QUASI-INDEPENDENCE MODELS WITH RATIONAL MAXIMUM LIKELIHOOD ESTIMATES

JANE IVY COONS AND SETH SULLIVANT

ABSTRACT. We classify the two-way independence quasi-independence models (or independence models with structural zeros) that have rational maximum likelihood estimators, or MLEs. We give a necessary and sufficient condition on the bipartite graph associated to the model for the MLE to be rational. In this case, we give an explicit formula for the MLE in terms of combinatorial features of this graph. We also use the Horn uniformization to show that for general log-linear models \( M \) with rational MLE, any model obtained by restricting to a face of the cone of sufficient statistics of \( M \) also has rational MLE.

1. INTRODUCTION

Huh [12] classified the varieties with rational maximum likelihood estimate using Kapranov’s Horn uniformization [14]. In spite of the classification, it can be difficult to tell a priori whether a given model has rational MLE, or not. Duarte, Marigliano, and Sturmfels [8] have since applied Huh’s ideas to varieties that are the closure of discrete statistical models. In the present paper, we study this problem for a family of discrete statistical model called quasi-independence models, also commonly known as independence models with structural zeros. Because quasi-independence models have a simple structure whose description is determined by a bipartite graph, this is a natural test case for trying to apply Huh’s theory. Our complete classification of quasi-independence models with rational MLE is the main result of the present paper (Theorems 1.3 and 5.4).

Let \( X \) and \( Y \) be two discrete random variables with \( m \) and \( n \) states, respectively. Quasi-independence models describe the situation in which some combinations of states of \( X \) and \( Y \) cannot occur together, but \( X \) and \( Y \) are otherwise independent of one another. This condition is known as quasi-independence in the statistics literature [4]. Quasi-independence models are basic models that arise in data analysis with log-linear models. For example, quasi-independence models arise in the biomedical field as rater agreement models [11, 15] and in engineering to model system failures at nuclear plants [6]. There is a great deal of literature regarding hypothesis testing under the assumption of quasi-independence, see, for example, [2, 10, 17]. Results about existence and uniqueness of the maximum likelihood estimate in quasi-independence models as well as explicit computations in some cases can be found in [4, Chapter 5].

In order to define quasi-independence models, let \( S \subset [m] \times [n] \) be a set of indices, where \( [m] = \{1, 2, \ldots, m\} \). These correspond to a matrix with structural zeros whose observed entries are given by the indices in \( S \). We often use \( S \) to refer to both the set of indices and the matrix representation of this set and abbreviate the ordered pairs \((i, j)\) in

Date: June 15, 2020.
S by $ij$. For all $r$, we denote by $\Delta_{r-1}$ the open $(r - 1)$-dimensional probability simplex in $\mathbb{R}^r$,
\[ \Delta_{r-1} := \{ x \in \mathbb{R}^r \mid x_i > 0 \text{ for all } i \text{ and } \sum_{i=1}^r x_i = 1 \}. \]

**Definition 1.1.** Let $S \subset [m] \times [n]$. Index the coordinates of $\mathbb{R}^{m+n}$ by $(s_1, \ldots, s_m, t_1, \ldots, t_n) = (s, t)$. Let $\mathbb{R}^S$ denote the real vector space of dimension $\#S$ whose coordinates are indexed by $S$. Define the monomial map $\phi^S : \mathbb{R}^{m+n} \to \mathbb{R}^S$ by
\[ \phi^S_{ij}(s, t) = s_it_j. \]

The quasi-independence model associated to $S$ is the model,
\[ M_S := \phi^S(\mathbb{R}^{m+n}) \cap \Delta_{\#S-1}. \]

We note that the Zariski closure of $M_S$ is a toric variety since it is parametrized by monomials. To any quasi-independence model, we can associate a bipartite graph in the following way.

**Definition 1.2.** The bipartite graph associated to $S$, denoted $G_S$, is the bipartite graph with independent sets $[m]$ and $[n]$ with an edge between $i$ and $j$ if and only if $(i, j) \in S$. The graph $G_S$ is chordal bipartite if every cycle of length greater than or equal to 6 has a chord. The graph $G_S$ is doubly chordal bipartite if every cycle of length greater than or equal 6 has at least two chords. We say that $S$ is doubly chordal bipartite if $G_S$ is doubly chordal bipartite.

Let $u \in \mathbb{N}^S$ be a vector of counts of independent, identically distributed (iid) data. The maximum likelihood estimate, or MLE, for $u$ in $M_S$ is the distribution $\hat{p} \in M_S$ that maximizes the probability of observing the data $u$ over all distributions in the model. We describe the maximum likelihood estimation problem in more detail in Section 2. We say that $M_S$ has rational MLE if for generic choices of $u$, the MLE for $u$ in $M_S$ can be written as a rational function in the entries of $u$. We can now state the key result of this paper.

**Theorem 1.3.** Let $S \subset [m] \times [n]$ and let $M_S$ be the associated quasi-independence model. Let $G_S$ be the bipartite graph associated to $S$. Then $M_S$ has rational maximum likelihood estimate if and only if $G_S$ is doubly chordal bipartite.

Theorem 5.4 is a strengthened version of Theorem 1.3 in which we give an explicit formula for the MLE when $G_S$ is doubly chordal bipartite. The outline of the rest of the paper is as follows. In Section 2 we introduce general log-linear models and their MLEs and discuss some key results on these topics. In Section 3 we discuss the notion of a facial submodel of a log-linear model and prove that facial submodels of models with rational MLE also have rational MLE. In Section 4 we apply the results of Section 3 to show that if $G_S$ is not doubly chordal bipartite, then $M_S$ does not have rational MLE. The main bulk of the paper is in Sections 5, 6 and 7 where we show that if $G_S$ is doubly chordal bipartite, then the MLE is rational and we give an explicit formula for it. Section 5 covers combinatorial features of doubly chordal bipartite graphs and gives the statement of the main Theorem 5.4. Sections 6 and 7 are concerned with the verification that the formula for the MLE is correct.
2. LOG-LINEAR MODELS AND THEIR MAXIMUM LIKELIHOOD ESTIMATES

In this section, we collect some results from the literature on log-linear models and maximum likelihood estimation in these models. These results will be important tools in the proof of Theorem 5.4.

Let $A \in \mathbb{Z}^{d \times r}$ with entries $a_{ij}$. Denote by $\mathbf{1}$ the vector of all ones in $\mathbb{Z}^r$. We assume throughout that $\mathbf{1} \in \text{rowspan}(A)$.

**Definition 2.1.** The log-linear model associated to $A$ is the set of probability distributions,

$$M_A := \{ p \in \Delta_{r-1} \mid \log p \in \text{rowspan}(A) \}.$$ 

Algebraic and combinatorial tools are well-suited for the study of log-linear models since these models have monomial parametrizations. Define the map $\phi^A : \mathbb{R}^d \to \mathbb{R}^r$ by

$$\phi^A_j(t_1, \ldots, t_d) = \prod_{i=1}^d t_i^{a_{ij}}.$$ 

Then we have that $M_A = \phi^A(\mathbb{R}^d) \cap \Delta_{r-1}$. Background on log-linear models can be found in [18, Chapter 6.2]. Denote by $C[p] := \mathbb{C}[p_1, \ldots, p_r]$ the polynomial ring in $r$ indeterminates. Let $I_A \subset C[p]$ denote the vanishing ideal of $\phi^A(\mathbb{R}^d)$ over the algebraically closed field $\mathbb{C}$. Since $\phi^A$ is a monomial map, $I_A$ is a toric ideal. For this reason, $M_A$ is also known as a toric model. Some key properties of $I_A$ are summarized in the following proposition.

**Proposition 2.2** ([18], Proposition 6.2.4). The toric ideal $I_A$ is a binomial ideal and

$$I_A = \langle p^u - p^v \mid u, v \in \mathbb{N}^r \text{ and } Au = Av \rangle.$$ 

If $\mathbf{1} \in \text{rowspan}(A)$, then $I_A$ is homogeneous.

Note that the quasi-independence model associated to a set $S \subset [m] \times [n]$ is a log-linear model with respect to matrix $A(S)$ constructed in the following way. We have $A(S) \in \mathbb{Z}^{(m+n) \times \#S}$. The $ij$ column of $A(S)$, denoted $a^{ij}$, has $k$th entry:

$$a_{ik}^{ij} = \begin{cases} 1, & \text{if } k = i \\ 1, & \text{if } k = m + j \\ 0, & \text{otherwise}. \end{cases}$$

In this way, $M_S = M_{A(S)}$. Note that $\mathbf{1} \in \text{rowspan}(A(S))$ for all $S$, since it can be written as the sum of the first $m$ rows of $A(S)$.

Given independent, identically distributed (iid) data $u \in \mathbb{N}^r$, we wish to infer the distribution $p \in M_A$ that is “most likely” to have generated it. This is the central problem of maximum likelihood estimation.

**Definition 2.3.** Let $M$ be a discrete statistical model in $\mathbb{R}^r$ and let $u \in \mathbb{N}^r$ be an iid vector of counts. The likelihood function is

$$L(p \mid u) = \prod_{i=1}^r p_i^{u_i}.$$
The \textit{maximum likelihood estimate}, or MLE, for \( u \) is the distribution in \( M \) that maximizes the likelihood function; that is, it is the distribution
\[
\hat{p} = \text{argmax}_{p \in M} L(p \mid u).
\]

Note that for a fixed \( p \in M \), \( L(p \mid u) \) is exactly the probability of observing \( u \) from the distribution \( p \). Hence, the MLE for \( u \) is the distribution \( \hat{p} \in M \) that maximizes the probability of observing \( u \). The map \( u \mapsto \hat{p} \) is a function of the data known as the \textit{maximum likelihood estimator}. We are particularly interested in the case when the coordinate functions of the maximum likelihood estimator are rational functions of the data. In this case, we say that \( M \) has \textit{rational MLE}.

The \textit{log-likelihood function} \( \ell(p \mid u) \) is the natural logarithm of \( L(p \mid u) \). Note that since the natural log is a concave function, \( \ell(p \mid u) \) and \( L(p \mid u) \) have the same maximizers. We define the \textit{maximum likelihood degree} of \( M \) to be the number of critical points of \( \ell(p \mid u) \) for generic \( u \). Huh and Sturmfels \cite{13} show that the maximum likelihood degree is well-defined. In particular, \( M \) has maximum likelihood degree 1 if and only if it has rational maximum likelihood estimator \cite{12}. The following result of Huh gives a characterization of the form of this maximum likelihood estimator, when it exists.

\textbf{Theorem 2.4 (Huh, \cite{12}).} A discrete statistical model \( M \) has maximum likelihood degree 1 if and only if there exists \( h = (h_1, \ldots, h_r) \in (\mathbb{C}^*)^r \), a positive integer \( d \), and a matrix \( B \in \mathbb{Z}^{d \times r} \) with entries \( b_{ij} \) whose column sums are zero such that the map
\[
\Psi : \mathbb{P}^{r-1} \longrightarrow (\mathbb{C}^*)^r
\]
with coordinate function
\[
\Psi_k(u_1, \ldots, u_r) = h_k \prod_{i=1}^d \left( \sum_{j=1}^r b_{ij} u_j \right)^{h_k}
\]
maps dominantly onto \( \overline{M} \). In this case, the function \( \Psi \) is the maximum likelihood estimator for \( M \).

In this context, the pair \((B, h)\) is called the \textit{Horn pair} that defines \( \Psi \), and \( \Psi \) is called the \textit{Horn map}. For more details about the Horn map and its connection to the theory of \( A \)-discriminants, we refer the reader to \cite{8} and \cite{12}.

\textbf{Example 2.5.} Consider the quasi-independence model associated to
\[
S = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2)\}.
\]
This is the log-linear model whose defining matrix is
\[
A = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}.
\]
We index the columns of \( A \) by the ordered pairs in \( S \) in the given order. Note that we have \( M_S = M_{A(S)} \). Let \( u \in \mathbb{N}^S \) be a vector of counts of iid data for the model \( M_S \).
According to Theorem 1.3, $\mathcal{M}_S$ has rational MLE. Theorem 5.4 shows that the associated Horn pair is

$$
B = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \\
-1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix}
$$

with $h = (-1, -1, 1, -1, -1, 1, 1, 1)$. The columns of $B$ and $h$ are also indexed by the elements of $S$. We can use this Horn pair to write the MLE as a rational function of the data. Denote by $u_{++}$ the sum of all entries of $u$, and abbreviate each ordered pair $(i, j) \in S$ by $ij$. Then for example, the $(1, 3)$ coordinate of the MLE is

$$
\hat{p}_{13} = \frac{h_{13}(u_{11} + u_{12} + u_{13})^1(u_{13} + u_{23})^1(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})^{-1}u_{++}^{-1}}{u_{++}(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})}.
$$

Similarly, the $(2, 3)$ coordinate is

$$
\hat{p}_{23} = \frac{(u_{21} + u_{22} + u_{23})(u_{13} + u_{23})}{u_{++}(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})}.
$$

The following theorem, known as Birch’s Theorem, says that the maximum likelihood estimate for $u$ in a log-linear model $\mathcal{M}_A$, if it exists, is the unique distribution $\hat{p}$ in $\mathcal{M}_A$ with the same sufficient statistics as the normalized data. A proof of this result can be found in [18, Chapter 7].

**Theorem 2.6** (Birch’s Theorem). Let $A \in \mathbb{Z}^{n \times r}$ such that $1 \in \text{rowspan}(A)$. Let $u \in \mathbb{R}_{\geq 0}^r$ and let $u_+ = u_1 + \cdots + u_r$. Then the maximum likelihood estimate in the log-linear model $\mathcal{M}_A$ given data $u$ is the unique solution, if it exists, to the equations $Au = u_+Ap$ subject to $p \in \mathcal{M}_A$.

**Example** (Example 2.5 continued). Consider the last row $a_6$ of the matrix $A$. One sufficient statistic of $\mathcal{M}_A$ is $a_6 \cdot u = u_{13} + u_{23}$. We must check that $a_6 \cdot u = u_+a_6 \cdot \hat{p}$. Indeed, we compute that

$$
a_6 \cdot \hat{p} = \frac{(u_{11} + u_{12} + u_{13})(u_{13} + u_{23})}{u_+(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})} + \frac{(u_{21} + u_{22} + u_{23})(u_{13} + u_{23})}{u_+(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})}
$$

$$
= (u_{13} + u_{23}) \frac{(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})}{u_+(u_{11} + u_{12} + u_{13} + u_{21} + u_{22} + u_{23})}
$$

$$
= \frac{u_{13} + u_{23}}{u_+},
$$

as needed.
3. Facial Submodels of Log-Linear Models

In order to prove that a quasi-independence model with rational MLE must have a doubly chordal bipartite associated graph $G_S$, we first prove a result that applies to general log-linear models with rational MLE. Let $A \in \mathbb{Z}^{n \times r}$ be the matrix defining the monomial map for the log-linear model $\mathcal{M}_A$. Let $I_A$ denote the vanishing ideal of the Zariski closure of $\mathcal{M}_A$. We assume throughout that $1 \in \text{rowspan}(A)$. Let $P_A = \text{conv}(A)$, where $\text{conv}(A)$ denotes the convex hull of the columns $a_1, \ldots, a_r$ of $A$.

We assume throughout that $P_A$ has $n$ facets, $F_1, \ldots, F_n$, and that the $ij$ entry of $A$, denoted $a_{ij}$ is equal to the lattice distance between the $j$th column of $A$ and facet $F_i$. This is not a restriction, since one can always reparametrize a log-linear model in this way \cite[Theorem 27]{16}. Indeed, given a polytope $Q$, a matrix $A$ that satisfies the above condition is a slack matrix of $Q$, and the convex hull of the columns of $A$ is affinely isomorphic to $Q$ \cite[11]{11}.

Let $\overline{A}$ be a matrix whose columns are a subset of $A$. Without loss of generality, assume that the columns of $\overline{A}$ are $a_1, \ldots, a_s$.

**Definition 3.1.** The submatrix $\overline{A}$ is called a facial submatrix of $A$ if $P_{\overline{A}}$ is a face of $P_A$. The corresponding statistical model $\mathcal{M}_{\overline{A}}$ is called a facial submodel of $\mathcal{M}_A$.

Let $e_i$ denote the $i$th standard basis vector in $\mathbb{R}^n$. Then $e_i \cdot a_j = 0$ if $a_j$ lies on $F_i$ and $e_i \cdot a_j \geq 1$ otherwise. So under our assumptions on $A$, this definition of a facial submatrix of $A$ aligns with the one given in \cite{9} and \cite{16}. We prove the following result concerning the maximum likelihood estimator for $\mathcal{M}_{\overline{A}}$ when $\overline{A}$ is a facial submatrix of $A$. This result was used implicitly in the proof of Theorem 4.4 of \cite{9}.

**Theorem 3.2.** Let $A \in \mathbb{Z}^{n \times r}$ and let $\overline{A} \in \mathbb{Z}^{n \times s}$ consist of the first $s$ columns of $A$. Suppose that $\overline{A}$ is a facial submatrix of $A$. Let $\mathcal{M}_A$ have rational maximum likelihood estimator $\Psi$ given by the Horn pair $(B, h)$. Let $\overline{B}$ denote the submatrix consisting of the first $s$ columns of $B$ and let $\overline{h} = (h_1, \ldots, h_s)$. Then $\mathcal{M}_{\overline{A}}$ has rational maximum likelihood estimator $\overline{\Psi}$ given by the Horn pair $(\overline{B}, \overline{h})$.

In order to prove Theorem 3.2 we check the conditions of Birch’s theorem. We do this using the following Lemmas.

**Lemma 3.3.** Let $\overline{\Psi}$ be as in Theorem 3.2. Then for generic $\overline{u} \in \mathbb{R}^s_{\geq 0}$, $\overline{\Psi}(\overline{u})$ is defined. In this case, $\overline{\Psi}(\overline{u})$ is in the Zariski closure of $\mathcal{M}_{\overline{A}}$.

**Proof.** Let $u \in \mathbb{R}^n_{\geq 0}$ be given by $u_i = \pi_i$ if $i \leq s$ and $u_i = 0$ if $i > s$. We claim that when $\Psi(u)$ is defined, $\overline{\Psi}_k(u) = \Psi_k(u)$ for $k \leq s$. Indeed, each factor of $\Psi_k(u)$ is of the form

$$\left( \sum_{j=1}^{r} b_{ij} u_j \right)^{b_{ik}}$$

for each $i = 1, \ldots, d$. If the $i$th factor of $\Psi_k$ is not identically equal to one, then $b_{ik} \neq 0$. So the $i$th factor has the nonzero summand $b_{ik} u_k$ and is generically nonzero when evaluated

\footnote{Note that the term “facial submodel” is a slight abuse of terminology because $\mathcal{M}_{\overline{A}}$ is not a submodel of $\mathcal{M}_A$. This is because the log-linear model $\mathcal{M}_A$ does not include distributions on the boundary of the probability simplex. Technically, $\mathcal{M}_{\overline{A}}$ is a submodel of the closure of $\mathcal{M}_A$.}
at a point $u$ of the given form. In particular, this implies that $\Psi_k(u)$ is defined for a generic $u$ of the given form since having $u_j = 0$ for $j > s$ does not make any factor of $\Psi_k$ identically equal to zero. Setting each $b_{ij} = 0$ when $j > s$ gives that $\overline{\Psi}_k(\pi) = \Psi_k(u)$ when $k \leq s$.

The elements of $I_A$ are those elements of $I_A$ that belong to the polynomial ring $k[p_1, \ldots, p_s]$. Let $f \in I_A$. Since $f \in I_A$ as well, $f(\overline{\Psi}((\pi))) = f(\Psi(u)) = 0$, as needed. □

Next we check that the sufficient statistics $\overline{\Psi}/\overline{\pi}_+$ are equal to those of $\overline{\Psi}((\pi))$.

**Lemma 3.4.** Let $\pi$ be a row of $A$. Then

$$\frac{\pi \cdot \pi}{\overline{\pi}_+} = \pi \cdot \overline{\Psi}((\pi)).$$

**Proof.** Let $c$ be the row of $A$ corresponding to $\pi$. Define a sequence $u^{(i)} \in \mathbb{R}_{\geq 0}$ by

$$u^{(i)}_j = \begin{cases} \pi_j & \text{if } j \leq s \\ c_j & \text{if } j > s, \end{cases}$$

where $\lim_{i \to \infty} \epsilon^{(i)}_j = 0$ for each $j$. We choose each $\epsilon^{(i)}_j > 0$ generically so that $\Psi(u^{(i)})$ is defined for all $i$.

Since $\pi$ is generic, we have that $\lim_{i \to \infty} u^{(i)}_+ = \pi_+ \neq 0$. Similarly, we have that $\lim_{i \to \infty} c \cdot u^{(i)} = \pi \cdot \pi$. So

$$\lim_{i \to \infty} \frac{c \cdot u^{(i)}}{u^{(i)}_+} = \frac{\pi \cdot \pi}{\overline{\pi}_+}$$

Since $\Psi(u^{(i)})$ is the maximum likelihood estimate in $\mathcal{M}_A$ for each $u^{(i)}$, by Birch’s theorem we have that

$$\frac{c \cdot u^{(i)}}{u^{(i)}_+} = c \cdot \Psi(u^{(i)})$$

$$= \sum_{i=1}^s c_i \Psi_i(u^{(i)}) + \sum_{j=s+1}^r c_j \Psi_j(u^{(i)}).$$

By the arguments in the proof of Lemma 3.3 when $k \leq s$, no factor of $\Psi_k(u^{(i)})$ involves only summands $u^{(i)}_j$ for $j > s$. So $\lim_{i \to \infty} \Psi_k(u^{(i)}) = \overline{\Psi}(\pi)$.

Finally, we claim that for $k > s$, $\lim_{i \to \infty} \Psi_k(u^{(i)}) = 0$. Without loss of generality, we may assume that $P_k$ is a facet of $P_A$. Indeed, if it were not, we could simply iterate these arguments over a saturated chain of faces between $P_k$ and $P_A$ in the face lattice of $P_A$.

Let $a = (a_1, \ldots, a_n)$ be the row of $A$ corresponding to the facet $P_k$ of $P_A$. Then $a_j = 0$ if $j \leq s$ and $a_j \geq 1$ if $j > s$. Since $\Psi(u^{(i)})$ is the maximum likelihood estimate in $\mathcal{M}_A$ for $u^{(i)}$, by Birch’s theorem we have that

$$a \cdot \Psi(u^{(i)}) = \frac{1}{u^{(i)}_+} (a_{s+1} u^{(i)}_{s+1} + \cdots + a_r u^{(i)}_r)$$

$$= \frac{1}{u^{(i)}_+} (a_{s+1} e^{(i)}_{s+1} + \cdots + a_r e^{(i)}_r).$$
Since \( \overline{u}_+ \neq 0 \), we also have that
\[
\lim_{i \to \infty} \mathbf{a} \cdot \Psi(u^{(i)}) = \lim_{i \to \infty} (a_{s+1} \Psi_{s+1}(u^{(i)}) + \cdots + a_r \Psi_r(u^{(i)})) \\
= \lim_{i \to \infty} \frac{1}{u^{(i)}_+} (a_{s+1} \epsilon^{(i)}_{s+1} + \cdots + a_r \epsilon^{(i)}_r) \\
= \frac{1}{u_+} \lim_{i \to \infty} (a_{s+1} \epsilon^{(i)}_{s+1} + \cdots + a_r \epsilon^{(i)}_r) \\
= 0.
\]

Furthermore, for all \( i \) and \( k \), \( \Psi_k(u^{(i)}) > 0 \). So \( \lim_{i \to \infty} \Psi_k(u^{(i)}) \geq 0 \). Since each \( a_i > 0 \) for \( i > s \), this implies that \( \lim_{i \to \infty} \Psi_k(u^{(i)}) = 0 \) for all \( k > s \).

So we have that
\[
\overline{c} \cdot \overline{u} = \lim_{i \to \infty} \frac{\overline{c} \cdot u^{(i)}}{u^{(i)}_+} \\
= \lim_{i \to \infty} \overline{c} \cdot \Psi(u^{(i)}) \\
= \overline{c} \cdot \overline{\Psi}(\overline{u}) + \sum_{j=s+1}^{r} c_j \left( \lim_{i \to \infty} \Psi_j(u^{(i)}) \right) \\
= \overline{c} \cdot \overline{\Psi}(\overline{u})
\]
as needed. \( \square \)

**Proof of Theorem 3.2.** First, note that \( \overline{\Psi} \) is still a rational function of degree zero since deleting columns of \( B \) does not affect the remaining column sums. So \((\overline{B}, \overline{h})\) is a Horn pair.

By Lemma 3.3, we have that \( \overline{\Psi}(\overline{u}) \in \overline{\mathcal{M}_A} \). Since \( 1 \in \text{rowspan}(\overline{A}) \), it follows from Lemma 3.4 that \( \sum_{k=1}^{s} \Psi_k(\overline{u}) = 1 \). Defining a sequence \( \{u^{(i)}\}_{i=1}^{\infty} \) as in the proof of Lemma 3.4, we have that \( \overline{\Psi}(\overline{u}) = \lim_{i \to \infty} \Psi_k(u^{(i)}) \). So \( \overline{\Psi}(\overline{u}) \geq 0 \) since each \( \Psi_k(u^{(i)}) > 0 \). Furthermore, for generic choices of \( \overline{u} \), we cannot have \( \overline{\Psi}(\overline{u}) = 0 \). Indeed, for \( k \leq s \), the \( i \)th factor of \( \overline{\Psi}_k(\overline{u}) \) has nonzero summand \( b_{ik}u_k \). So none of these factors is zero for generic choices of \( u \) of the given form. Therefore \( \overline{\Psi}(\overline{u}) \in \mathcal{M}_A = \overline{\mathcal{M}_A} \cap \Delta_{s-1} \).

By Lemma 3.4
\[
\overline{A} \cdot \overline{u} = \overline{\Psi}(\overline{u}).
\]
So by Birch’s theorem, \( \overline{\Psi} \) is the maximum likelihood estimator for \( \mathcal{M}_A \). \( \square \)

Note that \( \overline{\Psi} \) is a dominant map. Indeed, for generic \( p \in \mathcal{M}_A \), \( \overline{\Psi}(p) \) is defined. Since \( p \) is a probability distribution, \( p_+ = 1 \). By Birch’s Theorem, \( p \) is the MLE for data vector \( p \). So \( \overline{\Psi}(p) = p \).

We close this section by noting that we believe that a natural generalization of Theorem 3.2 is also true.

**Conjecture 3.5.** Let \( A \in \mathbb{Z}^{n \times r} \) and \( \overline{A} \in \mathbb{Z}^{n \times s} \) a facial submatrix of \( A \). Then the maximum likelihood degree of \( \mathcal{M}_A \) is greater than or equal to the maximum likelihood degree of \( \mathcal{M}_\overline{A} \).
4. QUASI-INDEPENDENCE MODELS WITH NON-RATIONAL MLE

In this section, we show that when $S$ is not doubly chordal bipartite, the ML-degree of $\mathcal{M}_S$ is strictly greater than one. We can apply Theorem 3.2 to quasi-independence models whose associated bipartite graphs are not doubly chordal bipartite using cycles and the following “double square” structure.

**Example 4.1.** The minimal example of a chordal bipartite graph that is not doubly chordal bipartite is the double-square graph. The matrix of the double-square graph has the form

$$\begin{bmatrix}
  * & * & 0 \\
  * & * & * \\
  0 & * & *
\end{bmatrix},$$

or any permutation of the rows and columns of this matrix. The resulting graph, pictured in Figure 1, is two squares joined along an edge. This is a 6-cycle with exactly one chord and as such, is not doubly chordal bipartite.

**Remark 4.2.** A bipartite graph is doubly chordal bipartite if and only if it is chordal bipartite and does not have the double-square graph as an induced subgraph.

We now compute the maximum likelihood degree of models associated to the double square and to cycles of length greater than or equal to 6.

**Proposition 4.3.** The maximum likelihood degree of the quasi-independence model whose associated graph is the double square is 2.

**Proof.** Without loss of generality, let

$$S = \{11, 12, 21, 22, 23, 32, 33\},$$

so that $G_S$ is a double-square graph. Then the vanishing ideal of $\mathcal{M}_S$ is the ideal $I(\mathcal{M}_S) \subset \mathbb{C}[p_{ij} \mid ij \in S]$ given by

$$I(\mathcal{M}_S) = \langle p_{11}p_{22} - p_{12}p_{21}, p_{22}p_{33} - p_{23}p_{33} \rangle.$$

Define the hyperplane arrangement

$$\mathcal{H} := \{ p \in \mathbb{C}^S \mid p_{++} \prod_{ij \in S} p_{ij} = 0 \},$$

where $p_{++}$ denotes the sum of all the coordinates of $p$. Then Proposition 7 of [2] implies that the ML-degree of $\mathcal{M}_S$ is the number of solutions to the system

$$I(\mathcal{M}_S) + \langle A(S)u + u_+A(S)p \rangle$$
that lie outside of $\mathcal{H}$ for generic $u$. Since $A(S)$ encodes the row and column marginals of $u$, this is equivalent to solving for $\alpha$ and $\beta$ in the system

\[
(u_{11} + \alpha)(u_{22} + \alpha + \beta) - (u_{12} - \alpha)(u_{21} - \alpha) = 0
\]

\[
(u_{22} + \alpha + \beta)(u_{33} + \beta) - (u_{23} - \beta)(u_{32} - \beta) = 0.
\]

Expanding gives two equations of the form

\[
\alpha \beta + c_1 \alpha + c_2 \beta + c_3 = 0
\]

\[
\alpha \beta + d_1 \alpha + d_2 \beta + d_3 = 0,
\]

where each $c_i, d_i$ are polynomials in the entries of $u$.

Solving for $\alpha = -(c_2 \beta + c_3)/(\beta + c_1)$ in the first equation of (1) and substituting into the second gives a degree 2 function of $\beta$, which can have at most two solutions. Indeed, for generic choices of $u$, this equation has exactly two solutions, neither of which lie on $\mathcal{H}$. For example, take $u_{11} = u_{12} = u_{21} = u_{22} = 1$ and $u_{23} = u_{32} = u_{33} = 2$. By performing this substitution in (1) with these values for $u$, we obtain the degree 2 equation

\[
\frac{-\beta^2}{\beta + 4} + \frac{7\beta}{\beta + 4} + 2\beta - 2 = 0.
\]

After clearing denominators, we obtain that $\beta^2 + 13\beta - 8 = 0$. This polynomial has two distinct roots neither of which lie on $\mathcal{H}$, and (2) is defined at both of these roots. These are generic conditions on the data; so since there exists a $u$ for which (1) has exactly two solutions, the ML-degree of $\mathcal{M}_S$ is 2.

\[\square\]

**Proposition 4.4.** Let $S_k \subset [k] \times [k]$ be a collection of indices such that $G_{S_k}$ is a cycle of length $2k$. Then the ML-degree of $\mathcal{M}_{S_k}$ is $k$ if $k$ is odd and $(k - 1)$ if $k$ is even.

**Proof.** Without loss of generality, we may assume that $S_k = \{(i, i) \mid i \in [k]\} \cup \{(i, i + 1) \mid i \in [k - 1]\} \cup \{(k, 1)\}$. Since $G_{S_k}$ consists of a single cycle, the ideal $I(\mathcal{M}_{S_k})$ is principle. Indeed, it is given by

\[
I(\mathcal{M}_{S_k}) = \langle \prod_{i=1}^{k} p_{i,i} - \prod_{i=1}^{k} p_{i,i+1} \rangle,
\]

where we set $p_{k,k+1} = p_{k,1}$. Let $\mathcal{H}$ be the hyperplane arrangement,

\[
\mathcal{H} = \{p \mid p_{++} \prod_{ij \in S} p_{ij} = 0\}.
\]

By Proposition 7 of [2], ML-degree of $\mathcal{M}_{S_k}$ is the number of solutions to

\[
I(\mathcal{M}_{S_k}) + \langle A(S_k)u - u_+ A(S_k)p \rangle.
\]

that lie outside of $\mathcal{H}$.

The sufficient statistics of $u$ are of the form $u_{i,i} + u_{i,i+1}$ and $u_{i-1,i} + u_{i,i}$ where we set $u_{0,1} = u_{k,1}$. So computing solutions to Equation (4) is equivalent to solving for $\alpha \in \mathbb{C}$ in the equation

\[
\prod_{i=1}^{k} (u_{i,i} + \alpha) - \prod_{i=1}^{k} (u_{i,i+1} - \alpha) = 0.
\]
The degree of this polynomial is \( k \) when \( k \) is odd and \( k - 1 \) when \( k \) is even.

Furthermore, we claim that for generic \( u \), none of these solutions lie in \( \mathcal{H} \). Indeed, without loss of generality, suppose that \( \bar{p} \) is a solution to \((\text{4})\) with \( \bar{p}_{1,1} = 0 \). Then we have that \( \alpha = -u_{1,1} \). So the first term of \((\text{4})\) is 0. But then there exists an \( i \) such that

\[
u_{i,i+1} - \alpha = u_{i,i+1} + u_{1,1} = 0,
\]

which is a non-generic condition on \( u \). Similarly, since \( u \) is generic, we may assume that \( u_+ \neq 0 \). But if \( \bar{p}_+ = 0 \), then since each \( \bar{p}_{i,i} = (u_{i,i} + \alpha)/u_+ \) and \( \bar{p}_{i,i+1} = (u_{i,i+1} - \alpha)/u_+ \), this implies that \( u_+ = 0 \), which is a contradiction. So for generic values of \( u \), the roots of \((\text{4})\) give rise to exactly \( k \), resp. \( k - 1 \), solutions to \((\text{4})\) that lie outside of \( \mathcal{H} \). So the ML-degree of \( \mathcal{M}_S \) is \( k \) if \( k \) is odd and \( k - 1 \) if \( k \) is even.

**Theorem 4.5.** Let \( S \) be such that \( G_S \) is not doubly chordal bipartite. Then \( \mathcal{M}_S \) does not have rational MLE.

**Proof.** Suppose that \( G_S \) is not doubly chordal bipartite. Then it has an induced subgraph \( H \) that is either a double square or a cycle of length greater than or equal to 6. Without loss of generality, let the edge set \( E(H) \) be a subset of \([k] \times [k] \). Let \( A = A(S) \) and let \( \bar{A} \) be the submatrix of \( A \) consisting of the columns indexed by elements of \( E(H) \).

Let the coordinates of \( P_A \) and \( P_{\bar{A}} \) be indexed by \((x_1, \ldots, x_m, y_1, \ldots, y_n) \). We claim that \( \bar{A} \) is a facial submatrix of \( A \). Indeed, \( \bar{A} \) consists of exactly the vertices of \( P_A \) that satisfy \( x_i = 0 \) for \( k < i \leq m \) and \( y_j = 0 \) for \( k < j \leq n \). Since \( P_A \) is a 0/1 polytope, each of these inequalities is valid. So this constitutes a face of \( P_A \).

Therefore, by Propositions 4.3 and 4.4, \( A \) has a facial submatrix \( \bar{A} \) such that \( \mathcal{M}_\bar{A} \) has ML-degree strictly greater than 1. So by Theorem 3.2, the ML-degree of \( \mathcal{M}_A = \mathcal{M}_S \) is also strictly greater than 1, as needed.

## 5. The Clique Formula for the MLE

In this section we state the main result of the paper, which gives the specific form of the rational maximum likelihood estimates for quasi-independence models when they exist. These are described in terms of the complete bipartite subgraphs of the associated graph \( G_S \). A complete bipartite subgraph of \( G_S \) corresponds to an entirely nonzero submatrix of \( S \). This motivates our use of the word “clique” in the following definition.

**Definition 5.1.** A set of indices \( C = \{i_1, \ldots, i_r\} \times \{j_1, \ldots, j_s\} \) is a **clique** in \( S \) if \((i_\alpha, j_\beta) \in S \) for all \( 1 \leq \alpha \leq r \) and \( 1 \leq \beta \leq s \). A clique \( C \) is maximal if it is not contained in any other clique in \( S \).

We now describe some important sets of cliques in \( S \).

**Notation 5.2.** For every pair of indices \((i, j) \in S \), we let \( \text{Max}(ij) \) be the set of all maximal cliques in \( S \) that contain \((i, j) \). We let \( \text{Int}(ij) \) be the set of all containment-maximal pairwise intersections of elements of \( \text{Max}(ij) \). Similarly, we let \( \text{Max}(S) \) denote the set of all maximal cliques in \( S \) and \( \text{Int}(S) \) denote the set of all maximal intersections of maximal cliques in \( S \).

**Example 5.3.** Consider the set of indices

\[
S = \{11, 12, 21, 22, 23, 28, 31, 32, 33, 34, 41, 45, 45, 51, 56, 57, 65, 76, 86, 87, 89\},
\]
Figure 2. The bipartite graph associated the matrix $S$ in Example 5.3

where we replace $(i, j)$ with $ij$ for the sake of brevity. The corresponding matrix with structural zeros is

$$
\begin{bmatrix}
* & * & 0 & 0 & 0 & 0 & 0 & 0 \\
* & * & * & 0 & 0 & 0 & * & 0 \\
* & * & * & 0 & 0 & 0 & 0 & 0 \\
* & 0 & 0 & 0 & * & 0 & 0 & 0 \\
* & 0 & 0 & 0 & 0 & * & 0 & 0 \\
0 & 0 & 0 & 0 & * & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & * & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & * & * & 0 & * \\
\end{bmatrix}.
$$

We will use this as a running example. The bipartite graph $G_S$ associated to $S$ is pictured in Figure 2. In this figure, we use white circles to denote vertices corresponding to rows in $S$ and black squares to denote vertices corresponding to columns in $S$. Note that $G_S$ is doubly chordal bipartite since its only cycle of length 6 has two chords.

In this case, the set of maximal cliques in $S$ is

$$\text{Max}(S) = \{\{11, 21, 31, 41, 51\}, \{11, 12, 21, 22, 31, 32\}, \{21, 22, 23, 31, 32, 33\}, \{21, 22, 23, 28\}, \{31, 32, 33, 34\}, \{41, 45\}, \{51, 56, 57\}, \{45, 65\}, \{56, 76, 86\}, \{56, 57, 86, 87\}, \{86, 87, 89\}\}.$$

The set of maximal intersections of maximal cliques in $S$ is

$$\text{Int}(S) = \{\{11, 21, 31\}, \{21, 22, 31, 32\}, \{21, 22, 23\}, \{31, 32, 33\}, \{41\}, \{51\}, \{45\}, \{56, 57\}, \{56, 86\}, \{86, 87\}\}.$$

Note, for example, that $\{31, 32\}$ is the intersection of the two maximal cliques $\{11, 12, 21, 22, 31, 32\}$ and $\{31, 32, 33, 34\}$. However it is not in $\text{Int}(S)$ because it is properly contained in the intersection of maximal cliques,

$$\{11, 12, 21, 22, 31, 32\} \cap \{21, 22, 23, 31, 32, 33\} = \{21, 22, 31, 32\}.$$
Let \( u = (u_{ij} \mid (i, j) \in S) \) be a matrix of counts. For any \( C \subset S \), we let \( C^+ \) denote the sum of all the entries of \( U \) whose indices are in \( C \). That is,

\[
C^+ = \sum_{(i, j) \in C} u_{ij}.
\]

Similarly, we denote the row and column marginals \( u_i^+ = \sum_{j: (i, j) \in S} u_{ij} \) and \( u^+_j = \sum_{i: (i, j) \in S} u_{ij} \).

**Theorem 5.4.** Let \( S \subset [m] \times [n] \) be a set of indices with associated bipartite graph \( G_S \) and quasi-independence model \( \mathcal{M}_S \). Then \( \mathcal{M}_S \) has rational maximum likelihood estimate if and only if \( G_S \) is doubly chordal bipartite. In particular, if \( u = (u_{ij} \mid (i, j) \in S) \) is a matrix of counts, the maximum likelihood estimate for \( u \) has \( ij \)th entry

\[
\hat{p}_{ij} = \frac{u_{i+}u_{+j}}{u_{++}} \prod_{C \in \text{Int}(ij)} C^+ \prod_{D \in \text{Max}(ij)} D^+.
\]

Over the course of the next two sections, we prove various lemmas that ultimately allow us to prove Theorem 5.4.

**Example 5.5.** Consider the set of indices \( S \) from Example 5.3. Let \( u \) be a matrix of counts. Consider the maximum likelihood estimate for the \((2, 1)\) entry, \( \hat{p}_{21} \). The maximal cliques that contain \( 21 \) are \{11, 21, 31, 41, 51\}, \{11, 12, 21, 22, 31, 32\} and \{21, 22, 23, 31, 32, 33\}. The maximal intersections of maximal cliques that contain \( 21 \) are \{11, 21, 31\} and \{21, 22, 31, 32\}. Since \( S \) is doubly chordal bipartite, we apply Theorem 5.4 to obtain that the numerator of \( \hat{p}_{21} \) is

\[
(u_{21} + u_{22} + u_{23})(u_{11} + u_{21} + u_{31} + u_{41} + u_{51})(u_{11} + u_{21} + u_{31})(u_{21} + u_{22} + u_{31} + u_{32}).
\]

The denominator of \( \hat{p}_{21} \) is

\[
u_{++}(u_{11} + u_{21} + u_{31} + u_{41} + u_{51})(u_{11} + u_{12} + u_{21} + u_{22} + u_{31} + u_{32})(u_{21} + u_{22} + u_{23} + u_{31} + u_{32} + u_{33}).
\]

We note that when a maximal clique is a single row or column, as is the case with \{11, 21, 31, 41, 51\}, we have cancellation between the numerator and denominator.

In order to prove Theorem 5.4, we show that \( \hat{p}_{ij} \) satisfies the conditions of Birch’s theorem. First, we investigate the intersections of a fixed column of the matrix with structural zeros with maximal cliques and their intersections. We prove useful lemmas about the form that these maximal cliques have that allow us to show that the conditions of Birch’s Theorem are satisfied. In particular, we use them to prove Corollary 7.1 which states that the column marginal of the formula in Theorem 5.4 given by the fixed column is equal to that of the normalized data.

6. **Intersections of Cliques with a Fixed Column**

In this section we prove some results that will set the stage for the proof of Theorem 5.4 that appears in Section 7. To prove that our formulas satisfy Birch’s theorem, we need to understand what happens to sums of these formulas over certain sets of indices.
Let $S \subset [m] \times [n]$ and let $j_0 \in [n]$. Without loss of generality, we assume that $(1, j_0), \ldots, (r, j_0) \in S$, and that the last $(i, j_0) \not\in S$ for all $i > r$. We let $N_{j_0} = \{(1, j_0), \ldots, (r, j_0)\}$. We consider $j_0$ to be the index of a column in the matrix representation of $S$, and $1, \ldots, r$ to be the indices of its nonzero rows. Now let $T_0 \mid \cdots \mid T_h$ be the coarsest partition of $[n]$ with the property that whenever $j, k \in T_\ell$,

$$\{i \in [r] \mid (i, j) \in S\} = \{i \in [r] \mid (i, k) \in S\}.$$ 

In the matrix representation of $S$, each $T_\ell$ corresponds to a set of columns whose first $r$ rows are identical. The fact that we take $T_0 \mid \cdots \mid T_h$ to be the coarsest such partition ensures that the supports of the columns in distinct parts of the partition are distinct.

Define the partition $B_0 \mid \cdots \mid B_h$ of $S \cap ([r] \times [n])$ by $B_\ell = \{(i, j) \mid j \in T_\ell\}$. Note that one of the $B_\ell$ may be empty, in which case we exclude it from the partition. We call these $B_\ell$ the blocks of $S$ corresponding to column $j_0$. We fix $j_0$ and $B_0, \ldots, B_h$ for the entirety of this section, and we assume without loss of generality that $j_0 \in T_0$.

Denote by rows$^{j_0}(B_\alpha)$ the set of all $i \in [r]$ such that $(i, j) \in B_\alpha$ for some column index $j$. Note that this is a subset of the first $1, \ldots, r$ rows of $S$, and that in the matrix representation of $S$, the columns whose indices are in $B_\alpha$ may not have the same zero patterns in rows $r + 1, \ldots m$. Similarly, for each $j \in [n]$, define rows$^{j_0}(j)$ to be the set of all $i \in [r]$ such that $(i, j) \in S$; that is, the elements of rows$^{j_0}(j)$ are the row indices of the nonzero entries of column $j$ in the first $r$ rows of $S$. Note that the dependence on $j_0$ in this notation stems from the fact that the column $j_0$ is used to obtained the partition $B_0 \mid \cdots \mid B_h$.

**Example 6.1.** Consider the running example $S$ from Example 5.3 and let $j_0 = 1$ be the first column of $S$. In this case, $r = 5$ since only the first 5 rows entries of column $j_0$ are nonzero. Then the blocks associated to $j_0$ consist of the following columns.

- $T_0 = \{j_0\} = \{1\}$
- $T_2 = \{3\}$
- $T_4 = \{5\}$
- $T_6 = \{8\}$
- $T_1 = \{2\}$
- $T_3 = \{4\}$
- $T_5 = \{6, 7\}$
- $T_7 = \{9\}$.

We note that although columns 6 and 7 are not the same over the whole matrix, their first five rows are the same. Since these are the nonzero rows of column $j_0$, columns 6 and 7 belong to the same block.

The $B_\ell$ associated to each of these sets of column indices are

- $B_0 = \{11, 21, 31, 41, 51\}$
- $B_2 = \{23, 33\}$
- $B_4 = \{45\}$
- $B_6 = \{28\}$
- $B_1 = \{12, 22, 32\}$
- $B_3 = \{34\}$
- $B_5 = \{56, 57\}$
- $B_7 = \emptyset$

For instance, rows$^{j_0}(B_1) = \{1, 2, 3\}$ and rows$^{j_0}(B_5) = \{5\}$.

The following proposition characterizes what configurations of the rows of the $B_\alpha$'s are allowable in order to avoid a cycle with exactly one chord. We call the condition outlined in Proposition 6.2 the double-squarefree, or DS-free condition.
Proposition 6.2 (DS-free condition). Let \( S \) be doubly chordal bipartite. Let \( \alpha, \beta \in [h] \). If \( \text{rows}^{j_0}(B_\alpha) \cap \text{rows}^{j_0}(B_\beta) \) is nonempty, then \( \text{rows}^{j_0}(B_\alpha) \subset \text{rows}^{j_0}(B_\beta) \) or \( \text{rows}^{j_0}(B_\beta) \subset \text{rows}^{j_0}(B_\alpha) \).

Proof. For the sake of contradiction, suppose without loss of generality that \( \text{rows}^{j_0}(B_1) \cap \text{rows}^{j_0}(B_2) \) is nonempty but neither is contained in the other. Then let \( i_0, i_1 \in \text{rows}^{j_0}(B_1) \) and \( i_1, i_2 \in \text{rows}^{j_0}(B_2) \) so that \( i_0 \not\in \text{rows}^{j_0}(B_2) \) and \( i_2 \not\in \text{rows}^{j_0}(B_1) \). We have \( i_0, i_1, i_2 \in \text{rows}^{j_0}(j_0) \) by definition. Let \( j_2 \in T_1 \) and \( j_3 \in T_2 \). Then the \( \{i_0, i_1, i_2\} \times \{j_0, j_1, j_2\} \) submatrix of \( S \) is the matrix of a double-square, which contradicts that \( S \) is doubly chordal bipartite. \( \square \)

Proposition 6.2 implies that the sets \( \text{rows}^{j_0}(B_\alpha) \) over all \( \alpha \) have a tree structure ordered by containment. In fact, we will see that this gives a tree structure on the maximal cliques in \( S \) that intersect \( N_{j_0} \). (Recall that \( N_{j_0} = \{(i, j_0) \in S\} = [r] \times \{j_0\} \).

Example 6.3. The matrix \( S \) from Example 5.3 is doubly chordal bipartite, and as such, satisfies the DS-free condition. If we append a tenth column, \( (0, *, *, *, 0, 0, 0, 0)^T \) to obtain a matrix \( S' \), this introduces a new block \( B_8 \) which just contains column 10. This matrix violates the DS-free condition since \( \text{rows}^{j_0}(B_1) = \{1, 2, 3\} \) and \( \text{rows}^{j_0}(B_8) = \{2, 3, 4\} \). Their intersection is nonempty, but neither is contained in another. Indeed, the \( \{1, 2, 4\} \times \{1, 2, 10\} \) submatrix of \( S' \) is the matrix of a double-square.

For each pair of indices \( ij \) such that \( (i, j) \in S \), let \( x_{ij} \) be the polynomial obtained by simultaneously clearing the denominators of all \( \hat{p}_{k\ell} \). That is,

\[
x_{ij} = u_{i+} u_{j+} \prod_{C \in \text{Int}(ij)} C^+ \prod_{D \in \text{Max}(S) \setminus \text{Max}(ij)} D^+.
\]

Our main goal in this section is to derive a formula for the sum,

\[
\sum_{i \in \text{rows}^{j_0}(B_\alpha)} x_{ij_0}.
\]

This is the content of Lemma 6.9. This formula allows us to verify that the \( j_0 \) column marginal of \( \hat{p} \) matches that of the normalized data. In order to simplify this sum, we must first understand how maximal cliques and their intersections intersect \( N_{j_0} \). For each \( B_\alpha \) with \( \alpha \in [h] \) and \( \text{rows}^{j_0}(B_\alpha) \neq \emptyset \), we let \( D_\alpha \) be the clique,

\[
D_\alpha = \{(i, j) \mid i \in \text{rows}^{j_0}(B_\alpha) \text{ and } \text{rows}^{j_0}(B_\alpha) \subset \text{rows}^{j_0}(j)\}.
\]

In other words, \( D_\alpha \) is the largest clique that contains \( B_\alpha \) and intersects \( N_{j_0} \). We call \( D_\alpha \) the clique induced by \( B_\alpha \).

Example 6.4. Consider our running example \( S \) with \( j_0 = 1 \) and blocks \( B_0, \ldots, B_7 \) as described in Example 6.1. Then \( N_{j_0} = N_1 = \{11, 21, 31, 41, 51\} \). The cliques induced by
$B_0, \ldots, B_7$ are

\begin{align*}
D_0 &= \{11, 21, 31, 41, 51\}, \\
D_1 &= \{11, 12, 21, 22, 31, 32\}, \\
D_2 &= \{21, 22, 23, 31, 32, 33\}, \\
D_3 &= \{31, 32, 33, 34\}, \\
D_4 &= \{41, 45\}, \\
D_5 &= \{51, 56, 57\}, \text{ and} \\
D_6 &= \{21, 22, 23, 28\}.
\end{align*}

There is no $D_7$ since the block $B_7$ is empty. Note that these are exactly the maximal cliques in $S$ that intersect $N_1$. The next proposition proves that this is the case for all DS-free matrices with structural zeros.

We note that when $D_\alpha$ is the clique induced by $B_\alpha$, all of the nonzero rows of $D_\alpha$ lie in $[r]$ by definition of an induced clique. We continue to use the notation $\text{rows}^{j_0}(D_\alpha)$ since the formation of the set $B_\alpha$ depends on the specified column $j_0$. For any clique $C$, let $\text{cols}(C) = \{j \mid (i, j) \in C \text{ for some } j\}$.

**Proposition 6.5.** For all $\alpha \in [h]$, $D_\alpha$ is maximal. Furthermore, any maximal clique that has nonempty intersection with $\alpha$ is induced by some $B_\alpha$.

**Proof.** We will show that $D_\alpha$ is maximal by showing that we cannot add any rows or columns to it. We cannot add any columns to $D_\alpha$ by definition. We cannot add any of rows $1, \ldots, r$ to $D_\alpha$ since all nonzero rows of $B_\alpha$ are already contained in $D_\alpha$. We cannot add any of rows $r + 1, \ldots, m$ to $D_\alpha$ since $j_0$ is a column of $D_\alpha$ whose entries in rows $r + 1, \ldots, m$ are zero. Note that if we can add one element $(i, j)$ to $D_\alpha$, then by definition of a clique, we must either be able to add all of $\{i\} \times \text{cols}(D_\alpha)$ or $\text{rows}^{j_0}(D_\alpha) \times \{j\}$ to the clique. Since we cannot add any rows or columns to $D_\alpha$, it is a maximal clique.

Now let $D$ be a maximal clique that intersects $\alpha$. For the sake of contradiction, suppose that $D \neq D_\alpha$ for each $\alpha \in [h]$.

Let $j_1$ be a column in $D$ such that $\text{rows}^{j_0}(j_1)$ is minimal among all columns of $D$. We must have that $(i, j_1) \in B_\alpha$ for some $i \in [r]$ and $\alpha \in [h]$. Since $D \neq D_\alpha$, it must be the case that column $j_1$ has a nonzero row $i_1 \in [r]$ that is not in $D$. Since $D$ is maximal, there must exist another column $j_2$ in $D$ that has a zero in row $i_1$. Therefore, we have that $\text{rows}^{j_0}(j_1) \nsubseteq \text{rows}^{j_0}(j_2)$. Furthermore, $\text{rows}^{j_0}(j_2) \nsubseteq \text{rows}^{j_0}(j_1)$ by the minimality of $j_1$. But since $D$ is nonempty, the intersection of $\text{rows}^{j_0}(j_1)$ and $\text{rows}^{j_0}(j_2)$ must be nonempty. This contradicts Proposition 6.2 as needed. \hfill \Box

Proposition 6.5 shows that the maximal cliques that intersect $\alpha$ are exactly the cliques that are induced by some $B_\alpha$. The DS-free condition gives a poset structure on the set of these maximal cliques $D_0, \ldots, D_h$ that intersect $\alpha$ nontrivially.

**Definition 6.6.** Let $P(j_0)$ denote the poset with ground set $\{D_0, \ldots, D_h\}$ and $D_\alpha \leq D_\beta$ if and only if $\text{rows}^{j_0}(D_\alpha) \subseteq \text{rows}^{j_0}(D_\beta)$.

In fact, the Hasse diagram of this poset is a tree since the DS-free condition implies that any $D_\alpha$ is covered by at most one maximal clique.
\[ \begin{align*}
\text{Figure 3. The poset } P(j_0) \text{ for Example 6.7.}
\end{align*} \]

**Example 6.7.** In our running example \( S \) with \( j_0 = 1 \) and blocks \( B_0, \ldots, B_6 \) and associated cliques \( D_0, \ldots, D_6 \), the Hasse diagram of the poset \( P(j_0) \) is pictured in Figure 3.

The next proposition shows that the cover relations in this poset, denoted \( D_\alpha \prec D_\beta \) correspond to maximal intersections of maximal cliques that intersect \( N_{j_0} \) nontrivially. Denote by \( \text{cols}(D_\alpha) \) the nonzero columns of the clique \( D_\alpha \). We note that if \( D_\alpha \prec D_\beta \), then \( \text{cols}(D_\beta) \subset \text{cols}(D_\alpha) \). In particular, this means that if \( C = D_\alpha \cap D_\beta \), then \( C = \text{rows}^{j_0}(D_\alpha) \times \text{cols}(D_\beta) \).

**Proposition 6.8.** Let \( C = D_\alpha \cap D_\beta \). Then \( C \) is maximal among all pairwise intersections of maximal cliques if and only if \( D_\alpha \prec D_\beta \) or \( D_\beta \prec D_\alpha \) in \( P(j_0) \).

**Proof.** Suppose without loss of generality that \( D_\alpha \prec D_\beta \) in \( P(j_0) \). For the sake of contradiction, suppose that \( D_\alpha \cap D_\beta \not\in \text{Int}(S) \). Then there exists another maximal clique that contains \( D_\alpha \cap D_\beta \). By Proposition 6.5 and the fact that \( D_\alpha \cap D_\beta \) intersects \( N_{j_0} \) nontrivially, we can write this maximal clique as \( D_\gamma \) for some \( \gamma \in [h] \).

Note that we have \( \text{rows}^{j_0}(C) = \text{rows}^{j_0}(D_\alpha) \) and \( \text{cols}(C) = \text{cols}(D_\beta) \). Therefore \( C = \text{rows}^{j_0}(D_\alpha) \times \text{cols}(D_\beta) \). So \( \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\gamma) \) and \( \text{cols}(D_\beta) \subset \text{cols}(D_\gamma) \). In particular, this second inclusion implies that \( \text{rows}^{j_0}(D_\gamma) \subset \text{rows}^{j_0}(D_\beta) \). Indeed, suppose that \( i \leq r \) is a row of \( D_\gamma \) that is not a row of \( D_\beta \). Then there exists a column \( j \) of \( D_\beta \) for which \( (i, j) \notin S \). But since \( j \) is also a column of \( D_\gamma \), this contradicts that \( D_\gamma \) is a clique. So we have the proper containment
\[
\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\gamma) \subset \text{rows}^{j_0}(D_\beta),
\]
which contradicts that \( D_\alpha \prec D_\beta \) in \( P(j_0) \). So \( D_\alpha \cap D_\beta \) must be maximal.

Now let \( C = D_\alpha \cap D_\beta \in \text{Int}(S) \). For the sake of contradiction, suppose that \( D_\alpha \) does not cover \( D_\beta \) or vice versa. Since \( C \) is nonempty, without loss of generality we must have \( \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\beta) \) by the DS-free condition. So there exists a \( D_\gamma \) such that \( D_\alpha \prec D_\gamma \prec D_\beta \) in \( P(j_0) \). Therefore we have that
\[
\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\gamma) \subset \text{rows}^{j_0}(D_\beta),
\]
which implies that \( D_\alpha \prec D_\beta \) in \( P(j_0) \). So \( D_\alpha \cap D_\beta \) must be maximal.

Let \( (i, j) \in C \). Then \( i \) is a row of \( D_\alpha \), so it is a row of \( D_\gamma \). Furthermore, since \( j \) is a column of \( D_\beta \), \( \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(j) \). So \( \text{rows}^{j_0}(D_\gamma) \subset \text{rows}^{j_0}(j) \) and \( j \) is a column of \( D_\gamma \). Therefore, \( C \subset D_\gamma \cap D_\beta \). This containment is proper since \( \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\gamma) \). So we have contradicted that \( C \) is maximal. \[\square\]
We can now state the key lemma regarding the sum of the $x_{ij}$s over $\{i : (i, j_0) \in D_\alpha\}$ for any $\alpha \in [h]$.

**Lemma 6.9.** Let $S$ be DS-free and let $D_\alpha$ be a maximal clique that intersects $N_{j_0}$. Then
\begin{equation}
\sum_{i \in \text{rows}^{j_0}(D_\alpha)} x_{ij_0} = u_{+, j_0} \left( \prod_{\substack{C \in \text{Int}(S) \\ C \cap N_{j_0} \neq \emptyset}} C^+ \right) \left( \prod_{\substack{D \in \text{Max}(S) \\ D \cap N_{j_0} \neq \emptyset}} D^+ \right) \left( \prod_{\substack{E \in \text{Max}(S) \\ N_{j_0} \cap D_\alpha \cap E = \emptyset}} E^+ \right)
\end{equation}

In order to prove this, we will sum the entries $x_{ij_0}$ over all $i \in \text{rows}^{j_0}(D_\beta)$ for each $\beta$. We will do this inductively from the bottom of $P(j_0)$. The key idea of this induction is as follows.

**Remark 6.10.** If $D_{\alpha_1}, \ldots, D_{\alpha_k}$ are covered by $D_\beta$ in $P(j_0)$, then the rows of $N_{j_0} \cap D_\beta$ are partitioned by each $N_{j_0} \cap D_{\alpha_k}$ along with the set of rows that are in $D_\beta$ and not in any $D_{\alpha_k}$. The fact that this is a partition follows from the DS-free condition. Therefore, summing the $x_{ij_0}$ that belong to each clique covered by $D_\beta$ and adding in the $x_{ij_0}$s for rows $i$ that are not in any clique covered by $D_\beta$ will give us the sum of $x_{ij_0}$ over all $i \in \text{rows}^{j_0}(D_\beta)$.

The next proposition focuses on the factors of the left-hand side of Equation (5) that correspond to elements of $\text{Int}(S)$. It will be used to show that when we perform the induction and move upwards by one cover relation from $D_\alpha$ to $D_\beta$ in the poset $P(j_0)$, all but one of these factors stays the same. The only one that no longer appears in the product corresponds to the maximal intersection $D_\alpha \cap D_\beta$.

**Proposition 6.11.** Let $D_\alpha \prec D_\beta$ in $P(j_0)$. Let $C \in \text{Int}(S)$ intersect $N_{j_0}$ nontrivially so that $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(C)$. Then either $C = D_\alpha \cap D_\beta$ or $\text{rows}^{j_0}(D_\beta) \subset \text{rows}^{j_0}(C)$.

**Proof.** Without less of generality, let $C = D_1 \cap D_2$. Proposition 6.8 tells us that $C$ must be of this form. By the same proposition, we may assume without loss of generality that $D_2 \prec D_1$, so $\text{rows}^{j_0}(C) = \text{rows}^{j_0}(D_2)$. Suppose that $\text{rows}^{j_0}(D_\beta) \not\subset \text{rows}^{j_0}(D_2)$. Since $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_2)$ and $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\beta)$, we must have that $\text{rows}^{j_0}(D_2) \cap \text{rows}^{j_0}(D_\beta)$ is nonempty. So by the DS-free condition, $\text{rows}^{j_0}(D_2) \not\subset \text{rows}^{j_0}(D_\beta)$. So we have the chain of inclusions,
\[ \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_2) \not\subset \text{rows}^{j_0}(D_\beta). \]

But since $D_\beta$ covers $D_\alpha$ in $P(j_0)$, and every element of $P(j_0)$ is covered by at most one element, this implies that $\alpha = 1$ and $\beta = 2$, so $C = D_\alpha \cap D_\beta$, as needed. \qed

The following proposition focuses on the factors of the left-hand side of Equation (5) that correspond to elements of $\text{Max}(S)$. It gives a correspondence between the factors of this product for $D_\alpha$ and all but one of the factors of this product for $D_\beta$ when we have the cover relation $D_\alpha \prec D_\beta$ in $P(j_0)$. 

Proposition 6.12. For any $D_\alpha, D_\beta \in P(j_0)$, define the following sets:

$$R_\alpha = \{ D \in \text{Max}(S) \mid D \cap N_{j_0} \neq \emptyset, \text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\alpha) \}$$

$$\cup \{ E \in \text{Max}(S) \mid N_{j_0} \cap D_\alpha \cap E = \emptyset \}$$

$$\overline{R}_\beta = \{ D \in \text{Max}(S) \mid D \cap N_{j_0} \neq \emptyset, \text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\beta) \}$$

$$\cup \{ E \in \text{Max}(S) \mid N_{j_0} \cap D_\beta \cap E = \emptyset \}.$$

If $D_\alpha \subset D_\beta$ in $P(j_0)$, then $R_\alpha = \overline{R}_\beta$.

Proof. First let $D \in R_\alpha$. If $\text{rows}^{j_0}(D) \subset \text{rows}^{j_0}(D_\alpha)$ and $D \cap N_{j_0} \neq \emptyset$, then since $\text{rows}^{j_0}(D_\alpha) \subsetneq \text{rows}^{j_0}(D_\beta)$, we have that $\text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\beta)$. So $D \in \overline{R}_\beta$.

Otherwise, we have $N_{j_0} \cap D_\alpha \cap D = \emptyset$. There are now two cases.

Case 1: If $N_{j_0} \cap D = \emptyset$, then $N_{j_0} \cap D_\beta \cap D = \emptyset$ as well. So $D \in \overline{R}_\beta$.

Case 2: Suppose that $N_{j_0} \cap D \neq \emptyset$ and $D_\alpha \cap D = \emptyset$. If $D_\beta \cap D$ is empty as well, then $D \in \overline{R}_\beta$.

Otherwise, suppose $D_\beta \cap D \neq \emptyset$. Then we must have that $\text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\beta)$ by the fact that $\text{rows}^{j_0}(D_\beta) \not\subset \text{rows}^{j_0}(D)$ and the DS-free condition. So $D \in \overline{R}_\beta$ in this case as well. Note that it is never the case that $N_{j_0} \cap D \neq \emptyset$ and $D_\alpha \cap D \neq \emptyset$ but $N_{j_0} \cap D \cap D_\alpha = \emptyset$ since $j_0$ is a column of $D_\alpha$. So we have shown that $R_\alpha \subset \overline{R}_\beta$.

Now let $D \in \overline{R}_\beta$. We have two cases again.

Case 1: First, consider the case in which $\text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\beta)$ and $D \cap N_{j_0} \neq \emptyset$. If $\text{rows}^{j_0}(D) \subset \text{rows}^{j_0}(D_\alpha)$, then $D \in R_\alpha$, as needed. Otherwise, by the DS-free condition, there are two cases.

Case 1a: If $\text{rows}^{j_0}(D_\alpha) \subsetneq \text{rows}^{j_0}(D)$, then we have the chain of containments,

$$\text{rows}^{j_0}(D_\alpha) \subsetneq \text{rows}^{j_0}(D) \subsetneq \text{rows}^{j_0}(D_\beta),$$

which contradicts that $D_\alpha \subset D_\beta$ in $P(j_0)$. So this case cannot actually occur.

Case 1b: If $\text{rows}^{j_0}(D_\alpha) \cap \text{rows}^{j_0}(D) = \emptyset$, then we have that $N_{j_0} \cap D_\alpha \cap D = \emptyset$. Therefore, $D \in R_\alpha$, as needed.

Case 2: The final case is when $N_{j_0} \cap D_\beta \cap D = \emptyset$. In this case, since $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\beta)$, we have that $N_{j_0} \cap D_\alpha \cap D = \emptyset$ as well. So $D \in R_\alpha$. So we have shown that $\overline{R}_\beta \subset R_\alpha$, as needed. \qed

Remark 6.13. Note that Proposition 6.12 implies that whenever $D_\alpha$ and $D_\beta$ are covered by the same element of $P(j_0)$, we have that $R_\alpha = R_\beta$. This shows that the left-hand side of Equation (5) for $D_\alpha$ and $D_\beta$ consist of the same terms that come from cliques in $\text{Max}(S)$.

Let $D_{\alpha_1}, \ldots, D_{\alpha_\ell} \subset D_\beta$. As we discussed in Remark 6.10 in order to sum the values of $x_{ij_0}$ over $N_{j_0} \cap D_\beta$, we must understand the sum over $x_{ij_0}$ for those rows $i$ such that $i \in \text{rows}^{j_0}(D_\beta)$ but $i \not\in \text{rows}^{j_0}(D_{\alpha_k})$ for all $k$. The following proposition concerns the sum of the $x_{ij_0}$ over these values of $i$. 
Proposition 6.14. Let $D_{\alpha_1}, \ldots, D_{\alpha_\ell} \subset D_{\beta}$. Let $r_1, \ldots, r_a$ be the rows of $D_{\beta}$ that are not in any $D_{\alpha_k}$ for $k = 1, \ldots, \ell$. Then

$$\sum_{i=1}^{a} x_{r_i,j_0} = \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^{a} u_{r_i} \right)$$

(6)

Proof. Without loss of generality, we will let $D_2, \ldots, D_\ell \subset D_1$, and let rows $1, \ldots, a$ be the rows of $D_1$ that are not rows of any $D_\alpha$ for $\alpha = 2, \ldots, \ell$. Let $i \in [a]$. Recall that

$$x_{i,j_0} = u_{i,j_0} \prod_{C \in \text{Int}(i,j_0)} C^+ \prod_{D \in \text{Max}(S) \setminus \text{Max}(i,j_0)} D^+.$$ 

We first consider the maximum cliques $D$ with $(i,j_0) \not\subset D$. If $N_{j_0} \cap D_1 \cap D = \emptyset$, then $D^+$ is a term of $x_{i,j_0}$ for all $i = 1, \ldots, a$. So $D^+$ is a factor of both the left-hand and right-hand sides of Equation (6).

Otherwise, we have $N_{j_0} \cap D_1 \cap D \neq \emptyset$. In particular, this means that $N_{j_0} \cap D \neq \emptyset$. Since $\text{rows}^0(D_1) \cap \text{rows}^0(D) \neq \emptyset$, and $(i,j_0) \not\subset D$, we must have $\text{rows}^0(D) \subset \text{rows}^0(D_1)$ by the DS-free condition. Furthermore, for all maximal cliques $D$ with $\text{rows}^0(D) \subset \text{rows}^0(D_1)$, we have $(i,j_0) \not\subset D$. Indeed, $D \cap N_{j_0} \neq \emptyset$, so by Proposition 6.5, we have $D = D_\gamma$ for some $\gamma$ with $D_\gamma \subset D_1$ in $P(j_0)$. So $\text{rows}^0(D_\gamma) \subset \text{rows}^0(D_\alpha)$ for some $\alpha \in \{2, \ldots, \ell\}$. Since $(i,j_0) \not\subset D_\alpha$, we have $(i,j_0) \not\subset D_\gamma$ as well.

Therefore, the factors $D^+$ corresponding to maximal cliques in each $x_{i,j_0}$ are the same for all $i \in [a]$, and are exactly those with $\text{rows}^0(D) \subset \text{rows}^0(D_1)$ or $N_{j_0} \cap D_1 \cap D = \emptyset$.

Now let $(i,j_0) \in C$ where $C \in \text{Int}(S)$ and $C \cap N_{j_0} \neq \emptyset$. By Proposition 6.8, we have $C = D_\gamma \cap D_\delta$ where $D_\gamma < D_\delta$ in $P(j_0)$. Since $C \cap D_1$ is nonempty, we must have that $\text{rows}^0(D_\gamma) \subset \text{rows}^0(D_1)$ or $\text{rows}^0(D_\delta) \subset \text{rows}^0(D_\delta)$, and similarly for $D_\delta$. But since $i \in [a]$, $(i,j_0) \not\subset D_\alpha$ for any $D_\alpha < D_1$. So we must have $\text{rows}^0(D_1) \subset \text{rows}^0(D_\gamma) \subset \text{rows}^0(D_\delta)$. Therefore, $\text{rows}^0(D_1) \subset \text{rows}^0(C)$.

Furthermore, we have that $(i,j_0) \in C$ for all $C \in \text{Int}(S)$ with $\text{rows}^0(D_1) \subset \text{rows}^0(C)$. So the factors $C^+$ corresponding to maximal intersections of maximal cliques are exactly those with $\text{rows}^0(D_1) \subset \text{rows}^0(C)$ in each $x_{i,j_0}$.
Therefore, we have that
\[
\sum_{i=1}^a x_{ij_0} = \sum_{i=1}^a u_{ij} + u_{j_0} \left( \prod_{C \in \text{Int}(j_0)} C^+ \right) \left( \prod_{D \in \text{Max}(S) \setminus \text{Max}(j_0)} D^+ \right) 
\]
\[
= \left( \prod_{C \in \text{Int}(S) \setminus N_j \neq \emptyset} C^+ \right) \left( \prod_{D \in \text{Max}(S) \setminus D \cap N_j \neq \emptyset} D^+ \right) \left( \prod_{E \in \text{Max}(S) \setminus \overline{N_j \cap D \cap E = \emptyset}} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right) 
\]
as needed. \(\square\)

Finally, the following proposition gives a way to write \(D^+\) as a sum over its intersections with the elements of \(P(j_0)\) that it covers, along with the rows of \(D_\beta\) that are not rows of any clique that it covers.

**Proposition 6.15.** Let \(D_{\alpha_1}, \ldots, D_{\alpha_\ell} < D_\beta\). Let \(r_1, \ldots, r_\ell\) be the rows of \(D_\beta\) that are not in any \(D_{\alpha_i}\) for \(i = 1, \ldots, \ell\). Then

\[
(7) \quad D^+_\beta = \sum_{i=1}^a u_{r_i} + \sum_{i=1}^\ell (D_{\alpha_i} \cap D_\beta)^+. 
\]

**Proof.** Without loss of generality, we will let \(D_2, \ldots, D_\ell < D_1\), and let rows \(1, \ldots, a\) be the rows of \(D_1\) that are not rows of any \(D_\alpha\) for \(\alpha = 2, \ldots, \ell\). First note that each \(u_{ij}\) that appears on the right-hand side of Equation (7) is a term of \(D^+_1\). Indeed, if \((i, j) \in D_{\alpha} \cap D_1\) for some \(\alpha = 2, \ldots, \ell\), this is clear.

Otherwise, we have \(i \in [a]\). For the sake of contradiction, suppose that there exists a column \(j\) so that \((i, j) \notin D_1\) but \((i, j) \in S\). But then \(\text{rows}^{j_0}(D_1) \cap \text{rows}^{j_0}(j)\) is non-empty. So by the DS-free condition, either \(\text{rows}^{j_0}(D_1) \subset \text{rows}^{j_0}(j)\) or \(\text{rows}^{j_0}(j) \subset \text{rows}^{j_0}(D_1)\). If \(\text{rows}^{j_0}(D_1) \subset \text{rows}^{j_0}(j)\), then \(j\) is a column of \(D_1\) by definition, which is a contradiction. If \(\text{rows}^{j_0}(j) \subset \text{rows}^{j_0}(D_1)\), then column \(j\) belongs to some block \(B_\beta\) with \(\text{rows}^{j_0}(D_\beta) \subset \text{rows}^{j_0}(D_1)\). But this contradicts that row \(i\) is not in any \(D_\alpha\) for \(\alpha = 2, \ldots, \ell\).

Now it remains to show that all the terms in \(D^+_1\) appear in the right-hand side of Equation (7). Let \((i, j) \in D_1\). If \(i \in [a]\), then \(u_{ij}\) is a term in the right-hand side, as needed. Otherwise, \(i \in \text{rows}^{j_0}(D_\alpha)\) for some \(\alpha \in \{2, \ldots, \ell\}\). Since \(\text{cols}(D_1) \subset \text{cols}(D_\alpha)\) by definition, we must have \(j \in \text{cols}(D_\alpha)\). So \((i, j) \in D_\alpha\). Therefore, \(u_{ij}\) is a term in \((D_\alpha \cap D_1)^+\). Finally, since \(D_\gamma \cap D_\delta = \emptyset\) for all \(\gamma, \delta \in \{2, \ldots, \ell\}\) with \(\gamma \neq \delta\), no term is repeated. \(\square\)

We can now use these propositions to prove Lemma 6.9.
Proof of Lemma 6.9. We will induct over the poset $P(j_0)$. For the base case, we let $D_\alpha$ be minimal in $P(j_0)$. First, we claim that

$$D_\alpha^+ = \sum_{i \in \text{rows}^{j_0}(D_\alpha)} u_{i+}.$$ 

Indeed, if this were not the case, there would be a column $j$ such that for two rows $i_1, i_2 \in \text{rows}^{j_0}(D_\alpha)$, we have $i_1 \in \text{rows}^{j_0}(j)$ and $i_2 \notin \text{rows}^{j_0}(j)$. Since $\text{rows}^{j_0}(D_\alpha) \cap \text{rows}^{j_0}(j) \neq \emptyset$, and $\text{rows}^{j_0}(D_\alpha) \nsubseteq \text{rows}^{j_0}(j)$, this implies that $\text{rows}^{j_0}(j) \subset \text{rows}^{j_0}(D_\alpha)$ by the DS-free condition. But this contradicts that $D_\alpha$ is minimal in $P(j_0)$.

Let $C \in \text{Int}(S)$ with $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(C)$ such that $C \cap N_{j_0} \neq \emptyset$. Then for any $i \in \text{rows}^{j_0}(D_\alpha), (i, j_0) \in C$. So $C \in \text{Int}(i, j_0)$ and $C^+$ is a factor of $x_{ij_0}$.

If $C \in \text{Int}(i, j_0)$, then $N_{j_0} \cap C \neq \emptyset$. It remains to be shown that all factors of $x_{ij_0}, C^+$ corresponding to maximal intersections of maximal cliques have $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(C)$. Let $(i, j_0) \in D_\alpha$. Let $D_\beta$ and $D_\gamma$ be maximal cliques such that $C = D_\beta \cap D_\gamma \in \text{Int}(i, j_0)$. Then we have $\text{rows}^{j_0}(D_\alpha) \cap \text{rows}^{j_0}(D_\beta)$ and $\text{rows}^{j_0}(D_\alpha) \cap \text{rows}^{j_0}(D_\gamma)$ nonempty. Since $D_\alpha$ is minimal, this implies that $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(D_\beta), \text{rows}^{j_0}(D_\gamma)$. So $\text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(C)$. Therefore the factors of each $x_{ij_0}$ with $(i, j_0) \in D_\alpha$ that correspond to maximal intersections of maximal cliques are

$$\prod_{\substack{C \in \text{Int}(S) \\ C \cap N_{j_0} \neq \emptyset \\ \text{rows}^{j_0}(D_\alpha) \subset \text{rows}^{j_0}(C)}} C^+,$$

as needed.

The other factors of each $x_{ij_0}$ for $(i, j_0) \in D_\alpha$ are of the form

$$\prod_{\substack{E \in \text{Max}(S) \\ (i, j_0) \notin E}} E^+.$$ 

Since all $(i, j_0) \in D_\alpha$ are contained in the same maximal cliques when $D_\alpha$ is minimal in $P(j_0)$, the terms corresponding to maximal cliques in each $x_{ij_0}$ are of the form

$$\prod_{\substack{E \in \text{Max}(S) \\ N_{j_0} \cap D_\alpha \cap E = \emptyset}} E^+,$$

as needed.
So we have that

\[
\sum_{(i,j_0) \in D_\alpha} x_{i0} = \sum_{(i,j_0) \in D_\alpha} u_{i_0}^+ u_{j_0}^+ \left( \prod_{C \in \text{Int}(S) \cap N_{j_0} \neq \emptyset} C^+ \right) \left( \prod_{E \in \text{Max}(S) \cap N_{j_0} \cap D_\alpha \cap E = \emptyset} E^+ \right)
\]

\[
= u_{j_0}^+ \left( \sum_{i \in \text{rows}^{0}(D_\alpha)} u_{i_0}^+ \right) \left( \prod_{C \in \text{Int}(S) \cap N_{j_0} \neq \emptyset} C^+ \right) \left( \prod_{E \in \text{Max}(S) \cap N_{j_0} \cap D_\alpha \cap E = \emptyset} E^+ \right)
\]

\[
= u_{j_0}^+ D_\alpha^+ \left( \prod_{C \in \text{Int}(S) \cap N_{j_0} \neq \emptyset} C^+ \right) \left( \prod_{E \in \text{Max}(S) \cap N_{j_0} \cap D_\alpha \cap E = \emptyset} E^+ \right).
\]

Since $D_\alpha$ is the only maximal clique whose rows are contained in $D_\alpha$, we have that

\[
D_\alpha^+ = \prod_{D \in \text{Max}(S) \cap N_{j_0} \neq \emptyset} D^+.
\]

So the lemma holds for the base case.

Without loss of generality, let $D_2, \ldots, D_\ell \ll D_1$ in $P(j_0)$. Let rows $1, \ldots, a$ be the rows of $D_1$ that are not in any $D_\alpha$ with $\alpha = 2, \ldots, \ell$. We have the following chain of equalities.
\[
\sum_{(i,j_0)\in D_1} x_{ij_0} = u_{+jo} \sum_{i=1}^a x_{ij_0} + \ell \sum_{\alpha=2}^{\ell} \sum_{(i,j_0)\in D_\alpha} x_{ij_0}
\]

\[
= u_{+jo} \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^p u_{i+} \right)
\]

\[
+ \sum_{\alpha=2}^{\ell} \sum_{(i,j_0)\in D_\alpha} x_{ij_0}
\]

\[
= u_{+jo} \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right)
\]

\[
+ \sum_{\alpha=2}^{\ell} u_{+jo} \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right)
\]

\[
= u_{+jo} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right)
\]

\[
+ \left( \sum_{i=1}^a u_{i+} \times \prod_{C \in \text{Int}(S)} C^+ \right) + \left( \sum_{\alpha=2}^{\ell} \prod_{(D_\alpha)\cap \text{rows}^{\alpha}_0 (C)} C^+ \right)
\]

\[
+ \sum_{\alpha=2}^{\ell} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right) + \sum_{\alpha=2}^{\ell} \left( \prod_{(D_\alpha)\cap \text{rows}^{\alpha}_0 (D_1)} C^+ \right)
\]

\[
= u_{+jo} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} + \sum_{\alpha=2}^{\ell} (D_\alpha \cap D_1)^+ \right)
\]

\[
+ \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \sum_{i=1}^a u_{i+} \right) + \sum_{\alpha=2}^{\ell} \left( \prod_{(D_\alpha)\cap \text{rows}^{\alpha}_0 (D_1)} C^+ \right)
\]

\[
= u_{+jo} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \prod_{C \in \text{Int}(S)} C^+ \right) \left( \sum_{i=1}^a u_{i+} \right)
\]

\[
+ \sum_{\alpha=2}^{\ell} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right) \left( \sum_{i=1}^a u_{i+} \right) + \sum_{\alpha=2}^{\ell} \left( \prod_{(D_\alpha)\cap \text{rows}^{\alpha}_0 (D_1)} C^+ \right)
\]
THE SECOND EQUALITY FOLLOWS FROM PROPOSITION 6.14. THE THIRD EQUALITY IS AN APPLICATION OF THE INDUCTIVE HYPOTHESIS. THE FOURTH EQUALITY FOLLOWS FROM PROPOSITION 6.12 ALONG WITH REMARK 6.13. THE FIFTH EQUALITY FOLLOWS FROM PROPOSITION 6.11. THE SIXTH EQUALITY FOLLOWS FROM PROPOSITION 6.15. THE SEVENTH INEQUALITY FOLLOWS FROM THE FACT THAT $D_1$ IS THE ONLY CLIQUE WHOSE ROWS ARE EQUAL TO ROWS $j_0(D_1)$. THIS COMPLETES OUR PROOF BY INDUCTION.

7. CHECKING THE CONDITIONS OF BIRCH’S THEOREM

In the previous section, we wrote a formula for the sum of $x_{ij_0}$ where $i$ ranges over the rows of some maximal clique $D_0$. Since the block $B_0$ induces its own maximal clique, lemma 6.9 allows us to write the sum of the $x_{ij_0}$’s for $1 \leq i \leq r$ in the following concise way. This in turn verifies that the proposed maximum likelihood estimate $\hat{p}$ has the same sufficient statistics as the normalized data $u/u_{++}$, which is one of the conditions of Birch’s theorem.

**Corollary 7.1.** Let $S$ be $DS$-free. Then for any column $j_0$,

$$\sum_{i=1}^{r} x_{ij_0} = u_{+j_0} \prod_{D \in \text{Max}(S)} D^+.\prod_{E \in \text{Max}(S)} E^+.$$\]

**Proof.** The poset $P(j_0)$ has a unique maximal element $D_0$ with rows $j_0(D_0) = rows^{j_0}(j_0)$. Note that $D_0$ may include more columns than $j_0$ since it may have columns whose nonzero rows are the same as or contain those of $j_0$.

By Proposition 6.8 there are no maximal intersections of maximal cliques $C$ with rows $j_0(D_0) \subset rows^{j_0}(C)$, since $D_0$ is maximal in $P(j_0)$. It follows from Proposition 6.5 that a maximal clique $D$ intersects $N_{j_0}$ if and only if it has rows $j_0(D) \subset rows^{j_0}(D_0)$.

Since $N_{j_0} \subset D_0$, we have that $N_{j_0} \cap D_0 \cap E = N_{j_0} \cap E$ for any clique $E$. By Lemma 6.9 we have

$$\sum_{i=1}^{r} x_{ij_0} = u_{+j_0} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right)$$

$$= u_{+j_0} \left( \prod_{D \in \text{Max}(S)} D^+ \right) \left( \prod_{E \in \text{Max}(S)} E^+ \right)$$

$$= u_{+j_0} \prod_{D \in \text{Max}(S)} D^+,$$

as needed. □

Now we will address the condition of Birch’s theorem which states that the maximum likelihood estimate must satisfy the equations defining $\mathcal{M}_S$.

**Lemma 7.2.** Let $S$ be doubly chordal bipartite. Let $u \in \mathbb{R}^S$ be a generic matrix of counts. Then the point $(\hat{p}_{ij} \mid (i, j) \in S)$ specified in Theorem 5.4 is in the Zariski closure of $\mathcal{M}_S$.\]
In order to prove this lemma, we must first describe the vanishing ideal of $\mathcal{M}_S$. We denote this ideal $\mathcal{I}(\mathcal{M}_S)$. It is a subset of the polynomial ring in $\# S$ variables,

$$R = \mathbb{C}[p_{ij} \mid ij \in S].$$

**Proposition 7.3.** Let $S$ be chordal bipartite. Then $\mathcal{I}(\mathcal{M}_S)$ is generated by the $2 \times 2$ minors of the matrix form of $S$ that contain no zeros. That is, $\mathcal{I}(\mathcal{M}_S)$ is generated by all binomials of the form

$$p_{ij}p_{k\ell} - p_{i\ell}p_{kj},$$

such that $(i, j), (k, \ell), (i, \ell), (k, j) \in S$.

**Proof.** This follows from results in [3, Chapter 10.1]. The loops on $S$ correspond to cycles in $G_S$. The df 1 loops are those whose support does not properly contain the support of any other loop; that is, they correspond to cycles in $G_S$ with no chords. Since $G_S$ is chordal bipartite, each of these cycles contain exactly four edges. Therefore the df 1 loops form a Markov basis for $\mathcal{M}_S$. Therefore, by the Fundamental Theorem of Markov Bases [7, Theorem 3.1], the $2 \times 2$ minors of $S$ form a generating set for $\mathcal{I}(\mathcal{M}_S)$.

**Example 7.4.** Consider the matrix $S$ from Example 5.3. In Figure 2, we see that $G(S)$ has exactly one cycle. This cycle corresponds to the only $2 \times 2$ minor in $S$ that contains no zeros, which is the $\{2, 3\} \times \{1, 2\}$ submatrix. Therefore the (complex) Zariski closure of $\mathcal{M}_S$ is the variety of the ideal generated by the polynomial $p_{21}p_{32} - p_{31}p_{22}$.

**Proposition 7.5.** Let $S$ be set of indices such that $G_S$ is doubly chordal bipartite. Let $\{i_1, i_2\} \times \{j_1, j_2\}$ be a set of indices that corresponds to a $2 \times 2$ minor of $S$ that contains no zeros. Let $\hat{p}_{i_1j_1}, \hat{p}_{i_2j_1}, \hat{p}_{i_1j_2}, \hat{p}_{i_2j_2}$ be as defined in Theorem 7.4. Then

$$(8) \quad \hat{p}_{i_1j_1}\hat{p}_{i_1j_2} = \hat{p}_{i_1j_2}\hat{p}_{i_2j_2}$$

**Proof.** The terms $u_{i_1+}, u_{i_2+}, u_{j_1+}, u_{j_2+}, u_{+j_1}$ and $u_{+j_2}$ each appear once in the numerator on each side of Equation (8), and $u_{++}^2$ appears in both denominators. Furthermore if $(i_1, j_1)$ and $(i_2, j_2)$ are both contained in any clique in $S$, then $(i_1, j_2)$ and $(i_2, j_1)$ are also in the clique by definition. So any term that is squared in the numerator or denominator on one side of Equation (8) is also squared on the other side. Therefore it suffices to show that Max$(i_1j_1) \cup$ Max$(i_2j_2) = \text{Max}(i_1j_2) \cup \text{Max}(i_2j_1)$ and Int$(i_1j_1) \cup \text{Int}(i_1j_2) = \text{Int}(i_2j_2) \cup \text{Int}(i_2j_1)$.

First, we will show that Max$(i_1j_1) \cup$ Max$(i_2j_2) = \text{Max}(i_1j_2) \cup \text{Max}(i_2j_1)$. Let $D \in \text{Max}(i_1j_1)$. If $(i_2, j_1) \in D$, then we are done.

Now suppose that $(i_2, j_1) \notin D$. Since $D$ intersects column $j_1$, by Proposition 6.5 we know that $D$ has the form $D_a$ for some block of columns $B_a$ that are identical on rows $j_1$. Let rows $j_1$ $(D_a)$ denote the set of nonzero rows of $D_a$ that are also nonzero rows of $j_1$. Since $(i_2, j_1) \notin D$, we have that $i_2 \notin \text{rows}^{j_1}(D_a)$ while $i_1 \in \text{rows}^{j_1}(D_a)$. Since rows $j_1$ $(D_a)$ $\cap$ rows $j_1$ $(D_a)$ is nonempty, and since rows $j_1$ $(D_a)$ $\not\subset$ rows $j_1$ $(D_a)$, we must have that rows $j_1$ $(D_a)$ $\subset$ rows $j_1$ $(D_a)$ by the DS-free condition. Therefore $(i_1, j_2) \in D_a$ by definition of $D_a$. So $D_a = D \in \text{Max}(i_1, j_2)$, as needed.

Switching the roles of $i_1$ and $i_2$ or the roles of $j_1$ and $j_2$ yields the desired equality.
Now let \( C \in \Int(i_1, j_1) \). Then \( C = D_\alpha \cap D_\beta \) where \( D_\beta \prec D_\alpha \) in the poset \( P(j_1) \) by Proposition 6.8. If \((i_2, j_1) \in C\), then we are done.

Now suppose that \((i_2, j_1) \not\in C\). Then we have that \( i_2 \not\in \text{rows}^{j_1}(D_\beta) \), whereas \( i_1 \in \text{rows}^{j_1}(D_\alpha) \) and \( i_1 \in \text{rows}^{j_1}(D_\beta) \). So we must have that \( \text{rows}^{j_1}(D_\alpha) \not\subseteq \text{rows}^{j_1}(D_\beta) \) by the DS-free condition. Since \( \text{rows}^{j_1}(D_\alpha) \cap \text{rows}^{j_1}(j_2) \) is nonempty, we must have that \( \text{rows}^{j_1}(D_\alpha) \subset \text{rows}^{j_1}(j_2) \). This follows from the DS-free condition and the fact that \( D_\alpha \) covers \( D_\beta \) in the poset \( P(j_1) \). Therefore \((i_1, j_2) \in D_\alpha, D_\beta \) by definition of these cliques.

Again, switching the roles of \( i_1 \) and \( i_2 \) or the roles of \( j_1 \) and \( j_2 \) in the above proof yields the desired equality.

**Proof of Lemma 7.2.** By Proposition 7.3 the vanishing ideal of \( M_\mathcal{S} \) consists of all fully-observed \( 2 \times 2 \) minors of \( S \). By Proposition 7.5 each of these \( 2 \times 2 \) minors vanishes when evaluated on \( \hat{p} \).

We can now prove Theorem 5.4.

**Proof of Theorem 5.4.** Let \( G_\mathcal{S} \) be doubly chordal bipartite. Let \( u \in \mathbb{R}_+^S \) be a matrix of counts. By Corollary 7.1 the column marginals of \( u_{++}\hat{p} \) are equal to those of \( u \). Switching the roles of rows and columns in all of the proofs used to obtain this corollary shows that the row marginals are also equal. Corollary 7.1 also implies that \( \hat{p}_{++} = 1 \). So by Lemma 7.2 and the fact each \( \hat{p} \) is positive, \( \hat{p} \in M_\mathcal{S} \). Hence by Birch’s theorem, \( \hat{p} \) is the maximum likelihood estimate for \( U \). The other direction is exactly the contrapositive of Theorem 4.5.

**Acknowledgments**

Jane Coons was partially supported by the US National Science Foundation (DGE1746939). Seth Sullivant was partially supported by the US National Science Foundation (DMS 1615660).

**References**

[1] Agresti, A. Modelling patterns of agreement and disagreement. *Statistical methods in medical research* 1, 2 (1992), 201–218.

[2] Améndola, C., Bliss, N., Burke, I., Gibbons, C. R., Helmer, M., Hoşten, S., Nash, E. D., Rodriguez, J. I., and Smolkin, D. The maximum likelihood degree of toric varieties. *J. Symbolic Comput.* 92 (2019), 222–242.

[3] Aoki, S., Hara, H., and Takemura, A. *Markov bases in algebraic statistics*, vol. 199. Springer Science & Business Media, 2012.

[4] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. *Discrete multivariate analysis: theory and practice*. Springer, New York, 2007. With the collaboration of Richard J. Light and Frederick Mosteller, Reprint of the 1975 original.

[5] Bocci, C., and Rapallo, F. Exact tests to compare contingency tables under quasi-independence and quasi-symmetry. *J. Algebr. Stat.* 10, 1 (2019), 13–29.

[6] Colombo, A., and Ihm, P. A quasi-independence model to estimate failure rates. *Reliability Engineering & System Safety* 21, 4 (1988), 309–318.

[7] Diaconis, P., and Sturmfels, B. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics* 26, 1 (1998), 363–397.

[8] Duarte, E., Marigliano, O., and Sturmfels, B. Discrete statistical models with rational maximum likelihood estimator. arXiv:1903.06110, 2019.
[9] Geiger, D., Meek, C., and Sturmfels, B. On the toric algebra of graphical models. *Ann. Statist.* 34, 3 (2006), 1463–1492.

[10] Goodman, L. A. On quasi-independence and quasi-dependence in contingency tables, with special reference to ordinal triangular contingency tables. *Journal of the American Statistical Association* 89, 427 (1994), 1059–1063.

[11] Gouveia, J. A., Grappe, R., Kaibel, V., Pashkovich, K., Robinson, R. Z., and Thomas, R. R. Which nonnegative matrices are slack matrices? *Linear Algebra Appl.* 439, 10 (2013), 2921–2933.

[12] Huh, J. Varieties with maximum likelihood degree one. *J. Algebr. Stat.* 5, 1 (2014), 1–17.

[13] Huh, J., and Sturmfels, B. Likelihood geometry. In *Combinatorial algebraic geometry*, vol. 2108 of *Lecture Notes in Math.* Springer, Cham, 2014, pp. 63–117.

[14] Kapranov, M. M. A characterization of $A$-discriminantal hypersurfaces in terms of the logarithmic Gauss map. *Math. Ann.* 290, 2 (1991), 277–285.

[15] Rapallo, F. Algebraic exact inference for rater agreement models. *Stat. Methods Appl.* 14, 1 (2005), 45–66.

[16] Rauh, J., Kahle, T., and Ay, N. Support sets in exponential families and oriented matroid theory. *Internat. J. Approx. Reason.* 52, 5 (2011), 613–626.

[17] Smith, P. W., and McDonald, J. W. Exact conditional tests for incomplete contingency tables: estimating attained significance levels. *Statistics and Computing* 5, 3 (1995), 253–256.

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