Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Discrete Statistical Models with Rational Maximum Likelihood Estimator

by

Eliana Maria Duarte Gelvez, Orlando Marigliano, and Bernd Sturmfels

Preprint no.: 29 2019

Discrete Statistical Models with Rational Maximum Likelihood Estimator

Eliana Duarte, Orlando Marigliano and Bernd Sturmfels

Abstract

A discrete statistical model is a subset of a probability simplex. Its maximum likelihood estimator (MLE) is a retraction from that simplex onto the model. We characterize all models for which this retraction is a rational function. This is a contribution via real algebraic geometry which rests on results due to Huh and Kapranov on Horn uniformization. We present an algorithm for constructing models with rational MLE, and we demonstrate it on a range of instances. Our focus lies on models familiar to statisticians, like Bayesian networks, decomposable graphical models, and staged trees.

1 Introduction

A *discrete statistical model* is a subset $\mathcal{M}$ of the open probability simplex Δ_n. Each point p in Δ_n is a probability distribution on the finite state space $\{0, 1, \ldots, n\}$, i.e. $p = (p_0, p_1, \ldots, p_n)$, where the p_i are positive real numbers that satisfy $p_0 + p_1 + \cdots + p_n = 1$. The model $\mathcal{M}$ is the set of all distributions $p \in \Delta_n$ that are relevant for the particular application of interest.

In data analysis we are given an empirical distribution $u = (u_0, u_1, \ldots, u_n)$. This is the point in the simplex Δ_n whose ith coordinate u_i is the fraction of samples observed to be in state i. The *maximum likelihood estimator* (MLE) of $\mathcal{M}$ is a function $\Phi\colon \Delta_n \to \mathcal{M}$ that takes the empirical distribution u to a distribution $\hat{p} = (\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_n)$ that best explains the given observations. Here "best" is understood in the sense of likelihood inference. This means that $\hat{p} = \Phi(u)$ is the point in $\mathcal{M}$ that maximizes the *log-likelihood function* $p \mapsto \sum_{i=0}^{n} u_i \cdot \log(p_i)$. By convention, for any vector u in $\mathbb{R}^{n+1}_{>0}$, we set $\Phi(u) := \Phi(u/|u|)$ where $|u| = u_0 + \cdots + u_n$.

Likelihood inference is consistent. This means that $\Phi(u) = u$ for $u \in \mathcal{M}$. This follows from the fact that the log-likelihood function is strictly concave on Δ_n and its unique maximizer is $p = u$. Therefore, the MLE Φ is a retraction from the simplex onto the model.

This point is fundamental for two fields at the crossroads of mathematics and data science. *Information Geometry* [1] views the MLE as the nearest point map of a Riemannian metric on Δ_n, given by the Kullback-Leibler divergence of probability distributions. *Algebraic Statistics* [5] is concerned with models $\mathcal{M}$ whose MLE Φ is an algebraic function of u. This happens precisely when the constraints that define $\mathcal{M}$ can be expressed in terms of polynomials in p. In this article we address a question that is fundamental for both fields: *For which models $\mathcal{M}$ is the MLE Φ a rational function in the empirical distribution u?*

The most basic example where this happens is the independence model for two binary random variables ($n = 3$). Here $\mathcal{M}$ is a surface in the tetrahedron Δ_3. That surface is a familiar picture that serves as a point of entry for both Information Geometry and Algebraic Statistics. Points in $\mathcal{M}$ are positive rank one 2×2 matrices $\begin{bmatrix} p_0 & p_1 \\ p_2 & p_3 \end{bmatrix}$ whose entries sum to one. The data takes the form of a nonnegative integer 2×2 matrix u of counts of observed frequencies. Hence $|u| = u_0+u_1+u_2+u_3$ is the sample size, and $u/|u|$ is the empirical distribution. The MLE $\hat{p} = \Phi(u)$ is evaluated by multiplying the row and column sums of u:

$$\hat{p}_0 = \frac{(u_0+u_1)(u_0+u_2)}{|u|^2},\ \hat{p}_1 = \frac{(u_0+u_1)(u_1+u_3)}{|u|^2},\ \hat{p}_2 = \frac{(u_2+u_3)(u_0+u_2)}{|u|^2},\ \hat{p}_3 = \frac{(u_2+u_3)(u_1+u_3)}{|u|^2}.$$

These four expressions are rational, homogeneous of degree zero, and their sum is equal to 1. We refer to [10, Example 2] for a discussion of these formulas from our present perspective.

The independence model belongs to the class of graphical models [14]. Fix an undirected graph G whose nodes represent random variables with finitely many states. The undirected graphical model $\mathcal{M}_G$ is a subset of Δ_n, where $n+1$ is the number of states in the joint distribution. The graphical model $\mathcal{M}_G$ is *decomposable* if and only if the graph G is chordal. In this case, each coordinate $\hat{p}_i$ of the MLE is an alternating product of linear forms given by maximal cliques and minimal separators of G. A similar formula exists for directed graphical models, also known as Bayesian networks. In both cases, the coordinates of the MLE are not only rational functions, but even alternating products of linear forms in $u = (u_0, u_1, \ldots, u_n)$.

This is no coincidence. Huh [10] proved that if Φ is a rational function then each of its coordinates is an alternating product of linear forms, with numerator and denominator of the same degree. Huh further showed that this alternating product must take a very specific shape. That shape was discovered by Kapranov [12], who named it the *Horn uniformization*. The results by Kapranov and Huh are valid for arbitrary complex algebraic varieties. They make no reference to a context where the coordinates are real, positive, and add up to 1.

The present paper makes the leap from complex varieties back to statistical models. Building on the remarkable constructions found by Kapranov and Huh, we here work in the setting of real algebraic geometry that is required for statistical applications. Our main result (Theorem 1) characterizes all models $\mathcal{M}$ in Δ_n whose MLE is a rational function. It is stated in Section 2 and all its ingredients are presented in a self-contained manner.

In Section 3 we examine models with rational MLE that are familiar to statisticians, such as decomposable graphical models and Bayesian networks. Our focus lies on *staged tree models*, a far-reaching generalization of discrete Bayesian networks, as described in the book by Collazo, Görgen and Smith [3]. We explain how our main result applies to these models.

The proof of Theorem 1 is presented in Section 4. This is the technical heart of our paper, building on the likelihood geometry of [11, §3]. We also discuss the connection to toric geometry and geometric modeling that appeared in recent work of Clarke and Cox [2].

In Section 5 we present our algorithm for constructing models with rational MLE, and we discuss its implementation and some experiments. The input is an integer matrix representing a toric variety, and the output is a list of models derived from that matrix. Our results suggest that only a small fraction of Huh's varieties in [10] are statistical models.

2 How to be Rational

Let $\mathcal{M}$ be a discrete statistical model in the open simplex Δ_n that has a well-defined maximum likelihood estimator $\Phi : \Delta_n \to \mathcal{M}$. We also write $\Phi : \mathbb{R}^{n+1}_{>0} \to \mathcal{M}$ for the induced map $u \mapsto \Phi(u/|u|)$ on all positive vectors. If the $n+1$ coordinates of Φ are rational functions in u, then we say that $\mathcal{M}$ *has rational MLE*. The following is our main result in this paper.

Theorem 1. *The following are equivalent for a discrete statistical model $\mathcal{M}$ with MLE Φ:*

(1) The model $\mathcal{M}$ has **rational MLE**.

(2) There exists a **Horn pair** (H, λ) *such that $\mathcal{M}$ is the image of the Horn map $\varphi_{(H,\lambda)}$.*

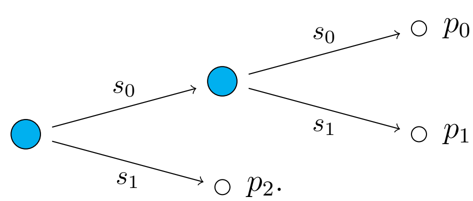
(3) There exists a **discriminantal triple** $(A, \Delta, \mathbf{m})$ *such that $\mathcal{M}$ is the image under the monomial map $\phi_{(\Delta,\mathbf{m})}$ of precisely one orthant (9) of the dual toric variety Y_A^*.*

The MLE of the model satisfies the following relation on the open orthant $\mathbb{R}^{n+1}_{>0}$:

$$\Phi \;=\; \varphi_{(H,\lambda)} \;=\; \phi_{(\Delta,\mathbf{m})} \circ H. \tag{1}$$

The goal of this section is to define all the terms seen in parts (2) and (3) of this theorem.

Example 2. We first discuss Theorem 1 for a very simple experiment: *Flip a biased coin. If it shows heads, flip it again.* This is a statistical model with $n = 2$ given by the tree diagram



The model $\mathcal{M}$ is a curve in the probability triangle Δ_2. The tree shows its parametrization

$$\Delta_1 \to \Delta_2\,,\;\; (s_0, s_1) \mapsto (s_0^2, s_0 s_1, s_1) \qquad \text{where } s_0, s_1 > 0 \text{ and } s_0 + s_1 = 1.$$

The implicit representation of the curve $\mathcal{M}$ is the quadratic equation $p_0p_2 - (p_0 + p_1)p_1 = 0$.

Let (u_0, u_1, u_2) be the counts from repeated experiments. A total of $2u_0 + 2u_1 + u_2$ coin tosses were made. We estimate the parameters as the empirical frequency of heads resp. tails:

$$\hat{s}_0 \;=\; \frac{2u_0 + u_1}{2u_0 + 2u_1 + u_2} \quad \text{and} \quad \hat{s}_1 \;=\; \frac{u_1 + u_2}{2u_0 + 2u_1 + u_2}.$$

The MLE is the retraction from the triangle Δ_2 to the curve $\mathcal{M}$ given by the rational formula

$$\Phi(u_0, u_1, u_2) \;=\; (\hat{s}_0^2, \hat{s}_0\hat{s}_1, \hat{s}_1) \;=\; \left(\frac{(2u_0 + u_1)^2}{(2u_0{+}2u_1{+}u_2)^2}\,,\, \frac{(2u_0{+}u_1)(u_1{+}u_2)}{(2u_0 + 2u_1 + u_2)^2}\,,\, \frac{u_1 + u_2}{2u_0{+}2u_1{+}u_2} \right).$$

Hence $\mathcal{M}$ has rational MLE. The corresponding Horn pair from part (2) in Theorem 1 has

$$H = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ -2 & -2 & -1 \end{pmatrix} \quad \text{and} \quad \lambda = (1, 1, -1).$$

We next exhibit the discriminantal triple $(A, \Delta, \mathbf{m})$ in part (3) of Theorem 1. The matrix $A = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$ gives a basis of the left kernel of H. The second entry is the polynomial

$$\Delta \;=\; x_3^2 - x_1^2 - x_1x_2 + x_2x_3 \;=\; (x_3 - x_1)(x_1 + x_2 + x_3). \tag{2}$$

The third entry marks the leading term $\mathbf{m} = x_3^2$. These data define the monomial map

$$\phi_{(\Delta,\mathbf{m})} \;:\; (x_1, x_2, x_3) \;\mapsto\; \left(\frac{x_1^2}{x_3^2}, \frac{x_1x_2}{x_3^2}, -\frac{x_2}{x_3} \right).$$

The toric variety of the matrix A is the point $Y_A = \{(1 : 1 : 1)\}$ in $\mathbb{P}^2$. Our polynomial Δ vanishes on the line $Y_A^* = \{x_1 + x_2 + x_3 = 0\}$ that is dual to Y_A. The relevant orthant is the open line segment $Y_{A,\sigma}^* := \{(x_1 : x_2 : x_3) \in Y_A^* \;:\; x_1, x_2 > 0 \text{ and } x_3 < 0\}$. Part (3) in Theorem 1 says that $\mathcal{M}$ is the image of $Y_{A,\sigma}^*$ under $\phi_{(\Delta,\mathbf{m})}$. The MLE is $\Phi = \phi_{(\Delta,\mathbf{m})} \circ H$.

We now come to the definitions needed for Theorem 1. Let $H = (h_{ij})$ be an $m \times (n{+}1)$ integer matrix whose columns sum to zero, i.e. $\sum_{i=1}^m h_{ij} = 0$ for $j = 0, \ldots, n$. We call such a matrix a *Horn matrix*. The following alternating products of linear forms have degree zero:

$$(Hu)^{h_j} \;:=\; \prod_{i=1}^{m} \big(h_{i0}u_0 + h_{i1}u_1 + \cdots + h_{in}u_n\big)^{h_{ij}} \qquad \text{for } j = 0, 1, \ldots, n.$$

The Horn matrix H is *friendly* if there exists a real vector $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_n)$ with $\lambda_i \neq 0$ for all i such that the following identity holds in the rational function field $\mathbb{R}(u_0, u_1, \ldots, u_n)$:

$$\lambda_0(Hu)^{h_0} + \lambda_1(Hu)^{h_1} + \cdots + \lambda_n(Hu)^{h_n} \;=\; 1. \tag{3}$$

If this holds, then we say that (H, λ) is a *friendly pair*, and we consider the rational function

$$\mathbb{R}^{n+1} \to \mathbb{R}^{n+1},\; u \mapsto \big(\lambda_0(Hu)^{h_0},\, \lambda_1(Hu)^{h_1},\, \ldots,\, \lambda_n(Hu)^{h_n}\big). \tag{4}$$

The friendly pair (H, λ) is called a *Horn pair* if no row of H is zero or is a multiple of another row, the function (4) is defined for all positive vectors, and it maps these to positive vectors. If these conditions hold then we write $\varphi_{(H,\lambda)} : \mathbb{R}_{>0}^{n+1} \to \mathbb{R}_{>0}^{n+1}$ for the restriction of (4) to the positive orthant. We call $\varphi_{(H,\lambda)}$ the *Horn map* associated to the Horn pair (H, λ).

Remark 3. Let (H, λ) be a friendly pair satisfying the positivity condition for the function (4). To it we associate a Horn pair $(\tilde{H}, \tilde{\lambda})$ by aggregating its collinear rows by summing them together, deleting the zero rows of H, and defining $\tilde{\lambda}$ as in [2, Proposition 6.11]. The pairs (H, λ) and $(\tilde{H}, \tilde{\lambda})$ define the same rational function (4). Furthermore, every Horn pair $(\tilde{H}, \tilde{\lambda})$ can be uniquely recovered, up to permutation of its rows, from its Horn map $\varphi_{(\tilde{H},\tilde{\lambda})}$.

Example 4. We illustrate the equivalence of (1) and (2) in Theorem 1 for the model described in [11, Example 3.11]. Here $n = 3$ and $m = 4$ and the Horn matrix equals

$$H \quad = \quad \begin{pmatrix} -1 & -1 & -2 & -2 \\ 1 & 0 & 3 & 2 \\ 1 & 3 & 0 & 2 \\ -1 & -2 & -1 & -2 \end{pmatrix}. \tag{5}$$

This Horn matrix is friendly because the following vector satisfies the identity (3):

$$\lambda \;=\; (\lambda_0, \lambda_1, \lambda_2, \lambda_3) \;=\; \left(\frac{2}{3}, -\frac{4}{27}, -\frac{4}{27}, \frac{1}{27}\right). \tag{6}$$

The pair (H, λ) is a Horn pair, with associated Horn map

$$\varphi_{(H,\lambda)} : \mathbb{R}^4_{>0} \to \mathbb{R}^4_{>0}, \quad \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} \mapsto \begin{pmatrix} \frac{2(u_0+3u_2+2u_3)(u_0+3u_1+2u_3)}{3(u_0+u_1+2u_2+2u_3)(u_0+2u_1+u_2+2u_3)} \\ \frac{4(u_0+3u_2+2u_3)^3}{27(u_0+u_1+2u_2+2u_3)^2(u_0+2u_1+u_2+2u_3)} \\ \frac{4(u_0+3u_1+2u_3)^3}{27(u_0+u_1+2u_2+2u_3)(u_0+2u_1+u_2+2u_3)^2} \\ \frac{(u_0+3u_2+2u_3)^2(u_0+3u_1+2u_3)^2}{27(u_0+u_1+2u_2+2u_3)^2(u_0+2u_1+u_2+2u_3)^2} \end{pmatrix}. \tag{7}$$

Indeed, this rational function evidently takes positive vectors to positive vectors. The image of the map $\varphi_{(H,\lambda)}$ is a subset $\mathcal{M}$ of the tetrahedron $\Delta_3 = \{p \in \mathbb{R}^4_{>0} : p_0 + p_1 + p_2 + p_3 = 1\}$. We regard the subset $\mathcal{M}$ as a discrete statistical model on the state space $\{0, 1, 2, 3\}$. The model $\mathcal{M}$ is the rational space curve of degree 4 defined by the two quadratic equations

$$9p_1p_2 - 8p_0p_3 \;=\; p_0^2 - 12(p_0 + p_1 + p_2 + p_3)p_3 \;=\; 0.$$

As in [11, Example 3.11], one verifies that the curve $\mathcal{M}$ has rational MLE, namely $\Phi = \varphi_{(H,\lambda)}$.

We next define all the terms that are used in part (3) of Theorem 1. Fix a matrix $A = (a_{ij}) \in \mathbb{Z}^{r \times m}$ of rank r that has the vector $(1, \ldots, 1)$ in its row span. The connection to (2) will be that the rows of A span the left kernel of H. We identify the columns of A with Laurent monomials in r unknowns $t_1, \ldots, t_r$. The corresponding monomial map is

$$\gamma_A \;:\; (\mathbb{R}^*)^r \to \mathbb{RP}^{m-1}, \;\; (t_1, \ldots, t_r) \mapsto \left(\prod_{i=1}^r t_i^{a_{i1}} : \prod_{i=1}^r t_i^{a_{i2}} : \;\cdots\; : \prod_{i=1}^r t_i^{a_{im}}\right). \tag{8}$$

Here $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ and $\mathbb{RP}^{m-1}$ denotes the real projective space of dimension $m-1$. Let Y_A be the closure of the image of γ_A. This is the projective toric variety given by the matrix A.

Every point $x = (x_1 : \cdots : x_m)$ in the dual projective space $(\mathbb{RP}^{m-1})^\vee$ corresponds to a hyperplane H_x in $\mathbb{RP}^{m-1}$. The dual variety Y_A^* to the toric variety Y_A is the closure of the set

$$\left\{ x \in \mathbb{RP}^{m-1} \mid \gamma_A^{-1}(H_x \cap Y_A) \text{ is singular} \right\}.$$

A general point x in the *dual toric variety* Y_A^* corresponds to a hyperplane H_x that is tangent to the toric variety Y_A at a point $\gamma_A(t)$ with nonzero coordinates. We identify sign vectors

$\sigma \in \{-1,+1\}^m$ with orthants in $\mathbb{R}^m$. These map in a 2-to-1 manner to orthants in $\mathbb{RP}^{m-1}$. If we intersect them with Y_A^*, then we get the *orthants* of the dual toric variety:

$$Y_{A,\sigma}^* \;=\; \bigl\{\, x \in Y_A^* \,:\, \sigma_i \cdot x_i > 0 \text{ for } i = 1,2,\ldots,m \,\bigr\} \quad \subset \quad \mathbb{RP}^{m-1}. \tag{9}$$

One of these is the distinguished orthant in Theorem 1, part (3).

Example 5. Fix $m=4$, $r=2$, and the following matrix with $(1,1,1,1)$ in its row span:

$$A \;=\; \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}. \tag{10}$$

As in [11, Example 3.9], the toric variety of A is the *twisted cubic curve* in projective 3-space:

$$Y_A \;=\; \overline{\bigl\{(t_1^3 : t_1^2 t_2 : t_1 t_2^2 : t_2^3) \in \mathbb{RP}^3 \;:\; t_1, t_2 \in \mathbb{R}^*\bigr\}}.$$

The dual toric variety Y_A^* is a surface in $(\mathbb{RP}^3)^\vee$. Its points x represent planes in $\mathbb{RP}^3$ that are tangent to the curve Y_A. Such a tangent plane corresponds to a cubic $x_1 t^3 + x_2 t^2 + x_3 t + x_4$ with a double root. Hence, Y_A^* is the surface of degree 4 in $(\mathbb{RP}^3)^\vee$ defined by the discriminant

$$\Delta_A \;=\; \underline{27 x_1^2 x_4^2} - 18 x_1 x_2 x_3 x_4 + 4 x_1 x_3^3 + 4 x_2^3 x_4 - x_2^2 x_3^2. \tag{11}$$

All eight orthants $Y_{A,\sigma}^*$ are non-empty. Representatives x for the orthants are the eight cubics

$$\begin{array}{c} (t+1)^2(t+3),\ (t+5)^2(t-1),\ (t-1)^2(t+3),\ (t+5)^2(t-8), \\ (t-3)^2(t+1),\ (t-1)^2(t-3),\ \underline{(t-2)^2(t+3)},\ (t+1)^2(t-3). \end{array}$$

The underlined cubic is the point $x = (1,-1,-8,12)$ in $Y_{A,\sigma}^*$, where $\sigma = (1,-1,-1,1)$.

We now present the key definition that is needed for part (3) of Theorem 1. Let Δ be a homogeneous polynomial with $n+2$ monomials, and let $\mathbf{m}$ be one of these monomials. If we divide Δ by $\mathbf{m}$, then we obtain a homogeneous Laurent polynomial of degree zero:

$$\frac{1}{\mathbf{m}}\Delta \;=\; 1 \,-\, \lambda_0 x_1^{h_{10}} x_2^{h_{20}} \cdots x_m^{h_{m0}} \,-\, \lambda_1 x_1^{h_{11}} x_2^{h_{21}} \cdots x_m^{h_{m1}} \,-\, \cdots \,-\, \lambda_n x_1^{h_{1n}} x_2^{h_{2n}} \cdots x_m^{h_{mn}}.$$

We write $H_{(\Delta,\mathbf{m})}$ for the $m \times (n+1)$ integer matrix with entries h_{ij}. Its column vectors are denoted $h_j = (h_{1j}, h_{2j}, \ldots, h_{mj})$ for $j = 0,1,\ldots,n$. These data define the monomial map

$$\phi_{(\Delta,\mathbf{m})} \;:\; (\mathbb{R}^*)^m \to \mathbb{R}^{n+1},\ x \;\mapsto\; \bigl(\lambda_0 x^{h_0},\, \lambda_1 x^{h_1},\, \ldots,\, \lambda_n x^{h_n}\bigr).$$

Definition 6. A *discriminantal triple* $(A, \Delta, \mathbf{m})$ consists of

1. an $r \times m$ integer matrix A of rank r having $(1,1,\ldots,1)$ in its row span,
2. an A-homogeneous polynomial Δ that vanishes on the dual toric variety Y_A^*,
3. a distinguished term $\mathbf{m}$ among those that occur in the polynomial Δ,

such that $H_{(\Delta,\mathbf{m})} = \tilde{H}_{(\Delta,\mathbf{m})}$, the sign vector $\sigma \coloneqq \mathrm{sign}(H_{(\Delta,\mathbf{m})} \cdot u)$ is the same for all positive column vectors $u \in \mathbb{R}^{n+1}_{>0}$, and it satisfies the condition

$$\lambda_i \cdot \sigma^{h_i} > 0 \text{ for all } i = 1, 2, \ldots, m. \tag{12}$$

Remark 7. Let $(A, \Delta, \mathbf{m})$ be a triple as in Definition 6, 1–3, such that for its associated Horn matrix $H_{(\Delta,\mathbf{m})}$ we have that $\mathrm{sign}(\tilde{H}_{(\Delta,\mathbf{m})} \cdot u)$ is the same for all positive u and the pair $(\tilde{H}_{(\Delta,\mathbf{m})}, \tilde{\lambda})$ satisfies (12). As in Remark 3 we associate to $(A, \Delta, \mathbf{m})$ a discriminantal triple $(\tilde{A}, \tilde{\Delta}, \tilde{\mathbf{m}})$ with $\mathrm{Im}\, \phi_{(\Delta,\mathbf{m})} = \mathrm{Im}\, \phi_{(\tilde{\Delta},\tilde{\mathbf{m}})}$.

All definitions are now complete. We next illustrate Definition 6 for our running example.

Example 8. Let A be the 2×4 matrix in (10), $\Delta = \Delta_A$ its discriminant in (11), and $\mathbf{m} = 27x_1^2x_4^2$ the underlined term. Then $(A, \Delta, \mathbf{m})$ is a discriminantal triple with associated sign vector $\sigma = (-1, +1, +1, -1)$. The orthant $Y^*_{A,\sigma}$ was highlighted in Example 5. It is a semialgebraic surface inside $Y^*_A \subset \mathbb{RP}^3$. This surface is mapped into the tetrahedron Δ_3 by

$$\phi_{(\Delta,\mathbf{m})} \,:\, (x_1, x_2, x_3, x_4) \,\mapsto\, \left(\frac{2}{3}\frac{x_2x_3}{x_1x_4}, -\frac{4}{27}\frac{x_3^3}{x_1x_4^2}, -\frac{4}{27}\frac{x_2^3}{x_1^2x_4}, \frac{1}{27}\frac{x_2^2x_3^2}{x_1^2x_4^2}\right). \tag{13}$$

The image of this map is a curve in Δ_3, namely the model $\mathcal{M}$ in Example 4. We verify (1) by comparing (7) with (13). The former is obtained from the latter by setting $x = Hu$.

We close this section with two remarks on Horn matrices, Horn pairs and Horn maps.

3 Staged Trees

We consider contingency tables $u = (u_{i_1 i_2 \cdots i_m})$ of format $r_1 \times r_2 \times \cdots \times r_m$. Following [5, 14], these represent joint distributions of discrete statistical models with $n + 1 = r_1 r_2 \cdots r_m$ states. For any subset $C \subset \{1, \ldots, m\}$, one considers the marginal table u_C that is obtained by summing out all indices not in C. The entries of the marginal table u_C are sums of entries in u. Namely, to obtain the entry $u_{I,C}$ of u_C for any state $I = (i_1, i_2, \ldots, i_m)$, we fix the indices of the states in C and sum over the indices not in C. For example, if $m = 4$, $C = \{1, 3\}$, $I = (i, j, k, l)$, then u_C is the $r_1 \times r_3$ matrix with entries

$$u_{I,C} \;=\; u_{i+k+} \;=\; \sum_{j=1}^{r_2}\sum_{l=1}^{r_4} u_{ijkl}.$$

Such linear forms are the basic building blocks for the familiar models with rational MLE.

Consider an undirected graph G with vertex set $\{1, \ldots, m\}$ which is assumed to be *chordal.* The associated *decomposable graphical model* $\mathcal{M}_G$ in Δ_n has the rational MLE

$$\hat{p}_I \;=\; \frac{\prod_C u_{I,C}}{\prod_S u_{I,S}}, \tag{14}$$

where the product in the numerator is over all maximal cliques C of G, and the product in the denominator is over all separators S in a junction tree for G. See [14, §4.4.1]. In what follows we regard G as a directed graph, with edge directions given by a perfect elimination ordering on the vertex set $\{1, \ldots, m\}$. This turns $\mathcal{M}_G$ into a Bayesian network.

More generally, a *Bayesian network* $\mathcal{M}_G$ is given by a directed acyclic graph G. We write $\mathrm{pa}(j)$ for the set of parents of the node j. The model $\mathcal{M}_G$ in Δ_n has the rational MLE

$$\hat{p}_I \;=\; \prod_{j=1}^{m} \frac{u_{I,\mathrm{pa}(j)\cup\{j\}}}{u_{I,\mathrm{pa}(j)}}. \tag{15}$$

If G comes from an undirected chordal graph then (14) arises from (15) by cancellations.

Example 9 ($m = 4$)**.** We revisit two examples that were discussed on page 36 in [5, §2.1]. The *star graph* $G = [14][24][34]$ is chordal. The MLE for $\mathcal{M}_G$ is the map Φ with coordinates

$$\hat{p}_{ijkl} \;=\; \frac{u_{i++l} \cdot u_{+j+l} \cdot u_{++kl}}{u_{++++} \cdot u_{+++l}^2} \;=\; \frac{u_{i+++}}{u_{++++}} \cdot \frac{u_{+j+l}}{u_{+++l}} \cdot \frac{u_{++kl}}{u_{+++l}} \cdot \frac{u_{i++l}}{u_{i+++}}.$$

The left expression is (14). The right is (15) for the directed graph $1 \to 4$, $4 \to 2$, $4 \to 3$.

The *chain graph* $G = [12][23][34]$ is chordal. Its MLE is the map Φ with coordinates

$$\hat{p}_{ijkl} \;=\; \frac{u_{i++l} \cdot u_{+jk+} \cdot u_{++kl}}{u_{+j++} \cdot u_{++k+}} \;=\; \varphi_{(H,\lambda)}(u)_{ijkl}.$$

This is the Horn map in Proposition 12, given by the specific pair (H, λ) in Example 11.

The formulas (14) and (15) are familiar to statisticians. Theorem 1 places them into a larger context. However, some readers may find our approach too algebraic and too general. Our aim in this section is lay out a useful middle ground: models given by staged trees.

Staged trees were introduced by Smith and Anderson [16] as a generalization of discrete Bayesian networks. They furnish an intuitive representation of many situations that the above graphs G cannot capture. In spite of their wide scope, staged tree models are appealing because of their intuitive formalism for encoding events. We refer to the textbook [3] for an introduction. In what follows we study parts (1) and (2) in Theorem 1 for staged trees.

To define a *staged tree model*, we start with a directed rooted tree $\mathcal{T}$ having at least two edges emanating from each non-leaf vertex, a label set $S = \{s_i \mid i \in I\}$, and a labeling $\theta\colon \mathrm{E}(\mathcal{T}) \to S$ of the edges of the tree. Each vertex of $\mathcal{T}$ has a corresponding *floret*, which is the multiset of edge labels emanating from it. The labeled tree $\mathcal{T}$ is a *staged tree* if any two florets are either equal or disjoint. Two vertices in $\mathcal{T}$ are in the same stage if their corresponding florets are the same. From this point on, F denotes the set of florets of $\mathcal{T}$.

Definition 10. Let J be the set of root-to-leaf paths in the tree $\mathcal{T}$. We set $|J| = n + 1$. For $i \in I$ and $j \in J$, let μ_{ij} denote the number of times edge label s_i appears in the j-th root-to-leaf path. The *staged tree model* $\mathcal{M}_\mathcal{T}$ is the image of the parametrization

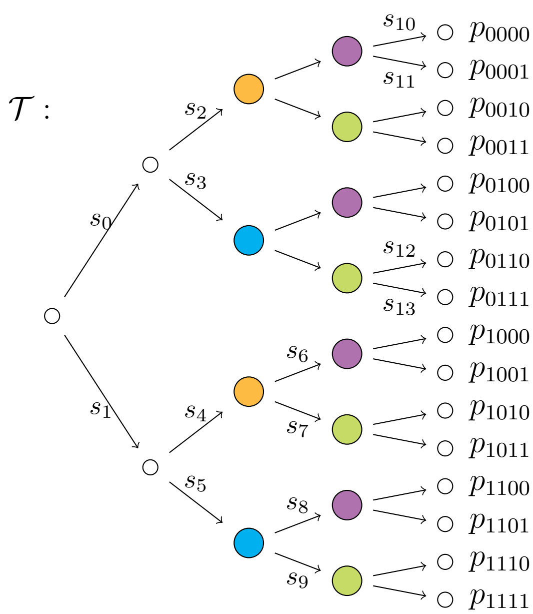
$$\phi_\mathcal{T} : \Theta \to \Delta_n\,,\;\; (s_i)_{i\in I} \mapsto (p_j)_{j\in J},$$

where $\Theta := \Big\{(s_i)_{i\in I} \in (0,1)^{|I|} : \sum_{s_i\in f} s_i = 1 \text{ for all florets } f \in F\Big\}$ is the parameter space of $\mathcal{M}_\mathcal{T}$, and $p_j = \prod_{i\in I} s_i^{\mu_{ij}}$ is the product of the edge parameters on the j-th root-to-leaf path.

In the model $\mathcal{M}_\mathcal{T}$, the tree $\mathcal{T}$ represents possible sequences of events. The parameter s_i associated to an edge vv' is the transition probability from v to v'. All parameter labels in a floret sum to 1. The fact that distinct nodes in $\mathcal{T}$ can have the same floret of parameter labels enables staged tree models to encode conditional independence statements [16]. This property allows us to represent any discrete Bayesian network or decomposable model as a staged tree model. Our first staged tree was seen in Example 2. Here is another specimen.

Example 11 ($n = 15$)**.** Consider the decomposable model for binary variables given by the 4-chain $G = [12][23][34]$. Figure 1 shows a realization of $\mathcal{M}_G$ as a staged tree model $\mathcal{M}_\mathcal{T}$. The leaves of $\mathcal{T}$ represent the outcome space $\{0,1\}^4$. Nodes with the same color have the same associated floret. The blank nodes all have different florets. The seven florets of $\mathcal{T}$ are

$$f_1{=}\{s_0,s_1\},\ f_2{=}\{s_2,s_3\},\ f_3{=}\{s_4,s_5\},\ f_4{=}\{s_6,s_7\},\ f_5{=}\{s_8,s_9\},\ f_6{=}\{s_{10},s_{11}\},\ f_7{=}\{s_{12},s_{13}\}.$$



$$H = \begin{array}{c} s_0 \\ s_1 \\ f_1 \\ s_2 \\ s_3 \\ f_2 \\ s_4 \\ s_5 \\ f_3 \\ s_6 \\ s_7 \\ f_4 \\ s_8 \\ s_9 \\ f_5 \\ s_{10} \\ s_{11} \\ f_6 \\ s_{12} \\ s_{13} \\ f_7 \end{array} \begin{pmatrix}
1&1&1&1&1&1&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&1&1&1&1&1&1 \\
-&-&-&-&-&-&-&-&-&-&-&-&-&-&-&- \\
1&1&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&1&1&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot \\
-&-&-&-&-&-&-&-&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&1&1&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&1&1 \\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&-&-&-&-&-&-&-&- \\
1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&\cdot&\cdot&\cdot&\cdot \\
-&-&-&-&\cdot&\cdot&\cdot&\cdot&-&-&-&-&\cdot&\cdot&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&\cdot&\cdot \\
\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&1 \\
\cdot&\cdot&\cdot&\cdot&-&-&-&-&\cdot&\cdot&\cdot&\cdot&-&-&-&- \\
1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot \\
\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot \\
-&-&\cdot&\cdot&-&-&\cdot&\cdot&-&-&\cdot&\cdot&-&-&\cdot&\cdot \\
\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot \\
\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&1 \\
\cdot&\cdot&-&-&\cdot&\cdot&-&-&\cdot&\cdot&-&-&\cdot&\cdot&-&-
\end{pmatrix}$$

Figure 1: A staged tree $\mathcal{T}$ and its Horn matrix H in Proposition 12. Entries $-$ indicate -1.

Next we show that staged tree models have rational MLE, so they satisfy part (1) of Theorem 1. Our formula for Φ uses the notation for I, J and μ_{ij} introduced in Definition 10. This formula is known in the literature on chain event graphs (see e.g. [15]).

Proposition 12. *Let $\mathcal{M}_\mathcal{T}$ be a staged tree model, and let $u = (u_j)_{j\in J}$ be a vector of counts. For $i \in I$, let f be the floret containing the label s_i, and define the estimates*

$$\hat{s}_i \;:=\; \frac{\sum_j \mu_{ij} u_j}{\sum_{s_\ell \in f} \sum_j \mu_{\ell j} u_j} \quad \text{and} \quad \hat{p}_j \;:=\; \prod_{i\in I} (\hat{s}_i)^{\mu_{ij}}.$$

The rational function Φ that sends $(u_j)_{j\in J}$ to $(\hat{p}_j)_{j\in J}$ is the MLE of the model $\mathcal{M}_\mathcal{T}$.

Proof. We prove that the likelihood function $L(p,u)$ has a unique maximum at $p = (\hat{p}_j)_{j\in J}$. For $f \in F$, we fix the vector of parameters $s_f = (s_i)_{s_i\in f}$. Associated with the floret f is the local likelihood function $L_f(s_f,u) = \prod_{s_i\in f} s_i^{\alpha_i}$, where $\alpha_i = \sum_j \mu_{ij}u_j$. We have

$$L(p,u) \;=\; \prod_j p_j^{u_j} \;=\; \prod_j \prod_i s_i^{u_j\mu_{ij}} \;=\; \prod_i s_i^{\alpha_i} \;=\; \prod_{f\in F} L_f(s_f,u).$$

Since the L_f depend on disjoint sets of unknowns, maximizing L is achieved by maximizing the factors L_f separately. But L_f is the likelihood function of the full model $\Delta_{|f|-1}$, given the data vector $(\alpha_i)_{s_i\in f}$. The MLE of that model is given by $\hat{s}_i = \alpha_i/\sum_{s_\ell\in f}\alpha_\ell$, where $s_i \in f$. We conclude that $\operatorname{argmax}_{s_f}\big(L_f(s_f,u)\big) = (\hat{s}_i)_{s_i\in f}$ and $\operatorname{argmax}_p\big(L(p,u)\big) = (\hat{p}_j)_{j\in J}$. $\square$

Remark 13. Here is a method for evaluating the MLE in Proposition 12. Let $[v] \subset J$ be the set of root-to-leaf paths through a fixed node v in the tree $\mathcal{T}$ and define $u_{[v]} = \sum_{j\in[v]} u_j$. The quotient $\frac{u_{[v']}}{u_{[v]}}$ is the empirical transition probability from v to v' given arrival at v. To obtain $\hat{s}_i$ we first compute the quotients $\frac{u_{[v']}}{u_{[v]}}$ for all edges vv' with parameter label s_i. Then we aggregate them by adding their numerators and denominators separately. We obtain $s_i = (\sum u_{[v']})/(\sum u_{[v]})$, where both sums range over all edges vv' with parameter label s_i.

Proposition 12 yields an explicit description of the Horn pair $(\tilde{H},\tilde{\lambda})$ associated to $\mathcal{M}_\mathcal{T}$.

Corollary 14. *Fix a staged tree model $\mathcal{M}_\mathcal{T}$ as above. Let H be the $(|I|+|F|)\times|J|$ matrix whose rows are indexed by the set $I \sqcup F$ and entries are given by*

$$\begin{aligned} h_{ij} \;&=\; \mu_{ij} \text{ for } i \in I, \text{ and} \\ h_{fj} \;&=\; -\sum_{s_\ell\in f}\mu_{\ell j} \text{ for } f \in F. \end{aligned}$$

Define the vector $\lambda \in \{-1,+1\}^{|J|}$ by $\lambda_j = (-1)^{\sum_f h_{fj}}$. Then $(\tilde{H},\tilde{\lambda})$ is the Horn pair of $\mathcal{M}_\mathcal{T}$, using the mapping $H \mapsto \tilde{H}$ defined in Remark 3.

Given a staged tree $\mathcal{T}$, we call the matrix H in Corollary 14 the *Horn matrix* of $\mathcal{T}$.

Remark 15. In Corollary 14, for a floret $f \in F$, let H_f be the submatrix of H with row indices $\{i : s_i \in f\}\cup\{f\}$. Then H is the vertical concatenation of the matrices H_f for $f \in F$. The matrix $\tilde{H}$ is obtained from H by the row operations described in Remark 3.

Example 16. For the tree $\mathcal{T}$ in Example 11, the Horn matrix H of $\mathcal{M}_\mathcal{T}$ is given in Figure 1. Its rows indices are $(s_0,s_1,f_1,s_2,s_3,f_2,s_4,s_5,f_3,s_6,s_7,f_4,s_8,s_9,f_5,s_{10},s_{11},f_6,s_{12},s_{13},f_7)$. The vector λ for the friendly Horn matrix H is the vector of ones $(1,\ldots,1)\in\mathbb{R}^{16}$. Note that (H,λ) is not a Horn pair. We can delete the rows s_0,s_1,f_2,f_3 of the matrix H by summing the pairs (s_0,f_2) and (s_1,f_3) and deleting zero rows. The result is the Horn pair $(\tilde{H},\tilde{\lambda})$.

Following [8], two staged trees $\mathcal{T}$ and $\mathcal{T}'$ are called *statistically equivalent* if there exists a bijection between the sets of root-to-leaf paths of $\mathcal{T}$ and $\mathcal{T}'$ such that, after applying this bijection, $\mathcal{M}_\mathcal{T} = \mathcal{M}_{\mathcal{T}'}$ inside the open simplex Δ_n. Any staged tree model may have different but statisticaly equivalent tree representations. In [8, Theorem 1], the authors show that statistical equivalence of staged trees can be determined by doing a sequence of operations on the trees, named *swap* and *resize*. One of the advantages of describing a staged tree model via its Horn pair is that it gives a new criterion to decide whether two staged trees are statistically equivalent. This is simpler to implement than the criterion formulated in [8].

Corollary 17. *Two staged trees are statistically equivalent if and only if their their Horn pairs* $(\tilde{H}, \tilde{\lambda})$ *agree.*

One natural operation on a staged tree $\mathcal{T}$ is identifying two florets of the same size. This gives a new tree $\mathcal{T}'$ and model $\mathcal{M}_{\mathcal{T}'}$ whose Horn matrix is readily obtained from that of $\mathcal{T}$.

Corollary 18. *Let $\mathcal{T}'$ be a staged tree arising from $\mathcal{T}$ by identifying two florets f and f', say by the bijection $(-)' \colon f \to f'$. Then the Horn matrix H' of $\mathcal{M}_{\mathcal{T}'}$ arises from the Horn matrix H of $\mathcal{M}_\mathcal{T}$ by replacing the blocks H_f and $H_{f'}$ in H by the block H'_f defined by*

$$\begin{aligned} h'_{ij} &= h_{ij} + h_{i'j} \quad \textit{for } s_i \in f, \\ h'_{fj} &= h_{fj} + h_{f'j}. \end{aligned}$$

Proof. This follows from the definition of the Horn matrices for $\mathcal{M}_\mathcal{T}$ and $\mathcal{M}_{\mathcal{T}'}$. □

Example 19. Let $\mathcal{T}'$ be the tree obtained from $\mathcal{T}$ in Example 11 by identifying the florets f_4 and f_5. Then $\mathcal{M}_{\mathcal{T}'}$ is the independence model of two random variables with four states.

Now we turn to part (3) of Theorem 1. We describe the triple $(A, \Delta, \mathbf{m})$ for a staged tree model $\mathcal{M}_\mathcal{T}$ giving rise to its discriminantal triple $(\tilde{A}, \tilde{\Delta}, \tilde{\mathbf{m}})$ as in Remark 7. The pair (H, λ) was given in Corollary 14. Let A be any matrix whose rows span the left kernel of H, set $m = |I| + |J|$, and write s for the m-tuple of parameters $(s_i, s_f)_{i \in I, f \in F}$. From the Horn matrix in Corollary 14 we see that

$$\Delta = \mathbf{m} \cdot \left(1 - \sum_j (-1)^{\epsilon_j} \prod_i \left(\frac{s_i}{s_f} \right)^{\mu_{ij}} \right),$$

where f depends on i, $\mathbf{m} = \operatorname{lcm}(\prod_i s_f^{\mu_{ij}} : f \in F)$ and $\epsilon_j = \sum_i \mu_{ij}$. The sign vector σ for the triple $(A, \Delta, \mathbf{m})$ is given by $\sigma_i = +1$ for $i \in I$ and $\sigma_f = -1$ for $f \in F$. Then $Y^*_{A,\sigma}$ gets mapped to $\mathcal{M}_\mathcal{T}$ via $\phi_{(\Delta,\mathbf{m})}$. Moreover, the map $\phi_\mathcal{T}$ from Definition 10 factors through $\phi_{(\Delta,\mathbf{m})}$. Indeed, if we define $\iota : \Theta \to Y^*_{A,\sigma}$ by $(s_i)_{i \in I} \mapsto (s_i, -1)_{i \in I, f \in F}$, then $\phi_\mathcal{T} = \phi_{(\Delta,\mathbf{m})} \circ \iota$.

The derivation in the following example is an extension of that in [11, Example 3.13].

Example 20. Let $\mathcal{M}_\mathcal{T}$ be the 4-chain model in Example 11. Its associated discriminant is

$$\begin{aligned} \Delta = f_1 f_2 f_3 f_4 f_5 f_6 f_7 &- s_0 s_2 s_6 s_{10} f_3 f_5 f_7 - s_0 s_2 s_6 s_{11} f_3 f_5 f_7 - s_0 s_2 s_7 s_{12} f_3 f_5 f_6 - s_0 s_2 s_7 s_{13} f_3 f_5 f_6 \\ &- s_0 s_3 s_8 s_{10} f_3 f_4 f_7 - s_0 s_3 s_8 s_{11} f_3 f_4 f_7 - s_0 s_3 s_9 s_{12} f_3 f_4 f_6 - s_0 s_3 s_9 s_{13} f_3 f_4 f_6 \\ &- s_1 s_4 s_6 s_{10} f_2 f_5 f_7 - s_1 s_4 s_6 s_{11} f_2 f_5 f_7 - s_1 s_4 s_7 s_{12} f_2 f_5 f_6 - s_1 s_4 s_7 s_{13} f_2 f_5 f_6 \\ &- s_1 s_5 s_8 s_{10} f_2 f_4 f_7 - s_1 s_5 s_8 s_{11} f_2 f_4 f_7 - s_1 s_5 s_9 s_{12} f_2 f_4 f_6 - s_1 s_5 s_9 s_{13} f_2 f_4 f_6. \end{aligned}$$

Our notation for the parameters matches the row labels of the Horn matrix H in Figure 1. This polynomial of degree 7 is irreducible, so it equals the A-discriminant: $\Delta = \Delta_A$. The underlying matrix A has format 13×21, and we represent it by its associated toric ideal

$$\begin{aligned} I_A \;\; = \;\; & \big\langle\, s_{10} - s_{11}\,,\ s_1 s_5 f_2 - s_0 s_3 f_3\,,\ s_1 s_4 f_2 - s_0 s_2 f_3\,,\ s_5 s_9 f_4 - s_4 s_7 f_5\,,\ s_3 s_9 f_4 - s_2 s_7 f_5, \\ & s_{12} - s_{13},\ s_5 s_8 f_4 - s_4 s_6 f_5, s_3 s_8 f_4 - s_2 s_6 f_5,\ s_9 s_{13} f_6 - s_8 s_{11} f_7,\ s_7 s_{13} f_6 - s_6 s_{11} f_7, \\ & s_0 s_2 s_6 s_{11} - f_1 f_2 f_4 f_6, s_0 s_2 s_7 s_{13} - f_1 f_2 f_4 f_7, s_0 s_3 s_8 s_{11} - f_1 f_2 f_5 f_6, s_0 s_3 s_9 s_{13} - f_1 f_2 f_5 f_7, \\ & s_1 s_4 s_6 s_{11} - f_1 f_3 f_4 f_6, s_1 s_4 s_7 s_{13} - f_1 f_3 f_4 f_7, s_1 s_5 s_9 s_{13} - f_1 f_3 f_5 f_7, s_1 s_5 s_8 s_{11} - f_1 f_3 f_5 f_6 \big\rangle. \end{aligned}$$

The toric variety $Y_A = \mathcal{V}(I_A)$ has dimension 12 and degree 141. It lives in a linear space of codimension 2 in $\mathbb{P}^{20}$, where it is defined by eight cubics and eight quartics. The dual variety $Y_A^* = \mathcal{V}(\Delta_A)$ is the above hypersurface of degree seven. We have $\mathbf{m} = f_1 f_2 f_3 f_4 f_5 f_6 f_7$, and σ is the vector in $\{-1, +1\}^{21}$ that has entry $+1$ at the indices corresponding to the s_i and entry -1 at the indices corresponding to the f_i. To obtain the discriminant $\tilde{\Delta}$ associated to the Horn pair in Example 16, we substitute 1 for s_0, s_1, f_2, f_3 in the polynomial Δ and change all the minus signs to plus signs. See also the discussion in Remark 3.

It would be interesting to study the combinatorics of the discriminantal triples for staged tree models. Our computations suggest that, for many such models, the polynomial Δ is irreducible and is equal to the A-discriminant Δ_A of the underlying configuration A. However, this is not true for all staged trees, as seen in equation (2) of Example 2. We close this section with a familiar class of models with rational MLE whose associated Δ factor.

Example 21. The *multinomial distribution* encodes the experiment of rolling a k-sided die m times. The associated model $\mathcal{M}$ is the independence model for m identically distributed random variables on k states. We have $n + 1 = \binom{k+m-1}{m}$. The Horn matrix H is the $(k+1) \times (n+1)$ matrix whose columns are the vectors $(-m, i_1, i_2, \ldots, i_k)^T$ where $i_1, i_2, \ldots, i_k$ are nonnegative integers whose sum equals m. Here, $A = (1\ 1\ 1 \cdots 1)$, so the A-discriminant is the linear form $\Delta_A = x_0 + x_1 + \cdots + x_k$. The following polynomial is a multiple of Δ_A:

$$\Delta \;=\; (-x_0)^m - (x_1 + x_2 + \cdots + x_k)^m.$$

This Δ, with its marked term $\mathbf{m} = (-x_0)^m$, encodes the MLE for the model $\mathcal{M}$.

4 Proof of the Main Theorem

In this section we prove Theorem 1. This involves making precise how the objects in the three parts correspond to each other. Namely, models with rational MLE correspond to Horn pairs (H, λ), and these correspond to pairs $(\Delta, \mathbf{m})$ in a discriminantal triple.

For a pair (H, λ) consisting of a Horn matrix H and a coefficient vector λ, we denote by φ the rational map defined in (4). We recall that its i-th coordinate is

$$\varphi_i(v) \;=\; \lambda_i \prod_{j=1}^{m} \left(\sum_{k=0}^{n} h_{jk} v_k \right)^{h_{ji}}. \tag{16}$$

We also define the likelihood function $L_u\colon \mathbb{R}^{n+1} \to \mathbb{R}$ for the image of φ:

$$L_u(v) \coloneqq \prod_{i=0}^{n} \varphi_i(v)^{u_i}. \tag{17}$$

Here $u \in \mathbb{N}^{n+1}$ is an arbitrary fixed data vector. We start with the following key lemma.

Lemma 22. *Let $H = (h_{ij})$ be a Horn matrix, λ a vector satisfying (3), and $u \in \mathbb{N}^{n+1}$. The vector u is the unique critical point of its own likelihood function L_u, up to scaling.*

Proof. We compute the partial derivatives of L_u. For $\ell = 0, \ldots, n$ we find

$$\begin{aligned}
\frac{\partial}{\partial v_\ell} L_u(v) &= \sum_{i=0}^{n} u_i \frac{L_u(v)}{\varphi_i(v)} \frac{\partial}{\partial v_\ell} \varphi_i(v) \\
&= \sum_{i=0}^{n} u_i \frac{L_u(v)}{\varphi_i(v)} \sum_{j=1}^{m} h_{ji} \frac{\varphi_i(v)}{\sum_{k=0}^{n} h_{jk} v_k} h_{j\ell} \\
&= L_u(v) \sum_{j=1}^{m} \sum_{i=0}^{n} \frac{u_i\, h_{ji}\, h_{j\ell}}{\sum_{k=0}^{n} h_{jk} v_k} \quad = \quad L_u(v) \sum_{j=1}^{m} \frac{h_{j\ell} \sum_{i=0}^{n} h_{ji} u_i}{\sum_{k=0}^{n} h_{jk} v_k}.
\end{aligned}$$

For $v = u$, this evaluates to zero, since the sums in the fraction cancel and the ℓ-th column of H sums to zero. The uniqueness of the critical point up to scaling follows from the fact that the projective variety given by the image of φ has ML-degree one, by [10, Theorem 1]. $\square$

We use [10] to explain the relation between models with rational MLE and Horn pairs.

Proof of Theorem 1, Equivalence of (1) and (2). Let $\mathcal{M}$ be a model with rational MLE Φ. The Zariski closure of $\mathcal{M}$ is a variety of ML-degree one. By [10, Theorem 1], there exists a Horn matrix H and a coefficient vector λ such that $\varphi = \Phi$. Now, the required sum-to-one and positivity conditions for φ are satisfied because they are satisfied by the MLE Φ. Indeed, the MLE of any discrete statistical model maps positive vectors u in $\mathbb{R}^{n+1}_{>0}$ into the simplex Δ_n. Conversely, we claim that every Horn pair (H, λ) specifies a nonempty model $\mathcal{M}$ with rational MLE. Indeed, define $\mathcal{M}$ to be the image of $\varphi_{(H,\lambda)}$. By the defining properties of the Horn pair, we have $\mathcal{M} \subset \Delta_n$. Lemma 22 shows that $\varphi_{(H,\lambda)}$ is the MLE of $\mathcal{M}$. $\square$

Next, we relate Horn pairs to discriminantal triples $(A, \Delta, \mathbf{m})$. The pair $(\Delta, \mathbf{m})$ is the data that defines $\mathcal{M}$ as an algebraic variety. The matrix A and the derived sign vector σ are witnesses of special properties of $(\Delta, \mathbf{m})$. Namely, the polynomial Δ is A-homogeneous and vanishes on some dual toric variety, Y_A^*, whose σ-orthant maps onto the model $\mathcal{M}$ via the map $\phi_{(\Delta,\mathbf{m})}$. The positivity condition of a Horn pair is supposed to translate into the positivity condition in (12). This translation is a consequence of the following key lemma.

Lemma 23. *Let (H, λ) be a friendly pair. If there exists a vector $u \in \mathbb{R}^{n+1}$ such that $\varphi(u) > 0$, then we have $\varphi(v) > 0$ for all v in $\mathbb{R}^{n+1}_{>0}$ where it is defined.*

Proof. The function φ is homogeneous of degree zero. It suffices to prove each coordinate of $\varphi(v)$ is a positive real number, for all vectors v with positive integer entries. Indeed, every positive v in $\mathbb{R}^{n+1}$ can be approximated by rational vectors, which can be scaled to be integral. The open subset $U = \varphi^{-1}(\Delta_n)$ of $\mathbb{R}^{n+1}$ contains u. If $U = \mathbb{R}^{n+1}$, then we are done. Else, U has a nonempty boundary ∂U. By continuity, $\partial U \subseteq \varphi^{-1}(\partial \Delta_n)$. The likelihood function L_v for the data vector v vanishes on ∂U.

We claim that L_v has a critical point in U. The closed subset $\overline{U}$ is homogeneous. After passing to projective space $\mathbb{P}^n$, it becomes compact. The likelihood function L_v is well defined on this compact set in $\mathbb{P}^n$, since it is homogeneous of degree zero, and L_v vanishes on the boundary. Hence the restriction $L_v|_U$ is either identically zero or it has a critical point in U. But, since $u \in U$ is a point with $L_v(u) \neq 0$, the second statement must be true.

Since U is an open subset of $\mathbb{R}^{n+1}$, a critical point of the restriction $L_v|_U$ is also a critical point of the function L_v itself. By Lemma 22, this critical point must be v. Hence $v \in U$. $\square$

Corollary 24. *Let (H, λ) be a friendly pair, with $H = \tilde{H}$ as in Remark 3. Fix any positive vector u in $\mathbb{R}^{n+1}_{>0}$. Then (H, λ) is a Horn pair if and only if $\lambda_i (Hu)^{h_i} > 0$ for $i = 0, 1 \ldots, n$. If this holds then the nonzero entries in each row of H have the same sign. In particular, the sign vector $\sigma = \mathrm{sign}(Hu)$ is independent of the choice of u.*

Proof. The coordinates of Hv are the linear factors of the numerators and denominators of $\varphi(v)$. We have shown in Lemma 23 that none of these numerators or denominators vanish on Δ_n, and hence the same holds for the coordinates of Hv. This implies that the rows of H have the desired sign property. The characterization of Horn pairs now follows from (4). $\square$

We prove the rest of Theorem 1 by first explaining how to turn (H, λ) into a pair $(\Delta, \mathbf{m})$ and then examining how the constraints on Horn pairs and discriminantal triples are related.

Proof of Theorem 1, Equivalence of (2) and (3). Let (H, λ) be a pair consisting of a Horn matrix and a coefficient vector. We construct a pair $(\Delta, \mathbf{m})$ consisting of a polynomial Δ and a monomial $\mathbf{m}$ appearing in Δ as follows. For $k = 0, \ldots, n+1$ let h_k denote the columns of H, and write h_k^+ resp. h_k^- for the positive part resp. the negative part of h_k, so that $h_k = h_k^+ - h_k^-$. In addition, let $\max_k(h_k^-)$ be the entrywise maximum of the h_k^-. We define

$$\mathbf{m} \,=\, x^{\max_k(h_k^-)} \quad \text{and} \quad \Delta \,=\, \mathbf{m} \cdot \left(1 - \sum_{k=0}^{n} \lambda_k x^{h_k}\right). \tag{18}$$

Conversely, from any pair $(\Delta, \mathbf{m})$ as above, we construct a pair (H, λ) by the equation on the right hand side. This specifies $H = (h_k)_k$ and $\lambda = (\lambda_k)_k$ uniquely. We next proceed with comparing the defining properties for Horn pairs with those for discriminantal triples.

Claim. *If (H, λ) is friendly and if the r columns of an integer matrix A with $AH = 0$ span $\mathbb{Z}^r$, then Δ is A-homogeneous and vanishes on the dual toric variety Y_A^*. Conversely, if Δ is A-homogeneous and vanishes on Y_A^* for some integer matrix A, then (H, λ) is friendly.*

Proof of Claim. Let (H, λ) be friendly and A a matrix as above. The Laurent polynomial $q := \Delta/\mathbf{m}$ from (18) is a rational function on $\mathbb{P}^{m-1}$ that vanishes on the dual toric variety

Y_A^*. To see this, consider the exponentiation map $\varphi_2\colon \mathbb{P}^{m-1} \to \mathbb{R}^{n+1}$ that is defined by $\varphi_2(x) = \lambda * x^H$, where $*$ is the entrywise product. Let $f = 1 - (p_0 + \cdots + p_n)$. We have $q = f \circ \varphi_2$. By [10, Theorems 1 and 2], the function φ_2 maps an open dense subset of Y_A^* dominantly to the closure $\mathcal{M}$ of the image of $\varphi_{(H,\lambda)}$. Since $f = 0$ on $\mathcal{M}$, we have $f \circ \varphi_2 = 0$ on an open dense subset of Y_A^*, hence $q = 0$ on Y_A^*, so $\Delta = 0$ there as well.

Conversely, let Δ be A-homogeneous and vanish on Y_A^* for some A. We claim that $q(x)$ is zero for all $x = Hu$ in the image of the linear map H. We may assume $\mathbf{m}(x) \neq 0$. We must prove that x is in the dual toric variety Y_A^*, since Δ vanishes on it. So, let $x_i = \sum_{j=0}^n h_{ij}u_j$ for $i = 0, \ldots n+1$. We claim that $t = (1, \ldots, 1)$ is a singular point of the hypersurface

$$\gamma_A^{-1}(H_x \cap Y_A) \;=\; \left\{ t \in \mathbb{C}^r \mid \sum_{i=1}^m x_i t^{a_i} = 0 \right\}.$$

First, the point t lies on that hypersurface since the columns of H sum to zero:

$$\sum_{i=1}^m x_i \;=\; \sum_{i=1}^m \sum_{j=0}^n h_{ij}u_j \;=\; \sum_{j=0}^m u_j \sum_{i=1}^m h_{ij} \;=\; 0.$$

For $s = 1, \ldots, r$ we have $\frac{\partial}{\partial t_s} t^{a_i} = a_{si} t^{a_i - e_s}$, with e_s the s-th canonical basis vector of $\mathbb{Z}^r$, and

$$\frac{\partial}{\partial t_s} \sum_{i=1}^n x_i t^{a_i} \;=\; \sum_{i=1}^m \sum_{j=0}^n h_{ij} u_j a_{si} t^{a_i - e_s} \;=\; \sum_{j=0}^n u_j \sum_{i=1}^m a_{si} h_{ij} t^{a_i - e_s}.$$

This is zero at $t = (1, \ldots, 1)$ because $AH = 0$. ■

Next comes the point where we incorporate positivity. If a friendly pair (H, λ) with $H = \tilde{H}$ is a Horn pair then the sign vector σ satisfies (12). But conversely, if $(A, \Delta, \mathbf{m})$ is a discriminantal triple then (12) holds, and Corollary 24 tells us that (H, λ) is a Horn pair. To complete the proof, let $\phi_{(\Delta,\mathbf{m})}(x) \coloneqq \lambda * x^H$. We have $\overline{\varphi_{(H,\lambda)}(\mathbb{R}^{n+1}_{>0})} = \overline{\phi_{(\Delta,\mathbf{m})}(Y^*_{A,\sigma})}$ by [10, Theorems 1 and 2], and we have $\varphi_{(H,\lambda)} = \phi_{(\Delta,\mathbf{m})} \circ H$ by construction. □

We proved that every model with rational MLE **arises from** a toric variety Y_A. In some cases, the model **is itself** a toric variety Y_C. It is crucial to distinguish the two matrices A and C. The two toric structures are very different. For instance, every undirected graphical model is toric [5, Proposition 3.3.3]. The toric varieties Y_C among staged tree models $\mathcal{M}_\mathcal{T}$ were classified in [4]. The 4-chain model $\mathcal{M}_\mathcal{T} = Y_C$ **is itself** a toric variety of dimension 7 in $\mathbb{P}^{15}$. But it **arises from** a toric variety Y_A of dimension 12 in $\mathbb{P}^{20}$, as seen in Example 20.

Toric models with rational MLE play an important role in *geometric modeling* [2, 6]. Given an integer matrix $C \in \mathbb{Z}^{r \times (n+1)}$ and a vector of weights $w \in \mathbb{R}^{n+1}_{>0}$, one considers the *scaled projective toric variety* $Y_{C,w}$ in $\mathbb{RP}^n$. This is defined as the closure of the image of

$$\gamma_{C,w} \;:\; (\mathbb{R}^*)^r \to \mathbb{RP}^n\,,\;\; (t_1, \ldots, t_r) \;\mapsto\; \left(w_1 \prod_{i=1}^r t_i^{c_{i1}},\, w_2 \prod_{i=1}^r t_i^{c_{i2}},\, \ldots\,,\, w_m \prod_{i=1}^r t_i^{c_{im}} \right). \qquad (19)$$

The set of positive points in this toric variety is a discrete statistical model $\mathcal{M}_{C,w}$ in Δ_n. There is a natural homeomorphism from the toric model $\mathcal{M}_{C,w}$ onto the polytope of C. This is known among geometers as the *moment map*, and as *Birch's Theorem* in Algebraic Statistics. In geometric modeling the pair (C, w) is used to define *toric blending functions* [13].

It is highly desirable for the toric blending functions to have *rational linear precision* [2, 13]. The property is rare and it depends in a subtle way on (C, w). Garcia-Puente and Sottile [6] established the connection to algebraic statistics. They showed that rational linear precision holds for (C, w) if and only if the statistical model $\mathcal{M}_{C,w}$ has rational MLE.

Example 25. The most classical blending functions with rational linear precision live on the triangle $\{x \in \mathbb{R}^3_{>0} : x_1+x_2+x_3 = 1\}$. They are the *Bernstein basis polynomials* of degree m:

$$\frac{m!}{i!j!(m-i-j)!} x_1^i x_2^j x_3^{m-i-j} \quad \text{for} \quad i, j \geq 0,\, i+j \leq m. \tag{20}$$

Here C is the $3 \times \binom{m+1}{2}$ matrix whose columns are the vectors $(i, j, m-i-j)$. The weights are $w_{(i,j)} = \frac{m!}{i!j!(m-i-j)!}$. The associated toric model $\mathcal{M}_{C,w}$ is the multinomial family, where (20) is the probability of observing i times 1, j times 2 and $m-i-j$ times 3 in m trials. This model is seen in Example 21 and it has rational MLE. Again, notice the distinction between the two toric varieties. Here, Y_A is a point in $\mathbb{P}^m$, whereas Y_C is a surface in $\mathbb{P}^{\binom{m}{2}-1}$.

Clarke and Cox [2] raise the problem of characterizing all pairs (C, w) with rational linear precision. This was solved by Duarte and Görgen [4] for pairs arising from staged trees. While the problem remains open in general, our theory in this paper offers new tools. We may ask for a characterization of discriminantal triples whose models are toric.

5 Constructing Models with Rational MLE

Part (3) in Theorem 1 allows us to construct models with rational MLE starting from a matrix A that defines a projective toric variety Y_A. In most cases, the dual variety Y_A^* is a hypersurface, and we can compute its defining polynomial Δ_A, the *discriminant* [7]. The polynomial Δ in a discriminantal triple can be any homogeneous multiple of Δ_A, but we just take $\Delta = \Delta_A$ in Algorithm 1. For all terms $\mathbf{m}$ in Δ_A, we check whether $(A, \Delta_A, \mathbf{m})$ is a discriminantal triple and, if so, we identify σ. We implemented this algorithm in `Macaulay2`.

Lines 1 and 18 of Algorithm 1 are computations that rely on Gröbner bases. The execution of Line 18 can be very slow. It may be omitted if one is satisfied with obtaining the parametric description and MLE $\Phi^{(\ell)}$ of the model $\mathcal{M}_\ell$. For the check in Line 17, one does not need to compute $\Phi_i(v)$ numerically. Instead, one can just examine the signs and parities of the entries of H.

Example 26 $(r = 2, m = 4)$**.** For distinct positive integers α, β, γ with $\gcd(\alpha, \beta, \gamma) = 1$, let

$$A_{\alpha,\beta,\gamma} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & \alpha & \beta & \gamma \end{pmatrix}.$$

Algorithm 1: From toric varieties to statistical models

Input : An integer matrix A of size $r \times m$ with $(1, \ldots, 1)$ in its row span
Output: An integer n and a collection of statistical models $\mathcal{M}^{(\ell)} = (\Phi^{(\ell)}, I^{(\ell)})$, where $\Phi^{(\ell)} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is a rational MLE for $\mathcal{M}^{(\ell)}$, and $I^{(\ell)} \subseteq \mathbb{R}[p_0, \ldots, p_n]$ is the defining prime ideal of $\mathcal{M}^{(\ell)}$.

1 Compute the A-discriminant $\Delta_A \in \mathbb{Z}[x_1, \ldots, x_m]$;
2 $n \leftarrow \#\mathrm{terms}(\Delta_A) - 2$;
3 models $\leftarrow \{\}$;
4 **for** $0 \leq \ell \leq n$ **do**
5 $\quad \mathbf{m} \leftarrow \mathrm{terms}(\Delta_A)_\ell$;
6 $\quad q \leftarrow 1 - \Delta_A / \mathbf{m}$;
7 $\quad$ **for** $0 \leq i \leq n$ **do**
8 $\quad\quad \lambda_i \leftarrow \mathrm{coefficients}(q)_i$;
9 $\quad\quad h_i \leftarrow \mathrm{exponent\_vectors}(q)_i$;
10 $\quad\quad \Phi_i^{(\ell)} \leftarrow (u \mapsto \lambda_i \prod_{j=1}^m (\sum_{k=0}^n h_{jk} u_k)^{h_{ji}})$;
11 $\quad$ **end**
12 $\quad H \leftarrow (h_i)_i$;
13 $\quad$ **if** $\mathrm{sign}(\tilde{H})$ *is not well-defined* **then**
14 $\quad\quad$ discard this instance and **continue loop**;
15 $\quad$ **end**
16 $\quad$ Choose any positive vector v in $\mathbb{R}^{n+1}_{>0}$;
17 $\quad$ **if** $\Phi_i^{(\ell)}(v) > 0$ *for* $i = 0, 1, \ldots, n$ **then**
18 $\quad\quad$ Compute the ideal $I^{(\ell)}$ of the image of $\Phi^{(\ell)}$;
19 $\quad\quad$ models $\leftarrow$ models $\cup \{(\Phi^{(\ell)}, I^{(\ell)})\}$;
20 $\quad$ **end**
21 **end**
22 **return** models;

We ran Algorithm 1 for all 613 such matrices with $0 < \alpha < \beta < \gamma \leq 17$. Line 1 computes the discriminant Δ_A of the univariate polynomial $f(t) = x_1 + x_2 t^\alpha + x_3 t^\beta + x_4 t^\gamma$. The number $n+2$ of terms of these discriminants equals $7927/613 = 12.93$ on average. Thus a total of 7927 candidate triples $(A, \Delta_A, \mathbf{m})$ were tested in Lines 12 to 21. Precisely 123 of these were found to be discriminantal triples. This is a fraction of 1.55 %. In other words, only 1.55 % of the resulting complex varieties permitted by [10] are actually statistical models.

Here is a typical model that was discovered. Take $\alpha = 1, \beta = 4, \gamma = 7$. The discriminant

$$\begin{aligned} \Delta_A \;\; = \;\; & 729x_2^4x_3^6 - 6912x_1^3x_3^7 - 8748x_2^5x_3^4x_4 + 84672x_1^3x_2x_3^5x_4 + 34992x_2^6x_3^2x_4^2 \\ & -351918x_1^3x_2^2x_3^3x_4^2 - 46656x_2^7x_4^3 + 518616x_1^3x_2^3x_3x_4^3 \; \underline{- 823543x_1^6x_4^4} \end{aligned}$$

has 9 terms, so $n = 7$. The special term $\mathbf{m}$ is underlined. The associated model is a curve of degree ten in Δ_7. Its prime ideal $I^{(\ell)}$ is generated by 18 quadrics. Among them are 15 binomials that define a toric surface of degree six: $49p_1p_2 - 48p_0p_3, 3p_0p_4 - p_2^2, \ldots, 361p_3p_7 - 128p_5^2$.

Inside that surface, our curve is cut out by three other quadrics, like $26068p_2^2 + 73728p_0p_5 + 703836p_0p_6 + 234612p_2p_6 + 78204p_4p_6 + 612864p_0p_7 + 212268p_2p_7 + 78204p_4p_7 - 8379p_7^2$.

Example 27 $(r = 3, m = 6)$**.** For any positive integers $\alpha, \beta, \gamma, \varepsilon$, we consider the matrix

$$A = \begin{pmatrix} 0 & \alpha & \beta & 0 & \gamma & \varepsilon \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

The discriminant Δ_A is the *resultant* of two trinomials $x_1 + x_2 t^\alpha + x_3 t^\beta$ and $x_4 + x_5 t^\gamma + x_6 t^\varepsilon$ in one variable t. We ran Algorithm 1 for all 138 such matrices with $0 < \alpha < \beta \leq 17$, $0 < \gamma < \varepsilon \leq 17$, $\gcd(\alpha, \beta) = \gcd(\gamma, \varepsilon) = 1$. The number $n + 2$ of terms of these discriminants equals $2665/138 = 19.31$ on average. Thus a total of 2665 candidate triples $(A, \Delta_A, \mathbf{m})$ were tested in Line 13. Precisely 93 of these are discriminantal triples. This is a fraction of 3.49 %.

We now shift gears by looking at polynomials Δ that are multiples of the A-discriminant.

Example 28 $(r = 1, m = 4)$**.** We saw in Examples 2 and 21 that interesting models can arise from the matrix $A = (1\ 1\ \cdots\ 1)$ whose toric variety is just one point. Any homogeneous multiple Δ of the linear form $\Delta_A = x_1 + x_2 + \cdots + x_m$ can be used as input in Line 1 of Algorithm 1. Here, taking $\Delta = \Delta_A$ results in the model given by the full simplex Δ_{m-2}.

Let $m = 4$ and abbreviate $x^a = x_1^{a_1} x_2^{a_2} x_3^{a_3} x_4^{a_4}$ and $|a| = a_1 + a_2 + a_3 + a_4$ for $a \in \mathbb{N}^4$. We conducted experiments with two families of multiples. The first uses binomial multipliers:

$$\Delta \;=\; (x^a + x^b)\Delta_A \text{ or } (x^a - x^b)\Delta_A, \quad \text{where } |a| = |b| \in \{1, 2, \ldots, 8\} \text{ and } \gcd(x^a, x^b) = 1.$$

This gives 1028 polynomials Δ. The numbers of polynomials of degree $2, 3, 4, 6, 7, 8, 9, 10$ is 6, 21, 46, 81, 126, 181, 246, 321. For the second family we use the trinomial multiples

$$\Delta = (x^a + x^b + x^c)\Delta_A \text{ or } (x^a + x^b - x^c)\Delta_A, \text{ where } |a| = |b| = |c| \in \{1, 2, 3\} \text{ and } \gcd(x^a, x^b, x^c) = 1.$$

Each list contains 4 quadrics, 104 cubics and 684 quartics. We report our findings in a table:

Family	Pairs $(\Delta, \mathbf{m})$	Horn pairs	Percentage
$(x^a - x^b)\Delta_A$	8212	12	0.15%
$(x^a + x^b)\Delta_A$	8218	0	0%
$(x^a + x^b - x^c)\Delta_A$	8678	8	0.01%
$(x^a + x^b + x^c)\Delta_A$	8968	0	0%

All 12 Horn pairs in the first family represent the same model, up to a permutation of coordinates. All are coming from the six quadrics of the family. The model is the surface in Δ_4 defined by the 2×2 minors of the matrix $\begin{pmatrix} p_0 & p_1 & p_2 \\ p_0 + p_1 + p_2 & p_3 & p_4 \end{pmatrix}$. This is a staged tree model similar to Example 2, but now with three choices at each blue node instead of two.

In our construction of models with rational MLE, we start with families where r and m are fixed. However, as the entries of the matrix A go up, the number $n+1$ of states increases. This suggests the possibility of listing all models for fixed small values of n. Is this list finite?

Problem. Suppose that n is fixed. Are there only finitely many models with rational MLE in the simplex Δ_n? Can we find absolute bounds, depending only on n, for the dimension, degree and number of ideal generators of the associated varieties in $\mathbb{P}^n$?

Algorithm 1 is a tool for studying these questions experimentally. At present, however, we do not have any clear answers, even for $n = 3$, where the models are curves in a triangle.

References

[1] Nihat Ay, Jürgen Jost, Hông Vân Lê and Lorenz Schwachhöfer: *Information Geometry*, Springer Verlag, New York, 2017.

[2] Patrick Clarke and David A Cox: Moment maps, strict linear precision, and maximum likelihood degree one, `arXiv:1810.03672`.

[3] Rodrigo Collazo, Christiane Görgen and Jim Smith: *Chain Event Graphs*, CRC Press, Taylor & Francis, 2018.

[4] Eliana Duarte and Christiane Görgen: Equations defining probability tree models, *Journal of Symbolic Computation,* to appear. `arXiv:1802.04511`.

[5] Mathias Drton, Bernd Sturmfels and Seth Sullivant: *Lectures on Algebraic Statistics*, Oberwolfach Seminars, Vol 40, Birkhäuser, Basel, 2009.

[6] Luis Garcia-Puente and Frank Sottile: Linear precision for parametric patches, *Advances in Computational Mathematics* **33** (2010) 191–214.

[7] Israel M. Gelfand, Mikhail Kapranov, and Andrei Zelevinsky: *Discriminants, Resultants, and Multidimensional Determinants*, Birkhäuser, Boston, 1994.

[8] Christiane Görgen and Jim Q. Smith: Equivalence classes of staged trees, *Bernoulli* **24** (2018) 2676–2692.

[9] Daniel Grayson and Michael Stillman: *Macaulay2, a software system for research in algebraic geometry*, available at `http://www.math.uiuc.edu/Macaulay2/`.

[10] June Huh: Varieties with maximum likelihood degree one, *J. Algebr. Statistics* **5** (2014) 1–17.

[11] June Huh and Bernd Sturmfels: Likelihood geometry, in *Combinatorial Algebraic Geometry*, pages 63–117. Springer, 2014.

[12] Mikhael Kapranov: A characterization of A-discriminantal hypersurfaces in terms of the logarithmic Gauss map, *Mathematische Annalen* **290** (1991) 277–285.

[13] Rimvydas Krasauskas: Toric surface patches, *Advances Comput. Math.* **17** (2002) 89–113.

[14] Steffen L. Lauritzen: *Graphical Models*, Oxford University Press, 1996.

[15] Tomi Silander and Tze-Yun Leong: A dynamic programming algorithm for learning chain event graphs, in J. Fürnkranz, E. Hüllermeier and T. Higuchi (eds.): *Discovery Science*, Volume 8140 of *Lecture Notes in Computer Science*, pages 201–216. Springer, Berlin-Heidelberg, 2013.

[16] Jim Q. Smith and Paul E. Anderson: Conditional independence and chain event graphs, *Artificial Intelligence* **172** (2008) 42–68.

Authors' addresses:

Eliana Duarte, MPI-MiS Leipzig and OvG Universität Magdeburg, `eliana.duarte@ovgu.de`

Orlando Marigliano, MPI-MiS Leipzig `Orlando.Marigliano@mis.mpg.de`

Bernd Sturmfels, MPI-MiS Leipzig and UC Berkeley `bernd@mis.mpg.de`