COMMUTATIVE ALGEBRA OF STATISTICAL RANKING

BERND STURMFELS AND VOLKMAR WELKER

Abstract. A model for statistical ranking is a family of probability distributions whose states are orderings of a fixed finite set of items. We represent the orderings as maximal chains in a graded poset. The most widely used ranking models are parameterized by rational function in the model parameters, so they define algebraic varieties. We study these varieties from the perspective of combinatorial commutative algebra. One of our models, the Plackett-Luce model, is non-toric. Five others are toric: the Birkhoff model, the ascending model, the Csiszár model, the inversion model, and the Bradley-Terry model. For these models we examine the toric algebra, its lattice polytope, and its Markov basis.

1. Introduction

A statistical model for ranked data is a family $\mathcal{M}$ of probability distribution on the symmetric group $\mathfrak{S}_n$. Each distribution $p(\theta)$ in $\mathcal{M}$ depends on some model parameters θ and it associates a probability $p_\pi(\theta)$ to each permutation π of $[n] = \{1, 2, \ldots, n\}$. Thus the model $\mathcal{M}$ is a parametrized subset of the $(n!-1)$-dimensional standard simplex $\Delta_{\mathfrak{S}_n}$.

In algebraic statistics, one assumes that the probabilities $p_\pi(\theta)$ are rational functions in the model parameters θ, so that $\mathcal{M}$ is a semi-algebraic set in $\Delta_{\mathfrak{S}_n}$, and a primary goal is to characterize the prime ideal $I_\mathcal{M}$ of all polynomials that vanish on $\mathcal{M}$. In fact, one of the origins of the field of algebraic statistics was the *spectral analysis for permutation data* described by Diaconis and Sturmfels in [11, §6.1]. We call the corresponding model $\mathcal{M}$ the *Birkhoff model* because it is the toric variety of the Birkhoff polytope. This polytope consists of all bistochastic matrices and it is the convex hull of all $n \times n$ permutation matrices. There has been a considerable amount of research on finding the geometric invariants of the Birkhoff model $\mathcal{M}$. The simplest such invariant is its dimension, $\dim(\mathcal{M}) = (n-1)^2$. The *degree* of $\mathcal{M}$ is the normalized volume of the Birkhoff polytope, a topic of independent interest in combinatorics [5]. With regard to the central problem of identifying a *Markov basis* of $\mathcal{M}$, i.e. minimal generators of the toric ideal $I_\mathcal{M}$, the state of the art is the work of Matthias Lenz [20] who resolved a conjecture of Diaconis and Eriksson [10] by showing that the Markov basis of the Birkhoff model consists of binomials of degree ≤ 3.

Besides the Birkhoff model, there are many other models for ranked data that are both relevant for statistical analysis and have an interesting algebraic structure. It is the objective of this article to conduct a comparative study of such models from the perspectives of commutative algebra and geometric combinatorics. Both toric models and non-toric models are of interest. The former include the models introduced by Csiszár [8, 9], and the latter include the Plackett-Luce model [7, 21, 25] and the generalized Bradley-Terry models [18].

The organization of this paper is as follows. In Section 2 we give an informal introduction to all our models. We write out formulas for the probabilities for the six permutations of $n = 3$ items, and we discuss the subsets they parametrize in the 5-dimensional simplex $\Delta_{\mathfrak{S}_3}$. Precise formal definitions for the four toric models are given in Section 3. We represent the states as maximal chains in a graded poset Q. Typically, Q is the distributive lattice induced by some order constraints on the n items to be ranked. If there are no such constraints then $Q = 2^{[n]}$ is the Boolean lattice whose maximal chains are all $n!$ permutations in $\mathfrak{S}_n$. Non-trivial order constraints arise frequently in applications of ranking models, for instance in computational biology [3] and machine learning [7]. Our algebraic framework based on graded posets Q is well-suited for such contemporary applications of statistical ranking.

While the Birkhoff model has already received a lot of attention in the literature, we here focus on the *Csiszár model* (Section 4), the *ascending model* (Section 5) and the *inversion model* (Section 6). For each of these toric varieties, we characterize the corresponding lattice polytope and its Markov bases, that is, binomials that generate the toric ideal.

Section 7 is concerned with the *Plackett-Luce model*, which is not a toric model, but is parametrized by certain conditional probabilities that are not monomials. In algebraic geometry language, this model is obtained by blowing up the projective space $\mathbb{P}^{n-1}$ along a family of linear subspaces of codimension 2, and we study its coordinate ring. We also examine marginalizations of our models, including the widely used *Bradley-Terry model*.

2. Toric Models: A Sneak Preview

A toric model for complete permutation data is specified by a non-negative integer matrix A with $n!$ columns that all have the same sum. These column vectors A_π are indexed by permutations $\pi \in \mathfrak{S}_n$ and they represent the *sufficient statistics* of the model. The article [15] serves as our general reference for toric models in statistics, their relationship with exponential families, and the role of the matrix A. For an introduction to algebraic statistics in general, and for further reading on toric models, we refer to the books [12, 24].

If $r = \text{rank}(A)$ then the convex hull of the columns A_π is a lattice polytope of dimension $r - 1$. We refer to it as the *model polytope*. The toric model can be identified with the set of non-negative points on the projective toric variety associated with the model polytope.

Any data set is summarized as a function $u : \mathfrak{S}_n \mapsto \mathbb{N}$, where $u(\pi)$ is the number of times the permutation π has been observed. Thinking of u as a row vector, we can form the matrix-vector product Au, whose entries are the sufficient statistics of the data set u. Then the sum of the entries in the vector Au coincides with the sample size $N = \sum_{\pi \in \mathfrak{S}_n} u(\pi)$.

In subsequent sections we will generalize to the case where $\mathfrak{S}_n$ is replaced by a proper subset, in which case A has fewer than $n!$ columns, but still labeled by permutations. These will be the linear extensions of a given partial order on $[n] = \{1, 2, \ldots, n\}$. In fact, for some models we can even take the set of maximal chains in an arbitrary ranked poset. But for a first look we confine ourselves to the situation described above, where A has $n!$ columns.

We now define four toric models for probability distributions on $\mathfrak{S}_n$. We do this by way of a verbal description of the sufficient statistics in each model. These sufficient statistics are numerical functions on the permutations π of the given set $[n]$ of items to be ranked.

(a) In the *ascending model*, the sufficient statistics Au record, for each subset $I \subset [n]$, the number of samples π in the data u that have the set I at the bottom. Here, the set I being *at the bottom* means that $(i \in I$ and $j \notin I)$ implies $\pi(i) < \pi(j)$.
(b) In the *Csiszár model*, the sufficient statistics Au count, for each $i \in I \subset [n]$, the number of samples that have I at the bottom but with i as winner in the group I. This is the model studied by Villõ Csiszár [8, 9] under the name "L-decomposable".
(c) In the *Birkhoff model* of [11, §6.1], the sufficient statistics Au of a data set u record, for each $i, j \in [n]$, the number of samples π in which object i is ranked in place j,
(d) In the *inversion model*, the sufficient statistics Au count, for each ordered pair $i < j$ in $[n]$, the number of samples π in which that pair is an inversion, meaning $\pi^{-1}(i) > \pi^{-1}(j)$. This model is a multivariate version of the *Mallows model* [22].

To illustrate the differences between these models let us consider the simplest case $n = 3$. In each case the toric ideal of the model is the kernel of a square-free monomial map from the polynomial ring $\mathbb{K}[p_{123}, p_{132}, p_{213}, p_{231}, p_{312}, p_{321}]$ representing the probabilities to another polynomial ring $\mathbb{K}[a, b, \dots]$ that represents the model parameters. The model polytope is the convex hull of the six 0-1 vectors corresponding to the square-free monomials:

	p_{123}	p_{132}	p_{213}	p_{231}	p_{312}	p_{321}
Birkhoff	$a_{11}a_{22}a_{33}$	$a_{11}a_{23}a_{32}$	$a_{12}a_{21}a_{33}$	$a_{12}a_{23}a_{31}$	$a_{13}a_{21}a_{32}$	$a_{13}a_{22}a_{31}$
inversion	$b_{12}b_{13}b_{23}$	$b_{12}b_{13}q_{23}$	$q_{12}b_{13}b_{23}$	$q_{12}q_{13}b_{23}$	$b_{12}q_{13}q_{23}$	$q_{12}q_{13}q_{23}$
ascending	$c_1c_{12}c_{123}$	$c_1c_{13}c_{123}$	$c_2c_{12}c_{123}$	$c_2c_{23}c_{123}$	$c_3c_{13}c_{123}$	$c_3c_{23}c_{123}$
Csiszár	$d_{\|1}d_{1\|2}d_{12\|3}$	$d_{\|1}d_{1\|3}d_{13\|2}$	$d_{\|2}d_{2\|1}d_{12\|3}$	$d_{\|2}d_{2\|3}d_{23\|1}$	$d_{\|3}d_{3\|1}d_{13\|2}$	$d_{\|3}d_{3\|2}d_{23\|1}$

The toric ideals record the algebraic relations among these square-free monomials:

$$\begin{array}{rcll} I_{\text{inv}} & = & \langle\, p_{132}p_{231} - p_{123}p_{321},\; p_{213}p_{312} - p_{123}p_{321} \,\rangle & \text{has codimension 2,} \\ I_{\text{birk}} = I_{\text{asc}} & = & \langle\, p_{123}p_{231}p_{312} - p_{132}p_{213}p_{321} \,\rangle & \text{has codimension 1,} \\ I_{\text{csi}} & = & \langle\ 0\ \rangle & \text{has codimension 0.} \end{array}$$

For each model, the matrix A has six columns, indexed by $\mathfrak{S}_3$, and its rows are labeled by the model parameters. For example, for the ascending model, the matrix has seven rows:

$$\mathrm{As}_3 \quad = \quad \begin{array}{c} \\ c_1 \\ c_2 \\ c_3 \\ c_{12} \\ c_{13} \\ c_{23} \\ c_{123} \end{array} \begin{array}{c} \begin{array}{cccccc} p_{123} & p_{132} & p_{213} & p_{231} & p_{312} & p_{321} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

Here we use the same notation for both the matrix and the model polytope, which is the convex hull of the columns. From the equality of ideals, $I_{\text{birk}} = I_{\text{asc}}$, we infer that the polytope As_3 is affinely isomorphic to the 3×3-Birkhoff polytope, which is a cyclic 4-polytope with six vertices. The ideal I_{inv} reveals that the model polytope for the inversion model is a *regular octahedron*, while the polytope for the Csiszár model is the full 5-*simplex*.

To see that no two of our four models agree, we need to go to $n \geq 4$. For instance, let us fix $n = 4$. Then all four model polytopes have 24 vertices but their dimensions are different. For $n = 4$, the inversion model has dimension 6, the Csiszár model has dimension 7, the Birkhoff model has dimension 9, and the ascending model has dimension 11. Theorem Theorem 3.1 will explain the precise relationships and inclusions among the four models.

Our work on this project started by trying to understand a certain model whose toric closure is the ascending model. Here *toric closure* refers to the smallest toric model containing a given model. That non-toric model for ranking is the *Plackett-Luce model* [7, 21, 25]. It can be obtained from the ascending model by the following specialization of parameters:

$$c_i \mapsto \frac{1}{\theta_i}, \; c_{ij} \mapsto \frac{1}{\theta_i + \theta_j}, \; c_{ijk} \mapsto \frac{1}{\theta_i + \theta_j + \theta_k}, \; \ldots .$$

The prime ideal of algebraic relations among the p_π is a non-toric ideal which contains the toric ideal I_{asc}. The case $n = 3$ is worked out explicitly in Example 7.1. Geometrically, that smallest Plackett-Luce model corresponds to blowing up $\mathbb{P}^2$ at the nine points in (18).

3. Toric Models: Definitions and General Results

Let Q be a poset on finite ground set Ω. A *Q-ranking* is a maximal chain $a_0 < \cdots < a_n$ in Q. A chain $a_0 < \cdots < a_n$ being *maximal* means that a_0 is minimal in Q, a_n is maximal, and $a_i < a_{i+1}$ is a cover relation for $0 \leq i \leq n-1$. We write $\mathrm{M}(Q)$ for the set of maximal chains in Q and $\mathrm{Cov}(Q)$ for the set of cover relations in Q. If $Q = 2^{[n]}$ is the Boolean lattice of all subsets of $[n]$ ordered by inclusion then the maximal chains in Q are in bijection with the permutations in $\mathfrak{S}_n$, and the models below coincide with the ones described in Section 2.

We shall define four toric models whose states are the maximal chains $\pi \in \mathrm{M}(Q)$. The probability of π is a represented by an indeterminate p_π. Each toric model for Q-rankings is defined by a non-negative integer matrix A whose columns are indexed by $\mathrm{M}(Q)$ and all have the same sum. The integer matrix A represents a monomial map from the polynomial ring $\mathbb{K}[p]$ in the unknowns p_π, $\pi \in \mathrm{M}(Q)$, to a suitably chosen second polynomial ring.

Any data set gives a function $u : \mathrm{M}(Q) \mapsto \mathbb{N}$, where $u(\pi)$ is the number of times the permutation π has been observed. Thinking of u as a row vector, we can form the matrix-vector product Au, whose entries are the sufficient statistics of the data set u. Note that the sum of the entries in the vector Au coincides with the sample size $N = \sum_{\pi \in \mathrm{M}(Q)} u(\pi)$.

(a) In the *ascending model*, the sufficient statistic Au records, for any given poset element $a \in Q$, the number of observed maximal chains π that pass though a. The model parameters are represented by unknowns c_a, and the monomial map is

$$p_\pi \;\mapsto\; c_{a_0} c_{a_1} \cdots c_{a_n} \qquad \text{for } \pi = (a_0 {<} a_1 {<} \cdots {<} a_n).$$

(b) In the *Csiszár model,* the sufficient statistic Au records, for any cover $a < b$, the number of observed maximal chains π passing though a and b. The model parameters are represented by unknowns $d_{a<b}$ for $(a{<}b) \in \mathrm{Cov}(Q)$. The monomial map is

$$p_\pi \;\mapsto\; d_{a_0<a_1} d_{a_1<a_2} \cdots d_{a_{n-1}<a_n} \qquad \text{for } \pi = (a_0 {<} a_1 {<} \cdots {<} a_n).$$

If $Q = 2^{[n]}$, the Boolean lattice of subsets $a \subseteq [n]$, then the maximal chains π in Q are identified with permutations in $\mathfrak{S}_n$, and we recover the ascending model as defined in Section 2. Likewise we recover the Csiszár model on $\mathfrak{S}_n$ by setting $d_{a<b} = d_{a|i}$ for $\{i\} = b\backslash a$.

The Birkhoff and inversion model cannot be formulated in the above generality. For these we need assume that the poset Q is a *distributive lattice*. This means that $Q = \mathcal{O}(\mathcal{P})$ is the poset of order ideals in a given partial order $\mathcal{P}$ on $[n]$. We refer to $\mathcal{P}$ as the *constraint poset*. The constraint $i{<}j$ stipulates that item i must always be ranked before item j. The maximal chains π in $Q = \mathcal{O}(\mathcal{P})$ are the permutations of $[n]$ that respect all constraints in $\mathcal{P}$. See [3] for an introduction to distributive lattices in a context of statistical interest.

The compatible permutations π are known as *linear extensions* of $\mathcal{P}$. From now on we abbreviate $\mathcal{L}(\mathcal{P}) = \mathrm{M}(\mathcal{O}(\mathcal{P}))$, and we identify elements of $\mathcal{L}(\mathcal{P})$ with permutations $\pi \in \mathfrak{S}_n$ that represent linear extensions of $\mathcal{P}$. This allows us to define our other two toric models:

(c) In the *Birkhoff model*, the sufficient statistic Au records, for all $i, j \in [n]$, the number of samples $\pi \in \mathcal{L}(\mathcal{P})$ for which object j is ranked in position i. The model parameters are represented by unknowns a_{ij} for $i, j \in [n]$. The monomial map is

$$p_\pi \;\mapsto\; a_{1\pi(1)} a_{2\pi(1)} \cdots a_{n\pi(n)} \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}).$$

(d) In the *inversion model*, the sufficient statistics Au records, for each ordered pair i, j in $[n]$, the number of samples $\pi \in \mathcal{L}(\mathcal{P})$ for which $i < j$ but $\pi^{-1}(i) > \pi^{-1}(j)$. The model parameters are represented by unknowns u_{ij} and v_{ij}. The monomial map is

$$p_\pi \;\mapsto\; \prod_{\substack{1 \le i < j \le n \\ \pi^{-1}(i) < \pi^{-1}(j)}} u_{ij} \prod_{\substack{1 \le i < j \le n \\ \pi^{-1}(i) > \pi^{-1}(j)}} v_{ij} \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}).$$

In general, we have the following inclusions among the four toric models (a)-(d). These inclusions of toric varieties correspond to linear projections among the model polytopes.

Theorem 3.1. (i) *The ascending model and the Csiszár model on a poset Q satisfy*

$$\mathcal{M}_{\mathrm{asc}} \subseteq \mathcal{M}_{\mathrm{csi}},$$

provided Q has either a unique minimal element $\hat{0}$ or a unique maximal element $\hat{1}$.

(ii) *If $Q = \mathcal{O}(\mathcal{P})$ is a distributive lattice, then the Birkhoff model $\mathcal{M}_{\mathrm{birk}}$, the inversion model $\mathcal{M}_{\mathrm{inv}}$, the ascending model $\mathcal{M}_{\mathrm{asc}}$ and the Csiszár model $\mathcal{M}_{\mathrm{csi}}$ satisfy*

$$\mathcal{M}_{\mathrm{inv}} \subseteq \mathcal{M}_{\mathrm{csi}} \;\; \textit{and} \;\; \mathcal{M}_{\mathrm{birk}} \subseteq \mathcal{M}_{\mathrm{asc}} \subseteq \mathcal{M}_{\mathrm{csi}}.$$

(iii) *The inclusions (ii) are strict in general. In particular, if $n \geq 4$ and $Q = 2^{[n]}$ then*

$$\mathcal{M}_{\mathrm{inv}} \not\subseteq \mathcal{M}_{\mathrm{asc}} \;\; \textit{and} \;\; \mathcal{M}_{\mathrm{birk}} \not\subseteq \mathcal{M}_{\mathrm{inv}}.$$

Proof. We begin by establishing (iii) for $n = 4$. A direct computation as in Section 6 reveals that the inversion model $\mathcal{M}_{\mathrm{inv}}$ is a projective toric variety of dimension 6 and degree 180 in $\mathbb{P}^{23}$. The Markov basis of I_{inv} consists of 81 quadrics. Since $\mathcal{M}_{\mathrm{birk}}$ has dimension 9, we conclude that $\mathcal{M}_{\mathrm{birk}} \not\subseteq \mathcal{M}_{\mathrm{inv}}$. An explicit point p in $\mathcal{M}_{\mathrm{birk}} \backslash \mathcal{M}_{\mathrm{inv}}$ is the uniform distribution on the nine derangements. This arises by setting $a_{ij} = 0$ and $a_{ij} = 1/\sqrt{3}$ for all $i \neq j$. The the quadric $p_{1243}p_{4321} - p_{2143}p_{4312} \in I_{\mathrm{inv}}$ does not vanish for this particular distribution.

The ascending model $\mathcal{M}_{\mathrm{asc}}$ has dimension 11 and degree 808. The Markov basis of its toric ideal I_{asc} consists of six quadrics, 64 cubics and 93 quartics. One of the cubics is

$$p_{1234}p_{1342}p_{1423} \,-\, p_{1243}p_{1324}p_{1432} \quad \in \; I_{\mathrm{asc}}. \tag{1}$$

An example of a point in $\mathcal{M}_{\mathrm{inv}} \backslash \mathcal{M}_{\mathrm{asc}}$ is obtained by taking the parameter values

$$u_{12} = u_{13} = u_{14} = 0,\, u_{23} = u_{24} = u_{34} = v_{12} = v_{13} = v_{23} = v_{24} = 1,\, v_{34} = 2,\, v_{14} = 1/9.$$

The resulting distribution is supported on the six permutations in (1). Its coordinates are

$$p_{1234} = p_{1342} = p_{1423} = 2/9 \quad \text{and} \quad p_{1243} = p_{1324} = p_{1432} = 1/9.$$

This distribution is not a zero of (1), and hence it is not in the ascending model $\mathcal{M}_{\mathrm{asc}}$.

The two probability distributions on permutations seen above can be lifted to similar counterexamples for $n \geq 5$, and we conclude that the non-inclusions are valid for all $n \geq 4$.

The inclusion $\mathcal{M}_{\mathrm{asc}} \subset \mathcal{M}_{\mathrm{csi}}$ in (i) is seen by the specialization of parameters that sends $d_{a<b}$ to c_a if Q has a unique maximal element $\hat{1}$ and to c_b if Q has a unique minimum $\hat{0}$.

We lastly prove the inclusions in (ii). The parameters for the Csiszár model $\mathcal{M}_{\mathrm{csi}}$ are $d_{a<b}$ where $a < b \in \mathrm{Cov}(Q)$. If $\mathrm{M}(Q) = \mathcal{L}(\mathcal{P})$ then the cover relation $a < b$ means $b = a \cup \{j\}$. Thus the following specialization of parameters gives the parameterization of $\mathcal{M}_{\mathrm{inv}}$:

$$d_{a<b} \;\mapsto\; \prod_{i \in a, i<j} u_{ij} \prod_{i \in a, i>j} v_{ij}.$$

This shows that the inversion model $\mathcal{M}_{\mathrm{inv}}$ is a subvariety of the Csiszár model $\mathcal{M}_{\mathrm{csi}}$.

It remains to show that $\mathcal{M}_{\mathrm{birk}} \subset \mathcal{M}_{\mathrm{asc}}$. To do this, we let A denote the model matrix for $\mathcal{M}_{\mathrm{birk}}$ and B the model matrix for $\mathcal{M}_{\mathrm{asc}}$. Both matrices have their entries in $\{0, 1\}$ and they have $|\mathcal{L}(\mathcal{P})|$ columns. The rows A_{ij} of A are indexed by unordered pairs $i, j \in [n] \times [n]$, and the rows B_I of B are indexed by subsets of $[n]$. We have the identity

$$A_{ij} \;=\; \sum\Big\{B_I \,:\, I \in \tbinom{[n]}{j} \text{ and } i \in I\Big\} \,-\, \sum\Big\{B_I \,:\, I \in \tbinom{[n]}{j-1} \text{ and } i \in I\Big\}.$$

This shows that every row of A is a $\mathbb{Z}$-linear combination of the rows of B. Hence, the kernel of A contains the kernel of B, and this implies that the toric ideal $I_A = I_{\mathrm{birk}}$ contains the toric ideal $I_B = I_{\mathrm{asc}}$. We conclude that $\mathcal{M}_{\mathrm{birk}}$ is a submodel of $\mathcal{M}_{\mathrm{asc}}$. $\square$

In the rest of this paper we consider the ascending and Csiszár models only in the graded situation, that is, when the monomial images of all the unknowns p_c, $c \in \mathrm{M}(Q)$ have the same total degree. The latter is equivalent to requiring that all maximal chains in Q have the same cardinality, which in turn is equivalent to Q being graded. For a graded poset Q we denote by $\mathrm{rk} : Q \to \mathbb{N}$ its rank function and write Q_i for the set of its elements of rank i. By $\mathrm{rk}(Q)$ we denote the rank of Q, which is the maximal rank of any of its elements.

In the next three sections we undertake a detailed study of the models (b), (a) and (d), in this order. The Birkhoff model (c) has already received considerable attention in the literature [10, 11], at least for $\mathcal{L}(\mathcal{P}) = \mathfrak{S}_n$, and we content ourselves with a few brief remarks. Its model polytope, the Birkhoff polytope of doubly stochastic matrices, is a key player in combinatorial optimization, and it is linked to many fields of pure mathematics.

The restriction of the Birkhoff model and its polytope to proper subsets $\mathcal{L}(\mathcal{P})$ of $\mathfrak{S}_n$ has been studied only in some special cases. For example, Chan, Robbins and Yuen [6] considered this polytope for the constraint poset $\mathcal{P}$ given by the transitive closure of $j > j-2$ and $j > j-3$ for $3 \leq j \leq n$. They stated a conjecture on its volume which was proved by Zeilberger [28]. We close by noting a formula for the dimension of these polytopes.

Proposition 3.2. *Let $\mathcal{P}$ be an arbitrary constraint poset on $[n] = \{1, 2, \ldots, n\}$. Set*

$$Z \;=\; \{\, (i,j) \in [n] \times [n] \mid \pi(i) \neq j \textit{ for all } \pi \in \mathcal{L}(\mathcal{P}) \,\}$$

$$\textit{and} \qquad C \;=\; \left\{ (i,j) \in [n] \times [n] \mid (i,j) \notin Z \textit{ and } \begin{array}{l} (i,j') \in Z \textit{ for } j' > j \textit{ or} \\ (i',j) \in Z \textit{ for } i' > j \end{array} \right\}.$$

The model polytope Bi *of the Birkhoff model, expressed using coordinates x_{ij} on $\mathbb{R}^{n\times n}$, equals the face of the classical Birkhoff polytope of bistochastic $n\times n$-matrices defined by*

$$x_{ij} = 0 \quad \textit{for all } (i,j) \in Z. \tag{2}$$

In particular, the dimension of the Birkhoff model polytope is $\dim(\mathrm{Bi}) \;=\; n^2 - |Z| - |C|$.

Proof. Clearly, the model polytope Bi of the Birkhoff model is contained in the classical Birkhoff polytope. Equally obvious is that all equations (2) are valid for the model polytope. Hence Bi is contained in the polytope cut out from the classical Birkhoff polytope by (2).

Following the lines of the Birkhoff-von Neumann Theorem (see e.g. [1, (5.2)]), we note that the vertices of the polytope cut out by (2) from the classical Birkhoff polytope are the permutation matrices of the permutation $\pi \in \mathcal{L}(\mathcal{P})$. The first assertion now follows.

The linear relations on the Birkhoff polytope state that all row and column sums are 1. In these relations we set $x_{ij} = 0$ for $(i,j) \in Z$. In the resulting linear relations precisely the variables x_{ij} for $(i,j) \in C$ are the leading terms. This proves the dimension statement. □

We illustrate Proposition 3.2 with two simple examples. If $\mathcal{P}$ is an n-element antichain then $Z = \emptyset$ and $C = \{(1,n), (2,n), \ldots, (n,n), (n,n-1), \ldots (n,1)\}$. Here our formula gives the dimension $n^2 - 0 - (2n-1) = (n-1)^2$ of the classical Birkhoff polytope. If $\mathcal{P}$ is the n-chain $1{<}2{<}\cdots{<}n$ then $Z = \{(i,j) \in [n] \times [n] \mid i \neq j\}$ and $C = \{(i,i) | i \in [n]\}$. Here the model polytope is joint one point, since $\dim(\mathrm{Bi}) = n^2 - |Z| - |C| = n^2 - n(n-1) - n = 0$.

4. The Csiszár model

The Csiszár model for the Boolean lattice $Q = 2^{[n]}$ was studied by Villõ Csiszár in [8, 9]. She calls it the *L-decomposable model* where the letter "L" refers to Luce [21]. We prefer to call it the *Csiszár model*, to credit her work for introducing this model in its current form in algebraic statistics. We note that the Csiszár model for $Q = 2^{[n]}$ also appears in work on multiple testing procedures by Hommel *et. al.* [17], but with a different coordinatization of its model polytope. Throughout this section, we fix a graded poset Q of positive rank.

We begin by defining a 0-1-matrix A = Ci that represents the Csiszár model. Our construction is based on the technique employed for $Q = 2^{[n]}$ in Csiszár's proof of [8, Theorem 1]. The columns of Ci are indexed by the unknown probabilities p_π where $\pi \in$

$\mathrm{M}(Q)$, and the rows of Ci are indexed by the model parameters $d_{a<b}$ where $(a<b) \in \mathrm{Cov}(Q)$. We write $\mathrm{MinCov}(Q)$ for the set of cover relations $a < b$ for some element $a \in Q_0$ of rank 0.

Consider the discrete undirected graphical model [12, 15] given by the n-chain graph $G = ([n], E)$ with edge set $E = \{\{i, i+1\} \mid 1 \leq i \leq n-1\}$. We take as the states of node i the set Q_i of all elements of rank i in Q. The n-chain graph G is chordal (or decomposable), so the five equivalent conditions of [15, Theorem 4.4] hold for G. Let A_G denote the associated model matrix [15, §2.2]. It has $\prod_{i=0}^{n} |Q_i|$ columns indexed by tuples $(a_0, \ldots, a_n)$ of elements $a_i \in Q_i$ and $\sum_{i=0}^{n-1} |Q_i| \cdot |Q_{i+1}|$ rows indexed by pairs (a, b) of elements of Q from consecutive ranks. Its entries are 0 or 1 according to the pattern for an undirected graphical model. The key fact we will be using is that the cone spanned by the columns of A_G is equal to the cone of all non-negative vectors in the column space of A_G.

As in Csiszár's proof of [8, Theorem 1], we focus on the submatrix A'_G of A_G whose column labels $(a_0, \ldots, a_n)$ correspond to maximal chains $a_0 < \cdots < a_n$ from $\mathrm{M}(Q)$. Many of the rows of A'_G are entirely zero, namely, all those rows indexed by pairs (a, b), where a is not covered by b in Q. Let A''_G denote the matrix obtained from A'_G by deleting all such zero rows. The remaining rows are indexed by pairs $(a, b) \in Q_i \times Q_{i+1}$ for some i. Equivalently, the rows of A''_G are indexed by $\mathrm{Cov}(Q)$. This shows that the toric model A''_G is precisely our Csiszár model, and, with this identification of coordinates our polytope Ci coincides with the convex hull of the columns of A''_G. Now we are in a position to give a description of the model polytope Ci in terms of linear equalities and inequalities.

Theorem 4.1. *Let Q be a graded poset of rank ≥ 1 and* $\mathrm{Ci} \subseteq \mathbb{R}^{\mathrm{Cov}(Q)}$ *the model polytope of its Csiszár model, with coordinates $x_{a<b}$ indexed by cover relations $a < b$ in* $\mathrm{Cov}(Q)$. *Then* Ci *is of dimension* $|\mathrm{Cov}(Q)| - |Q| + |Q_n| + |Q_0| - 1$. *Inside the orthant defined by*

$$x_{a<b} \geq 0 \qquad \text{for} \;\; (a<b) \in \mathrm{Cov}(Q), \tag{3}$$

the polytope Ci *is the solution set of the inhomogeneous linear equation*

$$\sum_{a<b \in \mathrm{MinCov}(Q)} x_{a<b} \;=\; 1 \tag{4}$$

together with the system of linear homogeneous equations

$$\sum_{b \in \nabla a} x_{a<b} \;=\; \sum_{b \in \Delta a} x_{b<a} \qquad \text{for} \;\; a \in Q \backslash (Q_0 \cup Q_n), \tag{5}$$

where ∇a is the set of b that cover a, and Δa is the set of b that are covered by a.

Proof. Let $G = ([n], E)$ with $E = \{\{i, i+1\} \mid 1 \leq i \leq n-1\}$ be the n-chain and A_G the defining matrix of its graphical model as discussed above. Also let A'_G and A''_G be as above.

The n-chain graph G is decomposable, so the five equivalent conditions in [15, Theorem 4.4] are true. The fifth condition, that the exponential family is closed in the probability simplex, is equivalent to the statement that the model polytope of that n-chain model is defined by linear equations and non-negativity constraints only. See [19] for a toric algebra perspective. We have shown that the toric model of A''_G is our Csiszár model. With this identification, the model polytope Ci coincides with the convex hull of the columns of A''_G.

The matrix A'_G was constructed so that its columns are precisely the points on a face of the model polytope for A_G. Hence the model polytope of the Csiszár model is obtained from the earlier polytope by simply setting some of the non-negative coordinates to zero. This implies that Ci inherits all the desirable properties spelled out in Theorem 4.4 of [15]. In particular, its exponential family is closed, and the polytope Ci coincides with the set of all non-negative points in the affine space spanned by the columns of the matrix A''_G.

At this stage we only need to show that the affine span of the columns of A''_G equals the solution space of (4) and (5). The equation (4) holds for a vertex of the model polytope because any maximal chain contains exactly one cover relation involving an element of rank 0 and an element of rank 1. The equations (5) holds for a vertex of the model polytope because, given any element $a \in Q$, a maximal chain either contains no cover relation involving a or exactly two, one of the form $b < a$ and one of the form $a < b'$. Hence each column of A''_G satisfies (4) and (5). Conversely, any 0-1-solution of these equations must come from a maximal chain in Q, and hence is among the columns of A''_G. □

The toric ideal $I_{\rm csi}$ of the Csiszár model is the kernel of the ring homomorphism

$$\mathbb{K}[p] \to \mathbb{K}[d]\,, \;\; p_\pi \;\mapsto\; d_{a_0<a_1} d_{a_1<a_2} \,\cdots\, d_{a_{n-1}<a_n} \qquad \text{for } \; \pi = (a_0{<}a_1{<}\cdots{<}a_n).$$

The minimal generators of $I_{\rm csi}$ form the Markov basis of $\mathcal{M}_{\rm csi}$. As shown in the proof of Theorem 4.1, the Csiszár model polytope $\mathrm{Ci} = A''_G$ inherits the equivalent conditions (b),(c),(d),(e) in [15, Theorem 4.4] from the larger model A_G. In particular, the toric ideal $I_{\rm csi}$ has a Gröbner basis consisting of quadratic binomials. We shall now describe this Gröbner basis explicitly. It generalizes the Markov basis for $Q = 2^{[n]}$ in [8, Theorem 3.1].

Theorem 4.2. *A Gröbner basis for the toric ideal $I_{\rm csi}$ of the Csiszár model on a graded poset Q is given by all quadratic binomials of the form*

$$p_{\pi_1\pi_2} \cdot p_{\pi'_1\pi'_2} \;-\; p_{\pi_1\pi'_2} \cdot p_{\pi'_1\pi_2}, \tag{6}$$

where the chains π_1 and π'_1 have the same ending point and both π_2 and π'_2 start there.

Proof. It is easy to check that the binomial quadrics that lie in the ideal $I_{\rm csi}$ are precisely the quadrics (6). These are inherited from the conditional independence statements valid for the n-chain graphical model G. These conditional independence statements translate into a quadratic Gröbner basis for the toric ideal of the matrix A_G. The leading terms of that Gröbner basis are squarefree, so they define a regular unimodular triangulation of the convex hull of the columns of A_G. Since $\mathrm{Ci} = A''_G$ is a face of that polytope, that face inherits the regular unimodular triangulation from A_G. We conclude that the Gröbner basis which specifies this regular triangulation of Ci consists precisely of the quadrics (6). □

The Gröbner basis (6) reveals that the Csiszár model has desirable algebraic properties:

Corollary 4.3. *The coordinate ring $\mathbb{K}[\,p\,]/I_{\rm csi}$ of the Csiszár model over any field $\mathbb{K}$ is Cohen-Macaulay and Koszul. Its Krull dimension equals $|\mathrm{Cov}(Q)| - |Q| + |Q_n| + |Q_0|$.*

Proof. Since $I_{\rm csi}$ has a quadratic Gröbner basis, by Theorem 4.2, it follows that $\mathbb{K}[\,p\,]/I_{\rm csi}$ is Koszul. Since Ci has a regular unimodular triangulation, the semigroup algebra $\mathbb{K}[\,p\,]/I_{\rm csi}$

is normal and hence Cohen-Macaulay, by Hochster's Theorem. The dimension of this semigroup algebra is one more than the dimension of its polytope, given in Theorem 4.1. □

For computations it is convenient to represent the quadrics in (6) as the 2×2-minors of certain natural matrices M_q that are indexed by the elements q of the poset Q. The row labels of the matrix M_q are the maximal chains in the order ideal $Q_{\leq q} = \{a \in Q : a \leq q\}$ and the column labels of M_q are the maximal chains in the filter $Q_{\geq q} = \{b \in Q : q \leq b\}$. Thus M_q is a matrix of format $|\mathrm{M}(Q_{\leq q})| \times |\mathrm{M}(Q_{\geq q})|$. We define M_q as follows. The entry of M_q in the row labeled $\pi_1 \in \mathrm{M}(Q_{\leq q})$ and the column labeled $\pi_2 \in Q_{\geq q}$ is the unknown p_π where π denotes the maximal chain of Q that is obtained by concatenating π_1 and π_2.

Corollary 4.4. *The Markov basis of the Csiszár ideal I_{csi} consists of the 2×2-minors of the matrices M_q, where q runs over Q. This Markov basis is also a Gröbner basis.*

Proof. Each 2×2-minor of M_q has the form required in (6), and, conversely, each binomial in (6) occurs as a 2×2-minor of M_q for some q. Note that this element $q \in Q$ is generally not unique for a given binomial. The Gröbner basis statement is part of Theorem 4.2. □

We illustrate our results for the case when $Q = 2^{[n]}$ is the Boolean lattice, with $n \leq 6$. For $n = 3$, the ideal I_{csi} is zero as seen in Section 2. For $n = 4$, the ideal I_{csi} is the complete intersection of six quadrics, namely, the determinants of the six 2×2-matrices $M_{\{i,j\}}$. Geometrically, these correspond to the six square faces of the 3-dimensional permutahedron:

$$\begin{aligned} I_{\mathrm{csi}} \;=\; \langle\; & p_{1243}p_{2134} - p_{1234}p_{2143},\;\; p_{1342}p_{3124} - p_{1324}p_{3142},\;\; p_{1432}p_{4123} - p_{1423}p_{4132},\\ & p_{2341}p_{3214} - p_{2314}p_{3241},\;\; p_{2431}p_{4213} - p_{2413}p_{4231},\;\; p_{3421}p_{4312} - p_{3412}p_{4321} \;\rangle. \end{aligned}$$

We conclude that the Csiszár model for $n = 4$ has dimension 17, as predicted by Theorem 4.1. As a projective variety, this model has degree 32 since it is a complete intersection. For $n = 5$, the Markov basis consists of the 2×2-minors of the ten 2×6-matrices $M_{\{1,2\}}, M_{\{1,3\}}, \ldots, M_{\{4,5\}}$ and ten 6×2-matrices $M_{\{1,2,3\}}, M_{\{1,2,4\}}, \ldots, M_{\{3,4,5\}}$. For example,

$$M_{\{2,4\}} \quad = \quad \begin{pmatrix} p_{24135} & p_{24153} & p_{24315} & p_{24351} & p_{24513} & p_{24531} \\ p_{42135} & p_{42153} & p_{42315} & p_{42351} & p_{42513} & p_{42531} \end{pmatrix}.$$

Altogether, these matrices have 300 maximal minors but 30 of the minors occur in two matrices, so the total number of distinct Markov basis elements is 270. The dimension of this model is 49, and its degree equals 50493797160. The Hilbert series of $\mathbb{K}[p]/I_{\mathrm{csi}}$ equals

$$\begin{aligned} &(1 + 70t + 2215t^2 + 42020t^3 + 534635t^4 + 4837694t^5 + 32227985t^6 + 161529320t^7 \\ &+617560160t^8 + 1816401720t^9 + 4129171068t^{10} + 7265606880t^{11} + 9880962560t^{12} \\ &+10337876480t^{13} + 8250364160t^{14} + 4953798656t^{15} + 2189864960t^{16} \\ &+688455680t^{17} + 145162240t^{18} + 18350080t^{19} + 1048576t^{20})/(1-t)^{50}. \end{aligned}$$

For $n = 6$, the Markov basis is represented by the fifteen 2×24-matrices $M_{\{i,j\}}$, the twenty 6×6-matrices $M_{\{i,j,k\}}$ and the fifteen 24×2-matrices $M_{\{i,j,k,l\}}$. Altogether, these 50 matrices have 12780 minors of size 2×2 but only 10980 of the binomial quadrics are distinct.

A systematic way of understanding our matrices M_q is furnished by Sullivant's theory of *toric fiber products* [27]. These matrices represent the "glueing quadrics" used in method for constructing larger toric ideals from smaller ones. In particular, we could give an alternative proof of Theorem 4.2 by applying the general results in [27]. Namely, the toric algebra $\mathbb{K}[p]/I_{\mathrm{csi}}$ can be obtained as an iterated toric fiber product of suitably graded smaller polynomial rings that are attached to the pieces in a decomposition of Q into antichains.

5. The ascending model

At the end of [8, p. 233] it is asserted that a Markov basis for the ascending model on $Q = 2^{[n]}$ can be obtained in a similar way as was done for the standard Csiszár model, but no details are given. However, simple examples show that it does not suffice to consider quadratic binomials for the generating set and it is not clear from [8] which properties the defining ideals of the ascending and Csiszár model have in common. The defining ideal and the model polytope of the ascending model seem to be complicated and more interesting than those of the Csiszár model. These are the structures to be explored in this section.

Generalizing the notation introduced in the preceding section, for any subset $A \subseteq Q$, we consider the set of elements of A that cover an element from A:

$$\nabla A \;:=\; \{b \in Q \mid a < b \in \mathrm{Cov}(Q) \text{ for some } a \in A\}.$$

We also consider the set of elements covered by an element from A:

$$\Delta A := \{b \in Q \mid b < a \in \mathrm{Cov}(Q) \text{ for some } a \in A\}.$$

Theorem 5.1. *Fix a graded poset Q of rank n. The model polytope* As *of the ascending model is the set of solutions in the space $\mathbb{R}^{|Q|}$, with coordinates x_a for $a \in Q$, of the equations*

$$\sum_{a \in Q_i} x_a \;=\; 1, \quad 0 \le i \le n, \tag{7}$$

and the inequalities

$$x_a \;\ge\; 0, \quad a \in Q, \tag{8}$$

$$-\sum_{a \in A} x_a + \sum_{a \in \nabla A} x_a \;\ge\; 0, \quad A \subseteq Q_i\,,\ 0 \le i \le n-1. \tag{9}$$

Proof. Equations (7) are valid on every vertex of As because every maximal chain in P has exactly one element of rank i for all $0 \le i \le n$. The inequalities (9) express the fact that if a maximal chain passes through an element of $A \subseteq Q_i$ then it must also pass through a unique element of ∇A. Inequalities (8) are obviously valid for As. Hence As is contained in the intersection of the linear spaces defined by (7) and the halfspaces defined by (8) and (9).

For the converse we proceed by induction on n. If $n = 0$ then As is a simplex of dimension $|Q| - 1$, defined by (7) and (8). If $n = 1$ then the result follows from [23, Corollary 1.8 (b)].

Assume $n \ge 2$. Let $\mathbf{x} = (x_a)_{a \in Q} \in \mathbb{R}^Q$ be any vector satisfying (7), (8) and (9). Let $\mathbf{x}'$ be the projection of $\mathbf{x}$ onto the coordinates in $Q' = Q_0 \cup \cdots \cup Q_{n-1}$ and $\mathbf{x}''$ the projection of $\mathbf{x}$ onto $Q'' = Q_{n-1} \cup Q_n$. By induction, $\mathbf{x}'$ and $\mathbf{x}''$ lie in the model polytopes of the ascending model for Q' and Q''. Hence we can write $\mathbf{x}$ and $\mathbf{x}'$ as convex linear combinations:

$$\mathbf{x}' \;=\; \textstyle\sum_{c' \in \mathrm{M}(Q')} \lambda_{c'} c' \qquad \text{and} \qquad \mathbf{x}'' \;=\; \textstyle\sum_{c'' \in \mathrm{M}(Q'')} \lambda_{c''} c''.$$

Consider a fixed element $a \in Q_{n-1}$. Let $c'_1, \ldots, c'_r$ be the chains from the above expansion of $\mathbf{x}'$ for which $\lambda_{c'} > 0$ and let $c''_1, \ldots, c''_s$ be the chains from the above expansion of $\mathbf{x}''$ for which $\lambda_{c''} > 0$. The coordinate x'_a of $\mathbf{x}'$ then equals $\sum \lambda_{c'_i}$ and the coordinate x''_a of $\mathbf{x}''$ equals $\sum \lambda_{c''_i}$. Since x'_a and x''_a coincide with the coordinate x_a of $\mathbf{x}$, we have $\sum \lambda_{c'_i} = \sum \lambda_{c''_i}$. After relabeling (and possibly swapping $\mathbf{x}'$ and $\mathbf{x}''$) we may assume that $\lambda_{c'_1}$ is the minimum of $\{\lambda_{c'_1}, \ldots, \lambda_{c'_r}, \lambda_{c''_1}, \ldots, \lambda_{c''_s}\}$. Then we replace $\lambda_{c''_1}$ by $\lambda_{c''_1} - \lambda_{c'_1}$. Let $c_1 \in \mathrm{M}(Q)$ be the concatenation of c'_1 and c''_1. Now set $\lambda_{c_1} = \lambda_{c'_1}$ and proceed with the new coefficients and the chains $c'_2, \ldots, c'_r$ and $c''_1, \ldots, c''_s$. Clearly the sums of the coefficients of $c'_2, \ldots, c'_r$ and $c''_1, \ldots, c''_s$ still coincide. Proceeding by induction and summing over all $a \in Q_{n-1}$, one constructs an expansion $\sum \lambda_i c_i$ in terms of chains in $\mathrm{M}(Q)$ whose projection onto Q' equals $\mathbf{x}'$ and whose projection onto Q'' equals $\mathbf{x}''$. Hence $\mathbf{x} = \sum \lambda_i c_i$, and we have $\lambda_i \geq 0$ and $\sum \lambda_i = \sum_{a \in Q_{n-1}} x_a = 1$ by (7). This proves that $\mathbf{x} \in \mathrm{As}$. $\square$

The equations (7) are complete and independent when $Q = 2^{[n]}$ in the Boolean lattice, so in that case the dimension of the model polytope As is equal to $2^n - n - 1$.

Now we turn to the toric ideal I_{asc} of the ascending model. It is the kernel of the map

$$\mathbb{K}[p] \to \mathbb{K}[t], \; p_\pi \;\mapsto\; t_{a_0} t_{a_1} \cdots t_{a_n} \quad \text{for } \pi = (a_0 < \cdots < a_n) \in \mathrm{M}(Q). \tag{10}$$

If $\mathrm{rk}(Q) = 0$ then this map is injective and $I_{\mathrm{asc}} = \{0\}$, so we assume $\mathrm{rk}(Q) \geq 1$ from now on. The case $\mathrm{rk}(Q) = 1$ serves as the base case for our inductive constructions. Here the poset Q is identified with a bipartite graph on Q_0 and Q_1, and the monomial map $p_\pi \mapsto t_{a_0} t_{a_1}$ defines the toric ring associated with a bipartite graph in commutative algebra. The kernel of this map was determined in [23, Lemma 1.1]. This is well-known in algebraic statistics:

Lemma 5.2 (Ohsuhi-Hibi [23])**.** *Let Q be a graded poset of rank 1. Then a Gröbner basis of the toric ideal I_{asc} consists of all cycles in the bipartite graph Q, expressed as binomials*

$$p_{a_0<a_1} p_{a_2<a_3} \cdots p_{a_{2n-2}<a_{2n-1}} \;-\; p_{a_2<a_1} p_{a_4<a_3} \cdots p_{a_{2n}<a_{2n-1}}$$

where $a_{2n} = a_0$ and the a_i are pairwise distinct otherwise.

Now we are in a position to describe a Gröbner basis for I_{asc} when $\mathrm{rank}(Q) \geq 1$.

Theorem 5.3. *A Gröbner basis for the toric ideal I_{asc} of the ascending model on a graded poset Q is given by two classes of binomials, where the first class consists of the quadrics*

$$p_{\pi_1} \cdot p_{\pi_2} \;-\; p_{\pi'_1} \cdot p_{\pi'_2}, \tag{11}$$

where $\pi_1, \pi'_1, \pi_2, \pi'_2$ are distinct chains of at least three elements, such that $\pi_1 \cup \pi_2 = \pi'_1 \cup \pi'_2$ as multisets and $\pi_1 \cap \pi_2 = \pi'_1 \cap \pi'_2$ is nonempty. The second class consists of all binomials

$$p_{\pi_1} p_{\pi_2} \cdots p_{\pi_s} \;-\; p_{\pi'_1} p_{\pi'_2} \cdots p_{\pi'_s}, \tag{12}$$

where $\pi_1, \pi'_1, \ldots, \pi_s, \pi'_s$ are constructed as follows: Fix $i \in \{0, 1, \ldots, n-1\}$ and take any path $\gamma = (a_0 {<} a_1 {>} a_2 {<} \cdots {<} a_{2n-1} {>} a_{2n} {=} a_0)$ in the subposet $Q_{i,i+1}$ of all elements having rank i or $i+1$ in Q. Then the maximal chains π_j, π'_j for $1 \leq j \leq s$ are chosen such that

$$\begin{array}{rcl} \pi_j & = & (\, u_{j,0} < \cdots < u_{j,i} = a_{2j} < a_{2j+1} = u_{j,i+1} < \cdots < u_{j,n}\,) \\ \textit{and} \quad \pi'_j & = & (\, u'_{j,0} < \cdots < u'_{j,i} = a_{2j} < a_{2j-1} = u'_{j,i+1} < \cdots < u'_{j,n}\,). \end{array}$$

For the proof of this result we shall employ Sullivant's theory of *toric fiber products* from [27]. We briefly review that theory. Consider two polynomial rings $\mathbb{K}[\,p'\,]$ and $\mathbb{K}[\,p''\,]$ and a surjective multigrading $\phi : \{p'\} \cup \{p''\} \to \mathcal{A} \subseteq \mathbb{R}^d$, called the $\mathcal{A}$-*grading*. Then choose new variables $z_{\pi,\tau}$ for all π and τ such that $\phi(\pi) = \phi(\tau)$. For ideals I in $\mathbb{K}[\,p'\,]$ and J in $\mathbb{K}[\,p''\,]$ that are $\mathcal{A}$-homogeneous we let $I \times_{\mathcal{A}} J$ denote the kernel of the map $z_{\pi,\tau} \mapsto p'_\pi \otimes p''_\tau$ from $\mathbb{K}[\,z\,]$ to the tensor product $\mathbb{K}[\,p'\,]/I \otimes \mathbb{K}[\,p''\,]/J$.

In order to describe a Gröbner basis of $I \times_{\mathcal{A}} J$ in terms of Gröbner bases of I and J, the concept of lifting monomials turns out to be crucial [27, p. 567]. A lift of a variable p'_π is $z_{\pi\tau}$ for some τ with $\phi(\pi) = \phi(\tau)$. Now assume that $\mathcal{A}$ is linearly independent. Let $f \in \mathbb{K}[\,p'\,]$ be an $\mathcal{A}$-homogeneous polynomial. Each monomial m in f factors as $m_{a_1} \dots m_{a_r}$ where $\mathcal{A} = \{a_1, \dots, a_r\}$ and $\phi(m_{a_i}) = \deg(m_{a_i})\, a_i$. Moreover, since $\mathcal{A}$ is linearly independent, each monomial m in f gives the same number $d_i := \deg(m_{a_i})$ of variables of degree a_i (counted with multiplicity). Now choose a multisets of d_i variables p'' of degree a_i. A lift of f is then any polynomial obtained from the above choices when lifting the variables in each monomial from f in such a way that for all monomials the chosen multisets are exhausted.

Proof. We proceed by induction on $n = \operatorname{rank}(Q)$. If $n = 1$ then (11) describes an empty set of binomials and the set in (12) coincides with the Gröbner basis given in Lemma 5.2.

Now assume $n \geq 2$. As in the proof of Theorem 5.1 we split Q into the subposet $Q' = Q_0 \cup \dots \cup Q_{n-1}$ consisting of ranks $0, \dots, n-1$ and the bipartite poset $Q'' = Q_{n-1} \cup Q_n$ consisting of ranks $n-1$ and n. Assume $Q_{n-1} = \{b_1, \dots, b_r\}$. Any chain in $\mathrm{M}(Q')$ ends in an element from Q_{n-1}, and any chain from $\mathrm{M}(Q'')$ starts in an element from Q_{n-1}. We consider the polynomial ring $\mathbb{K}[\,p'\,]$ with variables p'_π for $\pi \in \mathrm{M}(Q')$ and $\mathbb{K}[\,p''\,]$ with variables p''_π for $\pi \in \mathrm{M}(Q'')$. Then we grade p'_π by $e_i \in \mathbb{R}^r$ if π ends in a_i and p''_c by $e_i \in \mathbb{R}^r$ if π begins in a_i. Note that the set of degrees $\mathcal{A} = \{e_1, \dots, e_r\}$ is linearly independent.

We write I'_{asc} for the ideal of the ascending model of Q' and I''_{asc} for the ideal of the ascending model of Q''. The toric ideal of interest to us is the fiber product $I_{\mathrm{asc}} = I'_{\mathrm{asc}} \times_{\mathcal{A}} I''_{\mathrm{asc}}$. Since $\mathcal{A}$ is linear independent, we can apply [27, Theorem 12] and the induction hypothesis to prove the claim. Sullivant's result tells us that a Gröbner basis of I_{asc} can be found by lifting Gröbner bases of the ideals I'_{asc} and I''_{asc} and by adding some quadratic relations.

By induction, I'_{asc} has a Gröbner basis $\mathcal{G}'$ consisting of elements (11) and (12). We shall lift these to binomials in I_{asc}. Let $p_{\pi_1} p_{\pi_2} - p_{\pi'_1} p_{\pi'_2}$ be a quadric (11) in $\mathcal{G}'$. Since it is $\mathcal{A}$-homogeneous, the multisets of endpoints of π_1, π_2 and π'_1, π'_2 coincide. Suppose π_1 and π'_1 have the same endpoint. In the lifting described above we need to distinguish two cases.

Case 1: π_1 and π_2 end in different endpoints. Then, for any two chains τ, τ' in Q'' starting in the endpoints of π_1 and π_2 respectively, the unique lift for these choices is

$$p_{\pi_1\tau} \cdot p_{\pi_2\tau'} \;-\; p_{\pi'_1\tau} \cdot p_{\pi'_2\tau'} \quad \in \; I_{\mathrm{asc}}. \tag{13}$$

Case 2: π_1 and π_2 end is the same endpoint. Then, for any two chains τ, τ' in Q'' starting in the common endpoint of π_1 and π_2, besides the lift (13) we also have the lift

$$p_{\pi_1\tau} \cdot p_{\pi_2\tau'} \;-\; p_{\pi'_1\tau'} \cdot p_{\pi'_2\tau} \quad \in \; I_{\mathrm{asc}}. \tag{14}$$

One easily checks that the binomials from (13) and (14) satisfy the conditions from (11).

Next, consider any binomial $p_{\pi_1}\cdots p_{\pi_s} - p_{\pi'_1}\cdots p_{\pi'_s}$ of type (12) in the Gröbner basis $\mathcal{G}'$. Since it is $\mathcal{A}$-homogeneous, the multisets $\{\phi(\pi_1),\ldots,\phi(\pi_s)\}$ and $\{\phi(\pi'_1),\ldots,\phi(\pi'_s)\}$ coincide. Now choose maximal chains $\pi''_1,\ldots,\pi''_s$ from Q'' with the same multiset of $\mathcal{A}$-degrees $\{\phi(\pi''_1),\ldots,\phi(\pi''_s)\}$. For any $\tau \in \mathfrak{S}_s$ such that $\phi(\pi'_j) = \phi(\pi''_{\tau(j)})$, the binomial

$$p_{\pi_1\pi''_1}\cdots p_{\pi_s\pi''_s} \;-\; p_{\pi'_1\pi''_{\tau(1)}}\cdots p_{\pi'_s\pi''_{\tau(s)}}$$

lies in I'_{asc} and is of type (12). All the binomials constructed by these liftings from $\mathcal{G}'$ are among the binomials described in (11) and (12) for the ideal I_{asc} we seek to generate.

We next consider a Gröbner basis $\mathcal{G}''$ for I''_{asc} that consists of binomials of type (12). Note that there are no binomials of type (11) in I''_{asc} because the poset Q'' has only rank 1. For the binomials in $\mathcal{G}''$ we use the same argument as that for the lifting of the type (12) binomials in $\mathcal{G}'$. The resulting binomials lie in I_{asc} and they are of type (12).

Finally, we add the quadratic binomials $p_{\pi_1\pi_2}p_{\pi'_1\pi'_2} - p_{\pi_1\pi'_2}p_{\pi'_1\pi_2}$ for all maximal chains $\pi_1,\pi'_1 \in \mathrm{M}(Q')$ and $\pi_2,\pi'_2 \in \mathrm{M}(Q'')$ whose $\mathcal{A}$-degrees coincide. These binomials lie in I_{asc} and they have type (11). We have shown that the lifting of the Gröbner bases $\mathcal{G}'$ for I'_{asc} and $\mathcal{G}''$ for I''_{asc} plus the additional quadrics are a subset of the binomials described in (11) and (12). Using [27, Theorem 12], we conclude that the binomials from (11) and (12) form a Gröbner basis of I_{asc}. Actually, the following converse is true as well: all binomials (11) and (12) in I_{asc} arise from I'_{asc} and I''_{asc} using the lifting procedure we described. $\square$

Corollary 5.4. *The toric algebra $\mathbb{K}[p]/I_{\mathrm{asc}}$ is normal and Cohen-Macaulay.*

Proof. Theorem 5.3 gave a Gröbner basis for I_{asc} whose leading monomials are squarefree, showing that $\mathbb{K}[p]/I_{\mathrm{asc}}$ is normal. Hochster's Theorem implies Cohen-Macaulayness. $\square$

We close with some brief remarks on the ascending model for the Boolean lattice $Q = 2^{[n]}$. In Section 2 we saw that, for $n = 3$, the ideal I_{asc} is principal with generator $p_{123}p_{231}p_{312} - p_{132}p_{213}p_{321}$. This cubic is of type (12). It represents the unique cycle in the hexagon $Q_{1,2}$.

For $n = 4$, the minimal Markov basis of the ascending model consists of 6 quadrics, 64 cubics and 93 quartics. Thus, here we encounter binomials of both types (11) and (12). The Hilbert series of the Cohen-Macaulay ring $\mathbb{K}[p]/I_{\mathrm{asc}}$ for $Q = 2^{[4]}$ is found to be

$$\frac{1 + 12t + 72t^2 + 228t^3 + 291t^4 + 168t^5 + 36t^6}{(1-t)^{12}}.$$

Based on this data we conjecture that the a-invariant of $\mathbb{K}[p]/I_{\mathrm{asc}}$ for $Q = 2^{[n]}$ is $-n!/2$.

6. The inversion Model

The inversion model is defined only in the case when Q is the distributive lattice associated with a constraint poset $\mathcal{P}$ on $[n]$. The maximal chains in Q correspond to linear extensions $\pi \in \mathcal{L}(\mathcal{P})$ of the constraint poset. These are the permutations $\pi \in \mathfrak{S}_n$ that are compatible with $\mathcal{P}$. Fix unknowns u_{ij} and v_{ij} for $1 \le i < j \le n$. Algebraically, the inversion model is defined by the toric ideal which is the kernel of the monomial map

$$p_\pi \;\mapsto\; \prod_{\substack{1\le i<j\le n \\ \pi^{-1}(i)<\pi^{-1}(j)}} u_{ij} \prod_{\substack{1\le i<j\le n \\ \pi^{-1}(i)>\pi^{-1}(j)}} v_{ij}.$$

We begin considering the *unconstrained inversion model*. By this we mean the case when $\mathcal{P}$ is an n-element antichain, so there are no constraints at all. In that unconstrained case, we have $Q = 2^{[n]}$ and our state space $\mathrm{M}(Q) = \mathfrak{S}_n = \mathcal{L}(\mathcal{P})$ consists of all $n!$ permutations.

The *Mallows model* [22] is a natural specialization of the unconstrained inversion model to a single parameter q. It is obtained by setting $u_{ij} := 1$ and $v_{ij} := q$. So, in this model, the probability of observing the permutation π is $P(\pi) = Z^{-1} q^{|\mathrm{inv}(\pi)|}$, where

$$\mathrm{inv}(\pi) \;=\; \big\{(i,j) \,:\, 1 \leq i < j \leq n,\, \pi^{-1}(i) > \pi^{-1}(j)\big\}$$

is the set of inversions of π, and Z is a normalizing constant. In contrast, our inversion model permits different parameters for the various inversions occurring in a permutation.

The model polytope for the unconstrained inversion model is a familiar object in combinatorial optimization, where it is known as the *linear ordering polytope* [13, 16]. It is known that optimizing a general linear function over the linear ordering polytope is an NP-hard problem [16]. This mirrors the fact that the facial structure of this polytope is very complicated and a complete description appears out of reach. As a result of this, we expect the toric rings associated with the inversion models to be more complicated than those studied in the previous two sections. Our study was limited to finding some computational results.

Theorem 6.1. *For $n \leq 6$ the toric ring of the unconstrained inversion model is normal and hence Cohen-Macaulay. For $n \leq 5$ it is Gorenstein and its Markov basis consists of quadrics. For $n = 6$ it is not Gorenstein and there exists a Markov basis element of degree* 3.

Proof. Computations using `4ti2` [14] show that the Markov basis for $n = 3, 4, 5$ consists of $2, 81, 3029$ quadratic binomials. We do not know whether there is a quadratic Gröbner basis for $n = 5$, or whether the ring is Koszul. The Hilbert series for $n \leq 5$ are

n	Hilbert Series
3	$(1+2t+t^2)/(1-t)^4$
4	$(1+17t+72t^2+72t^3+17t^4+t^5)/(1-t)^7$
5	$(1{+}109t{+}2966t^2{+}22958t^3{+}61026t^4{+}61026t^5{+}22958t^6{+}2966t^7{+}109t^8{+}t^9)/(1-t)^{11}$

All three numerator polynomials are symmetric. Using `normaliz` [4] one checks that the toric ring is normal in each case. Hochster's Theorem implies that it is Cohen-Macaulay. The Gorenstein property now follows from the general result that any Cohen Macaulay domain whose Hilbert series has a symmetric numerator polynomial is Gorenstein.

For $n = 6$, the computations are much harder, and they reveal that the above nice properties no longer hold. The software also found that the Hilbert series of this unconstrained inversion model is the product of $1/(1-t)^{16}$ and the remarkable numerator polynomial

$$\begin{gathered} 1 + 704\,t + 117783\,t^2 + 5125328\,t^3 + 76415229\,t^4 \\ +475189840\,t^5 + 1372165343\,t^6 + 1943081264\,t^7 + 1372165343\,t^8 + 475189840\,t^9 \\ +76416069\,t^{10} + 5127008\,t^{11} + 118623\,t^{12} + 704\,t^{14} + t^{14}. \end{gathered}$$

This polynomial is close to symmetric but not symmetric, so the ring is not Gorenstein.

In addition to 130377 quadrics, a Markov basis for $n = 6$ must contain the cubic binomial

$$p_{123456}\,p_{123645}\,p_{416253} \;-\; p_{123465}\,p_{162345}\,p_{412536}. \tag{15}$$

Indeed, a direct computation shows that these are only two cubic monomials in the fiber given by the multiset of inversions $\{(1{,}4),(2{,}4),(2{,}6),(3{,}4),(3{,}5),(3{,}6),(4{,}6),(5{,}6),(5{,}6)\}$. $\square$

We do not know whether normality holds for $n \geq 7$, but we suspect not. To address this question, we return to the general situation of an underlying constraint poset $\mathcal{P}$. The states π of the *$\mathcal{P}$-constrained inversion model* are elements of the subset $\mathcal{L}(\mathcal{P}) \subset \mathfrak{S}_n$. This inclusion corresponds to passing to some coordinate hyperplanes in the ambient space of the model polytopes. Therefore, the model polytope for the $\mathcal{P}$-constrained model is a face of the model polytope for the unconstrained model. Hence, to answer our question about normality for $n \geq 7$, it would suffice to show that the toric ring for $\mathcal{P}$ is not normal.

At present our state of knowledge about the $\mathcal{P}$-constrained inversion models is rather limited. We do not yet even have useful formula for the dimension of its model polytope. By contrast, the dimension of the unconstrained model equals $\binom{n}{2}$, as this is the dimension of the linear ordering polytope. This was shown, for example, in [26, Proposition 3.10].

We wish to mention a family of constraint posets that is important for applications of statistical ranking in data mining, e.g. in recent work of Cheng *et al.* [7]. For that application one would take $\mathcal{P}$ to be any disjoint union of a chain and an antichain.

Example 6.2. Let $n \geq 4$ and $\mathcal{P}$ be the poset consisting of the 3-chain $1 < 2 < 3$ and $n-3$ incomparable elements. If $n = 4$ then $\mathcal{L}(\mathcal{P}) = \{1234, 1243, 1423, 4123\}$ and the toric ideal I_{inv} is the zero ideal in the polynomial ring in four unknowns. If $n = 5$ then the number of states is 20 and the model polytope has dimension 7, degree 82, and the Hilbert series is

$$\frac{1 + 12t + 38t^2 + 28t^3 + 3t^4}{(1-t)^8}.$$

The Markov basis for this $\mathcal{P}$-constrained model consists of 40 quadrics:

$p_{41523}p_{51423} - p_{14523}p_{54123}$	$p_{41253}p_{51423} - p_{14253}p_{54123}$	$p_{41235}p_{51423} - p_{14235}p_{54123}$	$p_{41253}p_{51243} - p_{12453}p_{54123}$
$p_{41235}p_{51243} - p_{12435}p_{54123}$	$p_{15423}p_{51243} - p_{15243}p_{51423}$	$p_{14253}p_{51243} - p_{12453}p_{51423}$	$p_{14235}p_{51243} - p_{12435}p_{51423}$
$p_{41235}p_{51234} - p_{12345}p_{54123}$	$p_{15423}p_{51234} - p_{15234}p_{51423}$	$p_{15243}p_{51234} - p_{15234}p_{51243}$	$p_{14235}p_{51234} - p_{12345}p_{51423}$
$p_{12543}p_{51234} - p_{12534}p_{51243}$	$p_{12435}p_{51234} - p_{12345}p_{51243}$	$p_{15423}p_{45123} - p_{14523}p_{54123}$	$p_{15243}p_{45123} - p_{41523}p_{51243}$
$p_{15234}p_{45123} - p_{41523}p_{51234}$	$p_{12543}p_{45123} - p_{12453}p_{54123}$	$p_{12534}p_{45123} - p_{41253}p_{51234}$	$p_{12354}p_{45123} - p_{12345}p_{54123}$
$p_{15243}p_{41253} - p_{12543}p_{41523}$	$p_{15234}p_{41253} - p_{12534}p_{41523}$	$p_{14523}p_{41253} - p_{14253}p_{41523}$	$p_{15234}p_{41235} - p_{12354}p_{41523}$
$p_{14523}p_{41235} - p_{14235}p_{41523}$	$p_{14253}p_{41235} - p_{14235}p_{41253}$	$p_{12534}p_{41235} - p_{12354}p_{41253}$	$p_{12453}p_{41235} - p_{12435}p_{41253}$
$p_{14253}p_{15243} - p_{12453}p_{15423}$	$p_{14235}p_{15243} - p_{12435}p_{15423}$	$p_{14235}p_{15234} - p_{12345}p_{15423}$	$p_{12543}p_{15234} - p_{12534}p_{15243}$
$p_{12435}p_{15234} - p_{12345}p_{15243}$	$p_{12543}p_{14523} - p_{12453}p_{15423}$	$p_{12534}p_{14523} - p_{14253}p_{15234}$	$p_{12354}p_{14523} - p_{12345}p_{15423}$
$p_{12534}p_{14235} - p_{12354}p_{14253}$	$p_{12453}p_{14235} - p_{12435}p_{14253}$	$p_{12435}p_{12534} - p_{12345}p_{12543}$	$p_{12354}p_{12453} - p_{12345}p_{12543}$

It can be asked which $\mathcal{P}$-constrained inversion model have a Markov basis of quadrics and, more generally, which degrees appear in a Markov basis. We confirmed the quadratic Markov basis for all posets $\mathcal{P}$ on $n \leq 4$ elements, all on $n = 5$ elements arising by adding one incomparable element to a poset on 4 elements, and all unconstrained models for $n \leq 5$.

Interestingly, the notion of inversion model changes if we define $i < j$ to be an inversion if $\pi(i) > \pi(j)$. The defining monomial map for this *alternative inversion model* equals

$$p_\pi \;\mapsto \prod_{\substack{1 \leq i < j \leq n \\ \pi(i) < \pi(j)}} u_{ij} \prod_{\substack{1 \leq i < j \leq n \\ \pi(i) > \pi(j)}} v_{ij} \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}).$$

For the 3-chain $1 < 2 < 3$ with two incomparable elements, the Markov basis now consists of

$$\begin{array}{llll}
p_{15243}p_{51423} - p_{12543}p_{54123} & p_{15234}p_{51423} - p_{12534}p_{54123} & p_{15423}p_{51243} - p_{12543}p_{54123} & p_{15234}p_{51243} - p_{12354}p_{54123} \\
p_{12534}p_{51243} - p_{12354}p_{51423} & p_{15423}p_{51234} - p_{12534}p_{54123} & p_{15243}p_{51234} - p_{12354}p_{54123} & p_{15234}p_{51234} - p_{12345}p_{54123} \\
p_{12543}p_{51234} - p_{12354}p_{51423} & p_{12534}p_{51234} - p_{12345}p_{51423} & p_{12354}p_{51234} - p_{12345}p_{51243} & p_{12534}p_{15243} - p_{12354}p_{15423} \\
p_{12543}p_{15234} - p_{12354}p_{15423} & p_{12534}p_{15234} - p_{12345}p_{15423} & p_{12354}p_{15234} - p_{12345}p_{15243} & p_{12354}p_{12534} - p_{12345}p_{12543} \\
p_{12435}p_{12453} - p_{12345}p_{12543}, & & &
\end{array}$$

and $p_{14235}p_{14253}p_{14523} - p_{12345}p_{15243}p_{15423}$, and $p_{41235}p_{41253}p_{41523}p_{45123} - p_{12345}p_{51243}p_{51423}p_{54123}$. So, unlike in Example 6.2, this Markov basis is not quadratic. The Hilbert series equals

$$\frac{1 + 9t + 28t^2 + 51t^3 + 66t^4 + 63t^5 + 44t^6 + 21t^7 + 5t^8}{(1-t)^{11}}.$$

Note that, if $\mathcal{L}(\mathcal{P})$ is closed under taking inversions, then this model coincides with the normal $\mathcal{P}$-constraint inversion model up to a relabeling. This holds for the unconstrained inversion model. All examples tested in this alternative model had normal model polytopes.

7. Plackett-Luce Model and Bradley-Terry model

The Plackett-Luce model is a non-toric model on the set $\mathcal{L}(\mathcal{P})$ of permutations $\pi \in \mathfrak{S}_n$ that are consistent with a given constraint poset $\mathcal{P}$ on $[n]$. It can be defined by the map

$$p_\pi \;\mapsto\; \prod_{i=1}^{n-1} \frac{1}{\sum_{j=1}^{i} \theta_{\pi(j)}} \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}). \tag{16}$$

We denote this model by $\mathrm{PL}_\mathcal{P}$ and its homogeneous ideal by $I_{\mathrm{PL}_\mathcal{P}}$. Thus $I_{\mathrm{PL}_\mathcal{P}}$ is the kernel of the ring map $\mathbb{R}[\,p_\pi : \pi \in \mathcal{L}(\mathcal{P})\,] \to \mathbb{R}(\theta_1, \theta_2, \ldots, \theta_n)$ defined by the formula (16). The formula shows that the Plackett-Luce model is a submodel of the ascending model on $\mathcal{L}(\mathcal{P})$. In fact, the ascending model is the *toric closure* of the Plackett-Luce model, by which we mean that $\mathrm{As}_\mathcal{P}$ is the smallest toric model containing $\mathrm{PL}_\mathcal{P}$. The specialization map is

$$t_{\pi(\{1,2,\ldots,i\})} \quad\mapsto\quad \big(\theta_{\pi(1)} + \theta_{\pi(2)} + \cdots + \theta_{\pi(i)}\big)^{-1}.$$

We fix $\mathbb{K} = \mathbb{C}$ and regard the Plackett-Luce model $\mathrm{PL}_\mathcal{P}$ as a projective variety in $\mathbb{P}^{|\mathcal{L}(\mathcal{P})|-1}$.

Of course, in order for $\mathrm{PL}_\mathcal{P}$ to be properly defined as a statistical model, its probabilities should sum to 1. For this we would need to identify the normalizing constant, which is the image of $\sum_{\pi \in \mathcal{L}(\mathcal{P})} p_\pi$ under the map (16). A formula for this quantity can be derived, for many situations of interest, from equations (25) and (26) in Hunter's article [18]. Still, it is instructive to verify that $\sum_{\pi \in \mathfrak{S}_n} p_\pi$ is mapped to $\frac{1}{\theta_1 \cdots \theta_n}$ under the ring map in (16).

Let us begin by examining the unconstrained case when $\mathcal{P}$ is an antichain, $Q = 2^{[n]}$ and $\mathcal{L}(\mathcal{P}) = \mathrm{M}(Q) = \mathfrak{S}_n$. This is the Plackett-Luce model PL_n familiar from the statistics literature [18, 21, 25]. With the correct normalizing constant, its parametrization equals

$$p_\pi \;\mapsto\; \prod_{i=1}^{n} \frac{\theta_{\pi(i)}}{\sum_{j=1}^{i} \theta_{\pi(j)}} \qquad \text{for } \pi \in \mathfrak{S}_n. \tag{17}$$

This defines a polynomial map from the non-negative orthant $\mathbb{R}^n_{\geq 0}$ to the $(n!-1)$-dimensional simplex of probability distributions on the symmetric group $\mathfrak{S}_n$. We shall regard PL_n as a complex projective variety in the ambient $\mathbb{P}^{n!-1}$. The dimension of this variety is $n-1$.

Example 7.1 ($n = 3$). The Plackett-Luce model PL_3 is a surface of degree 7 embedded in 5-dimensional projective space $\mathbb{P}^5$. The parameterization (16) of that surface is equivalent to

$$\begin{array}{lll} p_{123} \mapsto \theta_2\theta_3(\theta_1{+}\theta_3)(\theta_2{+}\theta_3), & p_{132} \mapsto \theta_2\theta_3(\theta_1{+}\theta_2)(\theta_2{+}\theta_3), & p_{213} \mapsto \theta_1\theta_3(\theta_1{+}\theta_3)(\theta_2{+}\theta_3), \\ p_{231} \mapsto \theta_1\theta_3(\theta_1{+}\theta_2)(\theta_1{+}\theta_3), & p_{312} \mapsto \theta_1\theta_2(\theta_1{+}\theta_2)(\theta_2{+}\theta_3), & p_{321} \mapsto \theta_1\theta_2(\theta_1{+}\theta_2)(\theta_1{+}\theta_3). \end{array}$$

The defining ideal I_{PL_3} of PL_3 is minimally generated by three quadratic polynomials, in addition to the familiar cubic binomial that specifies the ambient ascending model:

$$I_{\mathrm{PL}_3} \;=\; \left\langle \begin{array}{c} p_{123}(p_{321}+p_{231}) - p_{213}(p_{132}+p_{312}),\; p_{312}(p_{123}+p_{213}) - p_{132}(p_{231}+p_{321}), \\ p_{231}(p_{132}+p_{312}) - p_{321}(p_{123}+p_{213}), \quad p_{123}p_{231}p_{312} - p_{132}p_{321}p_{213} \end{array} \right\rangle.$$

The singular locus of PL_3 consists of the three isolated points $e_{321} - e_{231}$, $e_{123} - e_{213}$ and $e_{132} - e_{312}$ in $\mathbb{P}^5$. In particular, there are no singular points with non-negative coordinates, so this statistical model is a smooth surface in the 5-dimensional probability simplex.

From the point of view of algebraic geometry, our parametrization map represents the blow-up of the projective plane $\mathbb{P}^2$ at the following configuration of nine special points:

$$\begin{array}{ccc} (0:0:1) & (0:1:0) & (1:0:0) \\ (1:-1:0) & (1:0:-1) & (0:1:-1) \\ (1:1:-1) & (1:-1:1) & (-1:1:1) \end{array} \tag{18}$$

This configuration has three 4-point lines and four 3-point lines. The map blows down the three 4-point lines, and this creates a rational surface in $\mathbb{P}^5$ with three singular points.

From the point of view of commutative algebra, one might ask whether the four generators of the ideal I_{PL_3} form a Gröbner basis with respect to some term order. A computation reveals that this is not the case. However, we do get a square-free Gröbner basis for the lexicographic term order with $p_{123}{>}p_{132}{>}p_{213}{>}p_{231}{>}p_{312}{>}p_{321}$. The initial ideal equals

$$\begin{aligned} \mathrm{in}_{\mathrm{lex}}(I_{\mathrm{PL}_3}) \;=\; & \langle p_{123}, p_{132}, p_{231}\rangle \;\cap\; \langle p_{123}, p_{132}, p_{312}\rangle \;\cap\; \langle p_{123}, p_{132}, p_{213}\rangle \;\cap \\ & \langle p_{123}, p_{213}, p_{231}\rangle \;\cap\; \langle p_{123}, p_{213}, p_{312}\rangle \;\cap\; \langle p_{123}, p_{312}, p_{321}\rangle \;\cap\; \langle p_{231}, p_{312}, p_{321}\rangle. \end{aligned}$$

This represents a simplicial complex of seven triangles, listed in a shelling order, so I_{PL_3} is Cohen-Macaulay. The Hilbert series of the ring $\mathbb{R}[p]/I_{\mathrm{PL}_3}$ equals $(1{+}3t{+}3t^2)/(1-t)^3$. □

Example 7.2 ($n = 4$). The Plackett-Luce model PL_4 is a threefold of degree 191 in $\mathbb{P}^{23}$. It is obtained from $\mathbb{P}^3$ by blowing up 55 lines. The homogeneous prime ideal I_{PL_4} that defines PL_4 is minimally generated by 105 quadrics and 75 cubics. Its Hilbert series equals

$$\frac{1 + 20t + 105t^2 + 65t^3}{(1-t)^4}.$$

It is tempting to boldly conjecture that I_{PL_n} is generated in degree 2 and 3 for $n \geq 5$. □

Let us now turn to the general Plackett-Luce model with a given constraint poset $\mathcal{P}$, so only permutations π in $\mathcal{L}(\mathcal{P})$ are allowed. The model $\mathrm{PL}_{\mathcal{P}}$ is obtained from PL_n by projecting onto those coordinates. Algebraically, the prime ideal $I_{\mathcal{P}}$ is obtained from I_{PL_n} by eliminating all unknowns p_π where π is a permutation that is not compatible with $\mathcal{P}$.

Example 7.3. Let $n = 4$ and let $\mathcal{P}$ be the poset with two covering relations 1<2 and 3<4. The corresponding distributive lattice $\mathcal{L}(\mathcal{P})$ is the product of two chains of length 3. Note that $\mathcal{L}(\mathcal{P})$ has six maximal chains, namely, the permutations that respect $1 < 2$ and $3 < 4$. The corresponding unknowns are mapped to products of four linear forms as follows:

$$\begin{array}{ll} p_{1234} \mapsto \theta_3(\theta_1+\theta_3)(\theta_3+\theta_4)(\theta_1+\theta_3+\theta_4), & p_{1324} \mapsto \theta_3(\theta_1+\theta_2)(\theta_3+\theta_4)(\theta_1+\theta_3+\theta_4), \\ p_{1342} \mapsto \theta_3(\theta_1+\theta_2)(\theta_3+\theta_4)(\theta_1+\theta_2+\theta_3), & p_{3124} \mapsto \theta_1(\theta_1+\theta_2)(\theta_3+\theta_4)(\theta_1+\theta_3+\theta_4), \\ p_{3142} \mapsto \theta_1(\theta_1+\theta_2)(\theta_3+\theta_4)(\theta_1+\theta_2+\theta_3), & p_{3412} \mapsto \theta_1(\theta_1+\theta_2)(\theta_1+\theta_3)(\theta_1+\theta_2+\theta_3). \end{array}$$

These reducible quartics meet in nine lines in $\mathbb{P}^3$, so the parametrization of $\mathrm{PL}_{\mathcal{P}}$ blows these up. The ideal $I_{\mathcal{P}}$ is complete intersection. Its minimal generators are the cubic

$$p_{1234}p_{1342}p_{3142} + p_{1234}p_{3142}^2 + p_{1234}p_{3142}p_{3412} - p_{1234}p_{1324}p_{3412} - p_{1324}^2p_{3412} - p_{1324}p_{3124}p_{3412}$$

and the binomial quadric $p_{1342}p_{3124} - p_{1324}p_{3142}$ that defines the bottom model on $\mathcal{P}$. □

The following is our first result in this section. It should be useful for obtaining information about the $(n-1)$-dimensional variety $\mathrm{PL}_{\mathcal{P}}$ and its homogeneous prime ideal $I_{\mathcal{P}}$.

Theorem 7.4. *The parameterization $\mathbb{P}^{n-1} \to \mathrm{PL}_{\mathcal{P}} \subset \mathbb{P}^{|\mathcal{L}(\mathcal{P})|-1}$ of the Plackett-Luce model on the poset $\mathcal{P}$ is given geometrically as the blowing up of $\mathbb{P}^{n-1}$ along an arrangement of linear subspaces of codimension* 2. *These subspaces are defined by the equations $\sum_{i\in A}\theta_i = \sum_{j\in B}\theta_j = 0$ where $\{A, B\}$ runs over all incomparable pairs in the distributive lattice on $\mathcal{P}$.*

Proof. Let $\mathbb{R}[t]$ denote the polynomial ring of parameters in the ascending model (10). Its indeterminates are t_A where A runs over subsets of $[n]$ that are order ideals in $\mathcal{P}$. We define M to be the Stanley-Reisner ideal of the distributive lattice of order ideals in $\mathcal{P}$. This is the ideal in $\mathbb{R}[p]$ generated by products t_At_B where A and B are incomparable, meaning that neither $A \subset B$ nor $B \subset A$ holds. The *Alexander dual* of M is the monomial ideal

$$M^* \;=\; \bigcap_{\{A,B\}} \langle\, t_A \,,\, t_B \,\rangle,$$

where the intersection is over all incomparable pairs of order ideals. The generators of M^* correspond to the associated primes of M, so they are indexed by compatible permutations $\pi \in \mathcal{L}(\mathcal{P})$. Interpreting π as a maximal chain of order ideals, that correspondence is

$$p_\pi \;\mapsto\; \prod_{A\notin\pi} t_A \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}). \tag{19}$$

The arrangement of subspaces described in the statement of Theorem 7.4 is the image of the variety of M^* under the map to $\mathbb{P}^{n-1}$ defined by $t_A = \sum_{i\in A}\theta_i$. By substituting this into (19) we see that the blow-up along that subspace arrangement is defined by the map

$$p_\pi \;\mapsto\; \prod_{A\notin\pi}\Big(\sum_{i\in A}\theta_i\Big) \;=\; \mathrm{const}\cdot\prod_{A\in\pi}\frac{1}{\sum_{i\in A}\theta_i} \qquad \text{for } \pi \in \mathcal{L}(\mathcal{P}). \tag{20}$$

This is precisely the defining parametrization (16) of the Plackett-Luce model $\mathrm{PL}_{\mathcal{P}}$. □

Example 7.5. Let $n = 4$ and $\mathcal{P}$ as in Example 7.3. Then the above Stanley-Reisner ideal is

$$M \;=\; \langle\, t_1t_3,\, t_3t_{12},\, t_{12}t_{13},\, t_1t_{34},\, t_{12}t_{34},\, t_{13}t_{34},\, t_{34}t_{123},\, t_{12}t_{134},\, t_{123}t_{134}\,\rangle$$

and its Alexander dual is

$$M^* \;=\; \langle\, t_3t_{13}t_{34}t_{134},\, t_3t_{12}t_{34}t_{123},\, t_1t_{12}t_{34}t_{123},\, t_3t_{12}t_{34}t_{134},\, t_1t_{12}t_{34}t_{134},\, t_1t_{12}t_{13}t_{123}\,\rangle.$$

The model $\mathrm{PL}_{\mathcal{P}}$ is the blow-up of $\mathbb{P}^3$ at nine lines, one for each of the generators of M. $\square$

Each of our unconstrained ranking models was considered as a subvariety of the complex projective space $\mathbb{P}^{n!-1}$. If K is any k-element subset of $[n]$ then we obtain a natural rational map $\mathbb{P}^{n!-1} \to \mathbb{P}^{k!-1}$ which records the probabilities for each of the k orderings of K only. Statistically, this map corresponds to *marginalization* for the induced orderings on K. We can now take the direct product of all of these maps, where K runs over all $\binom{n}{k}$ subsets of cardinality k in $[n]$. The resulting map into a product of projective spaces,

$$\mathbb{P}^{n!-1} \;\longrightarrow\; (\mathbb{P}^{k!-1})^{\binom{n}{k}}, \tag{21}$$

is called the *complete marginalization map of order* k. For example, if $n = 3$ and $k = 2$ then we are mapping into a product of three projective lines, with coordinates $(q_{12} : q_{21})$, $(q_{13} : q_{31})$ and $(q_{23} : q_{32})$ respectively, and the complete marginalization is the rational map $\mathbb{P}^5 \longrightarrow \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ which is given in coordinates as follows:

$$\begin{aligned}
(q_{12} : q_{21}) \;&=\; (p_{123} + p_{132} + p_{312} : p_{213} + p_{231} + p_{321})\,,\\
(q_{13} : q_{31}) \;&=\; (p_{132} + p_{123} + p_{213} : p_{312} + p_{321} + p_{231})\,,\\
(q_{23} : q_{32}) \;&=\; (p_{123} + p_{213} + p_{231} : p_{132} + p_{312} + p_{321}).
\end{aligned}$$

We shall refer to the complete marginalization of order 2 as the *pairwise marginalization*.

Example 7.6. The pairwise marginalization of the Plackett-Luce surface $\mathrm{PL}_3 \subset \mathbb{P}^5$ is the surface in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ that is defined by the binomial equation $q_{12}q_{23}q_{31} \;=\; q_{21}q_{32}q_{13}$. The composition of the map in Example 7.1 with the map in (21) is a toric rational map $\mathbb{P}^2 \to \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ that blows up the three coordinate points (1:0:0), (0:1:0) and (0:0:1). $\square$

It is worthwhile, both algebraically and statistically, to study the various marginalizations of the Csiszár model, ascending model, the inversion model and the Plackett-Luce model. Of particular interest is the pairwise marginalization of the Plackett-Luce model. This is known in the literature as the *Bradley-Terry model* [18]. All of these marginalized models make sense relative to a fixed constraint poset $\mathcal{P}$. Here, we regard each k-set K as subposet of $\mathcal{P}$ and we write the corresponding marginalization map as

$$\mathbb{P}^{|\mathcal{L}(\mathcal{P})|-1} \;\longrightarrow\; \mathbb{P}^{|\mathcal{L}(K)|-1}. \tag{22}$$

The complete k-th marginalization is the image of the direct product of these maps, as K runs over all k-sets. For convenience, we shall here remove those k-sets K that are totally ordered in $\mathcal{P}$ because the corresponding maps in (22) are constant when $|\mathcal{L}(K)| = 1$.

We conclude this article with the following algebraic characterization of the Bradley-Terry model. We write $\mathcal{P}^c$ for the bidirected graph on $[n]$ where (i, j) is a directed edge if

i and j are incomparable in $\mathcal{P}$. Each circuit $i_1, i_2, \ldots, i_r, i_1$ in $\mathcal{P}^c$ is encoded as a binomial:

$$q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{r-1} i_r} q_{i_r i_1} \; - \; q_{i_2 i_1} q_{i_3 i_2} \cdots q_{i_r i_{r-1}} q_{i_1 i_r}. \tag{23}$$

These binomials define hypersurfaces in $\mathbb{P}^{\binom{n}{2}}$. For instance, the model in Example 7.6 is the toric hypersurface in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ thus associated to a 3-cycle.

The theorem below refers to *unimodular Lawrence ideals*. This class of toric ideals was introduced and studied by Bayer *et al.* in [2]. The associated toric varieties live naturally in a product of projective lines $\mathbb{P}^1 \times \cdots \times \mathbb{P}^1$. The case of interest here is that of unimodular Lawrence ideals arising from graphs. For these ideals and their syzygies we refer to [2, §5].

Theorem 7.7. *The Bradley-Terry model with constraints $\mathcal{P}$ is toric. It is defined by the unimodular Lawrence ideal whose generators are the circuits (23) in the bidirected graph $\mathcal{P}^c$.*

From this result we can now determine the commutative algebra invariants of the Bradley-Terry model, such as its Hilbert series in the $\mathbb{Z}^n$-grading and its multidegree.

Proof. Following [18], the parametrization of the Bradley-Terry model can be written as

$$q_{ij} \;\mapsto\; \frac{\theta_j}{\theta_i + \theta_j} \qquad \text{for } i, j \text{ incomparable in } \mathcal{P}. \tag{24}$$

Let $\rho_{\{i,j\}}$ be new unknowns indexed by unordered pairs $\{i,j\} \subset [n]$. The unimodular Lawrence ideal associated with the bidirected graph $\mathcal{P}^c$ is the kernel of the monomial map

$$q_{ij} \;\mapsto\; \rho_{\{i,j\}} \cdot \theta_j \qquad \text{for } i, j \text{ incomparable in } \mathcal{P}. \tag{25}$$

The specialization $\rho_{\{i,j\}} = (\theta_i + \theta_j)^{-1}$ shows that the ideal $I_{\mathrm{BT}_\mathcal{P}}$ of the Bradley-Terry model is contained the unimodular Lawrence ideal generated by the circuits (23). In addition, the ideal $I_{\mathrm{BT}_\mathcal{P}}$ contains the linear polynomials $q_{ij} + q_{ji} - 1$. These represent the fact that, in any compatible ranking π, either item i ranks before item j or vice versa, but not both.

Let J be the ideal generated by the circuits (23) and these linear polynomials. We have seen that $J \subseteq I_{\mathrm{BT}_\mathcal{P}}$, and we are claiming that equality holds. But this follows by observing that both ideals are prime, and their varieties have the same dimension, namely $n-1$. □

Acknowledgments

We very grateful to Winfried Bruns and Raymond Hemmecke for their substantial help with the computational results in Theorem 6.1. Using the developers' versions of `Normaliz` [4] and `4ti2` [14] respectively, they succeeded in computing the Hilbert series of the inversion model for $n = 6$ and in finding the cubic Markov basis element (15). We also thank Eyke Hüllermeier and Seth Sullivant for helpful conversations. Bernd Sturmfels was partially supported by the National Science Foundation (DMS-0757207 and DMS-0968882).

References

[1] A. Barvinok: *A Course in Convexity*, Graduate Studies in Mathematics, **54**, AMS, Providence, 2002.

[2] D. Bayer, S. Popescu and B. Sturmfels: Syzygies of unimodular Lawrence ideals, *J. Reine Angew. Math.* **534** (2001) 169–186.

[3] N. Beerenwinkel, N. Eriksson and B. Sturmfels: Evolution on distributive lattices, *Journal of Theoretical Biology* **242** (2006) 409–420.

[4] W. Bruns, B. Ichim and C. Söger: *Normaliz* – software for affine monoids, vector configurations, lattice polytopes, and rational cones, `http://www.mathematik.uni-osnabrueck.de/normaliz/`, 2010.

[5] E.R. Canfield and B.D. McKay: The asymptotic volume of the Birkhoff polytope, *J. Analytic Comb.* **4** (2009) article #2.

[6] C.S. Chan, D.P. Robbins and D.S. Yuen: On the volume of a certain polytope, *Experiment. Math.* **9** (2000) 91–99.

[7] W. Cheng, K. Dembczynski and E. Hüllermeier: Label ranking based on the Placket-Luce model, *Proc. ICML-2010, International Conference on Machine Learning*, Haifa, Israel, June 2010.

[8] V. Csiszár: Markov bases of conditional independence models for permutations, *Kybernetica* **45** (2009) 249-260.

[9] V. Csiszár: On L-decomposability of random permutations, *J. Math. Psychology* **53** (2009) 294-297.

[10] P. Diaconis and N. Eriksson: Markov bases for noncommutative Fourier analysis of ranked data, *J. of Symbolic Computation* **41** (2006) 182–195.

[11] P. Diaconis and B. Sturmfels: Algebraic algorithms for sampling from conditional distributions, *Ann. Stat.* **26** (1998) 363-397.

[12] M. Drton, B. Sturmfels and S. Sullivant: *Lectures on Algebraic Statistics*, Oberwolfach Seminars, Vol 39, Birkhäuser, Basel, 2009.

[13] S. Fiorini: $\{0, \frac{1}{2}\}$-cuts and the linear ordering problem: surfaces that define facets, *SIAM J. Discrete Math.* **20** (2006), 893–912.

[14] 4ti2 team: *4ti2* – A software package for algebraic, geometric and combinatorial problems in linear spaces, available at `www.4ti2.de`.

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

[16] M. Grötschel, M. Jünger and G. Reinelt: Facets of the linear ordering polytope, *Math. Program.* **33** (1985), 43–60.

[17] G. Hommel, F. Bretz and W. Maurer: Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies, *Statist. Med.* **26** (2007) 4063-4073.

[18] D.R. Hunter: MM algorithms for generalized Bradley-Terry models, *Ann. Stat.* **32** (2004) 384-406.

[19] A. Katsabekis and A. Thoma: Parametrizations of toric varieties over any field, *J. Algebra* **308** (2007) 751–763.

[20] M. Lenz: Toric ideals of flow polytopes, preprint, `arXiv:0801.0495`.

[21] R.D. Luce: *Individual Choice Behavior*, Wiley, New York, 1959.

[22] J.I. Marden: *Analyzing and Modeling Rank Data*, Monographs on Statistics and Applied Probability, **64**, Chapman & Hall, London, 1995

[23] H. Ohsugi and T. Hibi: Toric ideals generated by quadratic binomials, *J. Algebra* **218** (1998) 509–527.

[24] L. Pachter and B. Sturmfels: *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.

[25] R.L. Plackett: Random permutations. *J. R. Stat. Soc., Ser. B* **30** (1968) 517–534.

[26] V. Reiner, F. Saliola and V. Welker: Spectra of symmetrized shuffling operators, manuscript, 2010.

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

[28] D. Zeilberger: Proof of a conjecture of Chan, Robbins, and Yuen. In: Orthogonal polynomials: numerical and symbolic algorithms (Leganés, 1998), *Electron. Trans. Numer. Anal.* **9** (1999).

Department of Mathematics, University of California, Berkeley, CA 94720, USA
E-mail address: `bernd@math.berkeley.edu`

Fachbereich Mathematik und Informatik, Philipps-Universität, 35032 Marburg, Germany
E-mail address: `welker@mathematik.uni-marburg.de`