Foundations of Computational Mathematics (2023) 23:2151–2202
https://doi.org/10.1007/s10208-022-09583-7

Families of Polytopes with Rational Linear Precision in Higher Dimensions

Isobel Davies[1] · Eliana Duarte[2] · Irem Portakal[3] · Miruna-Ştefana Sorea[4,5]

Received: 27 September 2021 / Revised: 14 June 2022 / Accepted: 27 June 2022 /
Published online: 8 August 2022

Abstract

In this article, we introduce a new family of lattice polytopes with rational linear precision. For this purpose, we define a new class of discrete statistical models that we call multinomial staged tree models. We prove that these models have rational maximum likelihood estimators (MLE) and give a criterion for these models to be log-linear. Our main result is then obtained by applying Garcia-Puente and Sottile's theorem that establishes a correspondence between polytopes with rational linear precision and log-linear models with rational MLE. Throughout this article, we also study the interplay between the primitive collections of the normal fan of a polytope with rational linear precision and the shape of the Horn matrix of its corresponding statistical model. Finally, we investigate lattice polytopes arising from toric multinomial staged tree models, in terms of the combinatorics of their tree representations.

Communicated by Teresa Krick.

✉ Isobel Davies
isobel.davies@ovgu.de

Eliana Duarte
eliana.gelvez@fc.up.pt

Irem Portakal
mail@irem-portakal.de

Miruna-Ştefana Sorea
msorea@sissa.it

1 Otto-von-Guericke Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

2 Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

3 Technische Universität München, Lehrstuhl für Mathematische Statistik, Boltzmannstr. 3 Garching b, 85748 Munich, Germany

4 SISSA, via Bonomea, 265-34136 Trieste, Italy

5 RCMA Lucian Blaga University Sibiu, Bd-ul Victoriei nr.10, Sibiu 550024, Romania

Keywords Lattice polytope · Horn parametrisation · Rational linear precision · Primitive collection · Log-linear model · Maximum likelihood estimator · Toric variety · Staged tree

Mathematics Subject Classification Primary: 52B20 · Secondary: 14M25, 62R01, 65D17

1 Introduction

In geometric modelling, pieces of parametrised curves and surfaces are used as building blocks to describe geometric shapes in 2D and 3D. Some of the most widely used parametric units for this purpose are Bézier curves, triangular Bézier surfaces, and tensor product surfaces. These pieces of curves and surfaces are constructed using a set of polynomial blending functions defined on the convex hull of a set of points $\mathscr{A}$, together with a set of control points. Taking as inspiration the theory of toric varieties and the form of the blending functions for the previous examples, Krasauskas introduced the more general notion of a *toric patch* whose domain is a lattice polytope $P \subseteq \mathbb{R}^d$ [15]. The blending functions, $\{\beta_{w,m} : P \to \mathbb{R}\}_{m\in\mathscr{A}}$, of a toric patch are constructed from the set of lattice points $\mathscr{A} := P \cap \mathbb{Z}^d$ and a vector of positive weights w associated with each point in $\mathscr{A}$.

A significant difference between an arbitrary toric patch and one of the triangular or tensor product patches is that its blending functions do not necessarily satisfy the property of *linear precision*. A collection of blending functions $\{\beta_m : P \to \mathbb{R}\}_{m\in\mathscr{A}}$ has linear precision if for any affine function $\Lambda : \mathbb{R}^d \to \mathbb{R}$,

$$\Lambda(u) = \sum_{m\in\mathscr{A}} \Lambda(m)\beta_m(u), \text{ for all } u \in P.$$

Thus, linear precision is the ability of the blending functions to replicate affine functions and it is desirable from the practical standpoint [12]. To decide if the collection of blending functions associated with (P, w) has linear precision it is necessary and sufficient to check that the identity $p = \sum_{m\in\mathscr{A}} \beta_{w,m}(p)m$ holds for all $p \in P$ [12, Proposition 11], in this case we say the pair (P, w) has *strict linear precision*. If there exist rational blending functions $\{\hat{\beta}_{w,m} : P \to \mathbb{R}\}_{m\in\mathscr{A}}$ that are nonnegative on P, form a partition of unity, parametrise the same variety $X_{\mathscr{A},w}$ as the blending functions $\{\beta_{w,m} : P \to \mathbb{R}\}_{m\in\mathscr{A}}$, and also have linear precision, we say the pair (P, w) has *rational linear precision*.

It is an open problem, motivated by geometric modelling, to characterise all pairs (P, w) that have rational linear precision in dimension $d \geq 3$ [5, 15]. The classification of all such pairs in dimension $d = 2$ is given in [4].

Garcia-Puente and Sottile studied the property of rational linear precision for toric patches by associating a scaled projective toric variety $X_{\mathscr{A},w}$ to the pair (P, w) [12]. The variety $X_{\mathscr{A},w}$ is the image of the map $[w\chi]_{\mathscr{A}} : (\mathbb{C}^*)^d \to \mathbb{P}^{n-1}$ defined by $\mathbf{t} \mapsto [w_1\mathbf{t}^{m_1} : w_2\mathbf{t}^{m_2} : \ldots : w_s\mathbf{t}^{m_n}]$ where $\mathscr{A} = \{m_1, \ldots, m_n\}$. One of their main results states that a pair (P, w) has rational linear precision if and only if the variety

$X_{\mathscr{A},w}$, seen as a discrete statistical model, has rational maximum likelihood estimator (MLE). This result establishes a communication channel between geometric modelling and algebraic statistics. Thus, it is natural to use ideas from Algebraic Statistics to study the property of rational linear precision.

Models with rational MLE are algebraic varieties that admit a parametrisation known as Horn uniformisation [11, 14]. This parametrisation depends on a *Horn matrix* H and a coefficient for each column of H. In their recent study of moment maps of toric varieties [5], Clarke and Cox go one step further in strengthening the relationship between pairs (P, w) with rational linear precision and models $X_{\mathscr{A},w}$ with rational MLE by using Horn matrices to characterise all pairs (P, w) that have strict linear precision. They propose the use of Horn matrices to study polytopes with rational linear precision and state several questions and conjectures about the relationship between the Horn matrix of $X_{\mathscr{A},w}$ and the primitive collections of the normal fan of P.

In this article, we study the property of rational linear precision of pairs (P, w) from the point of view of Algebraic Statistics. Our main contribution is Theorem 4.1, which introduces a new family of polytopes (with associated weights) that has rational linear precision. We construct this family from a subclass of discrete statistical models introduced in Sect. 4 that we call *multinomial staged trees*. Looking at specific members of this family in 3D, we settle some of the questions raised in [5] related to Horn matrices and primitive collections.

This paper is structured as follows: In Sects. 2.1–2.4 we provide background material on rational linear precision, discrete statistical models with rational MLE, and Horn matrices. In Sect. 2.5, we state Questions 2.1 and 2.2 which guided our investigations related to Horn matrices and primitive collections. These questions are followed by a quick outline referring to the places in this article where they are addressed. In Sect. 3, we characterise the shape of the Horn matrix for pairs (P, w) in $2D$. We also present a family of pairs (P, w) in $3D$ that has rational linear precision and explain several aspects of this family that relate to Questions 2.1 and 2.2. In Sect. 4.1, we define multinomial staged tree models, we prove they have rational MLE in Sect. 4.6 and we characterise the subclass of these models that are toric varieties in Sect. 4.5. These results lead to our main theorem, Theorem 4.1. Finally, in Sect. 5, we show that the examples from Sect. 3 are all multinomial staged trees and prove our conjectures about the relationship between the combinatorics of the trees and primitive collections.

2 Preliminaries

We assume the reader is familiar with introductory material on computational algebraic geometry and toric geometry at the level of [7] and [8].

2.1 Notation and Conventions

We consider pairs (P, w) where P is a d-dimensional lattice polytope in $\mathbb{R}^d$, $\mathbb{Z}^d$ is the fixed lattice, $\mathscr{A} = P \cap \mathbb{Z}^d = \{m_1, \ldots, m_n\}$ and w is a vector of positive weights indexed by $\mathscr{A}$. Fix $n_1, \dots, n_r$ to be the inward facing primitive normal vectors of P corresponding to the facets $F_1, \dots, F_r$ of P and let $a_1, \dots, a_r$ be the corresponding integer translates in the facet presentation of P given by $P = \{p \in \mathbb{R}^d : \langle p, n_i \rangle \geq -a_i, \forall i \in \{1, \dots, r\}\}$. The lattice distance to the face F_i evaluated at $p \in \mathbb{R}^d$ is

$$h_i(p) = \langle p, n_i \rangle + a_i, \quad i = 1, \dots, r,$$

we record each of these values in the vector $h(p) = (h_1(p), \dots, h_r(p))$ The value $h_i(m_j)$ is the lattice distance from the j-th lattice point to the i-th facet. The matrix with ij entry equal to $h_i(m_j)$ is the *lattice distance matrix* of $\mathscr{A}$. We will often consider products of linear forms or variables whose exponents are given by vectors. For vectors $v = (v_1, \dots, v_N)$, $w = (w_1, \dots, w_N)$ we use v^w to denote the product $\prod_{i=1}^N v_i^{w_i}$ and use the convention that $0^0 = 1$. Common choices for v, w in the upcoming sections are the vectors $\mathbf{t} = (t_1, \dots, t_d)$, $h(p)$ and $h(m)$, $m \in \mathscr{A}$. If P is a polytope and $a \geq 1$ is an integer, aP denotes its dilation.

2.2 Rational Linear Precision

In this section, we follow closely the exposition in [5]. A more elementary introduction to this topic is available in [6, Chapter 3].

Definition 2.1 Let $P \subseteq \mathbb{R}^d$ be a full-dimensional polytope and let $w = (w_1, \dots, w_n)$ be a vector of positive weights.

1. For $1 \leq j \leq n$ and $p \in P$, $\beta_j(p) := h(p)^{h(m_j)} = \prod_{i=1}^r h_i(p)^{h_i(m_j)}$.
2. The functions $\beta_{w,j} := w_j \beta_j / \beta_w$ are the *toric blending functions* of (P, w), where $\beta_w(p) := \sum_{j=1}^n w_j \beta_j(p)$.
3. Given *control points* $\{Q_j\}_{1 \leq j \leq n} \in \mathbb{R}^\ell$, the *toric patch* $F : P \to \mathbb{R}^\ell$ is defined by

$$p \mapsto \frac{1}{\beta_w(p)} \sum_{j=1}^n w_j \beta_j(p) Q_j. \tag{1}$$

In part (3) of the previous definition, it is natural to choose the set of control points to be $\mathscr{A}$.

Definition 2.2 Let (P, w) be as in Definition 2.1.

1. The *tautological patch* $K_w : P \to P$ is the toric patch (1) where $\{Q_j = m_j\}_{1 \leq j \leq n}$.
2. The pair (P, w) has *strict linear precision* if K_w is the identity on P, that is

$$p = \frac{1}{\beta_w(p)} \sum_{j=1}^n w_j \beta_j(p) m_j, \text{ for all } p \in P.$$

3. The pair (P, w) has *rational linear precision* if there are rational functions $\hat{\beta}_1, \dots, \hat{\beta}_n$ on $\mathbb{C}^d$ satisfying:

(a) $\sum_{j=1}^{n} \hat{\beta}_j = 1$ as rational functions on $\mathbb{C}^d$.
(b) The map $\hat{\beta} : \mathbb{C}^d \dashrightarrow X_{\mathscr{A},w} \subset \mathbb{P}^{n-1}$, $\mathbf{t} \to (\hat{\beta}_1(\mathbf{t}), \dots, \hat{\beta}_n(\mathbf{t}))$ is a rational parametrisation of $X_{\mathscr{A},w}$.
(c) For every $p \in P \subset \mathbb{C}^d$, $\hat{\beta}_j(p)$ is defined and is a nonnegative real number.
(d) $\sum_{j=1}^{n} \hat{\beta}_j(p) m_j = p$ for all $p \in P$.

Remark 2.1 We are interested in the property of linear precision. By [12, Proposition 2.6], the blending functions $\{\beta_{w,j} : 1 \le j \le n\}$ have linear precision if and only if the pair (P, w) has strict linear precision. Rational linear precision requires the existence of rational functions $\{\hat{\beta}_j : P \to \mathbb{R} : 1 \le j \le n\}$ that have strict linear precision, and that are related to the blending functions of (P, w) via 3(b) in Definition 2.2.

Remark 2.2 An alternative way to specify a pair (P, w) is by using a homogeneous polynomial $F_{\mathscr{A},w}$ whose dehomogenisation $f_{\mathscr{A},w} = \sum_{j=1}^{n} w_j \mathbf{t}^{m_i}$ encodes the weights in the coefficients and the lattice points in $\mathscr{A}$ as exponents. We use this notation in Sect. 3 to describe toric patches in 2D and 3D.

Remark 2.3 If (P, w) has rational linear precision then $(aP, \tilde{w})$, $a \ge 1$, also has this property where $\tilde{w}$ is the vector of coefficients of $(f_{\mathscr{A},w})^a$. See [4, Lemma 2.2].

Example 2.1 Consider the trapezoid $P = conv((0,0), (3,0), (1,2), (0,2))$, with ordered set of lattice points $\mathscr{A}$ and vector of weights w, given as follows:

$$\begin{aligned} \mathscr{A} &= \{(0,2), (1,2), (0,1), (1,1), (2,1), (0,0), (1,0), (2,0), (3,0)\} \\ w &= (1, 1, 2, 4, 2, 1, 3, 3, 1). \end{aligned}$$

The polynomial $f_{\mathscr{A},w}(s,t) = (1+s)(1+s+t)^2$ encodes (P, w). The lattice distance functions for the facets of P are:

$$h_1(s,t) = s, \qquad h_2(s,t) = t, \qquad h_3(s,t) = 3 - t - s, \qquad h_4(s,t) = 2 - t.$$

The toric blending functions for (P, w) are

$$\beta_{w,(i,j)}(s,t) = \binom{2}{j}\binom{3-j}{i} \frac{s^i t^j (3-s-t)^{3-i-j}(2-t)^{2-j}}{6 - 4t + t^2}, \text{ where } (i,j) \in \mathscr{A}.$$

The pair (P, w) does not have strict linear precision, but it has rational linear precision. By Proposition 3.1 the parametrisation of the patch which has linear precision is given by:

$$\hat{\beta}_{w,(i,j)}(s,t) = \binom{2}{j}\binom{3-j}{i} \frac{s^i t^j (3-s-t)^{3-i-j}(2-t)^{2-j}}{4(3-t)^{3-j}}, \text{ where } (i,j) \in \mathscr{A}.$$

Example 2.2 Let $\Delta_d = \{x \in \mathbb{R}^d : x_1 + \cdots + x_d \le 1, \; x_i \ge 0\}$ be the standard simplex in $\mathbb{R}^d$ and $k\Delta_d$ be its dilation by the integer $k \ge 1$. To a point

$m = (a_1, \ldots, a_d) \in \mathscr{A} = k\Delta_d \cap \mathbb{Z}^d$ we associate the weight

$$w_m = \binom{k}{m} = \binom{k}{k-|m|, a_1, \ldots, a_d}, \text{ where } |m| = a_1 + \cdots + a_d.$$

The pair $(k\Delta_d, w)$ has strict linear precision, see [12, Example 4.7]. By [5, Example 4.6], the product of two pairs, (P, w) and $(Q, \tilde{w})$, with strict linear precision also has strict linear precision. Hence, *the Bézier simploids* [10], which are polytopes of the form $k_1\Delta_{d_1} \times \cdots \times k_r\Delta_{d_r}$ for positive integers $k_1, \ldots, k_r, n_1, \ldots n_r$, have strict linear precision. Conjecture 4.8 in [5] states that these are the only polytopes with strict linear precision.

2.3 Discrete Statistical Models with Rational MLE

A probability distribution of a discrete random variable X with outcome space $\{1, \ldots, n\}$ is a vector $(p_1, \ldots, p_n) \in \mathbb{R}^n$ where $p_i = P(X = i)$, $i \in \{1, \ldots, n\}$, $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. The open probability simplex

$$\Delta_{n-1}^\circ = \{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid p_i > 0, p_1 + \cdots + p_n = 1\}$$

consists of all strictly positive probability distributions for a discrete random variable with n outcomes. A discrete statistical model $\mathcal{M}$ is a subset of Δ_{n-1}°.

Given a set $\mathcal{D} = \{X_1, \ldots, X_N\}$ of independent and identically distributed observations of X, we let $u = (u_1, \ldots, u_n)$ be the vector where u_i is the number of times the outcome i appears in $\mathcal{D}$. The likelihood function $L(p, u) : \mathcal{M} \to \mathbb{R}_{\geq 0}$ defined by $(p_1, \ldots, p_n) \mapsto \prod p_i^{u_i}$ records the probability of observing the set $\mathcal{D}$. The *maximum likelihood estimator* (MLE) of the model $\mathcal{M}$ is the function $\Phi : \mathbb{R}^n \to \mathcal{M}$ that sends each vector $(u_1, \ldots, u_n)$ to the maximiser of $L(p, u)$, i.e.

$$\Phi(u) := \arg\max L(p, u).$$

For arbitrary $\mathcal{M}$, the problem of estimating $\arg\max L(p, u)$ is a difficult one. However, for special families, such as discrete exponential families, there are theorems that guarantee the existence and uniqueness of $\arg\max L(p, u)$ when u has nonzero entries. We are interested in the case where Φ is a rational function of u.

Definition 2.3 Let $\mathcal{M}$ be a discrete statistical model with MLE $\Phi : \mathbb{R}^n \to \mathcal{M}$, $u \mapsto \hat{p}$. The model $\mathcal{M}$ has *rational MLE* if the coordinate functions of Φ are rational functions in u.

Example 2.3 Consider the model $\mathcal{M}$ of two independent binary random variables X, Y, with outcome set $\{0, 1\}$ and $p_{ij} = P(X = i, Y = j)$. This model is the set of all points $(p_{00}, p_{01}, p_{10}, p_{11})$ in Δ_3° that satisfy the equation $p_{00}p_{11} - p_{10}p_{01} = 0$. The model has rational MLE $\Phi : \mathbb{R}^4 \to \mathcal{M}$ where

$$(u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2}\right)$$

and $u_{i+} = u_{i0} + u_{i1}, u_{+j} = u_{0j} + u_{1j}, u_{++} = \sum_{i,j\in\{0,1\}} u_{ij}$.

Definition 2.4 A *Horn matrix* is an integer matrix whose column sums are equal to zero. Given a Horn matrix H, with columns $h_1, \dots, h_n$, and a vector $\lambda \in \mathbb{R}^n$, the Horn parameterisation $\varphi_{(H,\lambda)} : \mathbb{R}^n \to \mathbb{R}^n$ is the rational map given by

$$u \mapsto (\lambda_1(Hu)^{h_1}, \lambda_2(Hu)^{h_2}, \dots, \lambda_n(Hu)^{h_n}).$$

Example 2.4 The MLE Φ in Example 2.3 is given by a Horn parametrisation $\varphi_{(H,\lambda)}$, where

$$u = \begin{pmatrix} u_{00} \\ u_{01} \\ u_{10} \\ u_{11} \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -2 & -2 & -2 & -2 \end{pmatrix}, \quad Hu = \begin{pmatrix} u_{0+} \\ u_{1+} \\ u_{+0} \\ u_{+1} \\ -2u_{++} \end{pmatrix}, \quad \text{and } \lambda = (4,4,4,4).$$

Definition 2.5 We say that (H, λ) is a Horn pair if: (1) the sum of the coordinates of $\varphi_{(H,\lambda)}$ as rational functions in u is equal to 1 and (2) the map $\varphi_{(H,\lambda)}$ is defined for all positive vectors and it sends these to positive vectors in $\mathbb{R}^r$.

Theorem 2.1 [11, Theorem 1] *A discrete statistical model $\mathcal{M}$ has rational MLE Φ if and only if there exists a Horn pair (H, λ) such that $\mathcal{M}$ is the image of the Horn parametrisation $\varphi_{(H,\lambda)}$ restricted to the open orthant $\mathbb{R}^n_{>0}$ and $\Phi = \varphi_{(H,\lambda)}$ on $\mathbb{R}^n_{>0}$.*

It is possible that two Horn parametrisations $\varphi_{(H,\lambda)}$ and $\varphi_{(\tilde{H},\tilde{\lambda})}$ are equal even if $H \neq \tilde{H}$ and $\lambda \neq \tilde{\lambda}$. A Horn matrix H is *minimal* if it has no zero rows and no two rows are linearly dependent. By [5, Proposition 6.11] there exists a unique, up to permutation of the rows, minimal Horn matrix that defines $\varphi_{(H,\lambda)}$. Any other pair (H, λ) that defines the same Horn parametrisation may be transformed into one where H is a minimal Horn matrix; this is done by adding collinear rows, deleting zero rows, and adjusting the vector λ accordingly, see [11, Lemma 3]. We end this section by noting that [11, Proposition 23] states that if (H, λ) is a minimal Horn pair, then every row of H has either all entries greater than or equal zero or all entries less than or equal to zero. We call the submatrix of H that consists of all rows with nonnegative entries, the positive part of H, and its complement the negative part of H.

2.4 The Links between Algebraic Statistics and Geometric Modelling

The links referred to in the title of this section are Theorems 2.3 and 2.4. Given a pair (P, w), the *scaled projective toric variety* $X_{\mathscr{A},w}$ is the image of the map $[w\chi]_{\mathscr{A}} : (\mathbb{C}^*)^d \to \mathbb{P}^{n-1}$ defined by $\mathbf{t} \mapsto [w_1\mathbf{t}^{m_1} : w_2\mathbf{t}^{m_2} : \dots : w_n\mathbf{t}^{m_n}]$. To consider the maximum likelihood estimation problem in the realm of complex algebraic geometry we consider the variety $W = V(x_1 \dots x_n(x_1 + \dots + x_n)) \subset \mathbb{P}^{n-1}$ and the map

$$X_{\mathscr{A},w} \setminus W \to (\mathbb{C}^*)^n, \quad [x_1 : \ldots : x_n] \mapsto \frac{1}{x_1 + \cdots + x_n}(x_1, \ldots, x_n).$$

The image of this map is closed and denoted by $Y_{\mathscr{A},w}$. We call $Y_{\mathscr{A},w}$ a *scaled very affine toric variety*. The set $\mathcal{M}_{\mathscr{A},w} = Y_{\mathscr{A},w} \cap \mathbb{R}^n_{>0}$ is a subset of the open simplex Δ°_{n-1} and as such it is a statistical model. This class of models, of the form $\mathcal{M}_{\mathscr{A},w}$, are known as *log-linear models*.

Remark 2.4 The variety $Y_{\mathscr{A},w}$ admits two parameterisations, one by monomials and one by toric blending functions [5, Proposition 5.2]. These are

$$\begin{aligned} \overline{w\chi_{\mathscr{A}}} : \mathbb{C}^d \dashrightarrow Y_{\mathscr{A},w}, \quad & \mathbf{t} \mapsto \left(\frac{w_1 \mathbf{t}^{m_1}}{\sum_{j=1}^{s} w_j \mathbf{t}^{m_j}}, \ldots, \frac{w_n \mathbf{t}^{m_n}}{\sum_{j=1}^{s} w_j \mathbf{t}^{m_j}} \right), \\ \overline{w\beta_{\mathscr{A}}} : \mathbb{C}^d \dashrightarrow Y_{\mathscr{A},w}, \quad & \mathbf{t} \mapsto \left(\frac{w_1 \beta_1(\mathbf{t})}{\beta_w(\mathbf{t})}, \ldots, \frac{w_n \beta_n(\mathbf{t})}{\beta_w(\mathbf{t})} \right). \end{aligned} \tag{2}$$

We now consider the maximum likelihood estimation problem for log-linear models. Given a vector of counts u, we let $\overline{u}:=u/|u| \in \Delta^\circ_{n-1}$ be the empirical distribution, where $|u| = \sum u_j$. We define the tautological map $\tau_{\mathscr{A}}$ following the convention in [5],

$$\tau_{\mathscr{A}} : \Delta^\circ_{n-1} \to P^\circ, \qquad (\overline{u}_1, \ldots, \overline{u}_n) \mapsto \sum_{j=1}^{n} \overline{u}_j m_j. \tag{3}$$

The maximum likelihood estimate of $\overline{u}$ for the model $\mathcal{M}_{\mathscr{A},w}$ exists and it is unique whenever all entries of $\overline{u}$ are positive.

Theorem 2.2 *[18, Corollary 7.3.9] The maximum likelihood estimate in $\mathcal{M}_{\mathscr{A},w}$ for the empirical distribution $\overline{u} \in \Delta^\circ_{n-1}$ is the unique point $\hat{p} \in \mathcal{M}_{\mathscr{A},w}$ that satisfies $\tau_{\mathscr{A}}(\hat{p}) = \tau_{\mathscr{A}}(\overline{u})$.*

In the Algebraic Statistics literature, models with rational MLE are also known as models with *maximum likelihood degree* equal to 1. Even though the previous theorem guarantees the existence and uniqueness of the MLE, it is not true that every log-linear model has rational MLE. We refer the reader to [1] for several examples of log-linear models that do not have rational MLE, or equivalently for examples of models with maximum likelihood degree greater than 1. We end this section by recalling two theorems that connect models with rational MLE and pairs with rational linear precision.

Theorem 2.3 [12, Proposition 5.1] *The pair (P, w) has rational linear precision if and only if the model $\mathcal{M}_{\mathscr{A},w}$ has rational MLE.*

Theorem 2.4 [5] *Set $a_P := \sum_{i=1}^r a_i$ and $n_P := \sum_{i=1}^r n_i$. The following are equivalent:*

1. *The pair (P, w) has strict linear precision.*
2. $n_P = 0$ *and* $\beta_w(p) = \sum_{j=1}^{n} w_j \beta_j(p) = \sum_{j=1}^{n} w_j \prod_{i=1}^{r} h_i(p)^{h_i(m_j)}$ *is a nonzero constant c.*
3. $\mathcal{M}_{\mathscr{A},w}$ *has rational MLE with minimal Horn pair (H, λ) given by*

$$H = \begin{pmatrix} h_1(m_1) & h_1(m_2) & \dots & h_1(m_n) \\ h_2(m_1) & h_2(m_2) & \dots & h_2(m_n) \\ \vdots & \vdots & & \vdots \\ h_r(m_1) & h_r(m_2) & \dots & h_r(m_n) \\ -a_P & -a_P & \dots & -a_P \end{pmatrix}, \quad \lambda_j = \frac{w_j}{c}(-a_P)^{a_P}$$

2.5 Primitive Collections and Horn Pairs

The notion of primitive collections was first introduced by Batyrev in [3] for a smooth and projective toric variety X_{Σ_P} of the polytope P. It provides an elegant description of the nef cone for X_{Σ_P}. This result has been generalised to the simplicial case and the definition of primitive collections for the non-simplicial case has been introduced in [9].

Definition 2.6 Let Σ_P be a normal fan. For $\sigma \in \Sigma_P$, $\sigma(1)$ denotes the 1-faces of σ. A subset $C \subseteq \Sigma_P(1)$ of 1-faces of Σ_P is called a *primitive collection* if

1. $C \nsubseteq \sigma(1)$ for all $\sigma \in \Sigma_P$.
2. For every proper subset $C' \subsetneq C$, there exists $\sigma \in \Sigma_P$ such that $C' \subseteq \sigma(1)$.

In particular, if Σ_P is simplicial, C is a primitive collection if C does not generate a cone of Σ_P but every proper subset does.

For strict linear precision, Theorem 2.4 gives the minimal Horn pair based only on the lattice distance functions of the facets of the polytope. The authors in [5] raise the question whether it is possible to obtain a similar description of minimal Horn pairs of polytopes with rational linear precision.

Question 2.1 Is the positive part of the minimal Horn matrix of a pair (P, w) with rational linear precision always equal to the lattice distance matrix of $\mathscr{A}$?

For pairs (P, w) in 2D with rational linear precision, and the family of prismatoids in Sect. 3.2, the answer to Question 2.1 is affirmative, see Theorem 3.1, Proposition 3.2, and Appendix A.

In [5], there are two examples, one of a trapezoid [5, Section 8.1] and one of a decomposable graphical model [5, Section 8.3], where the positive part of the Horn matrix is the lattice distance matrix of $\mathscr{A}$ and the negative rows are obtained via the primitive collections of the normal fan of P. These examples motivate the next definition and Question 2.2:

Definition 2.7 To a pair (P, w) we associate the matrix $M_{\mathscr{A},\Sigma_P}$ which consists of the lattice distance matrix of $\mathscr{A}$, with ij-th entry $h_i(m_j)$, together with negative rows given by summing the rows of the lattice distance functions $-h_i$, for which the facet normals n_i belong to the same primitive collection of Σ_P.

Question 2.2 For a pair (P, w) with rational linear precision is there a Horn pair (H, λ) for which $H = M_{\mathscr{A},\Sigma_P}$?

For pairs (P, w) in 2D with rational linear precision, the answer to Question 2.2 is affirmative, see Theorem 3.1. For the family of prismatoids in Sect. 3.2, Question 2.2 is affirmative only for certain subclasses, see Theorem 3.3. For an arbitrary pair (P, w) with rational linear precision, the matrix $M_{\mathscr{A},\Sigma_P}$ is not necessarily a Horn matrix, see Sect. 3.3.1. Even in the case that $M_{\mathscr{A},\Sigma_P}$ is a Horn matrix, it does not necessarily give rise to a Horn pair for (P, w), see Sect. 3.3.2. In Sect. 3, we see a number of special cases for which the answer to Question 2.2 is affirmative. In Sect. 5, we give a condition on (P, w) which guarantees the existence of a Horn pair (H, λ) with $H = M_{\mathscr{A},\Sigma_P}$. We also provide an explanation for the negative rows of the Horn matrix in the language of multinomial staged tree models—introduced in Sect. 4.

3 Examples of Horn Pairs in 2D and 3D

In this section, we present families of 2D and 3D pairs (P, w) with rational linear precision and explore the connection between the geometry of the polytope and the shape of its corresponding Horn pair. Throughout this section, we use (s, t), respectively, (s, t, v) to denote $\mathbf{t}$ in the 2D, respectively, 3D case.

3.1 Toric Surface Patches and Horn Pairs in 2D

By [4], the only 2D toric patches with rational linear precision are the Bézier triangles, tensor product patches and trapezoidal patches, seen in Fig. 1. This family of polygons, that we denote by $\mathcal{F}$, consists of all the Newton polytopes of the polynomials

$$f_{a,b,d}(\mathbf{t}) := (1+s)^a((1+s)^d + t)^b \quad \text{for } a, b, d \geq 0.$$

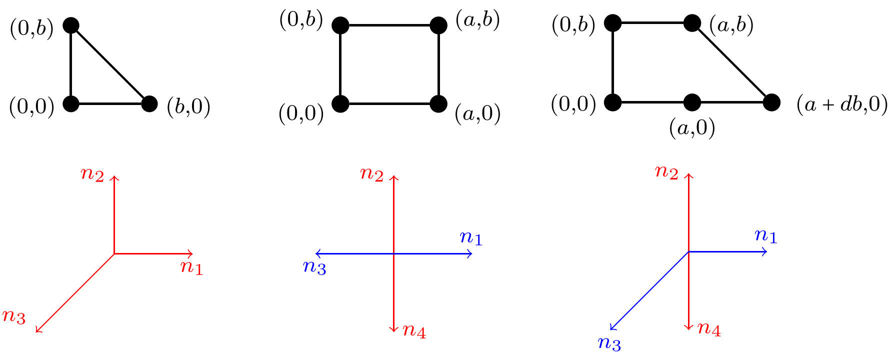


Fig. 1 Left: Bézier triangles. Middle: Tensor product patches. Right: Trapezoids. The normal fan of each polygon is displayed in the bottom row; two rays with the same colour are in the same primitive collection

For general a, b, d, the Newton polytope associated with $f_{a,b,d}$, which we will denote by $T_{a,b,d}$, will be a trapezoidal patch, in the special cases $T_{0,b,1} = b\Delta_2$ and $T_{a,b,0} = a\Delta_1 \times b\Delta_1$, we will have the more familiar Bézier triangles and tensor product patches. The lattice points in $T_{a,b,d} \cap \mathbb{Z}^2$ are $\mathscr{A} = \{(i, j) : 0 \leq j \leq b, 0 \leq i \leq a + d(b - j)\}$. By Theorems 2.1 and 2.3 we know that the statistical model associated with a pair in $\mathcal{F}$ admits a Horn pair.

Proposition 3.1 *A Horn pair (H, λ) of a polygon in the family $\mathcal{F}$ is given by:*

$$H = \begin{pmatrix} h_1(m_1) & \dots & h_1(m) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m) & \dots & h_4(m_n) \\ -(h_1+h_3)(m_1) & \dots & -(h_1+h_3)(m) & \dots & -(h_1+h_3)(m_n) \\ -(h_2+h_4)(m_1) & \dots & -(h_2+h_4)(m) & \dots & -(h_2+h_4)(m_n) \end{pmatrix},$$

$$\lambda_m = (-1)^{a+d(b-j)+b} \binom{(h_2+h_4)(m)}{j} \binom{(h_1+h_3)(m)}{i}$$

where $m{:=}(i, j) \in \mathscr{A}$ is a general lattice point, $m_1, \dots, m_n$ is an ordered list of elements in $\mathscr{A}$, $\mathbf{t}{:=}(s, t)$, and $h_1, \dots, h_4$ are

$$h_1(\mathbf{t}) = s, \quad h_2(\mathbf{t}) = t, \quad h_3(\mathbf{t}) = a + db - s - dt, \quad h_4(\mathbf{t}) = b - t.$$

Proof We use [5, Proposition 8.4]. The terms of the polynomial $f_{a,b,d}(\mathbf{t})$ specify weights and lattice points in $T_{a,b,d} \cap \mathbb{Z}^2$, i.e.

$$f_{a,b,d}(\mathbf{t}) = \sum_{m \in \mathscr{A}} w_m \mathbf{t}^m, \qquad w_m = \binom{b}{j}\binom{a+db-dj}{i}.$$

The monomial parametrisation (2) of $Y_{\mathscr{A},w}$ is

$$\overline{w\chi_{\mathscr{A}}}(\mathbf{t}) = \frac{1}{f_{a,b,d}(\mathbf{t})}(S_{m_1}, \dots, S_{m_n}),$$

where $S_m = w_m \mathbf{t}^m$. Composing the monomial parametrisation with the tautological map (3) gives the following birational map:

$$\begin{aligned}(\tau_{\mathscr{A}} \circ \overline{w\chi_{\mathscr{A}}})(\mathbf{t}) &= \left(\frac{s((a+db)(1+s)^d + at)}{((1+s)^d + t)(1+s)}, \frac{tb}{((1+s)^d + t)} \right) \\ &= \left(\frac{s((a+db)(1+s)^d + at)}{f_{1,1,d}(\mathbf{t})}, \frac{tb}{f_{0,1,d}(\mathbf{t})} \right)\end{aligned}$$

with the following inverse:

$$\varphi(\mathbf{t}) = \left(\frac{s}{a+db-s-dt}, \frac{(a+db-dt)^d t}{(a+db-s-dt)^d(b-t)}\right)$$
$$= \left(\frac{h_1(\mathbf{t})}{h_3(\mathbf{t})}, \frac{((h_1+h_3)(\mathbf{t}))^d h_2(\mathbf{t})}{(h_3(\mathbf{t}))^d h_4(\mathbf{t})}\right).$$

The component of the monomial parametrisation corresponding to a lattice point m, composed with $\varphi(\mathbf{t})$ is given by

$$\left((\overline{w\chi_{\mathscr{A}}} \circ \varphi)(\mathbf{t})\right)_m = \frac{S_m(\varphi(\mathbf{t}))}{f_{a,b,d}(\varphi(\mathbf{t}))}$$
$$= \binom{b}{j}\binom{a+db-dj}{i}\frac{s^i t^j (a+db-s-dt)^{a+db-i-dj}(b-t)^{b-j}}{(a+db-dt)^{a+db-dj} b^b}$$
$$= w_m(-1)^{a+d(b-j)+b} h(\mathbf{t})^{h(m)} \tag{4}$$

where $h(q) = (h_1(q), h_2(q), h_3(q), h_4(q), -h_5(q), -h_6(q))$ $(q \in \{\mathbf{t}, m\})$, $h_5 = h_1 + h_3$ and $h_6 = h_2 + h_4 = b$. It follows from [5, Proposition 8.4] that the Horn parametrisation is $(\overline{w\chi_{\mathscr{A}}} \circ \varphi)(p)$ where

$$p = \sum_{m\in\mathscr{A}} \frac{u_m}{u_+}(m), \qquad u_+ = \sum_{m\in\mathscr{A}} u_m.$$

Therefore, the columns of the Horn matrix are the exponents of

$$w_m(-1)^{a+d(b-j)+b} h(p)^{h(m)}, \quad m \in \mathscr{A},$$

namely $h(m)$. It follows that

$$\lambda_m = (-1)^{a+d(b-j)+b} w_m$$
$$= (-1)^{a+d(b-j)+b}\binom{(h_2+h_4)(m)}{j}\binom{(h_1+h_3)(m)}{i}.$$

□

Remark 3.1 The blending functions $\{\hat{\beta}_m : m \in \mathscr{A}\}$ for each pair (P, w) in $\mathcal{F}$ that satisfy Definition 2.2 (3) are given in equation (4) in the previous proof. For the case $a = d = 1$ and $b = 2$, these are written in Example 2.1.

Remark 3.2 For general a, b, d, Proposition 3.1 gives the minimal Horn pair for $T_{a,b,d}$; this is not the case for $T_{0,b,1}$ and $T_{a,b,0}$. For the last two cases, the minimal Horn pair is obtained after row reduction operations or from Theorem 2.4.

Using Proposition 3.1 and Theorem 2.4 we obtain an affirmative answer to Question 2.1 for pairs (P, w) in 2D. A closer look at the primitive collections in Fig. 1 also reveals an affirmative answer to Question 2.2. This is contained in the next theorem.

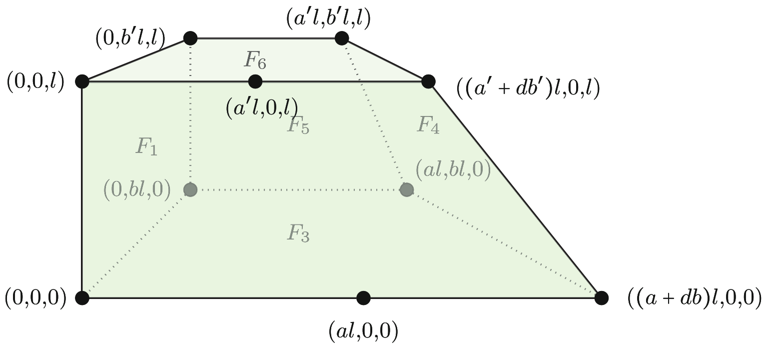


Fig. 2 The general representative of a prismatoid in $\mathcal{P}$ is the convex hull of two trapezoids, $conv(T_{a,b,d} \times \{0\}, T_{a',b',d} \times \{1\})$, dilated by l. For the labelling of facets, we refer to Notation 3.2

Theorem 3.1 *Every pair* (P, w) *in 2D with rational linear precision has a Horn pair* (H, λ) *with* $H = M_{\mathscr{A},\Sigma_P}$.

Proof The normal fans of the polygons in $\mathcal{F}$ are depicted in Fig. 1, in each subcase the shape of the normal fan and its primitive collections are independent of the values of a, b, d. The minimal Horn pair (H, λ) for the 2D simplex, $T_{0,b,1} = b\Delta_2$, given in Theorem 2.4 satisfies $H = M_{\mathscr{A},\Sigma_P}$. This follows because $b\Delta_2$ has one primitive collection, $\{n_1, n_2, n_3\}$ and hence $M_{\mathscr{A},\Sigma_P}$ has a single negative row. For the tensor product patch $T_{a,b,0} = a\Delta_1 \times b\Delta_1$ and the general trapezoid $T_{a,b,d}$, the primitive collections are $\{n_1, n_3\}$ and $\{n_2, n_4\}$. In these cases, the Horn pair (H, λ) in Proposition 3.1 satisfies $H = M_{\mathscr{A},\Sigma_P}$. □

3.2 A Family of Prismatoids with Rational Linear Precision

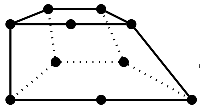
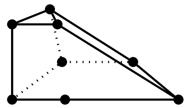
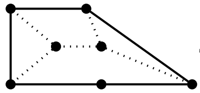
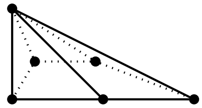
Unlike the 2D case, there is no classification for 3D lattice polytopes with rational linear precision. In this section, we consider the family of prismatoids

$$\begin{aligned}\mathcal{P}:=\{(P, w) : P \text{ is the Newton polytope of } f_{\mathscr{A},w}(\mathbf{t}) = (f_{a,b,d}(\mathbf{t}) + v f_{a',b',d}(\mathbf{t}))^l,\\ w \text{ is the vector of coefficients of } f_{\mathscr{A},w}(\mathbf{t}),\\ a, a', b, b', d, l \in \mathbb{Z}_{\geq 0} \text{ with } a' \leq a,\ b' \leq b\}.\end{aligned}$$

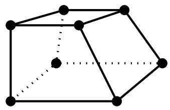
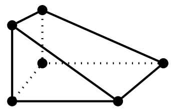
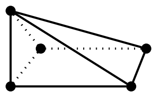
A general element of $\mathcal{P}$ is depicted in Fig. 2, prismatoids for different specialisations of a, a', b, b', d are displayed in Table 1. Note that some 3D Bézier simploids are also obtained by specialisation. Even though Remark 2.3 says it suffices to show that P has rational linear precision for $l = 1$, we do not use this extra assumption.

Table 1 Representative members of $\mathcal{P}$

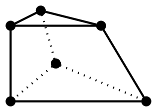
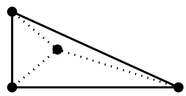
(A) Prismatoids with trapezoidal base $a > 0, b > 0, d > 0, l > 0$

Trapezoidal frusta $a' > 0, b' > 0$		Triangle top (simplex if $d = 1$) $a' = 0, b' > 0$	
Trapezoidal wedges $a' > 0, b' = 0$		Trapezoidal Pyramids $a' = 0, b' = 0$	

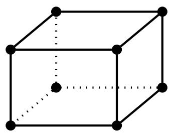
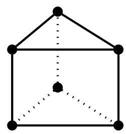
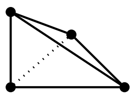
(B) Prismatoids with tensor product base $a > 0, b > 0, d = 0, l > 0$

Tensor product frusta $a' > 0,\ b' > 0$		Tensor product wedges $a' = 0, b' > 0$	
Tensor product pyramids $a' = 0,\ b' = 0$		Tensor product wedges $a' > 0, b' = 0$	

(C) Prismatoids with triangular base $a = 0, b > 0, d > 0, l > 0$

Triangular frusta (simplices if $d = 1$) $a' = 0, b' > 0$		Pyramid (simplex-based if $d = 1$) $a' = 0, b' = 0$	

(D) 3D Bézier simploids $l > 0$

3D Tensor Product $a' = a > 0$ $b' = b > 0$ $d = 0$ $la\Delta_1 \times lb\Delta_1 \times l\Delta_1$		Triangular Prism $a' = a = 0$ $b' = b > 0$ $d = 1$ $lb\Delta_2 \times l\Delta_1$	
3D simplex $a' = a = 0$ $b' = 0, b = 1$ $d = 1$ $l\Delta_3$			

The coordinates of the vertices of each polytope in this table are obtained by specialising the parameters a, a', b, b', d in the coordinates the vertices of the prismatoid in Fig. 2

Proposition 3.2 *The pairs in $\mathcal{P}$ have rational linear precision with a Horn pair (H, λ):*

$$H = \begin{pmatrix} h_1(m_1) & \ldots & h_1(m) & \ldots & h_1(m_n) \\ h_2(m_1) & \ldots & h_2(m) & \ldots & h_2(m_n) \\ h_3(m_1) & \ldots & h_3(m) & \ldots & h_3(m_n) \\ h_4(m_1) & \ldots & h_4(m) & \ldots & h_4(m_n) \\ h_5(m_1) & \ldots & h_5(m) & \ldots & h_5(m_n) \\ h_6(m_1) & \ldots & h_6(m) & \ldots & h_6(m_n) \\ -(h_1+h_4)(m_1) & \ldots & -(h_1+h_4)(m) & \ldots & -(h_1+h_4)(m_n) \\ -(h_2+h_5)(m_1) & \ldots & -(h_2+h_5)(m) & \ldots & -(h_2+h_5)(m_n) \\ -(h_3+h_6)(m_1) & \ldots & -(h_3+h_6)(m) & \ldots & -(h_3+h_6)(m_n) \end{pmatrix},$$

$$\lambda_m = (-1)^{(\sum_{\gamma=1}^{6} h_\gamma)(m)} \binom{(h_3+h_6)(m)}{k} \binom{(h_2+h_5)(m)}{j} \binom{(h_1+h_4)(m)}{i},$$

where $m:=(i, j, k) \in \mathscr{A}$ is a general lattice point, $m_1, \ldots, m_n$ is an ordered list of elements in $\mathscr{A}$, $\mathbf{t}:=(s, t, v)$, and $h_1, \ldots, h_6$ are

$$\begin{aligned} h_1(\mathbf{t}) &= s, & h_4(\mathbf{t}) &= (a+db)l - s - dt - ((a+db) - (a'+db'))v, \\ h_2(\mathbf{t}) &= t & h_5(\mathbf{t}) &= bl - t - (b-b')v, \\ h_3(\mathbf{t}) &= v, & h_6(\mathbf{t}) &= l - v. \end{aligned}$$

Proof The polynomial $f_{\mathscr{A},w}(\mathbf{t})$ in the definition of $\mathcal{P}$, can be expressed as a sum

$$f_{\mathscr{A},w}(\mathbf{t}) = \sum_{k=0}^{(h_3+h_6)(m)} \sum_{j=0}^{(h_2+h_5)(m)} \sum_{i=0}^{(h_1+h_4)(m)} S_m$$

where

$$S_m(\mathbf{t}) = \binom{(h_3+h_6)(m)}{k} \binom{(h_2+h_5)(m)}{j} \binom{(h_1+h_4)(m)}{i} \mathbf{t}^m.$$

We let

$$(\overline{w\chi_{\mathscr{A}}}(\mathbf{t}))_m = \frac{S_m(\mathbf{t})}{f_{\mathscr{A},w}(\mathbf{t})},$$

then, the vector of all $(\overline{w\chi_{\mathscr{A}}}(\mathbf{t}))_m$ gives the monomial parametrisation (2) of $Y_{\mathscr{A},w}$ with weights

$$w_m = \binom{(h_3+h_6)(m)}{k} \binom{(h_2+h_5)(m)}{j} \binom{(h_1+h_4)(m)}{i}.$$

Composing the monomial parametrisation with the tautological map (3) gives the following birational map:

$(\tau_{\mathscr{A}} \circ \overline{w\chi_{\mathscr{A}}})(\mathbf{t})$

$$= \frac{1}{f_{\mathscr{A},w}(\mathbf{t})} \left(\sum_{k=0}^{(h_3+h_6)(m)} \sum_{j=0}^{(h_2+h_5)(m)} \sum_{i=0}^{(h_1+h_4)(m)} S_m \right)(\mathbf{t})$$

$$= \Bigg(\frac{ls\left(((a+db)(1+s)^d + at) f_{a,b',d}(\mathbf{t}) + v(a'+db')(1+s)^d + a't) f_{a',b',d}(\mathbf{t})\right)}{f_{1,1,d}(\mathbf{t})\left(f_{a,b,d}(\mathbf{t})b + v f_{a',b',d}(\mathbf{t})\right)},$$
$$\frac{lt\left(b'v f_{a',b',d}(\mathbf{t}) + b f_{a,b,d}(\mathbf{t})\right)}{f_{0,1,d}(\mathbf{t})\left(f_{a,b,d}(\mathbf{t}) + v f_{a',b',d}(\mathbf{t})\right)}, \frac{lv f_{a',b',d}(\mathbf{t})}{\left(f_{a,b,d}(\mathbf{t}) + v f_{a',b',d}(\mathbf{t})\right)} \Bigg)$$

with the following inverse:

$$\varphi(\mathbf{t}) = \Bigg(\frac{h_1(\mathbf{t})}{h_4(\mathbf{t})}, \frac{h_2(\mathbf{t})((h_1+h_4)(\mathbf{t}))^d}{(h_4(\mathbf{t}))^d h_5(\mathbf{t})},$$
$$\frac{h_3(\mathbf{t})((h_1+h_4)(\mathbf{t}))^{(a+db)-(a'+db')}((h_2+h_5)(\mathbf{t}))^{b-b'}}{(h_4(\mathbf{t}))^{(a+db)-(a'+db')}(h_5(\mathbf{t}))^{b-b'}(h_6(\mathbf{t}))} \Bigg).$$

Composing $\varphi(\mathbf{t})$ with the monomial parametrisation gives

$$\frac{S_m(\varphi(\mathbf{t}))}{f_{\mathscr{A},w}(\varphi(\mathbf{t}))} = w_m(-1)^{(\sum_{\gamma=1}^6 h_\gamma)(m)} h(\mathbf{t})^{h(m)}$$

where $h(q) = (h_1(q), \cdots, h_6(q), -h_7(q), -h_8(q), -h_9(q))$, $(q \in \{\mathbf{t}, m\})$, the functions $h_1, \ldots, h_6$ are as in the statement of the theorem and $h_7 = h_1 + h_4$, $h_8 = h_2 + h_5$, and $h_9 = h_3 + h_6$.

According to [5, Proposition 8.4], the polytope has rational linear precision with weights w_m as defined above and the Horn parametrisation of $Y_{\mathscr{A},w}$ is given by:

$$\frac{S_m(\varphi(p))}{f_{\mathscr{A},w}(\varphi(p))} = w_m(-1)^{(\sum_{\gamma=1}^6 h_\gamma)(m)} h(p)^{h(m)}$$

where

$$p = \sum_{m \in \mathscr{A}} \frac{u_m}{u_+}(m), \qquad u_+ = \sum_{(m) \in \mathscr{A}} u_m.$$

Since the Horn parametrisation is, by definition, a product of linear forms whose exponents match their coefficients, we know that the columns of H are the vectors $h(m)$. It follows that $\lambda_m = (-1)^{(\sum_{\gamma=1}^6 h_\gamma)(m)} w_m$. □

3.3 Minimal Horn Pairs for Prismatoids in $\mathcal{P}$

We now study Questions 2.1 and 2.2 for elements in $\mathcal{P}$. Proposition 3.2 gives a Horn pair (H, λ) for each $(P, w) \in \mathcal{P}$ in Table 1, however H need not be the minimal Horn matrix in each case. By [11, Lemma 9], we can find *the minimal Horn matrix associated with* (P, w) using row reduction operations on H.

Notation 3.2 *We will denote the facets of a general element in $\mathcal{P}$ as follows:*

$$F_1 = \textit{left facet}, \quad F_2 = \textit{front facet}, \quad F_3 = \textit{bottom facet},$$
$$F_4 = \textit{right facet}, \quad F_5 = \textit{back facet}, \quad F_6 = \textit{upper facet}.$$

This labelling is used in Fig. 2. The normal vectors of each facet are:

$$n_1 = (1, 0, 0), \quad n_2 = (0, 1, 0), \quad n_3 = (0, 0, 1),$$
$$n_4 = (-1, -d, -((a + db) - (a' + db'))), \quad n_5 = (0, -1, -(b - b')), \quad n_6 = (0, 0, -1).$$

3.3.1 The Non-simple Prismatoids

The trapezoidal pyramids, tensor product pyramids and prismatoids with triangle on top, depicted in Table 1 (A) and (B), are all examples of non-simple polytopes in $\mathcal{P}$. Their primitive collections are:

Prismatoids with triangle on top	$\{n_1, n_3, n_4\},\ \{n_1, n_2, n_4\},\ \{n_2, n_5\},\ \{n_3, n_6\}$
Trapezoidal pyramids	$\{n_1, n_3, n_4\},\ \{n_2, n_3, n_5\}$
Tensor product pyramids	$\{n_1, n_3, n_4\},\ \{n_2, n_3, n_5\}.$

There is no n_6 for the two pyramids since the facet F_6 has collapsed to a point.

For a pair (P, w) in the subfamily of non-simple prismatoids in $\mathcal{P}$, the matrix $M_{\mathscr{A}, \Sigma_P}$ cannot be a Horn matrix since the primitive collections are not a partition of the 1-dimensional rays of the normal fan and therefore the columns cannot add to zero.

Example 3.1 It follows from Proposition 3.2 that the minimal Horn matrix associated with the tensor product pyramid in Table 1 (B) is:

$$H = \begin{pmatrix} h_1(m_1) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m_n) \\ h_5(m_1) & \dots & h_5(m_n) \\ -(h_1 + h_2 + h_4 + h_5 - h_6)(m_1) & \dots & -(h_1 + h_2 + h_4 + h_5 - h_6)(m_n) \\ -(h_3 + h_6)(m_1) & \dots & -(h_3 + h_6)(m_n) \end{pmatrix}$$

where $m_1, \dots m_n \in \mathscr{A}$, $\mathbf{t}:=(s, t, v)$ and $h_1, \dots, h_6$ are defined to be

$$h_1(\mathbf{t}) = s, \quad h_2(\mathbf{t}) = t, \quad h_3(\mathbf{t}) = v,$$
$$h_4(\mathbf{t}) = al - s - av \quad h_5(\mathbf{t}) = bl - t - bv, \quad h_6(\mathbf{t}) = l - v.$$

We were able to add h_6 to the negative rows $-(h_1 + h_4)$ and $-(h_2 + h_5)$ since all three rows are colinear in this case. As a result, the positive part of the minimal Horn matrix coincides with the lattice distance matrix of $\mathscr{A}$.

3.3.2 The Simple Prismatoids with Fewer Facets

The trapezoidal wedges (A), tensor product wedges (B), triangular frusta (C) and triangular-based pyramids (C) from Table 1 are simple prismatoids with less than 6 facets. The primitive collections in each case are:

Trapezoidal wedges	$\{n_1, n_4\},\ \{n_2, n_3, n_5\}$
Tensor product wedges $a' = 0$	$\{n_1, n_3, n_4\},\ \{n_2, n_5\}$
Tensor product wedges $b' = 0$	$\{n_1, n_4\},\ \{n_2, n_3, n_5\}$
Triangular-based pyramid	$\{n_1, n_2, n_3, n_4\}$
Triangular frusta	$\{n_1, n_2, n_4\},\ \{n_3, n_6\}$

None of the polytopes above, except the triangular frusta, have an upper facet F_6 and hence their normal fans and primitive collections do not include n_6. Also, the triangular-based pyramid and triangular frusta have no back facet F_5 and hence their normal fans and primitive collections do not include n_5. In each case, the primitive collections give a partition of the rays in the normal fan, hence the matrix $M_{\mathscr{A},\Sigma_P}$ associated with (P, w) is a Horn matrix for these cases. The question is whether this Horn matrix belongs to a Horn pair for (P, w).

Example 3.2 Proposition 3.2 gives a Horn pair for the trapezoidal wedge in Table 1 (A), which can be reduced to a Horn pair (H, λ), with:

$$H = \begin{pmatrix} h_1(m_1) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m_n) \\ h_5(m_1) & \dots & h_5(m_n) \\ -(h_1 + h_4)(m_1) & \dots & -(h_1 + h_4)(m_n) \\ -(h_2 + h_5 - h_6)(m_1) & \dots & -(h_2 + h_5 - h_6)(m_n) \\ -(h_3 + h_6)(m_1) & \dots & -(h_3 + h_6)(m_n) \end{pmatrix}$$

where $m_1, \dots m_n \in \mathscr{A}$, $\mathbf{t}{:=}(s, t, v)$ and $h_1, \dots, h_9$ are defined to be

$$\begin{aligned} h_1(\mathbf{t}) &= s, & h_4(\mathbf{t}) &= (a + db)l - s - dt - (a - a' + db)v, \\ h_2(\mathbf{t}) &= t & h_5(\mathbf{t}) &= bl - bv - t, \\ h_3(\mathbf{t}) &= v & h_6(\mathbf{t}) &= l - v. \end{aligned}$$

Let us compare H with the matrix $M_{\mathscr{A},\Sigma_P}$

$$M_{\mathscr{A},\Sigma_P} = \begin{pmatrix} h_1(m_1) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m_n) \\ h_5(m_1) & \dots & h_5(m_n) \\ -(h_1+h_4)(m_1) & \dots & -(h_1+h_4)(m_n) \\ -(h_2+h_3+h_5)(m_1) & \dots & -(h_2+h_3+h_5)(m_n) \end{pmatrix}$$

where $m_1, \dots m_n$ and $h_1, \dots, h_5$ are as in H. For $b = 1$, we see that $h_2 + h_5 - h_6 = 0$, $h_3 + h_6 = l$, and $h_2 + h_3 + h_5 = l$, thus $H = M_{\mathscr{A},\Sigma_P}$. For $b > 1$, H and $M_{\mathscr{A},\Sigma_P}$ are minimal Horn matrices, hence, by uniqueness, $M_{\mathscr{A},\Sigma_P}$ cannot give rise to a Horn pair for (P, w).

For all other examples of simple prismatoids with fewer facets, we noticed a similar phenomenon. Firstly, if n_i is not in the normal fan, then the positive row h_i is collinear with a negative row. In particular, for all these examples the positive part of the minimal Horn matrix coincides with the lattice distance matrix of $\mathscr{A}$. Below we summarise for which parameters the matrix $M_{\mathscr{A},\Sigma_P}$ gives rise to a Horn pair for (P, w), this is not true in general for these families of 'simple prismatoids with fewer facets'.

Trapezoidal wedges	$b = 1$
Tensor product wedges $a' = 0$	$a = 1$
Tensor product wedges $b' = 0$	$b = 1$
Triangular-based pyramid	$b = d = 1$
Triangular frusta	$d = 1$

These seemingly arbitrary constraints have a nice geometric interpretation. The constraint $b = 1$ forces the triangular facet F_1 in the trapezoidal wedges and tensor product wedges ($b' = 0$) to be a simplex. The constraint $a = 1$ forces the triangular facet F_2 in the tensor product wedges ($a' = 0$) to be a simplex. The constraint $b = d = 1$ on the triangular-based pyramid, means it is a 3D simplex and the constraint $d = 1$ on the triangular frusta forces the two triangular facets F_3 and F_6 to be simplices. All the prismatoids considered in this section, except the ones just described, are examples of polytopes with simplicial normal fans for which the answer to Question 2.2 is negative.

3.3.3 The Trapezoidal and Tensor Product Frusta

The primitive collections for the trapezoidal frusta and the tensor product frusta are:

$$\{n_1, n_4\} \qquad \{n_2, n_5\} \qquad \{n_3, n_6\}.$$

It follows easily that the Horn matrix given by Proposition 3.2 is $M_{\mathscr{A},\Sigma_P}$. This matrix is also the minimal Horn matrix for all trapezoidal frusta and for general tensor product frusta. However, there are cases of tensor product frusta, where two or more rows of this matrix are collinear and hence the minimal Horn matrix is not exactly $M_{\mathscr{A},\Sigma_P}$. For an overview of all minimal Horn matrices for the family $\mathcal{P}$ of prismatoids, see Table 3 in Appendix A.

Example 3.3 A Horn matrix associated with the tensor product frusta in Table 1 (B) is

$$M_{\mathscr{A},\Sigma_P} = \begin{pmatrix} h_1(m_1) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m_n) \\ h_5(m_1) & \dots & h_5(m_n) \\ h_6(m_1) & \dots & h_6(m_n) \\ -(h_1+h_4)(m_1) & \dots & -(h_1+h_4)(m_n) \\ -(h_2+h_5)(m_1) & \dots & -(h_2+h_5)(m_n) \\ -(h_3+h_6)(m_1) & \dots & -(h_3+h_6)(m_n) \end{pmatrix}$$

where $m_1, \dots m_n \in \mathscr{A}$, $\mathbf{t}{:=}(s, t, v)$ and $h_1, \dots, h_6$ are defined to be

$$\begin{aligned} h_1(\mathbf{t}) &= s, & h_4(\mathbf{t}) &= al - s - (a - a')v \\ h_2(\mathbf{t}) &= t, & h_5(\mathbf{t}) &= bl - t - (b - b')v, \\ h_3(\mathbf{t}) &= v, & h_6(\mathbf{t}) &= l - v. \end{aligned}$$

If we consider the subfamily of tensor product frusta such that $a = \lambda b,\ a' = \lambda b'$ for $\lambda \geq 1$ or $\lambda = \frac{1}{\mu}$ with $\mu \geq 1$, then the minimal Horn matrix is

$$H = \begin{pmatrix} h_1(m_1) & \dots & h_1(m_n) \\ h_2(m_1) & \dots & h_2(m_n) \\ h_3(m_1) & \dots & h_3(m_n) \\ h_4(m_1) & \dots & h_4(m_n) \\ h_5(m_1) & \dots & h_5(m_n) \\ h_6(m_1) & \dots & h_6(m_n) \\ -(h_1+h_2+h_4+h_5)(m_1) & \dots & -(h_1+h_2+h_4+h_5)(m_n) \\ -(h_3+h_6)(m_1) & \dots & -(h_3+h_6)(m_n) \end{pmatrix},$$

where $m_1, \dots m_n \in \mathscr{A}$ and $h_1, \dots, h_6$, are as above.

Theorem 3.3 *For all pairs in $\mathcal{P}$, the positive part of the minimal Horn matrix is the lattice distance matrix of $\mathscr{A}$. For the subfamilies of $\mathcal{P}$ in Table 2, the matrix $M_{\mathscr{A},\Sigma_P}$ gives rise to a Horn pair for (P, w).*

Proof Let $(P, w) \in \mathcal{P}$, if $M_{\mathscr{A},\Sigma_P}$ is a Horn matrix, then after row reduction operations, we get a minimal Horn matrix. Comparing this matrix with the minimal Horn matrix associated with (P, w) in Table 3 (Appendix A) and by uniqueness of minimal Horn matrices, one can verify both statements on the theorem. □

Table 2 Subfamilies of prismatoids for which there exists a Horn pair (H, λ) with $H = M_{\mathscr{A},\Sigma_P}$

Name of subfamily	Constraints on $a' \leq a, b' \leq b, d$
Trapezoidal wedges	$a' > 0,\ b' = 0,\ b = 1,\ d > 0$
Tensor product wedges ($a' = 0$)	$a' = 0,\ a = 1,\ b' > 0,\ d = 0$
Tensor product wedges ($b' = 0$)	$a' > 0,\ b' = 0,\ b = 1,\ d = 0$
3D simplex	$a' = a = b' = 0,\ b = 1,\ d = 1$
Triangular frusta	$a' = a = 0,\ b > 0,\ d = 1$
Tensor product frusta	$a' > 0,\ b' > 0,\ d = 0$
Trapezoidal frusta	$a' > 0,\ b' > 0,\ d > 0$

4 Multinomial Staged Tree Models

In this section, we define multinomial staged tree models, we prove that every such model has rational MLE and we give criteria to determine when such models are toric varieties for binary multinomial staged trees, see Theorem 4.4 and Theorem 4.3, respectively. To each toric binary multinomial staged tree, one can associate a polytope, by Theorem 2.3 such a polytope has rational linear precision. These results imply our main theorem:

Theorem 4.1 *Polytopes of toric binary multinomial staged trees have rational linear precision.*

Our motivation to introduce this model class arose from the observation that the Horn pairs of all 2D and 3D polytopes in Sect. 3 could be interpreted as a statistical model defined by an event tree with a specific choice of parametrisation. Multinomial staged tree models improve the understanding of polytopes with rational linear precision in $2D$ and $3D$. They also offer a generalisation for polytopes with rational linear precision in higher dimensions.

4.1 Definition of Multinomial Staged Trees

We start by introducing the multinomial model as an event tree. This model is the building block of multinomial staged tree models. Throughout this section m denotes a positive integer and $[m]{:=}\{1, 2, \ldots, m\}$, this differs from Sect. 3 where m was used for lattice point.

Example 4.1 The multinomial model encodes the experiment of rolling a q-sided die n independent times and recording the side that came up each time. The outcome space for this model is the set Ω of all tuples $K = (k_1, \cdots, k_q) \in \mathbb{N}^q$ whose entries sum to n. We can depict this model by a rooted tree $\mathcal{T} = (V, E)$ with vertices $V = \{r\} \cup \{r(K) : K \in \Omega\}$ and edges $E = \{r \to r(K) : K \in \Omega\}$. To keep track of the probability of each outcome we can further label $\mathcal{T}$ with monomials on the set of symbols $\{s_1, \ldots, s_q\}$. Each symbol s_i represents the probability that the die shows side i when rolled once. The monomial representing the probability of outcome K is the term with vector of exponents K in the multinomial expansion of

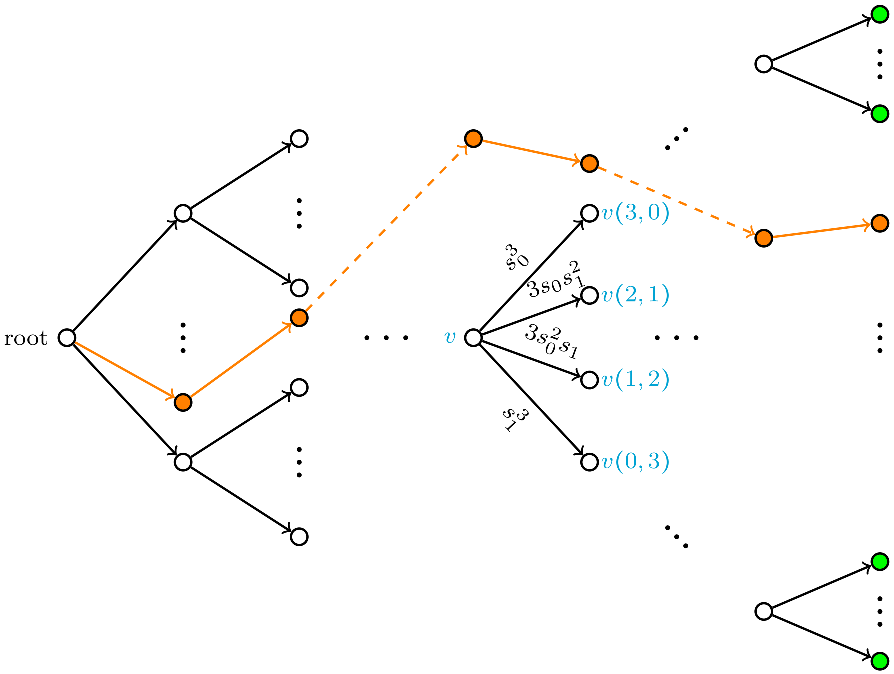


Fig. 3 General sketch of a multinomial staged tree. The vertex v is labelled by the floret of degree 3 on S_l, denoted by $f_{l,3}$. The green vertices are the leaves and a root-to-leaf path is shown in orange

$(s_1 + \ldots + s_q)^n$, namely $\binom{n}{K} \prod_{i=1}^n s_i^{k_i}$, where $\binom{n}{K} := \binom{n}{k_1, \cdots, k_m}$. The labelled tree $b\mathcal{T}_{\Delta_2}$ in Fig. 4, represents the multinomial model with $n = b$ and $q = 3$.

In general terms a multinomial staged tree, is a labelled and directed event tree such that at each vertex, the subsequent event is given by a multinomial model as in Example 4.1. To introduce this concept formally, we start with a rooted and directed tree $\mathcal{T} = (V, E)$ with vertex set V and edge set E such that edges are directed away from the root. The directed edge from v to w is denoted $v \to w$, the set of children of a vertex $v \in V$ is $\mathrm{ch}(v) := \{u \in V : v \to u \in E\}$ and the set of outgoing edges from v is $E(v) := \{v \to u : u \in \mathrm{ch}(v)\}$. If $\mathrm{ch}(v) = \emptyset$ then we say that v is a leaf and we let $\widetilde{V}$ denote the set of non-leaf vertices of $\mathcal{T}$.
Given a rooted and directed tree $\mathcal{T}$, we now explain how to label its edges using monomials terms. Figure 3 shows a general sketch of a multinomial staged tree.

Definition 4.1 Fix a set of symbols $S = \{s_i : i \in I\}$ indexed by a set I. Let $I_1, \ldots, I_m$ be a partition of I and $S_1, \ldots, S_m$ the induced partition in the set S.

(1) The sets $S_1, \ldots, S_m$ are called *stages*.
(2) For $a \in \mathbb{Z}_{\geq 1}$ and $\ell \in [m]$, a *floret of degree* a on S_ℓ is the set of terms in the multinomial expansion of the expression $(\sum_{i \in I_\ell} s_i)^a$, we denote this set by $f_{\ell,a}$.
(3) A function $\mathcal{L} : E \to \bigcup_{\ell \in [m], a \in \mathbb{Z}_{\geq 1}} f_{\ell,a}$ is a *labelling* of $\mathcal{T}$ if for every $v \in \widetilde{V}$, $\mathcal{L}(E(v)) = f_{\ell,a}$ for some $\ell \in [m]$, $a \in \mathbb{Z}_{\geq 1}$, and the restriction $\mathcal{L}_v : E(v) \to f_{\ell,a}$ is a bijection.

(4) A multinomial staged tree is a pair $(\mathcal{T}, \mathcal{L})$, where $\mathcal{T}$ is a rooted directed tree and $\mathcal{L}$ is a labelling of $\mathcal{T}$ as in condition (3).

In a multinomial staged tree $(\mathcal{T}, \mathcal{L})$, each $v \in \widetilde{V}$ is associated with the floret $f_{\ell,a}$ that satisfies $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$. In this case, we index the children of v by $v(K)$ where $K = (k_{i_1}, \ldots, k_{i_{|I_\ell|}}) \in \mathbb{N}^{|I_\ell|}$ is a tuple of nonnegative integers that add to a and $i_1, \ldots, i_{|I_\ell|}$ is a fixed ordering of the elements in I_ℓ. It follows that when $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$, then $E(v) = \{v \to v(K) : K \in \mathbb{N}^{|I_\ell|}, |K| = a\}$, where $|K| := \sum_{q=1}^{|I_\ell|} k_{i_q}$. We further assume that the indexing of the children v is compatible with the labelling $\mathcal{L}$, namely for all multinomial staged trees, $\mathcal{L}_v(v \to v(K)) = \binom{a}{K} \prod_{q=1}^{|I_\ell|} s_{i_q}^{k_{i_q}}$, where $\binom{a}{K} = \binom{a}{k_{i_1}, \ldots, k_{i_{|I_\ell|}}}$. It is important to note that this local description of the tree at the vertex v is the multinomial model described in Example 4.1 up to a change of notation. To clarify the notation just introduced we revisit Example 4.1 with a concrete choice of parameters.

Example 4.2 Consider the multinomial model for $q = 2$ and $n = 3$, the outcome space are all possible outcomes of flipping a coin 3 times. Here $S = S_1 = \{s_1, s_2\}$ and the root vertex v will have 4 children, all of which are leaves. The 4 edges of the tree will be labelled by the elements in the floret $f_{1,3} = \{s_1^3, 3s_1^2 s_2, 3s_1 s_2^2, 3s_2^3\}$. The sets of children and outgoing edges of v are then $\mathrm{ch}(v) = \{v(3,0), v(2,1), v(1,2), v(0,3)\}$ and $E(v) = \{v \to v(3,0), v \to v(2,1), v \to v(1,2), v \to v(0,3)\}$.

Remark 4.1 We will always consider a multinomial staged tree $(\mathcal{T}, \mathcal{L})$ as an embedded tree in the plane. This means the tree has a fixed ordering of its edges and vertices. The *level* of a vertex v in $\mathcal{T}$ is the number edges in a path from the root to v. All the trees we consider satisfy the property that two florets associated with two vertices in different levels must be on different stages. This implies that each root-to-leaf path contains at most one monomial term from each floret. Several figures in Sect. 5 contain multinomial staged trees, in these pictures, for simplicity, we omit the coefficients of the monomial edge labels.

Definition 4.2 Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree with index set $I = \sqcup_{\ell \in [m]} I_\ell$. Fix J to be the set of root-to-leaf paths in $\mathcal{T}$, with $|J| = n$. For $j \in J$, define p_j to be the product of all edge labels in the path j. Let c_j be the coefficient of p_j and a_j the exponent vector of the symbols $(s_i)_{i \in I}$ in p_j. With this notation, $p_j = c_j \prod_{i \in I} s_i^{a_{ij}}$, where a_{ij} are the entries of a_j. Define the parameter space

$$\Theta_{\mathcal{T}} := \{(\theta_i)_{i \in I} \in (0,1)^{|I|} : \sum_{i \in I_\ell} \theta_i = 1 \text{ for all } \ell \in [m]\}$$

The *multinomial staged tree model* $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ is the image of the parameterisation

$$\phi_{\mathcal{T}} : \Theta_{\mathcal{T}} \longrightarrow \Delta_{n-1}^{\circ}, \ (\theta_i)_{i \in I} \mapsto (c_j \prod_{i \in I} \theta_i^{a_{ij}})_{j \in J}.$$

Remark 4.2 The sum-to-one conditions on the parameter space $\Theta_{\mathcal{T}}$ imply that the image of $\phi_{\mathcal{T}}$ is contained in Δ°_{n-1}. The multinomial coefficients on the labels of $\mathcal{T}$ are necessary for this condition to hold. The model $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ is an algebraic variety inside Δ°_{n-1} with an explicit parameterisation given by $\phi_{\mathcal{T}}$. For $\theta \in \Theta_{\mathcal{T}}$, eval_{θ} is the evaluation map $s_i \mapsto \theta_i$. The j-th coordinate of $\phi_{\mathcal{T}}$ is $\mathrm{eval}_{\theta}(p_j)$, where $p_j = c_j \prod_{i\in I} s_i^{a_{ij}}$ (Definition 4.2). For this reason we also use p_j to denote the j-th coordinate in the probability simplex Δ°_{n-1}.

Remark 4.3 If all of the florets in a multinomial staged tree have degree one, then it is called a staged tree. Multinomial staged tree models are a generalisation of discrete Bayesian networks [16] and of staged tree models introduced in [17].

Example 4.3 Consider the following experiment with two independent coins: Toss the first coin b times and record the number of tails, say this number is j. Then, toss the second coin $a + d(b - j)$ times, record the number of tails, say it is i. An outcome of this experiment is a pair (i, j) where i is the number of tails in the second sequence of coin tosses and j is the number of tails in the first. This sequence of events may be represented by a multinomial staged tree $(\mathcal{T} = (V, E), \mathcal{L})$ where

$$V = \{r\} \cup \{r(j) : 0 \le j \le b\} \cup \{r(i, j) : 0 \le j \le b, 0 \le i \le a + d(b - j)\} \text{ and}$$
$$E = \{r \to r(j) : 0 \le j \le b\} \cup \{r(j) \to r(i, j) : 0 \le j \le b, 0 \le i \le a + d(b - j)\}.$$

This tree has two stages $S_1 = \{s_0, s_1\}$, $S_2 = \{s_2, s_3\}$ that are a formal representation of the parameters of the Bernoulli distributions of the two independent coins. The set $E(r)$ is labelled by the floret $f_{1,b}$ and the set $E(r(j))$ is labelled by the floret $f_{2,a+d(b-j)}$. Following the conventions set up earlier we see that $\mathcal{L}(r \to r(j)) = \binom{b}{j} s_0^j s_1^{b-j}$ and $\mathcal{L}(r(j) \to r(i, j)) = \binom{a+d(b-j)}{i} s_2^{a+d(b-j)-i} s_3^i$. The multinomial staged tree model $\mathcal{M}_{a,b,d} \subset \Delta_n$ associated with $(\mathcal{T}, \mathcal{L})$, is the statistical model consisting of all probability distributions that follow the experiment just described. Let p_{ij} denote the probability of the outcome (i, j). The model $\mathcal{M}_{a,b,d}$ is parameterised by the map $\phi : \Delta^{\circ}_1 \times \Delta^{\circ}_1 \to \mathcal{M}_{a,b,d}$,

$$(\theta_0, \theta_1) \times (\theta_2, \theta_3) \mapsto \big(p_{ij}\big)_{\substack{0\le j\le b\\ 0\le i\le a+d(b-j)}} \quad \text{where } p_{ij} = \binom{b}{j}\binom{a+d(b-j)}{i} \theta_0^j \theta_1^{b-j} \theta_2^i \theta_3^{a+d(b-j)-i}.$$

This model depends on two independent parameters, thus it has dimension two. The model $\mathcal{M}_{a,b,d}$ is a binary multinomial staged tree model, its tree representation $\mathcal{T}_{a,b,d}$ is displayed in Fig. 4.

Definition 4.3 Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree. Fix the polynomial rings $\mathbb{R}[P_j : j \in J]$, $\mathbb{R}[s_i : i \in I]$ and $\mathbb{R}[s_i : i \in I]/\mathfrak{q}$ where $\mathfrak{q} = \langle 1 - \sum_{i\in I_\ell} s_i : \ell \in [m]\rangle$. We define

$$\Psi^{\text{toric}}_{\mathcal{T}} : \mathbb{R}[P_j : j \in J] \to \mathbb{R}[s_i : i \in I] \text{ by } P_j \mapsto c_j \prod_{i\in I} s_i^{a_{ij}}, \text{ and}$$
$$\Psi_{\mathcal{T}} : \mathbb{R}[P_j : j \in J] \to \mathbb{R}[s_i : i \in I]/\mathfrak{q} \text{ by } \Psi_{\mathcal{T}} = \pi \circ \Psi^{\text{toric}}_{\mathcal{T}}$$

where $\pi : \mathbb{R}[s_i : i \in I] \to \mathbb{R}[s_i : i \in I]/\mathfrak{q}$ is the canonical projection to the quotient ring. The ideal $\ker(\Psi_{\mathcal{T}}^{\text{toric}})$ is the toric ideal associated with $(\mathcal{T}, \mathcal{L})$ and $\ker(\Psi_{\mathcal{T}})$ is the model ideal associated with $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$. Whenever $\ker(\Psi_{\mathcal{T}}) = \ker(\Psi_{\mathcal{T}}^{\text{toric}})$, we call $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ a toric model.

Remark 4.4 The ideal $\ker(\Psi_{\mathcal{T}})$ defines the model $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ implicitly, i.e. $\mathcal{M}_{(\mathcal{T},\mathcal{L})} = V(\ker(\Psi_{\mathcal{T}})) \cap \Delta_n^{\circ}$. Because of the containment $\ker(\Psi_{\mathcal{T}}^{\text{toric}}) \subset \ker(\Psi_{\mathcal{T}})$, $V(\ker(\Psi_{\mathcal{T}}^{\text{toric}}))$ is a toric variety that contains $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$. The polynomial $1 - \sum_{j \in J} P_j$ is always an element in $\ker(\Psi_{\mathcal{T}})$, hence using this polynomial as a homogenising element, we shall always consider $\ker(\Psi_{\mathcal{T}})$ as a homogeneous ideal in $\mathbb{R}[P_j : j \in J]$.

4.2 The Ideal of Model Invariants for $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$

As is common in algebraic geometry, finding the explicit equations of the prime ideal $\ker(\Psi_{\mathcal{T}})$ is hard. Luckily, the statistical insight of the problem allows us to find a nonprime ideal, usually referred to as the ideal of model invariants, that defines the model inside the probability simplex. We now define this ideal and postpone the proof that it has the aforementioned property to Sect. 4.4.

Definition 4.4 Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree. For a vertex $v \in V$, define $[v]:=\{j \in J : \text{ the path } j \text{ goes through the vertex } v\}$ and set $P_{[v]}:=\sum_{j \in [v]} P_j$.

$$
\begin{aligned}
I_{\text{stages}} := \langle & b P_{[w]} \left(\sum_{|K|=a, k_{i_q} \geq 1} k_{i_q} P_{[v(K)]} \right) - a P_{[v]} \left(\sum_{|K'|=b, k'_{i_q} \geq 1} k'_{i_q} P_{[w(K')]} \right) : \\
& v \sim w, \operatorname{im}(\mathcal{L}_v) = f_{\ell,a}, \operatorname{im}(\mathcal{L}_w) = f_{\ell,b}, \ell \in [m], 1 \leq q \leq |I_\ell| \rangle, \text{ and} \\
I_{\text{vertices}} := \langle & C_{(K^3,K^4)} P_{[v(K^1)]} P_{[v(K^2)]} - C_{(K^1,K^2)} P_{[v(K^3)]} p_{[v(K^4)]} : v \in \widetilde{V}, \\
& \operatorname{im}(\mathcal{L}_v) = f_{\ell,a}, K^1, K^2, K^3, K^4 \in \mathbb{N}^{|I_\ell|}, |K^1| = |K^2| = |K^3| = |K^4| = a, \\
& K^1 + K^2 = K^3 + K^4, C_{(K^i,K^j)} = \binom{a}{K^i}\binom{a}{K^j}, \ i = 1, 3, \ \ j = 2, 4 \rangle.
\end{aligned}
$$

The *ideal of model invariants of* $(\mathcal{T}, \mathcal{L})$ is $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} := I_{\text{stages}} + I_{\text{vertices}} + \langle 1 - \sum_{j \in J} P_j \rangle$.

The previous definition indicates, that there are equations that must hold for every pair of vertices with the same associated stage, and equations that must hold for every vertex. The motivation for this definition of the ideal of model invariants arises from the technical Lemma 7.1 in Appendix B.

Remark 4.5 The generators of I_{vertices} for each fixed vertex v are similar to the Veronese relations of the embedding $\nu_a : \mathbb{P}^{|I_\ell|-1} \to \mathbb{P}^M$ by monomials of total degree a. The only difference is in the coefficients, defined in Lemma 7.1 part (2), that are needed for cancellation.

Remark 4.6 By definition, $I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$ always contains the sum-to-one condition $1 - \sum_{j \in J} P_j$, thus in a similar way as for $\ker(\Psi_{\mathcal{T}})$ in Remark 4.4, we always consider $I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$ as a homogeneous ideal generated by I_{stages} and I_{vertices}.

4.3 Algebraic Lemmas for Multinomial Staged trees

To understand the defining equations of $\ker(\Psi_{\mathcal{T}})$ and the case when this ideal is toric, it is important to establish several lemmas that describe algebraic relations that hold in $\mathbb{R}[P_j : j \in J]$, $\mathbb{R}[s_i : i \in I]$, $\mathbb{R}[P_j : j \in J]/\ker(\Psi_{\mathcal{T}})$ and $\mathbb{R}[P_j : j \in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$. The reader may decide to skip this section and only get back to it when the lemmas are used in the proofs of Theorems 4.2 and 4.3.

Definition 4.5 Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree with $\mathcal{T} = (V, E)$. For $v \in V$, let Λ_v denote the set of all v-to-leaf paths in $\mathcal{T}$. A path $\lambda \in \Lambda_v$ is a sequence of edges $v \to v_1 \to \cdots \to v_\alpha$ where v_α is a leaf of $\mathcal{T}$. For each $v \in V$ we define the *interpolating polynomial of* v, $t(v) \in \mathbb{R}[s_i : i \in I]$, by

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

If v is a leaf, $t(v):=1$. We denote by $\overline{t(v)}$, the image of $t(v)$ under the canonical projection to $\mathbb{R}[s_i : i \in I]/\mathfrak{q}$. Note that for all $v \in V$, $\overline{t(v)} = 1$.

Lemma 4.1 *Let* $(\mathcal{T}, \mathcal{L})$ *be a multinomial staged tree where* $\mathcal{T} = (V, E)$ *and let* $v \in \widetilde{V}$ *be such that* $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$.

(1) *The polynomial* $t(v)$ *satisfies*

$$t(v) = \sum_{|K|=a} \binom{a}{K} \prod_{i \in I_\ell} s_i^{k_i} \cdot t(v(K));$$

(2) *The image of* $P_{[v]}$ *under* $\Psi_{\mathcal{T}}^{\mathrm{toric}}$ *is* $\left(\prod_{e \in \lambda_{r,v}} \mathcal{L}(e)\right) \cdot t(v)$, *where* $\lambda_{r,v}$ *is the set of edges in the root-to-*v *path in* $\mathcal{T}$. *Moreover* $\Psi_{\mathcal{T}}(P_{[v]}) = \prod_{e \in \lambda_{r,v}} \mathcal{L}(e)$.

Proof (1) Any path in Λ_v goes through a child $v(K)$ of v. The sum of all the edge products corresponding to the paths that go through child $v(K)$ is equal to the sum of all the edge products corresponding to the paths starting at $v(K)$ ($t(v(K))$) multiplied by the label of the edge from v to $v(K)$ ($\binom{a}{K} \prod_{i \in I_\ell} s_i^{k_i}$). Taking the sum of this expression over all children of v gives the desired result.
(2) Let j be a root-to-leaf path that goes through v. Then, j is the concatenation of a path from the root to v, denoted by $\lambda_{r,v}$ and a path from v to the leaf denoted by $\lambda_{v,j}$. Then,

$$\Psi_{\mathcal{T}}^{\mathrm{toric}}(P_{[v]}) = \sum_{j \in [v]} \prod_{e \in \lambda_{r,v}} \mathcal{L}(e) \prod_{e \in \lambda_{v,j}} \mathcal{L}(e) = \left(\prod_{e \in \lambda_{r,v}} \mathcal{L}(e)\right) \left(\sum_{j \in [v]} \prod_{e \in \lambda_{v,j}} \mathcal{L}(e)\right)$$
$$= \left(\prod_{e \in \lambda_{r,v}} \mathcal{L}(e)\right) t(v).$$

The second statement follows by noting that $\overline{t(v)} = 1$. □

Lemma 4.2 *Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree, there is a containment of ideals $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} \subset \ker(\Psi_{\mathcal{T}})$ in $\mathbb{R}[P_j : j \in J]$.*

Proof To show that $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} \subset \ker(\Psi_{\mathcal{T}})$, it suffices to show that the generators of I_{stages} and I_{vertices} are zero after applying $\Psi_{\mathcal{T}}$. We present the proof for the generators of I_{stages}, the proof for I_{vertices} is similar and also uses Lemma 4.1. A generator of $I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$ is of the form

$$bP_{[w]}\left(\sum_{|K|=a, k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}\right) - aP_{[v]}\left(\sum_{|K'|=b, k'_{i_q}\geq 1} k'_{i_q} P_{[w(K')]}\right),$$

where $v, w \in \widetilde{V}$, $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$, $\mathrm{im}(\mathcal{L}_w) = f_{\ell,b}$ for some $\ell \in [m]$ and a fixed q, $1 \leq q \leq |I_\ell|$.
Claim : $\Psi_{\mathcal{T}}(\sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q} P_{[v(K)]}) = as_{i_q} \prod_{e\in\lambda_{r,v}} \mathcal{L}(e)$.
Using Lemma 4.1, we compute $\Psi_{\mathcal{T}}(P_{[v(K)]})$.

$$\begin{aligned}
\Psi_{\mathcal{T}}(\sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q} P_{[v(K)]}) &= \sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q} \Psi_{\mathcal{T}}(P_{[v(K)]}) = \sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q}\left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right)\binom{a}{K}\prod_{\alpha=1}^{|I_\ell|} s_i^{k_{i_\alpha}} \\
&= \left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right)\left(\sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q}\binom{a}{K}\prod_{\alpha=1}^{|I_\ell|} s_i^{k_{i_\alpha}}\right) \\
&= \left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right)\left(\sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} k_{i_q}\frac{a(a-1)!}{k_{i_1}!\cdots k_{i_{|I_\ell|}}!} s_{i_q} s_{i_q}^{k_q-1}\prod_{\substack{\alpha=1\\ \alpha\neq q}}^{|I_\ell|} s_i^{k_{i_\alpha}}\right) \\
&= \left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right)\left(\sum_{|K|=a-1} as_{i_q}\binom{a-1}{K}\prod_{\alpha=1}^{|I_\ell|} s_i^{k_{i_\alpha}}\right) \\
&= \left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right) as_{i_q}(\sum_{\alpha=1}^{|I_\ell|} s_{i_\alpha})^{a-1} = \left(\prod_{e\in\lambda_{r,v}} \mathcal{L}(e)\right) as_{i_q}.
\end{aligned}$$

The last equality follows from the fact that $\sum_{\alpha=1}^{|I_\ell|} s_{i_\alpha} = 1$ in $\mathbb{R}[s_i : i \in I]/\mathfrak{q}$. The claim applied to $w \in \widetilde{V}$, implies $\Psi_{\mathcal{T}}(\sum_{\substack{|K|=b\\ k_{i_q}\geq 1}} k_{i_q} P_{[w(K)]}) = \left(\prod_{e\in\lambda_{r,w}} \mathcal{L}(e)\right) bs_{i_q}$. Thus, by

Lemma 4.1,

$$\Psi_{\mathcal{T}}\left(bP_{[w]}\left(\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}\right) - aP_{[v]}\left(\sum_{|K'|=b,k'_{i_q}\geq 1} k'_{i_q} P_{[w(K')]}\right)\right) =$$
$$b\prod_{e\in\lambda_{r,w}}\mathcal{L}(e)\cdot as_{i_q}\prod_{e\in\lambda_{r,v}}\mathcal{L}(e) - a\prod_{e\in\lambda_{r,v}}\mathcal{L}(e)\cdot bs_{i_q}\prod_{e\in\lambda_{r,w}}\mathcal{L}(e) = 0.$$

□

4.4 Defining Equations of Binary Multinomial Staged Trees

In this section and the next, we prove Theorems 4.2 and 4.3 for binary multinomial staged trees; despite being unable to provide a proof, we believe these statements also hold for non-binary multinomial staged trees. First we show that the ring homomorphism $\Psi_{\mathcal{T}}$ admits an inverse when localised at a suitable element. From this, it follows as a corollary that the ideal of model invariants defines $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ inside the probability simplex.

Theorem 4.2 *Let $(\mathcal{T},\mathcal{L})$ be a binary multinomial staged tree and define $\mathbf{P}:=\prod_{v\in V} P_{[v]}$. Then, the localised map*

$$(\Psi_{\mathcal{T}})_{\mathbf{P}} : \left(\mathbb{R}[P_j : j\in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}\right)_{\mathbf{P}} \to (\mathbb{R}[s_i : i\in I]/\mathfrak{q})_{\Psi_{\mathcal{T}}(\mathbf{P})},$$

is an isomorphism of $\mathbb{R}$-algebras. Therefore $(I_{\mathcal{M}(\mathcal{T},\mathcal{L})})_{\mathbf{P}} = (\ker(\Psi_{\mathcal{T}}))_{\mathbf{P}}$ and thus $(I_{\mathcal{M}(\mathcal{T},\mathcal{L})} : \mathbf{P}^{\infty}) = \ker(\Psi_{\mathcal{T}})$.

Proof We define a ring homomorphism

$$\varphi : (\mathbb{R}[s_i : i\in I]/\mathfrak{q})_{\Psi_{\mathcal{T}}(\mathbf{P})} \to \left(\mathbb{R}[P_j : j\in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}\right)_{\mathbf{P}}$$

and show that it is a two sided inverse for $(\Psi_{\mathcal{T}})_{\mathbf{P}}$. For $\ell\in[m]$ and $1\leq q\leq |I_\ell|$, let v be a vertex with $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$ and define

$$\varphi(s_{i_q}) = \frac{\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}}{aP_{[v]}}.$$

Note that φ is well defined: If w is another vertex with $\mathrm{im}(\mathcal{L}_w) = f_{\ell,b}$, then

$$\varphi(s_{i_q}) = \frac{\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}}{aP_{[v]}} = \frac{\sum_{|K'|=b,k'_{i_q}\geq 1} k_{i_q} P_{[w(K)]}}{bP_{[w]}}$$

because $bP_{[w]}\left(\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}\right) - aP_{[v]}\left(\sum_{|K'|=b,k'_{i_q}\geq 1} k'_{i_q} P_{[w(K')]}\right) \in I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$.
First, we check that $(\Psi_{\mathcal{T}})_{\mathbf{P}} \circ \varphi = \mathrm{Id}$,

$$
\begin{aligned}
(\Psi_{\mathcal{T}})_{\mathbf{P}}(\varphi(s_{i_q})) &= (\Psi_{\mathcal{T}})_{\mathbf{P}}\left(\frac{\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} P_{[v(K)]}}{aP_{[v]}}\right) \\
&= \sum_{|K|=a,k_{i_q}\geq 1} \frac{k_{i_q}}{a} (\Psi_{\mathcal{T}})_{\mathbf{P}}\left(\frac{P_{[v(K)]}}{P_{[v]}}\right) \\
&= \sum_{|K|=a,k_{i_q}\geq 1} \frac{k_{i_q}}{a} \frac{\Psi_{\mathcal{T}}(P_{[v(K)]})}{\Psi_{\mathcal{T}}(P_{[v]})} = \sum_{|K|=a,k_{i_q}\geq 1} \frac{k_{i_q}}{a} \binom{a}{K} \prod_{\alpha=1}^{|I_\ell|} s_{i_\alpha}^{k_{i_\alpha}} = s_{i_q}.
\end{aligned}
$$

The second to last equality follows by using the expression for $\Psi_{\mathcal{T}}(P_{[v]})$ presented in Lemma 4.1 part (2), the same result is used to compute $\Psi_{\mathcal{T}}(P_{[v(K)]})$, finally their quotient $\Psi_{\mathcal{T}}(P_{[v(K)]})/\Psi_{\mathcal{T}}(P_{[v]})$ is exactly $\binom{a}{K}\prod_{\alpha=1}^{|I_\ell|} s_{i_\alpha}^{k_{i_\alpha}}$. The last equality is obtained by using the same argument as in Lemma 7.1 part (3).
Next, we verify $\varphi \circ (\Psi_{\mathcal{T}})_{\mathbf{P}} = \mathrm{Id}$, which amounts to proving that $(\varphi \circ \Psi_{\mathcal{T},\mathbf{P}})(P_j) = P_j$ for each $j \in J$. From this point on we further assume that $|I_\ell| = 2$ for all $\ell \in [m]$. Fix $j \in J$ and let $v_1 \to v_2 \to \cdots \to v_\alpha$ be the root-to-leaf path j. By Definition 4.3

$$
(\Psi_{\mathcal{T}})_{\mathbf{P}}(P_j) = c_j \prod_{i\in I} s_i^{a_{ij}} = \mathcal{L}_{v_1}(v_1 \to v_2)\cdots\mathcal{L}_{v_{\alpha-1}}(v_{\alpha-1} \to v_\alpha), \tag{5}
$$

where for each $\gamma \in [\alpha-1]$, $\mathrm{im}(\mathcal{L}_{v_\gamma}) = f_{\ell_\gamma,a_\gamma}$ for some $\ell_\gamma \in [m]$ and $a_\gamma \in \mathbb{Z}_{\geq 1}$. By Remark 4.1, none of the florets f_{ℓ_γ,a_γ} share the same set of symbols. Moreover, for each $\gamma \in [\alpha-1]$,

$$
\mathcal{L}_{v_\gamma}(v_\gamma \to v_{\gamma+1}) = \binom{a_\gamma}{k_\gamma} s_{\gamma,i_1}^{k_\gamma} s_{\gamma,i_2}^{a-k_\gamma} \text{ where } S_{\ell_\gamma} = \{s_{\gamma,i_1}, s_{\gamma,i_2}\}, 0 \leq k_\gamma \leq a_\gamma. \tag{6}
$$

With this notation, we also deduce that $v_{\gamma+1} = v_\gamma(k_\gamma, a-k_\gamma)$. Now we apply φ to (5), use that φ is a ring homomorphism and use equation (6) to obtain

$$
\begin{aligned}
\varphi((\Psi_{\mathcal{T}})_{\mathbf{P}}(P_j)) &= \varphi(\mathcal{L}_{v_1}(v_1 \to v_2))\cdots\varphi(\mathcal{L}_{v_{\alpha-1}}(v_{\alpha-1} \to v_\alpha)) \\
&= \prod_{\gamma=1}^{\alpha-1} \binom{a_\gamma}{k_\gamma} \varphi(s_{\gamma,i_1})^{k_\gamma} \varphi(s_{\gamma,i_2})^{a-k_\gamma}.
\end{aligned} \tag{7}
$$

Using the definition of φ, for each $\gamma \in [\alpha-1]$,

$$
\varphi(s_{\gamma,i_1}) = \frac{\sum_{k=1}^{a_\gamma} kP_{[v_\gamma(k,a_\gamma-k)]}}{a_\gamma P_{[v_\gamma]}} \quad \text{and} \quad \varphi(s_{\gamma,i_2}) = \frac{\sum_{k=1}^{a_\gamma} kP_{[v_\gamma(a_\gamma-k,k)]}}{a_\gamma P_{[v_\gamma]}}.
$$

By Lemma 7.4, with $a = a_\gamma$, $l_1 = \sum_{k=1}^{a_\gamma} kP_{[v_\gamma(k,a_\gamma-k)]}$, $l_2 = \sum_{k=1}^{a_\gamma} kP_{[v_\gamma(a_\gamma-k,k)]}$ and $k_0 = k_\gamma$, we conclude that

$$\binom{a_\gamma}{k_\gamma}\varphi(s_{\gamma,i_1})^{k_\gamma}\varphi(s_{\gamma,i_1})^{a-k_\gamma} = \frac{P_{[v_\gamma(k_\gamma,a-k_\gamma)]}}{P_{[v_\gamma]}}.$$

Thus, continuing from (7) we have

$$\varphi((\Psi_\mathcal{T})_\mathbf{P}(P_j)) = \prod_{\gamma=1}^{\alpha-1}\frac{P_{[v_\gamma(k_\gamma,a-k_\gamma])]}}{P_{[v_\gamma]}}$$
$$= \frac{P_{[v_1(k_1,a_1-k_1)]}}{P_{[v_1]}}\frac{P_{[v_2(k_2,a_2-k_2)]}}{P_{[v_2]}}\cdots\frac{P_{[v_{\alpha-1}(k_{\alpha-1},a_{\alpha-1}-k_\alpha)]}}{P_{[v_{\alpha-1}]}} = \frac{P_{[v_\alpha]}}{P_{[v_1]}}.$$

To obtain the previous cancellation we used the fact that for each $\gamma \in [\alpha-1]$, $v_{\gamma+1} = v_\gamma(k_\gamma, a-k_\gamma)$, hence $P_{[v_\gamma(k_\gamma,a-k_\gamma)]} = P_{[v_{\gamma+1}]}$. Note that $P_{[v_1]} = 1$ by definition of $I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$ and $P_{[v_\alpha]} = P_j$ because v_α is the last vertex in the path j. Thus, $\varphi((\Psi_\mathcal{T})_\mathbf{P}(P_j)) = P_j$. The second statement of the theorem follows from the fact that $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} \subset \ker(\Psi_\mathcal{T})$ and that the localisation $(\Psi_\mathcal{T})_\mathbf{P}$ is an isomorphism. □

Corollary 4.1 *The ideal of model invariants defines the binary multinomial staged tree model inside the probability simplex, i.e.* $\mathcal{M}_{(\mathcal{T},\mathcal{L})} = V(I_{\mathcal{M}(\mathcal{T},\mathcal{L})}) \cap \Delta_{n-1}^\circ$.

Proof The variety $V(I_{\mathcal{M}(\mathcal{T},\mathcal{L})} : \mathbf{P}^\infty)$ exactly describes the points in $V(I_{\mathcal{M}(\mathcal{T},\mathcal{L})})$ that are not in $V(\mathbf{P})$. The latter variety contains the boundary of the simplex, hence restricting to positive points that add to one, yields $\mathcal{M}_{(\mathcal{T},\mathcal{L})} = V(I_{\mathcal{M}(\mathcal{T},\mathcal{L})}) \cap \Delta_{n-1}^\circ$. □

4.5 Toric Binary Multinomial Staged Tree Models

It is not true in general the ideal $\ker(\Psi_\mathcal{T})$ of a multinomial staged tree is toric. For the case of staged trees, a characterisation of when $\ker(\Psi_\mathcal{T})$ is equal to a subideal generated by binomials is available in [11]. The goal of this section is to establish a similar criterion, based on interpolating polynomials from Definition 4.5, for multinomial staged trees. This criterion will allow us to study the polyhedral geometry of these models in Sect. 5.

Definition 4.6 Let $(\mathcal{T}, \mathcal{L})$ be a multinomial staged tree and let v, w be two vertices in the same stage with $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$ and $\mathrm{im}(\mathcal{L}_w) = f_{\ell,b}$ for some $\ell \in [m]$.

(1) The vertex v is balanced if for all $K^1, K^2, K^3, K^4 \in \mathbb{N}^{|I_\ell|}$ with $|K^1| = |K^2| = |K^3| = |K^4| = a$ and $K^1+K^2 = K^3+K^4$, the next identity holds in $\mathbb{R}[s_i : i \in I]$

$$t(v(K^1))t(v(K^2)) = t(v(K^3))t(v(K^4)).$$

(2) The pair of vertices v, w is balanced if for all tuples $K, K', Q, Q' \in \mathbb{N}^{|I_\ell|}$ with $|K| = |K'| = a$ and $|Q| = |Q'| = b$ with $K + Q' = K' + Q$ the following

identity holds in $\mathbb{R}[s_i : i \in I]$

$$t(v(K)) \cdot t(w(Q')) = t(v(K')) \cdot t(w(Q)).$$

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

Remark 4.7 Condition (1) in Definition 4.6 is an empty condition for florets of degree one. For staged trees, condition (2) specialises to the definition of balanced stated in [2].

Remark 4.8 If all root-to-leaf paths in $(\mathcal{T}, \mathcal{L})$ have length 1, then $(\mathcal{T}, \mathcal{L})$ is vacuously balanced. If $(\mathcal{T}, \mathcal{L})$ has all root-to-leaf paths of length 2, such as $\mathcal{T}_{a,b,d}$ in Fig. 5, it suffices to check that the root is balanced. For the other vertices, the conditions in Definition 4.6 reduce to the trivial equality $1 \cdot 1 = 1 \cdot 1$.

Theorem 4.3 *Let $(\mathcal{T}, \mathcal{L})$ be a binary multinomial staged tree. The model $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ is toric if and only if $(\mathcal{T}, \mathcal{L})$ is balanced.*

Proof We prove that $\ker(\Psi_{\mathcal{T}}) = \ker(\Psi_{\mathcal{T}}^{\text{toric}})$ if and only if $(\mathcal{T}, \mathcal{L})$ is balanced. Define the ideal J to be generated by all polynomials of the form

$$C_{(K',Q)} P_{[v(K)]} P_{[w(Q')]} - C_{(K,Q')} P_{[v(K')]} P_{[w(Q)]},$$
$$C_{(K^3,K^4)} P_{[v(K^1)]} P_{[v(K^2)]} - C_{(K^1,K^2)} P_{[v(K^3)]} P_{[v(K^4)]}$$

where $v, w \in V$ are in the same stage and K, K', Q, Q' obey the condition (2) and K^1, K^2, K^3, K^4 obey the condition (1) in Definition 4.6.
Claim 1: $J \subset \ker(\Psi_{\mathcal{T}}^{\text{toric}})$ if and only if $(\mathcal{T}, \mathcal{L})$ is balanced.
By Lemma 4.1,

$$\Psi_{\mathcal{T}}^{\text{toric}}(P_{[v(K)]} P_{[w(Q')]}) = \left(\prod_{e \in \lambda_{r,v(K)}} \mathcal{L}(e)\right) t(v(K)) \left(\prod_{e \in \lambda_{r,w(Q')}} \mathcal{L}(e)\right) t(w(Q')) \text{ and} \tag{8}$$

$$\Psi_{\mathcal{T}}^{\text{toric}}(P_{[v(K')]} P_{[w(Q)]}) = \left(\prod_{e \in \lambda_{r,v(K')}} \mathcal{L}(e)\right) t(v(K')) \left(\prod_{e \in \lambda_{r,w(Q)}} \mathcal{L}(e)\right) t(w(Q)). \tag{9}$$

Note that the right-hand side of the two equations above share the common factor $\prod_{e \in \lambda_{v,w}} \mathcal{L}(e)$ where $\lambda_{v,w}$ is the set of edges in the path from v to w. Thus, we extract this factor from the two previous equations and multiply times the labels of the edges $v \to v(K), w \to w(Q)$ and $v \to v(K'), w \to w(Q')$, respectively, to further simplify the two expressions into

$$\Psi_{\mathcal{T}}^{\text{toric}}(P_{[v(K)]} P_{[w(Q')]}) = \prod_{e \in \lambda_{v,w}} \mathcal{L}(e) \left(\binom{a}{K}\binom{b}{Q'} \prod_{i \in I_\ell} s_i^{k_i + q_i'}\right) t(v(K)) t(w(Q')) \text{ and} \tag{10}$$

$$\Psi_{\mathcal{T}}^{\text{toric}}(P_{[v(K')]}P_{[w(Q)]}) = \prod_{e\in\lambda_{v,w}} \mathcal{L}(e) \left(\binom{a}{K'}\binom{b}{Q} \prod_{i\in I_\ell} s_i^{k_i'+q_i} \right) t(v(K'))t(w(Q)) \tag{11}$$

Finally, since $K + Q' = K' + Q$ and $(\mathcal{T}, \mathcal{L})$ is balanced, we obtain

$$\begin{aligned} &\Psi_{\mathcal{T}}^{\text{toric}}(C_{(K',Q)}P_{[v(K)]}P_{[w(Q')]} - C_{(K,Q')}P_{[v(K')]}P_{[w(Q)]}) \\ &\quad = C_{(K,Q')}C_{(K',Q)} \prod_{e\in\lambda_{v,w}} \mathcal{L}(e) \prod_{i\in I_\ell} s_i^{k_i+q_i'} \big(t(v(K))t(w(Q)) - t(v(K'))t(w(Q))\big) \\ &\quad = 0 \end{aligned}$$

A similar calculation shows that

$$\Psi_{\mathcal{T}}^{\text{toric}}(C_{(K^3,K^4)}P_{[v(K^1)]}P_{[v(K^2)]} - C_{(K^1,K^2)}P_{[v(K^3)]}P_{[v(K^4)]}) = 0$$

if $\mathcal{T}$ is balanced. Conversely, note that if $J \subset \ker(\Psi_{\mathcal{T}}^{\text{toric}})$, tracing these equations backwards implies that $\mathcal{T}$ must be balanced.

Claim 2: $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} \subset J$. The ideal $I_{\mathcal{M}(\mathcal{T},\mathcal{L}))}$ is the sum of I_{stages} and I_{vertices}. By definition, the generators of I_{vertices} are also generators of J. Hence, it suffices to show that the generators of I_{stages} are polynomial combinations of the generators of J. From this point on we further assume that $(\mathcal{T}, \mathcal{L})$ is binary. Suppose v, w are in the same stage, where $\text{im}(\mathcal{L}_v) = f_{\ell,a}$, $\text{im}(\mathcal{L}_w) = f_{\ell,b}$ and $|I_l| = 2$. There are two equations that hold for this stage, one for each element in I_ℓ. We will show that the equation

$$bP_{[w]}\left(\sum_{k_1=1}^{a} k_1 P_{[v(k_1,a-k_1)]}\right) - aP_{[v]}\left(\sum_{k_2=1}^{b} k_2 P_{[w(k_2,b-k_2)]}\right), \tag{12}$$

which is the equation for the first element in I_ℓ, is a combination the generators of J, defined at the beginning. The one for the second element in I_ℓ follows an analogous argument. We use the following two identities:

$$\begin{aligned} bP_{[w]} &= \sum_{k_2=1}^{b} k_2 P_{[w(k_2,b-k_2)]} + \sum_{k_2=1}^{b} k_2 P_{w[(b-k_2,k_2)]} \text{ and} \\ aP_{[v]} &= \sum_{k_1=1}^{a} k_1 P_{[v(k_1,a-k_1)]} + \sum_{k_1=1}^{a} k_1 P_{[v(a-k_1,k_1)]}. \end{aligned}$$

Working from equation (12), using the identities, we have

$$\left(\sum_{k_2=1}^{b} k_2 P_{[w(k_2,b-k_2)]} + \sum_{k_2=1}^{b} k_2 P_{[w(b-k_2,k_2)]}\right)\left(\sum_{k_1=1}^{a} k_1 P_{[v(k_1,a-k_1)]}\right)$$

$$
\begin{aligned}
&\quad - \left(\sum_{k_1=1}^{a} k_1 P_{[v(k_1,a-k_1)]} + \sum_{k_1=1}^{a} k_1 P_{[v(a-k_1,k_1)]} \right) \left(\sum_{k_2=1}^{b} k_2 P_{[w(k_2,b-k_2)]} \right) \\
&= \left(\sum_{k_2=1}^{b} k_2 P_{[w(b-k_2,k_2)]} \right) \left(\sum_{k_1=1}^{a} k_1 P_{[v(k_1,a-k_1)]} \right) \\
&\quad - \left(\sum_{k_1=1}^{a} k_1 P_{[v(a-k_1,k_1)]} \right) \left(\sum_{k_2=1}^{b} k_2 P_{[w(k_2,b-k_2)]} \right) \\
&= \left(\sum_{k_2=1}^{b} \sum_{k_1=1}^{a} k_2 k_1 P_{[w(b-k_2,k_2)]} P_{[v(k_1,a-k_1)]} \right) \\
&\quad - \left(\sum_{k_2=1}^{b} \sum_{k_1=1}^{a} k_1 k_2 P_{[w(k_2,b-k_2)]} P_{[v(a-k_1,k_1)]} \right) \qquad (13) \\
&= \sum_{k_2=1}^{b} \sum_{k_1=1}^{a} \Big(k_2 k_1 P_{[w(b-k_2,k_2)]} P_{[v(k_1,a-k_1)]} \\
&\quad - (b-(k_2-1))(a-(k_1-1)) P_{[w(b-(k_2-1),k_2-1)]} P_{[v(k_1-1,a-(k_1-1))]} \Big). \qquad (14)
\end{aligned}
$$

After rearranging the terms in (13) we get a single double summation. Finally, the (k_1, k_2) summand in (14) is a multiple of the generator of J, where $Q' = (b - k_2, k_2)$, $K = (k_1, a - k_1)$, $K' = (k_1 - 1, a - (k_1 - 1))$, $Q = (b - (k_2 - 1), k_2 - 1)$. The generator of J corresponding to this choice of K, K', Q, Q' is

$$
\begin{aligned}
&\binom{b}{k_2-1}\binom{a}{k_1-a} P_{[w(b-k_2,k_2)]} P_{[v(k_1,a-k_1)]} \\
&\quad - \binom{b}{k}\binom{a}{k_1} P_{[w(b-(k_2-1),k_2-1)]} P_{[v(k_1,a-(k_1-1))]}. \qquad (15)
\end{aligned}
$$

Note that $(b - k_2, k_2) + (k_1, a - k_1) = (b - (k_2 - 1), k_2 - 1) + (k_1 - 1, a - (k_1 - 1))$, thus $K + Q' = K' + Q$. Multiplying equation (15) times $\frac{(b-(k_2-1))!(a-(k_1-1)!k_1!k_2!)}{a!b!}$ gives the (k_1, k_2) summand in (14). This implies that (14) is a sum of multiples of the generators in J, hence $I_{\text{stages}} \subset J$.

Finally, combining Claim 1 and 2 we conclude that $I_{\mathcal{M}(\mathcal{T},\mathcal{L})} \subset J \subset \ker(\Psi_{\mathcal{T}}^{\text{toric}}) \subset \ker(\Psi_{\mathcal{T}})$ if and only if $\mathcal{T}$ is balanced. We now saturate this chain of ideals as in Theorem 4.2 to obtain $(I_{\mathcal{M}(\mathcal{T},\mathcal{L})} : \mathbf{P}^\infty) = (\ker(\Psi_{\mathcal{T}})^{\text{toric}} : \mathbf{P}^\infty) = (\ker(\Psi_{\mathcal{T}}) : \mathbf{P}^\infty)$. But $(\ker(\Psi_{\mathcal{T}})^{\text{toric}} : \mathbf{P}^\infty) = \ker(\Psi_{\mathcal{T}})^{\text{toric}}$ and $(\ker(\Psi_{\mathcal{T}}) : \mathbf{P}^\infty) = \ker(\Psi_{\mathcal{T}})$ because they are prime ideals. Hence, $\ker(\Psi_{\mathcal{T}}^{\text{toric}})$ and $\ker(\Psi_{\mathcal{T}})$ are equal. □

4.6 Multinomial Staged Tree Models have Rational MLE

In this last section on multinomial staged trees, we prove that they have rational MLE. This fact together with Theorem 4.3 establishes Theorem 4.1 and thus provides a new class of polytopes that have rational linear precision.

Theorem 4.4 *The multinomial staged tree model $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ has rational MLE Φ. The j-th coordinate of Φ is*

$$\Phi_j(u_1,\ldots,u_n) = c_j \prod_{i\in I} \hat{\theta}_i^{a_{ij}}\ , \ \textit{where for } i \in I_\ell, \hat{\theta}_i = \frac{\sum_{j\in J} u_j a_{ij}}{\sum_{i\in I_\ell}(\sum_{j\in J} u_j a_{ij})}.$$

Proof Let $u = (u_1,\ldots,u_n)$ be a vector of counts. The likelihood function for $\mathcal{M}_{\mathcal{T},\mathcal{L}}$ is

$$\begin{aligned} L(p|u) &= \prod_{j\in J} p_j^{u_j} = \prod_{j\in J}\left(\prod_{i\in I} c_j\theta_i^{a_{ij}}\right)^{u_j} \\ &= \prod_{j\in J}\left(\prod_{i\in I} c_j^{u_j}\theta_i^{a_{ij}u_j}\right) = \left(\prod_{i\in I} c_1^{u_1}\theta_i^{a_{i1}u_1}\right)\cdots\left(\prod_{i\in I} c_n^{u_n}\theta_i^{a_{in}u_n}\right) \\ &= \left(\prod_{j\in J} c_j^{u_j}\right)\left(\prod_{i\in I}\theta_i^{\sum_{j\in J} u_j a_{ij}}\right), \text{let } C = \prod_{j\in J} c_j^{u_j} \\ &= C\left(\prod_{i\in I_1}\theta_i^{\sum_{j\in J} u_j a_{ij}}\right)\cdots\left(\prod_{i\in I_k}\theta_i^{\sum_{j\in J} u_j a_{ij}}\right) = CL_1\cdots L_k, \end{aligned}$$

where $L_1,\ldots,L_m$ denote the factors before the last equality in the previous line. The function $L(p|u)$ is maximised when each factor is maximised. This is because the parameters are partitioned by $I_1,\ldots,I_m$ and hence each factor is independent. Thus, we find the maximisers of each factor. The function L_ℓ, $\ell\in[m]$, is the likelihood function of the saturated model $\Delta_{|I_\ell|-1}$ with parameters $(\theta_i)_{i\in I_\ell}$ and vector of counts $\left(\sum_{j\in J} u_j a_{ij}\right)_{i\in I_\ell}$. Therefore $\hat{\theta}_i = \frac{\sum_{j\in J} u_j a_{ij}}{\sum_{i\in I_\ell}\sum_{j\in J} u_j a_{ij}}$, $i \in I_\ell$. □

Corollary 4.2 *Let $H_{(\mathcal{T},\mathcal{L})}$ be the $(|I|+m)\times(|J|)$ matrix with entries*

$$h_{ij} = a_{ij},\quad i\in I,\quad h_{\ell j} = -\sum_{i\in I_\ell} a_{ij},\quad \ell\in[m],\ \textit{and}\ \lambda_j{:=}(-1)^{\sum_{i\in I} a_{ij}} c_j.$$

Then, $(H_{(\mathcal{T},\mathcal{L})},\lambda)$ is a Horn pair for $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$.

Proof It suffices to check that the j-th coordinate of $\varphi_{(H,\lambda)}$ is equal to Φ_j in Theorem 4.4. Let $u = (u_1,\ldots,u_n)$, then

$$(H_{(\mathcal{T},\mathcal{L})}u)^T = \left(\sum_{j\in J} a_{1j}u_j, \ldots, \sum_{j\in J} a_{|I|j}u_j, \ -\sum_{j\in J}(\sum_{i\in I_1} a_{ij})u_j, \ldots, -\sum_{j\in J}(\sum_{i\in I_m} a_{ij})u_j \right).$$

The j-th coordinate of $\varphi_{(H_{(\mathcal{T},\mathcal{L})},\lambda)}$ is

$$\begin{aligned}
\lambda_j(H_{(\mathcal{T},\mathcal{L})}u)^{h_j} &= \frac{(-1)^{\sum_{i\in I} a_{ij}} c_j \left(\sum_{j\in J} a_{1j}u_j\right)^{a_{1j}} \cdots \left(\sum_{j\in J} a_{|I|j}u_j\right)^{a_{|I|j}}}{\left(-\sum_{j\in J}(\sum_{i\in I_1} a_{ij})u_j\right)^{\sum_{i\in I_1} a_{ij}} \cdots \left(-\sum_{j\in J}(\sum_{i\in I_m} a_{ij})\right)^{\sum_{i\in I_k} a_{ij}}} \\
&= c_j \cdot \frac{\left(\sum_{j\in J} a_{1j}u_j\right)^{a_{1j}} \cdots \left(\sum_{j\in J} a_{|I|j}u_j\right)^{a_{|I|j}}}{\left(\sum_{j\in J}(\sum_{i\in I_1} a_{ij})u_j\right)^{\sum_{i\in I_1} a_{ij}} \cdots \left(\sum_{j\in J}(\sum_{i\in I_m} a_{ij})\right)^{\sum_{i\in I_k} a_{ij}}} \\
&= c_j \cdot \left(\frac{\sum_{j\in J} a_{1j}u_j}{\sum_{j\in J}(\sum_{i\in I_1} a_{ij})u_j}\right)^{a_{1j}} \cdots \left(\frac{\sum_{j\in J} a_{|I|j}u_j}{\sum_{j\in J}(\sum_{i\in I_m} a_{ij})u_j}\right)^{a_{|I|j}} \\
&= c_j \prod_{i\in I} \hat{\theta}_i^{a_{ij}} = \Phi_j.
\end{aligned}$$

□

5 Polytopes Arising from Toric Multinomial Staged Trees

The aim of this section is to bring together the examples of 2D and 3D polytopes with rational linear precision (Sect. 3) and multinomial staged trees (Sect. 4). To this end, we investigate certain properties of the lattice polytopes arising from toric multinomial staged trees. This leads to a better understanding of the negative part of the Horn matrix than that provided by the primitive collections. Recall that J denotes the set of root-to-leaf paths in $\mathcal{T}$. For $j \in J$, p_j is defined to be the product of all edge labels in the path j. We denote the stages of $(\mathcal{T}, \mathcal{L})$ by $S_1, \ldots, S_m$. Throughout this section m is a positive integer as in Sect. 4 and m_j (m with a subindex) denotes a lattice point as in Sect. 3.

Definition 5.1 The lattice polytope $P_{\mathcal{T}}$ of a balanced multinomial staged tree $(\mathcal{T}, \mathcal{L})$ is the convex hull of exponent vectors a_j of p_j for every root-to-leaf path j in $\mathcal{T}$.

Note that $P_{\mathcal{T}} \subset \mathbb{R}^d$ is not a full-dimensional polytope for $d = |S_1| + \cdots + |S_m|$. This can be observed, e.g. in Fig. 4 (left) for $P_{\mathcal{T}_{b\Delta_2}} \cong b\Delta_2$ (unimodularly equivalent). We call $(\mathcal{T}, \mathcal{L})$ a *multinomial staged tree representation* of a full-dimensional polytope $P \cong P_{\mathcal{T}}$.

5.1 Two-Dimensional Multinomial Staged Tree Models

The polytopes in 2D from Sect. 3.1 admit a multinomial staged tree representation.

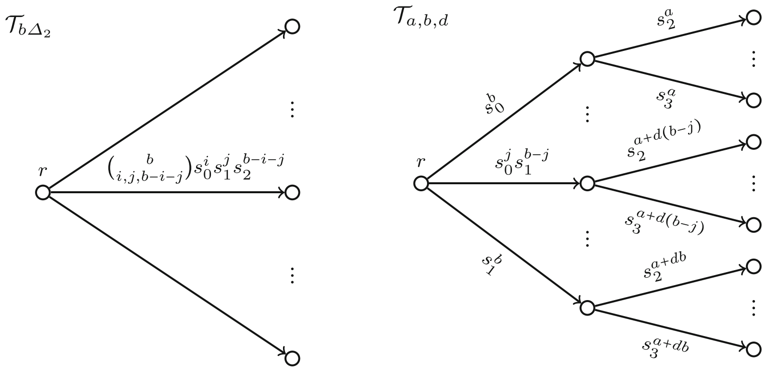


Fig. 4 The multinomial staged trees $\mathcal{T}_{b\Delta_2}$ and $\mathcal{T}_{a,b,d}$ represent the multinomial model with three outcomes and b trials and the model in Example 4.3, respectively

Proposition 5.1 *All statistical models associated with pairs (P, w) in $2D$ with rational linear precision are toric multinomial staged tree models. The multinomial staged tree representations for each family in $2D$ are described in Fig. 4.*

Proof For the model $b\Delta_2 = T_{0,b,1}$, it suffices to note that the polytope $b\Delta_n$ with weights given by multinomial coefficients has a Horn pair given by Theorem 2.4 which is equal to that one described in [11, Example 20] for multinomial models with b trials and $n+1$ outcomes. The statistical model for $T_{a,b,d}$ is the binary multinomial staged tree $\mathcal{M}_{a,b,d}$ in Example 4.3, denoted by $\mathcal{T}_{a,b,d}$ in Fig. 4. The Horn matrix in Proposition 3.1, associated with the model for $T_{a,b,d}$, is equal to the Horn matrix of the model $\mathcal{M}_{a,b,d}$. Firstly, in both cases the columns are indexed by pairs (i, j) such that $0 \leq j \leq b, 0 \leq i \leq a + d(b-j)$ so these matrices have the same number of columns. Using Corollary 4.2, we see that the column corresponding to the outcome (i, j) in $\mathcal{M}_{a,b,d}$ is $(i, j, a+d(b-j)-i, b-j, -(a+d(b-j)), -b)$, which equals the column associated with the lattice point (i, j) in Proposition 3.1. Uniqueness of the minimal Horn matrix, implies that the model associated with $T_{a,b,d}$ is $\mathcal{M}_{a,b,d}$.
It remains to show that $\mathcal{T}_{b\Delta_2}$ and $\mathcal{T}_{a,b,d}$ are balanced. By Remark 4.8, $\mathcal{T}_{b\Delta_2}$ is balanced because all root-to-leaf paths have length 1. For $\mathcal{T}_{a,b,d}$, it suffices to prove that the root r is balanced. Following the notation in Definition 4.6, let $K^1 = (j_1, b-j_1)$, $K^2 = (j_2, b-j_2)$, $K^3 = (j_3, b-j_3)$, and $K_4 = (j_4, b-j_4)$ be such that $K^1 + K^2 = K^3 + K^4$. Then,

$$\begin{aligned} t(r(K^1))t(r(K^2)) &= (s_2+s_3)^{a+d(b-j_1)}(s_2+s_3)^{a+d(b-j_2)} = (s_2+s_3)^{2a+2db-d(j_1+j_2)} \\ &= (s_2+s_3)^{2a+2db-d(j_3+j_4)} = (s_2+s_3)^{a+d(b-j_3)}(s_2+s_3)^{a+d(b-j_4)} \\ &= t(r(K^3))t(r(K^4)) \end{aligned}$$

□

Note that we obtain $P_{\mathcal{T}_{0,b,1}} \cong b\Delta_2$, i.e. we have two different tree representations of $b\Delta_2$: $\mathcal{T}_{b\Delta_2}$ and $\mathcal{T}_{0,b,1}$. For the investigation of the shape of a Horn matrix, we will be interested in those trees $\mathcal{T}$ where the positive part of the Horn matrix $H_{(\mathcal{T},\mathcal{L})}$ from Corollary 4.2 is the lattice distance matrix of $P_{\mathcal{T}}$ (Definition 5.3)(1). For simple polytopes $P_{\mathcal{T}}$, these trees with an additional property provide us an explanation for the negative part of $H_{(\mathcal{T},\mathcal{L})}$ in terms of primitive collections in Theorem 5.1.

5.2 Three-Dimensional Binary Multinomial Staged Tree Models

Before we examine the multinomial staged tree representations more generally, we present the multinomial staged trees for the family $\mathcal{P}$ in Sect. 3.

Proposition 5.2 *All statistical models associated with pairs in $\mathcal{P}$ are toric binary multinomial staged trees.*

Proof We first show that the Horn matrix of the statistical model associated with a general element in $\mathcal{P}$ is equal to the Horn matrix of a binary multinomial staged tree. The general element in $\mathcal{P}$ is a frustum with parameters $a, a', b, b', d, l > 0$. Let $S = \{\{s_0, s_1\}, \{s_2, s_3\}, \{s_4, s_5\}\}$ be a set of symbols. We define the labelled tree $(\mathcal{T}, \mathcal{L})$ by specifying its set of leaves, its set of root-to-leaf paths, and the labelling for the edges in each path. The set of leaves in $\mathcal{T}$ is

$$\begin{aligned} J = \{(i, j, k) : 0 \leq k \leq l,\ \ 0 \leq j \leq bl - (b - b')k, \\ 0 \leq i \leq (a + db)l - ((a + db) - (a' + db')k - dj)\}. \end{aligned}$$

The labelled root-to-leaf path that ends at leaf (i, j, k) is $r \to v \to w \to (i, j, k)$, where

$$\begin{aligned} \mathcal{L}(r \to v) &= \tbinom{l}{k} s_0^k s_1^{l-k}, \\ \mathcal{L}(v \to w) &= \tbinom{b-(b-b')k}{j} s_2^j s_3^{bl-(b-b')k-j}, \text{ and} \\ \mathcal{L}(w \to (i, j, k)) &= \tbinom{(a+db)l-((a+db)-(a+db'))k-dj}{i} \\ &\quad s_4^i s_5^{(a+db)l-((a+db)-(a+db'))k-dj-i}. \end{aligned}$$

Thus, $p_{(i,j,k)}$ is the product of the labels in this path. A picture of this tree when $l = 1$ is contained in Fig. 5 as $\mathcal{T}_{(A)_1}$. Using Corollary 4.2, applied to $(\mathcal{T}, \mathcal{L})$ just defined, we see that the column corresponding to (i, j, k) in the Horn matrix for $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ is equal to the column of the matrix in Proposition 3.2 evaluated at (i, j, k).
It remains to prove that $(\mathcal{T}, \mathcal{L})$ is balanced. Let us first prove that the root r of $(\mathcal{T}, \mathcal{L})$ is balanced. The exponents of the outgoing edges of r can be written as pairs of natural numbers that sum to the degree ℓ of the floret. Thus, they are pairs of the form $(k, \ell - k)$. Let us consider four such pairs, denoted by $Q_1 := (k_1, \ell - k_1)$, $Q_2 := (k_2, \ell - k_2)$, $Q_3 := (k_3, \ell - k_3)$, $Q_4 := (k_4, \ell - k_4)$. Suppose that $Q_1 + Q_2 = Q_3 + Q_4$. Then, we have $k_1 + k_2 = k_3 + k_4$.

We further need to check the following equality:

$$t\big(r(Q_1)\big)t\big(r(Q_2)\big) = t\big(r(Q_3)\big)t\big(r(Q_4)\big).$$

We have:

$$t\big(r(Q_1)\big) = \sum_{j=0}^{b\ell-(b-b')k_1} \left[\tbinom{b\ell-(b-b')k_1}{j} s_2^j s_3^{b\ell-(b-b')k_1-j} t\big(v(j, b\ell - (b - b')k_1 - j)\big) \right]$$

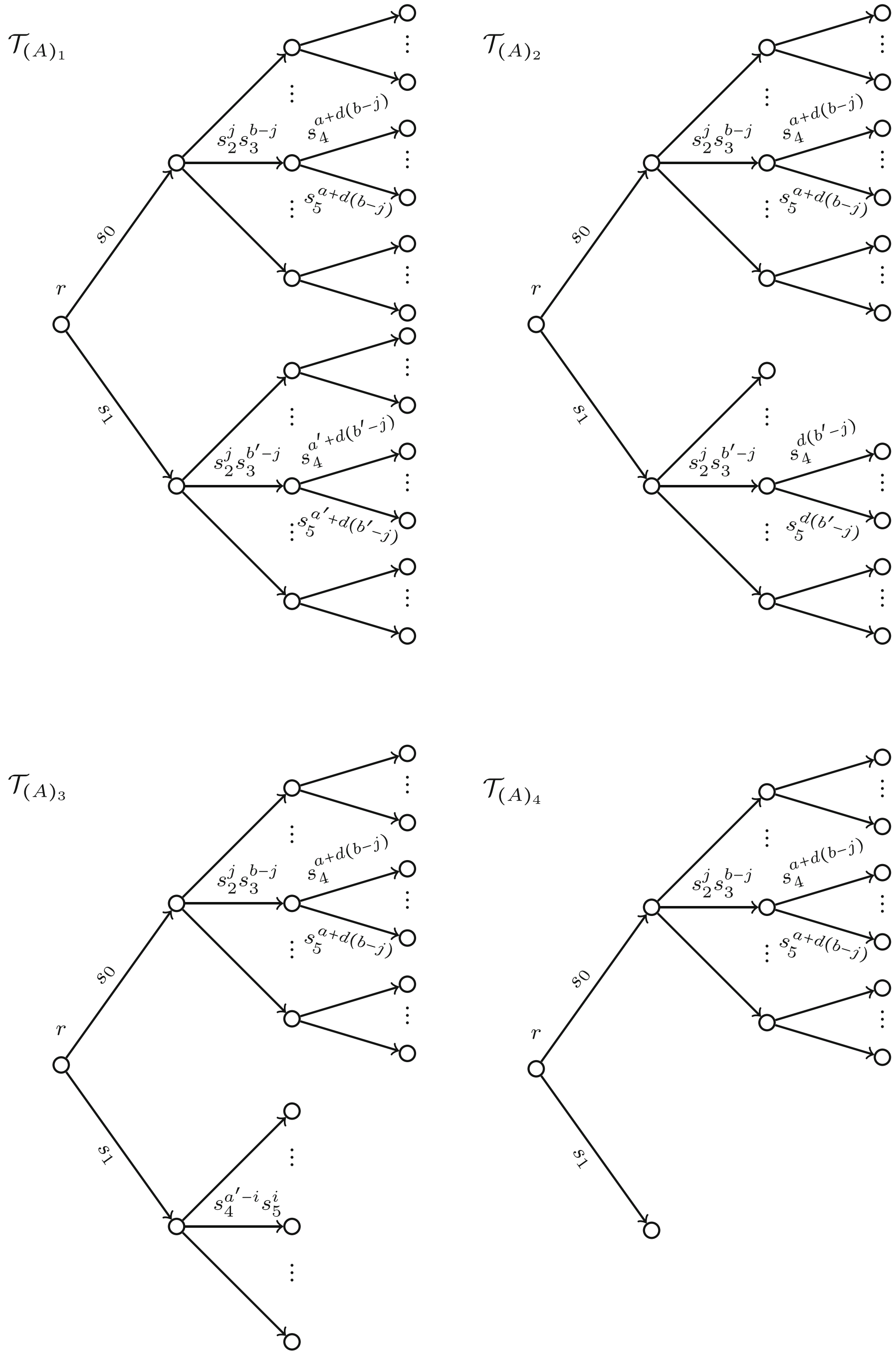


Fig. 5 Multinomial staged trees in 3D for pairs in Table 1(A)

$$= \sum_{j=0}^{b\ell-(b-b')k_1} \Bigg[\binom{b\ell-(b-b')k_1}{j} s_2^j s_3^{b\ell-(b-b')k_1-j}$$
$$(s_4+s_5)^{(a+db)\ell-((a+db)-(a'+db'))k_1-dj} \Bigg]$$
$$= (s_4+s_5)^{(a+db)\ell-((a+db)-(a'+db'))k_1}$$
$$\cdot \sum_{j=0}^{b\ell-(b-b')k_1} \Bigg[\binom{b\ell-(b-b')k_1}{j} \left(\frac{s_2}{(s_4+s_5)^d} \right)^j s_3^{b\ell-(b-b')k_1-j} \Bigg]$$
$$= (s_4+s_5)^{(a+db)\ell-((a+db)-(a'+db'))k_1} \left(\frac{s_2}{(s_4+s_5)^d} + s_3 \right)^{b\ell-(b-b')k_1}.$$

We obtain similar formulae for $t\big(r(Q_2)\big)$, $t\big(r(Q_3)\big)$ and $t\big(r(Q_4)\big)$. It follows that

$$\begin{aligned} t\big(r(Q_1)\big)t\big(r(Q_2)\big) &= (s_4+s_5)^{2(a+db)\ell-((a+db)-(a'+db'))(k_1+k_2)} \\ &\quad \cdot \left(\frac{s_2}{(s_4+s_5)^d} + s_3 \right)^{2b\ell-(b-b')(k_1+k_2)} \\ &= (s_4+s_5)^{2(a+db)\ell-((a+db)-(a'+db'))(k_3+k_4)} \\ &\quad \cdot \left(\frac{s_2}{(s_4+s_5)^d} + s_3 \right)^{2b\ell-(b-b')(k_3+k_4)} \\ &= t\big(r(Q_3)\big)t\big(r(Q_4)\big). \end{aligned}$$

A similar argument can be used to prove that the children of the root are balanced vertices. Next, let us denote by v such a vertex, whose parent is r. By Remark 4.8, any child of v is trivially balanced. Finally, we prove that all pairs of vertices in the same stage are balanced. There are three stages $S_1 = \{s_0, s_1\}$, $S_2 = \{s_2, s_3\}$ and $S_3 = \{s_4, s_5\}$. Denote by v and v' two children of the root r. The exponents of the outgoing edges of v can be written as pairs of natural numbers that sum to the degree $b\ell - (b-b')k_1$ of the floret. Thus, they are pairs of the form $(j, b\ell - (b-b')k_1 - j)$. Let us consider two such pairs, denoted by $Q_1{:=}(j_1, b\ell - (b-b')k_1 - j_1)$, $Q_2{:=}(j_2, b\ell - (b-b')k_1 - j_2)$. Similarly, we consider two children of v' and we denote by $Q_3{:=}(j_3, b\ell - (b-b')k_2 - j_3)$, $Q_4{:=}(j_4, b\ell - (b-b')k_2 - j_4)$. Suppose that $Q_1 + Q_4 = Q_2 + Q_3$. It follows that $j_1 + j_4 = j_2 + j_3$. We want to prove that

$$t\big(v(Q_1)\big)t\big(v(Q_4)\big) = t\big(v(Q_2)\big)t\big(v(Q_3)\big).$$

We obtain the following equality:

$$\begin{aligned} t\big(v(Q_1)\big) &= t\big(v(j_1, b\ell - (b-b')k_1 - j_1)\big) \\ &= (s_4+s_5)^{(a+db)\ell-((a+db)-(a'+db'))k_1-dj_1}, \end{aligned}$$

and its analogues for $t\big(v(Q_2)\big)$, $t\big(v(Q_3)\big)$ and $t\big(v(Q_4)\big)$. Then,

$$\begin{aligned} t\big(v(Q_1)\big)t\big(v(Q_4)\big) &= (s_4+s_5)^{2(a+db)\ell-((a+db)-(a'+db'))(k_1+k_2)-d(j_1+j_4)} \\ &= (s_4+s_5)^{2(a+db)\ell-((a+db)-(a'+db'))(k_1+k_2)-d(j_2+j_3)} \\ &= t\big(v(Q_2)\big)t\big(v(Q_3)\big). \end{aligned}$$

Therefore the pairs of vertices we considered are balanced. □

To obtain the multinomial staged tree representations for the models of the polytopes in Table 1, we use the tree in the proof of Proposition 5.2 and specialise the values of the parameters a, a', b, b', d, l accordingly. The trees for the family of prismatoids with trapezoidal base in Table 1 (A), with $l = 1$, are depicted in Fig. 5. The trapezoidal frusta is represented by $\mathcal{T}_{(A)_1}$, the upper branch is the model for $T_{a,b,d}$ and the lower branch is the model for $T_{a',b',d}$. The other trees, $\mathcal{T}_{(A)_2}$, $\mathcal{T}_{(A)_3}$ and $\mathcal{T}_{(A)_4}$, have the same upper branch as $\mathcal{T}_{(A)_1}$. For the prismatoid with simplex on top, the substitution $a' = 0$ has the effect of chopping a floret from $\mathcal{T}_{(A)_1}$, this gives $\mathcal{T}_{(A)_2}$. For the trapezoidal wedge, $b' = 0$, the edges in $\mathcal{T}_{(A)_1}$ that contain b' contract to a single vertex, yielding $\mathcal{T}_{(A)_3}$. For the trapezoidal pyramid, $a' = b' = 0$, we chop off the lower part of the tree after the edge labelled by s_1. The trees for the remaining part of Table 1 (B), (C), and (D) are obtained similarly.

5.3 Properties of the Polytope $P_{\mathcal{T}}$

In this section, we study certain properties of $P_{\mathcal{T}}$ that can be formulated in terms of the combinatorics of its tree $\mathcal{T}$. We start by looking at root-to-leaf paths in $\mathcal{T}$ that represent vertices of $P_{\mathcal{T}}$, this allows us to work with the normal fan $\Sigma_{P_{\mathcal{T}}}$. For simplicity we assume that $(\mathcal{T}, \mathcal{L})$ has a root-to-leaf path of length m where $S_1, \ldots, S_m$ are the stages of $\mathcal{T}$.

Definition 5.2 A root-to-leaf path j is *vertex representing* if the exponent vector a_j of p_j is a vertex of $P_{\mathcal{T}}$.

Lemma 5.1 *Let $P_{\mathcal{T}} \subset \mathbb{R}^d$ be a polytope where $(\mathcal{T}, \mathcal{L})$ is a multinomial staged tree from Proposition 5.1 or Proposition 5.2. Then, the vertex representing paths in $P_{\mathcal{T}}$ are those for which p_j is divisible by at most one symbol from each stage.*

Proof The upper and lower branches of the tree $\mathcal{T}_{(A)_1}$ are the same up to a choice of parameters, thus we prove it only for the upper branch. Consider a root-to-leaf path $\mathbf{j}$ in $\mathcal{T}_{(A)_1}$ such that $a_{\mathbf{j}} = (1, 0, j, b-j, k, a+d(b-j)-k)$ for $0 < j < b$ and $0 < k < a + d(b - j)$. Let $\mathbf{j}_1$ and $\mathbf{j}_2$ be two root-to-leaf paths such that $a_{\mathbf{j}_1} = (1, 0, j, b-j, a+d(b-j), 0)$ and $a_{\mathbf{j}_2} = (1, 0, j, b-j, 0, a+d(b-j))$. Then, we obtain the equality $a_{\mathbf{j}} = \frac{k}{a+d(b-j))} a_{\mathbf{j}_1} + \frac{a+d(b-j)-k}{a+d(b-j))} a_{\mathbf{j}_2}$ and hence $a_{\mathbf{j}}$ cannot be a vertex of $P_{\mathcal{T}_{(A)_1}}$. It remains to show that $a_{\mathbf{j}} = (1, 0, j, b-j, a+d(b-j), 0))$ is not a vertex. Let now $\mathbf{j}'_1$ and $\mathbf{j}'_2$ be two root-to-leaf paths such that $p_{\mathbf{j}'_1}$ and $p_{\mathbf{j}'_2}$ are divisible by one symbol from each stage and $a_{\mathbf{j}'_1} = (1, 0, b, 0, a, 0)$, $a_{\mathbf{j}'_2} = (1, 0, 0, b, a+db, 0)$.

Hence, $a_{\mathbf{j}} = \frac{j}{b} a_{\mathbf{j}'_1} + \frac{b-j}{b} a_{\mathbf{j}'_2}$ and $a_{\mathbf{j}}$ is not a vertex of $P_{\mathcal{T}_{(A)_1}}$. The proofs for the remaining trees follow similarly. □

Next, we investigate the relation between primitive collections, Horn matrices, and stages of multinomial staged trees. The following definition was motivated by the observations on the trees from Sects. 5.1, 5.2 and by an attempt to answer Question 2.2.

Definition 5.3 Let $(\mathcal{T}, \mathcal{L})$ be a balanced multinomial staged tree representation of $P \cong P_{\mathcal{T}}$. We say that the polytope $P_{\mathcal{T}}$ has property $(\star)$ if,

1. The positive part of the Horn matrix $H_{(\mathcal{T},\mathcal{L})}$ from Corollary 4.2 is the lattice distance matrix of $P_{\mathcal{T}}$.
2. The vertices of $P_{\mathcal{T}}$ satisfy the conclusion of Lemma 5.1.

Note that isomorphic polytopes have the same lattice distance matrices.

Lemma 5.2 *Let $P_{\mathcal{T}} \subset \mathbb{R}^d$ be a polytope with property $(\star)$. Then, $P_{\mathcal{T}}$ is simple if and only if all root-to-leaf paths have the same length m and $|S_1| + \cdots + |S_m| - m = \dim(P_{\mathcal{T}})$.*

Proof Recall that we assumed that there exists a root-to-leaf path of length m. All root-to-leaf paths have the same length m if and only if all vertex representing root-to-leaf paths have the same length m. Recall also that the vertices of $P_{\mathcal{T}}$ are in one-to-one correspondence with the maximal cones of the normal fan $\Sigma_{P_{\mathcal{T}}}$. First suppose that $P_{\mathcal{T}}$ is simple and there exists a vertex representing root-to-leaf path j' of length $m' < m$. Since the positive part of $H_{(\mathcal{T},\mathcal{L})}$ is the lattice distance matrix of $P_{\mathcal{T}}$, the symbols which do not divide $p_{j'}$ represent the facets of $P_{\mathcal{T}}$ which are lattice distance 0 to $a_{j'}$. Since $P_{\mathcal{T}}$ satisfies the conclusion of Lemma 5.1, the maximal cone associated with the vertex $a_{j'}$ in $\Sigma_{P_{\mathcal{T}}}$ has more 1-face (ray) generators than the one associated with a_j where j has length m. Thus, $\Sigma_{P_{\mathcal{T}}}$ is not simplicial, contradiction. Moreover we obtain that the maximal cone associated with a vertex a_j is generated by the normal vectors associated with $\bigcup_{l=1}^{m} S_l \backslash s_{i_l}$ for some $s_{i_l} \in S_l$ and where $\prod_{l=1}^{m} s_{i_l}$ divides p_j. This implies that $\dim(P_{\mathcal{T}}) = |S_1| + \cdots + |S_m| - m$. Now suppose that all vertex representing root-to-leaf paths have the same length m. Then, the number of symbols which do not divide p_j is $|S_1| + \cdots + |S_m| - m$ where j is a vertex representing root-to-leaf path. If this number is equal to $\dim(P_{\mathcal{T}})$, then $P_{\mathcal{T}}$ is simple. □

Remark that the equality $|S_1| + \cdots + |S_m| - m = \dim(P_{\mathcal{T}})$ holds for all models from Propositions 5.1 and 5.2.

Example 5.1 The multinomial staged tree $\mathcal{T}_{(A)_4}$ in Fig. 5 for the trapezoidal pyramid, does not satisfy Definition 5.3 (1). However when $b = 1$, we can find such a balanced multinomial staged tree representation for this polytope, it is shown in Fig. 6 (left). This tree $\mathcal{T}$ and $\mathcal{T}_{(A)_4}$ represent the same model because their minimal Horn matrices are equal. When $a = b = d = 1$, the tree and its polytope are in Fig. 6 (centre) and (right). There are five vertex representing root-to-leaf paths namely 1, 3, 4, 5, and 6 and thus a_1, a_3, a_4, a_5 and a_6 are the vertices of the trapezoidal pyramid. In particular $a_2 = \frac{1}{2}a_1 + \frac{1}{2}a_3$. Hence, $P_{\mathcal{T}}$ has property $(\star)$. Moreover, $P_{\mathcal{T}}$ is not simple by Lemma 5.2, since not all root-to-leaf paths have the length 2.

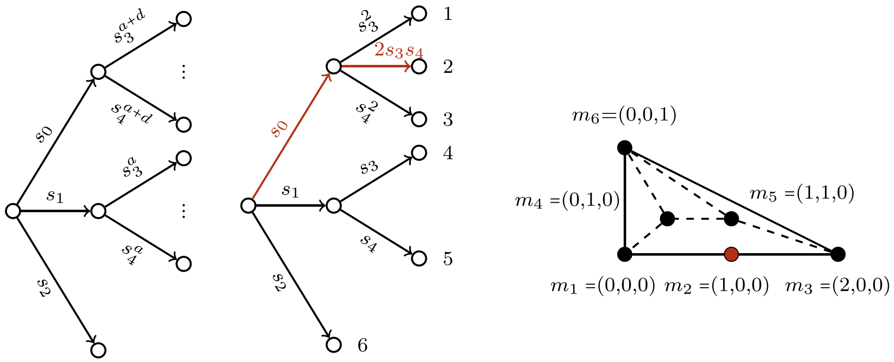


Fig. 6 Multinomial staged tree representation of the non-simple trapezoidal pyramid, $b = 1$

Furthermore, the minimal Horn matrix for this example with $b = 1$ (below left) coincides with $H_{(\mathcal{T},\mathcal{L})}$. As mentioned also in Sect. 3.3.1, the primitive collections $\{n_1, n_3, n_4\}$, $\{n_2, n_3, n_5\}$ do not offer an explanation for the negative part of the minimal Horn matrix, however the stages $\{s_0, s_1, s_2\}$,$\{s_3, s_4\}$ do.

	m_1	m_2	m_3	m_4	m_5	m_6
$s_0{=}h_5$	1	1	1	0	0	0
$s_1{=}h_2$	0	0	0	1	1	0
$s_2{=}h_3$	0	0	0	0	0	1
$s_3{=}h_4$	2	1	0	1	0	0
$s_4{=}h_1$	0	1	2	0	1	0
$-(s_0{+}s_1{+}s_2)$	−1	−1	−1	−1	−1	−1
$-(s_3{+}s_4)$	−2	−2	−2	−1	−1	0

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
$s_1{=}h_3$	0	0	0	0	0	0	0	0	0	1
$s_2{=}h_2$	0	0	0	0	2	2	1	1	1	0
$s_3{=}h_5$	2	2	2	2	0	0	1	1	1	0
$s_4{=}h_4$	3	2	1	0	1	0	2	1	0	0
$s_5{=}h_1$	0	1	2	3	0	1	0	1	2	0
$-(s_0{+}s_1)$	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
$s_0{-}(s_2{+}s_3)$	−1	−1	−1	−1	−1	−1	−1	−1	−1	0
$(s_4{+}s_5)$	−3	−3	−3	−3	−1	−1	−3	−2	−2	0

On the other hand, we observe that there exists no multinomial staged tree representation for $b = 2$, $a = d = 1$ fitting Definition 5.3 by looking at the lattice distance matrix seen in the positive part of its minimal Horn matrix (above right). This matrix can also be obtained by applying Corollary 4.2 to $\mathcal{T}_{(A)_4}$ and performing the row operations explained in [11, Lemma 3] eliminating the row $s_0 = h_6$. This demonstrates how multinomial staged trees provide a wider understanding for the negative part of the Horn matrix.

For simple polytopes $P_{\mathcal{T}}$ with property ($\star$) we show that the stages coincide with the primitive collections of $\Sigma_{P_{\mathcal{T}}}$.

Theorem 5.1 *Let $P_{\mathcal{T}} \subset \mathbb{R}^d$ be a simple polytope with property ($\star$). Then, the primitive collections of the simplicial normal fan $\Sigma_{P_{\mathcal{T}}}$ are represented by the stages $S_1, \ldots, S_m$.*

Proof By Definition 5.3(1), the symbols of the stages represent the facets of $P_{\mathcal{T}}$. Let now j be a vertex representing root-to-leaf path. Recall by the proof of Lemma 5.2, the maximal cone associated with a_j is generated by the normal vectors (1-faces) associated with $\bigcup_{l=1}^m S_l \backslash s_{i_l}$ for some $s_{i_l} \in S_l$ and where $\prod_{l=1}^m s_{i_l}$ divides p_j. Since any intersection of two cones in $\Sigma_{P_{\mathcal{T}}}$ is also a cone in $\Sigma_{P_{\mathcal{T}}}$, we obtain that $\bigcup_{l=1}^m S_l \backslash S_l'$ for all $S_l' \subseteq S_l$ with $|S_l'| \geq 1$ is a cone of $\Sigma_{P_{\mathcal{T}}}$. By Definition 2.6, since $\Sigma_{P_{\mathcal{T}}}$ is simplicial,

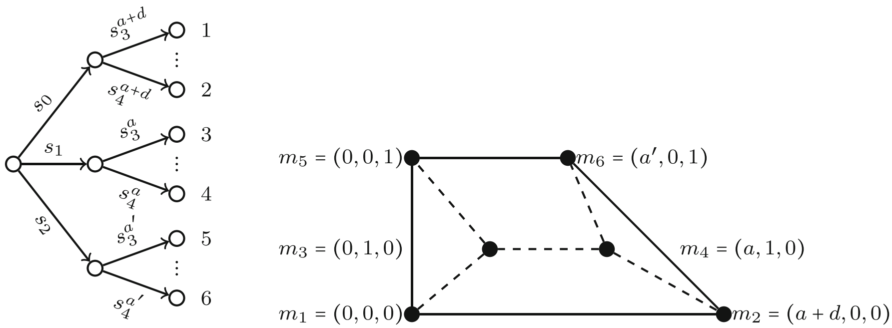


Fig. 7 Multinomial staged tree representation of the simple trapezoidal wedge with $b = 1$ considered in Example 5.3

a primitive collection is a set of 1-faces which does not generate a cone itself but any proper subset does. This concludes that the partition $S_1, \dots . S_m$ are the primitive collections of $\Sigma_{P_{\mathcal{T}}}$. □

The following corollary gives an affirmative answer to Question 2.2.

Corollary 5.1 *Let $P_{\mathcal{T}} \subset \mathbb{R}^d$ be a simple polytope with property* $(\star)$. *Then, we have $H_{(\mathcal{T},\mathcal{L})} = M_{\mathscr{A},\Sigma_{P_{\mathcal{T}}}}$, i.e. the negative rows are given by the primitive collections of $\Sigma_{P_{\mathcal{T}}}$.*

Proof It follows from Corollary 4.2 and Theorem 5.1. □

Example 5.2 The multinomial staged trees $\mathcal{T}_{b\Delta_2}$ and $\mathcal{T}_{a,b,d}$ satisfy Definition 5.3(1) for the simplex and trapezoid $(a, b, d > 0)$, respectively. That means the facets of the polytopes are in one-to-one correspondence with the symbols in the stages. Moreover $P_{\mathcal{T}_{b\Delta_2}}$ and $P_{\mathcal{T}_{a,b,d}}$ are simple polytopes. Hence, by Theorem 5.1 we obtain that the primitive collections are given by the partition of the stages. For the simplex $P_{\mathcal{T}_{b\Delta_2}} \cong a\Delta_2$ we have only one primitive collection $\{s_0, s_1, s_2\}$. Similarly, for $P_{\mathcal{T}_{a,b,d}} \cong T_{a,b,d}$ we have the partition of the stages as $\{s_0, s_1\}$ and $\{s_2, s_3\}$, which correspond exactly to the primitive collections obtained in Theorem 3.1.

Example 5.3 Let us consider the balanced multinomial staged tree representation, satisfying Definition 5.3, of the trapezoidal wedge from Table 1 (A) with $b = 1$ seen in Fig. 7. This tree representation encodes the same model as the tree $\mathcal{T}_{(A)_3}$ in Fig. 5, because they have the same minimal Horn matrix. In particular we observe by Lemma 5.2 that $P_{\mathcal{T}}$ is simple. By Theorem 5.1, the primitive collections are represented by the partition of the stages: $\{n_2, n_3, n_5\} = \{s_0, s_1, s_2\}$ and $\{n_1, n_4\} = \{s_3, s_4\}$. By Corollary 5.1 the negative part of the minimal Horn matrix $a = b = d = a' = 1$ (top), is explained by primitive collections. In Sect. 3.3.2, we saw that even for simple polytopes, the negative part of the Horn matrix is not always explained by the primitive collections. From the perspective of staged trees, for $a' = a = d = 1$, $b = 2$, this polytope can be represented by $\mathcal{T}_{(A)_3}$. After minimising the Horn matrix $H_{(\mathcal{T}_{(A)_3},\mathcal{L}_{(A)_3})}$, constructed by Corollary 4.2, we obtain the matrix (bottom) whose positive part is the lattice distance matrix of $P_{\mathcal{T}_{(A)_3}}$ (see Table 3). However a simple computation shows that there exists no tree $(\mathcal{T}, \mathcal{L})$ such that the positive part of $H_{(\mathcal{T},\mathcal{L})}$ is the lattice distance matrix of $P_{\mathcal{T}_{(A)_3}}$, i.e. satisfying Definition 5.3 (1).

$$
\begin{array}{c} h_5{=}s_0 \\ h_2{=}s_1 \\ h_3{=}s_2 \\ h_4{=}s_3 \\ h_1{=}s_4 \\ -(s_0{+}s_1{+}s_2) \\ -(s_3{+}s_4) \end{array}
\overset{\begin{array}{ccccccc} m_1 & m_2 & m_3 & m_4 & m_5 & m_6 & m_7 \end{array}}{\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 \\
2 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 2 & 0 & 1 & 0 & 1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-2 & -2 & -1 & -1 & -1 & -1 & -2
\end{bmatrix}}
$$

$$
\begin{array}{c} h_3{=}s_1 \\ h_5{=}s_2 \\ h_2{=}s_3 \\ h_4{=}s_4 \\ h_1{=}s_5 \\ -s_0-s_1 \\ s_0-s_2-s_3 \\ -s_4-s_5 \end{array}
\overset{\begin{array}{ccccccccccc} m_1 & m_2 & m_3 & m_4 & m_5 & m_6 & m_7 & m_8 & m_9 & m_{10} & m_{11} \end{array}}{\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
2 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 0 & 0 \\
3 & 2 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 2 & 3 & 0 & 1 & 2 & 0 & 1 & 0 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \\
-3 & -3 & -3 & -3 & -2 & -2 & -2 & -1 & -1 & -1 & -1
\end{bmatrix}}
$$

Acknowledgements Eliana Duarte was supported by the Deutsche Forschungsgemeinschaft DFG under grant 314838170, GRK 2297 MathCoRe, by the FCT grant 2020.01933.CEECIND, and partially supported by CMUP under the FCT grant UIDB/00144/2020. Computations using the software system Sagemath [19] and Macaulay2 [13] were crucial for the development of this paper. We thank Bernd Sturmfels for introducing us to each other and for encouraging us to work on this project.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix A

Table 3 Minimal Horn matrices for pairs in $\mathcal{P}$. The right column in the table contains the columns of the Horn matrix in terms of the lattice distance functions of each polytope

Name of subfamily	Columns of minimal Horn matrix
(A) Prismatoids with trapezoidal base $a > 0, b > 0, d > 0, l > 0$	
Trapezoidal frusta	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_4), -(h_2 + h_5), -(h_3 + h_6))$
Triangle top	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_4), -(h_2 + h_5), -(h_3 + h_6))$
Trapezoidal wedges with $b \neq 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_5 - h_6), -(h_3 + h_6))$
Trapezoidal wedges with $b = 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_3 + h_5))$
Trapezoidal pyramids with $b \neq 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_5 - h_6), -(h_3 + h_6))$
Trapezoidal pyramids with $b = 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_3 + h_5))$
(B) Prismatoids with tensor product base $a > 0, b > 0, d = 0, l > 0$	
General tensor product frusta	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_4), -(h_2 + h_5), -(h_3 + h_6))$
Tensor product frusta with $a' = a$	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_2 + h_5), -(h_1 + h_3 + h_4 + h_6))$
Tensor product frusta with $b' = b$	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_4), -(h_2 + h_3 + h_5 + h_6))$
3D Tensor Product	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_2 + h_3 + h_4 + h_5 + h_6))$
Tensor product frusta with $(a', a) = \lambda(b', b)$ or $\mu(a', a) = (b', b)$ $(\lambda, \mu \geq 1)$	$(h_1, h_2, h_3, h_4, h_5, h_6, -(h_1 + h_2 + h_4 + h_5), -(h_3 + h_6))$
Tensor product wedges $(a' = 0)$ with $a \neq 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4 - h_6), -(h_2 + h_5), -(h_3 + h_6))$
Tensor product wedges $(a' = 0)$ with $a = 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_2 + h_5), -(h_1 + h_3 + h_4))$
Tensor product wedges $(b' = 0)$ with $b \neq 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_5 - h_6), -(h_3 + h_6))$
Tensor product wedges $(b' = 0)$ with $b = 1$	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_4), -(h_2 + h_3 + h_5))$
Tensor product pyramids	$(h_1, h_2, h_3, h_4, h_5, -(h_1 + h_2 + h_4 + h_5 - h_6), -(h_3 + h_6))$

Table 3 continued

Name of subfamily	Columns of minimal Horn matrix
(C) Prismatoids with triangular base $a = 0, b > 0, d > 0, l > 0$	
Triangular frusta ($b' \neq b$) with $d \neq 1$	$(h_1, h_2, h_3, h_4, h_6, -(h_1+h_4-h_5), -(h_2+h_5), -(h_3+h_6))$
Triangular prism ($b' = b$) with $d \neq 1$	$(h_1, h_2, h_3, h_4, h_6, -(h_1+h_4-h_5), -(h_2+h_3+h_5+h_6))$
Triangular frusta ($b' \neq b$) with $d = 1$	$(h_1, h_2, h_3, h_4, h_6, -(h_1+h_2+h_4), -(h_3+h_6))$
Triangular prism ($b' = b$) with $d = 1$	$(h_1, h_2, h_3, h_4, h_6, -(h_1+h_2+h_3+h_4+h_6))$
Triangular-based pyramid with $b \neq 1$ and $d \neq 1$	$(h_1, h_2, h_3, h_4, -(h_1+h_4+-h_5), -(h_2+h_5-h_6), -(h_3+h_6))$
Triangular-based pyramid with $b = 1$ and $d \neq 1$	$(h_1, h_2, h_3, h_4, -(h_1+h_4+-h_5), -(h_2+h_3+h_5))$
Triangular-based pyramid with $b \neq 1$ and $d = 1$	$(h_1, h_2, h_3, h_4, -(h_1+h_2+h_4-h_6), -(h_3+h_6))$
3D simplex	$(h_1, h_2, h_3, h_4, -(h_1+h_2+h_3+h_4))$

Appendix B

The next lemma gives several equations that hold between a point in $\Theta_{\mathcal{T}}$ and its image under $\phi_{\mathcal{T}}$. We use Lemma 7.1 parts (2) and (4), to define the ideal of model invariants for a multinomial staged tree model.

Lemma 7.1 *Let $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ be a multinomial staged tree model where $\mathcal{T} = (V, E)$. Fix $v \in \widetilde{V}$ and suppose $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$ for $\ell \in [m]$, $a \in \mathbb{Z}_{\geq 1}$. Set $i_1, \dots, i_{|I_\ell|}$ to be a fixed ordering of the elements in I_ℓ. Let $(p_1, \dots, p_n) \in \mathcal{M}_{(\mathcal{T},\mathcal{L})}$ and $\theta = (\theta_i)_{i \in I} \in \Theta_{\mathcal{T}}$ be such that $\phi_{\mathcal{T}}(\theta) = (p_1, \dots, p_n)$.*

(1) *For each $K \in \mathbb{N}^{|I_\ell|}$ with $|K| = a$,*

$$\frac{p_{[v(K)]}}{p_{[v]}} = \binom{a}{k_{i_1}, \dots, k_{i_{|I_\ell|}}} \prod_{q=1}^{|I_\ell|} \theta_{i_q}^{k_{i_q}}.$$

(2) *Let $K^1, K^2, K^3, K^4 \in \mathbb{N}^{|I_\ell|}$ with $|K^1| = |K^2| = |K^3| = |K^4| = a$, be such that $K^1 + K^2 = K^3 + K^4$. Define $C_{(K^1,K^2)} := \binom{a}{K^1}\binom{a}{K^2}$ and similarly for $C_{(K^3,K^4)}$. Then,*

$$C_{(K^3,K^4)} p_{[v(K^1)]} p_{[v(K^2)]} - C_{(K^1,K^2)} p_{[v(K^3)]} p_{[v(K^4)]} = 0.$$

(3) *For each $i_q \in I_\ell$, $1 \leq q \leq |I_\ell|$*

$$\theta_{i_q} = \frac{\sum_{|K|=a, k_{i_q} \geq 1} k_{i_q} p_{[v(K)]}}{a p_{[v]}}.$$

(4) *Let* $w \in V$ *and* $\mathrm{im}(\mathcal{L}_w) = f_{\ell,b}$. *For all* $i_q \in I_\ell$*:*

$$bp_{[w]}\left(\sum_{|K|=a, k_{i_q}\geq 1} k_{i_q} p_{[v(K)]}\right) - ap_{[v]}\left(\sum_{|K'|=b, k'_{i_q}\geq 1} k'_{i_q} p_{[w(K')]}\right) = 0.$$

Proof (1) Since $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$ is a probability tree, the transition probability from v to $v(K)$ is the probability of arriving at $v(K)$ divided by the probability of arriving at v, namely $p_{[v(K)]}/p_{[v]}$. By definition of $\mathcal{M}_{(\mathcal{T},\mathcal{L})}$, and since $\mathcal{L}_v(v \to v(K)) = \binom{a}{k_{i_1},\ldots,k_{i_{|I_\ell|}}} \prod_{\alpha=1}^{|I_\ell|} s_{i_\alpha}^{k_{i_\alpha}}$, this probability is exactly $\binom{a}{k_{i_1},\ldots,k_{i_{|I_\ell|}}} \prod_{\alpha=1}^{|I_\ell|} \theta_{i_\alpha}^{k_{i_\alpha}}$.
(2) This equality follows by direct substitution for the values from (1) and by noting that the coefficients $C_{(K^1,K^2)}, C_{(K^3,K^4)}$ are needed to achieve cancellation.
(3) We start from the right-hand side, use (1), the fact that $\sum_{i\in I_\ell}\theta_i = 1$, and simplification with multinomial coefficients to arrive at $\theta_{i_q,}$:

$$\begin{aligned}\frac{\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} p_{[v(K)]}}{ap_{[v]}} &= \sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} \frac{k_{i_q}}{a}\frac{p_{[v(K)]}}{p_{[v]}} = \sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} \frac{k_{i_q}}{a}\binom{a}{k_{i_1},\ldots,k_{i_{|I_\ell|}}}\prod_{\alpha=1}^{|I_\ell|}\theta_{i_\alpha}^{k_{i_\alpha}}\\ &= \sum_{\substack{|K|=a\\ k_{i_q}\geq 1}} \frac{k_{i_q}a}{k_{i_q}a}\binom{a-1}{k_{i_1},\ldots,k_{i_q}-1,\ldots,k_{i_{|I_l|}}}\theta_{i_q}\theta_{i_q}^{k_{i_q}-1}\prod_{\substack{\alpha=1\\ \alpha\neq q}}^{|I_\ell|}\theta_{i_\alpha}^{k_{i_\alpha}}\\ &= \theta_{i_q}\sum_{|K|=a-1}\binom{a-1}{k_{i_1},\ldots,k_{i_q},\ldots,k_{i_{|I_l|}}}\prod_{\alpha=1}^{|I_\ell|}\theta_{i_\alpha}^{k_{i_\alpha}}\\ &= \theta_{i_q}(\sum_{\alpha=1}^{|I_\ell|}\theta_{i_\alpha})^{a-1} = \theta_{i_q}.\end{aligned}$$

(4) Applying part (3) to i_q for v and w separately yields

$$\theta_{i_q} = \frac{\sum_{|K|=a,k_{i_q}\geq 1} k_{i_q} p_{[v(K)]}}{ap_{[v]}} = \frac{\sum_{|K'|=b,k'_{i_q}\geq 1} k'_{i_q} p_{[w(K')]}}{bp_{[w]}}.$$

After cross-multiplication, we get the desired equation in (4). □

Binary Multinomial Staged Trees

From this point on, we assume in $(\mathcal{T},\mathcal{L})$ is a binary multinomial staged tree, and modify our notation according to this assumption.

Lemma 7.2 *Let* $(\mathcal{T},\mathcal{L})$ *be a binary multinomial staged tree where* $\mathcal{T} = (V, E)$ *and let* $v \in \widetilde{V}$ *be such that* $\mathrm{im}(\mathcal{L}_v) = f_{\ell,a}$.

Fix $k_0 \in \{0, \ldots, a\}$. The following equalities hold in $\mathbb{R}[P_j : j \in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$:

$$\binom{a}{k_0+k} P_{[v(k_0,a-k_0)]} l_1^{k_0+k} l_2^{a-(k_0+k)} = \binom{a}{k_0} P_{[v(k_0+k,a-(k_0+k))]} l_1^{k_0} l_2^{a-k_0} \quad , \tag{16}$$

for $1 \le k \le a - k_0$.

$$\binom{a}{k_0-k} P_{[v(k_0,a-k_0)]} l_1^{k_0-k} l_2^{a-(k_0-k)} = \binom{a}{k_0} P_{[v(k_0-k,a-(k_0-k))]} l_1^{k_0} l_2^{a-k_0} \quad , \tag{17}$$

for $1 \le k \le k_0$. where $l_1 := \sum_{k=1}^{a} k P_{[v(k,a-k)]}$ and $l_2 := \sum_{k=1}^{a} k P_{[v(a-k,k)]}$.

For the proof of Lemma 7.2, we use Lemma 7.3.

Lemma 7.3 *Under the hypotheses from Lemma 7.2, the following equality holds in $\mathbb{R}[P_j : j \in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$:*

$$(a - k_0) P_{[v(k_0,a-k_0)]} l_1 = (k_0 + 1) P_{[v(k_0+1,a-(k_0+1))]} l_2. \tag{18}$$

Proof By the part of Definition 4.4 involving I_{vertices}, we see that the equality

$$\binom{a}{k_0+1}\binom{a}{k-1} P_{[v(k_0,a-k_0)]} P_{[v(k,a-k)]} = \binom{a}{k_0}\binom{a}{k} P_{[v(k_0+1,a-(k_0+1))]} P_{[v(k-1,a-(k-1))]}$$

holds $\mathbb{R}[P_j : j \in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$. Note that

$$\binom{a}{k_0+1}\binom{a}{k-1} = \frac{a!a!}{(k_0+1)!(a-k_0)!k!(a-(k-1))!}(a-k_0)k \ , \text{ and}$$
$$\binom{a}{k_0}\binom{a}{k} = \frac{a!a!}{(k_0+1)!(a-k_0)!k!(a-(k-1))!}(k_0+1)(a-(k-1)).$$

Thus, we may cancel the constant $a!a! \ / \ (k_0+1)!(a-k_0)!k!(a-(k-1))!$ from the equality we started with in this proof, to obtain the simplified expression

$$P_{[v(k_0,a-k_0)]} P_{[v(k,a-k)]} = \frac{(k_0+1)(a-(k-1))}{(a-k_0)k} P_{[v(k_0+1,a-(k_0+1))]} P_{[v(k-1,a-(k-1))]}.$$

Using this identity and the definition of l_1 in Lemma 7.2, it follows that

$$\begin{aligned}(a-k_0) P_{[v(k_0,a-k_0)]} l_1 &= (a-k_0) P_{[v(k_0,a-k_0)]} \sum_{k=1}^{a} k P_{[v(k,a-k)]} \\ &= \sum_{k=1}^{a} (a-k_0) k P_{[v(k_0,a-k_0)]} P_{[v(k,a-k)]} \\ &= \sum_{k=1}^{a} (k_0+1)(a-(k-1)) P_{[v(k_0+1,a-(k_0+1))]} P_{[v(k-1,a-(k-1))]}\end{aligned}$$

$$= (k_0+1)P_{[v(k_0+1,a-(k_0+1))]}l_2.$$

□

We are now ready to prove Lemma 7.2.

Proof Let us first prove equality (16). We will do this by mathematical induction on k. First, we show that (16) holds for $k=1$:

$$\begin{aligned}\binom{a}{k_0+1}P_{[v(k_0,a-k_0)]}l_1^{k_0+1}l_2^{a-(k_0+1)} &\overset{(18)}{=} \binom{a}{k_0+1}\frac{k_0+1}{a-k_0}P_{[v(k_0+1,a-(k_0+1))]}l_1^{k_0}l_2^{a-k_0}\\ &= \binom{a}{k_0}P_{[v(k_0+1,a-(k_0+1))]}l_1^{k_0}l_2^{a-k_0}.\end{aligned}$$

Let us now suppose that (16) holds for k, and prove it for $k+1$.

$$\begin{aligned}&\binom{a}{k_0+k+1}P_{[v(k_0,a-k_0)]}l_1^{k_0+k+1}l_2^{a-(k_0+k+1)}\\ &\quad=\frac{a!(a-(k_0+k))}{(k_0+k+1)(k_0+k)!(a-(k_0+k))!}P_{[v(k_0,a-k_0)]}l_1^{k_0+k+1}l_2^{a-(k_0+k+1)}\\ &\quad=\binom{a}{k_0+k}\frac{a-(k_0+k)}{k_0+k+1}P_{[v(k_0,a-k_0)]}l_1 l_1^{k_0+k}l_2^{a-(k_0+k+1)}l_2^{-1}\end{aligned}$$

Using (16) for k, we further simplify to

$$\begin{aligned}&\overset{\text{hyp.}(16)}{=}\binom{a}{k_0}P_{[v(k_0+k,a-(k_0+k))]}l_1^{k_0}l_2^{a-k_0}\frac{a-(k_0+k)}{k_0+k+1}l_1l_2^{-1}\\ &\overset{(18)}{=}\binom{a}{k_0}\frac{a-(k_0+k)}{k_0+k+1}\frac{k_0+k+1}{a-(k_0+k)}P_{[v(k_0+k+1,a-(k_0+k+1))]}l_1^{k_0}l_2l_2^{a-k_0}l_2^{-1}\\ &=\binom{a}{k_0}P_{[v(k_0+k+1,a-(k_0+k+1))]}l_1^{k_0}l_2^{a-k_0}.\end{aligned}$$

Let us now prove equality (17). By (18), we have

$$(a-k_0-1)P_{[v(k_0-1,a-(k_0-1))]}l_1 = k_0P_{[v(k_0,a-k_0)]}l_2.$$

We will again use mathematical induction on k. First, we show that (17) holds for $k=1$:

$$\begin{aligned}&\binom{a}{k_0-1}P_{[v(k_0,a-k_0)]}l_1^{k_0-1}l_2^{a-(k_0-1)}\\ &\quad=\binom{a}{k_0-1}\frac{a-(k_0-1)}{k_0}P_{[v(k_0-1,a-(k_0-1))]}l_1l_2^{-1}l_1^{k_0-1}l_2^{a-(k_0-1)}\\ &\quad=\binom{a}{k_0}P_{[v(k_0-1,a-(k_0-1))]}l_1^{k_0}l_2^{a-k_0}.\end{aligned}$$

Let us now suppose that (17) holds for k, and prove it for $k+1$.

$$\begin{aligned}
&\binom{a}{k_0-(k+1)} P_{[v(k_0,a-k_0)]} l_1^{k_0-(k+1)} l_2^{a-(k_0-(k+1))} \\
&= \binom{a}{k_0-k} \frac{k_0-k}{a-(k_0-(k+1))} P_{[v(k_0,a-k_0)]} l_1^{k_0-k} l_1^{-1} l_2^{a-(k_0-k)} l_2 \\
&\overset{\text{hyp.(17)}}{=} \binom{a}{k_0} P_{[v(k_0-k,a-(k_0-k))]} l_1^{k_0} l_2^{a-k_0} \frac{k_0-k}{a-(k_0-(k+1))} l_1^{-1} l_2 \\
&\overset{(18)}{=} \binom{a}{k_0} \frac{k_0-k}{a-(k_0-(k+1))} \frac{a-(k_0-(k+1))}{k_0-k} l_1 P_{[v(k_0-k-1,a-(k_0-k-1))]} l_1^{k_0} l_2^{a-k_0} l_1^{-1} \\
&= \binom{a}{k_0} P_{[v(k_0-(k+1),a-(k_0-(k+1)))]} l_1^{k_0} l_2^{a-k_0}.
\end{aligned}$$

□

Lemma 7.4 *Under the same hypotheses as in Lemma 7.2, the following equalities hold in* $\mathbb{R}[P_j : j \in J]/I_{\mathcal{M}(\mathcal{T},\mathcal{L})}$*:*

$$\binom{a}{k_0}\left(\frac{l_1}{aP[v]}\right)^{k_0}\left(\frac{l_2}{aP[v]}\right)^{a-k_0} = \frac{P_{[v(k_0,a-k_0)]}}{P_{[v]}}. \tag{19}$$

Proof First, note that $aP_{[v]} = (l_1 + l_2)$. Indeed

$$\begin{aligned}
l_1 + l_2 &= (P_{[v(1,a-1)]} + 2P_{[v(2,a-2)]} + \cdots + (a-1)P_{[v(a-1,1)]} + aP_{[v(a,0)]}) \\
&\quad + (aP_{[v(0,a)]} + (a-1)P_{[v(1,a-1)]} + \cdots + 2P_{[v(a-2,2)]} + P_{[v(a-1,1)]}) \\
&= a(P_{[v(a,0)]} + P_{[v(1,a-1)]} + P_{[v(2,a-2)]} \\
&\quad + \cdots + P_{[v(1,a-1)]} + P_{[v(0,a)]}) = aP_{[v]}.
\end{aligned}$$

Therefore, we have $P_{[v(k_0,a-k_0)]}(aP_{[v]})^a = P_{[v(k_0,a-k_0)]}(l_1+l_2)^a$. Now, by Lemma 7.2, we obtain:

$$\begin{aligned}
P_{[v(k_0,a-k_0)]}(l_1+l_2)^a &= P_{[v(k_0,a-k_0)]} \sum_{k=0}^{a} \binom{a}{k} l_1^k l_2^{a-k} \\
&= \sum_{k=0}^{a} \binom{a}{k} P_{[v(k_0,a-k_0)]} l_1^k l_2^{a-k} \\
&= \sum_{k=1}^{k_0} \binom{a}{k_0-k} P_{[v(k_0,a-k_0)]} l_1^{k_0-k} l_2^{a-(k_0-k)} \\
&\quad + \binom{a}{k_0} P_{[v(k_0,a-k_0)]} l_1^{k_0} l_2^{a-k_0} \\
&\quad + \sum_{k=1}^{a-k_0} \binom{a}{k_0+k} P_{[v(k_0,a-k_0)]} l_1^{k_0+k} l_2^{a-(k_0+k)}
\end{aligned}$$

$$= \binom{a}{k_0} l_1^{k_0} l_2^{a-k_0} \left(\sum_{k=1}^{k_0} P_{[v(k_0-k, a-(k_0-k))]} + P_{[v(k_0, a-k_0)]} + \sum_{k=1}^{a-k_0} P_{[v(k_0+k, a-(k_0+k))]} \right)$$

$$= \binom{a}{k_0} l_1^{k_0} l_2^{a-k_0} P_{[v]}.$$

□

References

1. Améndola, C., Bliss, N., Burke, I., Gibbons, C.R., Helmer, M., Hoşten, S., Nash, E.D., Rodriguez, J.I., Smolkin, D.: The maximum likelihood degree of toric varieties. J. Symbolic Comput. **92**, 222–242 (2019). 10.1016/j.jsc.2018.04.016.
2. Ananiadi, L., Duarte, E.: Gröbner bases for staged trees. Algebr. Stat. **12**(1), 1–20 (2021). 10.2140/astat.2021.12.1.
3. Batyrev, V.V.: On the classification of smooth projective toric varieties. Tohoku Math. J. (2) **43**(4), 569–585 (1991). 10.2748/tmj/1178227429.
4. Graf-von-Bothmer, H.C., Ranestad, K., Sottile, F.: Linear precision for toric surface patches. Found. Comput. Math. **10**(1), 37–66 (2010). https://doi.org/10.1007/s10208-009-9052-6.
5. Clarke, P., Cox, D.A.: Moment maps, strict linear precision, and maximum likelihood degree one. Adv. Math. **370**, 107233 (2020). 10.1016/j.aim.2020.107233.
6. Cox, D.A.: Applications of polynomial systems, vol. 134. American Mathematical Soc. (2020)
7. Cox, D.A., Little, J., O'Shea, D.: Ideals, varieties, and algorithms, fourth edn. Undergraduate Texts in Mathematics. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16721-3. An introduction to computational algebraic geometry and commutative algebra
8. Cox, D.A., Little, J.B., Schenck, H.K.: Toric varieties, *Graduate Studies in Mathematics*, vol. 124. American Mathematical Society, Providence, RI (2011). https://doi.org/10.1090/gsm/124.
9. Cox, D.A., von Renesse, C.: Primitive collections and toric varieties. Tohoku Math. J. **61**(3), 309–332 (2009). https://doi.org/10.2748/tmj/1255700197.
10. DeRose, T.D., Goldman, R.N., Hagen, H., Mann, S.: Functional composition algorithms via blossoming. ACM Transactions on Graphics (TOG) **12**(2), 113–135 (1993)
11. Duarte, E., Marigliano, O., Sturmfels, B.: Discrete statistical models with rational maximum likelihood estimator. Bernoulli **27**(1), 135–154 (2021). 10.3150/20-BEJ1231.
12. Garcia-Puente, L.D., Sottile, F.: Linear precision for parametric patches. Adv. Comput. Math. **33**(2), 191–214 (2010). 10.1007/s10444-009-9126-7.
13. Grayson, D.R., Stillman, M.E.: Macaulay2, a software system for research in algebraic geometry. Available at https://math.uiuc.edu/Macaulay2/
14. Huh, J.: Varieties with maximum likelihood degree one. J. Algebr. Stat. **5**(1), 1–17 (2014). https://doi.org/10.18409/jas.v5i1.22
15. Krasauskas, R.: Toric surface patches. Advances in Computational Mathematics **17**(1), 89–113 (2002). 10.1023/A:1015289823859.
16. Lauritzen, S.L.: Graphical models, Oxford Statistical Science Series, vol. 17. The Clarendon Press, Oxford University Press, New York (1996). Oxford Science Publications
17. Smith, J.Q., Anderson, P.E.: Conditional independence and chain event graphs. Artificial Intelligence **172**(1), 42–68 (2008). 10.1016/j.artint.2007.05.004.
18. Sullivant, S.: Algebraic statistics, Graduate Studies in Mathematics, vol. 194. American Mathematical Society, Providence, RI (2018). https://doi.org/10.1090/gsm/194.
19. The Sage Developers: SageMath, the Sage Mathematics Software System (Version 9.1) (2020). https://www.sagemath.org