The squangles: The gift that just keeps on giving

Jeremy Sumner

School of Maths and Physics, University of Tasmania

Phylomania, 6-8 Nov 2013

Phylogenetic models

- Sequence evolution, cont-time Markov chain:

  JC, K2ST, K3ST, F81, GTR and GMM

  Stick this on a tree (and Charles is your uncle):

$$ {}^{1}_{2}\rangle\!-\!\langle^{3}_{4} \quad \text{or} \quad {}^{1}_{3}\rangle\!-\!\langle^{2}_{4} \quad \text{or} \quad {}^{1}_{4}\rangle\!-\!\langle^{2}_{3} \quad ??? $$

Phylogenetic models

- Sequence evolution, cont-time Markov chain:
  JC, K2ST, K3ST, F81, GTR and GMM
  Stick this on a tree (and Charles is your uncle):

$$\begin{matrix}1 & & 3\\ & \rangle\!\!-\!\!\langle & \\ 2 & & 4\end{matrix} \qquad \text{or} \qquad \begin{matrix}1 & & 2\\ & \rangle\!\!-\!\!\langle & \\ 3 & & 4\end{matrix} \qquad \text{or} \qquad \begin{matrix}1 & & 2\\ & \rangle\!\!-\!\!\langle & \\ 4 & & 3\end{matrix} \qquad ???$$

- Phylogenetic tensor $P = \{p_{ijkl}\}$ is joint probability distribution
  DATA: $p_{ijkl} \sim$ proportion of sites with pattern $ijkl$

Phylogenetic models

- Sequence evolution, cont-time Markov chain:

  JC, K2ST, K3ST, F81, GTR and GMM

  Stick this on a tree (and Charles is your uncle):

$$\begin{smallmatrix}1\\2\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}3\\4\end{smallmatrix} \quad \text{or} \quad \begin{smallmatrix}1\\3\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}2\\4\end{smallmatrix} \quad \text{or} \quad \begin{smallmatrix}1\\4\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}2\\3\end{smallmatrix} \quad ???$$

- Phylogenetic tensor $P = \{p_{ijkl}\}$ is joint probability distribution

  DATA: $p_{ijkl} \sim$ proportion of sites with pattern *ijkl*
- Think of $P = P(t_1, t_2, t_3, \ldots; \alpha, \beta, \gamma, \ldots)$ as a parameterized space that we can "move" around in

Phylogenetic models

- Sequence evolution, cont-time Markov chain:
  JC, K2ST, K3ST, F81, GTR and GMM
  Stick this on a tree (and Charles is your uncle):

  $\begin{smallmatrix}1\\ \\2\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}3\\ \\4\end{smallmatrix}$ or $\begin{smallmatrix}1\\ \\3\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}2\\ \\4\end{smallmatrix}$ or $\begin{smallmatrix}1\\ \\4\end{smallmatrix}\rangle\!\!-\!\!\langle\begin{smallmatrix}2\\ \\3\end{smallmatrix}$???

- Phylogenetic tensor $P = \{p_{ijkl}\}$ is joint probability distribution
  DATA: $p_{ijkl} \sim$ proportion of sites with pattern *ijkl*
- Think of $P = P(t_1, t_2, t_3, \ldots; \alpha, \beta, \gamma, \ldots)$ as a parameterized space that we can "move" around in

“Markov invariants”: 1D polynomial reps of Markov process

- “Invariant”: $q(P') = \lambda q(P)$, ie. one-dimensional

  Markov invariants, plethysms, and phylogenetics, Sumner, Charleston, Jermiin, Jarvis, *JTB*, 2008

  Markov inv. and the isotropy subgroup of a quartet tree, Sumner, Jarvis, *JTB*, 2009
- “log-det”: (Barry and Hartigan, 1988, Lockhart, Steel, Penny, Hendy et. al., 1994)

  Identifiability of tree topology (Steel, 1994)

  Early papers on the “tangle” gives triplet distances for GMM
- Semi-algebraic constraints: Klaere and Liebscher (2012), Allman, Jarvis, Rhodes, Sumner (2013)

Markov invariants! What are they good for?

- When did you sample the data? Today, yesterday, 10,000 years ago? Theoretical model: parameter changes on *leaf edges*: $P \rightarrow P'$

Markov invariants! What are they good for?

- When did you sample the data? Today, yesterday, 10,000 years ago?

  Theoretical model: parameter changes on *leaf edges*: $P \to P'$

  Markov invariants q_1, q_2, q_3:

  $$q_i(P) \to q_i(P') = \lambda q_i(P)$$

  λ depends only on the "difference" between P and P'

Markov invariants! What are they good for?

- When did you sample the data? Today, yesterday, 10,000 years ago?

  Theoretical model: parameter changes on *leaf edges*: $P \rightarrow P'$

  Markov invariants q_1, q_2, q_3:

$$q_i(P) \rightarrow q_i(P') = \lambda q_i(P)$$

  λ depends only on the "difference" between P and P'

$\Longrightarrow$ RELATIVE values of q_1, q_2, q_3 do not care about leaf-edge changes

Markov invariants! What are they good for?

- When did you sample the data? Today, yesterday, 10,000 years ago?
  Theoretical model: parameter changes on *leaf edges*: $P \to P'$
  Markov invariants q_1, q_2, q_3:

$$q_i(P) \to q_i(P') = \lambda q_i(P)$$

λ depends only on the “difference” between P and P'

$\Longrightarrow$ RELATIVE values of q_1, q_2, q_3 do not care about leaf-edge changes

$\Longrightarrow$ q_1, q_2, q_3 contain information about internal part of tree only

Markov invariants! What are they good for?

- When did you sample the data? Today, yesterday, 10,000 years ago?

  Theoretical model: parameter changes on *leaf edges*: $P \rightarrow P'$

  Markov invariants q_1, q_2, q_3:

  $$q_i(P) \rightarrow q_i(P') = \lambda q_i(P)$$

  λ depends only on the "difference" between P and P'

  $\Longrightarrow$ RELATIVE values of q_1, q_2, q_3 do not care about leaf-edge changes

  $\Longrightarrow$ q_1, q_2, q_3 contain information about internal part of tree only

- Isn't this what "phylogenetic" invariants are good for?

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$
- View $\boxed{P \to P'}$ as $P' := gP$, a linear transformation

  Induced transformation $\boxed{f \to f'}$ with $f'(P) := f(gP)$

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$
- View $\boxed{P \to P'}$ as $P' := gP$, a linear transformation

  Induced transformation $\boxed{f \to f'}$ with $f'(P) := f(gP)$

  $\implies$ Phylogenetic invariants never come singly: $\{f_1, f_2, \ldots\}$

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$
- View $P \to P'$ as $P' := gP$, a linear transformation

  Induced transformation $f \to f'$ with $f'(P) := f(gP)$

  $\implies$ Phylogenetic invariants never come singly: $\{f_1, f_2, \ldots\}$
- $\Delta_{\mathcal{T}}(P) = f_1(P)^2 + f_2(P)^2$

  Choose $\mathcal{T}$ such that $\Delta_{\mathcal{T}}$ is minimized

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$
- View $P \to P'$ as $P' := gP$, a linear transformation

  Induced transformation $f \to f'$ with $f'(P) := f(gP)$

  $\Longrightarrow$ Phylogenetic invariants never come singly: $\{f_1, f_2, \ldots\}$
- $\Delta_{\mathcal{T}}(P) = f_1(P)^2 + f_2(P)^2$

  Choose $\mathcal{T}$ such that $\Delta_{\mathcal{T}}$ is minimized
- **PROBLEM**: Under $P \to P'$ suppose

$$f_1 \to f_1' = \delta f_1 + \epsilon f_2$$
$$f_2 \to f_2' = \delta f_2 - \epsilon f_1$$

The problem with phylogenetic invariants

- $f(P) = 0$, all P on a fixed tree $\mathcal{T}$
- View $\boxed{P \to P'}$ as $P' := gP$, a linear transformation

  Induced transformation $\boxed{f \to f'}$ with $f'(P) := f(gP)$

  $\implies$ Phylogenetic invariants never come singly: $\{f_1, f_2, \ldots\}$
- $\Delta_{\mathcal{T}}(P) = f_1(P)^2 + f_2(P)^2$

  Choose $\mathcal{T}$ such that $\Delta_{\mathcal{T}}$ is minimized
- **PROBLEM**: Under $P \to P'$ suppose

$$f_1 \to f_1' = \delta f_1 + \epsilon f_2$$
$$f_2 \to f_2' = \delta f_2 - \epsilon f_1$$

- Clearly $\Delta_{\mathcal{T}}$ is NOT invariant to this change
- Answer depends on the day of the week!

Markov invariants fix this problem

- $q_i \to q_i' = \lambda q_i$

Markov invariants fix this problem

- $q_i \rightarrow q_i' = \lambda q_i$
- $\Delta_{\mathcal{T}}(P) := q_1(P)^2 + q_2(P)^2$ satisfies

$$\Delta_{\mathcal{T}}(P) \rightarrow \boxed{\Delta_{\mathcal{T}}(P') = \lambda^2 \Delta_{\mathcal{T}}(P)}$$

$$\implies \min(\Delta_{\mathcal{T}_1}(P), \Delta_{\mathcal{T}_2}(P), \ldots) = \min(\Delta_{\mathcal{T}_1}(P'), \Delta_{\mathcal{T}_2}(P'), \ldots)$$

Markov invariants fix this problem

- $q_i \to q_i' = \lambda q_i$
- $\Delta_{\mathcal{T}}(P) := q_1(P)^2 + q_2(P)^2$ satisfies

$$\Delta_{\mathcal{T}}(P) \to \boxed{\Delta_{\mathcal{T}}(P') = \lambda^2 \Delta_{\mathcal{T}}(P)}$$

$$\implies \min(\Delta_{\mathcal{T}_1}(P), \Delta_{\mathcal{T}_2}(P), \ldots) = \min(\Delta_{\mathcal{T}_1}(P'), \Delta_{\mathcal{T}_2}(P'), \ldots)$$

- Fantastic!

Markov invariants fix this problem

- $q_i \to q_i' = \lambda q_i$
- $\Delta_{\mathcal{T}}(P) := q_1(P)^2 + q_2(P)^2$ satisfies

$$\Delta_{\mathcal{T}}(P) \to \boxed{\Delta_{\mathcal{T}}(P') = \lambda^2 \Delta_{\mathcal{T}}(P)}$$

$$\implies \min(\Delta_{\mathcal{T}_1}(P), \Delta_{\mathcal{T}_2}(P), \ldots) = \min(\Delta_{\mathcal{T}_1}(P'), \Delta_{\mathcal{T}_2}(P'), \ldots)$$

- Fantastic!
- Markov invariants provide a method where the answer is the same on Monday, as it is on Tuesday, Wednesday, etc. etc. ...*

*Actual results may vary ... $\lambda \sim$ det is monotonically decreasing in time, so ability to choose correct minimum $\Delta_{\mathcal{T}}$ above stochastic noise weakens correspondingly

A tale of some polynomials that keep following me around. . .

The magical "squangles": Sumner, Jarvis (2009)

- *Four* Markov invariants on a quartet tree, degree 5 with 66,744 monomial terms: e.g.

$$q(P) = p_{\mathtt{AACA}}^2 p_{\mathtt{AACG}} p_{\mathtt{ACGA}} p_{\mathtt{CCCG}} + p_{\mathtt{ACCG}}^2 p_{\mathtt{ACCC}} p_{\mathtt{CGCG}} p_{\mathtt{GCCG}} + \dots$$

The magical "squangles": Sumner, Jarvis (2009)

- *Four* Markov invariants on a quartet tree, degree 5 with 66,744 monomial terms: e.g.

$$q(P) = p_{\mathtt{AACA}}^2 p_{\mathtt{AACG}} p_{\mathtt{ACGA}} p_{\mathtt{CCCG}} + p_{\mathtt{ACCG}}^2 p_{\mathtt{ACCC}} p_{\mathtt{CGCG}} p_{\mathtt{GCCG}} + \ldots$$

- $(1 \times 1 \times 1 \times 1)\underline{\otimes}(5) = \ldots + \boxed{4}\,(21^3 \times 21^3 \times 21^3 \times 21^3)$

The magical "squangles": **Sumner, Jarvis (2009)**

- *Four* Markov invariants on a quartet tree, degree 5 with 66,744 monomial terms: e.g.

$$q(P) = p_{\mathtt{AACA}}^2 p_{\mathtt{AACG}} p_{\mathtt{ACGA}} p_{\mathtt{CCCG}} + p_{\mathtt{ACCG}}^2 p_{\mathtt{ACCC}} p_{\mathtt{CGCG}} p_{\mathtt{GCCG}} + \ldots$$

- $(1 \times 1 \times 1 \times 1)\underline{\otimes}(5) = \ldots + \boxed{4}\,(21^3 \times 21^3 \times 21^3 \times 21^3)$
- *Two* of the four are phylogenetic invariants for a fixed quartet.

The magical "squangles": Sumner, Jarvis (2009)

- *Four* Markov invariants on a quartet tree, degree 5 with 66,744 monomial terms: e.g.

$$q(P) = p_{\texttt{AACA}}^2 p_{\texttt{AACG}} p_{\texttt{ACGA}} p_{\texttt{CCCG}} + p_{\texttt{ACCG}}^2 p_{\texttt{ACCC}} p_{\texttt{CGCG}} p_{\texttt{GCCG}} + \ldots$$

- $(1 \times 1 \times 1 \times 1)\underline{\otimes}(5) = \ldots + \boxed{4}\,(21^3 \times 21^3 \times 21^3 \times 21^3)$
- *Two* of the four are phylogenetic invariants for a fixed quartet.
- Symmetric choice: $\{q_1, q_2, q_3\}$ with $q_1 + q_2 + q_3 = 0$ and

	$\begin{smallmatrix}1 & & 3\\ & \rangle\!-\!\langle & \\ 2 & & 4\end{smallmatrix}$	$\begin{smallmatrix}1 & & 2\\ & \rangle\!-\!\langle & \\ 3 & & 4\end{smallmatrix}$	$\begin{smallmatrix}1 & & 2\\ & \rangle\!-\!\langle & \\ 4 & & 3\end{smallmatrix}$
q_1	0	v	$-w$
q_2	$-u$	0	w
q_3	u	$-v$	0

- Least squares application: $u, v, w \geq 0$ (Holland, Jarvis, Sumner 2012)

If the squangles are phylogenetic invariants...

- Degree 5 phylogenetic invariants occur as two types only:

i. Edge invariants, constructed from 5×5 minors of quartet flattenings

ii. Tripod invariants, constructed using "commutation" relations or discrete $\mathbb{Z}_3 wr \mathfrak{S}_4$ symmetry (see Phylomania 2012)

If the squangles are phylogenetic invariants...

- Degree 5 phylogenetic invariants occur as two types only:

i. Edge invariants, constructed from 5×5 minors of quartet flattenings

ii. Tripod invariants, constructed using "commutation" relations or discrete $\mathbb{Z}_3 wr \mathfrak{S}_4$ symmetry (see Phylomania 2012)

- The squangles are *not* edge invariants because, as $16 \times 16 \downarrow 4 \times 4 \times 4 \times 4$:

$$1^5 \times 1^5 \to (41 \times 21^3 \times 41 \times 21^3) + (41 \times 21^3 \times 32 \times 31^2) + \ldots$$

If the squangles are phylogenetic invariants...

- Degree 5 phylogenetic invariants occur as two types only:

i. Edge invariants, constructed from 5×5 minors of quartet flattenings

ii. Tripod invariants, constructed using "commutation" relations or discrete $\mathbb{Z}_3 wr \mathfrak{S}_4$ symmetry (see Phylomania 2012)

- The squangles are *not* edge invariants because, as $16 \times 16 \downarrow 4 \times 4 \times 4 \times 4$:

$$1^5 \times 1^5 \rightarrow (41 \times 21^3 \times 41 \times 21^3) + (41 \times 21^3 \times 32 \times 31^2) + \ldots$$

- The squangles *are* tripod invariants because, as $4 \times 4 \times 16 \downarrow 4 \times 4 \times 4 \times 4$:

$$21^3 \times 21^3 \times 31^2 \rightarrow (21^3 \times 21^3 \times 5 \times 31^2) + \ldots + (21^3 \times 21^3 \times 21^3 \times 21^3)$$

- For a fixed quartet, there are exactly TWO ways to do this.✓

That's all folks*. Watch out for Phylomania 2014!

*Please, no questions about rate variation across sites, thankyou.