JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Detecting Phylogenetic Learning Coefficient in General Identical

Zsinka Franciska
Dept. of Mathematics
American Mathematics Society
United States

We consider the problem of finding generators of the toric ideal associated to a combinatorial object called a staged tree. Our motivation to consider this problem originates from the use of staged trees to represent discrete statistical models such as conditional independence models and discrete Bayesian networks. The main theorem in this article states that toric ideals of staged trees that are balanced and stratified are generated by a quadratic Gröbner basis whose initial ideal is square-free. We apply this theorem to construct Gröbner bases of a subclass of discrete statistical models represented by staged trees. The proof of the main result is based on Sullivant's toric fiber product construction (*J. Algebra* **316**:2 (2007), 560–577).

1. Introduction

The study of toric ideals associated to statistical models was pioneered by the work of Diaconis and Sturmfels [4], who first used the generators of a toric ideal to formulate a sampling algorithm for discrete distributions. Since then, and with the subsequent work of [2; 17; 8], the study of toric ideals of discrete statistical models has been an active area of research in algebraic statistics. The books by Sullivant [18, Chapter 9] and Aoki, Hara and Takemura [1] are good references to learn about the role of toric ideals in statistics. A recent introduction to the topic, from the point of view of binomial ideals, can be found in [13, Chapter 9], which also contains a thorough list of references of previous contributions to this topic.

In 2008, Smith and Anderson [15] introduced a new graphical discrete statistical model called a *staged tree model*. This model is represented by an event tree together with an equivalence relation on its vertices. Staged tree models are useful to represent conditional independence relations among events and random variables, such as those coming from discrete graphical models. Hence, any discrete Bayesian network or decomposable model is also a staged tree model [15]. There are two properties that make staged tree models more general than Bayesian networks or decomposable models. First, the state space of a staged tree model does not have to be a cartesian product. Second, using staged tree models it is possible to represent extra context-specific conditional independence between events. The book of Collazo, Görgen and Smith [3] is a good reference to learn about these models.

In this article we define the toric ideal associated to a *staged tree* and study its properties from an algebraic and combinatorial point of view. We present Theorem 2.14, which states that toric ideals of staged trees that are *balanced* and *stratified* have quadratic Gröbner basis with square-free initial ideal. We apply Theorem 2.14 in Section 5 to obtain Gröbner bases for toric ideals of staged tree models. Our

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

results provide a new point of view on the construction of Gröbner bases for decomposable graphical models, some conditional independence models, as well as the construction of Gröbner bases for staged tree models whose underlying tree is *asymmetric*.

This article is organized as follows. In Section 2, we define the toric ideal associated to a staged tree. In Section 3, we formulate a toric fiber product construction for balanced and stratified staged trees. In Section 4, we prove our main result Theorem 2.14. Finally, in Section 5 we apply our results to compute Gröbner bases for several statistical models.

2. Staged trees

We start by defining our two objects of interest: a staged tree and its associated toric ideal. First, we set up the graph-theoretic notation and conventions. Let $\mathcal{T} = (V, E)$ be a directed rooted tree, with vertex set V and set E of directed edges. We only consider trees $\mathcal{T} = (V, E)$ where all elements in E are oriented away from the root. Since we only consider directed paths, we refer to any directed path in $\mathcal{T}$ simply as a path. For $v, w \in V$ the directed edge in E from v to w is denoted by (v, w), the set of children of v is $\mathrm{ch}(v) = \{u \mid (v, u) \in E\}$, and the set of outgoing edges from v is $E(v) = \{(v, u) \mid u \in \mathrm{ch}(v)\}$. A vertex $v \in V$ is a leaf if $\mathrm{ch}(v) = \varnothing$ and it is an nonroot vertex if it is different from the root.

Definition 2.1. Let $\mathcal{T} = (V, E)$ be a tree, $\mathcal{L}$ a finite set of labels, and $\theta : E \to \mathcal{L}$ a surjective function. For each $v \in V$, $\theta_v := \{\theta(e) \mid e \in E(v)\}$ is the set of labels associated to v. The pair $(\mathcal{T}, \theta)$ is a *staged tree* if

(i) for each $v \in V$, we have $|\theta_v| = |E(v)|$, and

(ii) for any two vertices $v, w \in V$ either $\theta_v = \theta_w$ or $\theta_v \cap \theta_w = \varnothing$.

Condition (ii) in Definition 2.1 defines an equivalence relation on the set of nonleaf vertices of $\mathcal{T}$. Namely, two nonleaf vertices $v, w \in V$ are equivalent if and only if $\theta_v = \theta_w$. We refer to the partition induced by this equivalence relation on the set of nonleaf vertices as the set of stages of $\mathcal{T}$ and to a single element in this partition as a *stage*. Condition (i) in Definition 2.1 guarantees that all edge labels associated to a single vertex are distinct. We use $(\mathcal{T}, \theta)$ to denote a staged tree with labeling rule θ. For simplicity we will often drop the use of θ and write $\mathcal{T}$ for a staged tree.

To define the toric ideal associated to $(\mathcal{T}, \theta)$ we define two polynomial rings. The first ring is $\mathbb{R}[p]_{\mathcal{T}} := \mathbb{R}[p_\lambda \mid \lambda \in \Lambda]$, where Λ is the set of root-to-leaf paths in $\mathcal{T}$. The second ring is $\mathbb{R}[\Theta]_{\mathcal{T}} := \mathbb{R}[z, \mathcal{L}]$, where the labels in $\mathcal{L}$ are indeterminates together with a homogenizing variable z. For a path γ in $\mathcal{T}$, $E(\gamma)$ is the set of edges in γ.

Definition 2.2. The *toric staged tree ideal* associated to $(\mathcal{T}, \theta)$ is the kernel of the ring homomorphism $\varphi_{\mathcal{T}} : \mathbb{R}[p]_{\mathcal{T}} \to \mathbb{R}[\Theta]_{\mathcal{T}}$ defined as

$$p_\lambda \mapsto z \cdot \prod_{e \in E(\lambda)} \theta(e). \tag{1}$$

The ideal $\ker(\varphi_{\mathcal{T}})$ defines the toric variety specified as the closure of the image of the monomial parametrization $\Phi_{\mathcal{T}} : (\mathbb{C}^*)^{|\mathcal{L}|} \to \mathbb{P}^{|\Lambda|-1}$ given by $(\theta(e) \in \mathcal{L}) \mapsto \big(z \cdot \prod_{e \in E(\lambda)} \theta(e)\big)_{\lambda \in \Lambda}$. We use the homogenizing variable z in Definition 2.2 to consider the projective toric variety in $\mathbb{P}^{|\Lambda|-1}$.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

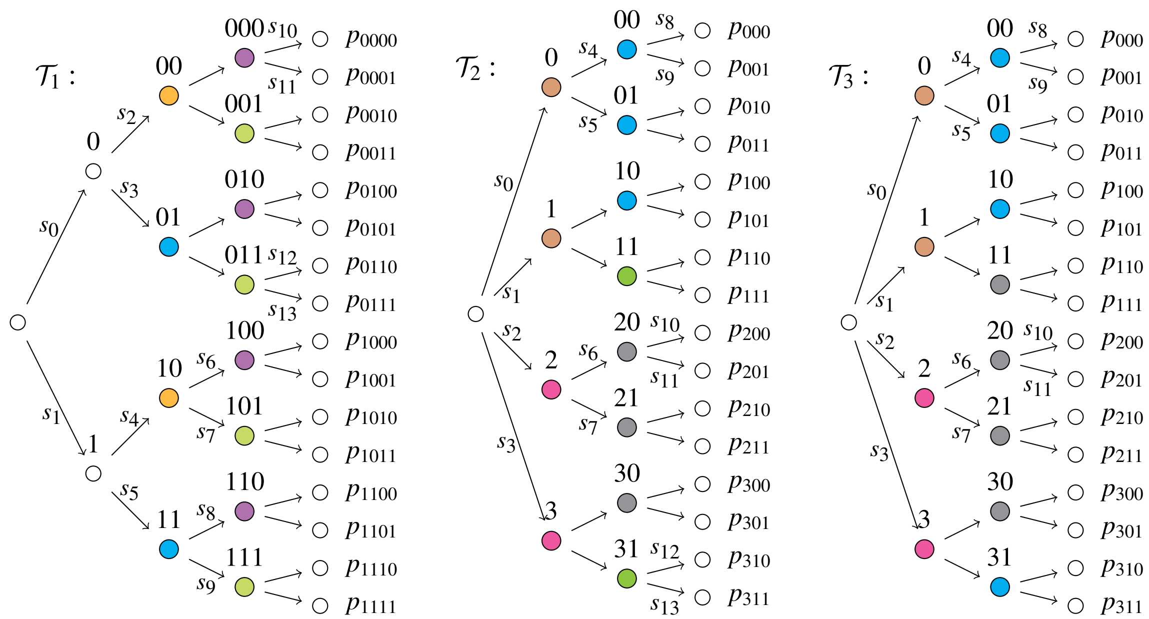

Figure 1. Three examples of staged trees. In each tree two vertices with the same color are in the same stage.

It is often useful to encode a monomial map between polynomial rings by using an exponent matrix. Let $B = (b_{ij})$ be a $d \times n$ matrix with nonnegative integer entries. The columns of B define a monomial map $\phi_B : \mathbb{R}[z_1, \dots, z_n] \to \mathbb{R}[t_1, \dots, t_d]$, $z_j \mapsto \prod_{i=1}^{d} t_i^{b_{ij}}$. The matrix B is the *exponent matrix* of ϕ_B. We will use this notation in Sections 3 and 4.

Example 2.3. The staged tree $\mathcal{T}_1$ in Figure 1 has label set $\mathcal{L} = \{s_0, \dots, s_{13}\}$. Each vertex in $\mathcal{T}_1$ is denoted by a string of 0s and 1s, and each edge has a label associated to it. The root-to-leaf paths in $\mathcal{T}_1$ are denoted by p_{ijkl}, where $i, j, k, l \in \{0, 1\}$. A vertex in $\mathcal{T}_1$ represented with a blank circle indicates a stage consisting of a single vertex. We use colors in the vertices of $\mathcal{T}_1$ to indicate which vertices are in the same stage. For instance, the set of purple vertices $\{000, 010, 100, 110\}$ are in the same stage and therefore they have the same set $\{s_{10}, s_{11}\}$ of associated edge labels. The map $\phi_{\mathcal{T}_1}$ sends $(s_0, \dots, s_{13})$ to

$$(s_0s_2s_6s_{10},\ s_0s_2s_6s_{11},\ s_0s_2s_7s_{12},\ s_0s_2s_7s_{13},\ s_0s_3s_8s_{10},\ s_0s_3s_8s_{11},\ s_0s_3s_9s_{12},\ s_0s_3s_9s_{13},$$
$$s_1s_4s_6s_{10},\ s_1s_4s_6s_{11},\ s_1s_4s_7s_{12},\ s_1s_4s_7s_{13},\ s_1s_5s_8s_{10},\ s_1s_5s_8s_{11},\ s_1s_5s_9s_{12},\ s_1s_5s_9s_{13}).$$

The toric ideal $\ker(\varphi_{\mathcal{T}_1})$ is generated by a quadratic Gröbner basis with square-free initial ideal.

Example 2.4. Consider the two staged trees $\mathcal{T}_2, \mathcal{T}_3$ depicted in Figure 1. For the staged tree $\mathcal{T}_2$, $\ker(\varphi_{\mathcal{T}_2})$ is generated by a quadratic Gröbner basis with square-free initial ideal. For $\mathcal{T}_3$, the ideal $\ker(\varphi_{\mathcal{T}_3})$ also has a Gröbner basis with square-free initial ideal but its elements are of degree 2 and degree 4.

We are interested in relating the combinatorial properties of the staged tree $(\mathcal{T}, \theta)$ with the properties of the toric ideal $\ker(\varphi_{\mathcal{T}})$. The two definitions that are relevant for the statement of the main theorem

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

are the definition of *balanced* staged tree and of *stratified* staged tree. Before stating the main theorem, Theorem 2.14, we look into the definition and consequences of these two notions.

Definition 2.5. Let $\mathcal{T}$ be a tree. For $v \in V$, the *level* of v is denoted by $\ell(v)$ and it is equal to the number of edges in the unique path from the root of $\mathcal{T}$ to v. If all the leaves in $\mathcal{T}$ have the same level then the level of $\mathcal{T}$ is equal to the level of any of its leaves. The staged tree $(\mathcal{T}, \theta)$ is *stratified* if all its leaves have the same level and if every two vertices in the same stage have the same level.

It is easy to check that all trees in Figure 1 are stratified. Namely, we only need to verify that every two vertices with the same color are also in the same level. Notice that the combinatorial condition of $(\mathcal{T}, \theta)$ being stratified implies the algebraic condition that the map $\varphi_{\mathcal{T}}$ is square-free.

We now turn our attention to the definition of a balanced staged tree. This definition is formulated in terms of polynomials associated to each vertex of the tree. We proceed to explain their notation and basic properties.

Definition 2.6. Let $(\mathcal{T}, \theta)$ be a staged tree, $v \in V$, and $\mathcal{T}_v$ the subtree of $\mathcal{T}$ rooted at v. The tree $\mathcal{T}_v$ is a staged tree with the induced labeling from $\mathcal{T}$. Let Λ_v be the set of v-to-leaf paths in $\mathcal{T}$. The *interpolating polynomial of* $\mathcal{T}_v$ is

$$t(v) := \sum_{\lambda \in \Lambda_v} \prod_{e \in E(\lambda)} \theta(e).$$

When v is the root of $\mathcal{T}$, the polynomial $t(v)$ is called the *interpolating polynomial* of $\mathcal{T}$. Two staged trees $(\mathcal{T}, \theta)$ and $(\mathcal{T}, \theta')$ with the same label set $\mathcal{L}$ are *polynomially equivalent* if their interpolating polynomials are equal.

The interpolating polynomial of a staged tree is useful to capture symmetries among subtrees. It is also an important tool in the study of the statistical properties of staged tree models. This polynomial was defined by Görgen and Smith in [10] and further studied by Görgen et al. in [11]. Although these two articles are written for a statistical audience, we would like to emphasize that their symbolic algebra approach to the study of statistical models proves to be very important for the use of these models in practice. We will define the statistical model associated to a staged tree and connect to other results in algebraic statistics in Section 5.

If $(\mathcal{T}, \theta)$ is a staged tree, the polynomials $t(\cdot)$ satisfy a recursive relation. This relation is useful to prove statements about the algebraic and combinatorial properties of $\mathcal{T}$. We state this property as a lemma.

Lemma 2.7 [11, Theorem 1]. *Let $(\mathcal{T}, \theta)$ be a staged tree, $v \in V$ and* $\mathrm{ch}(v) = \{v_0, \dots, v_k\}$. *Then the polynomial $t(v)$ admits the recursive representation $t(v) = \sum_{i=0}^{k} \theta(v, v_i) t(v_i)$.*

Example 2.8. Consider the staged tree $\mathcal{T}_1$ in Figure 1. If v and w are orange and blue vertices in $\mathcal{T}_1$ respectively and r is the root of $\mathcal{T}_1$ then

$$t(v) = s_6(s_{10} + s_{11}) + s_7(s_{12} + s_{13}),$$
$$t(w) = s_8(s_{10} + s_{11}) + s_9(s_{12} + s_{13}),$$
$$t(r) = (s_0 s_2 + s_1 s_4) t(v) + (s_0 s_3 + s_1 s_5) t(w).$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Definition 2.9. Let $(\mathcal{T}, \theta)$ be a staged tree and let v, w be two vertices in the same stage, with $\mathrm{ch}(v) = \{v_0, \dots, v_k\}$ and $\mathrm{ch}(w) = \{w_0, \dots, w_k\}$. After a possible reindexing of the elements in $\mathrm{ch}(w)$, we may assume that $\theta(v, v_i) = \theta(w, w_i)$ for all $i \in \{0, \dots, k\}$. The pair v, w is balanced if

$$t(v_i)t(w_j) = t(w_i)t(v_j) \quad \text{in } \mathbb{R}[\Theta]_{\mathcal{T}} \text{ for all } i \neq j \in \{0, \dots, k\}. \tag{2}$$

The staged tree $(\mathcal{T}, \theta)$ is *balanced* if every pair of vertices in the same stage is balanced.

Example 2.10. The two staged trees $\mathcal{T}_2, \mathcal{T}_3$ in Figure 1 are not balanced. The pair of pink vertices in $\mathcal{T}_2$ is not balanced because $(s_{10} + s_{11})(s_{12} + s_{13}) \neq (s_{10} + s_{11})^2$. By a similar argument we can check that $\mathcal{T}_3$ is also not balanced.

Although the balanced condition in Definition 2.9 seems to be algebraic and hard to check, in many cases it is very combinatorial. To formulate a precise instance where this is true we need the following definition.

Definition 2.11. Let $(\mathcal{T} = (V, E), \theta)$ be a staged tree. We say that two vertices $v, w \in V$ are in the same *position* if they are in the same stage and $t(v) = t(w)$.

The notion of position for vertices in the same stage was formulated in [15]. Intuitively it means that if we regard the subtrees $\mathcal{T}_v$ and $\mathcal{T}_w$ as representing the unfolding of a sequence of events, then the future of v and w is essentially the same. In the next lemma we use positions of vertices to provide a sufficient condition on a stratified staged tree $(\mathcal{T}, \theta)$ to be balanced.

Lemma 2.12. *Let $(\mathcal{T}, \theta)$ be a stratified staged tree. Suppose that every two vertices in $\mathcal{T}$ that are in the same stage are also in the same position. Then $(\mathcal{T}, \theta)$ is balanced.*

Proof. Following Definition 2.9, it suffices to prove that any pair of vertices in the same position is balanced. Let v, w be two vertices in the same position. We use the same notation in Definition 2.9 and assume without loss of generality that $\theta(v, v_i) = \theta(w, w_i)$. Using Lemma 2.7 we write

$$t(v) = t(w) \iff \sum_{i=0}^{k} \theta(v, v_i)t(v_i) = \sum_{i=0}^{k} \theta(w, w_i)t(w_i) \iff \sum_{i=0}^{k} \theta(v, v_i)(t(v_i) - t(w_i)) = 0.$$

Since $(\mathcal{T}, \theta)$ is stratified, the variables appearing in the polynomials $t(v_i), t(w_i)$ are disjoint from the set of variables $\{\theta(v, v_0), \dots, \theta(v, v_k)\}$. Thus $t(v_i) = t(w_i)$ for all $i \in \{0, \dots, k\}$. It follows that for all $i, j \in \{0, \dots, k\}$ the equality $t(v_i)t(w_j) = t(w_i)t(v_j)$ is true. Hence $(\mathcal{T}, \theta)$ is balanced. □

Example 2.13. The staged tree $\mathcal{T}_1$ in Figure 1 is balanced. This can be readily checked by noting that the blue vertices are in the same position and that the same is true for the orange vertices. The two trees in Figure 3 are examples of balanced staged trees in which the blue vertices are not in the same position.

We are now ready to state the main theorem.

Theorem 2.14. *If $(\mathcal{T}, \theta)$ is a balanced and stratified staged tree then* $\ker(\varphi_{\mathcal{T}})$ *is generated by a quadratic Gröbner basis with square-free initial ideal.*

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

We clarify that the conditions of $(\mathcal{T},\theta)$ being balanced and stratified in Theorem 2.14 are sufficient for $\ker(\varphi_{\mathcal{T}})$ to have a quadratic Gröbner basis but are not necessary. In the examples of staged trees in Figure 1, all of the trees $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ are stratified but only $\mathcal{T}_1$ is balanced. Even though $\mathcal{T}_2$ is not balanced, it has a quadratic Gröbner basis with square-free initial terms.

3. Toric fiber products

In this section we review toric fiber products following the exposition in [17]. We then use these results in Section 4 to prove Theorem 2.14.

Given a positive integer m, set $[m]=\{1,2,\dots,m\}$. Let r be a positive integer, and let s and t be two vectors of positive integers in $\mathbb{Z}^r_{>0}$. Consider the multigraded polynomial rings

$$\mathbb{K}[x] := \mathbb{K}[x^i_j \mid i\in[r],\, j\in[s_i]] \quad\text{and}\quad \mathbb{K}[y] := \mathbb{K}[y^i_k \mid i\in[r],\, k\in[t_i]]$$

graded by the same set $\mathcal{A}=\{\boldsymbol{a}^1,\dots,\boldsymbol{a}^r\}\subset\mathbb{Z}^d$, where

$$\deg(x^i_j)=\deg(y^i_k)=\boldsymbol{a}^i$$

and such that there exists a vector $w\in\mathbb{Q}^d$ such that $\langle w,\boldsymbol{a}^i\rangle=1$ for any $\boldsymbol{a}^i\in\mathcal{A}$. A polynomial in $\mathbb{K}[x]$ or $\mathbb{K}[y]$ is $\mathcal{A}$-*homogeneous* whenever it is homogeneous with respect to the multigrading given by $\mathcal{A}$. An ideal in $\mathbb{K}[x]$ or $\mathbb{K}[y]$ is $\mathcal{A}$-homogeneous if it is generated by $\mathcal{A}$-homogeneous elements. If $I\subseteq\mathbb{K}[x]$ and $J\subseteq\mathbb{K}[y]$ are $\mathcal{A}$-homogeneous ideals, then the quotient rings $R=\mathbb{K}[x]/I$ and $S=\mathbb{K}[y]/J$ are also multigraded rings. Let

$$\mathbb{K}[z] := \mathbb{K}[z^i_{jk} \mid i\in[r],\, j\in[s_i],\, k\in[t_i]]$$

and consider the ring homomorphism

$$\begin{aligned}\phi_{I,J}:\mathbb{K}[z]&\to R\otimes_{\mathbb{K}} S,\\ z^i_{jk}&\mapsto \bar{x}^i_j\otimes\bar{y}^i_k,\end{aligned}$$

where $\bar{x}^i_j$ and $\bar{y}^i_k$ are the equivalence classes of x^i_j and y^i_k respectively.

Definition 3.1. The *toric fiber product* of I and J is $I\times_{\mathcal{A}} J := \ker(\phi_{I,J})$.

We now recall results from [17] about the generators of the ideal $I\times_{\mathcal{A}} J$. The generators of $I\times_{\mathcal{A}} J$ come in two flavors; new quadratic generators that are created by the toric fiber product construction and new generators that are lifts of generators from I and J.

Consider the monomial parametrization

$$\begin{aligned}\phi_B:\mathbb{K}[z^i_{jk}\mid i\in[r],\, j\in[s_i],\, k\in[t_i]]&\to\mathbb{K}[x^i_j,\, y^i_k\mid i\in[r],\; j\in[s_i],\; k\in[t_i]],\\ z^i_{jk}&\mapsto x^i_j y^i_k,\end{aligned}$$

where B is the exponent matrix of ϕ_B. Let

$$\mathrm{Quad}_B := \langle \underline{z^i_{j_1k_2}z^i_{j_2k_1}} - z^i_{j_1k_1}z^i_{j_2k_2} \mid 1\le i\le r,\; 1\le j_1<j_2\le s_i,\; 1\le k_1<k_2\le t_i\rangle.$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

By [17, Proposition 10], the elements in Quad_B are a Gröbner basis of the ideal $I_B := \ker(\phi_B)$ with respect to any term order that selects the underlined terms as leading terms. The elements in Quad_B are new quadratic generators created by the toric fiber product construction of I and J.

The construction of the generators of $I \times_{\mathcal{A}} J$ that are lifts to the ring $\mathbb{K}[z]$ of elements in I and J is explained in full generality in [17]. Since we will only consider lifts of pure quadratic binomials, we restrict the definition from [17] to this case. We define lifts of $\mathcal{A}$-homogeneous elements in $\mathbb{K}[x]$, an analogous construction works to define lifts of elements in $\mathbb{K}[y]$.

Consider the $\mathcal{A}$-homogeneous polynomial

$$f = x_{a_1}^{i_1} x_{a_2}^{i_2} - x_{a_3}^{i_1} x_{a_4}^{i_2} \in \mathbb{K}[x],$$

where $i_1, i_2 \in [r],\ a_1, a_3 \in [s_{i_1}],\ a_2, a_4 \in [s_{i_2}]$ and $f \in I$. Set $k = (k_1, k_2)$ with $k_1 \in [t_{i_1}],\ k_2 \in [t_{i_2}]$ and consider $f_k \in \mathbb{K}[z]$ defined by

$$f_k = x_{a_1 k_1}^{i_1} x_{a_2 k_2}^{i_2} - x_{a_3 k_1}^{i_1} x_{a_4 k_2}^{i_2}.$$

The new $\mathcal{A}$-homogeneous polynomial f_k is in $I \times_{\mathcal{A}} J$ for all k because $f \in I$.

Definition 3.2. Let $\mathcal{A}$ be linearly independent and let $F \subset I$ be a collection of pure and quadratic $\mathcal{A}$-homogeneous polynomials. We associate to each $f \in F$ the set $T_f = [t_{i_1}] \times [t_{i_2}]$ of indices and define

$$\mathrm{Lift}(F) = \{f_k \mid f \in F,\ k \in T_f\}.$$

The set $\mathrm{Lift}(F)$ is called the lifting of F to $I \times_{\mathcal{A}} J$. For a collection H of $\mathcal{A}$-homogeneous elements of J, we define $\mathrm{Lift}(H)$ in a similar way.

We are now ready to state the result from [17] that we will use in the proof of Theorem 2.14. The important part of this theorem is that we can construct Gröbner basis of the toric fiber product $I \times_{\mathcal{A}} J$ by using lifts of Gröbner bases for I and J together with the elements in Quad_B.

Theorem 3.3 [17, Theorem 12]. *Suppose that $\mathcal{A}$ is linearly independent. Let $F \subset I$ be a homogeneous Gröbner basis for I with respect to the weight vector ω_1 and let $H \subset J$ be a homogeneous Gröbner basis for J with respect to the weight vector ω_2. Let ω be a weight vector such that* Quad_B *is a Gröbner basis for I_B. Then the set* $\mathrm{Lift}(F) \cup \mathrm{Lift}(H) \cup \mathrm{Quad}_B$ *is a Gröbner basis for $I \times_{\mathcal{A}} J$ with respect to the weight order $\phi_B^*(\omega_1, \omega_2) + \epsilon\omega$ for sufficiently small $\epsilon > 0$.*

4. Proof of main theorem

In the first part of this section we explain how to construct staged trees from smaller pieces and relate this construction to toric fiber products in Proposition 4.4. Then, in Proposition 4.11 we use this construction repeatedly for balanced and stratified staged trees. Finally, we use these results together with Theorem 3.3 to prove our main theorem.

Let $(\mathcal{T}, \theta)$ be a staged tree. We recursively define an indexing on the set of nonroot vertices of $\mathcal{T}$. The children of the root are indexed by $\{0, 1, \ldots, k\}$. If $\boldsymbol{a}$ is the index of a vertex v and $|E(v)| = j + 1$,

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

then we index the children of $\boldsymbol{a}$ by $\boldsymbol{a}0, \ldots, \boldsymbol{a}j$. This way each nonroot vertex in V is indexed by a finite sequence of nonnegative integers

$$\boldsymbol{a} = a_1 a_2 \cdots a_\ell,$$

where ℓ is the level of the vertex indexed by $\boldsymbol{a}$. From this point on we refer to any nonroot vertex in V via its index $\boldsymbol{a}$. All vertices of the trees in Figure 1 are indexed following this rule. In Figure 1 the index of each vertex is displayed immediately above each vertex and on the side for the leaves. We denote by $\boldsymbol{i}_{\mathcal{T}}$ the set of indices of the leaves in $\mathcal{T}$.

Definition 4.1. If a staged tree has level one we call it a *level-one tree*. We reserve for it the special notation $(\mathcal{B}, \epsilon)$, where $\mathcal{B} = (V, E)$ is the tree and ϵ its labeling rule. By condition (i) in Definition 2.1, the size of the label set of $\mathcal{B}$ is equal to $|E|$. Let $E = \{e_0, \ldots, e_m\}$. We use ϵ_k to denote the image of the k-th element in E under ϵ. We also use the notation $(\mathcal{B}, \{\epsilon_0, \ldots, \epsilon_m\})$ when we wish to emphasize the label set of the level-one tree.

Definition 4.2. Let $(\mathcal{T}, \theta)$ be a staged tree and $G = \{G_1, \ldots, G_r\}$ be a partition of the set of leaves $\boldsymbol{i}_{\mathcal{T}}$. We consider a collection $\{(\mathcal{B}_i, \epsilon^{(i)}) \mid i \in [r]\}$ of level-one trees such that their label sets are pairwise disjoint and disjoint from the label set of $(\mathcal{T}, \theta)$. The *gluing component* $\mathcal{T}_G$ associated to $\mathcal{T}$ and G is

$$\mathcal{T}_G := \bigsqcup_{i \in [r]} (\mathcal{B}_i, \epsilon^{(i)}).$$

The gluing component $\mathcal{T}_G$ is a forest of level-one trees; its label set is the union of the label sets of each $(\mathcal{B}_i, \epsilon^{(i)})$. We denote by $[\mathcal{T}, \mathcal{T}_G]$ the tree obtained by gluing $\mathcal{B}_i$ to the leaf $\boldsymbol{a}$ for all $\boldsymbol{a} \in G_i$ and all $i \in [r]$.

Remark 4.3. The tree $[\mathcal{T}, \mathcal{T}_G]$ is a staged tree. Its label set is the union of the labels sets of $(\mathcal{T}, \theta)$ and $\mathcal{T}_G$. The labeling rule is inherited from the labelings of $\mathcal{T}$ and $\mathcal{T}_G$ and it satisfies conditions (i), (ii) in Definition 2.1. Moreover, $\boldsymbol{i}_{[\mathcal{T}, \mathcal{T}_G]} = \{\boldsymbol{a}k \mid \boldsymbol{a} \in G_i,\ k \in \boldsymbol{i}_{\mathcal{B}_i},\ i \in [r]\}$. The stages in $[\mathcal{T}, \mathcal{T}_G]$ are the ones inherited from $\mathcal{T}$ union the new stages determined by G. This means that two vertices $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{i}_{\mathcal{T}}$ are in the same stage in $[\mathcal{T}, \mathcal{T}_G]$ provided $\boldsymbol{a}, \boldsymbol{b} \in G_i$.

We relate $\ker(\varphi_{[\mathcal{T}, \mathcal{T}_G]})$ to the toric fiber product of the two ideals $\ker(\varphi_{\mathcal{T}})$ and the zero ideal $\langle 0 \rangle$. Fix the notation for $\mathcal{T}$, G, $[\mathcal{T}, \mathcal{T}_G]$ as in Definition 4.2. We associate to $\mathcal{T}_G$ the rings

$$\mathbb{R}[p]_{\mathcal{T}_G} := \mathbb{R}[p^i_k \mid k \in \boldsymbol{i}_{\mathcal{B}_i},\ i \in [r]] \quad \text{and} \quad \mathbb{R}[\Theta]_{\mathcal{T}_G} := \mathbb{R}[\epsilon^{(i)}_k \mid i \in [r],\ k \in \boldsymbol{i}_{\mathcal{B}_i}],$$

and the ring map

$$\begin{aligned} \varphi_{\mathcal{T}_G} : \mathbb{R}[p]_{\mathcal{T}_G} &\to \mathbb{R}[\Theta]_{\mathcal{T}_G}, \\ p^i_k &\mapsto \epsilon^{(i)}_k. \end{aligned}$$

Since there is a one-to-one correspondence between the variables p^i_k and $\epsilon^{(i)}_k$, we see that $\varphi_{\mathcal{T}_G}$ is an isomorphism. In particular, $\ker(\varphi_{\mathcal{T}_G}) = \langle 0 \rangle$. Now, using G we regroup the variables in $\mathbb{R}[p]_{\mathcal{T}}$ by

$$\mathbb{R}[p]_{\mathcal{T}} = \mathbb{R}[p^i_{\boldsymbol{j}} \mid \boldsymbol{j} \in G_i,\ i \in [r]].$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

We define multigradings on the polynomial rings $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[p]_{\mathcal{T}_G}$ by

$$\deg(p^i_{\boldsymbol{j}}) = \deg(p^i_k) = \boldsymbol{e}_i \quad \text{for } \boldsymbol{j} \in G_i,\ k \in \boldsymbol{i}_{\mathcal{B}_i},\ i \in [r].$$

Here $\boldsymbol{e}_i$ is the i-th standard unit vector in $\mathbb{Z}^r$. If $\mathcal{A}$ is the set of all these multidegrees, then $\mathcal{A}$ is linearly independent as it is the collection of standard unit vectors in $\mathbb{Z}^r$.

Suppose $\ker(\varphi_{\mathcal{T}})$ is $\mathcal{A}$-homogeneous and fix $R = \mathbb{R}[p]_{\mathcal{T}}/\ker(\varphi_{\mathcal{T}})$, $S = \mathbb{R}[p]_{\mathcal{T}_G}/\ker(\varphi_{\mathcal{T}_G})$. Let $\mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]} = \mathbb{R}[p^i_{\boldsymbol{j}k} \mid \boldsymbol{j} \in G_i,\ k \in \boldsymbol{i}_{\mathcal{B}_i},\ i \in [r]]$ and consider the ring homomorphism

$$\begin{aligned} \psi : \mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]} &\to R \otimes_{\mathbb{R}} S, \\ p^i_{\boldsymbol{j}k} &\mapsto \bar{p}^i_{\boldsymbol{j}} \otimes \bar{p}^i_k \quad \text{for } \boldsymbol{j} \in G_i,\ k \in \boldsymbol{i}_{\mathcal{B}_i},\ i \in [r]. \end{aligned} \tag{3}$$

The ideal $\ker(\psi) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$ is the toric fiber product of $\ker(\varphi_{\mathcal{T}})$ and $\langle 0 \rangle$.

Proposition 4.4. *Let $\mathcal{T}$, G, and $\mathcal{T}_G$ be as in Definition 4.2. Suppose that $\ker(\varphi_{\mathcal{T}})$ is $\mathcal{A}$-homogeneous. Then*

$$\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]}) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle.$$

Proof. Consider the tensor product of maps $\bar{\varphi}_{\mathcal{T}} \otimes \bar{\varphi}_{\mathcal{T}_G} : R \otimes_{\mathbb{R}} S \to \mathbb{R}[\Theta]_{\mathcal{T}} \otimes_{\mathbb{R}} \mathbb{R}[\Theta]_{\mathcal{T}_G}$, where $\bar{\varphi}_{\mathcal{T}} : R \to \mathbb{R}[\Theta]_{\mathcal{T}}$ and $\bar{\varphi}_{\mathcal{T}_G} : S \to \mathbb{R}[\Theta]_{\mathcal{T}_G}$ are induced by $\varphi_{\mathcal{T}}$ and $\varphi_{\mathcal{T}_G}$ on the quotient rings R and S respectively. Note that there is a canonical isomorphism $\mathbb{R}[\Theta]_{\mathcal{T}} \otimes_{\mathbb{R}} \mathbb{R}[\Theta]_{\mathcal{T}_G} \cong \mathbb{R}[\Theta]_{[\mathcal{T},\mathcal{T}_G]}$. Under this isomorphism,

$$\bar{\varphi}_{\mathcal{T}} \otimes \bar{\varphi}_{\mathcal{T}_G}(\bar{p}^i_{\boldsymbol{j}} \otimes \bar{p}^i_k) = \varphi_{\mathcal{T}}(p^i_{\boldsymbol{j}}) \cdot \varphi_{\mathcal{T}_G}(p^i_k) = \left(z \cdot \prod_{e \in E(\lambda_{\boldsymbol{j}})} \theta(e)\right) \epsilon^{(i)}_k = \varphi_{[\mathcal{T},\mathcal{T}_G]}(p^i_{\boldsymbol{j}k}).$$

The last equality follows by the construction of $[\mathcal{T},\mathcal{T}_G]$. Hence $\varphi_{[\mathcal{T},\mathcal{T}_G]} = (\bar{\varphi}_{\mathcal{T}} \otimes \bar{\varphi}_{\mathcal{T}_G}) \circ \psi$. Since $\bar{\varphi}_{\mathcal{T}} \otimes \bar{\varphi}_{\mathcal{T}_G}$ is injective, $\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]}) = \ker(\psi)$. By (3), and since $\ker(\varphi_{\mathcal{T}})$ is $\mathcal{A}$-homogeneous, we conclude $\ker(\psi) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$. Thus $\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]}) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$. ☐

We make several remarks on the scope of Proposition 4.4 via the next set of examples.

Definition 4.5. Let $(\mathcal{T},\theta)$ be a stratified staged tree of level m. For $1 \le q \le m$ we define $V_{\le q} := \bigcup_{i=0}^{q} V_i$, where $V_q := \{v \in V \mid \ell(v) = q\}$. The tree $\mathcal{T}^{(q)} = (V_{\le q}, E_{\le q})$ is the induced subtree of $\mathcal{T}$ on the vertex set $V_{\le q}$. The restriction $\theta|_{E_{\le q}}$ defines a labeling on $\mathcal{T}^{(q)}$. The staged tree $(\mathcal{T}^{(q)}, \theta|_{E_{\le q}})$ is the *level-q subtree* of $(\mathcal{T},\theta)$.

Example 4.6. Consider the staged tree $\mathcal{T}_1$ in Figure 1 and let $\mathcal{T} = \mathcal{T}_1^{(3)}$ be the level-three subtree of $\mathcal{T}$. The label set of $\mathcal{T}$ is $\{s_0,\ldots,s_9\}$. Fix

$$\begin{aligned} G &= \{\{000, 010, 100, 110\}, \{001, 011, 101, 111\}\}, \\ \mathcal{T}_G &= (\mathcal{B}_1, \{s_{10}, s_{11}\}) \sqcup (\mathcal{B}_2, \{s_{12}, s_{13}\}). \end{aligned}$$

With this choice of $\mathcal{T}$, G and $\mathcal{T}_G$ we see that $\mathcal{T}_1 = [\mathcal{T},\mathcal{T}_G]$. Now $\mathbb{R}[p]_{\mathcal{T}} = \mathbb{R}[p^i_{\boldsymbol{a}} \mid \boldsymbol{a} \in G_i,\ i \in \{1,2\}]$; hence $\deg(p^1_{000}, p^1_{010}, p^1_{100}, p^1_{110}) = \boldsymbol{e}_1$ and $\deg(p^2_{001}, p^2_{011}, p^2_{101}, p^2_{111}) = \boldsymbol{e}_2$ so $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} \subset \mathbb{Z}^2$ is of full rank. The ideal

$$\ker(\varphi_{\mathcal{T}}) = \langle p^1_{000}p^2_{101} - p^1_{100}p^2_{011},\ p^1_{010}p^2_{111} - p^1_{110}p^2_{011} \rangle$$

is $\mathcal{A}$-homogeneous. Hence by Proposition 4.4 $\ker(\varphi_{\mathcal{T}_1}) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$.

Example 4.7. Let $\mathcal{T}_2$ be the staged tree from Figure 1. We proceed in a similar fashion as in Example 4.6. Set $\mathcal{T} = \mathcal{T}_2^{(2)}$,

$$G = \{\{00, 01, 10\}, \{11, 31\}, \{20, 21, 30\}\},$$
$$\mathcal{T}_G = (\mathcal{B}_1, \{s_8, s_9\}) \sqcup (\mathcal{B}_2, \{s_{12}, s_{13}\}) \sqcup (\mathcal{B}_3, \{s_{10}, s_{11}\}).$$

Then $\mathcal{T}_2 = [\mathcal{T}, \mathcal{T}_G]$. The set G defines a multigrading on $\mathbb{R}[p]_{\mathcal{T}}$ with $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\} \subset \mathbb{Z}^3$. The ideal

$$\ker(\varphi_{\mathcal{T}}) = \langle p^1_{00}p^2_{11} - p^1_{10}p^1_{01}, p^3_{20}p^2_{31} - p^3_{30}p^3_{21}\rangle$$

is not $\mathcal{A}$-homogeneous. Thus in this case $\ker(\varphi_{\mathcal{T}_2}) \neq \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0\rangle$.

Example 4.8. Let $\mathcal{T}_3$ be as in Figure 1 and $\mathcal{T} = \mathcal{T}_3^{(2)}$. Fix

$$G = \{\{00, 01, 10, 31\}, \{20, 21, 30, 11\}\},$$
$$\mathcal{T}_G = (\mathcal{B}_1, \{s_8, s_9\}) \sqcup (\mathcal{B}_2, \{s_{10}, s_{11}\})$$

so $\mathcal{T}_3 = [\mathcal{T}, \mathcal{T}_G]$. The set G defines a multigrading on $\mathbb{R}[p]_{\mathcal{T}}$ with $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} \subset \mathbb{Z}^2$. The ideal

$$\ker(\varphi_{\mathcal{T}}) = \langle p^1_{00}p^2_{11} - p^1_{10}p^1_{01}, p^2_{20}p^1_{31} - p^2_{30}p^2_{21}\rangle$$

is not $\mathcal{A}$-homogeneous. However there is a nonempty principal subideal of $\ker(\varphi_{\mathcal{T}})$ that is $\mathcal{A}$-homogeneous. This principal ideal Q is generated by the quartic $p^1_{00}p^2_{11}p^2_{20}p^1_{31} - p^1_{01}p^1_{10}p^2_{21}p^2_{30}$. In this case $\ker(\varphi_{\mathcal{T}_3}) = Q \times_{\mathcal{A}} \langle 0\rangle$. This does not fall in the context of Proposition 4.4 since $\ker(\varphi_{\mathcal{T}_3}) \neq \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0\rangle$.

Start with a level-one probability tree $\mathcal{T}_1$. Let G^1 be a partition of $\boldsymbol{i}_{\mathcal{T}_1}$, $\mathcal{T}_{G^1}$ be a gluing component and set $\mathcal{T}_2 = [\mathcal{T}_1, \mathcal{T}_{G^1}]$. In the inductive step, $\mathcal{T}_{j+1} = [\mathcal{T}_j, \mathcal{T}_{G^j}]$, with $r_j := |G^j|$. At each step j we also require that the label set of $\mathcal{T}_{G^j}$ is disjoint from the label set of $\mathcal{T}_j$. Note that G^j is a partition of the set $\boldsymbol{i}_{\mathcal{T}_j}$. After n iterations, we obtain a stratified staged tree $\mathcal{T}_n$ whose set of stages is exactly $\bigcup_{j=1}^{n-1} G^j$. Whenever a staged tree $(\mathcal{T}, \theta)$ is constructed in this way, so $\mathcal{T} = \mathcal{T}_n$ for some n, we say $\mathcal{T}$ is an *inductively constructed* staged tree.

Any stratified staged tree can be inductively constructed. This follows because the set of edge labels associated to the vertices in different levels are disjoint and because all leaves have the same level. Staged trees such as the one in Figure 3 (left) where the leaves have different levels do not fall into the definition of inductively constructed staged tree, even though the edge labels associated to vertices in two different levels are disjoint.

We will use Theorem 3.3 in Section 3 to write down the generators of inductively constructed staged trees that are balanced. The balanced condition is the combinatorial ingredient that makes the algebra of the toric fiber product work when considering iteratively constructed staged trees. This is already evident by looking at the trees in Figure 1 together with the Examples 4.6, 4.7, and 4.8. Although, all of the trees are stratified and can be inductively constructed, only $\ker(\varphi_{\mathcal{T}_1})$ can be constructed in steps by using toric fiber products. This is because $\mathcal{T}_1$ is balanced.

Let $\mathcal{T}_j$ be an inductively constructed staged tree and $\mathcal{T}_{j+1} = [\mathcal{T}_j, \mathcal{T}_{G^j}]$, with $r_j = |G^j|$. Consider the monomial map

$$\begin{aligned}\phi_{B_j} : \mathbb{R}[p]_{\mathcal{T}_{j+1}} &\to \mathbb{R}[p^i_{\boldsymbol{a}}, p^i_k \mid \boldsymbol{a} \in G^j_i,\ k \in \boldsymbol{i}_{\mathcal{B}^j_i},\ i \in [r_j]],\\ p^i_{\boldsymbol{a}k} &\mapsto p^i_{\boldsymbol{a}}p^i_k,\end{aligned} \tag{4}$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

where B_j denotes the exponent matrix of the monomial map ϕ_{B_j}. Set $I_{B_j} = \ker(\phi_{B_j})$ and

$$\mathrm{Quad}_{B_j} = \{\underline{p^i_{\boldsymbol{a}k_1} p^i_{\boldsymbol{b}k_2}} - p^i_{\boldsymbol{b}k_1} p^i_{\boldsymbol{a}k_2} \mid \boldsymbol{a}, \boldsymbol{b} \in G^j_i,\ k_1 \neq k_2 \in \boldsymbol{i}_{\mathcal{B}^j_i},\ i \in [r_j]\}.$$

By Proposition 10 in [17], Quad_{B_j} is a Gröbner basis for I_{B_j} with respect to any term order that selects the underlined terms as leading terms.

We now write down the of the elements in $\mathrm{Lift}(F)$ for $F \subset \mathbb{R}[p]_{\mathcal{T}}$. Fix $\mathcal{T}$, G, and $\mathcal{T}_G$ as in Definition 4.2 and denote by $\mathcal{A}$ the multigrading of the rings $\mathbb{R}[p]_{\mathcal{T}}$, $\mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ defined by G. Following Definition 3.2 for lifts of elements in the ring $\mathbb{R}[p]_{\mathcal{T}}$, we consider the $\mathcal{A}$-homogeneous polynomial

$$f = p^{i_1}_{\boldsymbol{a}_1} p^{i_2}_{\boldsymbol{a}_2} - p^{i_1}_{\boldsymbol{a}_3} p^{i_2}_{\boldsymbol{a}_4} \in \mathbb{R}[p]_{\mathcal{T}},$$

where $\boldsymbol{a}_1, \boldsymbol{a}_3 \in G_{i_1},\ \boldsymbol{a}_2, \boldsymbol{a}_4 \in G_{i_2}$ and $i_1, i_2 \in [r]$. Set $k = (k_1, k_2)$ with $k_1 \in \boldsymbol{i}_{\mathcal{B}_{i_1}},\ k_2 \in \boldsymbol{i}_{\mathcal{B}_{i_2}}$ and consider $f_k \in \mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ defined by

$$f_k = p^{i_1}_{\boldsymbol{a}_1 k_1} p^{i_2}_{\boldsymbol{a}_2 k_2} - p^{i_1}_{\boldsymbol{a}_3 k_1} p^{i_2}_{\boldsymbol{a}_4 k_2}.$$

For each $f \in F$, the set $T_f = \boldsymbol{i}_{\mathcal{B}_{i_1}} \times \boldsymbol{i}_{\mathcal{B}_{i_2}}$ is the set of associated indices and

$$\mathrm{Lift}(F) = \{f_k \mid f \in F, k \in T_f\}.$$

Definition 4.9. Let $\mathcal{T}_n$ be an inductively constructed staged tree with $\mathcal{T}_{i+1} = [\mathcal{T}_i, \mathcal{T}_{G^i}]$ for $1 \leq i \leq n-1$ and let $\mathcal{A}_i$ be the grading in $\mathbb{R}[p]_{\mathcal{T}_i}$ determined by G^i. Fix two nonnegative integers i, q with $1 \leq i+q \leq n-1$. We define

$$\mathrm{Lift}^q(\mathrm{Quad}_{B_i}) := \mathrm{Lift}_{\mathcal{A}_{i+q}}(\cdots(\mathrm{Lift}_{\mathcal{A}_{i+2}}(\mathrm{Lift}_{\mathcal{A}_{i+1}}(\mathrm{Quad}_{B_i})))\cdots),$$

where the subscript in $\mathrm{Lift}_{\mathcal{A}}(\cdot)$ indicates the grading of the argument is with respect to $\mathcal{A}$.

We formulate a lemma that says that level-q subtrees of balanced and stratified staged trees are also balanced. This leads us to consider interpolating polynomials of a vertex in two different rings. For a staged tree $(\mathcal{T} = (V, E), \theta)$ and a vertex $v \in V$ of level q, we write $t_{(j)}(v)$ for the interpolating polynomial of v in the level-j subtree $(\mathcal{T}^{(j)}, \theta|_{E_{\leq j}})$, where $q \leq j \leq m$ and m is the level of $\mathcal{T}$. Thus $t_{(j)}(v)$ is an element of $\mathbb{R}[\Theta]_{\mathcal{T}^{(j)}}$.

Lemma 4.10. *Let $(\mathcal{T}, \theta)$ be a staged tree of level m and let q be a positive integer with $1 \leq q \leq m-1$. If $(\mathcal{T}, \theta)$ is balanced and stratified, then the level-q subtree $\mathcal{T}^{(q)}$ of $\mathcal{T}$ is also balanced and stratified.*

Proof. We must prove that if $\boldsymbol{a}, \boldsymbol{b}$ are two vertices in $\mathcal{T}^{(q)}$ that are in the same stage, then they are a balanced pair in $\mathbb{R}[\Theta]_{\mathcal{T}^{(q)}}$. Since $\mathcal{T}$ is balanced, $\boldsymbol{a}, \boldsymbol{b}$ are a balanced pair in $\mathbb{R}[\Theta]_{\mathcal{T}}$. Namely,

$$t_{(m)}(\boldsymbol{a}k_1) t_{(m)}(\boldsymbol{b}k_2) = t_{(m)}(\boldsymbol{a}k_2) t_{(m)}(\boldsymbol{b}k_1) \quad \text{in } \mathbb{R}[\Theta]_{\mathcal{T}} \text{ for all } k_1, k_2 \in \{0, \ldots, |\mathrm{ch}(\boldsymbol{a})| - 1\}. \tag{5}$$

For a vertex v in $\mathcal{T}^{(q)}$ define $[v] = \{\boldsymbol{u} \in \boldsymbol{i}_{\mathcal{T}^{(q)}} \mid$ the root-to-$\boldsymbol{u}$ path in $\mathcal{T}^{(q)}$ goes through $v\}$. Then by a repeated use of Lemma 2.7, for $\boldsymbol{c} \in \{\boldsymbol{a}k_1, \boldsymbol{b}k_2, \boldsymbol{a}k_2, \boldsymbol{b}k_1\}$,

$$t_{(m)}(\boldsymbol{c}) = \sum_{\boldsymbol{u} \in [\boldsymbol{c}]} \prod_{e \in E(\lambda_{\boldsymbol{u}})} \theta(e) t_{(m)}(\boldsymbol{u}),$$

where $\lambda_{\boldsymbol{u}}$ is the $\boldsymbol{c}$ to $\boldsymbol{u}$ path in $\mathcal{T}^{(q)}$. Here $t_{(m)}(\boldsymbol{c})$ is an element of $\mathbb{R}[\Theta]_{\mathcal{T}}$. Denote by $t_m(\boldsymbol{c})|_{\mathcal{T}^{(q)}}$ the polynomial obtained from $t_{(m)}(\boldsymbol{c})$ by specializing $t_{(m)}(\boldsymbol{u}) = 1$ for all $\boldsymbol{u} \in [\boldsymbol{c}]$. This specialization is a polynomial in $\mathbb{R}[\Theta]_{\mathcal{T}^{(q)}}$. Since $\mathcal{T}$ is stratified, $t(\boldsymbol{c})|_{\mathcal{T}^{(q)}}$ is the interpolating polynomial $t_{(q)}(\boldsymbol{c})$ of $\boldsymbol{c}$ as a vertex in $\mathcal{T}^{(q)}$. Applying this specialization to (5) yields the balanced condition for the pair $\boldsymbol{a}, \boldsymbol{b}$ in $\mathbb{R}[\Theta]_{\mathcal{T}^{(q)}}$. □

Proposition 4.11. *Let $\mathcal{T}_i$ be a balanced and inductively constructed staged tree. Suppose $\mathcal{T}_{i+1} = [\mathcal{T}_i, \mathcal{T}_{G^i}]$ and $\mathcal{T}_{i+1}$ is balanced. Then the elements in*

$$\mathrm{Lift}^{i-2}(\mathrm{Quad}_{B_1}), \quad \mathrm{Lift}^{i-3}(\mathrm{Quad}_{B_2}), \quad \ldots, \quad \mathrm{Lift}(\mathrm{Quad}_{B_{i-2}}), \quad \mathrm{Quad}_{B_{i-1}}$$

are $\mathcal{A}_i$-homogeneous.

Proof. Since $\mathcal{T}_i$ is inductively constructed, there is a sequence of stratified trees and gluing components $(\mathcal{T}_1, \mathcal{T}_{G^1}), \ldots, (\mathcal{T}_{i-1}, \mathcal{T}_{G^{i-1}})$ from which $\mathcal{T}_i$ is constructed. Moreover, by Lemma 4.10 each of $\mathcal{T}_1, \ldots, \mathcal{T}_{i-1}$ is also balanced. Fix $q \in \{0, 1, \ldots, i-2\}$ and $j = i - q - 1$, we show that the binomials in $\mathrm{Lift}^q(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_i$-homogeneous. To this end we prove that for m such that $0 \le m \le q$, the elements in $\mathrm{Lift}^m(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_{j+m+1}$-homogeneous. The proof is by induction on m.

Fix $m = 0$. We will show that the elements in Quad_{B_j} are $\mathcal{A}_{j+1}$-homogeneous. The multidegrees in $\mathcal{A}_{j+1}$ are defined according to the partition G^{j+1} of the leaves of $\mathcal{T}_{j+1}$. If two leaves $\boldsymbol{c}, \boldsymbol{d} \in \boldsymbol{i}_{\mathcal{T}_{j+1}}$ in $\mathcal{T}_{j+1}$ are in the same set G^{j+1}_β of the partition G^{j+1}, then $\deg(p_{\boldsymbol{c}}) = \deg(p_{\boldsymbol{d}})$ in $\mathbb{R}[p]_{\mathcal{T}_{j+1}}$.

Since $\mathcal{T}_{j+2}$ is balanced, every pair of vertices in $\mathcal{T}_{j+2}$ in the same stage is balanced. In particular, this means that for all $\alpha \in \{1, \ldots, r_j\}$ and $\boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha$

$$t_{(j+2)}(\boldsymbol{a}k_1)t_{(j+2)}(\boldsymbol{b}k_2) = t_{(j+2)}(\boldsymbol{b}k_1)t_{(j+2)}(\boldsymbol{a}k_2) \quad \text{for } k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha} \text{ in } \mathbb{R}[\Theta]_{\mathcal{T}_{j+2}}, \tag{6}$$

where $\mathrm{ch}(\boldsymbol{a}) = \{\boldsymbol{a}k \mid k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$ and $\mathrm{ch}(\boldsymbol{b}) = \{\boldsymbol{b}k \mid k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$. Using the construction of $\mathcal{T}_{j+2}$ from $\mathcal{T}_{j+1}$ and $\mathcal{T}_{G^{j+1}}$, we know that for any index $\boldsymbol{c} \in \{\boldsymbol{a}k, \boldsymbol{b}k \mid k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$, we have $t_{(j+2)}(\boldsymbol{c}) = \epsilon_0^{(j+1)} + \cdots + \epsilon_{k'}^{(j+1)}$, where $\{\epsilon_0^{(j+1)}, \ldots, \epsilon_{k'}^{(j+1)}\}$ is the set of labels of some level-one probability tree $\mathcal{B}^{j+1}_\delta$ in $\mathcal{T}_{G^{j+1}}$. It follows that (6) can only involve at most two sets of variables associated to two level-one probability trees in $\mathcal{T}_{G^{j+1}}$, say $\mathcal{B}^{j+1}_\beta, \mathcal{B}^{j+1}_\gamma$. This implies that either $\{\boldsymbol{a}k_1, \boldsymbol{b}k_1\} \subset G^{j+1}_\beta$ and $\{\boldsymbol{a}k_2, \boldsymbol{b}k_1\} \subset G^{j+1}_\gamma$ or $\{\boldsymbol{a}k_1, \boldsymbol{a}k_2\} \subset G^{j+1}_\beta$ and $\{\boldsymbol{b}k_2, \boldsymbol{b}k_1\} \subset G^{j+1}_\gamma$. We use this fact to determine the multigrading of the elements in Quad_{B_j} with respect to $\mathcal{A}_{j+1}$. By definition,

$$\mathrm{Quad}_{B_j} = \bigcup_{\alpha=1}^{r_j} \{p_{\boldsymbol{a}k_1} p_{\boldsymbol{b}k_2} - p_{\boldsymbol{b}k_1} p_{\boldsymbol{a}k_2} \mid \boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha,\ k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}.$$

Thus we calculate that the element $p_{\boldsymbol{a}k_1} p_{\boldsymbol{b}k_2} - p_{\boldsymbol{b}k_1} p_{\boldsymbol{a}k_2}$ in Quad_{B_j} is $\mathcal{A}_{j+1}$-homogeneous of degree $\boldsymbol{e}_\beta + \boldsymbol{e}_\gamma$, where $\boldsymbol{e}_\beta$ and $\boldsymbol{e}_\gamma$ are the multidegrees in $\mathcal{A}_{j+1}$ associated to the sets G^{j+1}_β and G^{j+1}_γ, respectively. This completes the proof for $m = 0$. As a result, all the equations in Quad_{B_j} can be lifted to elements in $\ker(\varphi_{\mathcal{T}_{j+2}})$.

Suppose we have constructed $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ inductively by lifting the equations in Quad_{B_j} and at each step all equations lift. An element in $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ is a binomial of the form

$$f = p_{\boldsymbol{a}k_1 s} p_{\boldsymbol{b}k_2 u'} - p_{\boldsymbol{b}k_1 u} p_{\boldsymbol{a}k_2 s'}, \tag{7}$$

where $\alpha \in \{1, \ldots, r_j\}$, $\boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha$, $k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}$ and s, s', u, u' are sequences of nonnegative integers of length $m-1$ that arise as subindices after lifting $m-1$ times. Note that $\boldsymbol{a}k_1 s, \boldsymbol{b}k_2 u', \boldsymbol{b}k_1 u, \boldsymbol{a}k_2 s' \in \boldsymbol{i}_{\mathcal{T}_{j+m}}$The claim is that (7) is $\mathcal{A}_{j+m}$- homogeneous.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Following a similar argument as for $m=0$, we know that two elements in the same set of the partition G^{j+m} have the same multidegree with respect to $\mathcal{A}_{j+m}$. As before, this condition can be verified for f by checking that

$$t_{(j+m+1)}(\boldsymbol{a}k_1s)t_{(j+m+1)}(\boldsymbol{b}k_2u') = t_{(j+m+1)}(\boldsymbol{b}k_1u)t_{(j+m+1)}(\boldsymbol{a}k_2s') \quad \text{in } \mathbb{R}[\Theta]_{\mathcal{T}_{j+m+1}}. \tag{8}$$

For $\boldsymbol{c} \in \{\boldsymbol{a}k_1, \boldsymbol{b}k_2, \boldsymbol{a}k_2, \boldsymbol{b}k_1\}$, we have $[\boldsymbol{c}] := \{w \in \boldsymbol{i}_{\mathcal{T}_{j+m}} \mid \text{the root-to-}w \text{ path in } \mathcal{T}_{j+m} \text{ goes through } \boldsymbol{c}\}$. To check that (8) holds, consider (2) from Definition 2.9 for the vertices $\boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha$. This equation is

$$t_{(j+m+1)}(\boldsymbol{a}k_1)t_{(j+m+1)}(\boldsymbol{b}k_2) = t_{(j+m+1)}(\boldsymbol{b}k_1)t_{(j+m+1)}(\boldsymbol{a}k_2),$$

where $k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}$. We use Lemma 2.7 to rewrite this equation as

$$\begin{aligned}&\Bigg(\sum_{\boldsymbol{a}k_1s\in[\boldsymbol{a}k_1]}\Bigg(\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1s)}\theta(e)\Bigg)t_{j+m+1}(\boldsymbol{a}k_1s)\Bigg)\cdot\Bigg(\sum_{\boldsymbol{b}k_2u'\in[\boldsymbol{b}k_2]}\Bigg(\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2u')}\theta(e)\Bigg)t_{j+m+1}(\boldsymbol{b}k_2u')\Bigg)\\ &=\Bigg(\sum_{\boldsymbol{b}k_1u\in[\boldsymbol{b}k_1]}\Bigg(\prod_{e\in E(\boldsymbol{b}k_1\to\boldsymbol{b}k_1u)}\theta(e)\Bigg)t_{j+m+1}(\boldsymbol{b}k_1u)\Bigg)\cdot\Bigg(\sum_{\boldsymbol{a}k_2s'\in[\boldsymbol{a}k_2]}\Bigg(\prod_{e\in E(\boldsymbol{a}k_2\to\boldsymbol{a}k_2s')}\theta(e)\Bigg)t_{j+m+1}(\boldsymbol{a}k_2s')\Bigg).\end{aligned} \tag{9}$$

When we specialize $t_{(j+m+1)}(\boldsymbol{a}k_1s) = t_{(j+m+1)}(\boldsymbol{b}k_2u') = t_{(j+m+1)}(\boldsymbol{b}k_1u) = t_{(j+m+1)}(\boldsymbol{a}k_2s') = 1$ in each sum in (9) we get the interpolating polynomials $t_{(j+m)}(\boldsymbol{a}k_1)$, $t_{(j+m)}(\boldsymbol{b}k_2)$, $t_{(j+m)}(\boldsymbol{b}k_1)$, $t_{(j+m)}(\boldsymbol{a}k_2)$ in $\mathbb{R}[\Theta]_{\mathcal{T}_{j+m}}$. By Lemma 4.10, $\mathcal{T}_{j+m}$ is balanced; therefore

$$\begin{aligned}&\Bigg(\sum_{\boldsymbol{a}k_1s\in[\boldsymbol{a}k_1]}\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1s)}\theta(e)\Bigg)\cdot\Bigg(\sum_{\boldsymbol{b}k_2u'\in[\boldsymbol{b}k_2]}\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2u')}\theta(e)\Bigg)\\ &\qquad=\Bigg(\sum_{\boldsymbol{b}k_1u\in[\boldsymbol{b}k_1]}\prod_{e\in E(\boldsymbol{b}k_1\to\boldsymbol{b}k_1u)}\theta(e)\Bigg)\cdot\Bigg(\sum_{\boldsymbol{a}k_2s'\in[\boldsymbol{a}k_2]}\prod_{e\in E(\boldsymbol{a}k_2\to\boldsymbol{a}k_2s')}\theta(e)\Bigg).\end{aligned} \tag{10}$$

The factors in the above equality are sums of monomials all with coefficients equal to 1. Thus for every pair $\boldsymbol{a}k_1s \in [\boldsymbol{a}k_1]$, $\boldsymbol{b}k_2u' \in [\boldsymbol{b}k_2]$ in the product of the left-hand side of the equation, there exists a pair $\boldsymbol{a}k_2s' \in [\boldsymbol{a}k_2]$, $\boldsymbol{b}k_2u \in [\boldsymbol{b}k_1]$ in the product of the right-hand side of the equation such that

$$\Bigg(\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1s)}\theta(e)\Bigg)\cdot\Bigg(\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2u')}\theta(e)\Bigg)=\Bigg(\prod_{e\in E(\boldsymbol{b}k_1\to\boldsymbol{b}k_1u)}\theta(e)\Bigg)\cdot\Bigg(\prod_{e\in E(\boldsymbol{a}k_2\to\boldsymbol{a}k_2s')}\theta(e)\Bigg). \tag{11}$$

Hence (5) for the vertices $\boldsymbol{a}, \boldsymbol{b}$ in $\mathcal{T}_{j+m+1}$ can be rewritten as

$$\sum_{\substack{\boldsymbol{a}k_1s\in[\boldsymbol{a}k_1]\\ \boldsymbol{b}k_2u'\in[\boldsymbol{b}k_2]}}\Bigg(\prod_{\substack{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1s)\\ e'\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2u')}}\theta(e)\theta(e')\Bigg)(t_{(j+m+1)}(\boldsymbol{a}k_1s)t_{(j+m+1)}(\boldsymbol{b}k_2u')-t_{(j+m+1)}(\boldsymbol{b}k_1u)t_{(j+m+1)}(\boldsymbol{a}k_2s'))=0.$$

Since $\mathcal{T}_{j+m+1}$ is stratified, the variables that appear in the factored monomials above are different from the variables that appear in the factors of the form

$$t_{(j+m+1)}(\boldsymbol{a}k_1s)t_{(j+m+1)}(\boldsymbol{b}k_2u') - t_{(j+m+1)}(\boldsymbol{b}k_1u)t_{(j+m+1)}(\boldsymbol{a}k_2s').$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Hence this last equation is true only if

$$t_{(j+m+1)}(\boldsymbol{a}k_1s)t_{(j+m+1)}(\boldsymbol{b}k_2u') - t_{(j+m+1)}(\boldsymbol{b}k_1u)t_{(j+m+1)}(\boldsymbol{a}k_2s') = 0$$

for each summand. Following a similar argument as in the case for $m = 0$, this proves that the elements in $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_{j+m}$-homogeneous. □

We are now ready to prove our main result, Theorem 2.14, using toric fiber products for balanced and inductively constructed staged trees.

Proof of Theorem 2.14. If $\mathcal{T}$ is stratified, then $\mathcal{T}$ is an iteratively constructed staged tree and $\mathcal{T} = \mathcal{T}_n$ for some n. Set $F_n = \mathrm{Lift}^{n-2}(\mathrm{Quad}_{B_1}) \cup \mathrm{Lift}^{n-3}(\mathrm{Quad}_{B_2}) \cup \cdots \cup \mathrm{Quad}_{B_{n-1}}$. We prove by induction on n that $\ker(\varphi_{\mathcal{T}_n})$ is generated by F_n and that F_n is a Gröbner basis with square-free initial ideal. The first nontrivial case is $n = 2$. We have $F_2 = \mathrm{Quad}_{B_1}$ and from Proposition 10 in [17], F_2 is a Gröbner basis for the ideal $\ker(\varphi_{\mathcal{T}_2}) = \ker(\varphi_{\mathcal{T}_1}) \times_{\mathcal{A}_1} \langle 0 \rangle$. Suppose the statement is true for i, so the elements in F_i are a Gröbner basis for $\ker(\varphi_{\mathcal{T}_i})$. Since $\mathcal{T}_n$ is balanced, by Lemma 4.10 the trees $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ are also balanced. From Proposition 4.11 the elements in F_i are $\mathcal{A}_i$ homogeneous, so by Theorem 3.3 the set F_{i+1} is a Gröbner basis for $\ker(\varphi_{\mathcal{T}_{i+1}})$. Since the elements in F_n are all extensions of elements in Quad_{B_j} for j with $1 \le j \le n-1$ we see that all the terms in these binomials are square-free. Hence the initial ideal of $\langle F_n \rangle$ is square-free. □

Corollary 4.12. *Let $(\mathcal{T}, \theta)$ be a balanced and stratified staged tree. Fix Δ to be the polytope defined by the convex hull of the lattice points in the exponent matrix of the map $\varphi_{\mathcal{T}}$. Then Δ has a regular unimodular triangulation. In particular the toric variety defined by* $\ker(\varphi_{\mathcal{T}})$ *is Cohen–Macaulay.*

Proof. The ideal $\ker(\varphi_{\mathcal{T}})$ has a square-free quadratic Gröbner basis with respect to a term order $\prec$. From [16, Corollary 8.9], this induces a regular unimodular triangulation of Δ. □

5. Connections to discrete statistical models

Staged tree models are a class of graphical discrete statistical models introduced by Anderson and Smith in [15]. While Bayesian networks and decomposable models are defined via conditional independence statements on random variables corresponding to the vertices of a graph, staged tree models encode independence relations on the events of an outcome space represented by a tree. In the statistical literature these models are also referred to as chain event graphs. We refer the reader to the book [3] and to [19] to find out more about their statistical properties, practical implementation, and causal interpretation. In this section we give a formal definition of staged tree models and recall results from [6; 9] about their defining equations.

Given a discrete random variable X with state space $\{0, \dots, n\}$, a probability distribution on X is a vector $(p_0, \dots, p_n) \in \mathbb{R}^{n+1}$, where $p_i = P(X{=}i)$, $i \in \{0, \dots, n\}$, $p_i \ge 0$ and $\sum_{i=0}^n p_i = 1$. The open probability simplex

$$\Delta_n^\circ = \{(p_0, \dots, p_n) \in \mathbb{R}^{n+1} \mid p_i > 0,\, p_0 + \cdots + p_n = 1\}$$

consists of all the strictly positive probability distributions for a discrete random variable with state space $\{0, \dots, n\}$. A *discrete statistical model* is a subset of Δ_n°. In the next definition we associate a discrete statistical model to a given staged tree.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Definition 5.1. Let $(\mathcal{T}, \theta)$ be a staged tree. We define the parameter space

$$\Theta_{\mathcal{T}} := \left\{x \in \mathbb{R}^{|\mathcal{L}|} \;\middle|\; \text{for all } e \in E,\ x_{\theta(e)} \in (0,1) \text{ and for all } \boldsymbol{a} \in V,\ \textstyle\sum_{e\in E(\boldsymbol{a})} x_{\theta(e)} = 1\right\}.$$

Note that $\Theta_{\mathcal{T}}$ is a product of open probability simplices. A *staged tree model* $\mathcal{M}_{(\mathcal{T},\theta)}$ is the image of the map $\Psi_{\mathcal{T}} : \Theta_{\mathcal{T}} \to \Delta^{\circ}_{|\boldsymbol{i}_{\mathcal{T}}|-1}$ defined by

$$x \mapsto p_x = \left(\prod_{e\in E(\lambda_{\boldsymbol{j}})} x_{\theta(e)}\right)_{\boldsymbol{j}\in \boldsymbol{i}_{\mathcal{T}}}.$$

We can check that, for every $x \in \Theta_{\mathcal{T}},\ p_x$ is a probability distribution and therefore $\Psi(\Theta_{\mathcal{T}}) \subset \Delta^{\circ}_{|\boldsymbol{i}_{\mathcal{T}}|-1}$. Two staged trees $(\mathcal{T}, \theta)$ and $(\mathcal{T}', \theta')$ are *statistically equivalent* if there exists a bijection between $\Lambda_{\mathcal{T}}$ and $\Lambda_{\mathcal{T}'}$ in such a way that the image of $\Psi_{\mathcal{T}}$ is equal to the image of $\Psi_{\mathcal{T}'}$ under this bijection.

Example 5.2. The staged tree $\mathcal{T}_1$ in Figure 1 is the staged tree representation of the decomposable model associated to the undirected graph $G = [12][23][34]$ on four nodes.

Remark 5.3. For staged tree models, the root-to-leaf paths in the tree represent the possible unfoldings of a sequence of events. Given an edge (v, w) in $\mathcal{T}$, the label $\theta(v, w)$ is the transition probability from v to w given arrival at v.

Remark 5.4. A staged tree model $\mathcal{M}_{(\mathcal{T},\theta)}$ is a discrete statistical model parametrized by polynomials. The domain of this model is a semialgebraic set given by a product of simplices. As a consequence the image of $\Psi_{\mathcal{T}}$ is also a semialgebraic set. An important property of these models as noted in [9] is that the only inequality constraints of the image of $\Psi_{\mathcal{T}}$ are the ones imposed by the probability simplex, namely $0 \le p_{\boldsymbol{j}} \le 1$ for $\boldsymbol{j} \in \boldsymbol{i}_{\mathcal{T}}$ and $\sum_{\boldsymbol{j}\in\boldsymbol{i}_{\mathcal{T}}} p_{\boldsymbol{j}} = 1$.

In Definition 2.2 we defined the toric ideal associated to a staged tree $(\mathcal{T}, \theta)$. Now we define the ideal associated to a staged tree model $\mathcal{M}_{(\mathcal{T},\theta)}$. For this we use the rings $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[\Theta]_{\mathcal{T}}$ from Definition 2.2. Consider the ideal $\mathfrak{q}$ of $\mathbb{R}[\Theta]_{\mathcal{T}}$ generated by all sum-to-1 conditions $1 - \sum_{e\in E(\boldsymbol{a})} \theta(e)$ for $\boldsymbol{a} \in V$ and let $\mathbb{R}[\Theta]_{\mathcal{M}_{\mathcal{T}}} := \mathbb{R}[\Theta]_{\mathcal{T}}/\mathfrak{q}$. Denote by π the canonical projection from $\mathbb{R}[\Theta]_{\mathcal{T}}$ to the quotient ring $\mathbb{R}[\Theta]_{\mathcal{M}_{\mathcal{T}}}$.

Definition 5.5. Let $\mathcal{M}_{(\mathcal{T},\theta)}$ be a staged tree model and set $\bar{\varphi}_{\mathcal{T}} := \pi \circ \varphi_{\mathcal{T}}$. The ideal $\ker(\bar{\varphi}_{\mathcal{T}})$ is the *staged tree model ideal* associated to the model $\mathcal{M}_{(\mathcal{T},\theta)}$.

From the definition it follows that for every staged tree $(\mathcal{T}, \theta)$, the toric staged tree ideal is contained in the staged tree model ideal; i.e., $\ker(\varphi_{\mathcal{T}}) \subset \ker(\bar{\varphi}_{\mathcal{T}})$. It is not true in general that these two ideals are equal [6]. However, Theorem 10 in [6] states that if a staged tree $(\mathcal{T}, \theta)$ is balanced, then $\ker(\varphi_{\mathcal{T}}) = \ker(\bar{\varphi}_{\mathcal{T}})$. Combining this result with Theorem 2.14 we can obtain Gröbner bases for staged tree model ideals whose staged tree is balanced and stratified.

Corollary 5.6. *If $(\mathcal{T}, \theta)$ is a balanced and stratified staged tree, then the ideal* $\ker(\bar{\varphi}_{\mathcal{T}})$ *has a quadratic Gröbner basis with square-free initial ideal.*

Example 5.7. Consider the staged tree model defined by the tree $\mathcal{T}_1$ in Figure 1 as in Example 5.2. Since this staged tree model is equal to the decomposable model given by $G = [12][23][34]$, from [8] we know it has a quadratic Gröbner basis. We recover the same result from the perspective of staged trees by using Corollary 5.6.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

Corollary 5.6 is relevant in statistics because of the connection of Gröbner bases to sampling [1]. We presented Example 5.7, where a balanced and stratified staged tree represents an instance of a decomposable graphical model. We now provide more examples of staged tree models for which Corollary 5.6 holds. The first one is an explanation of the contraction axiom for conditional independence statements through the lens of staged trees. Before we present our examples we do a quick overview of discrete conditional independence models. Our exposition follows that in [18, Chapter 4]; for more details we refer the reader to [14; 5].

Let $X = (X_1, \dots, X_n)$ be a vector of discrete random variables, where X_i has state space $[d_i]$ for $i \in [n]$. The vector X has state space $\mathcal{X} = [d_1] \times \cdots \times [d_n]$ and we write $p_{u_1 \cdots u_n}$ for the probability $P(X_1{=}u_1, \dots, X_n{=}u_n)$. We consider only positive probability distributions of the random vector X. For each subset $A \subset [n]$, X_A is the subvector of X indexed by the elements in A. Similarly, $\mathcal{X}_A = \prod_{i \in A}[d_i]$ and for a vector $x \in \mathcal{X}$, x_A denotes the restriction of x to the indexes in A.

Definition 5.8. Let A, B, C be pairwise disjoint subsets of $[n]$. The random vector X_A is conditionally independent of X_B given X_C if for every $a \in X_A$, $b \in X_B$ and $c \in X_C$

$$P(X_A{=}a,\ X_B{=}b|X_C{=}c) = P(X_A{=}a|X_C = c) \cdot P(X_B{=}b|X_C{=}c).$$

The notation $X_A \perp\!\!\!\perp X_B \mid X_C$ is used to denote that the random vector X satisfies the *conditional independence statement* that X_A is conditionally independent on X_B given X_C. When C is the empty set this reduces to marginal independence between X_A and X_B.

If $\mathcal{C}$ is a list of conditional independence statements among variables in a vector X, the *conditional independence model* $\mathcal{M}_{\mathcal{C}}$ is the set of all probability distributions inside the open probability simplex $\Delta^\circ_{|\mathcal{X}|-1}$ that satisfy the conditional independence statements in $\mathcal{C}$. A conditional independence statement $X_A \perp\!\!\!\perp X_B \mid X_C$ translates into the condition that the joint probability distribution of the variables in X satisfies a set of quadratic equations. For elements $a \in X_A$, $b \in X_B$ and $c \in X_C$ we set $p_{a,b,c,+} = P(X_A{=}a,\ X_B{=}b,\ X_C{=}c)$.

Proposition 5.9 [18]. *If X is a discrete random vector, then the independence statement $X_A \perp\!\!\!\perp X_B \mid X_C$ holds for X if and only if the probability distribution of X satisfies*

$$p_{a_1,b_1,c,+}p_{a_2,b_2,c,+} - p_{a_1,b_2,c,+}p_{a_2,b_1,c,+} = 0$$

for all $a_1, a_2 \in \mathcal{X}_A$, $b_1, b_2 \in \mathcal{X}_B$ and $c \in \mathcal{X}_C$.

Let $\mathbb{R}[p_x \mid x \in \mathcal{X}]$ be the polynomial ring with one indeterminate for each element in the state space of X. The conditional independence ideal $I_{A\perp\!\!\!\perp B|C}$, is the ideal in $\mathbb{R}[p_x \mid x \in \mathcal{X}]$ generated by all quadratic relations in Proposition 5.9. If $\mathcal{C}$ is a list of conditional independence statements then we define $I_{\mathcal{C}}$ as the sum of all conditional independence ideals associated to statements in $\mathcal{C}$.

Example 5.10. We consider the contraction axiom for positive distributions using staged tree models. Fix three discrete random variables X_1, X_2, X_3 with state spaces $[d_1+1], [d_2+1], [d_3+1]$ respectively. The contraction axiom states that the set of conditional independence statements $\mathcal{C} = \{X_1 \perp\!\!\!\perp X_2 \mid X_3,\ X_2 \perp\!\!\!\perp X_3\}$ implies the statement $X_2 \perp\!\!\!\perp (X_1, X_3)$. A primary decomposition of the ideal $I_{\mathcal{C}}$ was obtained in [7, Theorem 1]. Here we provide a proof, using staged trees, that one of the primary components of $I_{\mathcal{C}}$ is

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

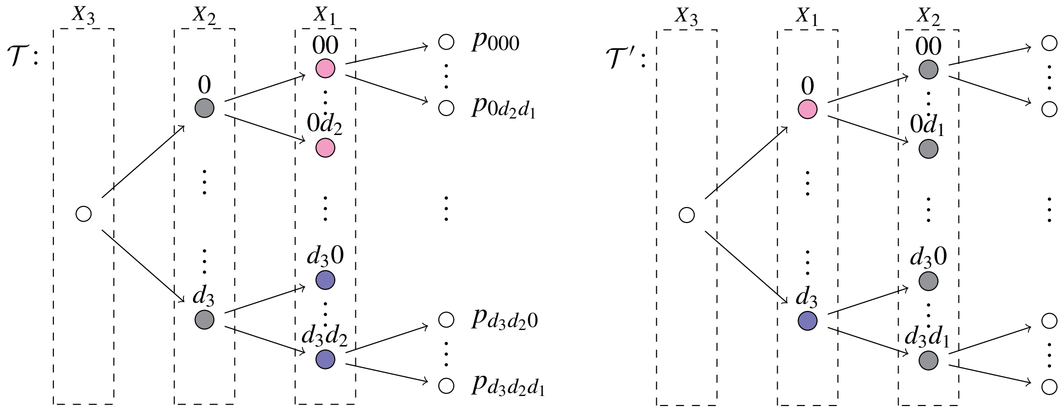

Figure 2. The staged trees $\mathcal{T}$ and $\mathcal{T}'$ are statistically equivalent, they represent the contraction axiom for three discrete random variables X_1, X_2 and X_3.

the prime binomial ideal $I_{X_2 \perp\!\!\!\perp (X_1,X_3)}$. As mentioned in [7] this is a well-known fact. First we explain how to represent the two statements in $\mathcal{C}$ with a staged tree. Consider the tree $\mathcal{T}$ in Figure 2. This tree represents the state space of the vector (X_3, X_2, X_1) as a sequence of events where X_3 takes place first, X_2 second and X_1 third. The vertices of $\mathcal{T}$ are indexed recursively as defined at the beginning of Section 4. The statement $X_2 \perp\!\!\!\perp X_3$ is represented by the stage consisting of the vertices $\{0, \dots, d_3\}$; these are colored gray in $\mathcal{T}$. The statement $X_1 \perp\!\!\!\perp X_2 \mid X_3$ is represented by the stages $S_0, \dots, S_{d_3}$, where $S_i = \{ij \mid j \in \{0, \dots, d_2\}\}$ and $i \in \{0, \dots, d_3\}$. These stages mean that for a given outcome of X_3, the unfolding of the event X_2 followed by X_1 behaves as an independence model on two random variables. In Figure 2 the stage S_0 is colored in pink and the stage S_{d_3} is colored in purple. Although the gray vertices are not in the same position, we can easily check that $\mathcal{T}$ is balanced and stratified. Therefore $\ker(\varphi_{\mathcal{T}})$ has a quadratic Gröbner basis. Following the proof of Theorem 2.14 we can construct this basis explicitly. It consists of a set of quadratic equations given by the elements in Quad_{B_2} coming from the stages in $S_0, \dots, S_{d_3}$ and the lifts of the equations Quad_{B_1} coming from the stage $\{0, \dots, d_3\}$. If we swap the order of X_1 and X_2 in $\mathcal{T}$, we obtain the staged tree $\mathcal{T}'$ in Figure 2. This tree represents the same statistical model as $\mathcal{T}$ now with the unfolding of events X_3, X_1, X_2. The gray stages in $\mathcal{T}'$ represent the statement $X_2 \perp\!\!\!\perp (X_1, X_3)$. Hence, after establishing the evident bijection between the leaves of $\mathcal{T}$ and $\mathcal{T}'$ we see that $I_{X_2 \perp\!\!\!\perp (X_1,X_3)} = \ker(\varphi_{\mathcal{T}'}) = \ker(\varphi_{\mathcal{T}})$.

One of the main differences between staged tree models and discrete Bayesian networks is that the state space of a Bayesian network is equal to the product of the state spaces of the random variables in the vertices of the graph, while the state space of a staged tree model does not necessarily have to equal a cartesian product. When $\mathcal{T}$ is not equal to the cartesian product of some finite sets we call the tree $\mathcal{T}$ *asymmetric*. The lemmas that follow are important to show that Theorem 2.14 also holds for the case when $\mathcal{T}$ is asymmetric. This implies that we can use Theorem 2.14 to construct quadratic Gröbner bases for staged tree models whose underlying tree does not necessarily represents the outcomes of a vector of discrete random variables.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

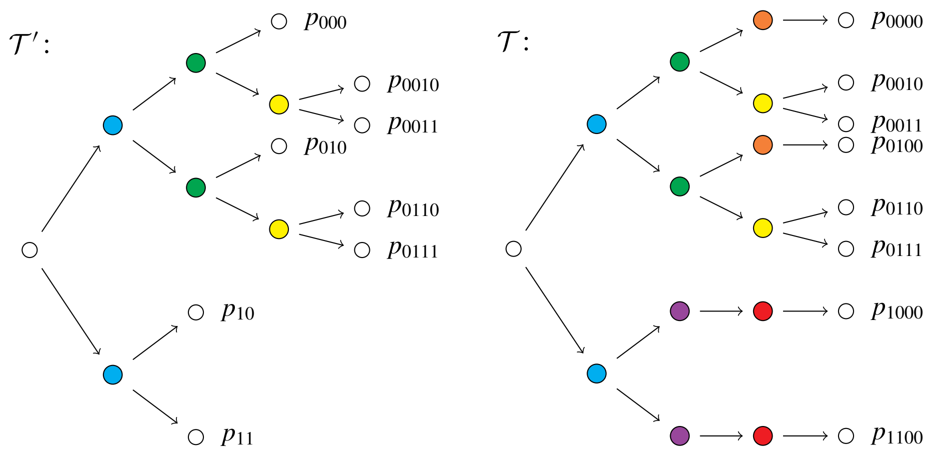

Figure 3. The staged trees $\mathcal{T}$ and $\mathcal{T}'$ are statistically equivalent.

The definition of staged tree in [9] requires that each vertex in $\mathcal{T}$ has either no or at least two outgoing edges from v. We stepped away from making this requirement for the staged trees we consider in Section 2. In the next lemmas we explain how this mild extension of the definition behaves with respect to the balanced condition for a pair of vertices, and how trees defined according to [9] are recovered from the more general trees we consider. Throughout the next lemmas, we fix a staged tree $(\mathcal{T}, \theta)$ with edge set E and define $E_1 = \{e \in E \mid E(v) = \{e\} \text{ for some } v \in V\}$. For the trees in Figure 3, $\mathcal{T}$ has $|E_1| = 6$, while for $\mathcal{T}'$, $|E_1| = 0$.

Lemma 5.11. *Suppose $(\mathcal{T}, \theta)$ is a staged tree. Let $\mathcal{T}'$ be the staged tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $\mathcal{M}_{(\mathcal{T},\theta)} = \mathcal{M}_{(\mathcal{T}',\theta)}$ and* $\ker(\bar{\varphi}_{\mathcal{T}}) = \ker(\bar{\varphi}_{\mathcal{T}'})$.

Proof. First, note that the number of root-to-leaf paths in $\mathcal{T}'$ is the same as in $\mathcal{T}$. Moreover, each root-to-leaf path λ' in $\mathcal{T}'$ is obtained from a unique root-to-leaf path λ in $\mathcal{T}$ by contracting the edges in E_1. Now let λ be a root-to-leaf path in $\mathcal{T}$. The λ-coordinate of the map $\Psi_{\mathcal{T}}$ applied to an element $\theta \in \Theta_{\mathcal{T}}$ is

$$[\Psi_{\mathcal{T}}(\theta)]_\lambda = \prod_{e \in E(\lambda)} \theta(e) = \prod_{e \in E(\lambda')} \theta(e) = [\Psi_{\mathcal{T}'}(\theta \mid_{\mathcal{T}'})]_{\lambda'}.$$

The second equality in the previous equation follows from taking a closer look at $\Theta_{\mathcal{T}}$. Indeed for all $e \in E_1$ we have $\theta(e) = 1$ because of the sum-to-1 conditions imposed on $\Theta_{\mathcal{T}}$ in Definition 5.1. For the third equality, $\theta \mid_{\mathcal{T}'}$ denotes the restriction of the vector θ to the edge labels of $\mathcal{T}'$. It follows from the equalities above that the coordinates of $\Psi_{\mathcal{T}}$ and $\Psi_{\mathcal{T}'}$ are equal. Therefore $\mathcal{M}_{(\mathcal{T},\theta)} = \mathcal{M}_{(\mathcal{T}',\theta)}$. A similar argument applied to the maps $\bar{\varphi}_{\mathcal{T}}$ and $\bar{\varphi}_{\mathcal{T}'}$ shows that $\ker(\bar{\varphi}_{\mathcal{T}}) = \ker(\bar{\varphi}_{\mathcal{T}'})$. To carry out this argument we need to reindex the leaves of the trees; this can be done by dropping the index of the elements in E_1. □

We illustrate Lemma 5.11 in Figure 3 where $\mathcal{T}'$ is obtained from $\mathcal{T}$ by contracting the six edges in E_1. The two staged trees in this figure define the same statistical model.

Remark 5.12. To prove Corollary 5.6 we used [6, Theorem 10]. The proof of Theorem 10 in [6] is presented for trees such that $E_1 = 0$. However the result still holds when $|E_1| > 1$ because the ideal I_{Paths} (from [6]) is contained in $\ker(\varphi_{\mathcal{T}})$ in this case also; see [6] for more details.

Lemma 5.13. *Suppose $(\mathcal{T}, \theta)$ is a balanced and stratified staged tree. Let $\mathcal{T}'$ be the tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $(\mathcal{T}', \theta)$ is also balanced.*

Proof. Suppose $\mathcal{T}$ is balanced and $\boldsymbol{a}, \boldsymbol{b}$ are in the same stage. Following the notation from Definition 2.9, we have $t(\boldsymbol{a}i)t(\boldsymbol{b}j) = t(\boldsymbol{b}j)t(\boldsymbol{a}j)$ in $\mathbb{R}[\Theta]_{\mathcal{T}}$ for all $i \neq j \in \{0, 1, \dots, k\}$. We specialize $\theta(e) = 1$ in this equation for all $e \in E_1$ to obtain $t(\boldsymbol{a}i)t(\boldsymbol{b}j)\mid_{\theta(e)=1, e\in E_1} = t(\boldsymbol{b}j)t(\boldsymbol{a}i)\mid_{\theta(e)=1, e\in E_1}$ in $\mathbb{R}[\Theta]_{\mathcal{T}'}$. Therefore $\mathcal{T}'$ is also balanced. □

Corollary 5.14. *Suppose $\mathcal{T}$ is a balanced and stratified staged tree. Let $\mathcal{T}'$ be the staged tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $\ker(\bar{\varphi}_{\mathcal{T}'})$ is a toric ideal with a quadratic Gröbner basis whose initial ideal is square-free.*

Proof. From Corollary 5.6 it follows that $\ker(\bar{\varphi}_{\mathcal{T}})$ is a toric ideal with a quadratic Gröbner basis and square-free initial ideal. After an appropriate bijection, by Lemma 5.11, $\ker(\bar{\varphi}_{\mathcal{T}}) = \ker(\bar{\varphi}_{\mathcal{T}'})$. □

We illustrate the result in Corollary 5.14 with an example.

Example 5.15. Fix $\mathcal{T}$ and $\mathcal{T}'$ to be the staged trees in Figure 3. The staged tree $\mathcal{T}'$ is considered in [6, Section 6] as an example of the possible unfolding of events in a cell culture. A thorough discussion of this example and its difference with other graphical models is also contained in [6, Section 6]. Here we explain how to obtain a Gröbner basis for $\ker(\varphi_{\mathcal{T}'})$ using Corollary 5.14. The tree $\mathcal{T}'$ is balanced and statistically equivalent to $\mathcal{T}$. By Corollary 5.6, $\mathcal{T}$ has a quadratic Gröbner basis with square-free initial ideal. Using the lemmas preceding this example, there is a bijection between the root-to-leaf paths in $\mathcal{T}$ and $\mathcal{T}'$; thus $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[p]_{\mathcal{T}'}$ are isomorphic. Under this isomorphism, the Gröbner basis for $\ker(\bar{\varphi}_{\mathcal{T}})$ is a Gröbner basis for $\ker(\bar{\varphi}_{\mathcal{T}'})$; its generators are

$$p_{0111}p_{10} - p_{0011}p_{110}, \quad p_{0011}p_{0110} - p_{0010}p_{0111}, \quad p_{0110}p_{10} - p_{0010}p_{110},$$
$$p_{0010}p_{010} - p_{000}p_{0110}, \quad p_{0011}p_{010} - p_{000}p_{0111}, \quad p_{010}p_{10} - p_{000}p_{110}.$$

Acknowledgements

Computations using the symbolic algebra software *Macaulay2* [12] were crucial for the development of this paper. We thank Thomas Kahle for his support in the completion of this project and Bernd Sturmfels for a careful reading of an earlier version of this manuscript. Both authors were supported by the Deutsche Forschungsgemeinschaft (314838170, GRK 2297 MathCoRe).

References

[1] S. Aoki, H. Hara, and A. Takemura, *Markov bases in algebraic statistics*, Springer, 2012.

[2] Y. Chen, I. H. Dinwoodie, and S. Sullivant, "Sequential importance sampling for multiway tables", *Ann. Statist.* **34**:1 (2006), 523–545.

[3] R. A. Collazo, C. Görgen, and J. Q. Smith, *Chain event graphs*, CRC Press, Boca Raton, FL, 2018.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 12, No. 1, 2021, p.01-20
https://publishoa.com
ISSN: 1309-3452

[4] P. Diaconis and B. Sturmfels, "Algebraic algorithms for sampling from conditional distributions", *Ann. Statist.* **26**:1 (1998), 363–397.

[5] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars **39**, Birkhäuser, Basel, 2009.

[6] E. Duarte and C. Görgen, "Equations defining probability tree models", *J. Symbolic Comput.* **99** (2020), 127–146.

[7] L. D. Garcia, M. Stillman, and B. Sturmfels, "Algebraic geometry of Bayesian networks", *J. Symbolic Comput.* **39**:3-4 (2005), 331–355.

[8] D. Geiger, C. Meek, and B. Sturmfels, "On the toric algebra of graphical models", *Ann. Statist.* **34**:3 (2006), 1463–1492.

[9] C. Görgen, *An algebraic characterisation of staged trees: their geometry and causal implications.*, Ph.D. thesis, University of Warwick, 2017, available at http://webcat.warwick.ac.uk/record=b3084201∼S15.

[10] C. Görgen and J. Q. Smith, "Equivalence classes of staged trees", *Bernoulli* **24**:4A (2018), 2676–2692.

[11] C. Görgen, A. Bigatti, E. Riccomagno, and J. Q. Smith, "Discovery of statistical equivalence classes using computer algebra", *Internat. J. Approx. Reason.* **95** (2018), 167–184.

[12] D. R. Grayson and M. E. Stillman, "Macaulay 2, a software system for research in algebraic geometry", available at http://www.math.uiuc.edu/Macaulay2.

[13] J. Herzog, T. Hibi, and H. Ohsugi, *Binomial ideals*, Graduate Texts in Mathematics **279**, Springer, 2018.

[14] T. Kahle, J. Rauh, and S. Sullivant, "Algebraic aspects of conditional independence and graphical models", pp. 61–80 in *Handbook of graphical models*, edited by M. Maathuis et al., CRC Press, Boca Raton, FL, 2019.

[15] J. Q. Smith and P. E. Anderson, "Conditional independence and chain event graphs", *Artificial Intelligence* **172**:1 (2008), 42–68.

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

[17] S. Sullivant, "Toric fiber products", *J. Algebra* **316**:2 (2007), 560–577.

[18] S. Sullivant, *Algebraic statistics*, Graduate Studies in Mathematics **194**, American Mathematical Society, Providence, RI, 2018.

[19] P. Thwaites, J. Q. Smith, and E. Riccomagno, "Causal analysis with chain event graphs", *Artificial Intelligence* **174**:12-13 (2010), 889–909.