The algebra of Markov models on phylogenetic split networks

Jeremy Sumner
School of Maths and Physics, University of Tasmania
Joint work with Peter Jarvis and Barbara Holland

Phylomania, 29-30 Oct, 2009

The algebra of Markov models on phylogenetic split networks

Jeremy Sumner
School of Maths and Physics, University of Tasmania
Joint work with Peter Jarvis and Barbara Holland

Phylomania, 29-30 Oct, 2009

Motivations

Motivations

- Phylogenetics: molecular sequences $\longrightarrow$ evolutionary history.

Motivations

- Phylogenetics: molecular sequences $\longrightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.

Motivations

- Phylogenetics: molecular sequences $\longrightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.
- This is modelled by taking stochastic independence across branches of the tree.

Motivations

- Phylogenetics: molecular sequences $\longrightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.
- This is modelled by taking stochastic independence across branches of the tree.
- It is well known that biological reality is more complicated.

Motivations

- Phylogenetics: molecular sequences $\rightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.
- This is modelled by taking stochastic independence across branches of the tree.
- It is well known that biological reality is more complicated.
- ie. Horizontal gene transfer, incomplete lineage sorting, hybridisation, recombination etc...

Motivations

- Phylogenetics: molecular sequences $\rightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.
- This is modelled by taking stochastic independence across branches of the tree.
- It is well known that biological reality is more complicated.
- ie. Horizontal gene transfer, incomplete lineage sorting, hybridisation, recombination etc...
- There are various ways this is coped with: split networks via distances, ie. incompatible distance metrics.

Motivations

- Phylogenetics: molecular sequences $\longrightarrow$ evolutionary history.
- It is usual to assume that the biological process corresponds strictly to evolution on a *tree*.
- This is modelled by taking stochastic independence across branches of the tree.
- It is well known that biological reality is more complicated.
- ie. Horizontal gene transfer, incomplete lineage sorting, hybridisation, recombination etc...
- There are various ways this is coped with: split networks via distances, ie. incompatible distance metrics.
- However it would be nice to generalize probability models themselves to arbitrary split networks.

Phylogenetic models as tensors

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \dots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \dots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \ldots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \ldots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.
- For a tree with n leaves we need $X^n := X \times X \times \ldots \times X$.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \ldots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.
- For a tree with n leaves we need $X^n := X \times X \times \ldots \times X$.
- Each element $(i_1, i_2, \ldots, i_n) \in X^n$ is a *pattern* and $p_{i_1 i_2 \ldots i_n}$ is an associated probability distribution.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \ldots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.
- For a tree with n leaves we need $X^n := X \times X \times \ldots \times X$.
- Each element $(i_1, i_2, \ldots, i_n) \in X^n$ is a *pattern* and $p_{i_1 i_2 \ldots i_n}$ is an associated probability distribution.
- The measures $\{\delta_{i_1} \otimes \delta_{i_2} \otimes \ldots \otimes \delta_{i_n}\}_{i_1, i_2, \ldots, i_n \in X}$ form a basis for the measures on X^n.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \dots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.
- For a tree with n leaves we need $X^n := X \times X \times \dots \times X$.
- Each element $(i_1, i_2, \dots, i_n) \in X^n$ is a *pattern* and $p_{i_1 i_2 \dots i_n}$ is an associated probability distribution.
- The measures $\{\delta_{i_1} \otimes \delta_{i_2} \otimes \dots \otimes \delta_{i_n}\}_{i_1, i_2, \dots, i_n \in X}$ form a basis for the measures on X^n.
- We take $e_i \equiv \delta_i$, where the e_i form a basis for the vector space $V \equiv \mathbb{C}^k$.

Phylogenetic models as tensors

- Take the state space $X = \{1, 2, \ldots, k\}$, ie. For DNA $k = 4$ and we write $X = \{A, C, G, T\}$.
- Consider the probability measures $\mu : X \to [0, 1]$.
- The measures δ_i defined as $\delta_i(j) = \delta_{ij}$ form a basis: $\mu = \sum_{i \in X} \mu(i)\delta_i$.
- For a tree with n leaves we need $X^n := X \times X \times \ldots \times X$.
- Each element $(i_1, i_2, \ldots, i_n) \in X^n$ is a *pattern* and $p_{i_1 i_2 \ldots i_n}$ is an associated probability distribution.
- The measures $\{\delta_{i_1} \otimes \delta_{i_2} \otimes \ldots \otimes \delta_{i_n}\}_{i_1, i_2, \ldots, i_n \in X}$ form a basis for the measures on X^n.
- We take $e_i \equiv \delta_i$, where the e_i form a basis for the vector space $V \equiv \mathbb{C}^k$.
- This allows us to speak of a *phylogenetic tensor* $P := \sum_{i_1, i_2, \ldots, i_n \in X} p_{i_1 i_2 \ldots i_n} e_{i_1} \otimes e_{i_2} \otimes \ldots \otimes e_{i_n}$ embedded in $V^{\otimes n}$.

Branching events

Branching events

- First of all let's deal with branching events (or speciation).

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \longrightarrow V \otimes V$.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \to V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \rightarrow V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.
- One taxa goes to two is described by $P^{(1)} \mapsto \delta \cdot P^{(1)}$.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \rightarrow V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.
- One taxa goes to two is described by $P^{(1)} \mapsto \delta \cdot P^{(1)}$.
- Two taxa goes to three: $P^{(2)} \mapsto 1 \otimes \delta \cdot P^{(2)}$.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \to V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.
- One taxa goes to two is described by $P^{(1)} \mapsto \delta \cdot P^{(1)}$.
- Two taxa goes to three: $P^{(2)} \mapsto 1 \otimes \delta \cdot P^{(2)}$.
- Meanwhile, the transformation rule under the Markov process is $P^{(1)} \to M_1 \cdot P^{(1)}$ and $P^{(2)} \to (M_1 \otimes M_2) \cdot P^{(1)}$.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \to V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.
- One taxa goes to two is described by $P^{(1)} \mapsto \delta \cdot P^{(1)}$.
- Two taxa goes to three: $P^{(2)} \mapsto 1 \otimes \delta \cdot P^{(2)}$.
- Meanwhile, the transformation rule under the Markov process is $P^{(1)} \to M_1 \cdot P^{(1)}$ and $P^{(2)} \to (M_1 \otimes M_2) \cdot P^{(1)}$.
- It is that $M_1 \otimes M_2$ occurs as a tensor product which gives stochastic independence across branches.

Branching events

- First of all let's deal with branching events (or speciation).
- We would like to do this at the level of abstract algebra.
- We need an operator $\delta : V \rightarrow V \otimes V$.
- Imposing conditional independence across branches is enough to specify $\delta \cdot e_i = e_i \otimes e_i$.
- One taxa goes to two is described by $P^{(1)} \mapsto \delta \cdot P^{(1)}$.
- Two taxa goes to three: $P^{(2)} \mapsto 1 \otimes \delta \cdot P^{(2)}$.
- Meanwhile, the transformation rule under the Markov process is $P^{(1)} \rightarrow M_1 \cdot P^{(1)}$ and $P^{(2)} \rightarrow (M_1 \otimes M_2) \cdot P^{(1)}$.
- It is that $M_1 \otimes M_2$ occurs as a tensor product which gives stochastic independence across branches.
- Relaxing the condition is central to generalizing to incompatible split systems.

The general Markov model and CTMCs

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.
- From a Lie group point of view this means the rate parameters become the coordinates of a “group” manifold.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.
- From a Lie group point of view this means the rate parameters become the coordinates of a "group" manifold.
- ie. For $k = 2$, forget t and take $M = e^{Q}$ with $Q = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.
- From a Lie group point of view this means the rate parameters become the coordinates of a "group" manifold.
- ie. For $k = 2$, forget t and take $M = e^{Q}$ with $Q = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$.
- Consider $L_\alpha = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$, $L_\beta = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$, so $Q = \alpha L_\alpha + \beta L_\beta$.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.
- From a Lie group point of view this means the rate parameters become the coordinates of a "group" manifold.
- ie. For $k = 2$, forget t and take $M = e^{Q}$ with $Q = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$.
- Consider $L_\alpha = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$, $L_\beta = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$, so $Q = \alpha L_\alpha + \beta L_\beta$.
- These matrices form a "Lie algebra": $[L_\alpha, L_\beta] = L_\beta - L_\alpha$.

The general Markov model and CTMCs

- In our work we like to always think of CTMCs.
- This amounts to taking $M = e^{Qt}$ where Q is a zero *column*-sum matrix.
- From a Lie group point of view this means the rate parameters become the coordinates of a "group" manifold.
- ie. For $k = 2$, forget t and take $M = e^{Q}$ with $Q = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$.
- Consider $L_\alpha = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$, $L_\beta = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$, so $Q = \alpha L_\alpha + \beta L_\beta$.
- These matrices form a "Lie algebra": $[L_\alpha, L_\beta] = L_\beta - L_\alpha$.
- This condition is exactly what is needed to ensure that a continuous-time model is *closed* (see PJ's talk).

The algebra of branching events

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.
- Q. Is it possible to "push through" δ so the branching event occurs at the root?

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.
- Q. Is it possible to "push through" δ so the branching event occurs at the root?
- A. Yes it is!

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.
- Q. Is it possible to “push through” δ so the branching event occurs at the root?
- A. Yes it is!   We can write $P = (M_1 \otimes M_2) \cdot \widetilde{M}_0 \cdot \delta \cdot \pi$ in this case.

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.
- Q. Is it possible to "push through" δ so the branching event occurs at the root?
- A. Yes it is! We can write $P = (M_1 \otimes M_2) \cdot \widetilde{M}_0 \cdot \delta \cdot \pi$ in this case.
- The key point is that $\widetilde{M}_0$ *cannot* be written as a tensor product (but it is a well defined linear operator on $V \otimes V$).

The algebra of branching events

- It is clear that we can uniquely define δ^n to generate multifurcating events.
- For instance, consider $\delta^2 := (1 \otimes \delta) \cdot \delta = (\delta \otimes 1) \cdot \delta$.
- Consider a simple case where we generate the tensor $P = (M_1 \otimes M_2) \cdot \delta \cdot M_0 \cdot \pi$.
- Q. Is it possible to "push through" δ so the branching event occurs at the root?
- A. Yes it is! We can write $P = (M_1 \otimes M_2) \cdot \widetilde{M}_0 \cdot \delta \cdot \pi$ in this case.
- The key point is that $\widetilde{M}_0$ *cannot* be written as a tensor product (but it is a well defined linear operator on $V \otimes V$).
- First of all, let's consider the simpler binary symmetric case...

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:
- First of all, write $e^{Qt} = e^{(-1+K)t} = e^{-t} e^{Kt}$, with

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:
- First of all, write $e^{Qt} = e^{(-1+K)t} = e^{-t} e^{Kt}$, with $\delta \cdot e^{Kt} = \delta \cdot \sum_{i=0}^{\infty} \frac{(Kt)^n}{n!} = \sum_{i=0}^{\infty} \frac{\delta \cdot (Kt)^n}{n!} = e^{(K \otimes K)t}$.

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:
- First of all, write $e^{Qt} = e^{(-1+K)t} = e^{-t} e^{Kt}$, with $\delta \cdot e^{Kt} = \delta \cdot \sum_{i=0}^{\infty} \frac{(Kt)^n}{n!} = \sum_{i=0}^{\infty} \frac{\delta \cdot (Kt)^n}{n!} = e^{(K \otimes K)t}$.
- Let's work through our simple 2-taxa example.

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:
- First of all, write $e^{Qt} = e^{(-1+K)t} = e^{-t}e^{Kt}$, with $\delta \cdot e^{Kt} = \delta \cdot \sum_{i=0}^{\infty} \frac{(Kt)^n}{n!} = \sum_{i=0}^{\infty} \frac{\delta \cdot (Kt)^n}{n!} = e^{(K \otimes K)t}$.
- Let's work through our simple 2-taxa example.
- We can write *any* phylogenetic tensor as $P = e^{-\lambda} \exp\left[\sum_{e \in T} K^{(e)} \tau_e\right]$.

Intertwinings: binary symmetric case ($\mathbb{Z}_2$).

- This was discussed in Bashford et. al. (2005) in the context of Lie algebraic approach and discussed by Bryant (2008).
- For binary symmetric case we have $Q = -1 + K$, where K is the permutation (12).
- The "intertwining property" is $\delta \cdot Q = (-1 \otimes 1 + K \otimes K) \cdot \delta$.
- Now, δ is a linear operator, so extending to transition matrices $M = e^{Qt}$ is easy:
- First of all, write $e^{Qt} = e^{(-1+K)t} = e^{-t}e^{Kt}$, with $\delta \cdot e^{Kt} = \delta \cdot \sum_{i=0}^{\infty} \frac{(Kt)^n}{n!} = \sum_{i=0}^{\infty} \frac{\delta \cdot (Kt)^n}{n!} = e^{(K \otimes K)t}$.
- Let's work through our simple 2-taxa example.
- We can write *any* phylogenetic tensor as $P = e^{-\lambda} \exp\left[\sum_{e \in T} K^{(e)} \tau_e\right]$.
- Here we have a "a ha!" moment and replace T with an arbitrary split system S.

Intertwinings: the general Markov model

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.
- The intertwining is: $\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha)$.

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.
- The intertwining is: $\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha)$.
- Extend to multifurcating events: $\delta^n L_\alpha = \left(\sum_{A\subset[n],A\neq\emptyset} L_\alpha^{(A)}\right) \cdot \delta^n := \mathcal{L}_\alpha^{[n]} \cdot \delta^n$.

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.
- The intertwining is: $\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha)$.
- Extend to multifurcating events:
  $\delta^n L_\alpha = \left(\sum_{A \subset [n], A \neq \emptyset} L_\alpha^{(A)}\right) \cdot \delta^n := \mathcal{L}_\alpha^{[n]} \cdot \delta^n$.
- Remember everything is linear, so we can mimic the action $\delta^n \cdot e^{\alpha L_\alpha + \alpha L_\beta} \cdot \pi$ with $e^{\alpha \mathcal{L}_\alpha^{[n]} + \beta \mathcal{L}_\beta^{[n]}} \delta^n$.

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.
- The intertwining is: $\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha)$.
- Extend to multifurcating events:
  $\delta^n L_\alpha = \left(\sum_{A \subset [n], A \neq \emptyset} L_\alpha^{(A)}\right) \cdot \delta^n := \mathcal{L}_\alpha^{[n]} \cdot \delta^n$.
- Remember everything is linear, so we can mimic the action
  $\delta^n \cdot e^{\alpha L_\alpha + \alpha L_\beta} \cdot \pi$ with $e^{\alpha \mathcal{L}_\alpha^{[n]} + \beta \mathcal{L}_\beta^{[n]}} \delta^n$.
- Working this through, (for a tree at least) it becomes possible to order by cardinality and write

Intertwinings: the general Markov model

- Recall that we have $Q = \alpha L_\alpha + \beta L_\beta$.
- The intertwining is: $\delta \cdot L_\alpha = (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha)$.
- Extend to multifurcating events: $\delta^n L_\alpha = \left(\sum_{A \subset [n], A \neq \emptyset} L_\alpha^{(A)} \right) \cdot \delta^n := \mathcal{L}_\alpha^{[n]} \cdot \delta^n$.
- Remember everything is linear, so we can mimic the action $\delta^n \cdot e^{\alpha L_\alpha + \alpha L_\beta} \cdot \pi$ with $e^{\alpha \mathcal{L}_\alpha^{[n]} + \beta \mathcal{L}_\beta^{[n]}} \delta^n$.
- Working this through, (for a tree at least) it becomes possible to order by cardinality and write

  $P = \exp[\mathcal{R}_1] \cdot \exp[\mathcal{R}_2] \cdot \ldots \cdot \exp[\mathcal{R}_{n-1}] \cdot \delta^n \cdot \pi$,

  where $\mathcal{R}_i = \sum_{A, |A| = i} \tau_A \mathcal{L}^{(A)}$.

These two approaches are different!

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

- We are missing the two cross terms $L_\alpha \otimes L_\beta$ and $L_\beta \otimes L_\alpha$.

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

- We are missing the two cross terms $L_\alpha \otimes L_\beta$ and $L_\beta \otimes L_\alpha$.
- The amazing fact is that on a tree this *does not* make any difference.

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

- We are missing the two cross terms $L_\alpha \otimes L_\beta$ and $L_\beta \otimes L_\alpha$.
- The amazing fact is that on a tree this *does not* make any difference.
- This is because the cross terms always annihilate any "mismatched terms".

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

- We are missing the two cross terms $L_\alpha \otimes L_\beta$ and $L_\beta \otimes L_\alpha$.
- The amazing fact is that on a tree this *does not* make any difference.
- This is because the cross terms always annihilate any "mismatched terms".
- eg. Consider the tree (1(23)) and the relevant operators.

These two approaches are different!

- Take the case where $\alpha = \beta = 1$, ie. $-1 + K = L_\alpha + L_\beta$.
- But the intertwining is not the same!

$$-1 \otimes 1 + K \otimes K \neq (L_\alpha \otimes L_\alpha + L_\alpha \otimes 1 + 1 \otimes L_\alpha) + (L_\beta \otimes L_\beta + L_\beta \otimes 1 + 1 \otimes L_\beta).$$

- We are missing the two cross terms $L_\alpha \otimes L_\beta$ and $L_\beta \otimes L_\alpha$.
- The amazing fact is that on a tree this *does not* make any difference.
- This is because the cross terms always annihilate any "mismatched terms".
- eg. Consider the tree (1(23)) and the relevant operators.
- Clearly, on an arbitrary split system the two approaches are different...

The L-rep vs. the K-rep

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".
- Consider a simple two-taxa case.

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".
- Consider a simple two-taxa case.
- Consider $t_{12}(\mathcal{L}^{(12)} + \mathcal{L}_{\beta}^{12}) + t_{23}(\mathcal{L}^{(23)} + \mathcal{L}_{\beta}^{23})$.

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".
- Consider a simple two-taxa case.
- Consider $t_{12}(\mathcal{L}^{(12)} + \mathcal{L}_{\beta}^{12}) + t_{23}(\mathcal{L}^{(23)} + \mathcal{L}_{\beta}^{23})$.
- This is compared to $t_{12}K \otimes K \otimes 1 + t_{23}1 \otimes K \otimes K$.

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".
- Consider a simple two-taxa case.
- Consider $t_{12}(\mathcal{L}^{(12)} + \mathcal{L}_{\beta}^{12}) + t_{23}(\mathcal{L}^{(23)} + \mathcal{L}_{\beta}^{23})$.
- This is compared to $t_{12}K \otimes K \otimes 1 + t_{23} 1 \otimes K \otimes K$.
- OK, great. But the K-rep has desirable mathematical features the L-rep doesn't share.

The L-rep vs. the K-rep

- The L-rep has features that the K-rep simply can't match.
- For instance the L-rep seems to have a natural notion of bringing edges "back together".
- Consider a simple two-taxa case.
- Consider $t_{12}(\mathcal{L}^{(12)} + \mathcal{L}_{\beta}^{12}) + t_{23}(\mathcal{L}^{(23)} + \mathcal{L}_{\beta}^{23})$.
- This is compared to $t_{12}K \otimes K \otimes 1 + t_{23}1 \otimes K \otimes K$.
- OK, great. But the K-rep has desirable mathematical features the L-rep doesn't share.
- The L-rep mis-behaves (badly!) under marginalizations and is chronically rooted (for split systems?=contradication).

Closing remarks

- At this point it seems wise to think more carefully about what we would like to *do* with these models.

Closing remarks

- At this point it seems wise to think more carefully about what we would like to *do* with these models.
- Hopefully the actual biological applications will be a useful guide here.

Closing remarks

- At this point it seems wise to think more carefully about what we would like to *do* with these models.
- Hopefully the actual biological applications will be a useful guide here.
- Special thanks to David Bryant, John Rhodes and Elizabeth Allman for additional discussions and comments.

Closing remarks

- At this point it seems wise to think more carefully about what we would like to *do* with these models.
- Hopefully the actual biological applications will be a useful guide here.
- Special thanks to David Bryant, John Rhodes and Elizabeth Allman for additional discussions and comments.
- Thanks for listening!