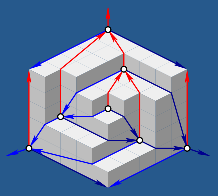


DFG Research Training Group
Methods for Discrete Structures
www.math.tu-berlin.de/MDS

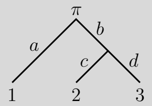
Lattice polytopes of phylogenetic trees

Kaie Kubjas
Department of Mathematics, Freie Universität Berlin, Germany
kubjas@math.fu-berlin.de
Supervisor: Christian Haase

Berlin Mathematical School

Phylogenetic models

Let T be a rooted tree with each vertex a random variable with 2 possible states (binary states $\{0,1\}$).



INPUT:	root distribution $\pi = (\pi_0, \pi_1)$ and edge transition matrices representing the probabilities of transition between the states $M_a = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$ $M_b = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}$ $M_c = \begin{pmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{pmatrix}$ $M_d = \begin{pmatrix} d_{00} & d_{01} \\ d_{10} & d_{11} \end{pmatrix}$
OUTPUT:	probability of observing the letter i at the leaf 1, the letter j at the leaf 2, and the letter k at the leaf 3 $\Phi_{ijk} = \pi_0 a_{0i} b_{00} c_{0j} d_{0k} + \pi_0 a_{01} b_{01} c_{1j} d_{1k} + \pi_1 a_{1i} b_{10} c_{0j} d_{0k} + \pi_1 a_{11} b_{11} c_{1j} d_{1k}$

- Eight polynomials Φ_{ijk} give the map

$$\Phi : \mathbb{C}^{18} \to \mathbb{C}^8.$$

- Fixing a parameter space $P \subseteq \mathbb{C}^{18}$ specifies the model.
- Object of interest is the Zariski closure of $\Phi(P)$.

Example. Jukes Cantor binary model has uniform root distribution $(\frac{1}{2}, \frac{1}{2})$ and transition matrices of the form

$$M_a = \begin{pmatrix} a_{00} & 1-a_{00} \\ 1-a_{00} & a_{00} \end{pmatrix} \quad M_b = \begin{pmatrix} b_{00} & 1-b_{00} \\ 1-b_{00} & b_{00} \end{pmatrix}$$
$$M_c = \begin{pmatrix} c_{00} & 1-c_{00} \\ 1-c_{00} & c_{00} \end{pmatrix} \quad M_d = \begin{pmatrix} d_{00} & 1-d_{00} \\ 1-d_{00} & d_{00} \end{pmatrix}$$

General setting (see [ERSS05]): rooted tree T, k states (usually 2 for the binary states or 4 for the nucleotides), root distribution π and edge transition matrices M_e.

Group-based models

- In case of group-based models transition matrices are invariant under a group action of G.
- Varieties $\overline{\Phi(P)}$ are toric and give lattice polytopes $P_{G,T}$ in the lattices $L_{G,T}$.
- Ehrhart polynomial is an important invariant of group-based models: $\mathrm{ehr}_{P_{G,T}}(k) = |kP_{G,T} \cap L_{G,T}|$ for $k \in \mathbb{Z}_{\geq 0}$.

Example. Transition matrices of the Jukes Cantor binary model are invariant under the $\mathbb{Z}_2$ action.

Example. The polytope $P_{\mathbb{Z}_2,T}$ of the Jukes Cantor binary model associated to a tree T is defined by

$$\mathrm{conv}(a \in \mathbb{Z}^E : a_e \in \{0,1\} \text{ and } \sum_{v \in e} a_e \in 2\mathbb{Z} \text{ for every } v \in I),$$

where E is the set of edges and I is the set of inner vertices of the tree T.

Jukes Cantor binary model

Theorem (Buczyńska and Wiśniewski [BW07]). *Let T_1 and T_2 be 3-valent trees with the same number of leaves. Then*

$$\mathrm{ehr}_{P_{\mathbb{Z}_2,T_1}}(t) = \mathrm{ehr}_{P_{\mathbb{Z}_2,T_2}}(t).$$

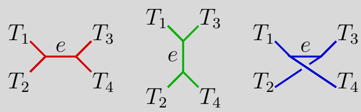
Example.

$$\mathrm{ehr}_{P_{\mathbb{Z}_2,\text{snowflake}}}(t) = \mathrm{ehr}_{P_{\mathbb{Z}_2,\text{3-caterpillar}}}(t)$$
$$= \tfrac{1}{22680}(t+1)(t+2)(t+3)$$
$$(31t^6 + 372t^5 + 1942t^4 + 5616t^3 + 9511t^2 + 8988t + 3780)$$

snowflake 3-caterpillar

Combinatorial proof (together with Haase and Paffenholz).

- The polytope of the 3-valent 4-leaf tree T_0 is the join of two squares. To each pair of non-isomorphic labelings of the leaves of T_0 we associate triangulations of the associated polytopes together with an isomorphism of the triangulations.
- Given four pointed trees T_i, where $i = 1, \ldots, 4$, we can produce a tree T by grafting the tree T_i along the i-th leaf of the labeled T_0. There is a natural projection from $P_{\mathbb{Z}_2,T}$ to $P_{\mathbb{Z}_2,T_0}$. The preimage of a triangulation of $P_{\mathbb{Z}_2,T_0}$ gives a subdivision of $P_{\mathbb{Z}_2,T}$.



trees differing by an elementary mutation along e

- Consider polytopes of two trees that differ by an elementary mutation. Their subdivisions are isomorphic via a piecewise unimodular transformation.
- Buczyńska and Wiśniewski showed that any two 3-valent trees with the same number of leaves differ by a sequence of elementary mutations. Composing corresponding maps gives a piecewise unimodular transformation between the associated polytopes.

Kimura 3-parameter model

- Transition matrices of the Kimura 3-parameter model are invariant under the $\mathbb{Z}_2 \times \mathbb{Z}_2$ action.
- Can we generalize the theorem of Buczyńska and Wiśniewski to the Kimura 3-parameter model?

Theorem (K. [K10]).

$$\mathrm{ehr}_{P_{\mathbb{Z}_2 \times \mathbb{Z}_2,\text{snowflake}}}(t) \neq \mathrm{ehr}_{P_{\mathbb{Z}_2 \times \mathbb{Z}_2,\text{3-caterpillar}}}(t).$$

Proof.

$$\mathrm{ehr}_{P_{\mathbb{Z}_2 \times \mathbb{Z}_2,\text{snowflake}}}(3) = 69248000$$
$$\neq 69324800 = \mathrm{ehr}_{P_{\mathbb{Z}_2 \times \mathbb{Z}_2,\text{3-caterpillar}}}(3).$$

□

References

[BW07] W. Buczyńska and J. Wiśniewski. On the geometry of binary symmetric models of phylogenetic trees. *J. Eur. Math. Soc.*, 9:609–635, 2007.

[ERSS05] N. Eriksson, K. Ranestad, B. Sturmfels and S. Sullivant. Phylogenetic algebraic geometry. In *Projective Varieties with Unexpected Properties*. Berlin, 2005.

[K10] K. Kubjas. Hilbert polynomial of the Kimura 3-parameter model. http://arxiv.org/abs/1007.3164, 2010.