The epic battle between Markov and phylogenetic invariants: equations

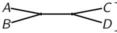
Jeremy Sumner

School of Physical Sciences
University of Tasmania, Australia

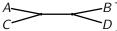
Phylomania 2014

The people I did this with

Symmetry in phylogenetics: Sequence and leaf permutations

$\Delta_1 = \{$*residual for* $T_1 =$ 

$\}$

$\Delta_2 = \{$*residual for* $T_2 =$ 

$\}$

$\Delta_3 = \{$*residual for* $T_3 =$ 

$\}$

ALGEBRA ALERT
i.e. a "representation" of $\mathfrak{S}_4$ on $\{T_1, T_2, T_3\}$

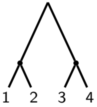
seqA
seqB
seqC
seqD
$\longrightarrow$ $\boxed{\textit{PHYLOGENETIC METHOD}}$ $\longrightarrow (\Delta_1, \Delta_2, \Delta_3)$

seqD
seqC
seqB
seqA
$\longrightarrow$ $\boxed{\textit{PHYLOGENETIC METHOD}}$ $\longrightarrow (\Delta_1, \Delta_2, \Delta_3)$

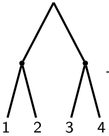
seqA
seqD
seqB
seqC
$\longrightarrow$ $\boxed{\textit{PHYLOGENETIC METHOD}}$ $\longrightarrow (\Delta_3, \Delta_1, \Delta_2)$

The leaf action

ALGEBRA ALERT
i.e. "action" of $\times^4 GL(n)$
with Markov matrices M_i



$\longrightarrow$ $\begin{matrix} seqA \\ seqB \\ seqC \\ seqD \end{matrix}$ $\longrightarrow$ $\boxed{\text{METHOD}}$ $\longrightarrow$ $(\Delta_1, \Delta_2, \Delta_3)$

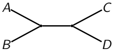


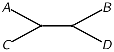
$\longrightarrow$ $\begin{matrix} seqA \\ seqB \\ seqC \\ seqD \end{matrix}$ $\longrightarrow$ $\boxed{\text{METHOD}}$ $\longrightarrow$ $(\Delta_1, \Delta_2, \Delta_3)$???

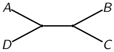
Ideally **tree support** should depend only on "internal" part of tree

Isn't this what "phylogenetic invariants" achieve?

Phylogenetic invariants

$T_1 =$ 

$T_2 =$ 

$T_3 =$ 

Consider polynomials $f(P) = f(p_{AAAA}, p_{AAAC}, p_{AAAG}, \ldots, p_{TTTT})$

Phylogenetic "invariant": (Cavender, Felsenstein, Lake, etc.)

$$\boxed{f(P_1) = 0 \quad f(P_2) \neq 0 \quad f(P_3) \neq 0}$$

Algebraic statistics (Sturmfels, Pacter, et. al.): *Ideals, varieties, etc.*

Our perspective: *Groups, modules, etc.*

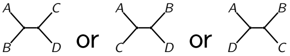
In either case f becomes an infinite space $\langle f_1, f_2, f_3, \ldots \rangle$

UNIVERSITY *of* TASMANIA

Back to sequence and leaf permutations

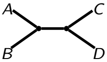
seqA
seqB
seqC
seqD

$\longrightarrow$ $\boxed{\textit{INVARIANTS METHOD}}$ $\longrightarrow$



Permute seqA $\leftrightarrow$ seqB $\implies (\Delta_1, \Delta_2, \Delta_3) \to (\Delta_1, \Delta_3, \Delta_2)$?

'*Biologically symmetric*' invariants (E 2009, R&H 2012)
'*Invariant*' invariants! (F-S pers. comm.)

Quartet	Stabilizer
	$\mathcal{G} = \mathfrak{S}_2 \wr \mathfrak{S}_2 = \langle (AB), (CD), (AC)(BD) \rangle$

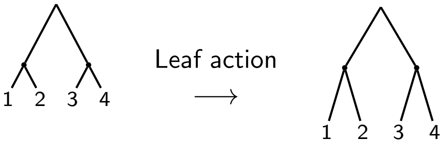
ALGEBRA ALERT

Irreducible representations of $\mathcal{G}$ provide distinguished basis for invariants (S&J 2009)

UNIVERSITY *of* TASMANIA

From an algebraic point of view, this is only half the story.
What about the leaf action?

The problem with phylogenetic invariants $\langle f_1, f_2, f_3, \ldots \rangle$



ALGEBRA ALERT
i.e. "action" of $\times^4 GL(n)$ with Markov matrices M_i

BIG INSIGHT

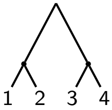
$$p'_{ijkl} = \sum \boxed{\text{Linear combination of } p_{ijkl}, \text{ coeffs from } M_i}$$

$$\Downarrow$$

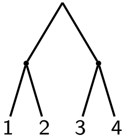
$$f'_i = \sum \boxed{\text{Linear combination of } f_i, \text{ coeffs from } M_i}$$

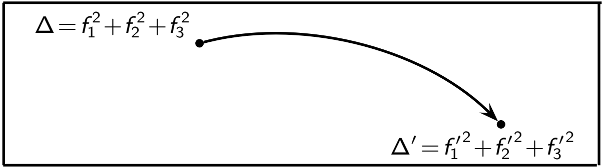
The problem with using phylogenetic invariants $\langle f_1, f_2, f_3, \ldots \rangle$

Tree "residual" $\Delta := \sum_i |f_i(P)|^2$ depends on choice of *basis* $\{f_1, f_2, f_3 \ldots\}$



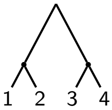
Leaf action $\longrightarrow$



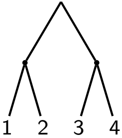


Any measure Δ entails a choice of phylogenetic "invariants" equivalent to alternative choice evaluated at a displaced P.
i.e. Δ is not invariant to leaf action

Markov invariants solve this problem!



Leaf action $\longrightarrow$



THE BIG INSIGHT: $f_i' = \sum \boxed{\text{Linear combination of } f_i, \text{ coeffs from the } M_i}$

Markov invariants: $q \to \lambda q$

Existence theorem (S,C,J,& J, 2009) : $\lambda =$ products of $\det(M_i)$

Surely Markov invariants are good because they don't depend on "internal" part of the tree?

Yes! Log-det and Hadamard q-coordinates; the magical squangles (H,S,& J 2012).

Binary quartet: The simplest case possible

$12|34$ "flattening" of quartet probability distribution P:

$$P = \begin{pmatrix} p_{00,00} & p_{00,01} & p_{00,10} & p_{00,11} \\ p_{01,00} & p_{01,01} & p_{01,10} & p_{01,11} \\ p_{10,00} & p_{10,01} & p_{10,10} & p_{10,11} \\ p_{11,00} & p_{11,01} & p_{11,10} & p_{11,11} \end{pmatrix}$$

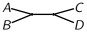
Initial value on "stubby" T_1, T_2, and T_3:

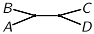
$$P_1 = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \!\!\begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} * & 0 & 0 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ * & 0 & 0 & * \end{pmatrix} \end{array} \quad P_2 = P_3 = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \!\!\begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix} \end{array}$$

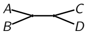
Notice 3×3 minors are 0 on T_1 and non-zero on T_2 and T_3

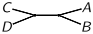
LEAF ACTION+BIG INSIGHT $\implies$ minors are **phylo invariants**

Minors and leaf permutations DETAIL OPTIONAL

Rotation: 

$\longrightarrow$ 

Reflection: 

$\longrightarrow$ 

Flattening: $P \longrightarrow KP$ and P^t, where $K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

3×3 minors under leaf permutations:

$$\begin{array}{c c} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \begin{pmatrix} \circ & \Diamond & \Diamond & \Box \\ \Diamond & \bigstar & \bigstar & \blacktriangle \\ \Diamond & \bigstar & \bigstar & \blacktriangle \\ \Box & \blacktriangle & \blacktriangle & * \end{pmatrix} \end{array}$$

Gives **six** possible leaf perm invariant residuals:
$\Delta = \circ^2 + \Diamond^2 + \Box^2 + \bigstar^2 + \blacktriangle^2 + *^2$

But what about the leaf action?

Leaf action for flattening: $P \to XPY^t$ (where $X = M_1 \otimes M_2$ and $Y = M_3 \otimes M_4$)

Under the leaf action the minors become all mixed up!

$$\circ, \Diamond, \Box, \bigstar, \blacktriangle, * \xrightarrow{\text{Leaf action}} \boxed{\textit{Linear combination of } \circ, \Diamond, \Box, \bigstar, \blacktriangle, *, \textit{ coeffs from the } M_i}$$

BADNESS!

Markov invariants, "the squangle": $q \xrightarrow{\text{Leaf action}} \lambda q$

$\Delta = q^2$ provides a tree-topology residual that is invariant to changes of parameter values at leaves of tree

AWESOMENESS!

UNIVERSITY of TASMANIA

But wait, there's more: Signed least squares

Three flattenings $\implies$ three squangles: q_1, q_2, q_3.

Rep theory: $q_1 + q_2 + q_3 = 0$ use $q_3 = -q_2$ in place of $q_1 = 0$.

Leaf action gives semi-algebraic constraints: $u, v, w > 0$

Hypothesis	$E[q_1]$	$E[q_2]$	$E[q_3]$
T_1	0	$-u$	u
T_2	v	0	$-v$
T_3	$-w$	w	0

Least squares estimate: $\widehat{u} = \frac{1}{2}(q_3 - q_2)$ or $\widehat{u} = 0$.

Residuals: $\Delta = \frac{1}{2}q_1^2$ or $q_2^2 + q_3^2$

Second case sends $q_2, q_3 \to 0$ as best estimate.

Analogous situation for minors

Origin of signed least squares

Leaf action: $P \rightarrow XPY^t$ (where $X = M_1 \otimes M_2$ and $Y = M_3 \otimes M_4$)

On matrix of minors, inverses get in the act:

$$M = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} \implies M^{-1} = \begin{pmatrix} + & - \\ - & + \end{pmatrix}$$

Leaf action on minors:

$$\begin{pmatrix} \circ & \Diamond & \Diamond & \square \\ \Diamond & \bigstar & \bigstar & \blacktriangle \\ \Diamond & \bigstar & \bigstar & \blacktriangle \\ \square & \blacktriangle & \blacktriangle & * \end{pmatrix} = \begin{pmatrix} + & - & - & + \\ - & + & + & - \\ - & + & + & - \\ + & - & - & + \end{pmatrix} \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix} \begin{pmatrix} + & - & - & + \\ - & + & + & - \\ - & + & + & - \\ + & - & - & + \end{pmatrix}$$

$$= \begin{pmatrix} + & - & - & + \\ - & + & + & - \\ - & + & + & - \\ + & - & - & + \end{pmatrix}$$

Take home message

- **Theory** says “Markov invariants” $>$ “phylogenetic invariants”
- In particular, the squangles should give stable tree residual function in face of changing rate parameters.
- Over to you Barbara...

REFS

Eriksson N. 2009. Using invariants for phylogenetic tree construction. IMA Vol. Math. Appl.

Rusinko JP, Hipp, B. 2013. Invariant based quartet puzzling. Alogrithms for Molecular Biology

Sumner JG, Charleston MA, Jermiin LS, Jarvis PD. 2008. Markov invariants, plethysms, and phylogenetics. JTB

Sumner JG, Jarvis PD. 2009. Markov invariants and the isotropy group of a quartet. JTB

Holland BR, Sumner JG, Jarvis PD. 2013. Low-Parameter Phylogenetic Inference Under the GM Model. Syst. Biol.

UNIVERSITY *of* TASMANIA