Markov invariants and phylogenetic modules

Jeremy Sumner

School of Physical Sciences
University of Tasmania, Australia

Algebraic Statistics, Chicago, May 2014
The only trick I know

- Suppose a problem has symmetry described by a group $G$
- Linear analysis implies $G$ action on vector space $V$, i.e. $\rho : G \to GL(V)$
- For “reductive” groups, $\rho$ breaks into “irreducible” pieces:
  \[
  \rho(g) = \begin{pmatrix}
  A_1(g) & 0 & 0 & \ldots & 0 \\
  0 & A_2(g) & 0 & \ldots & 0 \\
  0 & 0 & A_3(g) & \ldots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \ldots & A_r(g)
  \end{pmatrix}
  \]
- $V$ decomposes into irreducible “modules” $V = V_1 \oplus V_2 \oplus \ldots \oplus V_r$

We will see an example with $\dim(V) = 766, 480$
The only trick I know

- Suppose a problem has symmetry described by a group $G$
- Linear analysis implies $G$ action on vector space $V$, i.e. $\rho : G \to GL(V)$
- For “reductive” groups, $\rho$ breaks into “irreducible” pieces:

$$\rho(g) = \begin{pmatrix}
A_1(g) & 0 & 0 & \cdots & 0 \\
0 & A_2(g) & 0 & \cdots & 0 \\
0 & 0 & A_3(g) & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & A_r(g)
\end{pmatrix}$$

- $V$ decomposes into irreducible “modules” $V = V_1 \oplus V_2 \oplus \ldots \oplus V_r$

We will see an example with $dim(V) = 766,480$

*Representation theory helps us break the problem into smaller pieces*
Symmetry in phylogenetics: Markov process and nucleotide perms

Lie symmetry \(\text{(S,F-S& J 2012)}: \ e^{Qt} e^{Q't'} = e^{\hat{Q}(t+t')}? \)

i. \(Q + \lambda Q'\)
ii. \([Q, Q'] := QQ' - Q'Q\)

\text{Equivariant (D&K 2009) } \subset \text{ LMM}

\text{Lie Markov model 8.8 (and } \approx 35 \text{ others) } \text{(F-S,S,J,&W 2014)}

\[
Q = \begin{pmatrix}
\ast & a & u & u \\
 b & \ast & v & v \\
x & x & \ast & c \\
y & y & d & \ast \\
\end{pmatrix}
\]

Purine/Pyrimidine symmetry: AG vs. CT

Stabilizer = \(\mathbb{S}_2 \wr \mathbb{S}_2\)

= \(\langle (AG), (CT), (AC)(GT) \rangle\)

\[ [L_a, L_b] = L_a - L_b, \quad [L_a, L_v] = L_u - L_v \text{ etc.} \]
Symmetry in phylogenetics: Sequence and leaf permutations

<table>
<thead>
<tr>
<th>Quartet</th>
<th>Stabilizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( G = S_2 \wr S_2 )</td>
</tr>
<tr>
<td>( B )</td>
<td>( = \langle (AB), (CD), (AC)(BD) \rangle )</td>
</tr>
<tr>
<td>( C )</td>
<td></td>
</tr>
<tr>
<td>( D )</td>
<td></td>
</tr>
</tbody>
</table>

Irreducible reps of \( G \) provide distinguished basis for invariants \((S&J 2009)\)

seqA \( \leftrightarrow \) seqB \( \Rightarrow (w_1, w_2, w_3) \rightarrow (w_1, w_3, w_2) \)?

‘Biologically symmetric’ invariants \((E 2009, R&H 2012)\)

‘Invariant’ invariants! \((F-S \ pers. \ comm.)\)
The point of this talk is that accounting for symmetry under leaf/sequence permutations is only half the story.
Leaf action on phylogenetic tensors

‘Phylogenetic tensor’ $P \in \otimes^N V$, on $N$ taxa with $V = \langle A, G, C, T \rangle_C$

\[ g = M_1 \otimes M_2 \otimes M_3 \otimes M_4 \otimes M_5 \]

\[ P \rightarrow P' \]

\{Markov matrices\} $\subset GL(V') \ltimes V' < GL(V)$

\[ M = \begin{pmatrix} \text{unit column sum} \end{pmatrix} \rightarrow \begin{pmatrix} A & v' \\ 0 & 1 \end{pmatrix}, \quad A \in GL(V'), \ v' = \begin{pmatrix} v'_1 \\ v'_2 \\ v'_3 \end{pmatrix} \]
Leaf action on phylogenetic tensors

‘Phylogenetic tensor’ $P \in \otimes^N V$, on $N$ taxa with $V = \langle A, G, C, T \rangle_C$

$g = M_1 \otimes M_2 \otimes M_3 \otimes M_4 \otimes M_5$

$P \rightarrow P'$

{Markov matrices} $\subset GL(V') \ltimes V' < GL(V)$

$M = \begin{pmatrix} \text{unit column sum} \end{pmatrix} \rightarrow \begin{pmatrix} A & v' \\ 0 & 1 \end{pmatrix}$, $A \in GL(V')$, $v' = \begin{pmatrix} v'_1 \\ v'_2 \\ v'_3 \end{pmatrix}$

$P = M_1 \otimes M_2 \otimes M_3 \otimes M_4 \otimes M_5 \cdot \tilde{P}$

Reconstructing tree topology should depend only on $\tilde{P}$ with leaf action dealt with as a nuisance parameter.
Polynomial functions and plethysms for $\times^N GL(V)$

Polynomials: $\mathbb{C} [\otimes^N V] = \sum_d \mathbb{C}_d [\otimes^N V]$

Leaf action: $P \rightarrow gP$, $g \in \times^N GL(V)$

Induced action on polynomials

\[ f'(P) := f(gP) \]
\[ f \rightarrow f' = g^{-1} \circ f \]

e.g. $\mathbb{C}_4[\otimes^3 V] \cong (\{1\} \times \{1\} \times \{1\}) \otimes \{4\}$

\[ = \sum \left( \begin{array}{c} \text{irreducible poly} \\ \text{modules} \end{array} \right) \lambda_1 \times \lambda_2 \times \lambda_3 \]
\[ = \frac{4 \times 4 \times 4}{f f'} + \frac{4 \times 31 \times 31}{f f'} + \frac{31 \times 21^2 \times 1^4}{f f'} + \ldots \]

Dimensions

\[ \frac{64 \times 65 \times 66 \times 67}{4!} = 766,480 = 42,875 + 70,875 + 675 + \ldots \]
Markov invariants

Finite and classical groups \((GL(V), U(n), SO(n), Sp(n)\ldots)\) are reductive

Markov matrices \(M = \begin{pmatrix} A & v' \\ 0 & 1 \end{pmatrix} \in GL(V') \ltimes V'\)

This is not a reductive group!

\[
\lambda_1 \times \lambda_2 \times \lambda_3
\]

\[
f_1 \rightarrow f'_1 \rightarrow f_2 \rightarrow f'_2
\]

Final step is a 1D module iff \(\lambda_i = (r_i + s_i, r_i^{n-1})\), e.g.

We call \(q\) a Markov invariant \((S,C,J,&J 2008)\)

\(q \mapsto \lambda q\), with \(\lambda = \prod_i \det(A_i)^{r_i}\)
Phylogenetic invariants via group characters

- Don’t care about \( \{\text{Markov matrices}\} \subset GL(V') \ltimes V' < GL(V) \) (D& K 2009)
- Lemma 2: Phylogenetic module iff \( f(\tilde{P}) = 0 \) for all \( f \) in module

**Edge Invariants:** \( GL(\otimes^{N_1} V) \times GL(\otimes^{N_2} V) \downarrow \times^{N_1+N_2} GL(V) \)

**Tripod Invariants:**

\[
\tilde{P} = \begin{array}{ccc}
1 & 2 & 3 \\
\end{array}
\]

Leaf action stabilizer: \((\mathbb{C}^* \times \mathbb{C}^*) \wr \mathfrak{S}_4 \subset \times^3 GL(V)\) (K, 1977)

**Salmon conjecture:** degree = 5, 6, 9 (B&O 2011)

\[
\begin{align*}
\{31^2\} & \times \{21^3\} \times \{21^3\} & 1728 &= 3 \times 576 = 3 \times (36 \times 4^2) \\
\{31^3\} & \times \{2^3\} \times \{2^3\} & 3000 &= 3 \times 1000 = 3 \times (10 \times 10^2) \\
\{3^3\} & \times \{3^3\} \times \{3^3\} & 8000 &= 20^3
\end{align*}
\]
The problem with phylogenetic invariants modules

“Residual” measure: \( \Delta_T = \sum_i |f_i(P)^\ell | \)

\[
\Delta_T = f_1^\ell + f_2^\ell + f_3^\ell
\]

\[
\Delta'_T = f'_1^\ell + f'_2^\ell + f'_3^\ell
\]

Any measure \( \Delta_T \) entails a choice of phylogenetic “invariants” equivalent to alternative choice evaluated at a displaced \( P \). i.e. \( \Delta_T \) is not invariant to leaf action.
Markov invariants solve this problem

Quartet case: the (magical) “squangles”

\[ C_5[\otimes^4 V] = \ldots + 4\{21^3\} \times \{21^3\} \times \{21^3\} \times \{21^3\} + \ldots \implies \{q_1, q_2, q_3; q_4\} \]

Leaf action: \( q_i \mapsto \lambda q_i \) with \( \lambda = \det(g) \)

Leaf permutations: \( q_i \mapsto \begin{cases} \sgn(\sigma)q_i \\ q_j \end{cases} \)

\( \sigma \in Stab_{T_i} \implies q_i \) is phylo inv for \( T_i \)

Quartet reconstruction for general Markov model that is invariant to process at leaves

(H,S,&J 2013)

* Actual results may vary depending on \( \lambda \) and stochastic noise
OUTLOOK

- Grand Daddy Lemma?
- No Markov invariants in set of Edge invariants?
- Submodels? e.g. Lie algebra of Strand Symm is $sl_2 \oplus gl_1 \oplus gmm_2$

REFS


Allman ES, Jarvis PD, Rhodes JA, Sumner JG. 2013. Tensor rank, invariants, inequalities, and applications. SIMAX.


The Phylomanians

Come and visit our Theoretical Phylogenetics group in Hobart. Annual meeting *Phylomania 2014 (TBA)* will occur early November.
The Phylomanians
Come and visit our Theoretical Phylogenetics group in Hobart. Annual meeting *Phylomania 2014* (TBA) will occur early November.

Not impressed?
Check out our poo machine!