THE EPIC BATTLE BETWEEN MARKOV AND PHYLOGENETIC INVARIANTS

GRAPHS

Jeremy Sumner
Peter Jarvis
Amelia Taylor
Barbara Holland
Unsigned Sum of Squares

- Get one number for each tree
- Pick the one that is closest to zero
Signed Sum of Squares (SSS)

$q_1 + q_2 + q_3 = 0$

\[
\begin{array}{c|c|c|c|c|c|c|c}
 & q_1 & q_2 & q_3 \\
\hline
1 & 0 & -u & u \\
2 & v & 0 & -v \\
3 & -w & w & 0 \\
\end{array}
\]
\[\begin{array}{ccc}
q_1 & q_2 & q_3 \\
* & -u & u \\
v & 0 & -v \\
-w & w & 0
\end{array}\]
Signed Sum of Squares (SSS)

- Get three (really 2) numbers
- Consult the table of signs
- Pick the quartet tree that minimises the residual sum of squares (SSS)
The contestants

- **Phylogenetic Invariants (minors) with**
  - Unsigned Sum of Squares (USSm)
  - Signed Sum of Squares (SSSm)

- **Markov Invariants (squangles) with**
  - Unsigned Sum of Squares (USSs)
  - Signed Sum of Squares (SSSs)
Simulation - 1

Number correct vs. Sequence length

Phylogenetic invariants
Markov invariants

Tie between signed variants
Simulation - 2

Narrow victory for signed squangles

Phylogenetic invariants
Markov invariants
Narrow victory for signed squangles
Simulation – Felsenstein Zone

Narrow victory for signed squangles
Glorious victory for unsigned minors?
More from minors?

• We get 16 minors (for each flattening). So far we’ve summed the Residual Sum of Squares over all 16.

• What if some minors are more discriminating than others?

• Are there sensible subsets of the minors to look at? How do minors “move around” under leaf permutations?
Flattenings

1234
0000 a
0001 b
0010 c
0011 d
0100 e
0101 f
0110 g
0111 h
1000 i
1001 j
1010 k
1011 l
1100 m
1101 n
1110 o
1111 p

<table>
<thead>
<tr>
<th>00</th>
<th>01</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>D</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
</tr>
<tr>
<td>m</td>
<td>n</td>
<td>o</td>
<td>p</td>
</tr>
</tbody>
</table>
Flattenings and Permutations

<table>
<thead>
<tr>
<th>1234</th>
<th>2134</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000 a</td>
<td>0000 a</td>
</tr>
<tr>
<td>0001 b</td>
<td>0001 b</td>
</tr>
<tr>
<td>0010 c</td>
<td>0010 c</td>
</tr>
<tr>
<td>0011 d</td>
<td>0011 d</td>
</tr>
<tr>
<td>0100 e</td>
<td>1000 e</td>
</tr>
<tr>
<td>0101 f</td>
<td>1001 f</td>
</tr>
<tr>
<td>0110 g</td>
<td>1010 g</td>
</tr>
<tr>
<td>0111 h</td>
<td>1011 h</td>
</tr>
<tr>
<td>1000 i</td>
<td>0100 i</td>
</tr>
<tr>
<td>1001 j</td>
<td>0101 j</td>
</tr>
<tr>
<td>1010 k</td>
<td>0110 k</td>
</tr>
<tr>
<td>1011 l</td>
<td>0111 l</td>
</tr>
<tr>
<td>1100 m</td>
<td>1100 m</td>
</tr>
<tr>
<td>1101 n</td>
<td>1101 n</td>
</tr>
<tr>
<td>1110 o</td>
<td>1110 o</td>
</tr>
<tr>
<td>1111 p</td>
<td>1111 p</td>
</tr>
</tbody>
</table>

Flatten 12|34

<table>
<thead>
<tr>
<th>Flatten</th>
<th>00 01 10 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>a b c d</td>
</tr>
<tr>
<td>01</td>
<td>i j k l</td>
</tr>
<tr>
<td>10</td>
<td>e f g h</td>
</tr>
<tr>
<td>11</td>
<td>m n o p</td>
</tr>
</tbody>
</table>
Flattenings and Permutations

\[
\begin{align*}
1234 & \quad \quad \quad 3412 \\
0000 \quad a & \quad 0000 \quad a \\
0001 \quad b & \quad 0100 \quad b \\
0010 \quad c & \quad 1000 \quad c \\
0011 \quad d & \quad 1100 \quad d \\
0100 \quad e & \quad 0001 \quad e \\
0101 \quad f & \quad 0101 \quad f \\
0110 \quad g & \quad 1001 \quad g \\
0111 \quad h & \quad 1101 \quad h \\
1000 \quad i & \quad 0001 \quad i \\
1001 \quad j & \quad 0101 \quad j \\
1010 \quad k & \quad 1001 \quad k \\
1011 \quad l & \quad 1101 \quad l \\
1100 \quad m & \quad 0011 \quad m \\
1101 \quad n & \quad 0111 \quad n \\
1110 \quad o & \quad 1011 \quad o \\
1111 \quad p & \quad 1111 \quad p
\end{align*}
\]

Flatten 12|34

\[
\begin{array}{cccccc}
\text{Flatten} & 00 & 01 & 10 & 11 \\
00 & a & e & i & m \\
01 & b & f & j & n \\
10 & c & g & k & o \\
11 & d & h & l & p
\end{array}
\]
Accuracy

\begin{align*}
\text{a} & \quad 0.819 \\
\text{p} & \quad 0.822 \\
\text{dm} & \quad 0.803 \\
\text{bcei} & \quad 0.840 \\
\text{hlno} & \quad 0.849 \\
\text{fgjk} & \quad 0.829 \\
\text{sum} & \quad 0.867
\end{align*}
2 states to 4 states

- There are squangles for 4 state data
  - 5th order polynomials with 66,744 terms
  - 0.81 secs / quartet

- Squangles for 2 state data are
  - 3rd order polynomials with 6 terms
  - faster

- How effective is recoding followed by the simpler 2-state squangle?

<table>
<thead>
<tr>
<th>Short edge</th>
<th>Long edge</th>
<th>4 state squangle accuracy</th>
<th>2 state squangle accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.2</td>
<td>0.80</td>
<td>0.61</td>
</tr>
<tr>
<td>0.01</td>
<td>0.4</td>
<td>0.42</td>
<td>0.38</td>
</tr>
<tr>
<td>0.02</td>
<td>0.2</td>
<td>0.96</td>
<td>0.84</td>
</tr>
<tr>
<td>0.02</td>
<td>0.4</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td>0.04</td>
<td>0.2</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>0.04</td>
<td>0.4</td>
<td>0.65</td>
<td>0.69</td>
</tr>
</tbody>
</table>
Ready for a beer?

Yep, phylogenetics makes me thirsty.