Maximum Likelihood:

Maximum likelihood is a general statistical method for estimating unknown parameters of a probability model. A parameter is some descriptor of the model. A familiar model might be the normal distribution with two parameters: the mean and variance. In phylogenetics there are many parameters, including rates, differential transformation costs, and, most important, the tree itself

Likelihood is defined to be a quantity proportional to the probability of observing the data given the model, P(D|M). Thus, if we have a model, we can calculate the probability the observations would have actually been observed as a function of unknown parameters (like the branching pattern). We then examine this likelihood function to see where it is greatest, and the value of the parameter (the tree) at that point is the maximum likelihood estimate of the parameter.

Simple Coin Flip example:

The likelihood for heads probability p for a series of 11 tosses assumed to be independent-

HHTTHTHHTTT 5 heads (p), 6 tails (1-p)

$L= P(D|p)= pp(1-p)(1-p)p(1-p)pp(1-p)(1-p)(1-p)$

ML= 0.45454 = 5/11 This can be plotted i.e. brute force determination, or calculated by taking the derivative of the plot and looking for where the slope = 0.

Maximum Likelihood can be used as an optimality measure for choosing a preferred tree or set of trees. It evaluates a hypothesis (branching pattern), which is a proposed evolutionary history, in terms of the probability that the implemented model and the hypothesized history would have given rise to the observed data set. Essentially a pattern that has a higher probability preferred to one with lower probability.

Advantages and disadvantages of maximum likelihood methods:

Supposed advantages.
-lower variance than other methods (i.e. estimation method least affected by sampling error)
-robust to many violations of the assumptions in the evolutionary model, even with very short sequences it may outperform alternative methods such as parsimony or distance methods.
-the method is statistically well understood
-has explicit model of evolution
-evaluate different tree topologies (vs. NJ)
-use all the sequence information (vs. Distance)

Supposed disadvantages.
-very computationally intensive and so extremely slow
-the result is dependent on the model of evolution used
-philosophically less well established, especially in terms of probabilities and statistical measures of unique historical events (vs. Parsimony as a general principle)
"Although the true phylogeny is "unknowable" it can nonetheless be estimated…" Swofford et al.

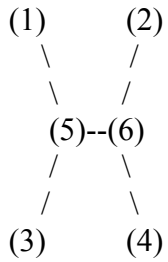
Branch lengths ML and Parsimony- The use of probability may provide information based on sites that would be uninformative under parsimony.

Simple tree example:

Assume that we have the aligned nucleotide sequences for four taxa:

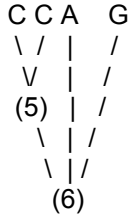
	1	**j**	....N
(1)	A G G C T	**C**	C A A ....A
(2)	A G G T T	**C**	G A A ....A
(3)	A G C C C	**A**	G A A.... A
(4)	A T T T C	**G**	G A A.... C

and we want to evaluate the likelihood of the unrooted tree represented by the nucleotides of site **j** in the sequence and shown below:



What is the probability that this tree would have generated the data presented in the sequence under the chosen model?

Since most of the models currently used are **time-reversible**, the likelihood of the tree is generally independent of the position of the root. Therefore it is convenient to root the tree at an arbitrary internal node as done below,



Under the assumption that nucleotide sites evolve independently (the Markovian model of evolution), we can calculate the likelihood for each site separately and combine the likelihood into a total value at the end. To calculate the likelihood for site **j**, we have to consider all the possible scenarios by which the nucleotides present at the tips of the tree could have evolved. So the likelihood for a particular site is the summation of the probabilities of every possible reconstruction of ancestral states, given some model of base substitution. So in this specific case all possible nucleotides A, G, C, and T occupying nodes (5) and (6), or 4 x 4 = 16 possibilities (see fig 10D from textbook).

Since any one of these scenarios could have led to the nucleotide configuration at the tip of the tree, we must calculate the probability of each and sum them to obtain the total probability for each site **j**.

The likelihood for the full tree then is product of the likelihood at each site.

$$L = L(1) \times L(2) \ldots\ldots \times L(N) = \prod_{j=1}^{N} L(j)$$

Since the individual likelihoods are extremely small numbers it is convenient to sum the log likelihoods at each site and report the likelihood of the entire tree as the log likelihood.

$$\ln L = \ln L(1) + \ln L(2) \ldots\ldots + \ln L(N) = \sum_{j=1}^{N} \ln L(j)$$

Models:

The basic form is a matrix: Q =

	A	C	G	T
A	$-\mu(a\pi_C+b\pi_G+c\pi_T)$	$\mu a\pi_C$	$\boldsymbol{\mu b\pi_G}$	$\mu c\pi_T$
C	$\mu g\pi_A$	$-\mu(g\pi_A+d\pi_G+e\pi_T)$	$\mu d\pi_G$	$\boldsymbol{\mu e\pi_T}$
G	$\boldsymbol{\mu h\pi_A}$	$\mu i\pi_C$	$-\mu(h\pi_C+j\pi_G+f\pi_T)$	$\mu f\pi_T$
T	$\mu i\pi_A$	$\boldsymbol{\mu k\pi_C}$	$\mu l\pi_G$	$-\mu(i\pi_C+k\pi_G+l\pi_T)$

Where μ= mean instantaneous substitution rate
a,b,c,…l =each possible transformation of one base to another
μa= rate parameter for a,b,c,…l
π_A= frequency of bases A, C,G, & T

Nearly all substitution models are rate tables that are variation of this general form, e.g. JC sets all rates equal to 1 (a…l =1) and frequency are equal (π_A, π_C, π_G, π_T all equal ¼), K2P where the observation that transitions and tranversions occur at different rates (b, e, h, k are adjusted by constant K).

Choosing a model:

As you might imagine, there are many models already available (ModelTest discussed below looks at 56!!) and an effectively infinite number are possible. How can one choose?

The program ModelTest (Posada & Crandal 1998) uses log likelihood scores to establish the model that best fits the data. Goodness of fit is tested using the likelihood ratio score.

$\underline{\max [L_0 \text{ (simpler model)} \mid \text{Data}]}$
$\max [L_1 \text{ (more complex model)} \mid \text{Data}]$

This is a nested comparison (i.e. L_0 is a special case of L_1)

Adding additional parameters will always result in a higher likelihood score. However, at some point adding additional parameters is no longer justified in terms of significant improvement in fit of a model to a particular dataset.

A simple example:

HKY85 -lnL = 1787.08
GTR -lnL = 1784.82

Then,

LR = 2 (1787.08 - 1784.82) = 4.52

degrees of freedom = 4 (GTR adds 4 additional parameters to HKY85)
critical value (P = 0.05) = 9.49

The added parameters are not justified by this significance test.

*some of the text is taken from online notes provided by M. Sanderson or the text book by Swofford et al. and other sources.