JOURNAL OF ALGEBRAIC STATISTICS
Vol. 3, No. 1, 2012, 64-69
ISSN 1309-3452 – www.jalgstat.com

Hilbert Polynomial of the Kimura 3-Parameter Model

Kaie Kubjas[1]

[1] *Institut für Mathematik, Fachbereich Mathematik und Informatik, Freie Universität Berlin, Berlin, Germany*

Abstract. In [2] Buczyńska and Wiśniewski showed that the Hilbert polynomial of the algebraic variety associated to the Jukes-Cantor binary model on a trivalent tree depends only on the number of leaves of the tree and not on its shape. We ask if this can be generalized to other group-based models. The Jukes-Cantor binary model has $\mathbb{Z}_2$ as the underlying group. We consider the Kimura 3-parameter model with $\mathbb{Z}_2 \times \mathbb{Z}_2$ as the underlying group. We show that the generalization of the statement about the Hilbert polynomials to the Kimura 3-parameter model is not possible as the Hilbert polynomial depends on the shape of a trivalent tree.

2000 Mathematics Subject Classifications: 13P99, 52B20

Key Words and Phrases: Kimura 3-parameter model, Hilbert polynomial, toric fiber products, lattice polytopes

---

1. Introduction

Phylogenetic algebraic geometry studies complex algebraic varieties arising from evolutionary models in biology (see [4] and [9] for introduction). Models invariant under an Abelian group action are called group-based models. The Jukes-Cantor binary model is the simplest group-based model with the underlying group $\mathbb{Z}_2$. Buczyńska and Wiśniewski showed in [2] that the Hilbert polynomial of the algebraic variety associated to the Jukes-Cantor binary model on a trivalent tree T depends only on the number of leaves of T and not on the shape. We ask if this property of the Hilbert polynomial can be generalized to more complex models.

The most natural generalization would be the Kimura 3-parameter model, which is a group-based model with the underlying group $\mathbb{Z}_2 \times \mathbb{Z}_2$. However, we show that the Hilbert polynomial of the algebraic variety associated to the Kimura 3-parameter model depends on the shape of a trivalent tree. We do this by considering two different trees with 6 leaves – the caterpillar tree with 6 leaves and the snowflake tree (see Figure 1). This is the smallest interesting case with more than one trivalent tree with the same number of leaves. The Kimura 3-parameter model being the closest model to the Jukes-Cantor

Email address: kubjas@zedat.fu-berlin.de (K. Kubjas)

binary model, it is unlikely that the property about Hilbert polynomials would hold for other models.

In Section 2 we recall the construction of the Kimura 3-parameter model. In Section 3 we show that the Hilbert polynomials of the algebraic varieties associated to the Kimura 3-parameter model on the caterpillar tree with 6 leaves and the snowflake tree have different values when evaluated at 3, and hence their Hilbert polynomials are different. The main idea is to decompose the original trees to smaller trees and use toric fiber products introduced by Sullivant in [11]. Finally we reduce the problem of evaluating the Hilbert polynomials of toric varieties to evaluating the Ehrhart polynomials of the corresponding polytopes. Computations are done with `polymake` [5], [6] and `Normaliz` [1].

2. Kimura 3-Parameter Model

First the toric ideals and the corresponding lattice polytopes of the Kimura 3-parameter model will be defined following Section 3.4 in [11]. Then the toric fiber product structure on these ideals will be explained following Section 1 and Section 3.4 in [11]. We will not give the parametric construction of the Kimura 3-parameter model coming from biology as it is not necessary for this article. For the parametric construction see [9]. The geometry of the Kimura 3-parameter model is studied by Casanellas and Fernandez-Sanchez in [3].

Let T be a tree with $n+1$ leaves labeled by $1,\dots,n+1$, let the root be at the leaf $n+1$, and direct the edges away from the root. A leaf denotes here a leaf edge. A leaf l is a descendant of an edge e if there is a directed path from e to l. Denote by $\mathrm{de}(e)$ the set of all descendants of the edge e.

For a sequence $g_1,\dots,g_n$ in $\mathbb{Z}_2\times\mathbb{Z}_2$, we define

$$g_e = \sum_{i\in\mathrm{de}(e)} g_i,$$

where e is an edge of T and the subindices i denote simultaneously leaves and their labels. Let

$$\mathbb{K}[q]=\mathbb{K}[q_{g_1,\dots,g_n}|g_i\in\mathbb{Z}_2\times\mathbb{Z}_2] \text{ and } \mathbb{K}[a]=\mathbb{K}[a_h^{(e)}|e\in E(T),h\in\mathbb{Z}_2\times\mathbb{Z}_2]$$

and consider the ring homomorphism

$$\begin{aligned}\phi_T : \mathbb{K}[q] &\to \mathbb{K}[a]\\ q_{g_1,\dots,g_n} &\mapsto \prod_{e\in E(T)} a_{g_e}^{(e)}.\end{aligned}$$

Definition 1. *The ideal of the Kimura 3-parameter model on a tree T is $I_T = ker(\phi_T)$.*

This is a projective toric ideal, and the corresponding lattice polytope can be defined.

Definition 2. *The lattice polytope of the Kimura 3-parameter model on a tree T is*

$$P_T = conv(\{\alpha\in\mathbb{Z}^{E(T)\times(\mathbb{Z}_2\times\mathbb{Z}_2)}|a^\alpha=\phi_T(q_{g_1,\dots,g_n}),q_{g_1,\dots,g_n}\in\mathbb{K}[q]\}).$$

Definition 3. *Define the lattice* $L_T \subseteq \mathbb{Z}^{E(T)\times(\mathbb{Z}_2\times\mathbb{Z}_2)}$ *to be the lattice generated by the vertices of* P_T.

Since T is an acyclic directed graph, there is an induced partial order on the edges of T. Namely $e < e'$ if there is a directed path from e' to e. Let T be a tree that contains an interior edge e. Then e induces a decomposition of T as $T_e^+ * T_e^-$ where T_e^- is a subtree of T consisting of all edges $e' \in T$ with $e' \leq e$ and T_e^+ consists of all edges $e' \in T$ with $e' \not< e$. Thus T_e^+ and T_e^- overlap in the single edge e. We root T_e^- by the tail of e, and keep the root of T_e^+ at the original root $n+1$. Without loss of generality, we may assume that the nonroot leaves of T_e^+ are $\{1, 2, \ldots, m\}$ and of T_e^- are $\{e, m+1, \ldots, n\}$.

Denote by $I_{T^+}^e$ and $I_{T^-}^e$ the ideals of the Kimura 3-parameter model on trees T_e^+ and T_e^-, and by $\mathbb{K}[q]_+$ and $\mathbb{K}[q]_-$ the ambient polynomial rings, respectively. For each variable $q_{\mathbf{g}}$ in $\mathbb{K}[q], \mathbb{K}[q]_+$ and $\mathbb{K}[q]_-$, let $\deg(q_{\mathbf{g}}) = e_{g_e}$. Let

$$\phi_{I_{T^+}^e, I_{T^-}^e} : \mathbb{K}[q] \to \mathbb{K}[q]_+/I_{T^+}^e \otimes_{\mathbb{K}} \mathbb{K}[q]_-/I_{T^-}^e$$

be the ring homomorphism such that

$$q_{g_1,\ldots,g_n} \mapsto q_{g_1,\ldots,g_m} \otimes q_{g_e,g_{m+1},\ldots,g_n}.$$

Note that

$$\deg(q_{g_1,\ldots,g_n}) = \deg(q_{g_1,\ldots,g_m}) = \deg(q_{g_e,g_{m+1},\ldots,g_n}) = e_{g_e} \in \{e_{(0,0)}, e_{(0,1)}, e_{(1,0)}, e_{(1,1)}\} =: \mathcal{A}.$$

Definition 4. *The toric fiber product of* $I_{T^+}^e$ *and* $I_{T^-}^e$, *denoted* $I_{T^+}^e \times_{\mathcal{A}} I_{T^-}^e$, *is the kernel of* $\phi_{I_{T^+}^e, I_{T^-}^e}$:

$$I_{T^+}^e \times_{\mathcal{A}} I_{T^-}^e = ker(\phi_{I_{T^+}^e, I_{T^-}^e}).$$

The following theorem about toric fiber products of ideals will be the basis for our computations in the next section.

Theorem 1 (Sullivant, Theorem 3.10 in [11])**.** *Let* T *be a tree with an interior edge* e, *and resulting decomposition* $T = T_e^+ * T_e^-$. *For each variable* $q_{\mathbf{g}}$ *in* $\mathbb{K}[q], \mathbb{K}[q]_+$ *and* $\mathbb{K}[q]_-$, *let* $deg(q_{\mathbf{g}}) = e_{g_e}$. *Then*

$$I_T = I_{T^+}^e \times_{\mathcal{A}} I_{T^-}^e.$$

3. Counting Lattice Points

Proposition 1. *The Hilbert polynomials of the ideals of the Kimura 3-parameter model on the caterpillar tree with 6 leaves and the snowflake tree are different.*

Proof. Let T be a trivalent tree. In [7] Michałek shows that the lattice polytope P_T is normal, hence its Ehrhart polynomial equals the Hilbert polynomial and the Hilbert function of I_T (see Theorem 13.11 in [8]). This allows us to use these notions interchangeably.

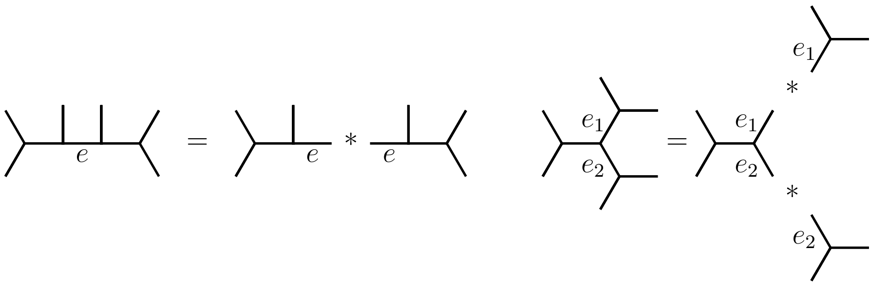

Figure 1: Decompositions of the caterpillar tree with 6 leaves and the snowflake tree

1. Since the polytopes of the caterpillar with 6 leaves and snowflake trees are too large to compute their lattice points directly, we decompose them into smaller trees like shown in Figure 1.

   Henceforth we use the abbreviations c6, sn, 3l, 4l for the caterpillar with 6 leaves, snowflake, 3-leaf and trivalent 4-leaf trees, respectively.

   In the decomposition of the caterpillar tree with 6 leaves define $\deg(q_{\mathbf{g}}) = e_{g_e}$ for $q_{\mathbf{g}}$ in $\mathcal{K}[q]_{\mathrm{c6}}$ and $\mathcal{K}[q]_{\mathrm{4l}}$. Then
   $$I_{\mathrm{c6}} = I_{\mathrm{4l}}^e \times_{\mathcal{A}} I_{\mathrm{4l}}^e$$
   with $\mathcal{A} = \{e_{(0,0)}, e_{(0,1)}, e_{(1,0)}, e_{(1,1)}\}$.

   In the decomposition of the snowflake tree define $\deg(q_{\mathbf{g}}) = e_{g_{e_1}, g_{e_2}}$ for $q_{\mathbf{g}}$ in $\mathcal{K}[q]_{\mathrm{sn}}$ and $\mathcal{K}[q]_{\mathrm{4l}}$ and define $\deg(q_{\mathbf{g}}) = e_{g_{e_i}}$ with $i \in \{1,2\}$ for $q_{\mathbf{g}}$ in $\mathcal{K}[q]_{\mathrm{3l}}$. Then
   $$I_{\mathrm{sn}} = (I_{\mathrm{4l}}^{e_1,e_2} \times_{\mathcal{A}} I_{\mathrm{3l}}^{e_1}) \times_{\mathcal{A}} I_{\mathrm{3l}}^{e_2}$$
   with $\mathcal{A} = \{e_{(0,0)}, e_{(0,1)}, e_{(1,0)}, e_{(1,1)}\}$. Abusing the notation slightly, the first toric fiber product corresponds to the decomposition of the 5-leaf tree into a 4-leaf tree and a 3-leaf tree with respect to the edge e_1 and the second toric fiber product corresponds to the decomposition of the snowflake tree into the 5-leaf tree of the first fiber product and a 3-leaf tree with respect to the edge e_2.

2. Denote the multigraded Hilbert function of $\mathbb{K}[q]/I$ by $h(\mathbb{K}[q]/I; u)$, where $u \in \mathbb{N}^{\mathbb{Z}_2 \times \mathbb{Z}_2}$. Corollary 2.12 in [11] gives a formula for computing multigraded Hilbert functions of toric fiber products. Applying this to the decompositions of Step 1 gives for $u, v \in \mathbb{N}^{\mathbb{Z}_2 \times \mathbb{Z}_2}$
   $$h(\mathbb{K}[q]_{\mathrm{c6}}/I_{\mathrm{c6}}; u) = h(\mathbb{K}[q]_{\mathrm{4l}}/I_{\mathrm{4l}}^e; u) h(\mathbb{K}[q]_{\mathrm{4l}}/I_{\mathrm{4l}}^e; u),$$
   $$h(\mathbb{K}[q]_{\mathrm{sn}}/I_{\mathrm{sn}}; u, v) = (h(\mathbb{K}[q]_{\mathrm{4l}}/I_{\mathrm{4l}}^{e_1,e_2}; u, v) h(\mathbb{K}[q]_{\mathrm{3l}}/I_{\mathrm{3l}}^{e_1}; u)) h(\mathbb{K}[q]_{\mathrm{3l}}/I_{\mathrm{3l}}^{e_2}; v).$$
   For the snowflake tree we apply the formula twice, and take into account that the edge e_2 in the 5-leaf tree belongs to the 4-leaf tree when decomposing the 5-leaf tree.

3. A monomial having multidegree $u \in \mathbb{N}^{\mathbb{Z}_2 \times \mathbb{Z}_2}$ has total degree $\sum_{h \in \mathbb{Z}_2 \times \mathbb{Z}_2} u_h$. Thus single graded Hilbert functions can be computed using multigraded Hilbert functions
   $$h(\mathbb{K}[q]_{\mathrm{c6}}/I_{\mathrm{c6}}; k) = \sum_{u : \sum u_h = k} h(\mathbb{K}[q]_{\mathrm{c6}}/I_{\mathrm{c6}}; u),$$

$$h(\mathbb{K}[q]_{\mathrm{sn}}/I_{\mathrm{sn}};k) = \sum_{u,v:\sum u_h=k,\sum v_h=k} h(\mathbb{K}[q]_{\mathrm{sn}}/I_{\mathrm{sn}};u,v).$$

4. Because of normality, $h(\mathbb{K}[q]/I_T;u)$ counts lattice points in the lattice L_T of the $\sum_{h\in\mathbb{Z}_2\times\mathbb{Z}_2} u_h$ dilation of the polytope P_T intersected with hyperplanes $\{x_h^e = u_h\}, h \in \mathbb{Z}_2\times\mathbb{Z}_2$. Denote the Ehrhart polynomial of a lattice polytope P by $\mathrm{ehr}_P(k)$, where $k\in\mathbb{N}$. Using Step 2 and Step 3 we get

$$\begin{aligned}
\mathrm{ehr}_{P_{\mathrm{c6}}}(k) &= \sum_{u:\sum u_h=k} \left| kP_{41}\bigcap\{x_h^e = u_h\}\bigcap L_{41}||kP_{41}\bigcap\{x_h^e=u_h\}\bigcap L_{41}\right|,\\
\mathrm{ehr}_{P_{\mathrm{sn}}}(k) &= \sum_{u,v:\sum u_h=k,\sum v_h=k} \left|kP_{41}\bigcap\{x_h^{e_1}=u_h\}\bigcap\{x_h^{e_2}=v_h\}\bigcap L_{41}\right|\\
&\cdot \left|kP_{31}\bigcap\{x_h^{e_1}=u_h\}\bigcap L_{31}\right|\left|kP_{31}\bigcap\{x_h^{e_2}=v_h\}\bigcap L_{31}\right|.
\end{aligned}$$

5. Using `polymake` and `Normaliz` we can compute $|3P_T\cap\{x_l^e=u_l\}\cap L_T|$ for 3-leaf and 4-leaf trees. It is important to do the basis transformation before counting lattice points, since these programs assume that the lattice is the standard lattice. Using formulas from Step 4 we get that in the 3rd dilation the polytope of the Kimura 3-parameter model on the caterpillar tree with 6 leaves has 69324800 and the polytope of the Kimura 3-parameter model on the snowflake tree has 69248000 lattice points. Hence their Ehrhart (and thus Hilbert) polynomials are different.

Remark 1. *Similar computations show that in the 2nd dilation the polytopes of the Kimura 3-parameter model on the caterpillar tree with 6 leaves and on the snowflake tree have both 396928 lattice points.*

Proposition 1 does not directly imply that for $n+1\geq 7$ there exist trivalent trees T' and T'' with $n+1$ leaves such that the Hilbert polynomials of the ideals of the Kimura 3-parameter model on T' and T'' are different. However, from Proposition 1 follows that the multigraded Hilbert function on $\mathbb{K}[q]_{41}/I_{41}$, where the multigrading is induced by the leaves of the 4-leaf tree, depends on the labelling of the leaves. For the Jukes-Cantor binary model this multigraded Hilbert function is independent of the labelling of the leaves by Corollary 7.12 in [10], which also explains the invariance of the Hilbert polynomial.

ACKNOWLEDGEMENTS We thank Andreas Paffenholz for technical advice and Christian Haase for independently checking our computations.

References

[1] Winfried Bruns and Bogdan Ichim. `Normaliz 2.1`. Available at `http://www.math.uos.de/normaliz`.

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

[3] Marta Casanellas and Jesus Fernandez-Sanchez. Geometry of the Kimura 3-parameter model. *Adv. Appl. Math*, 41:265–292, 2008.

[4] Nicholas Eriksson, Kristian Ranestad, Bernd Sturmfels, and Seth Sullivant. Phylogenetic algebraic geometry. In *Projective Varieties with Unexpected Properties*. Walter de Gruyter, Berlin, 2005.

[5] Michael Joswig and Ewgenij Gawrilow. `polymake 2.9.7`. Available at `http://www.polymake.de`.

[6] Michael Joswig, Benjamin Müller, and Andreas Paffenholz. Polymake and lattice polytopes. In *DMTCS Proceedings of FPSAC 2009*, 2009.

[7] Mateusz Michałek. Algebraic varieties representing group-based markov processes on trees. *J. Algebra*, 339:339–356, 2011.

[8] Bernd Sturmfels. *Gröbner bases and convex polytopes*. University Lecture Series. American Mathematical Society, Providence, RI, 1996.

[9] Bernd Sturmfels and Seth Sullivant. Toric ideals of phylogenetic invariants. *J. Comput. Biol.*, 12:204–228, 2005.

[10] Bernd Sturmfels and Zhiqiang Xu. Sagbi bases of Cox-Nagata rings. *J. Eur. Math. Soc.*, 12:429–459, 2010.

[11] Seth Sullivant. Toric fiber products. *J. Algebra*, 316:560–577, 2007.