1. Tree building process under maximum likelihood

2. Key differences between maximum likelihood and parsimony

3. Some fancy extras

“It [maximum likelihood] involves finding that evolutionary tree which yields the highest probability of evolving the observed data” (Felsenstein 1981).

Maximum Likelihood

Likelihood = \( P(D|M) \)
Maximum Likelihood

Likelihood = $P(D|M)$

- Probability
- Alignment
- Topology ($Ψ$), Branch lengths ($v$), Model Parameters ($θ$)

Toss a coin 10 times: H T H H T H H H H

Probability of heads is $p$
Probability of tails is $1-p$

What is the probability of our data given our model of a two-sided coin?

$$L = p \cdot (1-p) \cdot p \cdot p \cdot (1-p) \cdot p \cdot p \cdot p \cdot p$$

$$L = p^7 \cdot (1-p)^3$$

Our new approach

1. Build a model of character evolution
2. Use that model to calculate the likelihood of a particular phylogeny
3. Choose the tree/parameter values with the highest likelihood (maximum likelihood)
Our new approach: Step 1

• Build a model of character evolution
  – What is a model?
  – Parameters vs. parameter values

\[
\begin{array}{cccc}
A & T & C & G \\
\end{array}
\]

Modified from Huelsenbeck 2017

Our new approach: Step 1

• Build a model of character evolution
  – What is a model?
  – Parameters vs. parameter values

\[
\begin{array}{cccc}
A & T & C & G \\
A & - & a & a & a \\
T & a & - & a & a \\
C & a & a & - & a \\
G & a & a & a & - \\
\end{array}
\]

Continuous-Time Markov Chain

Modified from Huelsenbeck 2017
**Jukes Cantor (JC69)**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ/3</td>
<td>μ/3</td>
<td>μ/3</td>
<td>μ/3</td>
</tr>
</tbody>
</table>

**Kimura (K80)**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Felsenstein (F81)**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>πA</td>
<td>πG</td>
<td>πC</td>
<td>πT</td>
</tr>
</tbody>
</table>

**Hasegawa, Kishino, Yano (HKY85)**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>kπA</td>
<td>kπG</td>
<td>kπC</td>
<td>kπT</td>
</tr>
</tbody>
</table>

**General Time Reversible (GTR):**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>α π</td>
<td>β π</td>
<td>γ π</td>
<td>δ π</td>
</tr>
</tbody>
</table>

**Rate parameters:**
- α = A ↔ G
- β = A ↔ C
- γ = A ↔ T
- δ = C ↔ G
- ε = T ↔ G
- η = T ↔ C

**Stationary base frequencies:**
- πA = frequency of A
- πG = frequency of G
- πC = frequency of C
- πT = frequency of T

**Markov models are memoryless**

**Q =**

**General Time Reversible (GTR):**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>α π</td>
<td>β π</td>
<td>γ π</td>
<td>δ π</td>
</tr>
</tbody>
</table>

**Rate parameters:**
- α = A ↔ G
- β = A ↔ C
- γ = A ↔ T
- δ = C ↔ G
- ε = T ↔ G
- η = T ↔ C

**Stationary base frequencies:**
- πA = frequency of A
- πG = frequency of G
- πC = frequency of C
- πT = frequency of T

**COOL MODEL**

https://www.summitpost.org/giant-senecio/762750

https://www.summitpost.org/friljope-espeletia-pycnophylla/565302
Our new approach: Step 2

“It [maximum likelihood] involves finding that evolutionary tree which yields the highest probability of evolving the observed data” (Felsenstein 1981).

Let’s start with one site in the alignment

\[
P_{ij}(t) = e^{Qt}
\]

where do we get \(P_{ij}(t)\) and \(\pi_{s0}\)?

\[
\begin{array}{cccc}
A & G & C & T \\
-3\lambda & \lambda & \lambda & \lambda \\
\lambda & -3\lambda & \lambda & \lambda \\
\lambda & \lambda & -3\lambda & \lambda \\
\lambda & \lambda & \lambda & -3\lambda \\
\end{array}
\]

\[
P_0(t) = \frac{1}{4} + \frac{3}{4}e^{-4\lambda t}
\]

\[
P_1(t) = \frac{3}{4} - \frac{1}{4}e^{-4\lambda t}
\]
Overview:

- Pick a tree & parameter values
- Use Felsenstein’s Pruning Algorithm & your model of evolution to calculate the likelihood of a tree for a single site in an alignment
- Assume all sites in alignment are independent & multiply likelihoods of all sites to calculate the likelihood of the tree given the alignment
- Vary parameter values for that topology to find the maximum likelihood
- Pick a new tree, repeat!

This process is often untenable computationally

- Number of trees increases A LOT with increasing number of taxa
- multiplying small number -> smaller numbers so we use log likelihood

Main differences:

- Branch length matter!!!!!
- Estimation of parameters
- Model underlies the process

http://www.evolution.org/magazine/march-april-2009/beetle-mania
Main differences:
• Branch length matter!!!!!

What happened along the branches?

Main differences:
• Branch length matter!!!!!

C -> G
G -> C

What happened along the branches?

Main differences:
• Branch length matter!!!!!
• Estimation of parameters
• Model underlies the process

How does this compare to weighting in parsimony?
Fancy Things

- Error?
- Rate heterogeneity
- Model testing
- Additional models (codon, amino acid, etc.)

Get idea of error through bootstrapping

But what error is this measuring?

Rate Heterogeneity: Partitions

<table>
<thead>
<tr>
<th>gene1</th>
<th>gene2</th>
<th>gene3</th>
<th>gene4</th>
<th>gene5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-3λ</td>
<td>λ</td>
<td>λ</td>
</tr>
<tr>
<td>G</td>
<td>λ</td>
<td>-3λ</td>
<td>λ</td>
</tr>
<tr>
<td>C</td>
<td>λ</td>
<td>λ</td>
<td>-3λ</td>
</tr>
<tr>
<td>T</td>
<td>λ</td>
<td>λ</td>
<td>λ</td>
</tr>
</tbody>
</table>

Estimate model parameters separately for different genes!
Rate Heterogeneity: Partitions

Gene 1
Gene 2
Gene 3
Gene 4
Gene 5

Estimate model parameters separately for different codon positions!

Rate Heterogeneity: gamma ($\Gamma$)

Model testing

**Likelihood ratio test** for nested models:

\[
LR = 2 \times (-\ln L_1 - \ln L_2)
\]

HKY85 - $\ln L = 1787.08$
GTR - $\ln L = 1784.82$

\[
LR = 2(1787.08 - 1784.82) = 4.53, \text{ df } = 4
\]

critical value ($P = 0.05$) = 9.49 ($\chi^2$ distribution)

Model testing

Aikake Information Criterion (AIC) (non nested):

\[
AIC = 2k - 2\ln(\hat{L})
\]

Find model with lowest AIC

- Models need not be nested
- Can correct for small sample sizes (AICc)
- BIC is a similar alternative
Additional models

- Amino Acid
- Codon
- non-markov
- What would a model of morphology look like?