"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"
Integrative Biology 200 Spring 2014
University of California, Berkeley D.D. Ackerly

March 3, 2014. **Maximum Likelihood Principles, and Applications to Discrete Characters**

I. Introduction

Maximum likelihood (ML) principles provide a fundamental alternative to parsimony in the reconstruction of phylogenies and estimation of ancestral states. ML also represents an important shift in thinking from standard probabilistic statistics. In standard statistics we focus on the probability of a given observation under a null hypothesis. If the actual observations are considered very unlikely, then we reject the null hypothesis. However, we don't actually accept a particular alternative hypothesis. For example, consider a t-test of the following observations:

Treatment A: 5, 8, 10 (mean = 7.7, sd = 2.5)
Treatment B: 8, 12, 15 (mean = 11.7, sd = 3.5)

Null hypothesis:
• Assume that observations represent a finite sample from a normal distribution (in other words, the process in the natural world that generates these data would generate a normal distribution if you collected an infinite sample)
• assume that the means and variances of those distributions are equal in the two samples
• if these assumptions are true, the probability of drawing two samples that differ by as much or more as the two above is $\Pr(d|H_0) = 0.18$.
• note that we can describe the full distribution of possible outcomes, in terms of the differences between the two groups, which will be a normal distribution that sums to 1.

Maximum likelihood reverses the entire process. Let's assume that our data are real and true, and they reflect the outcome of some unknown process or model. Can we calculate the likelihood of the model, given these data, and compare that to the likelihood of alternative models. We are searching for the maximum likelihood model – the model of the world that best fits the data. The problem is that there are an infinite number of possible models, so unlike probability we can't describe the entire likelihood space as a distribution that sums to 1. So how can we calculate their relative likelihoods? The fundamental insight (Edwards 1972) that makes ML statistics possible is that:

$$L(m \mid d) \propto \Pr(d \mid m)$$

Thus, we can obtain relative likelihoods of alternative models and compare them. One of those alternatives could be the traditional null hypothesis, leading to the same significance test, but in general the ability to specify a range of alternatives enhances our ability to explore specific hypotheses with the data. One of the main drawbacks is that it can be quite difficult in some cases to find the best model under ML, if there is no analytic solution.

Likelihood estimation of ancestral states and rates of character evolution on a phylogeny:

We'll start with one of the simplest problem in phylogenetics: estimating the rates of character evolution for a binary trait. First consider the problem of evolution along a single branch:

α = the instantaneous forward transition rate from 0 -> 1
β = the instantaneous reverse transition rate from 1 -> 0

With a little calculus one can show that the probabilities of change along a branch of length t are:

	To:	
From:	0	1
0	$P_{00} = 1 - P_{01}$	$P_{01} = \frac{\alpha}{\alpha+\beta}\left(1-\exp\left[-(\alpha+\beta)t\right]\right)$
1	$P_{10} = \frac{\beta}{\alpha+\beta}\left(1-\exp\left[-(\alpha+\beta)t\right]\right)$	$P_{11} = 1 - P_{10}$

If one assumes that the backward and forward transition rates are the same, P_{01} and P_{10} also are the same and simplify considerably. As a quick exercise in R, write a function that returns the matrix of results, like those below, based on values of α, β and t.

Examples:
$\alpha = 0.5, \beta = 0.5, t = 1$

From: To:	0	1
0	0.684	0.316
1	0.316	0.684

$\alpha = 0.8, \beta = 0.2, t = 1$

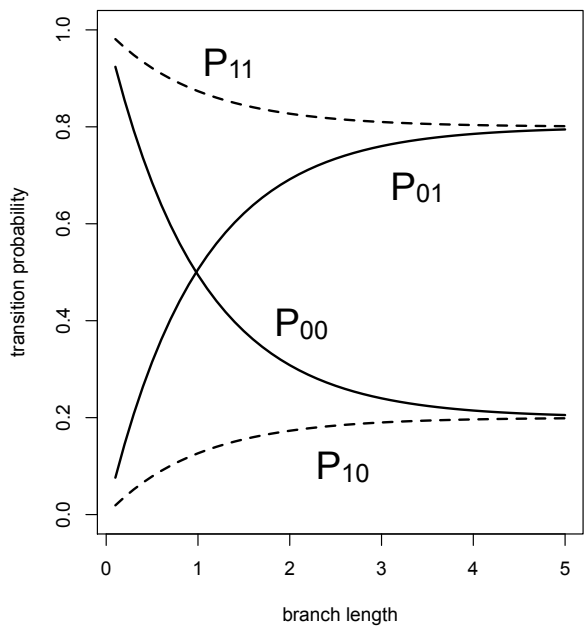
From: To:	0	1
0	0.494	0.506
1	0.126	0.874

$\alpha = 0.8, \beta = 0.0, t = 1$

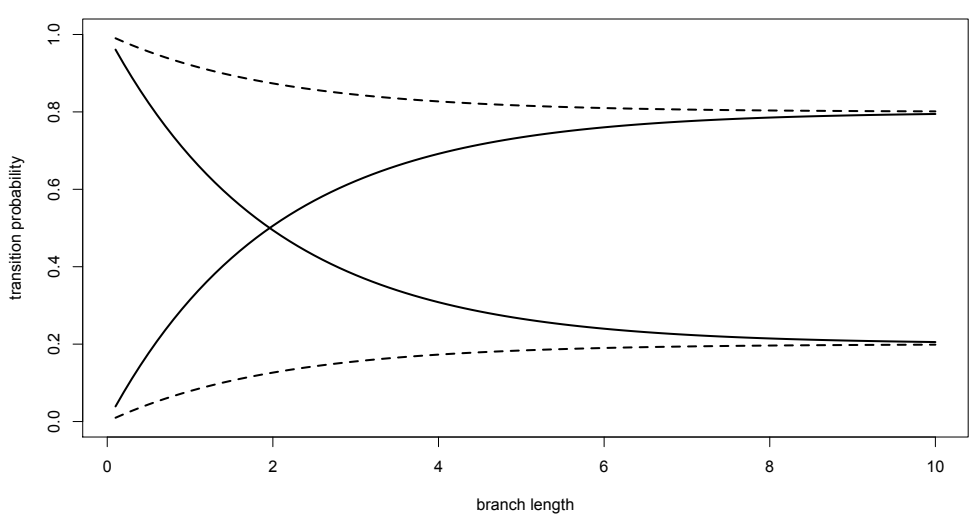
From: To:	0	1
0	0.449	0.551
1	0.0	1.0

The first REALLY important thing about the maximum likelihood view of evolutionary change is that branch lengths matter (unlike parsimony). Given instantaneous rates of change, α and β, a branch will eventually converge on a probability α that it ends in state 1 and probability β that it ends in state 0, regardless of the initial state. Another R exercise – reproduce these graphs!

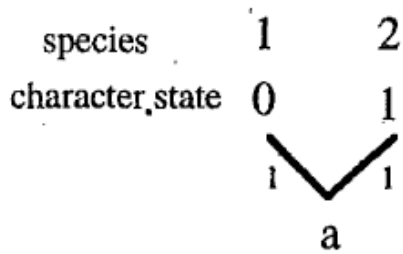
E.g. $\alpha = 0.8$, $\beta = 0.2$



If both rates are lower, but with a the same ratio to each other, the branches will converge to the same point, but it will take longer. For $\alpha = 0.4$, $\beta = 0.1$

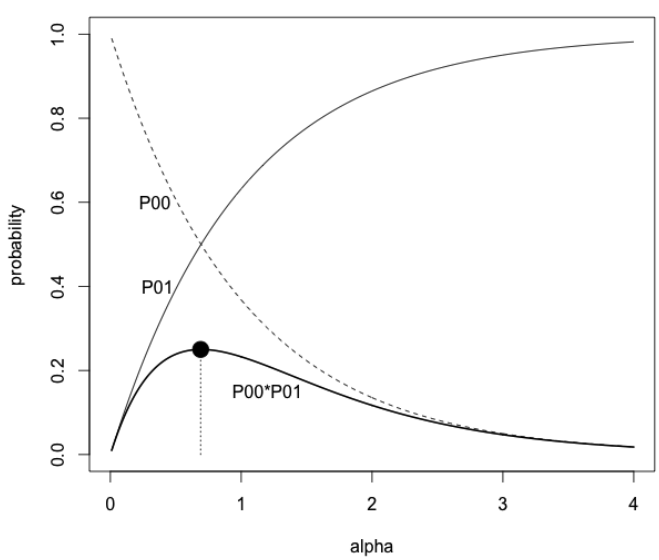


Now, let's consider the simplest ML problem: estimating the rates of change, and the ancestral states, for a binary trait on a phylogeny with two terminal taxa and their ancestor (a):

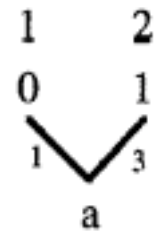


Two solve this and more complex problems we will consider later, we first have to recall two important rules of probability:
1) the joint probability that two independent events (A and B) will both occur is the product of their individual probabilities: P(A and B) = P(A) * P(B)
2) the probability that either of two independent events will occur is the sum of their individual probabilities: P(A or B) = P(A) + P(B)

So, to find the ML solution for the ancestral state a, we first need to solve for the ML values of the α and β rate parameters under each possible scenario (a = 0 and a = 1). Let's first examine the best solution, assuming the ancestral state a = 0. In that case, you can see that the best solution will have $\beta = 0$ and some positive α value, because we need to get a transition from 0 to 1, but it won't help if it switches back. But α can't be too high, because then the branch leading to species 1 won't stay in state 0. So, we can calculate P_{00}, P_{01} and the joint probability (P00 * P01) across a range of possible α values. As you can see, the ML solution is $\alpha = 0.69$ and the overall probability of the data given this model = 0.25. As you can imagine, if we switch and consider the case where a = 1, then the ML values are $\alpha = 0$ and $\beta = 0.69$. In other words, it's the symmetrical result, and thus we have discovered that the relative likelihood of the two ancestral values are the same!

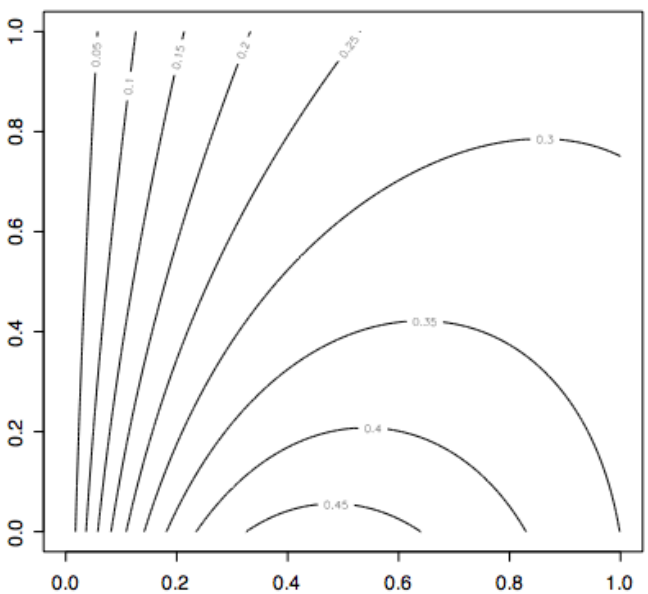


Now, it gets more interesting if the branch lengths are unequal.



For a = 0, the ML solution is $\alpha = 0.46$ and $\beta = 0$.
P00 = 0.63
P01 = 0.75

overall L $\propto$ 0.472



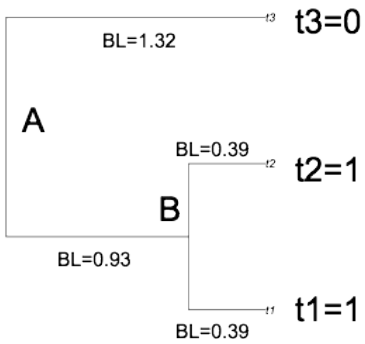
For a = 1, the ML solution is $\alpha = \inf$, $\beta = \inf$
P10 = 0.5
P11 = 0.5

overall L $\propto$ 0.25

So in this case, the ancestral state value is most likely to be 0, and the relative likelihood of a = 0 vs. a = 1 is about 1.7.

A 3-taxon example

phylogeny and species trait values



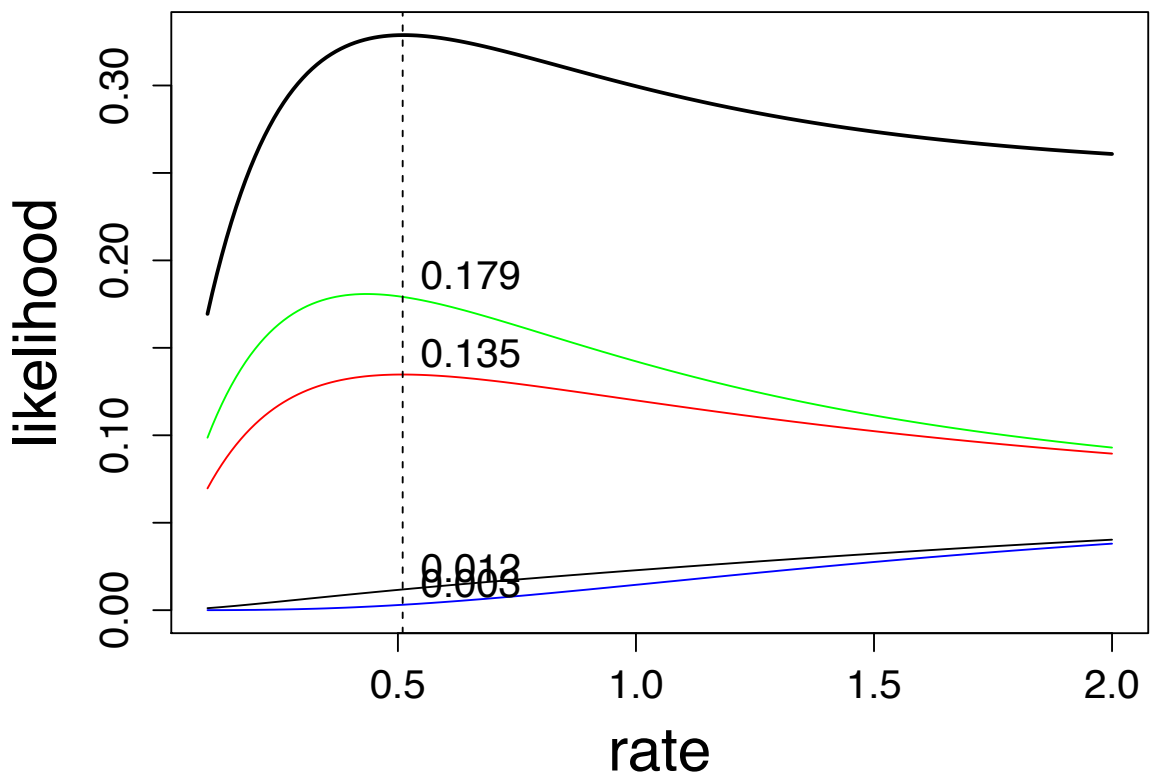
What is the likelihood of any one set of ancestral states given a hypothesized transition rate (assume equal forwards and backwards rates):

Let A = 0, B = 1, alpha = beta = 0.3

overall likelihood of this combination is the product of the individual likelihoods on each branch:

A->B = $P_{01,0.93}$ = 0.213
B->t1 = $P_{11,0.39}$ = 0.896
B->t2 = $P_{11,0.39}$ = 0.896
A->t3 = $P_{00,1.32}$ = 0.726

prod = 0.1246



Likelihood ratio tests

a measure of support for alternative hypotheses

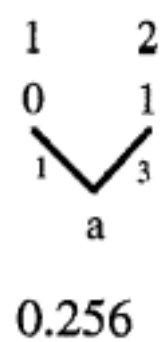
$LR = -2*\ln(L_1/L_2)$

For two hypotheses with the same number of parameters, there is no exact significance value attached to the LR. Values greater than 2 are considered 'strong support'

For nested hypotheses with different numbers of parameters, LR is distributed as a chi-square with df = the difference in number of parameters

For example:
If we find the maximum likelihood with alpha and beta fit independently, calculated over both possible ancestral states (0 and 1):

alpha = 0.59
beta = 0.31
L(m) $\propto$ 0.256.



If we allow only one transition rate, such that alpha=beta, then:

alpha = beta = inf
L $\propto$ 0.25

LR = -2*ln(.25/.256) = 0.05
chisq(0.05, df=1) = 0.82

So these data are insufficient to reject a single rates model.

Pagel's (1994) discrete test of correlated evolution:

Same idea as above, but test for parameters of dependence in trait change.

$$
\boldsymbol{I}+\boldsymbol{Q}_{\mathrm{D}} = \begin{array}{c} X, Y \\ 0,0 \\ \\ 0,1 \\ \\ 1,0 \\ \\ 1,1 \end{array}
\begin{array}{c}
\begin{array}{cccc} 0,0 & 0,1 & 1,0 & 1,1 \end{array} \\
\begin{bmatrix}
1-(q_{12}+q_{13}) & q_{12} & q_{13} & 0 \\
q_{21} & 1-(q_{21}+q_{24}) & 0 & q_{24} \\
q_{31} & 0 & 1-(q_{31}+q_{34}) & q_{34} \\
0 & q_{42} & q_{43} & 1-(q_{42}+q_{43})
\end{bmatrix}
\end{array},
$$

For example:
q_{12} is the probability that trait 2 changes from 0 -> 1, when trait 1 = 0
q_{34} is the probability that trait 2 changes from 0 -> 1, when trait 1 = 1

For an instantaneous model of change, assume they don't change simultaneously

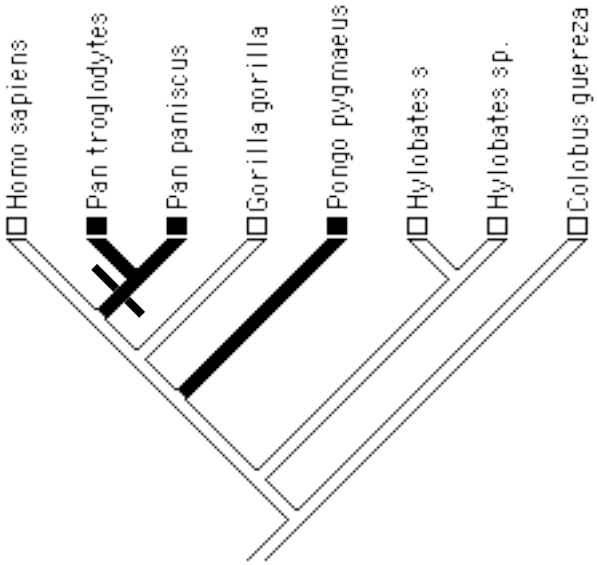
Model with full dependence has 8 parameters

If traits evolve independently, there are only 4 parameters, because:
$q_{12} = q_{34}$; $q_{13} = q_{24}$; $q_{31} = q_{42}$; $q_{21} = q_{43}$

Table 1. *Categories of hypothesis test using $L(I)$ and $L(D)$*

(The tests shown in the table do not exhaust the range of possible tests but represents ones likely to be of interest. Tests with more than 1 d.f. are also possible. The contingent change tests, temporal order/relative rate tests, and branch length transformation tests will have asymptotic χ^2 distributions but may also be tested by Monte Carlo simulation. The omnibus test, and tests of alternative models will not in general be distributed as χ^2. The test of branch length transformation can be applied to any likelihood, not just to the full model. Subscripts to $L(D)$ denote the number of parameters.)

test	hypothesis	description	statistic
omnibus χ^2 (4 d.f.)	$L(I) = L(D)$	correlated evolution	$L(I)$ versus $L(D)$
contingent change (1 d.f.)	$q_{12} \neq q_{34}$	change in Y depends upon state of X	$L(D_7)$ versus $L(D_8)$
	$q_{31} \neq q_{42}$	change in Y depends upon state of X	$L(D_7)$ versus $L(D_8)$
	$q_{12} \neq q_{34}$	change in X depends upon state of Y	$L(D_7)$ versus $L(D_8)$
	$q_{31} \neq q_{42}$	change in X depends upon state of Y	$L(D_7)$ versus $L(D_8)$
temporal order or relative rate (1 d.f.)	$q_{12} \neq q_{13}$	order of acquisition of X versus Y	$L(D_7)$ versus $L(D_8)$
	$q_{34} \neq q_{43}$	order of loss of X versus Y	$L(D_7)$ versus $L(D_8)$
alternative models	$q_{ij} = 0$	q_{ij} transitions excluded	$L(D_7)$ versus $L(D_8)$
branch transformation	$x \neq 1$	transformation of branch lengths	$L(D_n)$ versus $L(D_{n+1})$



ln(L(independent)) = -11.91
ln(L(dependent)) = -8.43
-2*ln(LI/LD) = 6.96
chisq(6.96,4) = 0.14

q12 = 0.29 gain of oestrous swellings in single-male breeding systems
q34 = 3.45 gain of OS in multi-male breeding systems
q13 = 1.87 gain of multi-male BS in absence of OS

CITATIONS:

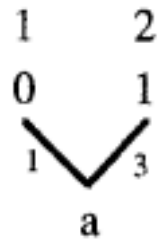
Pagel M. (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B, 255, 37-45

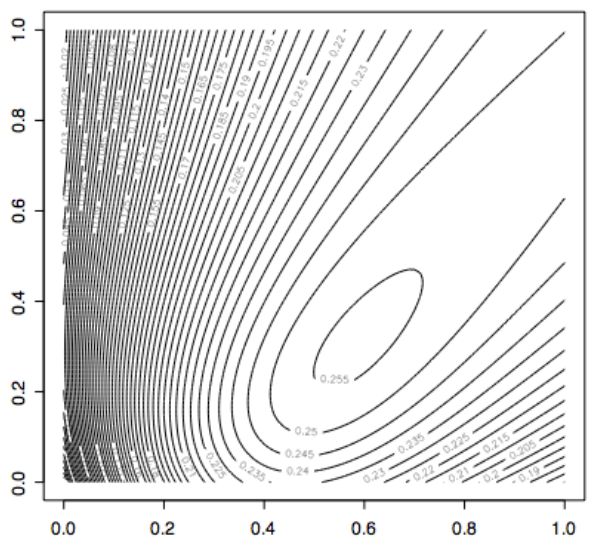
Pagel M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol., 48, 612-622

Schluter D., Price T., Mooers A. & Ludwig D. (1997) Likelihood of ancestor states in adaptive radiation. Evolution, 51, 1699-1711

THESE EXAMPLES USED IN 2007; IN 2009 REPLACED WITH EXAMPLE ON THREE TAXON TREE, IN FILE 'DISC ML.KEY'

What if one branch is longer than the other, providing more opportunity for change. Now, we find a finite solution from the ML surface, with alpha and beta < 1:





The solution is alpha = 0.59, beta = 0.31:

if ancestor = 0:
P00 = 0.61 (t = 1)
P01 = 0.61 (t = 3)

if ancestor = 1:
P10 = 0.20
P11 = 0.68

overall:
P = (0.5)(0.61*0.61) + (0.5)(0.2*0.68) = 0.256

This is somewhat counter intuitive that the forward transition rate is higher than the reverse. The individual probability of going from a = 1 to d = 0 is low, but all of the other 3 probabilities are quite high so this is the best overall solution.

As a final step, let us relax the assumption that the ancestor could be in either state. Instead, we can find the ML solution for alpha and beta, given that a = 0 or that a = 1, and therefore work out which ancestral state has a higher likelihood overall: