Algebraic Geometry of Bayesian Networks

Luis David Garcia, Michael Stillman and Bernd Sturmfels

Abstract

We study the algebraic varieties defined by the conditional independence statements of Bayesian networks. A complete algebraic classification is given for Bayesian networks on at most five random variables. Hidden variables are related to the geometry of higher secant varieties.

1   Introduction

The emerging field of *algebraic statistics* [16] advocates polynomial algebra as a tool in the statistical analysis of experiments and discrete data. Statistics textbooks define a *statistical model* as a family of probability distributions, and a closer look reveals that these families are often real algebraic varieties: they are the zeros of some polynomials in the probability simplex [7], [17].

In this paper we examine *directed graphical models* for discrete random variables. Such models are also known as *Bayesian networks* and they are widely used in machine learning, bioinformatics and many other applications [13], [15]. Our aim is to place Bayesian networks into the realm of algebraic statistics, by developing the necessary theory in algebraic geometry and by demonstrating the effectiveness of Gröbner bases for this class of models.

Bayesian networks can be described in two possible ways, either by a recursive factorization of probability distributions or by conditional independence statements (local and global Markov properties). This is an instance of the computer algebra principle that varieties can be presented either parametrically or implicitly [4, §3.3]. The equivalence of these two representations for Bayesian networks is a well-known theorem in statistics [13, Theorem 3.27], but, as we shall see, this theorem is surprisingly delicate and no longer holds when probabilities are replaced by negative reals or complex numbers. Hence in the usual setting of algebraic geometry, where the zeros lie in $\mathbb{C}^d$, there are many "distributions" which satisfy the global Markov property but which do not permit a recursive factorization. We explain this phenomenon using primary decomposition of polynomial ideals.

1

This paper is organized as follows. In Section 2 we review the algebraic theory of conditional independence, and we explicitly determine the Gröbner basis and primary decomposition arising from the *contraction axiom* [15], [21, §2.2.2]. This axiom is shown to fail for negative real numbers. In Section 3 we introduce the ideals $I_{\mathrm{local}(G)}$ and $I_{\mathrm{global}(G)}$ which represent a a Bayesian network G. When G is a forest then these ideals are the toric ideals derived from undirected graphs as in [8]; see Theorem 6 below.

The recursive factorization of a Bayesian network gives rise to a map between polynomial rings which is studied in Section 4. The kernel of this *factorization map* is the *distinguished prime ideal*. We prove that this prime is always a reduced primary component of $I_{\mathrm{local}(G)}$ and $I_{\mathrm{global}(G)}$. Our results in that section include the solutions to Problems 8.11 and 8.12 in [23].

In Sections 5 and 6 we present the results of our computational efforts: the complete algebraic classification of all Bayesian networks on four arbitrary random variables and all Bayesian networks on five binary random variables. The latter involved computing the primary decomposition of 301 ideals generated by a large number of quadrics in 32 unknowns. These large-scale primary decompositions were carried out in `Macaulay2` [10]. Some of the techniques and software tools we used are described in the Appendix.

The appearance of hidden variables in Bayesian networks leads to challenging problems in algebraic geometry. Statisticians have known for decades that the dimension of the corresponding varieties can unexpectedly drop [9], but the responsible singularities have been studied only quite recently, in [7] and [17]. In Section 7 we examine the elimination problem arising from hidden random variables, and we relate it to problems in projective algebraic geometry. We demonstrate that the *naive Bayes model* corresponds to the higher secant varieties of Segre varieties ([2], [3]), and we present several new results on the dimension and defining ideals of these secant varieties.

Our algebraic theory does not compete with but rather complements other approaches to conditional independence models. An impressive combinatorial theory of such models has been developed by Matúš [14] and Studený [21], culminating in their characterization of all realizable independence models on four random variables. Sharing many of the views expressed by these authors, we believe that exploring the precise relation between their work and ours will be a very fruitful research direction for the near future.

2   Ideals, Varieties and Independence Models

We begin by reviewing the general algebraic framework for independence models presented in [23, §8]. Let $X_1, \ldots, X_n$ be discrete random variables where X_i takes values in the finite set $[d_i] = \{1, 2, \ldots, d_i\}$. We write $D = [d_1] \times [d_2] \times \cdots \times [d_n]$ so that $\mathbb{R}^D$ denotes the real vector space of n-dimensional tables of format $d_1 \times \cdots \times d_n$. We introduce an indeterminate $p_{u_1 u_2 \cdots u_n}$ which represents the probability of the event $X_1 = u_1$, $X_2 = u_2, \ldots, X_n = u_n$. These indeterminates generate the ring $\mathbb{R}[D]$ of polynomial functions on the space of tables $\mathbb{R}^D$. A *conditional independence statement* has the form

$$A \text{ is independent of } B \text{ given } C \qquad (\text{ in symbols: } A \perp\!\!\!\perp B \mid C) \qquad (1)$$

where A, B and C are pairwise disjoint subsets of $\{X_1, \ldots, X_n\}$. If C is empty then (1) means that A is independent of B. By [23, Proposition 8.1], the statement (1) translates into a set of homogeneous quadratic polynomials in $\mathbb{R}[D]$, and we write $I_{A \perp\!\!\!\perp B|C}$ for the ideal generated by these polynomials.

Many statistical models (see e.g. [13], [21]) can be described by a finite set of independence statements (1). An *independence model* is any such set:

$$\mathcal{M} = \{A^{(1)} \perp\!\!\!\perp B^{(1)} \mid C^{(1)}, \ldots, A^{(m)} \perp\!\!\!\perp B^{(m)} \mid C^{(m)}\}.$$

The ideal of the independence model $\mathcal{M}$ is defined as the sum of ideals

$$I_{\mathcal{M}} = I_{A^{(1)} \perp\!\!\!\perp B^{(1)}|C^{(1)}} + \cdots + I_{A^{(m)} \perp\!\!\!\perp B^{(m)}|C^{(m)}}.$$

We wrote code in `Macaulay2` [10] and `Singular` [11] for generating the ideals $I_{\mathcal{M}}$. The *independence variety* is the set $V(I_{\mathcal{M}})$ of common zeros in $\mathbb{C}^D$ of the polynomials in $I_{\mathcal{M}}$. Equivalently, $V(I_{\mathcal{M}})$ is the set of all $d_1 \times \cdots \times d_n$-tables with complex number entries which satisfy the conditional independence statements in $\mathcal{M}$. The variety $V(I_{\mathcal{M}})$ has three natural subsets:

- the subset of real tables, denoted $V_{\mathbb{R}}(I_{\mathcal{M}})$,

- the non-negative tables, denoted $V_{\geq}(I_{\mathcal{M}})$,

- the non-negative tables whose entries sum to one, $V_{\geq}(I_{\mathcal{M}} + \langle p - 1 \rangle)$,

Here p denotes the sum of all unknowns $p_{u_1 \cdots u_n}$, so that $V_{\geq}(I_{\mathcal{M}} + \langle p - 1 \rangle)$ is the subset of the probability simplex specified by the model $\mathcal{M}$.

We illustrate these definitions by analyzing the independence model $\mathcal{M} = \{1 \perp\!\!\!\perp 2 \mid 3, \ 2 \perp\!\!\!\perp 3\}$ for $n = 3$ discrete random variables. Theorem 1 will be cited in Section 5 and it serves as a preview to Theorem 11.

3

The ideal $I_\mathcal{M}$ lies in the polynomial ring $\mathbb{R}[D]$ in $d_1 d_2 d_3$ unknowns p_{ijk}. Its minimal generators are $\binom{d_1}{2}\binom{d_2}{2}d_3$ quadrics of the form $p_{ijk}\,p_{rsk} - p_{isk}\,p_{rjk}$ and $\binom{d_2}{2}\binom{d_3}{2}$ quadrics of the form $p_{+jk}\,p_{+st} - p_{+jt}\,p_{+sk}$. We change coordinates in $\mathbb{R}[D]$ by replacing each unknown p_{1jk} by $p_{+jk} = \sum_{i=1}^{d_1} p_{ijk}$. This coordinate change transforms $I_\mathcal{M}$ into a binomial ideal in $\mathbb{R}[D]$.

Theorem 1. *The ideal $I_\mathcal{M}$ has a Gröbner basis consisting of squarefree binomials of degree two, three and four, and it is hence radical. It has $2^{d_3}-1$ minimal primes, each generated by the 2×2-minors of a generic matrix.*

Proof. The minimal primes of $I_\mathcal{M}$ will be indexed by proper subsets of $[d_3]$. For each such subset σ we introduce the monomial prime

$$M_\sigma \quad = \quad \langle\, p_{+jk} \mid j \in [d_2],\ k \in \sigma \,\rangle,$$

and the complementary monomial

$$m_\sigma \quad = \quad \prod_{j=1}^{d_2} \prod_{k \in [d_3]\backslash\sigma} p_{+jk}\,,$$

and we define the ideal

$$P_\sigma \quad := \quad \big((I_\mathcal{M} + M_\sigma) : m_\sigma^\infty\big).$$

It follows from the general theory of binomial ideals [6] that P_σ is a binomial prime ideal. A closer look reveals that P_σ is minimally generated by the $d_2 \cdot |\sigma|$ variables in M_σ together with all the 2×2-minors of the following two-dimensional matrices: the matrix (p_{ijk}) where the rows are indexed by $j \in [d_2]$ and the columns are indexed by pairs (i,k) with $i \in \{+, 2, 3, \ldots, d_1\}$ and $k \in [d_3]\backslash\sigma$, and for each $k \in \sigma$, the matrices (p_{ijk}) where the rows are indexed by $j \in [d_2]$ and the columns are indexed by $i \in \{2, 3, \ldots, d_1\}$.

We partition $V(I_\mathcal{M})$ into 2^{d_3} strata, each indexed by a subset σ of $[d_3]$. Namely, given a point (p_{ijk}) in $V(I_\mathcal{M})$ we define the subset σ of $[d_3]$ as the set of all indices k such that $(p_{+1k}, p_{+2k}, \ldots, p_{+d_2 k})$ is the zero vector. Note that two tables (p_{ijk}) lie in the same stratum if and only if they give the same σ. The stratum indexed by σ is a dense subset in $V(P_\sigma)$. When $\sigma = [d_3]$ the stratum consists of all tables such that the line sums p_{+jk} are all zero, and for each fixed k, the remaining $(d_1 - 1) \times d_2$-matrix (p_{ijk}) with $i \geq 2$ has rank ≤ 1. So this locus is defined by the prime ideal $P_{[d_3]}$. Any point in this stratum satisfies the defining equations of P_σ for any proper subset σ. So the stratum indexed by $[d_3]$ lies in the closure of all

4

other strata. But all remaining $2^{d_3} - 1$ strata have the property that no stratum lies in the closure of any other stratum, since the generic point of P_σ lies in exactly one stratum for any proper subset σ. Hence $V(I_\mathcal{M})$ is the irredundant union of the irreducible varieties $V(P_\sigma)$ where σ runs over all proper subsets of $[d_3]$. The second assertion in Theorem 1 now follows from Hilbert's Nullstellensatz.

To prove the first assertion, let us first note that $P_\emptyset$ is the prime ideal of 2×2-minors of the $d_2 \times (d_1 d_3)$-matrix (p_{ijk}) with rows indexed by $j \in [d_2]$ and columns indexed by pairs $(i, k) \in \{+, 2, 3, \ldots, d_1\} \times [d_3]$. Hence

$$P_\emptyset \quad = \quad \left(I_\mathcal{M} : m_\emptyset^\infty \right) \quad = \quad I_{2 \perp\!\!\!\perp \{1,3\}}. \tag{2}$$

It is well known (see e.g. [22, Proposition 5.4]) that the quadratic generators

$$p_{ijk} \, p_{rst} \quad - \quad p_{isk} \, p_{rjt} \tag{3}$$

form a reduced Gröbner basis for (2) with respect to the "diagonal term order". We modify this Gröbner basis to a Gröbner basis for $I_\mathcal{M}$ as follows:

- if $k = t$ take (3),

- if $i = +$ and $r = +$ take (3),

- if $i = +$ and $r \neq +$ and $k \neq t$ take (3) times p_{+jt} for any j,

- if $i \neq +$ and $r \neq +$ and $k \neq t$ take (3) times $p_{+jt} p_{+sk}$ for any j, s.

All of these binomials lie in $I_\mathcal{M}$ (this can be seen by taking S-pairs of the generators) and their S-pairs reduce to zero. By Buchberger's criterion, the given set of quadrics, cubics and quartics is a Gröbner basis, and the corresponding initial monomial ideal is square-free. This implies that $I_\mathcal{M}$ is radical (by [23, Proposition 5.3]), and the proof is complete. $\qquad \square$

The theorem above can be regarded as an algebraic refinement of the following well-known rule for conditional independence ([15], [21, §2.2.2]).

Corollary 2. (Contraction Axiom) *If a probability distribution on $[d_1] \times [d_2] \times [d_3]$ satisfies $1 \perp\!\!\!\perp 2 \mid 3$ and $2 \perp\!\!\!\perp 3$ then it also satisfies $2 \perp\!\!\!\perp \{1, 3\}$.*

Proof. The non-negative points satisfy $V_\geq(P_\sigma) \subseteq V_\geq(P_\emptyset)$, and this implies

$$V_\geq(I_\mathcal{M}) \quad = \quad V_\geq(I_{2 \perp\!\!\!\perp \{1,3\}}).$$

Intersecting with the probability simplex yields the assertion. $\qquad \square$

Theorem 1 shows that the Contraction Axiom fails to hold when probabilities are replaced by negative real numbers. Any general point on $V(P_\sigma)$ for $\sigma \neq \emptyset$ satisfies $1 \perp\!\!\!\perp 2 \,|\, 3$ and $2 \perp\!\!\!\perp 3$ but it does not satisfy $2 \perp\!\!\!\perp \{1, 3\}$.

3 Algebraic Representation of Bayesian Networks

A *Bayesian network* is an acyclic directed graph G with vertices $X_1, \ldots, X_n$. The following notation and terminology is consistent with Lauritzen's book [13]. The *local Markov property* on G is the set of independence statements

$$\text{local}(G) \quad = \quad \{X_i \perp\!\!\!\perp \text{nd}(X_i) \mid \text{pa}(X_i) \,:\, i = 1, 2, \ldots, n\},$$

where $\text{pa}(X_i)$ denotes the set of parents of X_i in G and $\text{nd}(X_i)$ denotes the set of nondescendents of X_i in G. Here X_j is a *nondescendent* of X_i if there is no directed path from X_i to X_j in G. The *global Markov property*, global(G), is the set of independence statements $A \perp\!\!\!\perp B \mid C$, for any triple A, B, C of subsets of vertices of G such that A and B are *d-separated* by C. Here two subsets A and B are said to be *d-separated* by C if all chains from A to B are blocked by C. A chain π from X_i to X_j in G is said to be *blocked* by a set C of nodes if it contains a vertex $X_b \in \pi$ such that either

- $X_b \in C$ and arrows of π do not meet head-to-head at X_b, or

- $X_b \notin C$ and X_b has no descendents in C, and arrows of π do meet head-to-head at X_b.

For any Bayesian network G, we have $\text{local}(G) \subseteq \text{global}(G)$, and this implies the following containment relations between ideals and varieties

$$I_{\text{local}(G)} \subseteq I_{\text{global}(G)} \qquad \text{and} \qquad V_{\text{local}(G)} \supseteq V_{\text{global}(G)}. \qquad (4)$$

The latter inclusion extends to the three real varieties listed above, and we shall discuss when equality holds. First, however, we give an algebraic version of the description of Bayesian networks by recursive factorizations.

Consider the set of parents of the j-th node, $\text{pa}(X_j) = \{X_{i_1}, \ldots, X_{i_r}\}$, and consider any event $X_j = u_0$ conditioned on $X_{i_1} = u_1, \ldots, X_{i_r} = u_r$, where $1 \leq u_0 \leq d_j, 1 \leq u_1 \leq d_{i_1}, \ldots, 1 \leq u_r \leq d_{i_r}$. We introduce an unknown $q_{u_0 u_1 \cdots u_r}^{(j)}$ to denote the conditional probability of this event, and we subject these unknowns to the linear relations $\sum_{v=1}^{d_j} q_{v u_1 \cdots u_r}^{(j)} = 1$ for all $1 \leq u_1 \leq d_{i_1}, \ldots, 1 \leq u_r \leq d_{i_r}$. Thus, we have introduced $(d_j - 1) d_{i_1} \cdots d_{i_r}$ unknowns for the vertex j. Let E denote the set of these unknowns $q_{u_0 u_1 \cdots u_r}^{(j)}$ for all $j \in \{1, \ldots, n\}$, and let $\mathbb{R}[E]$ denote the polynomial ring they generate.

6

If the n random variables are binary ($d_i = 2$ for all i) then the notation for $\mathbb{R}[E]$ can be simplified by dropping the first lower index and writing

$$q^j_{u_1 \cdots u_r} \quad := \quad q^{(j)}_{1u_1 \cdots u_r} \quad = \quad 1 - q^{(j)}_{2u_1 \cdots u_r}$$

In the binary case, $\mathbb{R}[E]$ is a polynomial ring in $\sum_{j=1}^n 2^{|\operatorname{pa}(X_j)|}$ unknowns.

The factorization of probability distributions according to G defines a polynomial map $\phi : \mathbb{R}^E \to \mathbb{R}^D$. By restricting to non-negative reals we get an induced map $\phi_{\geq 0}$. These maps are specified by the ring homomorphism $\Phi : \mathbb{R}[D] \to \mathbb{R}[E]$ which takes the unknown $p_{u_1 u_2 \cdots u_n}$ to the product of the expressions $q^{(j)}_{u_j u_{i_1} \cdots u_{i_r}}$ as j runs over $\{1, \ldots, n\}$. The image of ϕ lies in the independence variety $V_{\operatorname{global}(G)}$, or, equivalently, the independence ideal $I_{\operatorname{global}(G)}$ is contained in the prime ideal $\ker(\Phi)$. The *Factorization Theorem* for Bayesian networks [13, Theorem 3.27] states:

Theorem 3. *The following four subsets of the probability simplex coincide:*

$$V_{\geq}(I_{\operatorname{local}(G)} + \langle p - 1 \rangle) \quad = \quad V_{\geq}(I_{\operatorname{global}(G)} + \langle p - 1 \rangle)$$
$$= \quad V_{\geq}(\ker(\Phi)) \quad = \quad \operatorname{image}(\phi_{\geq}).$$

Example 4. Let G be the network on three binary random variables which has a single directed edge from 3 to 2. The parents and nondescendents are

$$\operatorname{pa}(1) = \emptyset, \operatorname{nd}(1) = \{2, 3\}, \ \operatorname{pa}(2) = \{3\}, \operatorname{nd}(2) = \{1\}, \ \operatorname{pa}(3) = \emptyset, \operatorname{nd}(3) = \{1\}.$$

The resulting conditional independence statements are

$$\operatorname{local}(G) \ = \ \operatorname{global}(G) \ = \ \{1 \perp\!\!\!\perp 3, \ 1 \perp\!\!\!\perp 2 \mid 3, \ 1 \perp\!\!\!\perp \{2, 3\}\}.$$

The ideal expressing the first two statements is contained in the ideal expressing the third statement, and we find that $I_{\operatorname{local}(G)} = I_{1 \perp\!\!\!\perp \{2,3\}}$ is the ideal generated by the six 2×2-subdeterminants of the 2×4-matrix

$$\begin{pmatrix} p_{111} & p_{112} & p_{121} & p_{122} \\ p_{211} & p_{212} & p_{221} & p_{222} \end{pmatrix} \tag{5}$$

This ideal is prime and its generators form a Gröbner basis. The Factorization Theorem is understood as follows for this example. We have $E = \{q^1, q^2_1, q^2_2, q^3\}$, and our ring map Φ takes the matrix (5) to

$$\begin{pmatrix} q^1 q^2_1 q^3 & q^1 q^2_2 (1 - q^3) & q^1 (1 - q^2_1) q^3 & q^1 (1 - q^2_2)(1 - q^3) \\ (1 - q^1) q^2_1 q^3 & (1 - q^1) q^2_2 (1 - q^3) & (1 - q^1)(1 - q^2_1) q^3 & (1 - q^1)(1 - q^2_2)(1 - q^3) \end{pmatrix}$$

The map ϕ from $\mathbb{R}^4$ to $\mathbb{R}^8$ corresponding to the ring map $\Phi : \mathbb{R}[D] \to \mathbb{R}[E]$ gives a parametrization of all 2×4-matrices of rank 1 whose entries sum to 1. The Factorization Theorem for G is the same statement for non-negative matrices. The kernel of Φ is exactly equal to $I_{\mathrm{local}(G)} + \langle p - 1 \rangle$. $\qquad \square$

Our aim is to decide to what extent the Factorization Theorem is valid over all real and all complex numbers. The corresponding algebraic question is to study the ideal $I_{\mathrm{local}(G)}$ and to determine its primary decomposition. Let us begin by considering all Bayesian networks on three random variables. We shall prove that for such small networks the ideal $I_{\mathrm{local}(G)}$ is always prime and coincides with the kernel of Φ. The following theorem is valid for arbitrary positive integers d_1, d_2, d_3. It is not restricted to the binary case.

Proposition 5. *For any Bayesian network G on three discrete random variables, the ideal $I_{\mathrm{local}(G)}$ is prime, and it has a quadratic Gröbner basis.*

Proof. We completely classify all possible cases. If G is the complete graph, directed acyclically, then $\mathrm{local}(G)$ contains no nontrivial independence statements, so $I_{\mathrm{local}(G)}$ is the zero ideal. In what follows we always exclude this case. There are five isomorphism types of (non-complete) directed acyclic graphs on three nodes. They correspond to the rows of the following table:

Graph	Local/Global Markov property	Independence ideal
3 2 1	$1 \perp\!\!\!\perp \{2,3\}$, $2 \perp\!\!\!\perp \{1,3\}$, $3 \perp\!\!\!\perp \{1,2\}$	I_{Segre}
$3 \longrightarrow 2 \quad 1$	$1 \perp\!\!\!\perp 3$, $1 \perp\!\!\!\perp 2 \mid 3$, $1 \perp\!\!\!\perp \{2,3\}$	$I_{1 \perp\!\!\!\perp \{2,3\}}$
$3 \longrightarrow 2 \longrightarrow 1$	$1 \perp\!\!\!\perp 3 \mid 2$	$I_{1 \perp\!\!\!\perp 3 \mid 2}$
$1 \longleftarrow 3 \longrightarrow 2$	$1 \perp\!\!\!\perp 2 \mid 3$	$I_{1 \perp\!\!\!\perp 2 \mid 3}$
$3 \longrightarrow 1 \longleftarrow 2$	$2 \perp\!\!\!\perp 3$	$I_{2 \perp\!\!\!\perp 3}$

The third and fourth network represent the same independence model. In all cases except for the first, the ideal $I_{\mathrm{local}(G)}$ is of the form $I_{A \perp\!\!\!\perp B \mid C}$, i.e., it is specified by a single independence statement. It was shown in [23, Lemma 8.2] that such ideals are prime. They are determinantal ideals and well known to possess a quadratic Gröbner basis. The only exceptional graph is the empty graph, which leads to the model of complete independence $1 \perp\!\!\!\perp \{2,3\}$, $2 \perp\!\!\!\perp \{1,3\}$, $3 \perp\!\!\!\perp \{1,2\}$. The corresponding ideal defines the Segre embedding of the product of three projective spaces $\mathbb{P}^{d_1-1} \times \mathbb{P}^{d_2-1} \times \mathbb{P}^{d_3-1}$ into $\mathbb{P}^{d_1 d_2 d_3 - 1}$. This ideal is prime and has a quadratic Gröbner basis. $\qquad \square$

A network G is a *directed forest* if every node has at most one parent. The conclusion of Proposition 5 also holds for directed forests on any number

of nodes. Proposition 16 will show that the direction of the edges is crucial: it is not sufficient to assume that the underlying undirected graph is a forest.

Theorem 6. *Let G be a directed forest. Then $I_{\text{global}(G)}$ is prime and has a quadratic Gröbner basis. These properties generally fail for $I_{\text{local}(G)}$.*

Proof. For a direct forest, the definition of a *blocked* chain reads as follows. A chain π from X_i to X_j in G is blocked by a set C if it contains a vertex $X_b \in \pi \cap C$. Hence, C d-separates A from B if and only if C separates A from B in the undirected graph underlying G. Thus, [8, Theorem 12] implies that $I_{\text{global}(G)}$ coincides with the distinguished prime ideal $\ker(\Phi)$, this ideal has a quadratic Gröbner basis. The second assertion is proved by the networks 18 and 26 in Table 1. See also [23, Example 8.8]. □

We close this section with a conjectured characterization of the global Markov property on a Bayesian network G in terms of commutative algebra.

Conjecture 7. *$I_{\text{global}(G)}$ is the ideal generated by all quadrics in $\ker(\Phi)$.*

4 The Distinguished Component

In what follows we shall assume that every edge (i, j) of the Bayesian network G satisfies $i > j$. In particular, the node 1 is always a sink and the node n is always a source. For any integer $r \in [n]$ and $u_i \in [d_i]$ as before, we abbreviate the *marginalization* over the first r random variables as follows:

$$p_{++\cdots+u_{r+1}\cdots u_n} \quad := \quad \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \cdots \sum_{i_r=1}^{d_r} p_{i_1 i_2 \cdots i_r u_{r+1} \cdots u_n}.$$

This is a linear form in our polynomial ring $\mathbb{R}[D]$. We denote by $\mathbf{p}$ the product of all of these linear forms. Thus the equation of $\mathbf{p} = 0$ defines a hyperplane arrangement in $\mathbb{R}^D$. We shall prove that the ideal $I_{\text{local}(G)}$ is prime locally outside this hyperplane arrangement, and hence so is $I_{\text{global}(G)}$. The following theorem provides the solution to [23, Problem 8.12].

Theorem 8. *The prime ideal $\ker(\Phi)$ is a minimal primary component of both of the ideals $I_{\text{local}(G)}$ and $I_{\text{global}(G)}$. More precisely,*

$$\left(I_{\text{local}(G)} : \mathbf{p}^\infty\right) \quad = \quad \left(I_{\text{global}(G)} : \mathbf{p}^\infty\right) \quad = \quad \ker(\Phi). \tag{6}$$

The prime ideal $\ker(\Phi)$ is called the *distinguished component*. It can be characterized as the set of all homogeneous polynomial functions on $\mathbb{R}^D$ which vanish on all probability distributions that factor according to G.

9

Proof. We relabel G so that $\mathrm{pa}(1) = \{2, 3, \ldots, r\}$ and $\mathrm{nd}(1) = \{r+1, \ldots, n\}$. Let A denote a set of $(d_1 - 1)d_2 \cdots d_r$ new unknowns $a_{i_1 i_2 \cdots i_r}$, for $i_1 > 1$ defining a polynomial ring $\mathbb{R}[A]$. Define $d_2 \cdots d_r$ linear polynomials

$$a_{1\,i_2 \cdots i_r} \quad = \quad 1 - \sum_{j=2}^{d_1} a_{j\,i_2 \cdots i_r}.$$

Let Q denote a set of $d_2 \cdots d_n$ new unknowns $q_{i_2 \cdots i_r i_{r+1} \cdots i_n} = q_{i_2 \cdots i_n}$, defining a polynomial ring $\mathbb{R}[Q]$. We introduce the partial factorization map

$$\Psi : \mathbb{R}[D] \to \mathbb{R}[A \cup Q]\,, \ p_{i_1 i_2 \cdots i_n} \ \mapsto \ a_{i_1 \cdots i_r} \cdot q_{i_2 \cdots i_n}. \tag{7}$$

The kernel of Ψ is precisely the ideal $I_1 := I_{1 \perp\!\!\!\perp \mathrm{nd}(1)|\mathrm{pa}(1)}$. Note that

$$q_{i_2 \cdots i_n} \ = \ \Psi(p_{+ i_2 \cdots i_n}).$$

Therefore Ψ becomes an epimorphism if we localize $\mathbb{R}[D]$ at the product $\mathbf{p}_1$ of the $p_{+ i_2 \cdots i_n}$ and we localize R at the product of the $q_{i_2 \cdots i_n}$. This implies that any ideal L in the polynomial ring $\mathbb{R}[D]$ satisfies the identity

$$\Psi^{-1}(\Psi(L)) \quad = \quad \big((L + I_1) : \mathbf{p}_1^\infty\big). \tag{8}$$

Let G' denote the graph obtained from G by removing the sink 1 and all edges incident to 1. We regard $I_{\mathrm{local}(G')}$ as an ideal in $\mathbb{R}[Q]$. We modify the set of independence statements $\mathrm{local}(G)$ by removing 1 from the sets $\mathrm{nd}(i)$ for any $i \geq 2$. Let $J \subset \mathbb{R}[D]$ be the ideal corresponding to these modified independence statements, so that $\Psi(J) = I_{\mathrm{local}(G')}$. Note that

$$J + I_1 \ \subseteq \ I_{\mathrm{local}(G)} \ \subseteq \ I_{\mathrm{global}(G)} \ \subseteq \ \ker(\Phi),$$

so it suffices to show that $(J + I_1) : \mathbf{p}^\infty = \ker(\Phi)$. The map Φ factors as

$$\mathbb{R}[D] \xrightarrow{\ \Psi\ } \mathbb{R}[A \cup Q] \xrightarrow{\ \Phi'\ } \mathbb{R}[A \cup E'] = \mathbb{R}[E], \tag{9}$$

where Φ' is the factorization map coming from the graph G', extended to be the identity on the variables A. By induction on the number of vertices, we may assume that Theorem 8 holds for the smaller graph G', i.e.,

$$\ker(\Phi') \quad = \quad (I_{\mathrm{local}(G')} : \mathbf{q}_2^\infty) \quad = \quad \Psi(J : \mathbf{p}_2^\infty), \tag{10}$$

where $\mathbf{q}_2 = \Psi(\mathbf{p}_2)$ and $\mathbf{p}_2$ is the product of the linear forms $p_{++\cdots+u_i \cdots u_n}$ with at least two initial $+$'s. Therefore

$$\ker(\Phi) \quad = \quad \Psi^{-1}(\Psi(J : \mathbf{p}_2^\infty)). \tag{11}$$

Applying (8), we get $\ker(\Phi) = ((J : \mathbf{p}_2^\infty) + I_1) : \mathbf{p}_1{}^\infty = (J + I_1) : \mathbf{p}^\infty$. $\qquad\square$

10

By following the technique of the proof, we can replace $\mathbf{p}_1$ by the product of a much smaller number of $p_{+u_2\cdots u_n}$. In fact, we need only take the linear forms $p_{+u_2\cdots u_r 11\cdots 1}$. Hence, by induction, $\mathbf{p}$ can be replaced by a much smaller product of linear forms. This observation proved to be crucial for computing some of the tough primary decompositions in Section 6.

As a corollary we derive an algebraic proof of the Factorization Theorem.

Proof of Theorem 3: We use induction on the number of nodes to show that every point in $V_{\geq}(I_{\mathrm{local}(G)} + \langle p - 1\rangle)$ also lies in $\mathrm{image}(\phi_{\geq})$. Such a point is a homomorphism $\tau : \mathbb{R}[D] \to \mathbb{R}$ with the property that τ is zero on $I_{\mathrm{local}(G)}$, and its values on the indeterminates $p_{u_1\cdots u_n}$ are non-negative and sum to 1. The map τ can be extended to a homomorphism $\tau' : \mathbb{R}[Q \cup A] \to \mathbb{R}$ as follows. We first set $\tau'(q_{i_2\cdots i_n}) = \tau(p_{+i_2\cdots i_n})$. If that real number is positive then we set $\tau'(a_{i_1\cdots i_r}) = \tau(p_{i_1 i_2\cdots i_n})/\tau(p_{+i_2\cdots i_n})$, and otherwise we set $\tau'(a_{i_1\cdots i_r}) = 0$. Our non-negativity hypothesis implies that τ coincides with the composition of τ' and Ψ, i.e., the point τ is the image of τ' under the induced map $\mathbb{R}^{A\cup Q} \to \mathbb{R}^D$. The conclusion now follows by induction. $\square$

We close this section by presenting our solution to [23, Problem 8.11].

Proposition 9. *There exists a Bayesian network G on five binary random variables such that the local Markov ideal $I_{\mathrm{local}}(G)$ is not radical.*

Proof. Let G be the complete bipartite network $K_{2,3}$ with nodes $\{1,5\}$ and $\{2,3,4\}$ and directed edges $(5,2),(5,3),(5,4),(2,1),(3,1),(4,1)$. Then

$$\mathrm{local}(G) = \big\{\, 1\perp\!\!\!\perp 5\,|\,\{2,3,4\},\ 2\perp\!\!\!\perp\{3,4\}\,|\,5,\ 3\perp\!\!\!\perp\{2,4\}\,|\,5,\ 4\perp\!\!\!\perp\{2,3\}\,|\,5 \,\big\}.$$

The polynomial ring $\mathbb{R}[E]$ has 32 indeterminates $p_{11111}, p_{11112}, \ldots, p_{22222}$. The ideal $I_{\mathrm{local}}(G)$ is minimally generated by eight binomial quadrics

$$p_{1u_2u_3u_41}\cdot p_{2u_2u_3u_42} - p_{1u_2u_3u_42}\cdot p_{2u_2u_3u_41}, \qquad u_2,u_3,u_4 \in \{1,2\},$$

and eighteen non-binomial quadrics

$$p_{+122u_5}\cdot p_{+221u_5} - p_{+121u_5}\cdot p_{+222u_5},\ p_{+212u_5}\cdot p_{+221u_5} - p_{+211u_5}\cdot p_{+222u_5},$$
$$p_{+112u_5}\cdot p_{+221u_5} - p_{+111u_5}\cdot p_{+222u_5},\ p_{+122u_5}\cdot p_{+212u_5} - p_{+112u_5}\cdot p_{+222u_5},$$
$$p_{+121u_5}\cdot p_{+212u_5} - p_{+111u_5}\cdot p_{+222u_5},\ p_{+122u_5}\cdot p_{+211u_5} - p_{+111u_5}\cdot p_{+222u_5},$$
$$p_{+112u_5}\cdot p_{+211u_5} - p_{+111u_5}\cdot p_{+212u_5},\ p_{+121u_5}\cdot p_{+211u_5} - p_{+111u_5}\cdot p_{+221u_5},$$
$$p_{+112u_5}\cdot p_{+121u_5} - p_{+111u_5}\cdot p_{+122u_5}, \qquad u_5 \in \{1,2\}.$$

These nine equations (for fixed value of u_5) define the Segre embedding of $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ in $\mathbb{P}^7$, as in [23, eqn. (8.6), page 103]. Consider the polynomial

$$f = p_{+1112}p_{+2222}(p_{12221}p_{12212}p_{12122}p_{12111} - p_{12112}p_{12121}p_{12211}p_{12222}).$$

By computing a Gröbner basis, it can be checked that f^2 lies in $I_{\text{local}(G)}$ but f does not lie in $I_{\text{local}(G)}$. Hence $I_{\text{local}(G)}$ is not a radical ideal. The primary decomposition of this ideal will be described in Example 18. $\qquad\square$

5 Networks on Four Random Variables

In this section we present the algebraic classification of all Bayesian networks on four random variables. In the binary case we have the following result.

Theorem 10. *The local and global Markov ideals of all Bayesian networks on four binary variables are radical. The hypothesis "binary" is essential.*

Thus the solution [23, Problem 8.11] is affirmative for networks on four binary nodes. Proposition 9 shows that the hypothesis "four" is essential. Theorem 10 is proved by exhaustive computations in `Macaulay2`. We summarize the results in Table 1. Each row represents one network G on four binary random variables along with some information about its two ideals

$$I_{\text{local}(G)} \ \subseteq \ I_{\text{global}(G)} \ \subseteq \ \mathbb{R}[p_{1111}, p_{1112}, \ldots, p_{2221}, p_{2222}].$$

Here G is represented by the list of sets of children $(\text{ch}(1), \text{ch}(2), \text{ch}(3), \text{ch}(4))$. The information given in the second column corresponds to the *codimension, degree, and number of minimal generators* of the ideal $I_{\text{local}(G)}$. For example, the network in the fourth row has four directed edges $(2,1), (3,1), (4,1)$ and $(4,2)$. Here $I_{\text{local}(G)} = I_{\text{global}(G)} = \ker(\Phi)$. This prime has codimension 3, degree 4 and is generated by the six 2×2-minors of the 2×4-matrix

$$\begin{pmatrix} p_{+111} & p_{+112} & p_{+211} & p_{+212} \\ p_{+121} & p_{+122} & p_{+221} & p_{+222} \end{pmatrix}.$$

Of the 30 local Markov ideals in Table 1 all but six are prime. The remaining six ideals are all radical, and the number of their minimal primes is listed. Hence all local Markov ideals are radical. The last column corresponds to the ideal $I_{\text{global}(G)}$. This ideal is equal to the distinguished component for all but two networks, namely `15` and `17`. For these two networks we have $I_{\text{local}(G)} = I_{\text{global}(G)}$. This proves the first assertion of Theorem 10.

The main point of this section is the second sentence in Theorem 10. Embedded components can appear when the number of levels increases. In the next theorem we let d_1, d_2, d_3 and d_4 be arbitrary positive integers.

Theorem 11. *Of the 30 local Markov ideals on four random variables, 22 are always prime, five are not prime but always radical (numbers `10`,`11`,`16`, `18`,`26` in Table 1) and three are not radical (numbers `15`,`17`,`21` in Table 1).*

Index	Information	Network	Local	Global
1	1, 2, 1	$\{\}, \{1\}, \{1, 2\}, \{1, 2\}$	prime	
2	2, 4, 2	$\{\}, \{1\}, \{1\}, \{1, 2, 3\}$	prime	
3	2, 4, 2	$\{\}, \{1\}, \{1, 2\}, \{1, 3\}$	prime	
4	3, 4, 6	$\{\}, \{1\}, \{1\}, \{1, 2\}$	prime	
5	4, 6, 9	$\{\}, \{1\}, \{1\}, \{1\}$	prime	
6	4, 16, 4	$\{\}, \{\}, \{1, 2\}, \{1, 2, 3\}$	prime	
7	4, 16, 4	$\{\}, \{1\}, \{1, 2\}, \{2, 3\}$	prime	
8	4, 16, 4	$\{\}, \{1\}, \{2\}, \{1, 2, 3\}$	prime	
9	5, 32, 5	$\{\}, \{\}, \{1, 2\}, \{1, 2\}$	prime	
10	5, 32, 5	$\{\}, \{1\}, \{1, 2\}, \{2\}$	prime	
11	6, 8, 10	$\{\}, \{1\}, \{1\}, \{2\}$	radical, 5 comp.	prime
12	6, 16, 12	$\{\}, \{\}, \{1\}, \{1, 2, 3\}$	prime	
13	6, 16, 12	$\{\}, \{\}, \{1, 2\}, \{2, 3\}$	prime	
14	6, 16, 12	$\{\}, \{1\}, \{2\}, \{2, 3\}$	prime	
15	6, 64, 6	$\{\}, \{1\}, \{1\}, \{2, 3\}$	radical, 5 comp.	radical
16	6, 64, 6	$\{\}, \{1\}, \{1, 2\}, \{3\}$	radical, 9 comp.	prime
17	6, 64, 6	$\{\}, \{1\}, \{2\}, \{1, 3\}$	radical, 5 comp.	radical
18	7, 8, 14	$\{\}, \{1\}, \{2\}, \{3\}$	radical, 3 comp.	prime
19	7, 8, 28	$\{\}, \{\}, \{1\}, \{1, 3\}$	prime	
20	7, 24, 16	$\{\}, \{\}, \{1\}, \{1, 2\}$	prime	
21	7, 32, 13	$\{\}, \{1\}, \{2\}, \{2\}$	prime	
22	8, 14, 31	$\{\}, \{\}, \{1\}, \{1\}$	prime	
23	8, 34, 20	$\{\}, \{\}, \{1\}, \{2, 3\}$	prime	
24	8, 36, 18	$\{\}, \{\}, \{\}, \{1, 2, 3\}$	prime	
25	8, 36, 18	$\{\}, \{\}, \{1, 2\}, \{3\}$	prime	
26	9, 20, 27	$\{\}, \{\}, \{1\}, \{2\}$	radical, 5 comp.	prime
27	9, 24, 34	$\{\}, \{\}, \{\}, \{1, 2\}$	prime	
28	9, 24, 34	$\{\}, \{\}, \{1\}, \{3\}$	prime	
29	10, 20, 46	$\{\}, \{\}, \{\}, \{1\}$	prime	
30	11, 24, 55	$\{\}, \{\}, \{\}, \{\}$	prime	

Table 1: All Bayesian Networks on Four Binary Random Variables

Proof. We prove this theorem by an exhaustive case analysis of all thirty networks. In most cases, the ideal $I_{\text{local}(G)}$ can be made binomial by a suitable coordinate change, just like in the proof of Theorem 1. In fact, let us start with a non-trivial case which is immediately taken care of by

Theorem 1.

The network 16*:* Here we have $\mathrm{local}(G) = \{1 \perp\!\!\!\perp 4 \,|\, \{2,3\}, \; 2 \perp\!\!\!\perp 4 \,|\, 3\}$. For fixed value of the third node we get the model $\{1 \perp\!\!\!\perp 4 \,|\, 2, \; 4 \perp\!\!\!\perp 2\}$ whose ideal was shown to be radical in Theorem 1. Hence $I_{\mathrm{local}(G)}$ is the ideal generated by d_3 copies of this radical ideal in disjoint sets of variables. We conclude that $I_{\mathrm{local}(G)}$ is radical and has $(2^{d_2} - 1)^{d_3}$ minimal primes.

The networks 1, 2, 3, 4, 6, 7, 8, 12, 13, 14*:* In each of these ten cases, the ideal $I_{\mathrm{local}(G)}$ is generated by quadratic polynomials corresponding to a single conditional independence statement. This observation implies that $I_{\mathrm{local}(G)}$ is a prime ideal, by [23, Lemma 8.2].

The network 5*:* Here $\mathrm{local}(G)$ specifies the model of complete independence for the random variables X_2, X_3 and X_4. This means that $I_{\mathrm{local}(G)}$ is the ideal of a Segre variety, which is prime and has a quadratic Gröbner basis.

The networks 24 *and* 25*:* Each of these two networks describes the join of d_4 and d_3 Segre varieties. The same reasoning as in case 5 applies.

The network 23*:* Observe that $I_{\mathrm{local}(G)} = I_{\mathrm{global}(G)} = I_{1 \perp\!\!\!\perp \{2,4\} | 3} + I_{2 \perp\!\!\!\perp \{1,3\} | 4}$. Since G is a directed tree, Theorem 6 implies that $I_{\mathrm{global}(G)}$ coincides with the distinguished prime ideal $\ker(\Phi)$. Therefore, $I_{\mathrm{local}(G)}$ is always prime.

The networks 19, 22, 27, 28, 29, 30*:* Each of these six networks has an isolated vertex. This means that $I_{\mathrm{local}(G)}$ is the ideal of the Segre embedding of the product of two smaller varieties namely, the projective space $\mathbb{P}^{d_i - 1}$ corresponding to the isolated vertex i and the scheme specified by the local ideal of the remaining network on three nodes. The latter ideal is prime and has a quadratic Gröbner basis, by Proposition 5, and hence so is $I_{\mathrm{local}(G)}$.

The network 20*:* The ideal $I_{\mathrm{local}(G)}$ is binomial in the coordinates p_{ijkl} with $i \in \{+, 2, \ldots, d_1\}$. Generators are $p_{i_1 j_2 kl} p_{i_2 j_1 kl} - p_{i_1 j_1 kl} p_{i_2 j_2 kl}$, $p_{i_1 j_2 k_1 l} p_{i_2 j_1 k_2 l} - p_{i_1 j_1 k_1 l} p_{i_2 j_2 k_2 l}$, and $p_{+ j_1 k_2 l_1} p_{+ j_2 k_1 l_2} - p_{+ j_1 k_1 l_1} p_{+ j_2 k_2 l_2}$. The S-pairs within each group reduce to zero by the Gröbner basis property of the 2×2-minors of a generic matrix. It can be checked easily that the crosswise reverse lexicographic S-pairs also reduce to zero. We conclude that the given set of irreducible quadrics is a reverse lexicographic Gröbner basis. In view of [22, Lemma 12.1], the lowest variable is not a zero-divisor, and hence by symmetry none of the variables p_{ijkl} is zero-divisor. It now follows from equation (6) in Theorem 8 that $I_{\mathrm{local}(G)}$ coincides with the prime ideal $\ker(\Phi)$.

The network 9*:* The ideal $I_{\mathrm{local}(G)}$ is generated by the quadratic polynomials $p_{i_1 j_2 kl} p_{i_2 j_1 kl} - p_{i_1 j_1 kl} p_{i_2 j_2 kl}$, $p_{++k_1 l_2} p_{++k_2 l_1} - p_{++k_1 l_1} p_{++k_2 l_2}$. These generators form a Gröbner basis in the reverse lexicographic order. Indeed,

assuming that $i_1 < i_2$, $j_1 < j_2$, $k_1 < k_2$, $l_1 < l_2$, the leading terms are $p_{i_1j_2kl}p_{i_2j_1kl}$ and $p_{11k_1l_2}p_{11k_2l_1}$. Hence no leading term from the first group of quadrics shares a variable with a leading term from the second group. Hence the crosswise S-pairs reduce to zero by [4, Prop. 4, §2.9]. The S-pairs within each group also reduce to zero by the Gröbner basis property of the 2×2-minors of a generic matrix. Hence the generators are a Gröbner basis. Since the leading terms are square-free, we see that the ideal is radical. An argument similar to the previous case shows that $I_{\mathrm{local}(G)}$ is prime.

The network 18: Here G is a directed chain of length four. We claim that $I_{\mathrm{local}(G)}$ is the irredundant intersection of $2^{d_2}-1$ primes, and it has a Gröbner basis consisting of square-free binomials of degree two, three and four. We give an outline of the proof. We first turn $I_{\mathrm{local}(G)}$ into a binomial ideal by taking the coordinates to be p_{ijkl} with $i \in \{+, 2, 3, \ldots, d_1\}$. The minimal primes are indexed by proper subsets of $[d_2]$. For each such subset σ we introduce the monomial prime $M_\sigma = \langle p_{+jkl} : j \in \sigma, k \in [d_3], l \in [d_4]\rangle$ and the complementary monomial $m_\sigma = \prod_{j\in[d_2]\backslash\sigma} \prod_{k\in[d_3]} \prod_{l\in[d_4]} p_{+jkl}$, and we define the ideal $P_\sigma = \big((I_{\mathrm{local}(G)} + M_\sigma) : m_\sigma^\infty\big)$. These ideals are prime, and the union of their varieties is irredundant and equals the variety of $I_{\mathrm{local}(G)}$. Using Buchberger's S-pair criterion, we check that the following four types of square-free binomials are a Gröbner basis:

- the generators $p_{i_1jk_1l_1}p_{i_2jk_2l_2} - p_{i_1jk_2l_2}p_{i_1jk_2l_2}$ encoding $1\perp\!\!\!\perp\{3,4\}\,|\,2$,

- the generators $p_{+j_1kl_1}p_{+j_2kl_2} - p_{+j_1kl_2}p_{+j_2kl_1}$ encoding $2\perp\!\!\!\perp 4\,|\,3$,

- the cubics $(p_{+j_1kl_1}p_{ij_2kl_2} - p_{+j_1kl_2}p_{ij_2kl_1}) \cdot p_{+j_2k_3l_3}$,

- the quartics $(p_{i_1j_1kl_1}p_{i_2j_2kl_2} - p_{i_1j_1kl_2}p_{i_2j_2kl_1}) \cdot p_{+j_1l_3k_3} \cdot p_{+j_2l_4k_4}$.

The network 10: The ideal $I_{\mathrm{local}(G)}$ is generated by $p_{i_1jkl_2}p_{i_2jkl_1} - p_{i_1jkl_1}p_{i_2jkl_2}$ and $p_{++k_1l_2}p_{++k_2l_1} - p_{++k_1l_1}p_{++k_2l_2}$. In general, this ideal is not prime, but it is always radical. If $d_4 = 2$ then the ideal is always prime, If $d_4 > 2$, $I_{\mathrm{local}(G)}$ is the intersection of the distingushed component and 2^{d_3-1} prime ideals indexed by all proper subsets $\sigma \subset [d_3]$ as in the previous network.

The network 11: Here, $\mathrm{local}(G) = \big\{1\perp\!\!\!\perp 4 \mid \{2,3\}, 2\perp\!\!\!\perp 3 \mid 4, 3\perp\!\!\!\perp\{2,4\}\big\}$. The ideal $I_{\mathrm{local}(G)}$ is binomial in the coordinates p_{ijkl} with $i \in \{+, 2, \ldots, d_1\}$. It is generated by the binomials $p_{i_1jkl_1}p_{i_2jkl_2} - p_{i_1jkl_2}p_{i_2jkl_1}$, $p_{+j_1k_1l_1}p_{+j_2k_2l_2} - p_{+j_1k_2l_1}p_{+j_2k_1l_2}$ encoding the first and third independent statements. The minimal primes are indexed by pairs of proper subsets of $[d_2]$ and $[d_3]$. For each such pair of subsets (σ, τ) we introduce the monomial prime $M_{(\sigma,\tau)} = $

$\langle p_{+jkl} : j \in \sigma, k \in \tau, l \in [d_4] \rangle$ and the complementary monomial $m_{(\sigma,\tau)} = \prod_{j \in [d_2] \setminus \sigma} \prod_{k \in [d_3] \setminus \tau} \prod_{l \in [d_4]} p_{+jkl}$, and we define the ideal $P_{(\sigma,\tau)} = \left((I_{\text{local}(G)} + M_{(\sigma,\tau)}) : m_{(\sigma,\tau)}^{\infty} \right)$. These ideals are prime, and the union of their varieties equals the variety of $I_{\text{local}(G)}$. Moreover, the ideal $I_{\text{local}(G)}$ is equal to the intersection of the minimal primes which are indexed by the following pairs: For each proper $\tau \subset [d_3]$ the pair $(\emptyset, \tau)$, and for each nonempty proper $\sigma \subset [d_2]$ the pairs (σ, τ) where $\tau \subset [d_3]$ is any subset of cardinality at most $d_3 - 2$. In particular, for $d_2 = d_3 = 3$, and arbitrary d_1, d_4, the ideal $I_{\text{local}(G)}$ has 31 prime components. For $d_2 = 2, d_3 = 4$, $I_{\text{local}(G)}$ has 37 prime components, and for $d_2 = 4, d_3 = 2$, $I_{\text{local}(G)}$ has 17 prime components.

The network 26: The ideal $I_{\text{local}(G)}$ is a radical ideal. The minimal primes are indexed by all pairs of proper subsets of $[d_3]$ and $[d_4]$. For each such pair (σ, τ) we introduce the monomial primes $M_\sigma = \langle p_{+jkl} : k \in \sigma, j \in [d_2], l \in [d_4] \rangle$, $M_\tau = \langle p_{i+kl} : l \in \tau, i \in [d_1], k \in [d_3] \rangle$, and $M_{(\sigma,\tau)} = M_\sigma + M_\tau$. Just as before, we introduce the complementary monomial $m_{(\sigma,\tau)}$, and the ideal $P_{(\sigma,\tau)} = \left((I_{\text{local}(G)} + M_{(\sigma,\tau)}) : m_{(\sigma,\tau)}^{\infty} \right)$. The ideal $I_{\text{local}(G)}$ is equal to the intersection of all these prime ideals.

The network 21: Here, $\text{local}(G) = \left\{ 1 \perp\!\!\!\perp \{3,4\} \,|\, 2, \ 3 \perp\!\!\!\perp 4 \right\}$. The ideal $I_{\text{local}(G)}$ is generated by the binomials $p_{i_1 j k_2 l_2} p_{i_2 j k_1 l_1} - p_{i_1 j k_1 l_1} p_{i_2 j k_2 l_2}$, and the polynomials $p_{++k_1 l_2} p_{++k_2 l_1} - p_{++k_1 l_1} p_{++k_2 l_2}$. This ideal is not radical, in general. The first counterexample occurs for the case $d_1 = d_2 = d_3 = 2$ and $d_4 = 3$. Here $I_{\text{local}(G)}$ is generated by 33 quadratic polynomials in 24 unknowns. The degree reverse lexicographic Gröbner basis of this ideal consists of 123 polynomials of degree up to 8. In this case, $I_{\text{local}(G)}$ is the intersection of the distinguished component and the P-primary ideal $Q = I_{1 \perp\!\!\!\perp \{3,4\} \,|\, 2} + P^2$, where P is the prime ideal generated by the 12 linear forms p_{+jkl}.

The networks 15 *and* 17: Here, after relabeling network 17, $\text{local}(G) = \left\{ 1 \perp\!\!\!\perp 4 \,|\, \{2,3\}, \ 2 \perp\!\!\!\perp 3 \,|\, 4 \right\}$. The ideal $I_{\text{local}(G)}$ is binomial in the coordinates p_{ijkl} with $i \in \{+, 2, \ldots, d_1\}$. It is generated by the binomials $p_{i_1 j k l_1} p_{i_2 j k l_2} - p_{i_1 j k l_2} p_{i_2 j k l_1}$, $p_{+j_1 k_1 l} p_{+j_2 k_2 l} - p_{+j_1 k_2 l} p_{+j_2 k_1 l}$. This ideal is not radical, in general. The first counterexample occurs for the case $d_1 = 2$ and $d_2 = d_3 = d_4 = 3$. Here $I_{\text{local}(G)}$ is generated by 54 quadratic binomials in 54 unknowns. The reverse lexicographic Gröbner basis consists of 13,038 binomials of degree up to 14. One of the elements in the Gröbner basis is

$$p_{+111} p_{+223} (p_{+331})^2 \cdot \left(p_{2122} p_{2133} p_{2323} p_{2332} - p_{2333} p_{2322} p_{2132} p_{2123} \right).$$

Removing the square from the third factor, we obtain a polynomial f of degree 7 such that that $f \notin I$ but $f^2 \in I$. This proves that I is not radical.

The number of minimal primes of $I_{\text{local}(G)}$ is equal to $2^{d_2} + 2^{d_3} - 3$. $\qquad\square$

In the 22 cases where I_{local} is prime, it follows from Theorem 8 that the global Markov ideal I_{global} is prime as well. Among the remaining cases, we have $I_{\text{local}(G)} = I_{\text{global}(G)}$ for networks $10, 15, 17, 21$, and we have $I_{\text{local}} \neq I_{\text{global}} = \ker(\Phi)$ for networks $11, 16, 18, 26$. This discussion implies:

Corollary 12. *Of the 30 global Markov ideals on four random variables, 26 are always prime, one is not prime but always radical (number 10 in Table 1) and three are not radical (numbers $15, 17, 21$ in Table 1).*

It is instructive to examine the distinguished prime ideal $P = \ker(\Phi)$ in the last case $15, 17$. Assume for simplicity that $d_1 = 2$ but d_2, d_3 and d_4 are arbitrary positive integers. We rename the unknowns $x_{jkl} = p_{2jkl}$ and $y_{jkl} = p_{+jkl}$. Then we can take Φ to be the following monomial map:

$$\mathbb{R}[x_{jkl}, y_{jkl}] \to \mathbb{R}[u_{jk}, v_{jl}, w_{kl}], \ x_{jkl} \mapsto u_{jk}v_{jl}w_{kl}, \ y_{jkl} \mapsto v_{jl}w_{kl}, \qquad (12)$$

For example, for $d_2 = d_3 = 3$ and $d_4 = 2$, the ideal $P = \ker(\Phi)$ has 361 minimal generators, of degrees ranging from two to seven. One generator is

$$x_{111}x_{132}x_{222}x_{312}x_{321}y_{221}y_{331} - x_{112}x_{131}x_{221}x_{311}x_{322}y_{232}y_{321}.$$

Among the 361 minimal generators, there are precisely 15 which do not contain any variable y_{ijk}, namely, there are nine quartics and six sextics like

$$x_{112}x_{121}x_{211}x_{232}x_{322}x_{331} - x_{111}x_{122}x_{212}x_{231}x_{321}x_{332}.$$

These 15 generators form the Markov basis for the $3 \times 3 \times 2$-tables in the no-three-way interaction model. See [22, Corollary 14.12] for a discussion.

The ideal for the no-three-way interaction model of $d_2 \times d_3 \times d_4$-tables always coincides with the elimination ideal $P \cap \mathbb{R}[x_{ijk}]$ and, moreover, every generating set of P contains a generating set for $P \cap \mathbb{R}[x_{ijk}]$. In view of [22, Proposition 14.14], this shows that the maximal degree among minimal generators of P exceeds any bound as d_2, d_3, d_4 increases. In practical terms, it is hard to compute these generators even for $d_2 = d_3 = d_4 = 4$. We refer to the web page `http://math.berkeley.edu/~seths/ccachallenge.html`.

6   Networks on Five Binary Random Variables

In this section we discuss the global Markov ideals of all Bayesian networks on five binary random variables. In each case we computed the primary

decomposition. In general, the built-in primary decomposition algorithms in current computer algebra systems cannot compute the primary decompositions of most of these ideals. In the Appendix, we outline some techniques that allowed us to compute these decompositions. The primary decompositions of the local Markov ideals of these networks could also be computed, but they have less regular structure and are in general more complicated.

There are 301 distinct non-complete networks on five random variables, up to isomorphism of directed graphs. We have placed descriptions of these networks and of the primary decompositions of their global Markov ideals on the website `http://math.cornell.edu/~mike/bayes/global5.html`. In this section, we refer to the graphs as $G_0, G_1, \ldots, G_{300}$, the indices matching the information on the website. We summarize our results in a theorem.

Theorem 13. *Of the* 301 *global Markov ideals on five binary random variables,* 220 *are prime,* 68 *are radical but not prime, and* 13 *are not radical.*

Proof. The proof is via direct computation with each of these ideals in `Macaulay2`. Some of these require little or no computation: if G is a directed forest, or if there is only one independence statement, then the ideal is prime. Others require substantial computation and some ingenuity to find the primary decomposition. Results are posted at the website cited above.

To prove primality, it suffices to compute the ideal quotient of $I = I_{\text{global}(G)}$ with respect to a small subset of the $p_{+++u_r\cdots u_n}$. Alternatively, one may birationally project I by eliminating variables, as in Proposition 23. In either case, if a zero divisor x is found, the ideal is not prime. If some ideal quotient satisfies $(I : x^2) \neq (I : x)$, then I is not radical. □

The numbers of prime components of the 288 radical global Markov ideals range from 1 to 39. The distribution is given in the following table:

# of components	1	3	5	7	17	25	29	33	39
# of ideals	220	8	41	3	9	1	2	3	1

Theorem 14. *Conjecture 7 is true for Bayesian networks G on five binary random variables. In each of the* 301 *cases, the distinguished prime ideal* $\ker(\Phi)$ *is generated by homogeneous polynomials of degree at most eight.*

Proof. We compute the distinguished component from $I_{\text{global}(G)}$ by saturation, and we check the result by using the techniques in the Appendix.

The computation of the distinguished component of the 81 non-prime examples yields that 64 of these ideals are generated in degrees ≤ 4, twelve are generated in degrees ≤ 6, and five are generated in degrees ≤ 8. $\square$

Theorem 8 says that we can decide primality or find the distinguished component of $I_{\mathrm{global}(G)}$ by inverting each of the $p_{+++u_i \cdots u_n}$. With some care, it is possible to reduce this to a smaller set. Still, the following is unexpected.

Proposition 15. *For all but two networks on five binary random variables, p_{+1111} is a non-zero divisor on $I = I_{\mathrm{global}(G)}$ if and only if I is prime. In all but these two examples, I is radical if and only if $(I : p_{+1111}^2) = (I : p_{+1111})$.*

Proof. The networks which do not satisfy the given property are $G_{201} = \big(\{\}, \{1\}, \{1,2\}, \{1,2\}, \{3,4\}\big)$ and $G_{214} = \big(\{\}, \{1\}, \{1,2\}, \{3\}, \{1,2,4\}\big)$. After permuting the nodes 4, 5, both the local and global independence statements of G_{214} are the same as those for G_{201}. The global independence statements for G_{201} are $\big\{\{1,2\} \perp\!\!\!\perp 5 \,|\, \{3,4\}, \; 3 \perp\!\!\!\perp 4 \,|\, 5\big\}$. The primary decomposition for the radical ideal $I = I_{\mathrm{global}(G_{201})}$ is

$$ I \;=\; \ker(\Phi) \cap (I + P_{++1\bullet\bullet}) \cap (I + P_{++2\bullet\bullet}) \cap (I + P_{++\bullet1\bullet}) \cap (I + P_{++\bullet2\bullet}), $$

where $\ker(\Phi)$ is the distinguished prime component,

$$ P_{++1\bullet\bullet} \quad = \quad \langle p_{++111}, p_{++112}, p_{++121}, p_{++122} \rangle, $$

and the other three components are defined in an analogous manner. Therefore, p_{+1111} is a non-zero divisor modulo I. By examining all 81 non-prime ideals, we see that all except these two have a minimal prime containing p_{+1111}. The final statement also follows from direct computation. $\square$

We have searched for conditions on the network which would characterize under what conditions the global Markov ideal is prime, or fails to be prime. Theorem 6 states that if the network is a directed forest, then the global Markov ideal is prime. Two possible conditions, the first for primality, and the second for non-primality, are close, but not quite right. We present them, with their counterexamples, in the following two propositions.

Proposition 16. *There is a unique network G on 5 binary nodes whose underlying undirected graph is a tree, but $I_{\mathrm{global}(G)}$ is not radical. Every other network whose underlying graph is a tree has prime global Markov ideal.*

Proof. The unique network is $G_{23} = (\{\}, \{1\}, \{2\}, \{2\}, \{2\})$. Its local and global Markov independent statements coincide and are equal to

$$\{1 \perp\!\!\!\perp \{3, 4, 5\} \,|\, 2, \ 3 \perp\!\!\!\perp \{4, 5\}, \ 4 \perp\!\!\!\perp \{3, 5\}, \ 5 \perp\!\!\!\perp \{3, 4\}\}.$$

Computation using `Macaulay2` reveals

$$I_{\text{global}(G_{23})} \quad = \quad \ker(\Phi) \cap (I_{\text{global}(G_{23})} + (P_{+\bullet\bullet\bullet\bullet})^2),$$

where $P_{+\bullet\bullet\bullet\bullet}$ is the ideal generated by the 16 linear forms $p_{+u_2 u_3 u_4 u_5}$. Inspecting the 81 non-prime ideals shows that G_{23} is the only example. $\square$

We say that the network G *has an induced r-cycle* if there is an induced subgraph H of G with r vertices which consists of two disjoint directed paths which share the same start point and end point.

Proposition 17. *Of the* 301 *networks on five nodes,* 70 *have an induced 4-cycle or 5-cycle. For exactly two of these, the ideal* $I_{\text{global}(G)}$ *is prime.*

Proof. Once again, this follows by examination of the 301 cases. The graphs which have an induced 4-cycle but whose global Markov ideal is prime are

$$G_{265} = \{\{\}, \{1\}, \{1, 2\}, \{1, 2\}, \{2, 3, 4\}\}$$
$$\text{and} \quad G_{269} = \{\{\}, \{1\}, \{1, 2\}, \{2, 3\}, \{1, 2, 4\}\}.$$

Removing node 2 results in a 4-cycle. The local and global Markov statements are all the same up to relabeling: $\{1 \perp\!\!\!\perp 5 \,|\, \{2, 3, 4\}, \ 3 \perp\!\!\!\perp 4 \,|\, 5\}$. $\square$

There are four graphs with three induced 4-cycles, namely, G_{138}, G_{139}, G_{150}, G_{157}. The first two graphs give rise to the same (global or local) independence statements, and similarly for the last two. The ideal $I_{\text{global}(G_{138})}$ has the most components of any of the 301 ideals considered in this section.

Example 18. The network $G_{138} = (\{\}, \{1\}, \{1\}, \{1\}, \{2, 3, 4\})$ is isomorphic to the one in Proposition 9. Its ideal $I_{\text{global}(G_{138})}$ has 207 minimal primes, and 37 embedded primes. Each of the 207 minimal primary components are prime. We will describe the structure of these components.

Let $F_{i_1 i_2 i_3} = \det \begin{pmatrix} p_{+i_1 i_2 i_3 1} & p_{+i_1 i_2 i_3 2} \\ p_{2 i_1 i_2 i_3 1} & p_{2 i_1 i_2 i_3 2} \end{pmatrix}$. Let J_i be the ideal generated by the 2×2 minors located in the first two rows or columns of the matrix

$$\begin{pmatrix} p_{+111i} & p_{+112i} & p_{+211i} & p_{+212i} \\ p_{+121i} & p_{+122i} & p_{+221i} & p_{+222i} \\ p_{+211i} & p_{+212i} & * & * \\ p_{+221i} & p_{+222i} & * & * \end{pmatrix}.$$

We have

$$I \quad := \quad I_{\text{global}(G_{138})} \quad = \quad J_1 + J_2 + \langle F_{111}, F_{112}, \ldots, F_{222} \rangle.$$

Each J_i is minimally generated by 9 quadrics, so that I is minimally generated by 26 quadrics. Each J_i is prime of codimension 4, and so $J_1 + J_2$ is prime of codimension 8. Since there are only 8 more quadrics, Krull's principal ideal theorem tells us that all minimal primes have codimension at most 16, which is also the codimension of the distinguished component. Note that I is a binomial ideal in the unknowns $p_{+u_2u_3u_4u_5}$ and $p_{2u_2u_3u_4u_5}$.

# primes	codim	degree	faces
6	14	48	(f, f), f a facet
12	14	4	(e, e), e an edge
24	16	15	(f_1, f_2), $f_1 \cap f_2$ is an edge
48	16	4	(f, e), $f \cap e$ is a point
12	16	1	(e_1, e_2), 2 antipodal edges
48	16	1	(e_1, e_2), 2 non-parallel disjoint edges
48	16	1	(e, p), point p on the edge antipodal to e
8	16	1	(p_1, p_2), antipodal points
1	16	2316	distinguished component

Table 2: All 207 minimal primes of the ideal $I_{\text{global}(G_{138})}$

Let Δ be the unit cube, with vertices $(1, 1, 1), (1, 1, 2), \ldots, (2, 2, 2)$. If $\sigma \subset \Delta$ is a face, define $P_{\sigma,i}$ to be the monomial prime generated by $\{p_{+vi} \mid v \notin \sigma\}$, for $i \in \{1, 2\}$. If P is a minimal prime of I, which is not the distinguished component, then P must contain some $p_{+v_1v_2v_31}$, and also contain some $p_{+u_1u_2u_32}$. Therefore, there are faces σ_1 and σ_2 of Δ such that P contains $P_{\sigma_1,1} + P_{\sigma_2,2}$, and does not contain any other elements p_{+vi}. Let $m_{\sigma_1\sigma_2}$ be the product of all of the p_{+vi} such that $v \in \sigma_i$ for $i = 1, 2$. It turns out that every minimal prime ideal of I has the form

$$P_{\sigma_1,\sigma_2} \quad := \quad \left((I + P_{\sigma_1,1} + P_{\sigma_2,2}) : m_{\sigma_1\sigma_2}^\infty \right)$$

for some pair σ_1, σ_2 of proper faces of the cube Δ. However, not all pairs of faces correspond to minimal primes. There are 27 proper faces of the cube, and so there are $27^2 = 729$ possible minimal primes. Only 206 of these occur. The list of minimal primes is given in Table 2. □

Bayesian networks give rise to very interesting (new and old) constructions in algebraic geometry. In the next section, we shall encounter secant

varieties. Here, we offer a generalization of Example 18 to arbitrary toric varieties. Let $I_A \subset \mathbb{R}[z_1, \ldots, z_n]$ be any *toric ideal*, specified as in [22] by a point configuration $A = \{a_1, \ldots, a_n\} \subset \mathbb{Z}^d$. Let Δ be the convex hull of A in $\mathbb{R}^d$. We define the *double join* of the toric ideal I_A to be the new ideal

$$I_A(x) + I_A(y) + \langle F_1, \ldots, F_n \rangle \subset \mathbb{R}[x_1, \ldots, x_n, y_1, \ldots, y_n, a_1, \ldots a_n, b_1, \ldots, b_n]$$

where $F_i = \det \begin{pmatrix} x_i & a_i \\ y_i & b_i \end{pmatrix}$, and $I_A(x)$ and $I_A(y)$ are generated by copies of I_A in $\mathbb{R}[x_1, \ldots, x_n]$ and $\mathbb{R}[y_1, \ldots, y_n]$ respectively. The ideal I in Example 18 is the double join of the Segre variety $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \subset \mathbb{P}^7$, which is the toric variety whose polytope Δ is the 3-cube. In general, the minimal primes of the double join of I_A are indexed by pairs of faces of the polytope Δ. We believe that this construction deserves the attention of algebraic geometers.

7 Hidden Variables and Higher Secant Varieties

Let G be a Bayesian network on n discrete random variables and let $P_G = \ker(\Phi)$ be its homogeneous prime ideal in the polynomial ring $\mathbb{R}[D]$, whose indeterminates $p_{i_1 i_2 \cdots i_n}$ represent probabilities of events $(i_1, i_2, \ldots, i_n) \in D$. We now consider the situation when some of the random variables are hidden. After relabeling we may assume that the variables corresponding to the nodes $r+1, \ldots, n$ are hidden, while the random variables corresponding to the nodes $1, \ldots, r$ are observed. Thus the *observable probabilities* are

$$p_{i_1 i_2 \cdots i_r ++\cdots+} = \sum_{j_{r+1} \in [d_{r+1}]} \sum_{j_{r+2} \in [d_{r+2}]} \cdots \sum_{j_n \in [d_n]} p_{i_1 i_2 \cdots i_r j_{r+1} j_{r+2} \cdots j_n}.$$

We write $D' = [d_1] \times \cdots \times [d_r]$ and $\mathbb{R}[D']$ for the polynomial subring of $\mathbb{R}[D]$ generated by the observable probabilities $p_{i_1 i_2 \cdots i_r ++\cdots+}$. Let $\pi : \mathbb{R}^D \to \mathbb{R}^{D'}$ denote the canonical linear epimorphism induced by the inclusion of $\mathbb{R}[D']$ in $\mathbb{R}[D]$. We are interested in the following inclusions of semi-algebraic sets:

$$\pi(V_{\geq 0}(P_G)) \subset \pi(V(P_G))_{\geq 0} \subset \pi(V(P_G)) \subset \overline{\pi(V(P_G))} \subset \mathbb{R}^{D'}. \quad (13)$$

These inclusions are generally all strict. In particular, the space $\pi(V_{\geq 0}(P_G))$ which consists of all *observable probability distributions* is often much smaller than the space $\pi(V(P_G))_{\geq 0}$ which consists of probability distributions on D' which would be observable if non-negative or complex numbers were allowed for the hidden parameters. However, they have the same Zariski closure:

Proposition 19. *The set of all polynomial functions which vanish on the space $\pi(V_{\geq 0}(P_G))$ of observable probability distributions is the prime ideal*

$$Q_G \quad = \quad P_G \cap \mathbb{R}[D']. \tag{14}$$

Proof. The elimination ideal $Q_G \subset \mathbb{R}[D']$ is prime because $P_G \subset \mathbb{R}[D]$ was a prime ideal. By the Closure Theorem of Elimination Theory [4, Theorem 3, §3.2], the ideal Q_G is the vanishing ideal of the image $\pi(V(P_G))$. Since $V_{\geq 0}(P_G)$ is Zariski dense in $V(P_G)$, by the Factorization Theorem 3, and π is a linear map, it follows that $\pi(V_{\geq 0}(P_G))$ is Zariski dense in $\pi(V(P_G))$. $\quad\square$

We wish to demonstrate how computational algebraic geometry can be used to study hidden random variables in Bayesian networks. To this end we apply the concepts introduced above to a standard example from the statistics literature [7], [17], [18]. We fix the network G which has $n+1$ random variables $F_1, \ldots, F_n, H$ and n directed edges (H, F_i), $i = 1, 2, \ldots, n$. This is the *naive Bayes model*. The variable H is the hidden variable, and its levels $1, 2, \ldots, d_{n+1} =: r$ are called the *classes*. The observed random variables $F_1, \ldots, F_n$ are the *features* of the model. In this example, the prime ideal P_G coincides with the local ideal $I_{\text{local}(G)}$ which is specified by requiring that, for each fixed class, the features are completely independent:

$$F_1 \perp\!\!\!\perp F_2 \perp\!\!\!\perp \cdots \perp\!\!\!\perp F_n \,|\, H.$$

This ideal is obtained as the kernel of the map $p_{i_1 i_2 \cdots i_n k} \mapsto x_{i_1} y_{i_2} \cdots z_{i_n}$, one copy for each fixed class k, and then adding up these r prime ideals. Equivalently, P_G is the ideal of the join of r copies of the *Segre variety*

$$X_{d_1, d_2, \ldots, d_n} \quad := \quad \mathbb{P}^{d_1-1} \times \mathbb{P}^{d_2-1} \times \cdots \times \mathbb{P}^{d_n-1} \quad \subset \quad \mathbb{P}^{d_1 d_2 \cdots d_n - 1}. \tag{15}$$

The points on $X_{d_1, d_2, \ldots, d_n}$ represent tensors of rank ≤ 1. Our linear map π takes an r-tuple of tensors of rank ≤ 1 and it computes their sum, which is a tensor of rank $\leq r$. The closure of the image of π is what is called a *higher secant variety* in the language of algebraic geometry [12, Example 11.30].

Corollary 20. *The naive Bayes model with r classes and n features corresponds to the r-th secant variety of a Segre product of n projective spaces:*

$$\overline{\pi(V(P_G))} \quad = \quad \text{Sec}^r(X_{d_1, d_2, \ldots, d_n}) \tag{16}$$

The case $n = 2$ of two features is a staple of classical projective geometry. In that special case, the image of π is closed, and $\pi(V(P_G)) = \text{Sec}^r(X_{d_1, d_2})$

consists of all real $d_1 \times d_2$-matrices of rank at most r. This variety has codimension $(d_1 - r)(d_2 - r)$, provided $r \le \min(d_1, d_2)$. Its ideal Q_G is generated by the $(r+1) \times (r+1)$-minors of the $d_1 \times d_2$ matrix (p_{ij+}). The dimension formula of Settimi and Smith [18, Theorem 1] follows immediately. For instance, in the case of two ternary features $(d_1 = d_2 = 3, r = 2)$, discussed in different guises in [18, §4.2] and [12, Example 11.26], the *observable space* is the cubic hypersurface defined by the 3×3-determinant $\det(p_{ij+})$.

The leftmost inclusion in (13) leads to difficult open problems even for $n = 2$ features. Here, $\pi(V(P_G))_{\ge 0}$ is the set of all non-negative $d_1 \times d_2$-matrices of rank at most r, while $\pi(V_{\ge 0}(P_G))$ is the subset consisting of all matrices of *non-negative rank* at most r. Their difference consists of non-negative matrices of rank $\le r$ which cannot be written as the sum of r non-negative matrices of rank 1. In spite of recent progress by Barradas and Solis [1], there is still no practical algorithm for computing the non-negative rank of a $d_1 \times d_2$-matrix. Things get even harder for $n \ge 3$, when testing membership in $\pi(V_{\ge 0}(P_G))$ means computing *non-negative tensor rank*.

We next discuss what is known about the case of $n \ge 3$ features. The *expected dimension* of the secant variety (16) is

$$r \cdot (d_1 + d_2 + \cdots + d_n - n + 1) \; - \; 1. \tag{17}$$

This number is always an upper bound, and it is an interesting problem, studied in the statistics literature in [7], to characterize those cases $(d_1, \ldots, d_n; r)$ when the dimension is less than the expected dimension. We note that the results on dimension in [7] are all special cases of results by Catalisano, Geramita and Gimigliano [3], and the results on singularities in [7] follow from the geometric fact that the r-th secant variety of any projective variety is always singular along the $(r-1)$-st secant variety. The statistical problem of *identifiability*, addressed in [17], is related to the beautiful work of Strassen [20] on tensor rank, notably his Theorem 2.7 on *optimal computations*.

In Table 3 we display the range of straightforward `Macaulay2` computations when $dim(X) = d_1 + \cdots + d_n - 1$ is small. First consider the case of two classes $(r = 2)$, which corresponds to secant lines on $X = \mathbb{P}^{d_1 - 1} \times \cdots \times \mathbb{P}^{d_n - 1}$. In each of these cases, the ideal Q_G is generated by cubic polynomials, and each of these cubic generators is the determinant of a two-dimensional matrix obtained by flattening the tensor $(p_{i_1 i_2 \cdots i_n})$. The column labeled "cubics" lists the number of minimal generators. For example, in the case $(d_1 = d_2 = d_3 = 3)$, we can flatten (p_{ijk}) in three possible ways to a 3×9-matrix, and these have $3 \cdot \binom{9}{3} = 252$ maximal subdeterminants. The vector

$dim(X)$	$dim(\mathrm{Sec}^2(X))$	$\prod_{i=1} d_i$	$(d_1,\ldots,d_n)$	$degree$	$cubics$
4	9	12	$(2,2,3)$	6	4
4	9	16	$(2,2,2,2)$	64	32
5	11	16	$(2,2,4)$	20	16
5	11	18	$(2,3,3)$	57	36
5	11	24	$(2,2,2,3)$	526	184
5	11	32	$(2,2,2,2,2)$	3256	768
6	13	20	$(2,2,5)$	50	40
6	13	24	$(2,3,4)$	276	120
6	13	27	$(3,3,3)$	783	222
6	13	32	$(2,2,2,4)$	2388	544
6	13	36	$(2,2,3,3)$	6144	932

Table 3: The prime ideal defining the secant lines to the Segre variety (15)

space spanned by these subdeterminants has dimension 222, the listed number of minimal generators. The column "degree" lists the degree of the projective variety $\mathrm{Sec}^2(X)$, which is 783 in the previous example. These computational results in Table 3 lead us to make the following conjecture:

Conjecture 21. *The prime ideal Q_G of any naive Bayes model G with $r = 2$ classes is generated by the 3×3-subdeterminants of any two-dimensional table obtained by flattening the n-dimensional table $(p_{i_1 i_2 \cdots i_n})$.*

It was proved by Catalisano, Geramita and Gimigliano that the variety $\mathrm{Sec}^2(X)$ always has the expected dimension (17) when $r = 2$. A well-known example (see [9, page 221]) when the dimension is less than expected occurs for four classes and three binary features ($r = 3, n = 4, d_1 = d_2 = d_3 = d_4 = 2$). Here (17) evaluates to 14, but $\dim(\mathrm{Sec}^3(X)) = 13$ for $X = \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. The corresponding ideal Q_G is a complete intersection generated by any two of the three 4×4-determinants obtained by flattening the $2 \times 2 \times 2 \times 2$-table (p_{ijkl}). The third is a signed sum of the other two.

The problem of identifying explicit generators of Q_G is much more difficult when $r \geq 3$, i.e., when the hidden variable has three or more levels. We present the complete solution for the case of three ternary features. Here (p_{ijk}) is an indeterminate $3 \times 3 \times 3$-tensor which we wish to write as a sum of r rank one tensors. The following solution is derived from a result of Strassen [20, Theorem 4.6]. Let $A = (p_{ij1})$, $B = (p_{ij2})$ and $C = (p_{ij3})$ be three 3×3-matrices obtained by taking slices of the $3 \times 3 \times 3$-table (p_{ijk}).

Proposition 22. *Let Q_G be the ideal of $\mathrm{Sec}^r(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)$, the naive Bayes model with $n = 3$ ternary features with r classes. If $r = 2$ then Q_G is generated by the cubics described in Conjecture 21. If $r = 3$ then Q_G is generated by the quartic entries of the various 3×3-matrices of the form $A \cdot \mathrm{adj}(B) \cdot C - C \cdot \mathrm{adj}(B) \cdot A$. If $r = 4$ then Q_G is the principal ideal generated by the following homogeneous polynomial of degree 9 with $9,216$ terms:*

$$\det(B)^2 \cdot \det\big(A \cdot B^{-1} \cdot C - C \cdot B^{-1} \cdot B\big).$$

If $r \geq 5$ then Q_G is the zero ideal.

References

[1] I. Barradas and F. Solis: Nonnegative rank, *International Journal of Mathematics* **1** (2002) 601–610.

[2] M. Catalano-Johnson: The homogeneous ideals of higher secant varieties, *Journal of Pure and Applied Algebra* **158** (2001) 123–129.

[3] M.V. Catalisano, A.V. Geramita and A. Gimigliano: Ranks of tensors, secant varieties and fat points, *Linear Algebra Appl.*, to appear.

[4] D. Cox, J. Little and D. O'Shea: *Ideals, Varieties and Algorithms*, Springer Undergraduate Texts in Mathematics, Second Edition, 1997.

[5] W. Decker, G.-M. Greuel and G. Pfister: Primary decomposition: algorithms and comparisons. *Algorithmic Algebra and Number Theory (Heidelberg, 1997)*, 187–220, Springer, Berlin, 1999

[6] D. Eisenbud and B. Sturmfels: Binomial ideals, *Duke Math. Journal* **84** (1996) 1–45.

[7] D. Geiger, D. Heckerman, H. King and C. Meek: Stratified exponential families: graphical models and model selection, *Annals of Statistics* **29** (2001) 505–529.

[8] D. Geiger, C. Meek and B. Sturmfels: On the toric algebra of graphical models, Manuscript, 2002.

[9] L. Goodman: Explanatory latent structure analysis using both identifiable and unidentifiable models, *Biometrika* **61** (1974) 215–231.

[10] D. Grayson and M. Stillman: Macaulay2: a system for computation in algebraic geometry and commutative algebra, 1996, available at `http://www.math.uiuc.edu/Macaulay2/`.

[11] G.-M. Greuel, G. Pfister and H. Schönemann: Singular 2.0: A computer algebra system for polynomial computations, University of Kaiserslautern, 2001, `http://www.singular.uni-kl.de`.

[12] J. Harris: *Algebraic Geometry: A First Course*, Springer Graduate Texts in Mathematics, 1992.

[13] S. L. Lauritzen: *Graphical Models*, Oxford University Press, 1996.

[14] F. Matúš: Conditional independences among four random variables. III. Final conclusion. *Combin. Probab. Comput.* **8** (1999) 269–276.

[15] J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, 1988

[16] G. Pistone, E. Riccomagno and H. Wynn: *Algebraic Statistics: Computational Commutative Algebra in Statistics*, Chapman and Hall, Boca Raton, 2001.

[17] R. Settimi and J.Q. Smith: Geometry, moments and conditional independence trees with hidden variables, *Annals of Statistics* **28** (2000) 1179–1205.

[18] R. Settimi and J.Q. Smith: On the geometry of Bayesian graphical models with hidden variables, In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, San Francisco, CA, 1998, pp. 472-479.

[19] T. Shimoyama and K. Yokoyama: Localization and primary decomposition of polynomial ideals. *J. Symbolic Comput.* **22** (1996) 247–277.

[20] V. Strassen: Rank and optimal computation of generic tensors, *Linear Algebra and its Applications* **52/53** (1983) 645–685.

[21] M. Studený: *On Mathematical Description of Probabilistic Conditional Independence Structures*, Dr.Sc. thesis, Prague, May 2001, posted at `http://www.utia.cas.cz/user_data/studeny/studeny_home.html`.

[22] B. Sturmfels: *Gröbner Bases and Convex Polytopes*, American Mathematical Society, University Lectures Series, No. 8, Providence, Rhode Island, 1996.

[23] B. Sturmfels, *Solving Systems of Polynomial Equations*, CBMS Lectures Series, American Mathematical Society, 2002.

Appendix: Techniques for Primary Decomposition

The ideals in this paper present a challenge for present day computer algebra systems. Their large number of variables (e.g. 32 in Section 6), combined with the sometimes long polynomials which arise are difficult to handle with built-in primary decomposition algorithms. Even the standard implementations of factorization of multivariate polynomials have difficulty with some of the long polynomials. This is only a problem with current implementations, which are generally not optimized for large numbers of variables.

For the computations performed in Sections 5 and 6, it was necessary to write special code (in `Macaulay2`) in order to compute the components and primary decompositions of these ideals. We also have some code in `Macaulay2` or `Singular` for generating the ideals $I_{\text{local}(G)}$ or $I_{\text{global}(G)}$ from the graph G and the integers $d_1, d_2, \ldots, d_n$. In this appendix we indicate some techniques and tricks that were used to compute with these ideals.

The first modification which simplifies the problems dramatically is to change coordinates so that the indeterminates are $p_{2u_2 \cdots u_n}$ and $p_{+u_2 \cdots u_n}$, instead of $p_{u_1 \cdots u_n}$. This change of variables sometimes takes a Markov ideal into a binomial ideal, which is generally much simpler to compute with. Computing any one Gröbner basis, ideal quotient, or intersection of our ideals is not too difficult. Therefore, our algorithms make use of these operations. All ideals examined in this project have the property that every component is rational. The distinguished component $\ker(\Phi)$ is more complicated than any of the other components, in terms of the number of generators and their degrees, and it cannot be computed by implicitization.

The first problem is to decide whether an ideal is prime (i.e. whether it equals the unknown ideal $\ker(\Phi)$). There are several known methods for deciding primality (see [5] for a nice exposition). The standard method is to reduce to a zero-dimensional problem. This entails either a generic change of coordinates, or factorization over extension fields. We found that the current implementations of these methods fail for the majority of the 301 examples in Section 6. The technique that did work for us is to search for birational projections. This either produces a zero divisor, or a proof that the ideal is prime. It can sometimes be used to count the components (both minimal and embedded), without actually producing the components.

28

The following result is proved by localizing with respect to powers of g. This defines a *birational projection* $(x_1, x_2, \ldots, x_n) \mapsto (x_2, \ldots, x_n)$ for J.

Proposition 23. *Let $J \subset \mathbb{R}[x_1, \ldots, x_n]$ be an ideal, containing a polynomial $f = gx_1 + h$, with g, h not involving x_1, and g a non-zero divisor modulo J. Let $J_1 = J \cap \mathbb{R}[x_2, \ldots, x_n]$ be the elimination ideal. Then*

(a) $J = \left(\langle J_1, gx_1 + h \rangle : g^\infty \right)$,

(b) J is prime if and only if J_1 is prime.

(c) J is primary if and only if J_1 is primary.

(d) Any irredundant primary decomposition of J_1 lifts to an irredundant primary decomposition of J.

Our algorithm to check primality starts by searching for variables which occur linearly, checking that its lead coefficient is not a zero divisor and then eliminating that variable as in Proposition 23. In almost all of the Markov ideals that we have studied, iterative use of this technique proves or disproves primality. A priori, one might not be able to find a birational projection at all, but this never happened for any of our examples.

The second problem is to compute the minimal primes or the primary decomposition. Finding the minimal primes is the first step in computing a primary decomposition, using the technique of [19], which is implemented in several computer algebra systems, including `Macaulay2`. Here, we have not found a single method that always works best. One method that worked in most cases is based on splitting the ideal into two parts. Given an ideal I, if there is an element f of its Gröbner basis which factors as $f = f_1 f_2$, then

$$\sqrt{I} \;=\; \sqrt{\langle I, f_1 \rangle} \,\cap\, \sqrt{\langle I, f_2 \rangle : f_1^\infty}.$$

We keep a list of ideals whose intersection has the same radical as I. We process this list of ideals by ascending order on its codimension. For each ideal, we keep a list of the elements that we have inverted by so far (e.g. f_1 in the ideal $\left(\langle I, f_2 \rangle : f_1^\infty \right)$) and saturate at each step with these elements.

If there is no element which factors, then we search for a variable to birationally project away from, as in Proposition 23. If its lead coefficient g is a zero divisor, use this element to split the ideal via

$$\sqrt{I} \;=\; \sqrt{I : g} \,\cap\, \sqrt{\langle I, g \rangle}.$$

As we go, we only process ideals which do not contain the intersection of all known components computed so far.

If we cannot find any birational projection or reducible polynomial, then we have no choice but to decompose the ideal using the built-in routines,

which are based on characteristic sets. However, in none of the examples of this paper was this final step reached. This method works in a reasonable amount of time for all but about 10 to 15 of the 301 ideals in Section 6.

Luis David Garcia, Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, VA 24061, USA, `lgarcia@vt.edu`.

Michael Stillman, Department of Mathematics, Cornell University, Ithaca, NY 14853-4201, USA, `mike@math.cornell.edu`.

Bernd Sturmfels, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA, `bernd@math.berkeley.edu`.