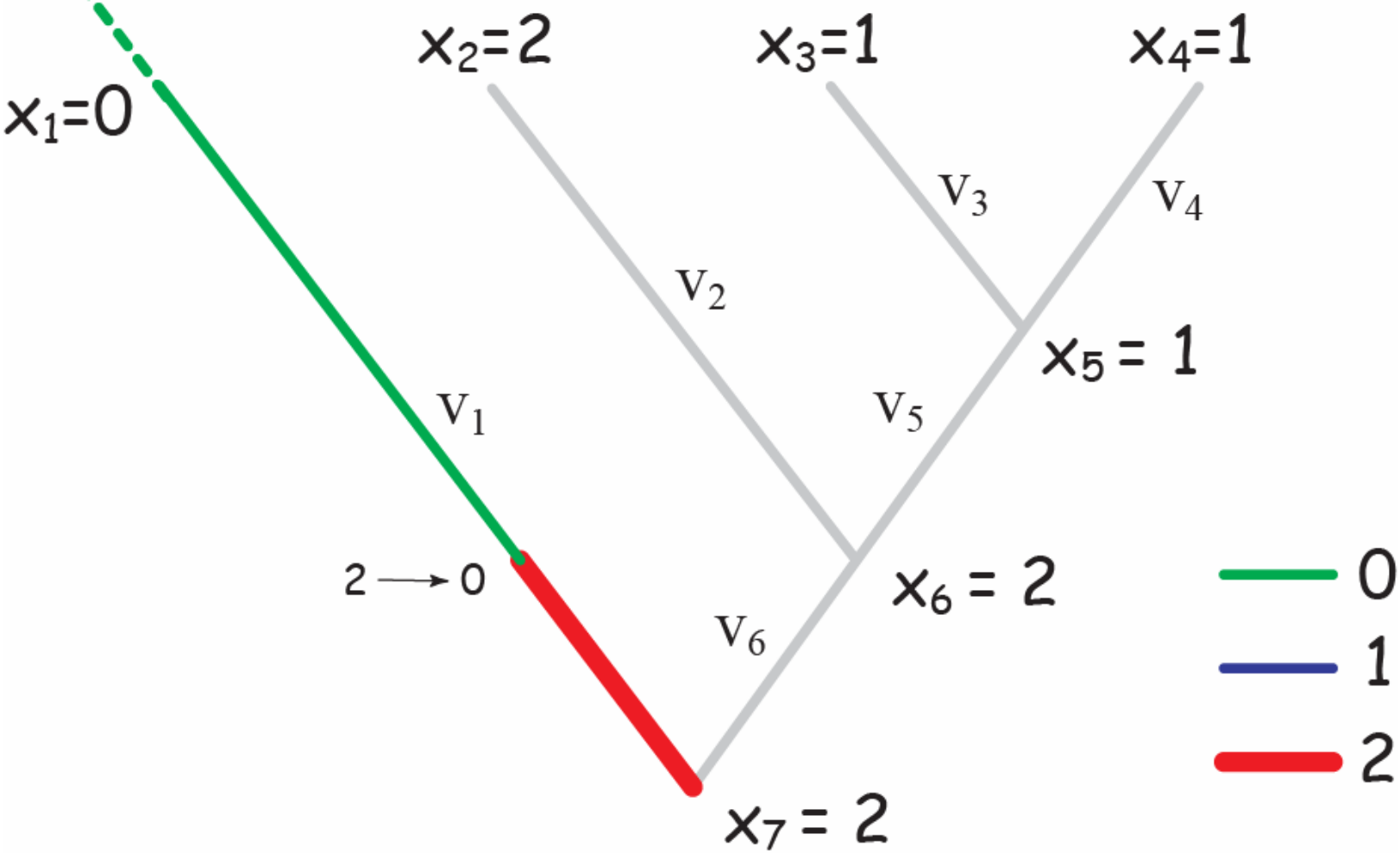
$$\Pr(0 \to j, t|0 \to 2, \nu_1)$$

Conditioning on the data

We simulate according to $\Pr(0 \to j, t | 0 \to 2, \nu_1)$ by rejection sampling:

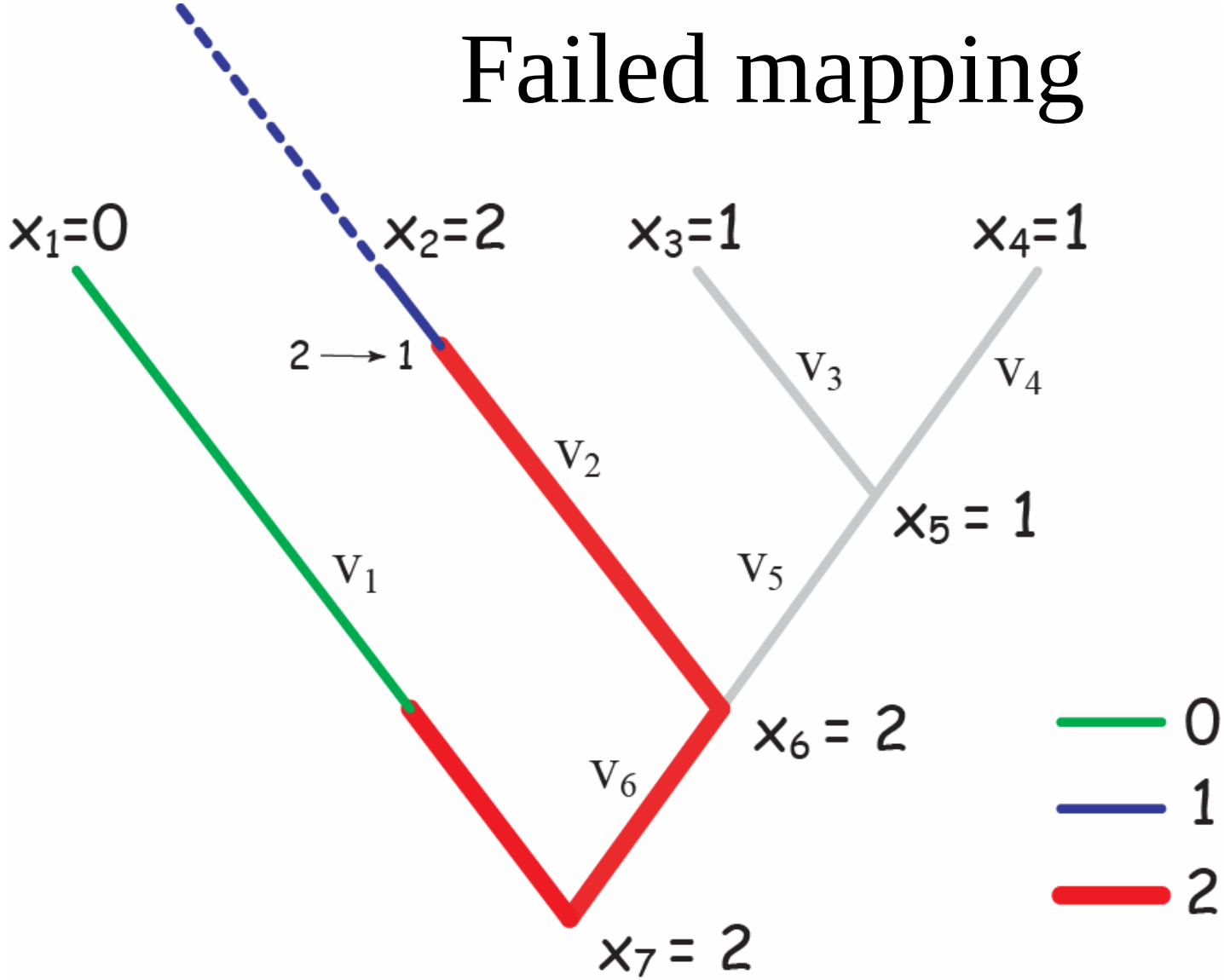
1. Simulate a series of draws of the form: $\Pr(0 \to j, t)$ until the sum of t exceeds ν_1,
2. If we ended up in the correct state, then accept this series of sojourns as the mapping.
3. If not, go back to step 1.

"Mapping" edge 1



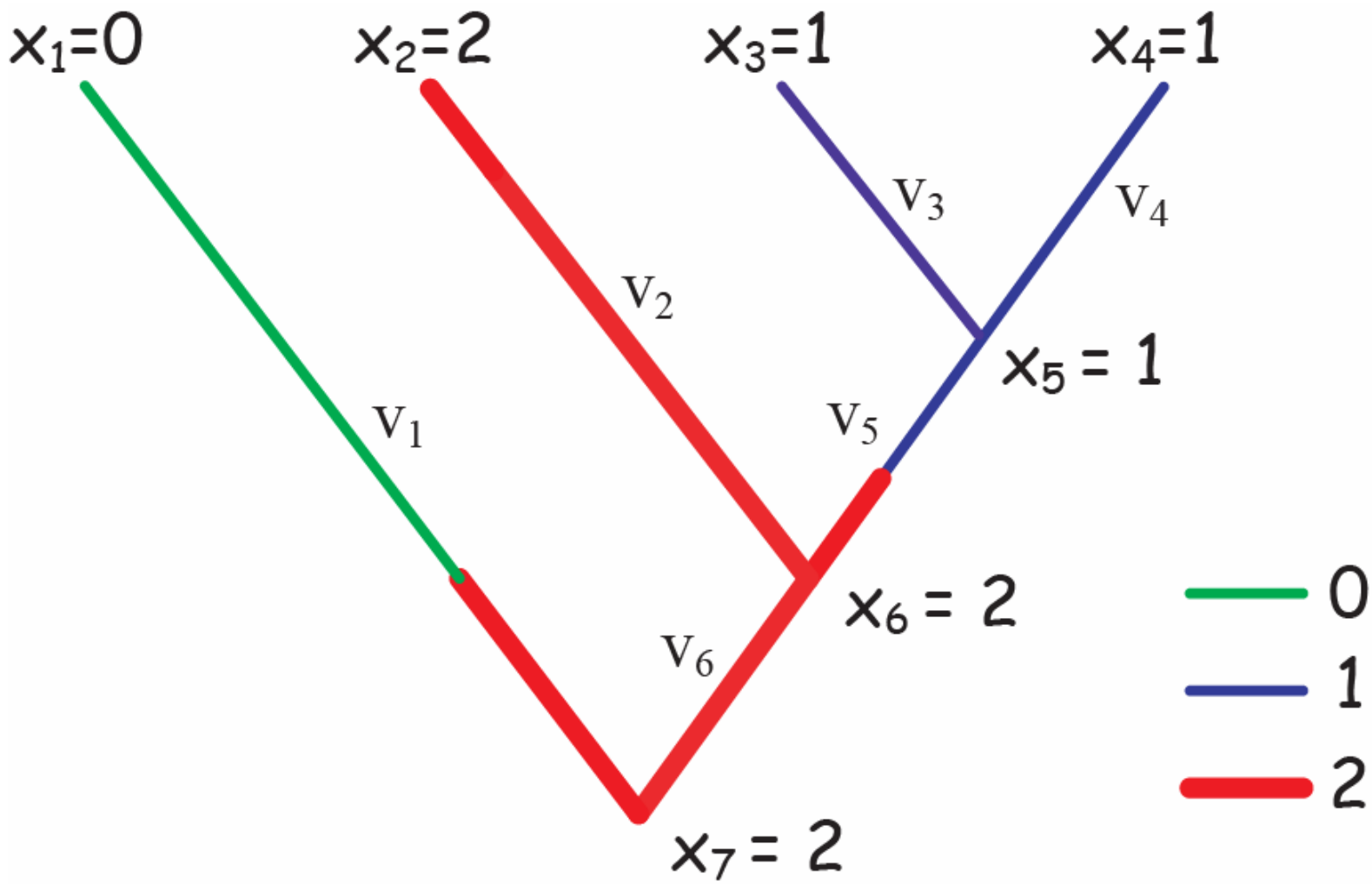
© 2007 by Paul O. Lewis

Failed mapping



© 2007 by Paul O. Lewis

Completed mapping



© 2007 by Paul O. Lewis

"Stochastic character mapping" or "Mutational Mapping"

- allows you to sample from the posterior distribution of histories for characters
- conditional on an assumed model
- conditional on the observed data
- implemented in Simmap (v1 by J. Bollback; v2 by J. Bollback and J. Yu) and Mesquite (module by P. Midford)

"Stochastic character mapping"

You can calculate summary statistics on the set of simulated histories to answer almost any type of question that depends on knowing ancestral character states or the number of changes:

- What proportion of the evolutionary history was 'spent' in state 0?
- What is the posterior probability that there were between 2 and 4 changes of state?
- Does the state in one character appear to affect the probability of change at another? – Map both characters and do a contingency analysis.

See Minin and Suchard (2008a,b); O'Brein et al. (2009) for rejection free approaches to generate mappings.

Often we are more interested in the "forces" driving evolutionary patterns than the outcome.

Perhaps we should focus on fitting a model of character evolution rather than the ancestral character state estimates.

Rather than:

1. estimate a tree,
2. estimate character states,
3. figure out what model of character evolution best fits those ancestral character state reconstructions.

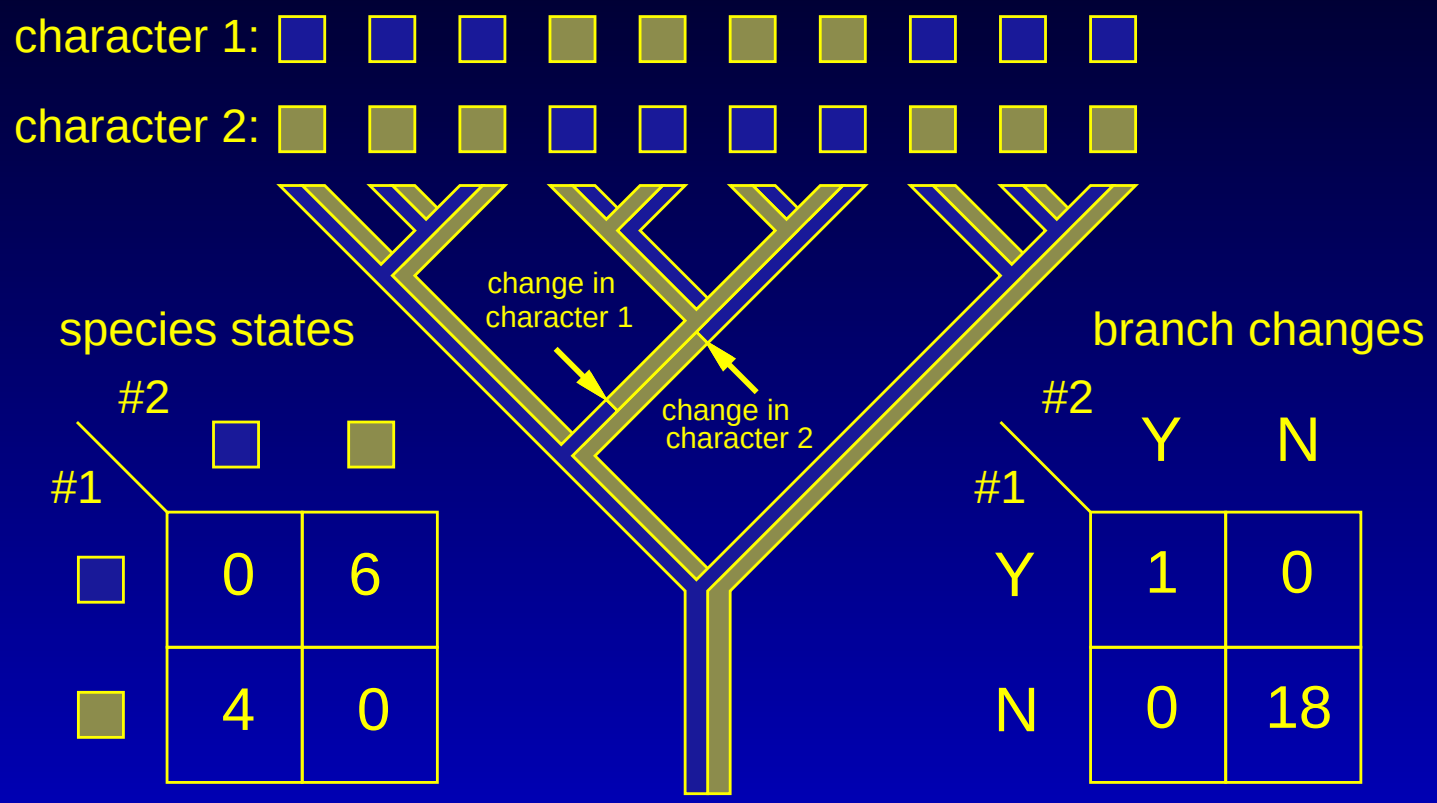
Rather than:

1. estimate a tree,
2. figure out what model of character evolution best fits tip data

Or even:

1. Jointly estimate character evolution in trees in MCMC (marginalizing over trees when answering questions about character evolution).

Correlation of states in a discrete-state model



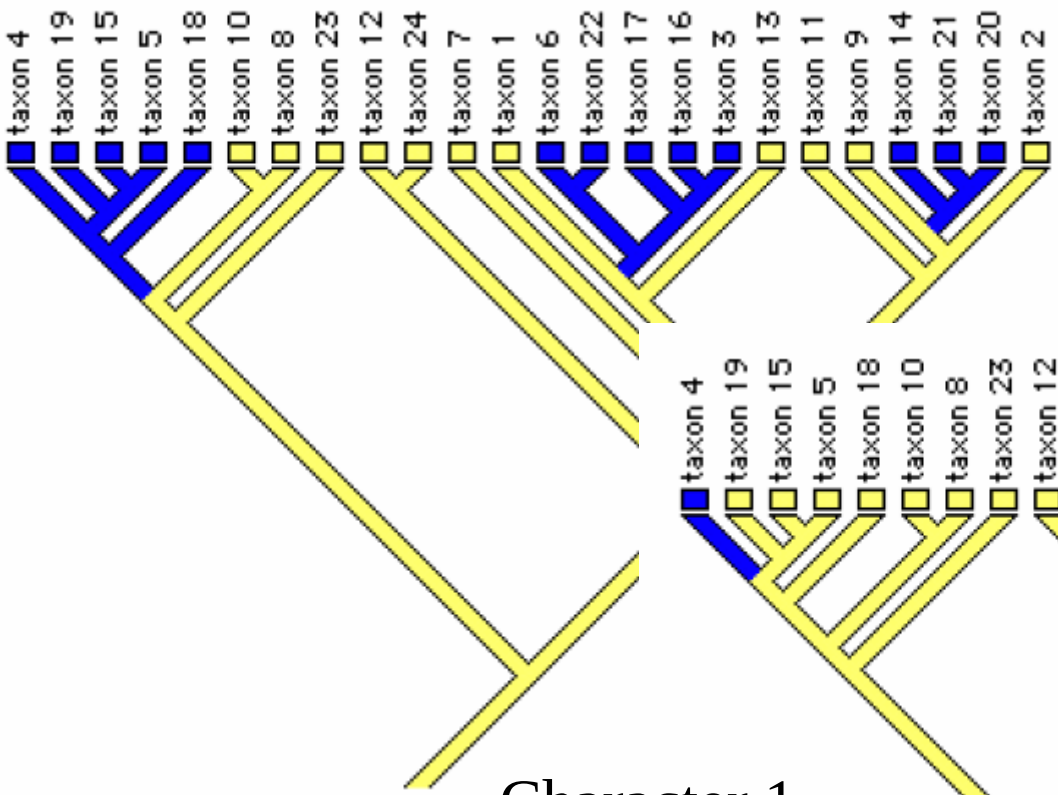
species states

#1 \ #2		
	0	6
	4	0

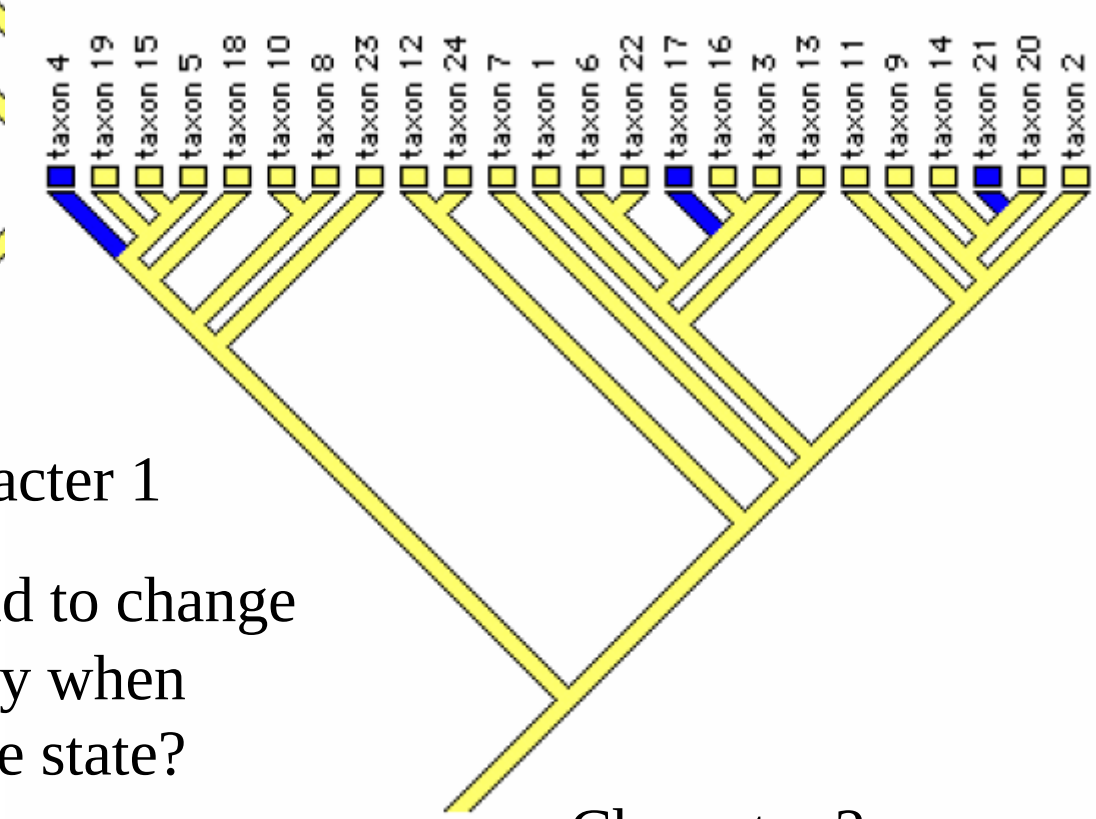
branch changes

#1 \ #2	Y	N
Y	1	0
N	0	18

Example from pp. 429-430 of the MacClade 4 manual.



Character 1

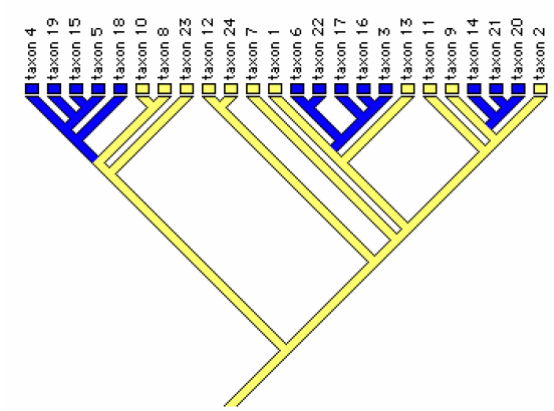


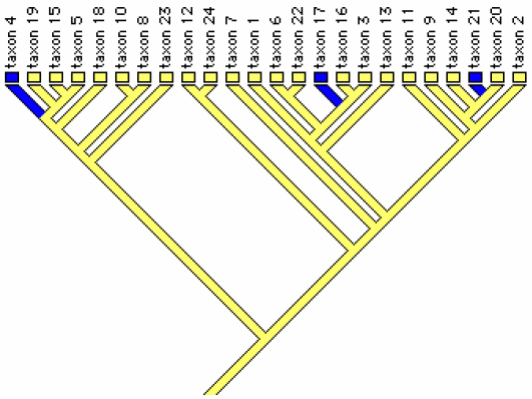
Character 2

Q: Does character 2 tend to change from yellow →blue only when character 1 is in the blue state?

© 2007 by Paul O. Lewis

Concentrated Changes Test





More precisely: how often would three yellow→blue changes occur in the blue areas of the cladogram on the left if these 3 changes were *thrown down at random* on branches of the tree?

Answer: 12.0671% of the time. Thus, we cannot reject the null hypothesis that the observed coincident changes in the two characters were simply the result of chance.

© 2007 by Paul O. Lewis

Model-based approach to detecting correlated evolution

View the problem as model-selection or use LRT if you prefer the hypothesis testing framework.

Consider two binary characters. Under the simplest model, there is only one rate:

		0 (0,0)	1 (0,1)	2 (1,0)	3 (1,1)
0	(0,0)	-	q	q	0
1	(0,1)	q	-	0	q
2	(1,0)	q	0	-	q
3	(1,1)	0	q	q	-

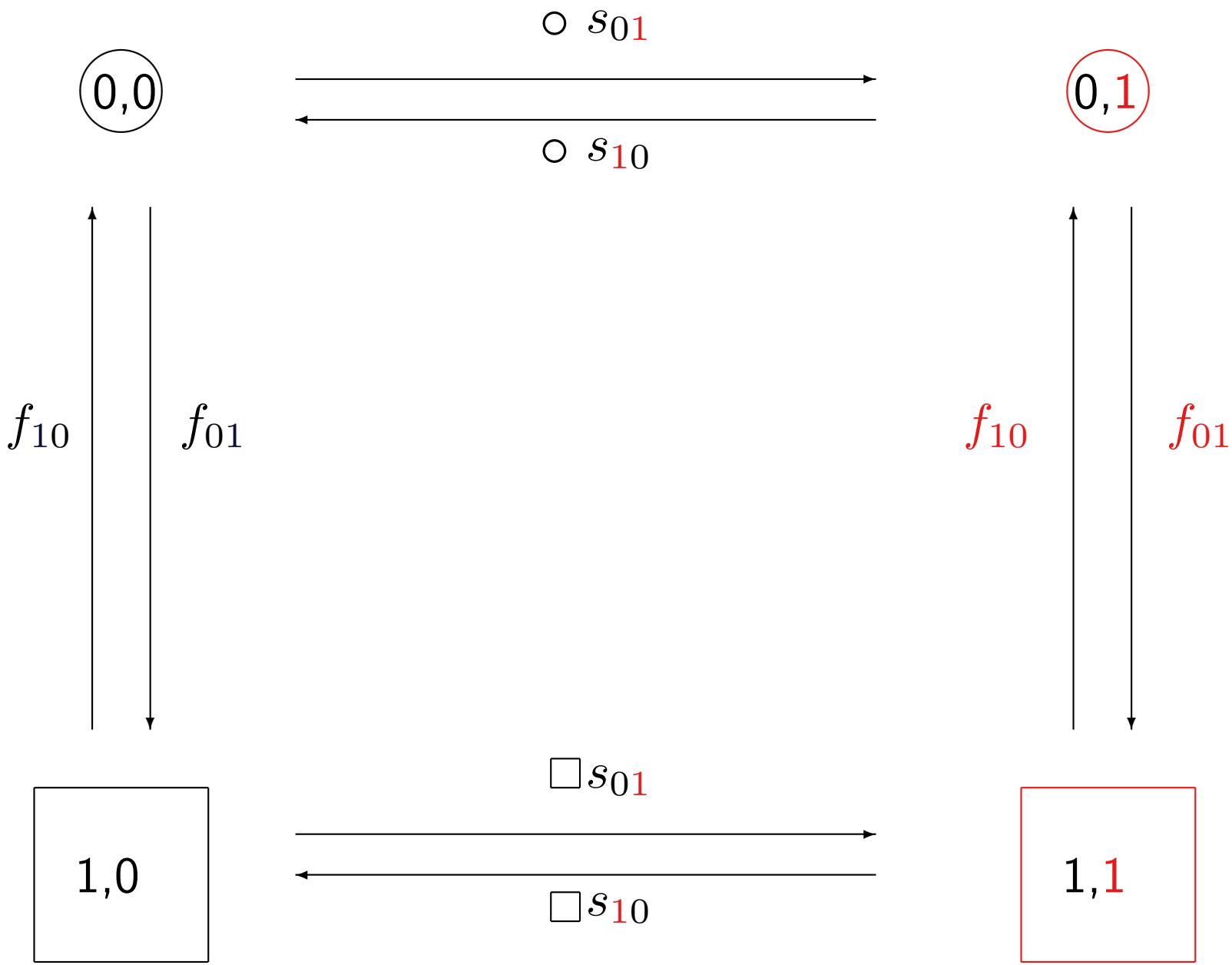
Why are there 0's?

Under a model assuming character independence:

		0 (0,0)	1 (0,1)	2 (1,0)	3 (1,1)
0	(0,0)	-	f_{01}	s_{01}	0
1	(0,1)	q_{*0}	-	0	q_{1*}
2	(1,0)	q_{0*}	0	-	q_{*1}
3	(1,1)	0	q_{0*}	q_{*0}	-

Under a the most general model:

		0 (0,0)	1 (0,1)	2 (1,0)	3 (1,1)
0	(0,0)	-	q_{01}	q_{02}	0
1	(0,1)	q_{10}	-	0	q_{13}
2	(1,0)	q_{20}	0	-	q_{23}
3	(1,1)	0	q_{31}	q_{32}	-


s_{01}
0,0
0,1
s_{10}
f_{10}
f_{01}
f_{10}
f_{01}
s_{01}
1,0
1,1
s_{10}

We can calculate maximum likelihood scores under any of these models and do model selection.

Or we could put priors on the parameter values and using Bayesian model selection (*e.g.* reversible jump methods in BayesTraits).

References

Minin, V. N. and Suchard, M. A. (2008a). Counting labeled transitions in continuous-time markov models of evolution. *J. Math. Biol*, 56:391–412.

Minin, V. N. and Suchard, M. A. (2008b). Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond, B, Biol Sci*, 363:3985–3995.

O'Brein, J. D., Minin, V., and Suchard, M. (2009). Learning to count: Robust estimates for labeled distances between molecular sequences. *Molecular Biology and Evolution*, 26(4):801–814.